Analytical Methods in Physics

Yi-Zen Chu

Contents

1 Preface

This work constitutes the free textbook project I initiated towards the end of Summer 2015, while preparing for the Fall 2015 *Analytical Methods in Physics* course I taught to upper level (mostly 2nd and 3rd year) undergraduates here at the University of Minnesota Duluth. During Fall 2017, I taught the graduate-level *Differential Geometry and Physics in Curved Spacetimes* here at National Central University, Taiwan; this has allowed me to further expand the text.

I assumed that the reader has taken the first three semesters of calculus, i.e., up to multi-variable calculus, as well as a first course in Linear Algebra and ordinary differential equations. (These are typical prerequisites for the Physics major within the US college curriculum.) My primary goal was to impart a good working knowledge of the mathematical tools that underlie fundamental physics – quantum mechanics and electromagnetism, in particular. This meant that Linear Algebra in its abstract formulation had to take a central role in these notes.[1] To this end, I first reviewed complex numbers and matrix algebra. The middle chapters cover calculus beyond the first three semesters: complex analysis and special/approximation/asymptotic methods. The latter, I feel, is not taught widely enough in the undergraduate setting. The final chapter is meant to give a solid introduction to the topic of linear partial differential equations (PDEs), which is crucial to the study of electromagnetism, linearized gravitation and quantum mechanics/field theory. But before tackling PDEs, I feel that having a good grounding in the basic elements of differential geometry not only helps streamlines one's fluency in multi-variable calculus; it also provides a stepping stone to the discussion of curved spacetime wave equations.

Some of the other distinctive features of this free textbook project are as follows.

Index notation and Einstein summation convention is widely used throughout the physics literature, so I have not shied away from introducing it early on, starting in §(3) on matrix algebra. In a similar spirit, I have phrased the abstract formulation of Linear Algebra in §(4) entirely in terms of P.A.M. Dirac's bra-ket notation. When discussing inner products, I do make a brief comparison of Dirac's notation against the one commonly found in math textbooks.

I made no pretense at making the material mathematically rigorous, but I strived to make the flow coherent, so that the reader comes away with a firm conceptual grasp of the overall structure of each major topic. For instance, while the full fledged study of continuous (as opposed to discrete) vector spaces can take up a whole math class of its own, I feel the physicist should be exposed to it right after learning the discrete case. For, the basics are not only accessible, the Fourier transform is in fact a physically important application of the continuous space spanned by the position eigenkets $\{|\vec{x}\rangle\}$. One key difference between Hermitian operators in discrete versus continuous vector spaces is the need to impose appropriate boundary conditions in the latter; this is highlighted in the Linear Algebra chapter as a prelude to the PDE chapter §(12), where the Laplacian and its spectrum plays a significant role. Additionally, while the Linear Algebra chapter was heavily inspired by the first chapter of Sakurai's *Modern Quantum Mechanics*, I have taken effort to emphasize that quantum mechanics is merely a very important application of the framework; for e.g., even the famous commutation relation $[X^i, P_j] = i\delta^i_j$ is not necessarily a quantum mechanical statement. This emphasis is based on the belief that the power of a given

[1] That the textbook originally assigned for this course relegated the axioms of Linear Algebra towards the very end of the discussion was one major reason why I decided to write these notes. This same book also cost nearly two hundred (US) dollars – a fine example of exorbitant textbook prices these days – so I am glad I saved my students quite a bit of their educational expenses that semester.

mathematical tool is very much tied to its versatility – this issue arises again in the JWKB discussion within §(7), where I highlight it is not merely some "semi-classical" limit of quantum mechanical problems, but really a general technique for solving differential equations.

Much of §(6) is a standard introduction to calculus on the complex plane and the theory of complex analytic functions. However, the Fourier transform application section gave me the chance to introduce the concept of the Green's function; specifically, that of the ordinary differential equation describing the damped harmonic oscillator. This (retarded) Green's function can be computed via the theory of residues – and through its key role in the initial value formulation of the ODE solution, allows the two linearly independent solutions to the associated homogeneous equation to be obtained for any value of the damping parameter.

Differential geometry may appear to be an advanced topic to many, but it really is not. From a practical standpoint, it cannot be overemphasized that most vector calculus operations can be readily carried out and the curved space(time) Laplacian/wave operator computed once the relevant metric is specified explicitly. I wrote much of §(9) in this "practical physicist" spirit. Although it deals primarily with curved spaces, teaching *Physics in Curved Spacetimes* during Fall 2017 at National Central University, Taiwan, gave me the opportunity to add its curved spacetime sequel, §(11), where I elaborated upon geometric concepts – the emergence of the Riemann tensor from parallel transporting a vector around an infinitesimal parallelogram, for instance – deliberately glossed over in §(9). It is my hope that §(9) and §(11) can be used to build the differential geometric tools one could then employ to understand General Relativity, Einstein's field equations for gravitation.

In §(12) on PDEs, I begin with the Poisson equation in curved space, followed by the enumeration of the eigensystem of the Laplacian in different flat spaces. By imposing Dirichlet or periodic boundary conditions for the most part, I view the development there as the culmination of the Linear Algebra of continuous spaces. The spectrum of the Laplacian also finds important applications in the solution of the heat and wave equations. I have deliberately discussed the heat instead of the Schrödinger equation because the two are similar enough, I hope when the reader learns about the latter in her/his quantum mechanics course, it will only serve to enrich her/his understanding when she/he compares it with the discourse here. Finally, the wave equation in Minkowski spacetime – *the* basis of electromagnetism and linearized gravitation – is discussed from both the position/real and Fourier/reciprocal space perspectives. The retarded Green's function plays a central role here, and I spend significant effort exploring different means of computing it. The tail effect is also highlighted there: classical waves associated with massless particles transmit physical information *within* the null cone in $(1+1)$D and all odd dimensions. Wave solutions are examined from different perspectives: in real/position space; in frequency space; in the non-relativistic/static limits; and with the multipole-expansion employed to extract leading order features. The final section contains a brief introduction to the variational principle for the classical field theories of the Poisson and wave equations.

Finally, I have interspersed problems throughout each chapter because this is how I personally like to engage with new material – read and "doodle" along the way, to make sure I am properly following the details. My hope is that these notes are concise but accessible enough that anyone can work through both the main text as well as the problems along the way; and discover they have indeed acquired a new set of mathematical tools to tackle physical problems.

By making this material available online, I view it as an ongoing project: I plan to update and add new material whenever time permits; for instance, illustrations/figures accompanying the main text may eventually show up at some point down the road. The most updated version can be found at the following URL:

http://www.stargazing.net/yizen/AnalyticalMethods_YZChu.pdf

I would very much welcome suggestions, questions, comments, error reports, etc.; please feel free to contact me at yizen [dot] chu @ gmail [dot] com.

– Yi-Zen Chu

2 Complex Numbers and Functions

[2]The motivational introduction to complex numbers, in particular the number i,[3] is the solution to the equation

$$i^2 = -1. \tag{2.0.1}$$

That is, "what's the square root of -1?" For us, we will simply take eq. (2.0.1) as the *defining* equation for the algebra obeyed by i. A general complex number z can then be expressed as

$$z = x + iy \tag{2.0.2}$$

where x and y are real numbers. The x is called the *real part* ($\equiv \mathrm{Re}(z)$) and y the *imaginary part* of z ($\equiv \mathrm{Im}(z)$).

Geometrically speaking z is a vector (x, y) on the 2-dimensional plane spanned by the real axis (the x part of z) and the imaginary axis (the iy part of z). Moreover, you may recall from (perhaps) multi-variable calculus, that if r is the distance between the origin and the point (x, y) and ϕ is the angle between the vector joining $(0, 0)$ to (x, y) and the positive horizontal axis – then

$$(x, y) = (r\cos\phi, r\sin\phi). \tag{2.0.3}$$

Therefore a complex number must be expressible as

$$z = x + iy = r(\cos\phi + i\sin\phi). \tag{2.0.4}$$

This actually takes a compact form using the exponential:

$$z = x + iy = r(\cos\phi + i\sin\phi) = re^{i\phi}, \qquad r \geq 0,\ 0 \leq \phi < 2\pi. \tag{2.0.5}$$

Some words on notation. The distance r between $(0, 0)$ and (x, y) in the complex number context is written as an absolute value, i.e.,

$$|z| = |x + iy| = r = \sqrt{x^2 + y^2}, \tag{2.0.6}$$

where the final equality follows from Pythagoras' Theorem. The angle ϕ is denoted as

$$\arg(z) = \arg(re^{i\phi}) = \phi. \tag{2.0.7}$$

The symbol $\mathbb{C}$ is often used to represent the 2D space of complex numbers.

$$z = |z|e^{i\arg(z)} \in \mathbb{C}. \tag{2.0.8}$$

Problem 2.1. Euler's formula. Assuming $\exp z$ can be defined through its Taylor series for any complex z, prove by Taylor expansion and eq. (2.0.1) that

$$e^{i\phi} = \cos(\phi) + i\sin(\phi), \qquad \phi \in \mathbb{R}. \tag{2.0.9}$$

[2]Some of the material in this section is based on James Nearing's *Mathematical Tools for Physics*.

[3]Engineers use j instead of i.

Arithmetic Addition and subtraction of complex numbers take place component-by-component, just like adding/subtracting 2D real vectors; for example, if

$$z_1 = x_1 + iy_1 \qquad \text{and} \qquad z_2 = x_2 + iy_2, \tag{2.0.10}$$

then

$$z_1 \pm z_2 = (x_1 \pm x_2) + i(y_1 \pm y_2). \tag{2.0.11}$$

Multiplication is more easily done in polar coordinates: if $z_1 = r_1 e^{i\phi_1}$ and $z_2 = r_2 e^{i\phi_2}$, their product amounts to adding their phases and multiplying their radii, namely

$$z_1 z_2 = r_1 r_2 e^{i(\phi_1+\phi_2)}. \tag{2.0.12}$$

To summarize:

> Complex numbers $\{z = x+iy = re^{i\phi} | x, y \in \mathbb{R}; r \geq 0, \phi \in \mathbb{R}\}$ are 2D real vectors as far as addition/subtraction goes – Cartesian coordinates are useful here (cf. (2.0.11)). It is their multiplication that the additional ingredient/algebra $i^2 \equiv -1$ comes into play. In particular, using polar coordinates to multiply two complex numbers (cf. (2.0.12)) allows us to see the result is a combination of a re-scaling of their radii plus a rotation.

Problem 2.2. If $z = x + iy$ what is z^2 in terms of x and y? □

Problem 2.3. Explain why multiplying a complex number $z = x + iy$ by i amounts to rotating the vector (x, y) on the complex plane counter-clockwise by $\pi/2$. Hint: first write i in polar coordinates. □

Problem 2.4. Describe the points on the complex z-plane satisfying $|z - z_0| < R$, where z_0 is some fixed complex number and $R > 0$ is a real number.

Problem 2.5. Use the polar form of the complex number to proof that multiplication of complex numbers is associative, i.e., $z_1 z_2 z_3 = z_1(z_2 z_3) = (z_1 z_2) z_3$. □

Problem 2.6. Explain why, for real a and b,

$$|a^{ib}| = 1. \tag{2.0.13}$$

Hint: $a = \exp \ln a$. □

Problem 2.7. Multiplication and Vector Calculus If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, show that

$$z_1^* z_2 = \vec{z}_1 \cdot \vec{z}_2 + i \left(\begin{bmatrix} \vec{z}_1 \\ 0 \end{bmatrix} \times \begin{bmatrix} \vec{z}_2 \\ 0 \end{bmatrix} \right) \cdot \widehat{e}_3. \tag{2.0.14}$$

Here, we have converted the complex numbers into 3D vectors via $\vec{z}_1 \equiv (x_1, y_1)^{\mathrm{T}}$ and $\vec{z}_2 \equiv (x_2, y_2)^{\mathrm{T}}$; whereas $\widehat{e}_3 \equiv (0, 0, 1)^{\mathrm{T}}$. □

Complex conjugation Taking the complex conjugate of $z = x + iy$ means we flip the sign of its imaginary part, i.e.,

$$z^* = x - iy; \tag{2.0.15}$$

it is also denoted as $\bar{z}$. In polar coordinates, if $z = re^{i\phi} = r(\cos\phi + i\sin\phi)$ then $z^* = re^{-i\phi}$ because

$$e^{-i\phi} = \cos(-\phi) + i\sin(-\phi) = \cos\phi - i\sin\phi. \tag{2.0.16}$$

The $\sin\phi \to -\sin\phi$ is what brings us from $x + iy$ to $x - iy$. Now

$$z^* z = zz^* = (x + iy)(x - iy) = x^2 + y^2 = |z|^2. \tag{2.0.17}$$

When we take the ratio of complex numbers, it is possible to ensure that the imaginary number i appears only in the numerator, by multiplying the numerator and denominator by the complex conjugate of the denominator. For x, y, a and b all real,

$$\frac{x + iy}{a + ib} = \frac{(a - ib)(x + iy)}{a^2 + b^2} = \frac{(ax + by) + i(ay - bx)}{a^2 + b^2}. \tag{2.0.18}$$

Problem 2.8. Is $(z_1 z_2)^* = z_1^* z_2^*$, i.e., is the complex conjugate of the product of 2 complex numbers equal to the product of their complex conjugates? What about $(z_1/z_2)^* = z_1^*/z_2^*$? Is $|z_1 z_2| = |z_1||z_2|$? What about $|z_1/z_2| = |z_1|/|z_2|$? Also show that $\arg(z_1 \cdot z_2) = \arg(z_1) + \arg(z_2)$. Strictly speaking, $\arg(z)$ is well defined only up to an additive multiple of 2π. Can you explain why? Hint: polar coordinates are very useful in this problem. □

Problem 2.9. Show that z is real if and only if $z = z^*$. Show that z is purely imaginary if and only if $z = -z^*$. Show that $z + z^* = 2\mathrm{Re}(z)$ and $z - z^* = 2i\mathrm{Im}(z)$. Hint: use Cartesian coordinates. □

Problem 2.10. Roots come in complex conjugate pairs Prove that the roots of a polynomial with real coefficients

$$P_N(z) \equiv c_0 + c_1 z + c_2 z^2 + \dots + c_N z^N, \qquad \{c_i \in \mathbb{R}\}, \tag{2.0.19}$$

come in complex conjugate pairs; i.e., if z is a root then so is z^*. Hint: If $P_N(z) = 0$, consider its complex conjugate. □

Trigonometric, Hyperbolic and Exponential functions Complex numbers allow us to connect trigonometric, hyperbolic and exponential (exp) functions. Start from

$$e^{\pm i\phi} = \cos\phi \pm i\sin\phi. \tag{2.0.20}$$

These two equations can be added and subtracted to yield

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin(z) = \frac{e^{iz} - e^{-iz}}{2i}, \qquad \tan(z) = \frac{\sin(z)}{\cos(z)}. \tag{2.0.21}$$

We have made the replacement $\phi \to z$. This change is cosmetic if $0 \leq z < 2\pi$, but we can in fact now use eq. (2.0.21) to *define* the trigonometric functions in terms of the exp function for any complex z. This $\exp z$, for $z = x + iy$, is exponentially dominant (suppressed) in magnitude for large positive (negative) $x \equiv \mathrm{Re}(z)$; and periodic along the $y \equiv \mathrm{Im}(z)$ direction; because

$$\exp z = e^{x+iy} = (\exp x)(\exp iy). \tag{2.0.22}$$

(Compare this form of the exponential with eq. (2.0.8).) Trigonometric identities can be readily obtained from their exponential definitions. For example, the addition formulas would now begin from

$$e^{i(\theta_1+\theta_2)} = e^{i\theta_1} e^{i\theta_2}. \tag{2.0.23}$$

Applying Euler's formula (eq. (2.0.9)) on both sides,

$$\begin{aligned} \cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2) &= (\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1). \end{aligned} \tag{2.0.24}$$

If we suppose $\theta_{1,2}$ are real angles, equating the real and imaginary parts of the left-hand-side and the last line tell us

$$\cos(\theta_1 + \theta_2) = \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2, \tag{2.0.25}$$
$$\sin(\theta_1 + \theta_2) = \sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1. \tag{2.0.26}$$

Problem 2.11. You are probably familiar with the hyperbolic functions, now defined as

$$\cosh(z) = \frac{e^z + e^{-z}}{2}, \qquad \sinh(z) = \frac{e^z - e^{-z}}{2}, \qquad \tanh(z) = \frac{\sinh(z)}{\cosh(z)}, \tag{2.0.27}$$

for any complex z. Show that

$$\cosh(iz) = \cos(z), \qquad \sinh(iz) = i\sin(z), \tag{2.0.28}$$
$$\cos(iz) = \cosh(z), \qquad \sin(iz) = i\sinh(z). \tag{2.0.29}$$

These relations tell us, the trigonometric and hyperbolic functions are really connected to each other via a $\pi/2$ rotation on the complex $z-$plane; i.e., $z \to iz$. $\square$

Problem 2.12. Calculate, for real θ and positive integer N:

$$\cos(\theta) + \cos(2\theta) + \cos(3\theta) + \cdots + \cos(N\theta) = ? \tag{2.0.30}$$
$$\sin(\theta) + \sin(2\theta) + \sin(3\theta) + \cdots + \sin(N\theta) = ? \tag{2.0.31}$$

Hint: consider the geometric series $e^{i\theta} + e^{2i\theta} + \cdots + e^{Ni\theta}$. $\square$

Problem 2.13. Starting from $(e^{i\theta})^n$, for arbitrary integer n, re-write $\cos(n\theta)$ and $\sin(n\theta)$ as a sum involving products/powers of $\sin\theta$ and $\cos\theta$. Hint: if the arbitrary n case is confusing at first, start with $n = 1, 2, 3$ first. $\square$

Roots of unity In polar coordinates, circling the origin n times bring us back to the same point,

$$z = re^{i\theta + i2\pi n}, \qquad n = 0, \pm 1, \pm 2, \pm 3, \dots . \tag{2.0.32}$$

This observation is useful for the following problem: what is mth root of 1, when m is a positive integer? Of course, 1 is an answer, but so are

$$1^{1/m} = e^{i2\pi n/m}, \qquad n = 0, 1, \dots, m-1. \tag{2.0.33}$$

The terms repeat themselves for $n \geq m$; the negative integers n do not give new solutions for m integer. If we replace $1/m$ with a/b where a and b are integers that do not share any common factors, then

$$1^{a/b} = e^{i2\pi n(a/b)} \qquad \text{for} \qquad n = 0, 1, \dots, b-1, \tag{2.0.34}$$

since when $n = b$ we will get back 1. If we replaced (a/b) with say $1/\pi$,

$$1^{1/\pi} = e^{i2\pi n/\pi} = e^{i2n}, \tag{2.0.35}$$

then there will be infinite number of solutions, because $1/\pi$ cannot be expressed as a ratio of integers – there is no way to get $2n = 2\pi n'$, for n' integer.

In general, when you are finding the mth root of a complex number z, you are actually solving for w in the polynomial equation $w^m = z$. The fundamental theorem of algebra tells us, if m is a positive integer, you are guaranteed m solutions – although not all of them may be distinct.

Square root of -1 What is $\sqrt{-1}$? Since $-1 = e^{i(\pi + 2\pi n)}$ for any integer n,

$$\left(e^{i(\pi+2\pi n)}\right)^{1/2} = e^{i\pi/2 + i\pi n} = \pm i. \qquad n = 0, 1. \tag{2.0.36}$$

Problem 2.14. Find all the solutions to $\sqrt{1-i}$. □

Logarithm and powers As we have just seen, whenever we take the root of some complex number z, we really have a multi-valued function. The inverse of the exponential is another such function. For $w = x + iy$, where x and y are real, we may consider

$$e^w = e^x e^{i(y+2\pi n)}, \qquad n = 0, \pm 1, \pm 2, \pm 3, \dots . \tag{2.0.37}$$

We define ln to be such that

$$\ln e^w = x + i(y + 2\pi n). \tag{2.0.38}$$

Another way of saying this is, for a general complex z,

$$\ln(z) = \ln|z| + i(\arg(z) + 2\pi n). \tag{2.0.39}$$

One way to make sense of how to raise a complex number $z = re^{i\theta}$ to the power of another complex number $w = x + iy$, namely z^w, is through the ln:

$$z^w = e^{w \ln z} = e^{(x+iy)(\ln(r) + i(\theta + 2\pi n))} = e^{x \ln r - y(\theta + 2\pi n)} e^{i(y \ln(r) + x(\theta + 2\pi n))}. \tag{2.0.40}$$

This is, of course, a multi-valued function. We will have more to say about such multi-valued functions when discussing their calculus in §(6).

Problem 2.15. Change-of-Base If $a > 0$ and $\theta \in \mathbb{R}$, explain why $|a^{i\theta}| = 1$. What is the change in θ swept out by a complete unit circle around the origin of the complex plane? Hint: $a = e^{\ln a}$. $\square$

Problem 2.16. Zeroes of trigonometric and hyperbolic functions Find the inverse hyperbolic functions of eq. (2.0.27) in terms of ln. Does $\sin(z) = 0$, $\cos(z) = 0$ and $\tan(z) = 0$ have any complex solutions? Hint: for the first question, write $e^z = w$ and $e^{-z} = 1/w$. Then solve for w. A similar strategy may be employed for the second question. $\square$

Taylor Expansion If $f(x)$ is a real function of the real variable x, and if all $n \geq 1$ derivatives of f exists at $x = x_0$ for some fixed x_0, then $f(x)$ in the neighborhood of $x \approx x_0$ may be approximated by a polynomial of degree N via the formula

$$f(x) \approx f(x_0) + \sum_{n=1}^{N} \frac{(x-x_0)^n}{n!} \frac{\mathrm{d}^n f(z=x_0)}{\mathrm{d}z^n}; \tag{2.0.41}$$

i.e., at zeroth order f is simply its value at $x = x_0$; at first order it is approximately a straight line passing through $(x_0, f(x_0))$ and tangent to the curve $(x, f(x))$; at second order it is a parabola; etc. The $N \to \infty$ limit provides an exact expression for $f(x)$ itself over the domain of x on the real line where the infinite series converges.

$$f(x) = f(x_0) + \sum_{n=1}^{\infty} \frac{(x-x_0)^n}{n!} \frac{\mathrm{d}^n f(z=x_0)}{\mathrm{d}z^n} \tag{2.0.42}$$

This infinite sum in eq. (2.0.42) is known as the *Taylor series* (or, Taylor expansion) of $f(x)$ about the point $x = x_0$; and is *unique* whenever it exists. The Taylor series of some commonly used functions are as follows. For those that converge on the entire real line, $x \in \mathbb{R}$,

$$\sin(x) = \sum_{n=1}^{\infty} \frac{(-)^{n+1} x^{2n-1}}{(2n-1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots, \tag{2.0.43}$$

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots, \tag{2.0.44}$$

$$e^x \equiv \exp x = \sum_{\ell=0}^{\infty} \frac{x^\ell}{\ell!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots; \tag{2.0.45}$$

and for $|x| < 1$,

$$\ln(1-x) = -\sum_{\ell=1}^{\infty} \frac{x^\ell}{\ell}. \tag{2.0.46}$$

Exponential Operator By comparison with eq. (2.0.45), notice the Taylor series in eq. (2.0.42) itself can be written as an exponential operator:

$$f(x+a) = f(x) + \sum_{n=1}^{\infty} \frac{a^n}{n!} \frac{\mathrm{d}^n f(x)}{\mathrm{d}x^n} \tag{2.0.47}$$

$$= \exp\left(a \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x). \tag{2.0.48}$$

In the linear algebra discussion below, we shall see that the derivative can be viewed as the 'generator' of translations; in this case, $x \to x + a$.

Problem 2.17. Multi-Variable Taylor Expansion Suppose the real function $F(\vec{x})$ depends on $D \geq 2$ real arguments $\vec{x} \equiv (x^1, x^2, x^3, \dots, x^D)$. (Note that x^i here, for $i \in \{1, 2, 3, \dots, D\}$, is not the ith power of x; but the ith coordinate in a given $D-$dimensional space.) Apply eq. (2.0.42) to each of argument of F to show that its Taylor series about $\vec{x} = \vec{x}_0$ is

$$F(\vec{x}) = F(\vec{x}_0) + \sum_{n=1}^{\infty} \frac{1}{n!}((\vec{x} - \vec{x}_0) \cdot \vec{\nabla}_{\vec{z}})^n F(\vec{z} = \vec{x}_0), \tag{2.0.49}$$

where the nth derivative is, more explicitly, given by

$$\begin{aligned} &((\vec{x} - \vec{x}_0) \cdot \vec{\nabla}_{\vec{z}})^n F(\vec{z} = \vec{x}_0) \\ &= \sum_{i_1, \dots, i_n = 1}^{D} (x^{i_1} - x_0^{i_1})(x^{i_2} - x_0^{i_2}) \dots (x^{i_n} - x_0^{i_n}) \partial_{z^{i_1}} \partial_{z^{i_2}} \dots \partial_{z^{i_n}} F(\vec{z} = \vec{x}_0). \end{aligned} \tag{2.0.50}$$

From this, further explain why

$$f(\vec{x} + \vec{a}) = \exp\left(\vec{a} \cdot \vec{\nabla}_{\vec{x}}\right) f(\vec{x}). \tag{2.0.51}$$

Hint: Start by performing induction on D. □

Problem 2.18. Let $\vec{\xi}$ and $\vec{\xi}'$ be vectors in a 2D Euclidean space, i.e., you may assume their Cartesian components are

$$\vec{\xi} = (x, y) = \rho(\cos\phi, \sin\phi), \qquad \vec{\xi}' = (x', y') = \rho'(\cos\phi', \sin\phi'). \tag{2.0.52}$$

Use complex numbers, and assume that the following complex version of eq. (2.0.46) holds; i.e., replacing $x \in \mathbb{R}$ with $z \in \mathbb{Z}$,

$$\ln(1 - z) = -\sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell}, \qquad |z| < 1; \tag{2.0.53}$$

to show that

$$\ln|\vec{\xi} - \vec{\xi}'| = \ln\rho_> - \sum_{\ell=1}^{\infty} \frac{1}{\ell}\left(\frac{\rho_<}{\rho_>}\right)^\ell \cos\left(\ell(\phi - \phi')\right), \tag{2.0.54}$$

where $\rho_>$ is the larger and $\rho_<$ is the smaller of the (ρ, ρ'), and $|\vec{\xi} - \vec{\xi}'|$ is the distance between the vectors $\vec{\xi}$ and $\vec{\xi}'$ – not the absolute value of some complex number. Here, $\ln|\vec{\xi} - \vec{\xi}'|$ is proportional to the electric or gravitational potential generated by a point charge/mass in 2-dimensional flat space. Hint: first let $z = \rho \cdot e^{i\phi}$ and $z' = \rho' e^{i\phi'}$; then consider $\ln(z - z')$ – how do you extract $\ln|\vec{\xi} - \vec{\xi}'|$ from it? □

3 Matrix Algebra

[4]In this section I will review some basic properties of matrices and matrix algebra, oftentimes using index notation. We will assume all matrices have complex entries unless otherwise stated. This is primarily intended to be warmup to the next section, where I will treat Linear Algebra from a more abstract point of view.

3.1 Basics, Matrix Operations, and Special Types of Matrices

Index notation, Einstein summation, Basic Matrix Operations Consider two matrices M and N. The ij component – the ith row and jth column of M and that of N can be written as

$$M^i{}_j \qquad \text{and} \qquad N^i{}_j. \tag{3.1.1}$$

As an example, if M is a 2×2 matrix, we have

$$M = \begin{bmatrix} M^1{}_1 & M^1{}_2 \\ M^2{}_1 & M^2{}_2 \end{bmatrix}. \tag{3.1.2}$$

I prefer to write one index up and one down, because as we shall see in the abstract formulation of linear algebra below, the row and column indices may transform 'oppositely'. However, it is common to see the notation M_{ij} and M^{ij}, etc., too.

A vector $\vec{v}$ can be written as

$$v^i = (v^1, v^2, \ldots, v^{D-1}, v^D). \tag{3.1.3}$$

Here, v^5 does not mean the fifth power of some quantity v, but rather the 5th component of the vector v.

The matrix multiplication $M \cdot N$ can be written as

$$(M \cdot N)^i{}_j = \sum_{k=1}^{D} M^i{}_k N^k{}_j \equiv M^i{}_k N^k{}_j. \tag{3.1.4}$$

In words: the ij component of the product MN, for a fixed i and fixed j, means we are taking the ith row of M and "dotting" it into the jth column of N. In the second equality we have employed Einstein's summation convention, which we will continue to do so in these notes: repeated indices are summed over their relevant range – in this case, $k \in \{1, 2, \ldots, D\}$. For example, if

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad N = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \tag{3.1.5}$$

then

$$M \cdot N = \begin{bmatrix} a + 3b & 2a + 4b \\ c + 3d & 2c + 4d \end{bmatrix}. \tag{3.1.6}$$

[4]Much of the material here in this section were based on Chapter 1 of Cahill's *Physical Mathematics*.

Note: $M^i{}_k N^k{}_j$ works for multiplication of non-square matrices M and N too, as long as the number of columns of M is equal to the number of rows of N, so that the sum involving k makes sense.

Addition of M and N; and multiplication of M by a complex number λ goes respectively as

$$(M+N)^i{}_j = M^i{}_j + N^i{}_j \tag{3.1.7}$$

and

$$(\lambda M)^i{}_j = \lambda M^i{}_j. \tag{3.1.8}$$

Associativity The associativity of matrix multiplication means $(AB)C = A(BC) = ABC$. This can be seen using index notation

$$A^i{}_k B^k{}_l C^l{}_j = (AB)^i{}_l C^l{}_j = A^i{}_k (BC)^k{}_j = (ABC)^i{}_j. \tag{3.1.9}$$

Tr $\mathrm{Tr}(A) \equiv A^i{}_i$ denotes the trace of a square matrix A. The index notation makes it clear the trace of AB is that of BA because

$$\mathrm{Tr}\left[A \cdot B\right] = A^l{}_k B^k{}_l = B^k{}_l A^l{}_k = \mathrm{Tr}\left[B \cdot A\right]. \tag{3.1.10}$$

This immediately implies the Tr is cyclic, in the sense that

$$\mathrm{Tr}\left[X_1 \cdot X_2 \cdots X_N\right] = \mathrm{Tr}\left[X_N \cdot X_1 \cdot X_2 \cdots X_{N-1}\right] = \mathrm{Tr}\left[X_2 \cdot X_3 \cdots X_N \cdot X_1\right]. \tag{3.1.11}$$

Problem 3.1. Prove the linearity of the Tr, namely for $D \times D$ matrices X and Y and complex number λ,

$$\mathrm{Tr}\left[X+Y\right] = \mathrm{Tr}\left[X\right] + \mathrm{Tr}\left[Y\right], \qquad \mathrm{Tr}\left[\lambda X\right] = \lambda \mathrm{Tr}\left[X\right]. \tag{3.1.12}$$

Comment on whether it makes sense to define $\mathrm{Tr}(A) \equiv A^i{}_i$, if A is not a square matrix. □

Identity and the Kronecker delta The $D \times D$ identity matrix $\mathbb{I}$ has 1 on each and every component on its diagonal and 0 everywhere else. This is also the Kronecker delta.

$$\begin{aligned} \mathbb{I}^i{}_j = \delta^i{}_j &= 1, \qquad i = j \\ &= 0, \qquad i \neq j \end{aligned} \tag{3.1.13}$$

The Kronecker delta is also the flat Euclidean metric in D spatial dimensions; in that context we would write it with both lower indices δ_{ij} and its inverse is δ^{ij}.

The Kronecker delta is also useful for representing *diagonal* matrices. These are matrices that have non-zero entries strictly on their diagonal, where row equals to column number. For example $A^i{}_j = a_i \delta^i_j = a_j \delta^i_j$ is the diagonal matrix with $a_1, a_2, \dots, a_D$ filling its diagonal components, from the upper left to the lower right. Diagonal matrices are also often denoted, for instance, as

$$A = \mathrm{diag}[a_1, \dots, a_D]. \tag{3.1.14}$$

Suppose we multiply AB, where B is also diagonal ($B^i{}_j = b_i \delta^i_j = b_j \delta^i_j$),

$$(AB)^i{}_j = \sum_l a_i \delta^i_l b_j \delta^l_j. \tag{3.1.15}$$

If $i \neq j$ there will be no l that is simultaneously equal to i and j; therefore either one or both the Kronecker deltas are zero and the entire sum is zero. If $i = j$ then when (and only when) $l = i = j$, the Kronecker deltas are both one, and

$$(AB)^i{}_j = a_i b_j. \tag{3.1.16}$$

This means we have shown, using index notation, that the product of diagonal matrices yields another diagonal matrix.

$$(AB)^i{}_j = a_i b_j \delta^i_j \qquad \text{(No sum over } i, j\text{)}. \tag{3.1.17}$$

Transpose The transpose of *any* matrix A, denoted a A^{T}, is defined via

$$(A^{\mathrm{T}})^i{}_j = A^j{}_i. \tag{3.1.18}$$

In words: the i row of A^{T} is the ith column of A; the jth column of A^{T} is the jth row of A. If A is a (square) $D \times D$ matrix, you reflect it along the diagonal to obtain A^{T}.

Problem 3.2. Show using index notation that $(A \cdot B)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$. □

Problem 3.3. Explain why the ij component of $A^{\mathrm{T}} B$ is

$$(A^{\mathrm{T}} B)^i{}_j = A^l{}_i B^l{}_j \tag{3.1.19}$$

whereas the ij component of AB^{T} is

$$(AB^{\mathrm{T}})^i{}_j = A^i{}_l B^j{}_l. \tag{3.1.20}$$

□

Adjoint The adjoint † of *any* matrix is given by

$$(A^\dagger)^i{}_j = (A^j{}_i)^* = (A^*)^j{}_i. \tag{3.1.21}$$

In other words, $A^\dagger = (A^{\mathrm{T}})^*$; to get $A^\dagger$, you start with A, take its transpose, then take its complex conjugate. An example is,

$$A = \begin{bmatrix} 1+i & e^{i\theta} \\ x+iy & \sqrt{10} \end{bmatrix}, \qquad 0 \leq \theta < 2\pi,\ x, y \in \mathbb{R} \tag{3.1.22}$$

$$A^{\mathrm{T}} = \begin{bmatrix} 1+i & x+iy \\ e^{i\theta} & \sqrt{10} \end{bmatrix}, \qquad A^\dagger = \begin{bmatrix} 1-i & x-iy \\ e^{-i\theta} & \sqrt{10} \end{bmatrix}. \tag{3.1.23}$$

Orthogonal, Unitary, Symmetric, and Hermitian A $D \times D$ matrix A is

1. Orthogonal if $A^{\mathrm{T}} A = A A^{\mathrm{T}} = \mathbb{I}$. The set of real orthogonal matrices implement rotations in a D-dimensional real (vector) space.

2. Unitary if $A^\dagger A = A A^\dagger = \mathbb{I}$. Thus, a real unitary matrix is orthogonal. Moreover, unitary matrices, like their real orthogonal counterparts, implement "rotations" in a D dimensional complex (vector) space.

3. Symmetric if $A^{\mathrm{T}} = A$; anti-symmetric if $A^{\mathrm{T}} = -A$.

4. Hermitian if $A^{\dagger} = A$; anti-hermitian if $A^{\dagger} = -A$.

Problem 3.4. Explain why, if A is an orthogonal matrix, it obeys the equation

$$A^i{}_k A^j{}_l \delta_{ij} = \delta_{kl}. \tag{3.1.24}$$

Now explain why, if A is a unitary matrix, it obeys the equation

$$(A^i{}_k)^* A^j{}_l \delta_{ij} = \delta_{kl}. \tag{3.1.25}$$

□

Problem 3.5. Prove that $(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$ and $(AB)^{\dagger} = B^{\dagger} A^{\dagger}$. This means if A and B are orthogonal, then AB is orthogonal; and if A and B are unitary AB is unitary. Can you explain why? □

Simple examples of a unitary, symmetric and Hermitian matrix are, respectively (from left to right):

$$\begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{i\delta} \end{bmatrix}, \qquad \begin{bmatrix} e^{i\theta} & X \\ X & e^{i\delta} \end{bmatrix}, \qquad \begin{bmatrix} \sqrt{109} & 1-i \\ 1+i & \theta^{\delta} \end{bmatrix}, \qquad \theta, \delta \in \mathbb{R}. \tag{3.1.26}$$

3.2 Determinants, Linear (In)dependence, Inverses, Eigensystems

Levi-Civita symbol and the Determinant We will now define the determinant of a $D \times D$ matrix A through the Levi-Civita symbol $\epsilon_{i_1 i_2 \dots i_{D-1} i_D}$, where every index runs from 1 through D:

$$\det A \equiv \epsilon_{i_1 i_2 \dots i_{D-1} i_D} A^{i_1}{}_1 A^{i_2}{}_2 \dots A^{i_{D-1}}{}_{D-1} A^{i_D}{}_D. \tag{3.2.1}$$

This definition is equivalent to the usual co-factor expansion definition.

The $D-$dimensional Levi-Civita symbol is defined through the following properties.

- It is completely antisymmetric in its indices. This means swapping any of the indices $i_a \leftrightarrow i_b$ (for $a \neq b$) will return

$$\epsilon_{i_1 i_2 \dots i_{a-1} i_a i_{a+1} \dots i_{b-1} i_b i_{b+1} \dots i_{D-1} i_D} = -\epsilon_{i_1 i_2 \dots i_{a-1} i_b i_{a+1} \dots i_{b-1} i_a i_{b+1} \dots i_{D-1} i_D}. \tag{3.2.2}$$

- In matrix algebra and flat Euclidean space, $\epsilon_{123\dots D} = \epsilon^{123\dots D} \equiv 1$.[5]

These are sufficient to define every component of the Levi-Civita symbol. Because ϵ is fully anti-symmetric, if any of its D indices are the same, say $i_a = i_b$, then the Levi-Civita symbol returns zero. (Why?) Whenever $i_1 \dots i_D$ are distinct indices, $\epsilon_{i_1 i_2 \dots i_{D-1} i_D}$ is really the sign of the permutation ($\equiv (-)^{\text{number of swaps of index pairs}}$) that brings $\{1, 2, \dots, D-1, D\}$ to $\{i_1, i_2, \dots, i_{D-1}, i_D\}$. Hence, $\epsilon_{i_1 i_2 \dots i_{D-1} i_D}$ is $+1$ when it takes zero/even number of swaps, and -1 when it takes odd.

[5] In Lorentzian flat spacetimes, the Levi-Civita tensor with upper indices will need to be carefully distinguished from its counterpart with lower indices.

For example, in the 2 dimensional case $\epsilon_{11} = \epsilon_{22} = 0$; whereas it takes one swap to go from 12 to 21. Therefore,

$$1 = \epsilon_{12} = -\epsilon_{21}. \tag{3.2.3}$$

In the 3 dimensional case,

$$1 = \epsilon_{123} = -\epsilon_{213} = -\epsilon_{321} = -\epsilon_{132} = \epsilon_{231} = \epsilon_{312}. \tag{3.2.4}$$

Properties of the determinant include

$$\det A^{\mathrm{T}} = \det A, \qquad \det(A \cdot B) = \det A \cdot \det B, \qquad \det A^{-1} = \frac{1}{\det A}, \tag{3.2.5}$$

for all square matrices A and B. As a simple example, let us use eq. (3.2.1) to calculate the determinant of

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{3.2.6}$$

Remember the only non-zero components of $\epsilon_{i_1 i_2}$ are $\epsilon_{12} = 1$ and $\epsilon_{21} = -1$.

$$\begin{aligned} \det A = \epsilon_{12} A^1{}_1 A^2{}_2 + \epsilon_{21} A^2{}_1 A^1{}_2 &= A^1{}_1 A^2{}_2 - A^2{}_1 A^1{}_2 \\ &= ad - bc. \end{aligned} \tag{3.2.7}$$

Problem 3.6. Inverse of 2×2 matrix By viewing ϵ as a 2×2 matrix, prove that, whenever the inverse of a matrix M exist, it can be written as

$$M^{-1} = -\frac{\epsilon \cdot M^{\mathrm{T}} \cdot \epsilon}{\det M} = \frac{\epsilon^\dagger \cdot M^{\mathrm{T}} \cdot \epsilon}{\det M} = \frac{\epsilon \cdot M^{\mathrm{T}} \cdot \epsilon^\dagger}{\det M}. \tag{3.2.8}$$

Hint: Can you explain why eq. (3.2.1) implies

$$\epsilon_{\mathrm{AB}} M^{\mathrm{A}}{}_{\mathrm{I}} M^{\mathrm{B}}{}_{\mathrm{J}} = \epsilon_{\mathrm{IJ}} \det M? \tag{3.2.9}$$

Then contract both sides with M^{-1} and use $\epsilon^2 = -\mathbb{I}$. Or, simply prove it by brute force. □

Problem 3.7. Generalization of eq. (3.2.1) Explain why eq. (3.2.1) implies

$$\epsilon_{i_1 i_2 \dots i_{D-1} i_D} A^{i_1}{}_{j_1} A^{i_2}{}_{j_2} \dots A^{i_{D-1}}{}_{j_{D-1}} A^{i_D}{}_{j_D} = \epsilon_{j_1 j_2 \dots j_{D-1} j_D} \det A. \tag{3.2.10}$$

Hint: What happens when you swap $A^{i_m}{}_m$ and $A^{i_n}{}_n$ in eq. (3.2.1)? □

Whenever the inverse of A in eq. (3.2.10) exists; i.e., there is a A^{-1} such that $(A^{-1})^i{}_l A^l{}_j = \delta^i{}_j = A^i{}_l (A^{-1})^l{}_j$; then by contracting both sides with N powers of A^{-1} one would arrive at

$$\epsilon_{i_1 \dots i_N k_1 \dots k_{D-N}} A^{k_1}{}_{j_1} \dots A^{k_{D-N}}{}_{j_{D-N}} = (A^{-1})^{l_1}{}_{i_1} \dots (A^{-1})^{l_N}{}_{i_N} \epsilon_{l_1 \dots l_N j_1 \dots j_{D-N}} (\det R). \tag{3.2.11}$$

This formula has applications in, for instance, rigid body motion in 3D.

Problem 3.8. Determinant of 2-Block Off Diagonal Matrix Consider the following $2N \times 2N$ matrix,

$$M = \begin{bmatrix} 0 & A_{N\times N} \\ B_{N\times N} & 0 \end{bmatrix}; \tag{3.2.12}$$

where A and B are $N \times N$ blocks. Prove that

$$\det M = (-)^{N^2} (\det A)(\det B). \tag{3.2.13}$$

Hint: You should find the leftmost N terms of the right hand side of eq. (3.2.1) to involve $\det B$ and the rightmost N terms $\det A$. $\square$

Linear (in)dependence Given a set of D vectors $\{v_1, \dots, v_D\}$, we say one of them is linearly dependent (say v_i) if we can express it in as a sum of multiples of the rest of the vectors,

$$v_i = \sum_{j\neq i}^{D-1} \chi_j v_j \qquad \text{for some} \qquad \chi_j \in \mathbb{C}. \tag{3.2.14}$$

We say the D vectors are linearly independent if none of the vectors are linearly dependent on the rest.

Determinant as test of linear independence If we view the columns or rows of a $D \times D$ matrix A as vectors and if these D vectors are linearly dependent, then the determinant of A is zero. This is because of the antisymmetric nature of the Levi-Civita symbol. Moreover, suppose $\det A \neq 0$. Cramer's rule (cf. eq. (3.2.30) below) tells us the inverse A^{-1} exists. In fact, for finite dimensional matrix A, its inverse A^{-1} is unique. That means the only solution to the D-component row (or column) vector w, obeying $w \cdot A = 0$ (or, $A \cdot w = 0$), is $w = 0$. And since $w \cdot A$ (or $A \cdot w$) describes the linear combination of the rows (or, columns) of A; this indicates they must be linearly independent whenever $\det A \neq 0$.

> For a square matrix A, $\det A = 0$ iff ($\equiv$ if and only if) its columns and rows are linearly dependent. Equivalently, $\det A \neq 0$ iff its columns and rows are linearly independent.

Problem 3.9. If the columns of a square matrix A are linearly dependent, use eq. (3.2.1) to prove that $\det A = 0$. Hint: use the antisymmetric nature of the Levi-Civita symbol.

Problem 3.10. Show that, for a $D \times D$ matrix A and some complex number λ,

$$\det(\lambda A) = \lambda^D \det A. \tag{3.2.15}$$

Hint: this follows almost directly from eq. (3.2.1). $\square$

Relation to cofactor expansion The co-factor expansion definition of the determinant is

$$\det A = \sum_{i=1}^{D} A^i{}_k C^i{}_k, \tag{3.2.16}$$

where k is an arbitrary integer from 1 through D. The $C^i{}_k$ is $(-)^{i+k}$ times the determinant of the $(D-1)\times(D-1)$ matrix formed from removing the ith row and kth column of A. (This definition sums over the row numbers; it is actually equally valid to define it as a sum over column numbers.)

As a 3×3 example, we have

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & l \end{bmatrix} = b(-)^{1+2}\det\begin{bmatrix} d & f \\ g & l \end{bmatrix} + e(-)^{2+2}\det\begin{bmatrix} a & c \\ g & l \end{bmatrix} + h(-)^{3+2}\det\begin{bmatrix} a & c \\ d & f \end{bmatrix}. \tag{3.2.17}$$

Problem 3.11. Prove that the cofactor expansion definition of the determinant of a matrix is equivalent to that in eq. (3.2.1). □

Pauli Matrices The 2×2 identity together with the Pauli matrices are Hermitian matrices.

$$\sigma^0 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \sigma^1 \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma^2 \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma^3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{3.2.18}$$

Moreover, *any* complex 2×2 matrix may be expressed as a linear combination of these $\{\sigma^\mu | \mu = 0,1,2,3\}$. This important fact has deep-ploughing applications, including the study of symmetries in 4D flat spacetime and (quantum) field theory.

Problem 3.12. Let $p_\mu \equiv (p_0, p_1, p_2, p_3)$ be a 4-component collection of complex numbers. Verify the following determinant, relevant for the study of Lorentz symmetry in 4-dimensional flat spacetime,

$$p_\mu \sigma^\mu = \begin{bmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{bmatrix}, \tag{3.2.19}$$

$$\det p_\mu \sigma^\mu = \sum_{0\leq\mu,\nu\leq 3} \eta^{\mu\nu} p_\mu p_\nu \equiv p^2, \tag{3.2.20}$$

where $p_\mu\sigma^\mu \equiv \sum_{0\leq\mu\leq 3} p_\mu \sigma^\mu$ and

$$\eta^{\mu\nu} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{3.2.21}$$

(This is the metric in 4 dimensional flat "Minkowski" spacetime.) Verify, for $i,j,k \in \{1,2,3\}$ and ϵ denoting the 2D Levi-Civita symbol,

$$\det \sigma^0 = 1, \qquad \det \sigma^i = -1, \qquad \mathrm{Tr}\left[\sigma^0\right] = 2, \qquad \mathrm{Tr}\left[\sigma^i\right] = 0 \tag{3.2.22}$$

$$\sigma^i\sigma^j = \delta^{ij}\mathbb{I} + i\sum_{1\leq k\leq 3}\epsilon^{ijk}\sigma^k, \qquad \epsilon\sigma^i\epsilon = (\sigma^i)^* = \epsilon(-\sigma^i)\epsilon^\dagger. \tag{3.2.23}$$

Also use the antisymmetric nature of the Levi-Civita symbol to aruge that

$$\theta_i \theta_j \epsilon^{ijk} = 0. \tag{3.2.24}$$

Defining the exponential of a matrix X via the Taylor series

$$\exp X = \sum_{\ell=0}^{\infty} \frac{X^\ell}{\ell!}, \tag{3.2.25}$$

use these facts to derive the result:

$$\begin{aligned} U(\vec{\theta}) &\equiv \exp\left[-\frac{i}{2}\sum_{j=1}^{3}\theta_j\sigma^j\right] \equiv e^{-(i/2)\vec{\theta}\cdot\vec{\sigma}} \\ &= \cos\left(\frac{1}{2}|\vec{\theta}|\right)\mathbb{I}_{2\times 2} - i\frac{\vec{\theta}\cdot\vec{\sigma}}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right), \qquad |\vec{\theta}| = \sqrt{\theta_i\theta_i} \equiv \sqrt{\vec{\theta}\cdot\vec{\theta}}, \qquad (3.2.26) \\ &= \begin{bmatrix} \cos\left(\frac{1}{2}|\vec{\theta}|\right) - i\frac{\theta_3}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right) & \frac{-i\theta_1-\theta_2}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right) \\ \frac{-i\theta_1+\theta_2}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right) & \cos\left(\frac{1}{2}|\vec{\theta}|\right) + i\frac{\theta_3}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right) \end{bmatrix}; \qquad (3.2.27) \end{aligned}$$

which is valid for complex $\{\theta_i\}$. (Hint: Applying the first relation in eq. (3.2.23) to each term in the Taylor series of the exponential.)

Show that any 2×2 complex matrix A can be built from $p_\mu\sigma^\mu$ by choosing the p_μs appropriately. Then compute $(1/2)\text{Tr}\left[p_\mu\sigma^\mu\sigma^\nu\right]$, for $\nu = 0, 1, 2, 3$, and comment on how the trace can be used, given A, to solve for the p_μ in the equation

$$p_\mu\sigma^\mu = A. \tag{3.2.28}$$

□

Inverse The inverse of the $D\times D$ matrix A is defined to be

$$A^{-1}A = AA^{-1} = \mathbb{I}. \tag{3.2.29}$$

The inverse A^{-1} of a finite dimensional matrix A is unique; moreover, the left $A^{-1}A = \mathbb{I}$ and right inverses $AA^{-1} = \mathbb{I}$ are the same object. The inverse exists if and only if ($\equiv$ iff) $\det A \neq 0$.

Problem 3.13. Cramer's rule Can you show the equivalence of equations (3.2.1) and (3.2.16)? Can you also show that

$$\delta_{kl}\det A = \sum_{i=1}^{D} A^i{}_k C^i{}_l? \tag{3.2.30}$$

That is, show that when $k \neq l$, the sum on the right hand side is zero. Explain why eq. (3.2.30) informs us that

$$(A^{-1})^l{}_i = (\det A)^{-1}\sum_{i=1}^{D} C^i{}_l. \tag{3.2.31}$$

Hint: start from the left-hand-side, namely

$$\begin{aligned}\det A &= \epsilon_{j_1 \dots j_D} A^{j_1}{}_1 \dots A^{j_D}{}_D \\ &= A^i{}_k \left(\epsilon_{j_1 \dots j_{k-1} i j_{k+1} \dots j_D} A^{j_1}{}_1 \dots A^{j_{k-1}}{}_{k-1} A^{j_{k+1}}{}_{k+1} \dots A^{j_D}{}_D \right),\end{aligned} \tag{3.2.32}$$

where k is an arbitrary integer in the set $\{1, 2, 3, \dots, D-1, D\}$. Examine the term in the parenthesis. First shift the index i, which is located at the kth slot from the left, to the ith slot. Then argue why the result is $(-)^{i+k}$ times the determinant of A with the ith row and kth column removed. Finally, remember $A^{-1} \cdot A = \mathbb{I}$. □

Problem 3.14. Why are the left and right inverses of (an invertible) matrix A the same? Hint: Consider $LA = \mathbb{I}$ and $AR = \mathbb{I}$; for the first, multiply R on both sides from the right. □

Problem 3.15. Prove that $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$ and $(A^{-1})^{\dagger} = (A^{\dagger})^{-1}$. □

3.3 Eigenvectors and Eigenvalues

If A is a $D \times D$ matrix, v is its (D-component) eigenvector with eigenvalue λ if it obeys

$$A^i{}_j v^j = \lambda v^i. \tag{3.3.1}$$

This means

$$(A^i{}_j - \lambda \delta^i{}_j) v^j = 0 \tag{3.3.2}$$

has non-trivial solutions iff

$$P_D(\lambda) \equiv \det (A - \lambda \mathbb{I}) = 0. \tag{3.3.3}$$

Equation (3.3.3) is known as the characteristic equation. For a $D \times D$ matrix, it gives us a Dth degree polynomial $P_D(\lambda)$ for λ, whose roots are the eigenvalues of the matrix λ – the set of all eigenvalues of a matrix is called its *spectrum*. For each solution for λ, we then proceed to solve for the v^i in eq. (3.3.2). That there is always at least one solution – there could be more – for v^i is because, since its determinant is zero, the columns of $A - \lambda\mathbb{I}$ are necessarily linearly dependent. As already discussed above, this amounts to the statement that there is some sum of multiples of these columns ($\equiv$ "linear combination") that yields zero – in fact, the components of v^i are precisely the coefficients in this sum. If $\{w_i\}$ are these columns of $A - \lambda\mathbb{I}$,

$$A - \lambda\mathbb{I} \equiv [w_1 w_2 \dots w_D] \qquad \Rightarrow \qquad (A - \lambda\mathbb{I})v = \sum_j w_j v^j = 0. \tag{3.3.4}$$

(Note that, if $\sum_j w_j v^j = 0$ then $\sum_j w_j (K v^j) = 0$ too, for any complex number K; in other words, eigenvectors are only defined up to an overall multiplicative constant.) Every $D \times D$ matrix has D eigenvalues from solving the Dth order polynomial equation (3.3.3); from that, you can then obtain D corresponding eigenvectors. Note, however, the eigenvalues can be repeated; when this occurs, it is known as a *degenerate* spectrum. Moreover, not all the eigenvectors are guaranteed to be linearly independent; i.e., some eigenvectors can turn out to be sums of multiples of other eigenvectors.

The **Cayley-Hamilton theorem** states that the matrix A satisfies its own characteristic equation. In detail, if we express eq. (3.3.3) as $\sum_{i=0}^{D} q_i \lambda^i = 0$ (for appropriate complex constants $\{q_i\}$), then replace $\lambda^i \to A^i$ (namely, the ith power of λ with the ith power of A), we would find

$$P_D(A) = 0. \tag{3.3.5}$$

Any $D \times D$ matrix A admits a **Schur decomposition**. Specifically, there is some unitary matrix U such that A can be brought to an upper triangular form, with its eigenvalues on the diagonal:

$$U^\dagger A U = \text{diag}(\lambda_1, \dots, \lambda_D) + N, \tag{3.3.6}$$

where N is strictly upper triangular, with $N^i{}_j = 0$ for $j \leq i$. The Schur decomposition can be proved via mathematical induction on the size of the matrix.

Diagonalization A special case of the Schur decomposition occurs when all the off-diagonal elements are zero. A $D \times D$ matrix A can be *diagonalized* if there is some unitary matrix U such that

$$U^\dagger A U = \text{diag}(\lambda_1, \dots, \lambda_D), \tag{3.3.7}$$

where the $\{\lambda_i\}$ are the eigenvalues of A. Each column of U is filled with a distinct unit length eigenvector of A. (Unit length means $v^\dagger v = (v^i)^* v^j \delta_{ij} = 1$.) In index notation,

$$A^i{}_j U^j{}_k = \lambda_k U^i{}_k = U^i{}_l \delta^l{}_k \lambda_k, \qquad \text{(No sum over } k\text{)}. \tag{3.3.8}$$

In matrix notation,

$$AU = U \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_{D-1}, \lambda_D]. \tag{3.3.9}$$

Here, $U^j{}_k$ for fixed k, is the kth eigenvector, and λ_k is the corresponding eigenvalue. By multiplying both sides with $U^\dagger$, we have

$$U^\dagger A U = D, \qquad D^j{}_l \equiv \lambda_l \delta^j{}_l \qquad \text{(No sum over } l\text{)}. \tag{3.3.10}$$

Equivalently,

$$A = U D U^\dagger. \tag{3.3.11}$$

Some jargon: the *null space* of a matrix M is the space spanned by all vectors $\{v_i\}$ obeying $M \cdot v_i = 0$. When we solve for the eigenvector of A by solving $(A - \lambda \mathbb{I}) \cdot v$, we are really solving for the null space of the matrix $M \equiv A - \lambda \mathbb{I}$, because for a fixed eigenvalue λ, there could be more than one solution – that's what we mean by degeneracy.

What types of matrices can be diagonalized? Real symmetric matrices can be always diagonalized via an orthogonal transformation. Complex Hermitian matrices can always be diagonalized via a unitary one. These statements can be proved readily using their Schur decomposition. For, let A be Hermitian and U be a unitary matrix such that

$$U A U^\dagger = \text{diag}(\lambda_1, \dots, \lambda_D) + N, \tag{3.3.12}$$

where N is strictly upper triangular. Now, if A is Hermitian, so is $UAU^\dagger$, because $(UAU^\dagger)^\dagger = (U^\dagger)^\dagger A^\dagger U^\dagger = UAU^\dagger$. Therefore,

$$(UAU^\dagger)^\dagger = UAU^\dagger \qquad \Rightarrow \qquad \text{diag}(\lambda_1^*, \dots, \lambda_D^*) + N^\dagger = \text{diag}(\lambda_1, \dots, \lambda_D) + N. \tag{3.3.13}$$

Because the transpose of a strictly upper triangular matrix returns a strictly lower triangular matrix, we have a strictly lower triangular matrix $N^\dagger$ plus a diagonal matrix (built out of the complex conjugate of the eigenvalues of A) equal to a diagonal one (built out of the eigenvalues of A) plus a strictly upper triangular N. That means $N = 0$ and $\lambda_l = \lambda_l^*$. That is, any Hermitian A is diagonalizable and all its eigenvalues are real.

Unitary matrices can also always be diagonalized. In fact, all its eigenvalues $\{\lambda_i\}$ lie on the unit circle on the complex plane, i.e., $|\lambda_i| = 1$. Suppose now A is unitary and U is another unitary matrix such that the Schur decomposition of A reads

$$UAU^\dagger = M, \tag{3.3.14}$$

where M is an upper triangular matrix with the eigenvalues of A on its diagonal. Now, if A is unitary, so is $UAU^\dagger$, because

$$\left(UAU^\dagger\right)^\dagger \left(UAU^\dagger\right) = UA^\dagger U^\dagger UAU^\dagger = UA^\dagger AU^\dagger = UU^\dagger = \mathbb{I}. \tag{3.3.15}$$

That means

$$M^\dagger M = \mathbb{I} \qquad \Rightarrow \qquad (M^\dagger M)^k{}_l = (M^\dagger)^k{}_s M^s{}_l = \sum_s \overline{M^s{}_k} M^s{}_l = \delta_{ij} \overline{M^i{}_k} M^j{}_l = \delta_{kl}, \tag{3.3.16}$$

where we have recalled eq. (3.1.25) in the last equality. If w_i denotes the ith column of M, the unitary nature of M implies all its columns are orthogonal to each other and each column has length one. Since M is upper triangular, we see that the only non-zero component of the first column is its first row, i.e., $w_1^i = M^i{}_1 = \lambda_1 \delta_1^i$. Unit length means $w_1^\dagger w_1 = 1 \Rightarrow |\lambda_1|^2 = 1$. That w_1 is orthogonal to every other column $w_{i>1}$ means the latter have their first rows equal to zero; $\overline{M^1{}_1} M^1{}_l = \overline{\lambda_1} M^1{}_l = 0 \Rightarrow M^1{}_l = 0$ for $l \neq 1$ – remember $\overline{M^1{}_1} = \overline{\lambda_1}$ itself cannot be zero because it lies on the unit circle on the complex plane. Now, since its first component is necessarily zero, the only non-zero component of the second column is its second row, i.e., $w_2^i = M^i{}_2 = \lambda_2 \delta_2^i$. Unit length again means $|\lambda_2|^2 = 1$. And, by demanding that w_2 be orthogonal to every other column means their second components are zero: $\overline{M^2{}_2} M^2{}_l = \overline{\lambda_2} M^2{}_l = 0 \Rightarrow M^2{}_l = 0$ for $l > 2$ – where, again, $\overline{M^2{}_2} = \overline{\lambda_2}$ cannot be zero because it lies on the complex plane unit circle. By induction on the column number, we see that the only non-zero component of the ith column is the ith row. That is, any unitary A is diagonalizable and all its eigenvalues lie on the circle: $|\lambda_{1 \leq i \leq D}| = 1$.

More generally, a complex square matrix A is diagonalizable if and only if it is *normal*, which in turn is defined as a matrix that commutes with its adjoint, namely

$$[A, A^\dagger] \equiv A \cdot A^\dagger - A^\dagger \cdot A = 0. \tag{3.3.17}$$

We prove this in §(4.7). Note that, if A is Hermitian, it must be normal:

$$[A, A^\dagger] = AA^\dagger - A^\dagger A = A^2 - A^2 = 0. \tag{3.3.18}$$

Likewise, unitary matrices are also normal; if $A^\dagger A = \mathbb{I} = AA^\dagger$,

$$[A, A^\dagger] = AA^\dagger - A^\dagger A = \mathbb{I} - \mathbb{I} = 0. \tag{3.3.19}$$

Diagonalization example As an example, let's diagonalize σ^2 from eq. (3.2.18).

$$P_2(\lambda) = \det\left[\sigma^2 - \lambda\mathbb{I}_{2\times 2}\right] = \det\begin{bmatrix} -\lambda & -i \\ i & -\lambda \end{bmatrix} = \lambda^2 - 1 = 0 \tag{3.3.20}$$

(We can even check Caley-Hamilton here: $P_2(\sigma^2) = (\sigma^2)^2 - \mathbb{I} = \mathbb{I} - \mathbb{I} = 0$; see eq. (3.2.23).) The solutions are $\lambda = \pm 1$ and

$$\begin{bmatrix} \mp 1 & -i \\ i & \mp 1 \end{bmatrix}\begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad v^1_\pm = \mp i v^2_\pm. \tag{3.3.21}$$

The subscripts on v refer to their eigenvalues, namely

$$\sigma^2 v_\pm = \pm v_\pm. \tag{3.3.22}$$

By choosing $v^2 = 1/\sqrt{2}$, we can check $(v^i_\pm)^* v^j_\pm \delta_{ij} = 1$ and therefore the normalized eigenvectors are

$$v_\pm = \frac{1}{\sqrt{2}}\begin{bmatrix} \mp i \\ 1 \end{bmatrix}. \tag{3.3.23}$$

Furthermore you can check directly that eq. (3.3.22) is satisfied. We therefore have

$$\underbrace{\left(\frac{1}{\sqrt{2}}\begin{bmatrix} i & 1 \\ -i & 1 \end{bmatrix}\right)}_{\equiv U^\dagger}\sigma^2\underbrace{\left(\frac{1}{\sqrt{2}}\begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}\right)}_{\equiv U} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.3.24}$$

An example of a matrix that cannot be diagonalized is

$$A \equiv \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}. \tag{3.3.25}$$

The characteristic equation is $\lambda^2 = 0$, so both eigenvalues are zero. Therefore $A - \lambda\mathbb{I} = A$, and

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad v^1 = 0,\ v^2 \text{ arbitrary}. \tag{3.3.26}$$

There is a repeated eigenvalue of 0, but there is only one linearly independent eigenvector $(0, 1)$. It is not possible to build a unitary 2×2 matrix U whose columns are distinct unit length eigenvectors of σ^2.

Problem 3.16. Show how to go from eq. (3.3.8) to eq. (3.3.10) using index notation. □

Problem 3.17. Use the Schur decomposition to explain why, for any matrix A, $\mathrm{Tr}\,[A]$ is equal to the sum of its eigenvalues and $\det A$ is equal to their product:

$$\mathrm{Tr}\,[A] = \sum_{l=1}^{D}\lambda_l, \qquad \det A = \prod_{l=1}^{D}\lambda_l. \tag{3.3.27}$$

Hint: For $\det A$, the key question is how to take the determinant of an upper triangular matrix. □

Problem 3.18. For a *strictly* upper triangular matrix N, prove that N multiplied to itself any number of times still returns a strictly upper triangular matrix. $\square$

Problem 3.19. Can a strictly upper triangular matrix be diagonalized? (Explain.) Hint: What is the eigensystem of such a matrix?

Problem 3.20. Suppose $A = UXU^\dagger$, where U is a unitary matrix. If $f(z)$ is a function of z that can be Taylor expanded about some point z_0, explain why $f(A) = Uf(X)U^\dagger$. Hint: Can you explain why $(UBU^\dagger)^\ell = UB^\ell U^\dagger$, for B some arbitrary matrix, U unitary, and $\ell = 1, 2, 3, \dots$? $\square$

Problem 3.21. Can you provide a simple explanation to why the eigenvalues $\{\lambda_l\}$ of a unitary matrix are always of unit absolute magnitude; i.e. why are the $|\lambda_l| = 1$? $\square$

Problem 3.22. Simplified example of neutrino oscillations. We begin with the observation that the solution to the first order equation

$$i\partial_t \psi(t) = E\psi(t), \tag{3.3.28}$$

for E some real constant, is

$$\psi(t) = e^{-iEt}\psi_0. \tag{3.3.29}$$

The ψ_0 is some arbitrary (possibly complex) constant, corresponding to the initial condition $\psi(t = 0)$. Now solve the matrix differential equation

$$i\partial_t N(t) = HN(t), \qquad N(t) \equiv \begin{bmatrix} \nu_1(t) \\ \nu_2(t) \end{bmatrix}, \tag{3.3.30}$$

with the initial condition – describing the production of ν_1-type of neutrino, say –

$$\begin{bmatrix} \nu_1(t=0) \\ \nu_2(t=0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{3.3.31}$$

where the Hamiltonian H is

$$H \equiv \begin{bmatrix} p & 0 \\ 0 & p \end{bmatrix} + \frac{1}{4p}M, \tag{3.3.32}$$

$$M \equiv \begin{bmatrix} m_1^2 + m_2^2 + (m_1^2 - m_2^2)\cos(2\theta) & (m_1^2 - m_2^2)\sin(2\theta) \\ (m_1^2 - m_2^2)\sin(2\theta) & m_1^2 + m_2^2 + (m_2^2 - m_1^2)\cos(2\theta) \end{bmatrix}. \tag{3.3.33}$$

The p is the magnitude of the momentum, $m_{1,2}$ are masses, and θ is the "mixing angle". Then calculate

$$P_{1\to 1} \equiv \left| N(t)^\dagger \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right|^2 \qquad \text{and} \qquad P_{1\to 2} \equiv \left| N(t)^\dagger \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right|^2. \tag{3.3.34}$$

Express $P_{1\to 1}$ and $P_{1\to 2}$ in terms of $\Delta m^2 \equiv m_1^2 - m_2^2$. (In quantum mechanics, they respectively correspond to the probability of observing the neutrinos ν_1 and ν_2 at time $t > 0$, given ν_1 was produced at $t = 0$.) Hint: Start by diagonalizing $M = U^{\mathrm{T}}AU$ where

$$U \equiv \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{3.3.35}$$

The $UN(t)$ is known as the "mass-eigenstate" basis. Can you comment on why? Note that, in the highly relativistic limit, the energy E of a particle of mass m is

$$E = \sqrt{p^2 + m^2} \to p + \frac{m^2}{2p} + \mathcal{O}(1/p^2). \tag{3.3.36}$$

Note: In this problem, we have implicitly set $\hbar = c = 1$, where $\hbar$ is the reduced Planck's constant and c is the speed of light in vacuum. $\Box$

Problem 3.23. Quadrupole Moments Show that, for $N \geq 1$ positive masses $\{m_\ell > 0\}$, real position vectors $\{\vec{x}_\ell\}$, and the $\vec{x}^2$ denoting the dot product $\vec{x} \cdot \vec{x}$, the second moments

$$A^{ij} = \sum_{\ell=1}^{N} m_\ell x_\ell^i x_\ell^j \qquad \text{and} \qquad B^{ij} = \sum_{\ell=1}^{N} m_\ell \left(\delta^{ij} \vec{x}_\ell^2 - x_\ell^i x_\ell^j \right) \tag{3.3.37}$$

have strictly non-negative eigenvalues. Hint: Both A^{ij} and B^{ij} are real and symmetric. For all eigenvectors $\{\vec{v}\}$, consider $v^i A^{ij} v^j$ or $v^i B^{ij} v^j$. $\Box$

3.4 ⋆2D Real Orthogonal Matrices

In this subsection we will illustrate what a real orthogonal matrix is by studying the 2D case in some detail. Let A be such a 2×2 real orthogonal matrix. We will begin by writing its components as follows

$$A \equiv \begin{bmatrix} v^1 & v^2 \\ w^1 & w^2 \end{bmatrix}. \tag{3.4.1}$$

(As we will see, it is useful to think of $v^{1,2}$ and $w^{1,2}$ as components of 2D vectors.) That A is orthogonal means $AA^{\mathrm{T}} = \mathbb{I}$.

$$\begin{bmatrix} v^1 & v^2 \\ w^1 & w^2 \end{bmatrix} \cdot \begin{bmatrix} v^1 & w^1 \\ v^2 & w^2 \end{bmatrix} = \begin{bmatrix} \vec{v} \cdot \vec{v} & \vec{v} \cdot \vec{w} \\ \vec{w} \cdot \vec{v} & \vec{w} \cdot \vec{w} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.4.2}$$

This translates to: $\vec{w}^2 \equiv \vec{w} \cdot \vec{w} = 1$, $\vec{v}^2 \equiv \vec{v} \cdot \vec{v} = 1$ (length of both the 2D vectors are one); and $\vec{w} \cdot \vec{v} = 0$ (the two vectors are perpendicular). In 2D any vector can be expressed in polar coordinates; for example, the Cartesian components of $\vec{v}$ are

$$v^i = r(\cos\phi, \sin\phi), \qquad r \geq 0, \ \phi \in [0, 2\pi). \tag{3.4.3}$$

But $\vec{v}^2 = 1$ means $r = 1$. Similarly,

$$w^i = (\cos\phi', \sin\phi'), \qquad \phi' \in [0, 2\pi). \tag{3.4.4}$$

Because $\vec{v}$ and $\vec{w}$ are perpendicular,

$$\vec{v} \cdot \vec{w} = \cos\phi \cdot \cos\phi' + \sin\phi \cdot \sin\phi' = \cos(\phi - \phi') = 0. \tag{3.4.5}$$

This means $\phi' = \phi \pm \pi/2$. (Why?) Furthermore

$$w^i = (\cos(\phi \pm \pi/2), \sin(\phi \pm \pi/2)) = (\mp\sin(\phi), \pm\cos(\phi)). \tag{3.4.6}$$

What we have figured out is that, any real orthogonal matrix can be parametrized by an angle $0 \leq \phi < 2\pi$; and for each ϕ there are two distinct solutions.

$$R_1(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}, \qquad R_2(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{bmatrix}. \tag{3.4.7}$$

By a direct calculation you can check that $R_1(\phi > 0)$ rotates an arbitrary 2D vector *clockwise* by ϕ. Whereas, $R_2(\phi > 0)$ rotates the vector, followed by flipping the sign of its y-component; this is because

$$R_2(\phi) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot R_1(\phi). \tag{3.4.8}$$

In other words, the $R_2(\phi = 0)$ in eq. (3.4.7) corresponds to a "parity flip" where the vector is reflected about the x-axis.

Problem 3.24. What about the matrix that reflects 2D vectors about the y-axis? What value of θ in $R_2(\theta)$ would it correspond to?

Find the determinants of $R_1(\phi)$ and $R_2(\phi)$. You should be able to use that to argue, there is no θ_0 such that $R_1(\theta_0) = R_2(\theta_0)$. Also verify that

$$R_1(\phi)R_1(\phi') = R_1(\phi + \phi'). \tag{3.4.9}$$

This makes geometric sense: rotating a vector clockwise by ϕ then by ϕ' should be the same as rotation by $\phi+\phi'$. Mathematically speaking, this composition law in eq. (3.4.9) tells us rotations form the SO_2 *group*. The set of $D \times D$ real orthogonal matrices obeying $R^{\mathrm{T}}R = \mathbb{I}$, including both rotations and reflections, forms the group O_D. The group involving only rotations is known as SO_D; where the 'S' stands for "special" ($\equiv$ determinant equals one). $\square$

Problem 3.25. 2×2 Unitary Matrices. Can you construct the most general 2×2 unitary matrix? First argue that the most general complex 2D vector $\vec{v}$ that satisfies $\vec{v}^\dagger \vec{v} = 1$ is

$$v^i = e^{i\phi_1}(\cos\theta, e^{i\phi_2}\sin\theta), \qquad \phi_{1,2}, \theta \in [0, 2\pi). \tag{3.4.10}$$

Then consider $\vec{v}^\dagger \vec{w} = 0$, where

$$w^i = e^{i\phi_1'}(\cos\theta', e^{i\phi_2'}\sin\theta'), \qquad \phi_{1,2}', \theta' \in [0, 2\pi). \tag{3.4.11}$$

You should arrive at

$$\sin(\theta)\sin(\theta')e^{i(\phi_2'-\phi_2)} + \cos(\theta)\cos(\theta') = 0. \tag{3.4.12}$$

By taking the real and imaginary parts of this equation, argue that

$$\phi_2' = \phi_2, \qquad \theta = \theta' \pm \frac{\pi}{2}. \tag{3.4.13}$$

or

$$\phi_2' = \phi_2 + \pi, \qquad \theta = -\theta' \pm \frac{\pi}{2}. \tag{3.4.14}$$

From these, deduce that the most general 2×2 unitary matrix U can be built from the most general real orthogonal one $O(\theta)$ via

$$U = \left[\begin{array}{cc} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{array}\right] \cdot O(\theta) \cdot \left[\begin{array}{cc} 1 & 0 \\ 0 & e^{i\phi_3} \end{array}\right]. \tag{3.4.15}$$

As a simple check: note that $\vec{v}^\dagger\vec{v} = \vec{w}^\dagger\vec{w} = 1$ together with $\vec{v}^\dagger\vec{w} = 0$ provides 4 constraints for 8 parameters – 4 complex entries of a 2×2 matrix – and therefore we should have 4 free parameters left.

Bonus problem: By imposing $\det U = 1$, can you connect eq. (3.4.15) to eq. (3.2.26)? □

3.5 ⋆2D Unitary Matrices

In this section we will construct the most general 2×2 unitary matrix $\widehat{U}$, which satisfy

$$\widehat{U}^\dagger\widehat{U} = \mathbb{I}_{2\times 2} = \widehat{U}\widehat{U}^\dagger. \tag{3.5.1}$$

If we parametrize the matrix as

$$\widehat{U} = \left[\begin{array}{cc} \vec{u} & \vec{v} \end{array}\right], \tag{3.5.2}$$

where $\vec{u}$ and $\vec{v}$ are to be viewed as 2−component complex vectors, then

$$\widehat{U}^\dagger\widehat{U} = \left[\begin{array}{cc} \vec{u}^\dagger\vec{u} & \vec{u}^\dagger\vec{v} \\ \vec{v}^\dagger\vec{u} & \vec{v}^\dagger\vec{u} \end{array}\right] = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]. \tag{3.5.3}$$

Notice, if $\widehat{U}$ is unitary, so is $e^{i\gamma}\widehat{U}$, for real γ; i.e., there is always an overall phase freedom. We first note: for a unit norm vector $\vec{a}$ obeying $\vec{a}^\dagger\vec{a} = |a^1|^2 + |a^2|^2 = 1$, its components may be parametrized as $\vec{a} = (e^{i\alpha_1}\cos\theta, e^{i\alpha_2}\sin\theta)$ for real angles $\alpha_{1,2}$ and θ. The $\vec{u}$ and $\vec{v}$ are therefore expressible

$$\vec{u} = \left(e^{i\alpha_1}\cos(\theta), e^{i\alpha_2}\sin(\theta)\right), \tag{3.5.4}$$
$$\vec{v} = \left(e^{i\alpha_1'}\cos(\theta'), e^{i\alpha_2'}\sin(\theta')\right); \tag{3.5.5}$$

for real angles $\alpha_{1,2}$, $\alpha'_{1,2}$ and θ, θ'. We may use the overall phase freedom of $\widehat{U}$ to set α_1 to 0. The orthogonality relation between $\vec{u}$ and $\vec{v}$ then reads

$$\vec{u}^\dagger\vec{v} = 0 = \vec{v}^\dagger\vec{u} \tag{3.5.6}$$
$$0 = e^{-i\alpha_1'}\cos(\theta)\cos(\theta') + e^{i(\alpha_2-\alpha_2')}\sin(\theta)\sin(\theta') \tag{3.5.7}$$
$$0 = \cos(\theta)\cos(\theta') + e^{i\phi}\sin(\theta)\sin(\theta'), \qquad \phi \equiv (\alpha_2 - \alpha_2') + \alpha_1'. \tag{3.5.8}$$

We may decompose this relation into the real part

$$0 = \cos(\theta)\cos(\theta') + \cos(\phi)\sin(\theta)\sin(\theta') \tag{3.5.9}$$

as well as the imaginary part

$$0 = \sin(\phi)\sin(\theta)\sin(\theta'). \tag{3.5.10}$$

We are trying to determine $\vec{u}$ in terms of the parameters of $\vec{v}$ (or vice versa) by making them orthogonal, without specializing to specific forms of $\vec{v}$. So we do not want to set $\sin\theta$ or $\sin\theta'$ to zero. But that means

$$(\alpha_2 - \alpha_2') + \alpha_1' = n\pi, \qquad n = 0, \pm 1, \pm 2, \dots . \tag{3.5.11}$$

In turn, we have for odd n,

$$0 = \cos(\theta)\cos(\theta') - \sin(\theta)\sin(\theta') = \cos(\theta + \theta'); \tag{3.5.12}$$

while for even n,

$$0 = \cos(\theta)\cos(\theta') + \sin(\theta)\sin(\theta') = \cos(\theta - \theta'). \tag{3.5.13}$$

Hence, for odd $n = \pm 1, \pm 3, \pm 5, \dots$,

$$\theta' = -\theta + \frac{m}{2}\pi, \qquad m = \pm 1, \pm 3, \pm 5, \pm 7, \dots; \tag{3.5.14}$$

and for even $n = 0, \pm 2, \pm 4, \dots$,

$$\theta' = \theta + \frac{m}{2}\pi. \tag{3.5.15}$$

At this point, our 2×2 unitary matrix takes one of the following four forms:

$$\widehat{U} = e^{i\gamma} \begin{bmatrix} \cos(\theta) & \mp e^{i\alpha_1'} \sin(\theta) \\ e^{i(\alpha_2' - \alpha_1')} \sin(\theta) & \pm e^{i\alpha_2'} \cos(\theta) \end{bmatrix} \tag{3.5.16}$$

or

$$\widehat{U} = e^{i\gamma} \begin{bmatrix} \cos(\theta) & \pm e^{i\alpha_1'} \sin(\theta) \\ e^{i(\alpha_2' - \alpha_1')} \sin(\theta) & \pm e^{i\alpha_2'} \cos(\theta) \end{bmatrix}. \tag{3.5.17}$$

Since $\alpha_{1,2}'$ are arbitrary, we may shift them by π to absorb/introduce an overall minus sign. That means we have

$$\widehat{U} = e^{i\gamma} \begin{bmatrix} \cos(\theta) & -e^{i\alpha_1'} \sin(\theta) \\ e^{i(\alpha_2' - \alpha_1')} \sin(\theta) & e^{i\alpha_2'} \cos(\theta) \end{bmatrix}, \qquad \gamma, \alpha_{1,2}', \theta \in \mathbb{R}. \tag{3.5.18}$$

YZ: The answer with $+$ in the $(1, 2)$ component is wrong. Why? Multiplying this by an appropriate phase factor, we obtain the general SU_2 matrix:

$$\widehat{U} = e^{i\gamma} \begin{bmatrix} e^{-i\beta} \cos(\theta) & -e^{i\alpha} \sin(\theta) \\ e^{-i\alpha} \sin(\theta) & e^{i\beta} \cos(\theta) \end{bmatrix}, \qquad \alpha, \beta, \gamma \in \mathbb{R}. \tag{3.5.19}$$

Problem 3.26. Special Unitary 2×2 Matrices: SU_2 Explain why the most general SU_2 matrix, with the 'S' $\equiv$ 'special' referring to an additional unit determinant $\det \widehat{U} = 1$ constraint, is given by eq. (3.5.19) with $\gamma = 0$:

$$\begin{aligned} \widehat{U} &= \begin{bmatrix} e^{-i\beta} \cos(\theta) & -e^{i\alpha} \sin(\theta) \\ e^{-i\alpha} \sin(\theta) & e^{i\beta} \cos(\theta) \end{bmatrix}, \qquad \gamma, \alpha_{1,2}', \theta \in \mathbb{R} \\ &= p_\mu \sigma^\mu; \end{aligned} \tag{3.5.20}$$

where

$$p_\mu = (\cos(\beta)\cos(\theta), -i\sin(\alpha)\sin(\theta), -i\cos(\alpha)\sin(\theta), -i\sin(\beta)\cos(\theta)) \,. \tag{3.5.21}$$

The $\{\sigma^\mu\}$ are the unit and the Pauli matrices in eq. (3.2.18). Explain using eq. (**??**) why

$$\eta^{\mu\nu} p_\mu p_\nu = 1. \tag{3.5.22}$$

$\square$

Problem 3.27. Relation Between SU$_2$ and SO$_2$ Matrices Show that the SU$_2$ matrix $\widehat{U}$ related to its SO$_2$ cousin $\widehat{O}$ via

$$\widehat{U} = e^{i\gamma} \begin{bmatrix} e^{-i\beta} & 0 \\ 0 & e^{i\alpha} \end{bmatrix} \cdot \widehat{O} \cdot \begin{bmatrix} 1 & 0 \\ 0 & e^{-i(\alpha-\beta)} \end{bmatrix}. \tag{3.5.23}$$

$\square$

4 Linear Algebra

4.1 Definition

Loosely speaking, the notion of a vector space – as the name suggests – amounts to abstracting the algebraic properties – addition of vectors, multiplication of a vector by a number, etc. – obeyed by the familiar $D \in \{1, 2, 3, \dots\}$ dimensional Euclidean space $\mathbb{R}^D$. We will discuss the linear algebra of vector spaces using Paul Dirac's bra-ket notation. This will not only help you understand the logical foundations of linear algebra and the matrix algebra you encountered earlier, it will also prepare you for the study of quantum theory, which is built entirely on the theory of both finite and infinite dimensional vector spaces.[6]

We will consider a vector space over complex numbers. A member of the vector space will be denoted as $|\alpha\rangle$; we will use the words "ket", "vector" and "state" interchangeably in what follows. We will allude to aspects of quantum theory, but point out everything we state here holds in a more general context; i.e., quantum theory is not necessary but merely an application – albeit a very important one for physics. For now α is just some arbitrary label, but later on it will often correspond to the eigenvalue of some linear operator. We may also use α as an enumeration label, where $|\alpha\rangle$ is the αth element in the collection of vectors. In quantum mechanics, a physical system is postulated to be completely described by some $|\alpha\rangle$ in a vector space, whose time evolution is governed by some Hamiltonian. (The latter is what Schrödinger's equation is about.)

Here is what defines a "vector space over complex numbers". It is a collection of states $\{|\alpha\rangle, |\beta\rangle, |\gamma\rangle, \dots\}$ endowed with the operations of addition and scalar multiplication subject to the following rules.

1. **Ax1: Addition** Any two vectors can be added to yield another vector

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle . \tag{4.1.1}$$

   Addition is commutative and associative:

$$|\alpha\rangle + |\beta\rangle = |\beta\rangle + |\alpha\rangle \tag{4.1.2}$$
$$|\alpha\rangle + (|\beta\rangle + |\gamma\rangle) = (|\alpha\rangle + |\beta\rangle) + |\gamma\rangle . \tag{4.1.3}$$

2. **Ax2: Additive identity (zero vector) and existence of inverse** There is a zero vector $|\text{zero}\rangle$ – which can be gotten by multiplying any vector by 0, i.e.,

$$0\,|\alpha\rangle = |\text{zero}\rangle \tag{4.1.4}$$

   – that acts as an additive identity.[7] Namely, adding $|\text{zero}\rangle$ to any vector returns the vector itself:

$$|\text{zero}\rangle + |\beta\rangle = |\beta\rangle . \tag{4.1.5}$$

[6]The material in this section of our notes was drawn heavily from the contents and problems provided in Chapter 1 of Sakurai's *Modern Quantum Mechanics*.

[7]In this section we will be careful and denote the zero vector as $|\text{zero}\rangle$. For the rest of the notes, whenever the context is clear, we will often use 0 to denote the zero vector.

For any vector $|\alpha\rangle$ there exists an additive inverse; if $+$ is the usual addition, then the inverse of $|\alpha\rangle$ is just $(-1)|\alpha\rangle$.

$$|\alpha\rangle + (-|\alpha\rangle) = |\text{zero}\rangle . \tag{4.1.6}$$

3. **Ax3: Multiplication by scalar** Any ket can be multiplied by an arbitrary complex number c to yield another vector

$$c|\alpha\rangle = |\gamma\rangle . \tag{4.1.7}$$

(In quantum theory, $|\alpha\rangle$ and $c|\alpha\rangle$ are postulated to describe the same system.) This multiplication is distributive with respect to both vector and scalar addition; if a and b are arbitrary complex numbers,

$$a(|\alpha\rangle + |\beta\rangle) = a|\alpha\rangle + a|\beta\rangle \tag{4.1.8}$$

$$(a+b)|\alpha\rangle = a|\alpha\rangle + b|\alpha\rangle . \tag{4.1.9}$$

Note: If you define a "vector space over scalars," where the scalars can be more general objects than complex numbers, then in addition to the above axioms, we have to add: (I) Associativity of scalar multiplication, where $a(b|\alpha\rangle) = (ab)|\alpha\rangle$ for any scalars a, b and vector $|\alpha\rangle$; (II) Existence of a scalar identity 1, where $1|\alpha\rangle = |\alpha\rangle$.

Examples The Euclidean space $\mathbb{R}^D$ itself, the space of D-tuples of real numbers

$$|\vec{a}\rangle \equiv (a^1, a^2, \dots, a^D), \tag{4.1.10}$$

with $+$ being the usual addition operation is, of course, *the* example of a vector space. We shall check explicitly that $\mathbb{R}^D$ does in fact satisfy all the above axioms. To begin, let

$$|\vec{v}\rangle = (v^1, v^2, \dots, v^D),$$

$$|\vec{w}\rangle = (w^1, w^2, \dots, w^D) \qquad \text{and} \tag{4.1.11}$$

$$|\vec{x}\rangle = (x^1, x^2, \dots, x^D) \tag{4.1.12}$$

be vectors in $\mathbb{R}^D$.

1. **Addition** Any two vectors can be added to yield another vector

$$|\vec{v}\rangle + |\vec{w}\rangle = (v^1 + w^1, \dots, v^D + w^D) \equiv |\vec{v} + \vec{w}\rangle . \tag{4.1.13}$$

Addition is commutative and associative because we are adding/subtracting the vectors component-by-component:

$$\begin{aligned} |\vec{v}\rangle + |\vec{w}\rangle &= |\vec{v} + \vec{w}\rangle = (v^1 + w^1, \dots, v^D + w^D) \\ &= (w^1 + v^1, \dots, w^D + v^D) \\ &= |\vec{w}\rangle + |\vec{v}\rangle = |\vec{w} + \vec{v}\rangle , \end{aligned} \tag{4.1.14}$$

$$\begin{aligned} |\vec{v}\rangle + |\vec{w}\rangle + |\vec{x}\rangle &= (v^1 + w^1 + x^1, \dots, v^D + w^D + x^D) \\ &= (v^1 + (w^1 + x^1), \dots, v^D + (w^D + x^D)) \\ &= ((v^1 + w^1) + x^1, \dots, (v^D + w^D) + x^D) \\ &= |\vec{v}\rangle + (|\vec{w}\rangle + |\vec{x}\rangle) = (|\vec{v}\rangle + |\vec{w}\rangle) + |\vec{x}\rangle = |\vec{v} + \vec{w} + \vec{x}\rangle . \end{aligned} \tag{4.1.15}$$

2. **Additive identity (zero vector) and existence of inverse** There is a zero vector $|\text{zero}\rangle$ – which can be gotten by multiplying any vector by 0, i.e.,

$$0 \, |\vec{v}\rangle = 0(v^1, \dots, v^D) = (0, \dots, 0) = |\text{zero}\rangle \tag{4.1.16}$$

– that acts as an additive identity. Namely, adding $|\text{zero}\rangle$ to any vector returns the vector itself:

$$|\text{zero}\rangle + |\vec{w}\rangle = (0, \dots, 0) + (w^1, \dots, w^D) = |\vec{w}\rangle \, . \tag{4.1.17}$$

For any vector $|\vec{x}\rangle$ there exists an additive inverse; in fact, the inverse of $|\vec{x}\rangle$ is just $(-1) \, |\vec{x}\rangle = |-\vec{x}\rangle$.

$$|\vec{x}\rangle + (- \, |\vec{x}\rangle) = (x^1, \dots, x^D) - (x^1, \dots, x^D) = |\text{zero}\rangle \, . \tag{4.1.18}$$

3. **Multiplication by scalar** Any ket can be multiplied by an arbitrary real number c to yield another vector

$$c \, |\vec{v}\rangle = c(v^1, \dots, v^D) = (cv^1, \dots, cv^D) \equiv |c\vec{v}\rangle \, . \tag{4.1.19}$$

This multiplication is distributive with respect to both vector and scalar addition; if a and b are arbitrary real numbers,

$$\begin{aligned} a(|\vec{v}\rangle + |\vec{w}\rangle) &= (av^1 + aw^1, av^2 + aw^2, \dots, av^D + aw^D) \\ &= |a\vec{v}\rangle + |a\vec{w}\rangle = a \, |\vec{v}\rangle + a \, |\vec{w}\rangle \, , \end{aligned} \tag{4.1.20}$$

$$\begin{aligned} (a + b) \, |\vec{x}\rangle &= (ax^1 + bx^1, \dots, ax^D + bx^D) \\ &= |a\vec{x}\rangle + |b\vec{x}\rangle = a \, |\vec{x}\rangle + b \, |\vec{x}\rangle \, . \end{aligned} \tag{4.1.21}$$

□

The following are some further examples of vector spaces.

1. The space of polynomials with complex coefficients.

2. The space of square integrable functions on $\mathbb{R}^D$ (where D is an arbitrary integer greater or equal to 1); i.e., all functions $f(\vec{x})$ such that $\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} |f(\vec{x})|^2 < \infty$.

3. The space of all homogeneous solutions to a linear (ordinary or partial) differential equation.

4. The space of $M \times N$ matrices of complex numbers, where M and N are arbitrary integers greater or equal to 1.

Problem 4.1. Prove that the examples in (1), (3), and (4) are indeed vector spaces, by running through the above axioms. □

Linear (in)dependence, Basis, Dimension Suppose we pick N vectors from a vector space, and find that one of them (say, $|N\rangle$) can be expressed as a *linear combination* (or, *superposition*) of the rest,

$$|N\rangle = \sum_{i=1}^{N-1} c^i |i\rangle \,, \tag{4.1.22}$$

where the $\{\chi^i\}$ are complex numbers. Then we say that this set of N vectors are linearly dependent. Equivalently, we may state that $|1\rangle$ through $|N\rangle$ are linearly dependent if a *non-trivial* superposition of them can be found to yield the zero vector:

$$\sum_{i=1}^{N} c^i |i\rangle = |\text{zero}\rangle \,, \qquad \exists\{\chi^i\}. \tag{4.1.23}$$

That equations (4.1.22) and (4.1.23) are equivalent, is because – by assumption, $c^N \neq 0$ – we can divide eq. (4.1.23) throughout by c^N; similarly, we may multiply eq. (4.1.22) by c^N.

Suppose we have picked D vectors $\{|1\rangle, |2\rangle, |3\rangle, \dots, |D\rangle\}$ such that they are linearly independent, i.e., no vector is a linear combination of any others, and suppose further that any arbitrary vector $|\alpha\rangle$ from the vector space can now be expressed as a superposition of these vectors

$$|\alpha\rangle = \sum_{i=1}^{D} c^i |i\rangle \,, \qquad \{\chi^i \in \mathbb{C}\}. \tag{4.1.24}$$

In other words, we now have a maximal number of linearly independent vectors – then, D is called the *dimension* of the vector space. The $\{|i\rangle \,|i = 1, 2, \dots, D\}$ is a complete set of *basis vectors*; and such a set of (basis) vectors is said to *span* the vector space.[8] It is worth reiterating, this is a maximal set because – if it were not, that would mean there is some additional vector $|\alpha\rangle$ that cannot be written as eq. (4.1.24).

Example For instance, for the D-tuple $|\vec{a}\rangle \equiv (a^1, \dots, a^D)$ from the real vector space of $\mathbb{R}^D$, we may choose

$$\begin{aligned} |1\rangle &= (1, 0, 0, \dots), \qquad |2\rangle = (0, 1, 0, 0, \dots), \\ |3\rangle &= (0, 0, 1, 0, 0, \dots), \qquad \dots \qquad |D\rangle = (0, 0, \dots, 0, 0, 1). \end{aligned} \tag{4.1.25}$$

Then, any arbitrary $|\vec{a}\rangle$ can be written as

$$|\vec{a}\rangle = (a^1, \dots, a^D) = \sum_{i=1}^{D} a^i |i\rangle \,. \tag{4.1.26}$$

The basis vectors are the $\{|i\rangle\}$ and the dimension is D. Additionally, if we define

$$|\vec{v}\rangle \equiv (1, 1, 0, \dots, 0) \,, \tag{4.1.27}$$
$$|\vec{w}\rangle \equiv (1, -1, 0, \dots, 0) \,, \tag{4.1.28}$$
$$|\vec{u}\rangle \equiv (1, 0, 0, \dots, 0) \,. \tag{4.1.29}$$

[8]The *span* of vectors $\{|1\rangle, \dots, |D\rangle\}$ is the space gotten by considering all possible linear combinations $\{\sum_{i=1}^{D} c^i |i\rangle \,| c^i \in \mathbb{C}\}$.

We see that $\{|\vec{v}\rangle, |\vec{w}\rangle\}$ are linearly independent – they are not proportional to each other – but $\{|\vec{v}\rangle, |\vec{w}\rangle, |\vec{u}\rangle\}$ are linearly dependent because

$$|\vec{u}\rangle = \frac{1}{2}|\vec{v}\rangle + \frac{1}{2}|\vec{w}\rangle. \tag{4.1.30}$$

Problem 4.2. Is the space of polynomials of complex coefficients of degree less than or equal to $(n \geq 1)$ a vector space? (Namely, this is the set of polynomials of the form $P_n(x) = c_0 + c_1 x + \dots + c_n x^n$, where the $\{c_i | i = 1, 2, \dots, n\}$ are complex numbers.) If so, write down a set of basis vectors. What is its dimension? Answer the same questions for the space of $D \times D$ matrices of complex numbers. □

Vector space within a vector space Before moving on to inner products, let us note that a subset of a vector space is itself a vector space – i.e., a subspace of the larger vector space – if it is closed under addition and multiplication by complex numbers. Closure means, if $|\alpha\rangle$ and $|\beta\rangle$ are members of the subset, then $c_1|\alpha\rangle + c_2|\beta\rangle$ are also members of the same subset for any pair of complex numbers $c_{1,2}$.

In principle, to understand why closure guarantees the subset is a subspace, we need to run through all the axioms in Ax1 through Ax3 above. But a brief glance tells us, the axioms in Ax1 and Ax3 are automatically satisfied when closure is obeyed. Furthermore, closure means $-|\alpha\rangle$ (i.e., the inverse of $|\alpha\rangle$) must lie within the subset whenever $|\alpha\rangle$ does, since the former is -1 times $|\alpha\rangle$. And that in turn teaches us, the zero vector gotten from superposing $|\alpha\rangle + (-1)|\alpha\rangle = |\text{zero}\rangle$ must also lie within the subset. Namely, the set of axioms in Ax2 are, too, satisfied.

Examples The space of vectors $\{|\vec{a}\rangle = (a^1, a^2)\}$ in a 2D real space is a subspace of the 3D counterpart $\{|\vec{a}\rangle = (a^1, a^2, a^3)\}$; the former can be thought of as the latter with the third component held fixed, $a^3 =$ same constant for all vectors. It is easy to check, the 2D vectors are closed under linear combination.

We have already noted that the set of $M \times M$ matrices form a vector space. Therefore, the subset of Hermitian matrices over real numbers; or (anti)symmetric matrices over complex numbers; must form subspaces. For, the superposition of Hermitian matrices $\{\widehat{H}_1, \widehat{H}_2, \dots\}$ with real coefficients yield another Hermitian matrix

$$\left(c_1\widehat{H}_1 + c_2\widehat{H}_2\right)^\dagger = c_1\widehat{H}_1 + c_2\widehat{H}_2, \qquad c_{1,2} \in \mathbb{R}; \tag{4.1.31}$$

whereas the superposition of (anti)symmetric ones with complex coefficients return another (anti)symmetric matrix:

$$\left(c_1\widehat{H}_1 + c_2\widehat{H}_2\right)^{\mathrm{T}} = c_1\widehat{H}_1 + c_2\widehat{H}_2, \qquad c_{1,2} \in \mathbb{C}, \widehat{H}_{1,2}^{\mathrm{T}} = \widehat{H}_{1,2}, \tag{4.1.32}$$

$$\left(c_1\widehat{H}_1 + c_2\widehat{H}_2\right)^{\mathrm{T}} = -(c_1\widehat{H}_1 + c_2\widehat{H}_2), \qquad c_{1,2} \in \mathbb{C}, \widehat{H}_{1,2}^{\mathrm{T}} = -\widehat{H}_{1,2}. \tag{4.1.33}$$

4.2 Inner Products

In Euclidean D-space $\mathbb{R}^D$ the ordinary dot product, between the real vectors $|\vec{a}\rangle \equiv (a^1, \dots, a^D)$ and $|\vec{b}\rangle \equiv (b^1, \dots, b^D)$, is defined as

$$\vec{a} \cdot \vec{b} \equiv \sum_{i=1}^{D} a^i b^i = \delta_{ij} a^i b^j. \tag{4.2.1}$$

The inner product of linear algebra is again an abstraction of this notion of the dot product, where the analog of $\vec{a} \cdot \vec{b}$ will be denoted as $\langle \vec{a} | \vec{b} \rangle$. Like the dot product for Euclidean space, the inner product will allow us to define a notion of the length of vectors and angles between different vectors.

Dual/'bra' space Given a vector space, an inner product is defined by first introducing a *dual space* (aka *bra space*) to this vector space. Specifically, given a vector $|\alpha\rangle$ we write its dual as $\langle\alpha|$. We also introduce the notation

$$|\alpha\rangle^\dagger \equiv \langle\alpha| \,. \tag{4.2.2}$$

Importantly, for some complex number c, the dual of $c\,|\alpha\rangle$ is

$$(c\,|\alpha\rangle)^\dagger \equiv c^* \,\langle\alpha| \,. \tag{4.2.3}$$

Moreover, for complex numbers a and b,

$$(a\,|\alpha\rangle + b\,|\beta\rangle)^\dagger \equiv a^* \,\langle\alpha| + b^* \,\langle\beta| \,. \tag{4.2.4}$$

Since there is a one-to-one correspondence between the vector space and its dual, observe that this dual space is itself a vector space.

Now, the primary purpose of these dual vectors is that they act on vectors of the original vector space to return a complex number:

$$\langle\alpha|\,\beta\rangle \in \mathbb{C}. \tag{4.2.5}$$

You will soon see a few examples below.

Definition. The inner product is now defined by the following properties. For an arbitrary complex number c,

$$\langle\alpha|\,(|\beta\rangle + |\gamma\rangle) = \langle\alpha|\,\beta\rangle + \langle\alpha|\,\gamma\rangle \tag{4.2.6}$$
$$\langle\alpha|\,(c\,|\beta\rangle) = c\,\langle\alpha|\,\beta\rangle \tag{4.2.7}$$
$$\langle\alpha|\,\beta\rangle^* = \overline{\langle\alpha|\,\beta\rangle} = \langle\beta|\,\alpha\rangle \tag{4.2.8}$$
$$\langle\alpha|\,\alpha\rangle \geq 0 \tag{4.2.9}$$

and

$$\langle\alpha|\,\alpha\rangle = 0 \tag{4.2.10}$$

if and only if $|\alpha\rangle$ is the zero vector.

Some words on notation here. Especially in the math literature, the bra-ket notation is not used. There, the inner product is often denoted by (α, β), where α and β are vectors. Then the defining properties of the inner product would read instead

$$(\alpha, b\beta + c\gamma) = b(\alpha, \beta) + c(\alpha, \gamma), \qquad \text{(for any constants } b \text{ and } c\text{)}, \tag{4.2.11}$$
$$(\alpha, \beta)^* = \overline{(\alpha, \beta)} = (\beta, \alpha), \tag{4.2.12}$$
$$(\alpha, \alpha) \geq 0; \tag{4.2.13}$$

and

$$(\alpha, \alpha) = 0 \tag{4.2.14}$$

if and only if α is the zero vector. In addition, notice from equations (4.2.11) and (4.2.12) that

$$(b\beta + c\gamma, \alpha) = b^*(\beta, \alpha) + c^*(\gamma, \alpha). \tag{4.2.15}$$

Example: Dot Product We may readily check that the ordinary dot product does, of course, satisfy all the axioms of the inner product. Firstly, let us denote

$$|\vec{a}\rangle = (a^1, a^2, \dots, a^D), \tag{4.2.16}$$
$$\left|\vec{b}\right\rangle = (b^1, b^2, \dots, b^D), \tag{4.2.17}$$
$$|\vec{c}\rangle = (c^1, c^2, \dots, c^D); \tag{4.2.18}$$

where all the components $a^i, b^i, \dots$ are now real. Next, define

$$\langle\vec{a}|\,\vec{b}\rangle = \vec{a}\cdot\vec{b} = \vec{a}^{\mathrm{T}}\vec{b}. \tag{4.2.19}$$

Then we may start with eq. (4.2.6): $\langle\vec{a}|\,(|b\rangle + |c\rangle) = \langle\vec{a}|\,\vec{b}+\vec{c}\rangle = \vec{a}\cdot\vec{b} + \vec{a}\cdot\vec{c} = \langle\vec{a}|\,\vec{b}\rangle + \langle\vec{a}|\,\vec{c}\rangle$. Second, $\langle\vec{a}|\,(c\left|\vec{b}\right\rangle) = \langle\vec{a}|\,c\vec{b}\rangle = c(\vec{a}\cdot\vec{b}) = c\,\langle\vec{a}|\,\vec{b}\rangle$. Third, $\langle\vec{a}|\,\vec{b}\rangle = \vec{a}\cdot\vec{b} = \vec{b}\cdot\vec{a} = (\vec{b}\cdot\vec{a})^* = \left\langle\vec{b}\middle|\,\vec{a}\right\rangle$. Fourth, $\langle\vec{a}|\,\vec{a}\rangle = \vec{a}\cdot\vec{a} = \sum_i (a^i)^2$ is a sum of squares and therefore non-negative. Finally, because $\langle\vec{a}|\,\vec{a}\rangle$ is a sum of squares the only way it can be zero is for every component of $\vec{a}$ to be zero; moreover, if $\vec{a}$ is $\vec{0}$ then $\langle\vec{a}|\,\vec{a}\rangle = 0$.

Problem 4.3. Prove that $\langle\alpha|\,\alpha\rangle$ is a real number. Hint: See eq. (4.2.8) □

The following are examples of inner products.

- Take the D-tuple of complex numbers $|\vec{\alpha}\rangle \equiv (\alpha^1, \dots, \alpha^D)$ and $\left|\vec{\beta}\right\rangle \equiv (\beta^1, \dots, \beta^D)$; and define the inner product to be

$$\langle\vec{\alpha}|\,\vec{\beta}\rangle \equiv \sum_{i=1}^{D} (\alpha^i)^*\beta^i = \delta_{ij}(\alpha^i)^*\beta^j = \vec{\alpha}^\dagger\vec{\beta}. \tag{4.2.20}$$

- Consider the space of $D\times D$ complex matrices. Consider two such matrices $\widehat{X}$ and $\widehat{Y}$ and define their inner product to be

$$\left\langle\widehat{X}\middle|\,\widehat{Y}\right\rangle \equiv \mathrm{Tr}\left[\widehat{X}^\dagger\widehat{Y}\right]. \tag{4.2.21}$$

  Here, Tr means the matrix trace – i.e., summation over the diagonal components –

$$\mathrm{Tr}\,[M] \equiv \sum_{i=1}^{D} M^i{}_i \equiv M^i{}_i; \tag{4.2.22}$$

  and $\widehat{X}^\dagger$ is the *matrix adjoint* of $\widehat{X}$.

- Consider the space of polynomials. Suppose $|f\rangle$ and $|g\rangle$ are two such polynomials of the vector space. Then

$$\langle f|\, g\rangle \equiv \int_{-1}^{1} \mathrm{d}x f(x)^* g(x) \tag{4.2.23}$$

  defines an inner product. Here, $f(x)$ and $g(x)$ indicates the polynomials are expressed in terms of the variable x.

Problem 4.4. Prove the above examples are indeed inner products. □

Problem 4.5. Prove the Cauchy-Schwarz inequality:

$$\langle\alpha|\,\alpha\rangle\,\langle\beta|\,\beta\rangle \geq |\langle\alpha|\,\beta\rangle|^2 . \tag{4.2.24}$$

The analogy in Euclidean space is $|\vec{x}|^2|\vec{y}|^2 \geq |\vec{x}\cdot\vec{y}|^2$. Hint: Start with

$$(\langle\alpha| + c^*\,\langle\beta|)\,(|\alpha\rangle + c\,|\beta\rangle) \geq 0. \tag{4.2.25}$$

for any complex number c. (Why is this true?) Now choose an appropriate c to prove the Schwarz inequality. □

Orthogonality Just as we would say two real vectors in $\mathbb{R}^D$ are perpendicular (aka orthogonal) when their dot product is zero, we may now define two vectors $|\alpha\rangle$ and $|\beta\rangle$ in a vector space to be orthogonal when their inner product is zero:

$$\langle\alpha|\,\beta\rangle = 0 = \langle\beta|\,\alpha\rangle . \tag{4.2.26}$$

We also call the positive square root $\sqrt{\langle\alpha|\,\alpha\rangle}$ the *norm* of the vector $|\alpha\rangle$; recall, in Euclidean space, the analogous $|\vec{x}| = \sqrt{\vec{x}\cdot\vec{x}}$. Given any vector $|\alpha\rangle$ that is not the zero vector, we can always construct a vector from it that is of unit length,

$$|\widehat{\alpha}\rangle \equiv \frac{|\alpha\rangle}{\sqrt{\langle\alpha|\,\alpha\rangle}} \qquad \Rightarrow \qquad \langle\widehat{\alpha}|\,\widehat{\alpha}\rangle = 1. \tag{4.2.27}$$

Orthonormal Basis Suppose we are given a set of basis vectors $\{|i'\rangle\}$ of a vector space. Through what is known as the Gram-Schmidt process, one can always build from them a set of orthonormal basis vectors $\{|i\rangle\}$; where every basis vector has unit norm and is orthogonal to every other basis vector,

$$\langle i|\, j\rangle = \delta^i_j. \tag{4.2.28}$$

As you will see, just as vector calculus problems are often easier to analyze when you choose an orthogonal coordinate system, linear algebra problems are often easier to study when you use an orthonormal basis to describe your vector space. If $\{|i\rangle\}$ form an orthonormal basis, any vector $|\gamma\rangle$ should be expandable as

$$|\gamma\rangle = \sum_i \widehat{\gamma}^i\,|i\rangle\,, \tag{4.2.29}$$

where the $\{\widehat{\gamma}^i\}$ are complex numbers. Projecting both sides with $\langle j|$ and exploiting the orthonormality condition $\langle j| i\rangle = \delta^j_i$,

$$\langle j| \gamma\rangle = \widehat{\gamma}^j. \tag{4.2.30}$$

This in turn means

$$|\gamma\rangle = \sum_i |j\rangle \langle j| \gamma\rangle . \tag{4.2.31}$$

Compare this with the vector calculus expression $\vec{v} = \sum_i v^i \widehat{e}_i$, where $\{\widehat{e}_i\}$ are the unit vectors in the i-th direction.

Problem 4.6. Suppose $|\alpha\rangle$ and $|\beta\rangle$ are linearly dependent – they are scalar multiples of each other. However, their inner product is zero. What are $|\alpha\rangle$ and $|\beta\rangle$? □

Problem 4.7. Projection Process If $\vec{v}$ and $\vec{w}$ are vectors in $\mathbb{R}^D$, verify that

$$\vec{v}_\perp \equiv \vec{v} - \frac{\vec{w}}{|\vec{w}|} (\vec{w} \cdot \vec{v}) \tag{4.2.32}$$

is perpendicular to $\vec{w}$. In index notation, if we denote $\widehat{w}^i \equiv w^i/|\vec{w}|$,

$$v^i_\perp \equiv v^i - \widehat{w}^i(\widehat{w}^j v^j) = \left(\delta^{ij} - \widehat{w}^j \widehat{w}^j\right) v^j. \tag{4.2.33}$$

(In Chapter §(9.7) below, we shall see that $\delta^{ij} - \widehat{w}^i\widehat{w}^j$ describes the geometry of the $D-1$ space orthogonal to $\widehat{w}$.) Write down the corresponding $\vec{w}_\perp$, the component of $\vec{w}$ perpendicular to $\vec{v}$.

Now, let $\{|1\rangle, |2\rangle, \dots, |N\rangle\}$ be a set of N orthonormal vectors. Let $|\alpha\rangle$ be an arbitrary vector lying in the same vector space. Show that the following vector constructed from $|\alpha\rangle$ is orthogonal to all the $\{|i\rangle\}$.

$$\left|\alpha^\perp\right\rangle \equiv |\alpha\rangle - \sum_{j=1}^{N} |j\rangle \langle j| \alpha\rangle \tag{4.2.34}$$

This result is key to the following Gram-Schmidt process. □

Gram-Schmidt Let $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_D\rangle\}$ be a set of D linearly independent vectors that spans some vector space. The Gram-Schmidt process is an iterative algorithm, based on the observation in eq. (4.2.34), to generate from it a set of orthonormal set of basis vectors.

1. Take the first vector $|\alpha_1\rangle$ and normalize it to unit length:

$$|\widehat{\alpha}_1\rangle = \frac{|\alpha_1\rangle}{\sqrt{\langle \alpha_1| \alpha_1\rangle}}. \tag{4.2.35}$$

2. Take the second vector $|\alpha_2\rangle$ and project out $|\widehat{\alpha}_1\rangle$:

$$\left|\alpha_2^\perp\right\rangle \equiv |\alpha_2\rangle - |\widehat{\alpha}_1\rangle \langle \widehat{\alpha}_1| \alpha_2\rangle , \tag{4.2.36}$$

and normalize it to unit length

$$|\widehat{\alpha}_2\rangle \equiv \frac{\left|\alpha_2^\perp\right\rangle}{\sqrt{\left\langle \alpha_2^\perp \middle| \alpha_2^\perp\right\rangle}}. \tag{4.2.37}$$

3. Take the third vector $|\alpha_3\rangle$ and project out $|\widehat{\alpha}_1\rangle$ and $|\widehat{\alpha}_2\rangle$:

$$\left|\alpha_3^\perp\right\rangle \equiv |\alpha_3\rangle - |\widehat{\alpha}_1\rangle \langle\widehat{\alpha}_1|\, \alpha_3\rangle - |\widehat{\alpha}_2\rangle \langle\widehat{\alpha}_2|\, \alpha_3\rangle\,, \tag{4.2.38}$$

then normalize it to unit length

$$|\widehat{\alpha}_3\rangle \equiv \frac{\left|\alpha_3^\perp\right\rangle}{\sqrt{\left\langle\alpha_3^\perp\right|\left.\alpha_3^\perp\right\rangle}}. \tag{4.2.39}$$

4. Repeat ... Take the ith vector $|\alpha_i\rangle$ and project out $|\widehat{\alpha}_1\rangle$ through $|\widehat{\alpha}_{i-1}\rangle$:

$$\left|\alpha_i^\perp\right\rangle \equiv |\alpha_i\rangle - \sum_{j=1}^{i-1} |\widehat{\alpha}_j\rangle \langle\widehat{\alpha}_j|\, \alpha_i\rangle\,, \tag{4.2.40}$$

then normalize it to unit length

$$|\widehat{\alpha}_i\rangle \equiv \frac{\left|\alpha_i^\perp\right\rangle}{\sqrt{\left\langle\alpha_i^\perp\right|\left.\alpha_i^\perp\right\rangle}}. \tag{4.2.41}$$

By construction, $|\widehat{\alpha}_i\rangle$ will be orthogonal to $|\widehat{\alpha}_1\rangle$ through $|\widehat{\alpha}_{i-1}\rangle$. Therefore, at the end of the process, we will have D mutually orthogonal and unit norm vectors. Because they are orthogonal they are linearly independent – hence, we have succeeded in constructing an orthonormal set of basis vectors.

Example Here is a simple example in 3D Euclidean space endowed with the usual dot product. Let us have

$$|\alpha_1\rangle \doteq (2,0,0), \qquad |\alpha_2\rangle \doteq (1,1,1), \qquad |\alpha_3\rangle \doteq (1,0,1). \tag{4.2.42}$$

You can check that these vectors are linearly independent by taking the determinant of the 3×3 matrix formed from them. Alternatively, the fact that they generate a set of basis vectors from the Gram-Schmidt process also implies they are linearly independent.

Normalizing $|\alpha_1\rangle$ to unity,

$$|\widehat{\alpha}_1\rangle = \frac{|\alpha_1\rangle}{\sqrt{\langle\alpha_1|\,\alpha_1\rangle}} = \frac{(2,0,0)}{2} = (1,0,0). \tag{4.2.43}$$

Next we project out $|\widehat{\alpha}_1\rangle$ from $|\alpha_2\rangle$.

$$\left|\alpha_2^\perp\right\rangle = |\alpha_2\rangle - |\widehat{\alpha}_1\rangle \langle\widehat{\alpha}_1|\, \alpha_2\rangle = (1,1,1) - (1,0,0)(1+0+0) = (0,1,1). \tag{4.2.44}$$

Then we normalize it to unit length.

$$|\widehat{\alpha}_2\rangle = \frac{\left|\alpha_2^\perp\right\rangle}{\sqrt{\left\langle\alpha_2^\perp\right|\left.\alpha_2^\perp\right\rangle}} = \frac{(0,1,1)}{\sqrt{2}}. \tag{4.2.45}$$

Next we project out $|\widehat{\alpha}_1\rangle$ and $|\widehat{\alpha}_2\rangle$ from $|\alpha_3\rangle$.

$$\begin{aligned}
\left|\alpha_3^\perp\right\rangle &= |\alpha_3\rangle - |\widehat{\alpha}_1\rangle \langle\widehat{\alpha}_1|\,\alpha_3\rangle - |\widehat{\alpha}_2\rangle \langle\widehat{\alpha}_2|\,\alpha_3\rangle \\
&= (1,0,1) - (1,0,0)(1+0+0) - \frac{(0,1,1)}{\sqrt{2}}\frac{0+0+1}{\sqrt{2}} \\
&= (1,0,1) - (1,0,0) - \frac{(0,1,1)}{2} = \left(0,-\frac{1}{2},\frac{1}{2}\right).
\end{aligned} \tag{4.2.46}$$

Then we normalize it to unit length.

$$|\widehat{\alpha}_3\rangle = \frac{\left|\alpha_3^\perp\right\rangle}{\sqrt{\left\langle\alpha_3^\perp\middle|\,\alpha_3^\perp\right\rangle}} = \frac{(0,-1,1)}{\sqrt{2}}. \tag{4.2.47}$$

To sum: you can check that

$$|\widehat{\alpha}_1\rangle = (1,0,0), \qquad |\widehat{\alpha}_2\rangle = \frac{(0,1,1)}{\sqrt{2}}, \qquad |\widehat{\alpha}_3\rangle = \frac{(0,-1,1)}{\sqrt{2}}, \tag{4.2.48}$$

are mutually perpendicular and of unit length.

Problem 4.8. Consider the space of polynomials with complex coefficients. Let the inner product be

$$\langle f|\,g\rangle \equiv \int_{-1}^{+1} \mathrm{d}x f(x)^* g(x). \tag{4.2.49}$$

Starting from the set $\{|0\rangle = 1, |1\rangle = x, |2\rangle = x^2\}$, construct from them a set of orthonormal basis vectors spanning the subspace of polynomials of degree equal to or less than 2. Compare your results with the Legendre polynomials

$$P_\ell(x) \equiv \frac{1}{2^\ell \ell!}\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\left(x^2-1\right)^\ell, \qquad \ell = 0,1,2. \tag{4.2.50}$$

$\square$

Orthogonality and Linear independence. We close this subsection with an observation. If a set of non-zero kets $\{|i\rangle\,|i = 1,2,\dots,N-1,N\}$ are orthogonal, then they are necessarily linearly independent. This can be proved readily by contradiction. Suppose these kets were linearly dependent. Then it must be possible to find non-zero complex numbers $\{C^i\}$ such that

$$\sum_{i=1}^{N} C^i\,|i\rangle = 0. \tag{4.2.51}$$

If we now act $\langle j|$ on this equation, for any $j \in \{1,2,3,\dots,N\}$,

$$\sum_{i=1}^{N} C^i\,\langle j|\,i\rangle = \sum_{i=1}^{N} C^i\delta_{ij}\,\langle j|\,j\rangle = C^j\,\langle j|\,j\rangle = 0. \tag{4.2.52}$$

That means all the $\{C^j | j = 1, 2, \dots, N\}$ are in fact zero.

A simple application of this observation is, if you have found D mutually orthogonal kets $\{|i\rangle\}$ in a D dimensional vector space, then these kets form a basis. By normalizing them to unit length, you'd have obtained an orthonormal basis. Such an example is that of the Pauli matrices $\{\sigma^\mu | \mu = 0, 1, 2, 3\}$ in eq. (3.2.18). The vector space of 2×2 complex matrices is 4-dimensional, since there are 4 independent components. Moreover, we have already seen that the trace $\mathrm{Tr}\left[X^\dagger Y\right]$ is one way to define an inner product of matrices X and Y. Since

$$\frac{1}{2}\mathrm{Tr}\left[(\sigma^\mu)^\dagger \sigma^\nu\right] = \frac{1}{2}\mathrm{Tr}\left[\sigma^\mu \sigma^\nu\right] = \delta^{\mu\nu}, \qquad \mu, \nu \in \{0, 1, 2, 3\}, \tag{4.2.53}$$

that means, by the argument just given, the 4 Pauli matrices $\{\sigma^\mu\}$ form an orthogonal set of basis vectors for the vector space of complex 2×2 matrices. That means it must be possible to choose $\{p_\mu\}$ such that the superposition $p_\mu \sigma^\mu$ is equal to any given 2×2 complex matrix A. In fact,

$$p_\mu \sigma^\mu = A \qquad \Leftrightarrow \qquad p_\mu = \frac{1}{2}\mathrm{Tr}\left[\sigma^\mu A\right]. \tag{4.2.54}$$

In quantum mechanics and quantum field theory, these $\{\sigma^\mu\}$ are fundamental to describing spinors and spin$-1/2$ systems.

Problem 4.9. SU_2 and Pauli Matrices Use eq. (4.2.54) to compute the coefficients $\{p_\mu\}$ in the Pauli-matrices expansion of the general SU_2 matrix in eq. (3.5.20). ☐

4.3 Linear Operators

4.3.1 Definitions and Fundamental Concepts

In quantum theory, a physical observable is associated with a (Hermitian) linear operator acting on the vector space. What defines a linear operator? If A is one, it is primarily defined by *how* it acts from the left on a vector to return another vector

$$A |\alpha\rangle = |\alpha'\rangle. \tag{4.3.1}$$

In other words, if you can tell me what the 'output' $|\alpha'\rangle$ is, after A acts on *any* vector of the vector space $|\alpha\rangle$ – you'd have defined A itself. But that's not all – linearity also means, for otherwise arbitrary operators A and B and complex numbers c and d,

$$\begin{aligned}(c\, A + d\, B) |\alpha\rangle &= c\, A |\alpha\rangle + d\, B |\alpha\rangle \\ A(c |\alpha\rangle + d |\beta\rangle) &= c\, A |\alpha\rangle + d\, A |\beta\rangle.\end{aligned} \tag{4.3.2}$$

If X and Y are both linear operators, since $Y |\alpha\rangle$ is a vector, we can apply X to it to obtain another vector, $X(Y |\alpha\rangle)$. This means we ought to be able to multiply operators, for e.g., XY. We will assume this multiplication is associative, namely

$$XYZ = (XY)Z = X(YZ). \tag{4.3.3}$$

Identity The identity operator obeys

$$\mathbb{I}\left|\gamma\right\rangle = \left|\gamma\right\rangle \qquad \text{for all } \left|\gamma\right\rangle. \tag{4.3.4}$$

Inverse The inverse of the operator X is still defined as one that obeys

$$X^{-1}X = XX^{-1} = \mathbb{I}. \tag{4.3.5}$$

Strictly speaking, we need to distinguish between the left and right inverse, but in finite dimensional vector spaces, they are the same object.

Adjoint Next, let us observe that an operator always acts on a bra from the right, and returns another bra,

$$\left\langle\alpha\right| A = \left\langle\alpha'\right|. \tag{4.3.6}$$

The reason is that a bra is something that acts linearly on an arbitrary vector and returns a complex number. Since that is what $\left\langle\alpha\right| A$ does, it must therefore some bra 'state'.

To formalize this further, we shall denote the adjoint of the linear operator X, namely $X^\dagger$, by taking the † of the ket $X^\dagger\left|\alpha\right\rangle$ in the following way:

$$(X^\dagger\left|\alpha\right\rangle)^\dagger = \left\langle\alpha\right| X. \tag{4.3.7}$$

If $\left|\alpha\right\rangle$ and $\left|\beta\right\rangle$ are arbitrary states,

$$\left\langle\beta\left|X\right|\alpha\right\rangle = \left\langle\beta\right|(X\left|\alpha\right\rangle) = \left(X^\dagger\left|\beta\right\rangle\right)^\dagger\left|\alpha\right\rangle. \tag{4.3.8}$$

In words: Given a linear operator X, its adjoint $X^\dagger$ is defined as the operator that – after acting upon $\left|\beta\right\rangle$ – would yield an inner product $(X^\dagger\left|\beta\right\rangle)^\dagger\left|\alpha\right\rangle$ which is equal to $\left\langle\beta\right|(X\left|\alpha\right\rangle)$. As we shall see below, an equivalent manner to define the adjoint is either

$$\overline{\left\langle\alpha\right| X\left|\beta\right\rangle} = \left\langle\beta\right| X^\dagger\left|\alpha\right\rangle \tag{4.3.9}$$

$$\text{or} \qquad \overline{\left\langle\alpha\right| X^\dagger\left|\beta\right\rangle} = \left\langle\beta\right| X\left|\alpha\right\rangle. \tag{4.3.10}$$

Why such a definition yields a unique operator $X^\dagger$ would require some explanation; in a similar vein, we shall see below that,

$$(X^\dagger)^\dagger = X. \tag{4.3.11}$$

Hence, we could also have began with the definition

$$(X\left|\alpha\right\rangle)^\dagger = \left\langle\alpha\right| X^\dagger. \tag{4.3.12}$$

In the math literature, where α and β denote the states and X is still some linear operator, the latter's adjoint is expressed through the inner product as

$$(\beta, X\alpha) = \left(X^\dagger\beta, \alpha\right). \tag{4.3.13}$$

Problem 4.10. Prove that

$$(XY)^\dagger = Y^\dagger X^\dagger. \tag{4.3.14}$$

Hint: take the adjoint of $(XY)^\dagger\left|\alpha\right\rangle$ and $Y^\dagger(X^\dagger\left|\alpha\right\rangle)$. $\square$

Eigenvectors and eigenvalues An eigenvector of some linear operator A is a vector that, when acted upon by A, returns the vector itself multiplied by a complex number a:

$$X|a\rangle = a|a\rangle . \tag{4.3.15}$$

This number a is called the eigenvalue of A.

Remark The eigenvector is not unique, in that we may always multiply it by an arbitrary complex number z and still obtain an eigenvector:

$$X(z|a\rangle) = a(z|a\rangle). \tag{4.3.16}$$

In quantum mechanics we require the state to be normalized to unity, i.e., $\langle a|a\rangle = 1 = \langle a|z^*z|a\rangle = |z|^2\langle a|a\rangle$. This $|z| = \pm 1$ constraint implies that unit norm eigenvectors may differ by a phase.

$$X\left(e^{i\theta}|a\rangle\right) = a(e^{i\theta}|a\rangle), \qquad \theta \in \mathbb{R}. \tag{4.3.17}$$

Ket-bra operator Notice that the product $|\alpha\rangle\langle\beta|$ can be considered a linear operator. To see this, we apply it on some arbitrary vector $|\gamma\rangle$ and observe it returns the vector $|\alpha\rangle$ multiplied by a complex number describing the projection of $|\gamma\rangle$ on $|\beta\rangle$,

$$(|\alpha\rangle\langle\beta|)|\gamma\rangle = |\alpha\rangle(\langle\beta|\gamma\rangle) = (\langle\beta|\gamma\rangle)\cdot|\alpha\rangle, \tag{4.3.18}$$

as long as we assume these products are associative. It obeys the following "linearity" rules. If $|\alpha\rangle\langle\beta|$ and $|\alpha'\rangle\langle\beta'|$ are two different ket-bra operators,

$$(|\alpha\rangle\langle\beta| + |\alpha'\rangle\langle\beta'|)|\gamma\rangle = |\alpha\rangle\langle\beta|\gamma\rangle + |\alpha'\rangle\langle\beta'|\gamma\rangle; \tag{4.3.19}$$

and for complex numbers c and d,

$$|\alpha\rangle\langle\beta|(c|\gamma\rangle + d|\gamma'\rangle) = c\;|\alpha\rangle\langle\beta|\gamma\rangle + d\;|\alpha\rangle\langle\beta|\gamma'\rangle. \tag{4.3.20}$$

Problem 4.11. Show that

$$(|\alpha\rangle\langle\beta|)^\dagger = |\beta\rangle\langle\alpha|. \tag{4.3.21}$$

Hint: Act $(|\alpha\rangle\langle\beta|)^\dagger$ on an arbitrary vector, and then take its adjoint. □

Projection operator The special case $|\alpha\rangle\langle\alpha|$ acting on any vector $|\gamma\rangle$ will return $|\alpha\rangle\langle\alpha|\gamma\rangle$. Thus, we can view it as a projection operator – it takes an arbitrary vector and extracts the portion of it "parallel" to $|\alpha\rangle$.

Identity Operator and Completeness Relations We will now see that (square) matrices can be viewed as representations of linear operators on a vector space. Let $\{|i\rangle\}$ denote the basis orthonormal vectors of the vector space, which obey $\langle i|j\rangle = \delta^i_j$. We may begin with eq. (4.2.31),

$$|\gamma\rangle = \sum_i |i\rangle\langle i|\gamma\rangle = \left(\sum_i |i\rangle\langle i|\right)|\gamma\rangle. \tag{4.3.22}$$

Since $|\gamma\rangle$ was arbitrary, we have identified the identity operator as

$$\mathbb{I} = \sum_i |i\rangle \langle i| \,. \tag{4.3.23}$$

This is also often known as a completeness relation: summing over the ket-bra projection operators built out of the orthonormal basis vectors of a vector space returns the unit (aka identity) operator. $\mathbb{I}$ acting on any vector yields the same vector.

Representations, Vector components, Matrix elements Once a set of orthonormal basis vectors are chosen, notice from the expansion in eq. (4.3.22), that to specify a vector $|\gamma\rangle$ all we need to do is to specify the complex numbers $\{\langle i|\gamma\rangle\}$. These can be arranged as a column vector; if the dimension of the vector space is D, then

$$|\gamma\rangle \doteq \begin{bmatrix} \langle 1|\gamma\rangle \\ \langle 2|\gamma\rangle \\ \langle 3|\gamma\rangle \\ \dots \\ \langle D|\gamma\rangle \end{bmatrix}. \tag{4.3.24}$$

The $\doteq$ is not quite an equality; rather it means "represented by," in that this column vector contains as much information as eq. (4.3.22), provided the orthonormal basis vectors are known.

We may also express an arbitrary bra through a superposition of the basis bras $\{\langle i|\}$, using the adjoint of eq. (4.3.23).

$$\langle\alpha| = \sum_i \langle\alpha|i\rangle \langle i| \,. \tag{4.3.25}$$

(According to eq. (4.3.23), this is simply $\langle\alpha|\mathbb{I}$.) In this case, the coefficients $\{\langle\alpha|i\rangle\}$ may be arranged as a row vector:

$$\langle\alpha| \doteq \begin{bmatrix} \langle\alpha|1\rangle & \langle\alpha|2\rangle & \dots & \langle\alpha|D\rangle \end{bmatrix}. \tag{4.3.26}$$

Inner products Let us consider the inner product $\langle\alpha|\gamma\rangle$. By inserting the completeness relation in eq. (4.3.23), we obtain

$$\langle\alpha|\gamma\rangle = \sum_i \langle\alpha|i\rangle \langle i|\gamma\rangle = \delta_{ij} \left(\widehat{\alpha}^i\right)^* \widehat{\gamma}^j = \widehat{\alpha}^\dagger \widehat{\gamma}, \tag{4.3.27}$$

$$\widehat{\alpha}^i \equiv \langle i|\alpha\rangle \,, \qquad \widehat{\gamma}^j \equiv \langle j|\gamma\rangle \,. \tag{4.3.28}$$

This is the reason for writing a ket $|\gamma\rangle$ as a column whose components are its representation (eq. (4.3.24)) and a bra $\langle\alpha|$ as a row whose components are the complex conjugate of its representation (eq. (4.3.26)) – their inner product is in fact the complex 'dot product'

$$\langle\alpha|\gamma\rangle = \begin{bmatrix} \langle 1|\alpha\rangle \\ \langle 2|\alpha\rangle \\ \dots \\ \langle D|\alpha\rangle \end{bmatrix}^\dagger \begin{bmatrix} \langle 1|\gamma\rangle \\ \langle 2|\gamma\rangle \\ \dots \\ \langle D|\gamma\rangle \end{bmatrix}, \tag{4.3.29}$$

where the dagger here refers to the matrix algebra operation of taking the transpose and complex conjugation, for e.g. $v^\dagger = (v^\mathrm{T})^*$. Furthermore, if $|\gamma\rangle$ has unit norm, then

$$1 = \langle\gamma|\gamma\rangle = \sum_i \langle\gamma|i\rangle\langle i|\gamma\rangle = \sum_i |\langle i|\gamma\rangle|^2 \doteq \delta_{ij}\left(\widehat{\gamma}^i\right)^* \widehat{\gamma}^j = \widehat{\gamma}^\dagger\widehat{\gamma}. \tag{4.3.30}$$

Linear operators Next, consider some operator X acting on an arbitrary vector $|\gamma\rangle$, expressed through the orthonormal basis vectors $\{|i\rangle\}$. We can insert identity operators, one from the left and another from the right of X,

$$X|\gamma\rangle = \sum_{i,j} |j\rangle\langle j|X|i\rangle\langle i|\gamma\rangle. \tag{4.3.31}$$

We can also apply the lth basis bra $\langle l|$ from the left on both sides and obtain

$$\langle l|X|\gamma\rangle = \sum_i \langle l|X|i\rangle\langle i|\gamma\rangle. \tag{4.3.32}$$

Just as we read off the components of the vector in eq. (4.3.22) as a column vector, we can do the same here. Again supposing a D dimensional vector space for notational convenience,

$$X|\gamma\rangle \doteq \begin{bmatrix} \langle 1|X|1\rangle & \langle 1|X|2\rangle & \dots & \langle 1|X|D\rangle \\ \langle 2|X|1\rangle & \langle 2|X|2\rangle & \dots & \langle 2|X|D\rangle \\ \dots & \dots & \dots & \dots \\ \langle D|X|1\rangle & \langle D|X|2\rangle & \dots & \langle D|X|D\rangle \end{bmatrix} \begin{bmatrix} \langle 1|\gamma\rangle \\ \langle 2|\gamma\rangle \\ \langle 3|\gamma\rangle \\ \dots \\ \langle D|\gamma\rangle \end{bmatrix}. \tag{4.3.33}$$

In words: X acting on some vector $|\gamma\rangle$ can be represented by the column vector gotten from acting the matrix $\langle j|X|i\rangle$, with row number j and column number i, acting on the column vector $\langle i|\gamma\rangle$. In index notation, with[9]

$$\widehat{X}^i{}_j \equiv \langle i|X|j\rangle \qquad \text{and} \qquad \widehat{\gamma}^j \equiv \langle j|\gamma\rangle, \tag{4.3.34}$$

we have

$$\langle i|X|\gamma\rangle \doteq \widehat{X}^i{}_j\widehat{\gamma}^j. \tag{4.3.35}$$

Since $|\gamma\rangle$ in eq. (4.3.31) was arbitrary, we may record that any linear operator X admits an ket-bra operator expansion:

$$X = \sum_{i,j} |j\rangle\langle j|X|i\rangle\langle i| = \sum_{i,j} |j\rangle\widehat{X}^j{}_i\langle i|. \tag{4.3.36}$$

We have already seen, this result follows from inserting the completeness relation in eq. (4.3.23) on the left and right of X. Importantly, notice that specifying the matrix $\widehat{X}^j{}_i$ amounts to defining the linear operator X itself, once a orthonormal basis has been picked.

[9] In this chapter on the abstract formulation of Linear Algebra, I use a $\widehat{\cdot}$ to denote a matrix (representation), in order to distinguish it from the linear operator itself.

As an example: what is the matrix representation of $|\beta\rangle\langle\alpha|$? We apply $\langle i|$ from the left and $|j\rangle$ from the right to obtain the ij component

$$\langle i|\left(|\alpha\rangle\langle\beta|\right)|j\rangle = \langle i|\,\alpha\rangle\langle\beta|\,j\rangle \doteq \widehat{\alpha}^i\left(\widehat{\beta}^j\right)^*. \tag{4.3.37}$$

Products of Linear Operators We can consider YX, where X and Y are linear operators. By inserting the completeness relation in eq. (4.3.23),

$$\begin{aligned} YX|\gamma\rangle &= \sum_{i,j,k}|k\rangle\langle k|Y|j\rangle\langle j|X|i\rangle\langle i|\,\gamma\rangle \\ &= \sum_k |k\rangle\,\widehat{Y}^k{}_j\widehat{X}^j{}_i\widehat{\gamma}^i. \end{aligned} \tag{4.3.38}$$

The product YX can therefore be represented as

$$YX \doteq \begin{bmatrix} \langle 1|Y|1\rangle & \langle 1|Y|2\rangle & \dots & \langle 1|Y|D\rangle \\ \langle 2|Y|1\rangle & \langle 2|Y|2\rangle & \dots & \langle 2|Y|D\rangle \\ \dots & \dots & \dots & \dots \\ \langle D|Y|1\rangle & \langle D|Y|2\rangle & \dots & \langle D|Y|D\rangle \end{bmatrix}\begin{bmatrix} \langle 1|X|1\rangle & \langle 1|X|2\rangle & \dots & \langle 1|X|D\rangle \\ \langle 2|X|1\rangle & \langle 2|X|2\rangle & \dots & \langle 2|X|D\rangle \\ \dots & \dots & \dots & \dots \\ \langle D|X|1\rangle & \langle D|X|2\rangle & \dots & \langle D|X|D\rangle \end{bmatrix}. \tag{4.3.39}$$

Notice how the rules of matrix multiplication emerges from this abstract formulation of linear operators acting on a vector space.

Adjoint Finally, we may now understand how to construct the matrix representation of the adjoint of a given linear operator X by starting from eq. (4.3.8) with orthonormal states $\{|i\rangle\}$. Firstly, from eq. (4.3.36),

$$X^\dagger|i\rangle = \sum_{a,b}|a\rangle(\widehat{X}^\dagger)^a{}_b\langle b|\,i\rangle = \sum_a|a\rangle(\widehat{X}^\dagger)^a{}_i. \tag{4.3.40}$$

Taking the † using eq. (4.2.3), and then applying it to $|j\rangle$,

$$(X^\dagger|i\rangle)^\dagger|j\rangle = \sum_a\overline{(\widehat{X}^\dagger)^a{}_i}\langle a|\,j\rangle = \overline{(\widehat{X}^\dagger)^j{}_i}. \tag{4.3.41}$$

According to eq. (4.3.8), this must be equal to $\langle i|X|j\rangle = \widehat{X}^i{}_j$. This, of course, coincides with the definition of the adjoint from matrix algebra: the representation of the adjoint of X is the complex conjugate and transpose of that of X:

$$\left\langle j\left|X^\dagger\right|i\right\rangle = \langle i|X|j\rangle^* \qquad \Leftrightarrow \qquad \widehat{X}^\dagger = \left(\widehat{X}^{\mathrm{T}}\right)^*. \tag{4.3.42}$$

Because the matrix representation of a linear operator within an orthonormal basis is unique, notice we have also provided a constructive proof of the uniqueness of $X^\dagger$ itself. We could also have obtained eq. (4.3.42) more directly by starting with the ket-bra expansion (cf. (4.3.36)) of X and then using equations (4.2.3) and (4.3.21) to directly implement † on the right hand side:

$$X^\dagger = \sum_{i,j}\left((\widehat{X}^i{}_j|i\rangle)\langle j|\right)^\dagger \tag{4.3.43}$$

$$= \sum_{i,j}|j\rangle\,\overline{\widehat{X}^i{}_j}\,\langle i| \equiv \sum_{i,j}|j\rangle(\widehat{X}^\dagger)^j{}_i\langle i|. \tag{4.3.44}$$

Problem 4.12. Adjoint of an adjoint Prove that $(X^\dagger)^\dagger = X$. $\square$

Now that you have shown that $(Y^\dagger)^\dagger = Y$ for any linear operator Y, for any states $|\alpha\rangle$ and $|\beta\rangle$; and for any linear operator X, we may recover eq. (4.3.9) from the property $\overline{\langle\alpha|\gamma\rangle} = \langle\gamma|\alpha\rangle$:

$$\overline{\langle\alpha| X |\beta\rangle} = (X|\beta\rangle)^\dagger |\alpha\rangle = ((X^\dagger)^\dagger |\beta\rangle)^\dagger |\alpha\rangle = \langle\beta| X^\dagger |\alpha\rangle . \tag{4.3.45}$$

Vector Space of Linear Operators You may step through the axioms of Linear Algebra to verify that the space of Linear operators is, in fact, a vector space itself. (More specifically, we have just seen that linear operators may be represented by $D \times D$ square matrices, which in turn span a vector space of dimension D^2.) Given an orthonormal basis $\{|i\rangle\}$ for the original vector space upon which these linear operators are acting, the expansion in eq. (4.3.36) – which holds for an arbitrary linear operator X – teaches us the set of D^2 ket-bra operators

$$\{|j\rangle \langle i| \, | j, i = 1, 2, 3, \dots, D\} \tag{4.3.46}$$

form the basis of the space of linear operators. The matrix elements $\langle j|X|i\rangle = \widehat{X}^j{}_i$ are the expansion coefficients.

> The set of linear operators $\{X\}$ acting on a $D-$dimensional vector space is itself a D^2 dimensional vector space, because they are represented by the set of $D \times D$ matrices $\{\widehat{X}^i{}_j = \langle i|X|j\rangle\}$.

Inner Product of Linear Operators We have already witnessed how the trace operation may be used to define an inner product between matrices: $\left\langle \widehat{A} \middle| \widehat{B} \right\rangle = \mathrm{Tr}\left[\widehat{A}^\dagger \widehat{B}\right]$. Let us now define the trace of a linear operator X to be

$$\mathrm{Tr}[X] \equiv \sum_{\ell=1}^{D} \langle \ell|X|\ell\rangle ; \tag{4.3.47}$$

where the $\{|\ell\rangle\}$ form an orthonormal basis. (That *any* orthonormal basis would do – i.e., this is a basis independent definition, as along as the basis is unit norm and mutually perpendicular – will be proven in the section on unitary operators below.) We may now define the inner product between two linear operators X and Y as

$$\langle X|Y\rangle \equiv \mathrm{Tr}\left[X^\dagger Y\right]. \tag{4.3.48}$$

This is in fact equivalent to the matrix trace inner product because, by inserting the completeness relation (4.3.23) between X and Y and employing eq. (4.3.47),

$$\langle X|Y\rangle = \sum_{i,j} \langle i| X^\dagger |j\rangle \langle j| Y |i\rangle = (\widehat{X}^\dagger)^i{}_j \widehat{Y}^j{}_i = \mathrm{Tr}\left[\widehat{X}^\dagger \widehat{Y}\right]. \tag{4.3.49}$$

With such a tool, it is now possible to sharpen the statement that the set of D^2 ket-bra operators $\{|i\rangle \langle j| \, |i, j \in 1, 2, 3, \dots, D\}$ form an orthonormal basis for the vector space of linear operators. Recall: since the dimension of such a space is D^2, all we have to show is the linear independence of this set. But this in turn follows if they are orthonormal. Hence, consider the inner product between $|i\rangle\langle j|$ and $|m\rangle\langle n|$. Utilizing the result that $(|i\rangle\langle j|)^\dagger = |j\rangle\langle i|$:

$$\mathrm{Tr}\left[(|i\rangle\langle j|)^\dagger(|m\rangle\langle n|)\right] = \sum_\ell \langle \ell | j\rangle \langle i | m\rangle \langle n | \ell\rangle . \tag{4.3.50}$$

Now, by assumption, $\langle \ell | j\rangle$ is non-zero only when $\ell = j$. Similarly, $\langle n | \ell\rangle$ is non-zero only when $\ell = n$. Therefore when $j \neq n$ the entire sum is zero because ℓ cannot be simultaneously equal to both j and n. But again by the orthonormal assumption, when $\ell = j = n$, $\langle \ell | j\rangle\langle n | \ell\rangle = 1$. In other words, the sum is proportional to δ^j_n; likewise $\langle i | m\rangle = \delta^i_m$ too. At this point, we have arrived at the orthonormality condition:

$$\mathrm{Tr}\left[(|i\rangle\langle j|)^\dagger(|m\rangle\langle n|)\right] = \delta^i_m \delta^j_n. \tag{4.3.51}$$

The kets must be identical and so must the bras; otherwise these ket-bra linear operators are perpendicular.

Problem 4.13. Throughout this section, we are focusing on linear operators that act on a ket and return another within the same vector space; hence, their matrix representations are $D \times D$ matrices. Suppose a linear operator acts on kets within a N dimensional vector space but returns a ket from a (different) M dimensional one. What is the size of the matrix representation? □

Mapping finite dimensional vector spaces to $\mathbb{C}^D$ Let us pause to summarize our preceding discussion. Even though it is possible to discuss finite dimensional vector spaces in the abstract, it is always possible to translate the setup at hand to one of the D-tuple of complex numbers, where D is the dimensionality. First choose a set of orthonormal basis vectors $\{|1\rangle, \dots, |D\rangle\}$. Then, every vector $|\alpha\rangle$ can be represented as a column vector; the ith component is the result of projecting the abstract vector on the ith basis vector $\langle i | \alpha\rangle$; conversely, writing a column of complex numbers can be interpreted to define a vector in this orthonormal basis. The inner product between two vectors $\langle \alpha | \beta\rangle = \sum_i \langle \alpha | i\rangle \langle i | \beta\rangle$ boils down to the complex conjugate of the $\langle i | \alpha\rangle$ column vector dotted into the $\langle i | \beta\rangle$ vector. Moreover, every linear operator O can be represented as a matrix with the element on the ith row and jth column given by $\langle i | O | j\rangle$; and conversely, writing any square matrix $\widehat{O}^i{}_j$ can be interpreted to define a linear operator, on this vector space, with matrix elements $\langle i | O | j\rangle$. Product of linear operators becomes products of matrices, with the usual rules of matrix multiplication.

Object	Representation
Vector/Ket: $\|\alpha\rangle = \sum_i \|i\rangle\langle i\|\alpha\rangle$	$\alpha^i = (\langle 1\|\alpha\rangle, \dots, \langle D\|\alpha\rangle)^{\mathrm{T}}$
Dual Vector/Bra: $\langle\alpha\| = \sum_i \langle\alpha\|i\rangle\langle i\|$	$(\alpha^\dagger)^i = (\langle\alpha\|1\rangle, \dots, \langle\alpha\|D\rangle)$
Inner product: $\langle\alpha\|\beta\rangle = \sum_i \langle\alpha\|i\rangle\langle i\|\beta\rangle$	$\alpha^\dagger\beta = \delta_{ij}\overline{\alpha^i}\beta^j$
Linear operator (LO): $X = \sum_{i,j} \|i\rangle\langle i\|X\|j\rangle\langle j\|$	$\widehat{X}^i{}_j = \langle i\|X\|j\rangle$
LO acting on ket: $X\|\gamma\rangle = \sum_{i,j} \|i\rangle\langle i\|X\|j\rangle\langle j\|\gamma\rangle$	$(\widehat{X}\gamma)^i = \widehat{X}^i{}_j\gamma^j$
Products of LOs: $XY = \sum_{i,j,k} \|i\rangle\langle i\|X\|j\rangle\langle j\|Y\|k\rangle\langle k\|$	$(\widehat{XY})^i{}_k = \widehat{X}^i{}_j\widehat{Y}^j{}_k$
Adjoint of LO: $X^\dagger = \sum_{i,j} \|j\rangle\overline{\langle i\|X\|j\rangle}\langle i\|$	$(\widehat{X}^\dagger)^j{}_i = \overline{\langle i\|X\|j\rangle} = \overline{(\widehat{X}^{\mathrm{T}})^j{}_i}$

Differentiating kets, bras, and linear operators Suppose a ket $|\psi(t)\rangle$ depends on a continuous real parameter t. Then it should make sense to define the limit

$$\partial_t |\psi(t)\rangle \equiv \lim_{\delta t \to 0} \frac{|\psi(t+\delta t)\rangle - |\psi(t)\rangle}{\delta t}. \tag{4.3.52}$$

Taking the adjoint on both sides hands us the corresponding definition for the derivative of the bra.

$$\partial_t \langle\psi(t)| \equiv \lim_{\delta t \to 0} \frac{\langle\psi(t+\delta t)| - \langle\psi(t)|}{\delta t} = \left(\partial_t |\psi(t)\rangle\right)^\dagger . \tag{4.3.53}$$

Likewise, the derivative of a linear operator $A(t)$ that depends on a real continuous parameter t may be defined as

$$\partial_t A(t) = \lim_{\delta t \to 0} \frac{A(t+\delta t) - A(t)}{\delta t}. \tag{4.3.54}$$

Problem 4.14. Product rule Can you prove the product rule holds for the derivative of matrix elements; i.e.,

$$\begin{aligned} &\partial_t \left(\langle\psi_1(t)| A(t) |\psi_2(t)\rangle\right) \\ &\quad = \left(\partial_t \langle\psi_1(t)|\right) A(t) |\psi_2(t)\rangle + \langle\psi_1(t)| \left(\partial_t A(t)\right) |\psi_2(t)\rangle + \langle\psi_1(t)| A(t) \partial_t |\psi_2(t)\rangle? \end{aligned} \tag{4.3.55}$$

Explain why the derivative of the adjoint of a linear operator is the adjoint of the derivative of the same operator: $(\partial_t A(t))^\dagger = \partial_t A^\dagger$. □

Next we highlight two special types of linear operators that play central roles in quantum theory: hermitian and unitary operators.

4.3.2 Hermitian Operators

A hermitian linear operator X is one that is equal to its own adjoint, namely

$$X^\dagger = X. \tag{4.3.56}$$

From eq. (4.3.9), we see that a linear operator X is hermitian if and only if

$$\langle\alpha| X |\beta\rangle = \langle\beta| X |\alpha\rangle^* \tag{4.3.57}$$

for arbitrary vectors $|\alpha\rangle$ and $|\beta\rangle$. In particular, if $\{|i\rangle \,|\, i = 1, 2, 3, \dots, D\}$ form an orthonormal basis, we recover the definition of a Hermitian matrix,

$$\langle j| X |i\rangle = \langle i| X |j\rangle^* . \tag{4.3.58}$$

We now turn to the following important facts about Hermitian operators.

> **Hermitian Operators Have Real Spectra:** If X is a Hermitian operator, all its eigenvalues are real and eigenvectors corresponding to different eigenvalues are orthogonal.

Proof Let $|a\rangle$ and $|a'\rangle$ be eigenvectors of X, i.e.,

$$X |a\rangle = a |a\rangle \tag{4.3.59}$$

Taking the adjoint of the analogous equation for $|a'\rangle$, and using $X = X^\dagger$,

$$\langle a'| X = a'^* \langle a'| . \tag{4.3.60}$$

We can multiply $\langle a'|$ from the left on both sides of eq. (4.3.59); and multiply $|a\rangle$ from the right on both sides of eq. (4.3.60).

$$\langle a'| X |a\rangle = a \langle a'| a\rangle , \qquad \langle a'| X |a\rangle = a'^* \langle a'| a\rangle \tag{4.3.61}$$

Subtracting these two equations,

$$0 = (a - a'^*) \langle a'| a\rangle . \tag{4.3.62}$$

Suppose the eigenvalues are the same, $a = a'$. Then $0 = (a - a^*) \langle a| a\rangle$; because $|a\rangle$ is not a null vector, this means $a = a^*$; eigenvalues of Hermitian operators are real. Suppose instead the eigenvalues are distinct, $a \neq a'$. Because we have just proven that a' can be assumed to be real, we have $0 = (a - a') \langle a'| a\rangle$. By assumption the factor $a - a'$ is not zero. Therefore $\langle a'| a\rangle = 0$, namely, eigenvectors corresponding to different eigenvalues of a Hermitian operator are orthogonal.

> **Completeness of Hermitian Eigensystem:**[10] The eigenkets $\{|\lambda_k\rangle |k = 1, 2, \dots, D\}$ of a Hermitian operator span the vector space upon which it is acting. The full set of eigenvalues $\{\lambda_k|k = 1, 2, \dots, D\}$ of some Hermitian operator is called its *spectrum*; and from eq. (4.3.23), completeness of its eigenvectors reads
>
> $$\mathbb{I} = \sum_{k=1}^{D} |\lambda_k\rangle \langle\lambda_k| . \tag{4.3.63}$$
>
> In the language of matrix algebra, we'd say that a Hermitian matrix is always diagonalizable via a unitary transformation.

In quantum theory, we postulate that (many) observables such as spin, position, momentum, etc., correspond to Hermitian operators; their eigenvalues are then the possible outcomes of the measurements of these observables. (It is not possible to obtain a measurement corresponding to X that is *not* its eigenvalue.) Since their spectrum are real, this guarantees we get a real number from performing a measurement on the system at hand. That the eigenstates of an observable span the given vector space also means the range of physical states corresponding to possible measurement outcomes may be employed to fully characterize the dynamics of the quantum system itself.

Degeneracy and Symmetry If more than one eigenket of A has the same eigenvalue, we say A's spectrum is degenerate. The simplest example is the identity operator itself: every basis vector is an eigenvector with eigenvalue 1. The matrix diag$[1, \pi, 2, 2]$ is degenerate: it acts on a 4D vector space with two repeated eigenvalues.

[10] The most general type of operator that is diagonalizable is a normal operator, defined as one that commutes with its own adjoint. Both Hermitian and Unitary (discussed in §(4.3.3) below) operators are normal. I prove the diagonalizability of normal operators in §(4.7).

When an operator is degenerate, the labeling of eigenkets using their eigenvalues become ambiguous – which eigenket does $|\lambda\rangle$ correspond to, if this subspace is 5 dimensional, say? What often happens is that one can find a different observable B to distinguish between the eigenkets of the same λ. For example, we will see below that the negative Laplacian on the 2-sphere – known as the "square of total angular momentum," when applied to quantum mechanics – will have eigenvalues $\ell(\ell+1)$, where $\ell \in \{0,1,2,3,\dots\}$. It will also turn out to be $(2\ell+1)$-fold degenerate, but this degeneracy can be labeled by an integer m, corresponding to the eigenvalues of the generator-of-rotation about the North pole $J(\phi)$ (where ϕ is the azimuthal angle). A closely related fact is that $[-\vec{\nabla}^2_{\mathbb{S}^2}, J(\phi)] = 0$, where $[X,Y] \equiv XY - YX$.

$$\begin{aligned} &-\vec{\nabla}^2_{\mathbb{S}^2} \left|\ell, m\right\rangle = \ell(\ell+1) \left|\ell, m\right\rangle, \\ &\ell \in \{0,1,2,\dots\}, \qquad m \in \{-\ell, -\ell+1, \dots, -1, 0, 1, \dots, \ell-1, \ell\}. \end{aligned} \tag{4.3.64}$$

It's worthwhile to mention, in the context of quantum theory – degeneracy in the spectrum is often associated with the presence of symmetry. For example, the Stark and Zeeman effects can be respectively thought of as the breaking of rotational symmetry of an atomic system by, respectively, a non-zero magnetic and electric field. Previously degenerate spectral lines become split into distinct ones, due to these $\vec{E}$ and $\vec{B}$ fields.[11] In the context of classical field theory, we will witness in the section on continuous vector spaces below, how the translation invariance of space leads to a degenerate spectrum of the Laplacian.

Problem 4.15. Let X be a linear operator with eigenvalues $\{\lambda_i | i = 1,2,3,\dots,D\}$ and orthonormal eigenvectors $\{|\lambda_i\rangle \,| i = 1,2,3,\dots,D\}$ that span the given vector space. Show that X can be expressed as

$$X = \sum_i \lambda_i \left|\lambda_i\right\rangle \left\langle\lambda_i\right|. \tag{4.3.65}$$

(Assume a non-degenerate spectra for now.) Verify that the right hand side is represented by a diagonal matrix in this basis $\{|\lambda_i\rangle\}$. Of course, a Hermitian linear operator is a special case of eq. (4.3.65), where all the $\{\lambda_i\}$ are real. Hint: Given that the eigenkets of X span the vector space, all you need to verify is that all possible matrix elements of X return what you expect. $\square$

How to diagonalize a Hermitian operator? To *diagonalize* a linear operator X means to get it in the form in eq. (4.3.65), where it is expanded in terms of projectors built out of its eigen-kets $\{|\lambda_i\rangle\}$. The matrix representation in such a basis is purely diagonal $\langle\lambda_i |X| \lambda_j\rangle = \lambda_i \delta^i_j$.

Suppose you are given a Hermitian operator H in some orthonormal basis $\{|i\rangle\}$, namely

$$H = \sum_{i,j} |i\rangle \widehat{H}^i{}_j \langle j| = \sum_{i,j} |i\rangle \langle i |H| j\rangle \langle j|. \tag{4.3.66}$$

How does one go about diagonalizing it? Here is where the matrix algebra you are familiar with comes in – recall the discussion leading up to eq. (3.3.11). By treating $\widehat{H}^i{}_j$ as a matrix, you can

[11]See Wikipedia articles on the Stark and Zeeman effects for plots of the energy levels vs. electric/magnetic field strengths.

find its eigenvectors and eigenvalues $\{\lambda_k\}$. Specifically, what you are solving for is the unitary matrix $\widehat{U}^j{}_k$, whose kth column is the kth unit length eigenvector of $\widehat{H}^i{}_j$, with eigenvalue λ_k:

$$\widehat{H}^i{}_j \widehat{U}^j{}_k = \lambda_k \widehat{U}^j{}_k \qquad \Leftrightarrow \qquad \sum_j \langle i |H| j\rangle \langle j|\lambda_k\rangle = \lambda_k \langle i|\lambda_k\rangle, \tag{4.3.67}$$

with

$$\langle i |H| j\rangle \equiv \widehat{H}^i{}_j \qquad \text{and} \qquad \langle j|\lambda_k\rangle \equiv \widehat{U}^j{}_k. \tag{4.3.68}$$

In other words,

$$\begin{aligned}\widehat{H}^i{}_j &= (\widehat{U}\mathrm{diag}[\lambda_1, \dots, \lambda_D]\widehat{U}^\dagger)^i{}_j \\ &= \sum_k \langle i|\lambda_k\rangle \lambda_k \langle \lambda_k| j\rangle .\end{aligned} \tag{4.3.69}$$

Once you have obtained the representation of the kth eigenket $\widehat{U}^i{}_k = (\langle 1|\lambda_k\rangle, \langle 2|\lambda_k\rangle, \dots, \langle D|\lambda_k\rangle)^{\mathrm{T}}$, you can then write the eigenket itself as

$$|\lambda_k\rangle = \sum_i |i\rangle \langle i|\lambda_k\rangle = \sum_i |i\rangle \widehat{U}^i{}_k. \tag{4.3.70}$$

The adjoint of the same eigenket is

$$\langle\lambda_k| = \sum_i \langle\lambda_k| i\rangle \langle i| = \sum_i \overline{\widehat{U}^i{}_k} \langle i| = \sum_i (\widehat{U}^\dagger)^k{}_i \langle i| . \tag{4.3.71}$$

The operator H has now been diagonalized as

$$H = \sum_k \lambda_k |\lambda_k\rangle \langle\lambda_k| \tag{4.3.72}$$

because according to eq. (4.3.69),

$$H = \sum_{i,j} |i\rangle \widehat{H}^i{}_j \langle j| = \sum_{i,j,a} |i\rangle \langle i|\lambda_a\rangle \lambda_a \langle\lambda_a| j\rangle \langle j| . \tag{4.3.73}$$

Using the completeness relation in eq. (4.3.23) then leads us to eq. (4.3.72).

In summary, with the relations in eq. (4.3.68),

$$\begin{aligned} H &= \sum_{i,j} |i\rangle \widehat{H}^i{}_j \langle j| = \sum_k \lambda_k |\lambda_k\rangle \langle\lambda_k| & (4.3.74)\\ &= \sum_{i,j,m,n} |i\rangle \widehat{U}^i{}_m \left(\mathrm{diag}\left[\lambda_1, \dots, \lambda_D\right]\right)^m{}_n (\widehat{U}^\dagger)^n{}_j \langle j| . & (4.3.75)\end{aligned}$$

This matrix algebra that guarantees every Hermitian matrix can be diagonalized through a unitary transformation, $\widehat{H} = \widehat{U}\mathrm{diag}[\lambda_1, \dots, \lambda_D]\widehat{U}^\dagger$, also amounts to a proof that all Hermitian operators acting on finite dimensional vector spaces have a complete spectra – since, the columns of $\widehat{U}$ are the (representation of) the orthonormal eigenbasis.

Problem 4.16. Consider a 2 dimension vector space with the orthonormal basis $\{|1\rangle, |2\rangle\}$. The operator H is defined through its actions:

$$H|1\rangle = a|1\rangle + ib|2\rangle, \tag{4.3.76}$$
$$H|2\rangle = -ib|1\rangle + a|2\rangle; \tag{4.3.77}$$

where a and b are real numbers. Is H hermitian? What are its eigenvectors and eigenvalues? □

Compatible observables Let X and Y be observables – aka Hermitian operators. We shall define compatible observables to be ones where the operators commute,

$$[A, B] \equiv AB - BA = 0. \tag{4.3.78}$$

They are incompatible when $[A, B] \neq 0$. Finding the maximal set of mutually compatible set of observables in a given physical system will tell us the range of eigenvalues that fully capture the quantum state of the system. To understand this we need the following result.

> **Theorem** Suppose X and Y are observables – they are Hermitian operators. Then X and Y are compatible (i.e., commute with each other) if and only if they are simultaneously diagonalizable.

Proof We will provide the proof for the case where the spectrum of X is non-degenerate. We have already stated earlier that if X is Hermitian we can expand it in its basis eigenkets.

$$X = \sum_a a|a\rangle\langle a| \tag{4.3.79}$$

In this basis X is already diagonal. But what about Y? Suppose $[X, Y] = 0$. We consider, for distinct eigenvalues a and a' of X,

$$\langle a'|[X, Y]|a\rangle = \langle a'|XY - YX|a\rangle = (a' - a)\langle a'|Y|a\rangle = 0. \tag{4.3.80}$$

Remember, all eigenvalues of X and Y are real because the operators are Hermitian; hence not only $X|a\rangle = a|a\rangle$ we also have $\langle a'|X = (X^\dagger|a'\rangle)^\dagger = (X|a'\rangle)^\dagger = (a'|a'\rangle)^\dagger = a'\langle a'|$. From the last equality, since $a - a' \neq 0$ by assumption, we must have $\langle a'|Y|a\rangle = 0$. That means the only non-zero matrix elements are the diagonal ones $\langle a|Y|a\rangle$.[12]

We have thus shown $[X, Y] = 0 \Rightarrow X$ and Y are simultaneously diagonalizable. We now turn to proving, if X and Y are simultaneously diagonalizable, then $[X, Y] = 0$. That is, suppose

$$X = \sum_{a,b} a|a, b\rangle\langle a, b| \qquad \text{and} \qquad Y = \sum_{a,b} b|a, b\rangle\langle a, b|, \tag{4.3.81}$$

[12]If the spectrum of X were N-fold degenerate, $\{|a; i\rangle \,|\, i = 1, 2, \dots, N\}$ with $X|a; i\rangle = a|a; i\rangle$, to extend the proof to this case, all we have to do is to diagonalize the $N \times N$ matrix $\langle a; i|Y|a; j\rangle$. That this is always possible is because Y is Hermitian. Within the subspace spanned by these $\{|a; i\rangle\}$, $X = \sum_i a|a; i\rangle\langle a; i| + \dots$ acts like a times the identity operator, and will therefore definitely commute with Y.

[13]let's compute the commutator

$$[X,Y] = \sum_{a,b,a',b'} ab' \left(|a,b\rangle \langle a,b| a',b'\rangle \langle a',b'| - |a',b'\rangle \langle a,b| \right). \tag{4.3.82}$$

Remember that eigenvectors corresponding to distinct eigenvalues are orthogonal, namely $\langle a,b| a',b'\rangle$ is unity only when $a = a'$ and $b = b'$ simultaneously. This means we may discard the summation over (a',b') and set $a = a'$ and $b = b'$ within the summand.

$$[X,Y] = \sum_{a,b} ab \left(|a,b\rangle \langle a,b| - |a,b\rangle \langle a,b| a,b\rangle \langle a,b| \right) = 0. \tag{4.3.83}$$

□

Problem 4.17. Assuming the spectrum of X is non-degenerate, show that the Y in the preceding theorem can be expanded in terms of the eigenkets of X as

$$Y = \sum_a |a\rangle \langle a |Y| a\rangle \langle a|. \tag{4.3.84}$$

Read off the eigenvalues. □

Problem 4.18. Properties of Commutators Show that, for linear operators A, B, and C, the following relations hold.

$$[AB,C] = A[B,C] + [A,C]B, \tag{4.3.85}$$
$$[A,BC] = B[A,C] + [A,B]C, \tag{4.3.86}$$
$$[A,B]^\dagger = -[A^\dagger, B^\dagger]; \tag{4.3.87}$$

and

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0. \tag{4.3.88}$$

If we have a collection of linear operators $\{A_{(i)}|i = 1,2,\dots,M\}$ and $\{B_{(i)}|i = 1,2,\dots,N\}$, explain why the commutator in linear in both slots, in that

$$\left[\sum_{i=1}^{M} A_{(i)}, \sum_{j=1}^{N} B_{(j)}\right] = \sum_{i=1}^{M}\sum_{j=1}^{N} \left[A_{(i)}, B_{(j)}\right]. \tag{4.3.89}$$

□

Uncertainty Relation If X and Y are incompatible observables, then they cannot be simultaneously diagonalized. The product of their 'variances', however, can be shown to have a lower limit provided by their commutator $[X,Y]$. (Hence, if X and Y were compatible, namely $[X,Y] = 0$, this lower limit would become zero.) This is the celebrated uncertainty relation.

[13]Remember, to say X or Y is diagonalized means it has been put in the form in eq. (4.3.65). To say both of them have been simultaneously diagonalized therefore means they can be put in the form in eq. (4.3.65) using the same set of eigenkets.

More precisely, we define the variance of an operator X with respect to a given state $|\psi\rangle$ via the relation

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \equiv \left\langle \psi \left| (X - \langle \psi | X | \psi \rangle)^2 \right| \psi \right\rangle; \tag{4.3.90}$$

i.e., $\Delta X \equiv X - \langle \psi | X | \psi \rangle$. Note that, since X is Hermitian and $\langle \psi | X | \psi \rangle$ is a real number; ΔX (and, similarly, ΔY) is Hermitian.

From the Cauchy-Schwarz inequality of eq. (4.2.24), if we identify $|\alpha\rangle = \Delta X |\psi\rangle$ and $|\beta\rangle = \Delta Y |\psi\rangle$, then

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta Y^2 \right| \psi \right\rangle \geq | \langle \psi | \Delta X \Delta Y | \psi \rangle |^2. \tag{4.3.91}$$

The product of two arbitrary operators A and B may be written as half of their commutator plus half of their anti-commutator:

$$AB = \frac{1}{2} [A, B] + \frac{1}{2} \{A, B\}; \tag{4.3.92}$$

where the anti-commutator itself is defined as

$$\{A, B\} \equiv AB + BA. \tag{4.3.93}$$

(If eq. (4.3.92) is not apparent, simply expand the right hand side.) Now, let us note that the commutator of two observables is anti-Hermitian, in that

$$[A, B]^\dagger = (AB)^\dagger - (BA)^\dagger \tag{4.3.94}$$

$$= BA - AB = -[A, B]. \tag{4.3.95}$$

Whereas the anti-commutator of a pair of observables is itself an observable:

$$\{A, B\}^\dagger = (AB)^\dagger + (BA)^\dagger \tag{4.3.96}$$

$$= BA + AB = \{A, B\}. \tag{4.3.97}$$

Additionally, if $A^\dagger = \pm A$; then $\langle \psi | A | \psi \rangle^* = \left\langle \psi \left| A^\dagger \right| \psi \right\rangle = \pm \langle \psi | A | \psi \rangle$.

> The expectation value of a Hermitian operator is purely real; that of an anti-Hermitian operator is purely imaginary.

Altogether, we learn that the expectation value of eq. (4.3.92), when $A = \Delta X$ and $B = \Delta Y$, reads

$$\langle \psi | \Delta X \Delta Y | \psi \rangle = \frac{1}{2} \langle \psi | [\Delta X, \Delta Y] | \psi \rangle + \frac{1}{2} \langle \psi | \{\Delta X, \Delta Y\} | \psi \rangle. \tag{4.3.98}$$

The first and second terms on the right hand side are, respectively, its real and imaginary parts. But since the modulus square of a complex number is the sum of the square of its real and imaginary pieces,

$$| \langle \psi | \Delta X \Delta Y | \psi \rangle |^2 = \frac{1}{4} | \langle \psi | [\Delta X, \Delta Y] | \psi \rangle |^2 + \frac{1}{4} | \langle \psi | \{\Delta X, \Delta Y\} | \psi \rangle |^2. \tag{4.3.99}$$

Plugging this result back into eq. (4.3.91),

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta Y^2 \right| \psi \right\rangle \geq \frac{1}{4} \left| \langle \psi | \left[\Delta X, \Delta Y \right] | \psi \rangle \right|^2 + \frac{1}{4} \left| \langle \psi | \left\{ \Delta X, \Delta Y \right\} | \psi \rangle \right|^2. \tag{4.3.100}$$

Note that $[\Delta X, \Delta Y] = [X + \langle X \rangle, Y + \langle Y \rangle] = [X, Y]$ because $\langle X \rangle$ and $\langle Y \rangle$ are numbers, which must commute with everything. Since the sum of two squares on the right hand side of eq. (4.3.100) must certainly larger or equal to the first commutator term or the second anti-commutator one, we arrive at

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta Y^2 \right| \psi \right\rangle \geq \frac{1}{4} \left| \langle \psi | \left[X, Y \right] | \psi \rangle \right|^2, \tag{4.3.101}$$

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta Y^2 \right| \psi \right\rangle \geq \frac{1}{4} \left| \langle \psi | \left\{ \Delta X, \Delta Y \right\} | \psi \rangle \right|^2. \tag{4.3.102}$$

Problem 4.19. Verify that, for $\Delta X \equiv X - \langle \psi | X | \psi \rangle$ and $\Delta Y \equiv Y - \langle \psi | Y | \psi \rangle$,

$$\frac{1}{2} \langle \psi | \{ \Delta X, \Delta Y \} | \psi \rangle = \frac{1}{2} \langle \psi | \{ X, Y \} | \psi \rangle - \langle \psi | X | \psi \rangle \langle \psi | Y | \psi \rangle ; \tag{4.3.103}$$

and use it to re-write eq. (4.3.102). □

Probabilities and Expectation value In the context of quantum theory, given a state $|\alpha\rangle$ and an observable O, we may expand the former in terms of the orthonormal eigenkets $\{|\lambda_i\rangle\}$ of the latter,

$$|\alpha\rangle = \sum_i |\lambda_i\rangle \langle \lambda_i | \alpha \rangle , \qquad O |\lambda_i\rangle = \lambda_i |\lambda_i\rangle . \tag{4.3.104}$$

It is a postulate of quantum theory that the probability of obtaining a specific λ_j in an experiment designed to observe O (which can be energy, spin, etc.) is given by $| \langle \lambda_j | \alpha \rangle |^2 = \langle \alpha | \lambda_i \rangle \langle \lambda_i | \alpha \rangle$; if the spectrum is degenerate, so that there are N eigenkets $\{ |\lambda_i; j\rangle \, | j = 1, 2, 3, \dots, N \}$ corresponding to λ_i, then the probability will be

$$P(\lambda_i) = \sum_j \langle \alpha | \lambda_i; j \rangle \langle \lambda_i; j | \alpha \rangle . \tag{4.3.105}$$

This is known as the Born rule.

The expectation value of some operator O with respect to some state $|\alpha\rangle$ is defined to be

$$\langle \alpha | O | \alpha \rangle . \tag{4.3.106}$$

If O is Hermitian, then the expectation value is real, since

$$\langle \alpha | O | \alpha \rangle^* = \left\langle \alpha \left| O^\dagger \right| \alpha \right\rangle = \langle \alpha | O | \alpha \rangle . \tag{4.3.107}$$

In the quantum context, because we may interpret O to be an observable, its expectation value with respect to some state can be viewed as the average value of the observable – the result of measuring it over $N \to \infty$ number of times. This can be seen by expanding $|\alpha\rangle$ in terms of the eigenstates of O.

$$\begin{aligned}
\langle \alpha |O| \alpha \rangle &= \sum_{i,j} \langle \alpha | \lambda_i \rangle \langle \lambda_i |O| \lambda_j \rangle \langle \lambda_j | \alpha \rangle \\
&= \sum_{i,j} \langle \alpha | \lambda_i \rangle \lambda_i \langle \lambda_i | \lambda_j \rangle \langle \lambda_j | \alpha \rangle \\
&= \sum_{i} | \langle \alpha | \lambda_i \rangle |^2 \lambda_i = \sum_i \lambda_i P(\lambda_i).
\end{aligned} \tag{4.3.108}$$

The probability of finding λ_i is $| \langle \alpha | \lambda_i \rangle |^2$, therefore the expectation value is an average. (In the sum here, we assume a non-degenerate spectrum for simplicity; otherwise, simply include the sum over all the relevant degenerate states.)

Suppose instead O is anti-Hermitian, $O^\dagger = -O$. Then we see its expectation value with respect to some state $|\alpha\rangle$ is purely imaginary.

$$\langle \alpha |O| \alpha \rangle^* = \left\langle \alpha \left| O^\dagger \right| \alpha \right\rangle = - \langle \alpha |O| \alpha \rangle \tag{4.3.109}$$

Hellmann-Feynman Whenever the Hermitian operator $A(\alpha_1, \alpha_2, \dots) \equiv A(\vec{\alpha})$ depends on a number of parameters $\{\alpha_i\}$, we expect its unit-norm eigenstates $\{|\lambda(\vec{\alpha})\rangle\}$ and eigenvalues $\{\lambda(\vec{\alpha})\}$ to also depend on them. We may express these eigenvalues through the expectation value

$$\lambda(\vec{\alpha}) = \langle \lambda(\vec{\alpha}) |A(\vec{\alpha})| \lambda(\vec{\alpha}) \rangle . \tag{4.3.110}$$

The result due to Hellmann and Feynman – which has applications in, say, the quantum mechanics of molecules – is that the derivative of this eigenvalue does not involve the derivatives of the states, namely

$$\frac{\partial \lambda(\vec{\alpha})}{\partial \alpha_i} = \left\langle \lambda(\vec{\alpha}) \left| \frac{\partial A(\vec{\alpha})}{\partial \alpha_i} \right| \lambda(\vec{\alpha}) \right\rangle, \qquad i = 1, 2, 3, \dots . \tag{4.3.111}$$

Proof A straightforward differentiation would confirm

$$\partial_{\alpha_i} \lambda = \left(\partial_{\alpha_i} \langle \lambda | \right) A |\lambda\rangle + \langle \lambda | A \partial_{\alpha_i} |\lambda\rangle + \langle \lambda |\partial_{\alpha_i} A| \lambda \rangle . \tag{4.3.112}$$

Keeping in mind $\langle \lambda | A = \lambda \langle \lambda |$ and $A |\lambda\rangle = \lambda |\lambda\rangle$, the result follows upon recognizing the unit-norm character of the $|\lambda\rangle$.

$$\begin{aligned}
\partial_{\alpha_i} \lambda &= \lambda \left\{ \left(\partial_{\alpha_i} \langle \lambda | \right) |\lambda\rangle + \langle \lambda | \partial_{\alpha_i} |\lambda\rangle \right\} + \langle \lambda |\partial_{\alpha_i} A| \lambda \rangle \\
&= \lambda \partial_{\alpha_i} \left(\langle \lambda | \lambda \rangle \right) + \langle \lambda |\partial_{\alpha_i} A| \lambda \rangle \\
&= \lambda \partial_{\alpha_i} (1) + \langle \lambda |\partial_{\alpha_i} A| \lambda \rangle .
\end{aligned} \tag{4.3.113}$$

Pauli matrices from their algebra. Before moving on to unitary operators, let us now try to construct (up to a phase) the Pauli matrices in eq. (3.2.18). We assume the following.

- The $\{\sigma^i | i = 1, 2, 3\}$ are Hermitian linear operators acting on a 2 dimensional vector space.

- They obey the algebra

$$\sigma^i \sigma^j = \delta^{ij}\mathbb{I} + i\sum_k \epsilon^{ijk}\sigma^k. \tag{4.3.114}$$

That this is consistent with the Hermitian nature of the $\{\sigma^i\}$ can be checked by taking † on both sides. We have $(\sigma^i\sigma^j)^\dagger = \sigma^j\sigma^i$ on the left-hand-side; whereas on the right-hand-side $(\delta^{ij}\mathbb{I} + i\sum_k \epsilon^{ijk}\sigma^k)^\dagger = \delta^{ij}\mathbb{I} - i\epsilon^{ijk}\sigma^k = \delta^{ij}\mathbb{I} + i\epsilon^{jik}\sigma^k = \sigma^j\sigma^i$.

We begin by noting

$$[\sigma^i, \sigma^j] = (\delta^{ij} - \delta^{ji})\mathbb{I} + \sum_k i(\epsilon^{ijk} - \epsilon^{jik})\sigma^k = 2i\sum_k \epsilon^{ijk}\sigma^k. \tag{4.3.115}$$

We then define the operators

$$\sigma^\pm \equiv \sigma^1 \pm i\sigma^2 \qquad \Rightarrow \qquad (\sigma^\pm)^\dagger = \sigma^\mp; \tag{4.3.116}$$

and calculate[14]

$$\begin{aligned} [\sigma^3, \sigma^\pm] = [\sigma^3, \sigma^1] \pm i[\sigma^3, \sigma^2] &= 2i\epsilon^{312}\sigma^2 \pm 2i^2\epsilon^{321}\sigma^1 \qquad (4.3.117) \\ &= 2i\sigma^2 \pm 2\sigma^1 = \pm 2(\sigma^1 \pm i\sigma^2), \\ \Rightarrow \qquad [\sigma^3, \sigma^\pm] &= \pm 2\sigma^\pm. \qquad (4.3.118) \end{aligned}$$

Also,

$$\begin{aligned} \sigma^\mp\sigma^\pm &= (\sigma^1 \mp i\sigma^2)(\sigma^1 \pm i\sigma^2) \\ &= (\sigma^1)^2 + (\mp i)(\pm i)(\sigma^2)^2 \mp i\sigma^2\sigma^1 \pm i\sigma^1\sigma^2 \\ &= 2\mathbb{I} \pm i(\sigma^1\sigma^2 - \sigma^2\sigma^1) = 2\mathbb{I} \pm i[\sigma^1, \sigma^2] = 2\mathbb{I} \pm 2i^2\epsilon^{123}\sigma^3 \\ \Rightarrow \qquad \sigma^\mp\sigma^\pm &= 2(\mathbb{I} \mp \sigma^3). \qquad (4.3.119) \end{aligned}$$

σ^3 and its Matrix representation. Suppose $|\lambda\rangle$ is a unit norm eigenket of σ^3. Using $\sigma^3|\lambda\rangle = \lambda|\lambda\rangle$ and $(\sigma^3)^2 = \mathbb{I}$,

$$1 = \langle\lambda|\,\lambda\rangle = \left\langle\lambda\left|\sigma^3\sigma^3\right|\lambda\right\rangle = \left(\sigma^3|\lambda\rangle\right)^\dagger\left(\sigma^3|\lambda\rangle\right) = \lambda^2\langle\lambda|\,\lambda\rangle = \lambda^2. \tag{4.3.120}$$

We see immediately that the spectrum is at most $\lambda_\pm = \pm 1$. (We will prove below that the vector space is indeed spanned by both $|\pm\rangle$.) Since the vector space is 2 dimensional, and since the eigenvectors of a Hermitian operator with distinct eigenvalues are necessarily orthogonal, we see that $|\pm\rangle$ span the space at hand. We may thus say

$$\sigma^3 = |+\rangle\langle+| - |-\rangle\langle-|, \tag{4.3.121}$$

which immediately allows us to read off its matrix representation in this basis $\{|\pm\rangle\}$, with $\langle+|\sigma^3|+\rangle$ being the top left hand corner entry:

$$\left\langle j\left|\sigma^3\right|i\right\rangle = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.3.122}$$

[14] The commutator is linear in that $[X, Y+Z] = X(Y+Z) - (Y+Z)X = (XY - YX) + (XZ - ZX) = [X,Y] + [X,Z]$.

Observe that we could have considered $\langle \lambda | \sigma^i \sigma^i | \lambda \rangle$ for any $i \in \{1,2,3\}$; we are just picking $i = 3$ for concreteness. In particular, we see from their algebraic properties that *all three Pauli operators* $\sigma^{1,2,3}$ *have the same spectrum* $\{+1, -1\}$. Moreover, since the σ^is do not commute, we already know they cannot be simultaneously diagonalized.

Raising and lowering (aka Ladder) operators $\sigma^\pm$*, and* $\sigma^{1,2}$*.* Let us now consider

$$\begin{aligned}\sigma^3 \sigma^\pm |\lambda\rangle &= (\sigma^3 \sigma^\pm - \sigma^\pm \sigma^3 + \sigma^\pm \sigma^3) |\lambda\rangle \\ &= ([\sigma^3, \sigma^\pm] + \sigma^\pm \sigma^3) |\lambda\rangle = (\pm 2\sigma^\pm + \lambda \sigma^\pm) |\lambda\rangle \\ &= (\lambda \pm 2)\sigma^\pm |\lambda\rangle \qquad \Rightarrow \qquad \sigma^\pm |\lambda\rangle = K_\lambda^\pm |\lambda \pm 2\rangle, \qquad K_\lambda^\pm \in \mathbb{C}. \end{aligned} \tag{4.3.123}$$

This is why the $\sigma^\pm$ are often called raising/lowering operators: when applied to the eigenket $|\lambda\rangle$ of σ^3 it returns an eigenket with eigenvalue raised/lowered by 2 relative to λ. This sort of algebraic reasoning is important for the study of group representations; solving the energy levels of the quantum harmonic oscillator and the Hydrogen atom[15]; and even the notion of particles in quantum field theory.

What is the norm of $\sigma^\pm |\lambda\rangle$?

$$\begin{aligned}\left\langle \lambda \left| \sigma^\mp \sigma^\pm \right| \lambda \right\rangle &= |K_\lambda^\pm|^2 \langle \lambda \pm 2 | \lambda \pm 2 \rangle \\ \left\langle \lambda \left| 2(\mathbb{I} \mp \sigma^3) \right| \lambda \right\rangle &= |K_\lambda^\pm|^2 \\ 2(1 \mp \lambda) &= |K_\lambda^\pm|^2. \end{aligned} \tag{4.3.124}$$

This means we can solve $K_\lambda^\pm$ up to a phase

$$K_\lambda^\pm = e^{i\delta_\pm^{(\lambda)}} \sqrt{2(1 \mp \lambda)}, \qquad \lambda \in \{-1, +1\}. \tag{4.3.125}$$

Note that $K_+^+ = e^{i\delta_+^{(+)}} \sqrt{2(1-(+1))} = 0$, and $K_-^- = e^{i\delta_-^{(-)}} \sqrt{2(1+(-1))} = 0$, which means

$$\sigma^+ |+\rangle = 0, \qquad \sigma^- |-\rangle = 0. \tag{4.3.126}$$

We can interpret this as saying, there are no larger eigenvalues than $+1$ and no smaller than -1 – this is consistent with our assumption that we have a 2-dimensional vector space. Moreover, $K_+^- = e^{i\delta_-^{(+)}} \sqrt{2(1+(+1))} = 2e^{i\delta_-^{(+)}}$ and $K_-^+ = e^{i\delta_+^{(-)}} \sqrt{2(1-(-1))} = 2e^{i\delta_-^{(+)}}$.

$$\sigma^+ |-\rangle = 2e^{i\delta_+^{(-)}} |+\rangle, \qquad \sigma^- |+\rangle = 2e^{i\delta_-^{(+)}} |-\rangle. \tag{4.3.127}$$

At this point, we have proved that the spectrum of σ^3 has to include both $|\pm\rangle$, because we can get from one to the other by applying $\sigma^\pm$ appropriately. In other words, if $|+\rangle$ exists, so does $|-\rangle \propto \sigma^- |+\rangle$; and if $|-\rangle$ exists, so does $|+\rangle \propto \sigma^+ |-\rangle$.

Also notice we have figured out how $\sigma^\pm$ acts on the basis kets (up to phases), just from their algebraic properties. We may now turn this around to write them in terms of the basis bras/kets:

$$\sigma^+ = 2e^{i\delta_+^{(-)}} |+\rangle \langle -|, \qquad \sigma^- = 2e^{i\delta_-^{(+)}} |-\rangle \langle +|. \tag{4.3.128}$$

[15]For the H atom, the algebraic derivation of its energy levels involve the quantum analog of the classical Laplace-Runge-Lenz vector.

Since $(\sigma^+)^\dagger = \sigma^-$, we must have $\delta_+^{(-)} = -\delta_-^{(+)} \equiv \delta$.

$$\sigma^+ = 2e^{i\delta} \left|+\right\rangle \left\langle-\right|, \qquad \sigma^- = 2e^{-i\delta} \left|-\right\rangle \left\langle+\right|. \tag{4.3.129}$$

with the corresponding matrix representations, with $\langle + |\sigma^\pm| + \rangle$ being the top left hand corner entry:

$$\langle j |\sigma^+| i \rangle = \begin{bmatrix} 0 & 2e^{i\delta} \\ 0 & 0 \end{bmatrix}, \qquad \langle j |\sigma^-| i \rangle = \begin{bmatrix} 0 & 0 \\ 2e^{-i\delta} & 0 \end{bmatrix}. \tag{4.3.130}$$

Now, we have $\sigma^\pm = \sigma^1 \pm i\sigma^2$, which means we can solve for

$$2\sigma^1 = \sigma^+ + \sigma^-, \qquad 2i\sigma^2 = \sigma^+ - \sigma^-. \tag{4.3.131}$$

We have

$$\sigma^1 = e^{i\delta} \left|+\right\rangle \left\langle-\right| + e^{-i\delta} \left|-\right\rangle \left\langle+\right|, \tag{4.3.132}$$
$$\sigma^2 = -ie^{i\delta} \left|+\right\rangle \left\langle-\right| + ie^{-i\delta} \left|-\right\rangle \left\langle+\right|, \qquad \delta \in \mathbb{R}, \tag{4.3.133}$$

with matrix representations

$$\langle j |\sigma^1| i \rangle = \begin{bmatrix} 0 & e^{i\delta} \\ e^{-i\delta} & 0 \end{bmatrix}, \qquad \langle j |\sigma^2| i \rangle = \begin{bmatrix} 0 & -ie^{i\delta} \\ ie^{-i\delta} & 0 \end{bmatrix}. \tag{4.3.134}$$

You can check explicitly that the algebra in eq. (4.3.114) holds for any δ. However, we can also use the fact that unit normal eigenkets can be re-scaled by a phase and still remain unit norm eigenkets.

$$\sigma^3 \left(e^{i\theta} \left|\pm\right\rangle\right) = \pm \left(e^{i\theta} \left|\pm\right\rangle\right), \qquad \left(e^{i\theta} \left|\pm\right\rangle\right)^\dagger \left(e^{i\theta} \left|\pm\right\rangle\right) = 1, \qquad \theta \in \mathbb{R}. \tag{4.3.135}$$

We re-group the phases occurring within our σ^3 and $\sigma^\pm$ as follows.

$$\sigma^3 = (e^{i\delta/2} \left|+\right\rangle)(e^{i\delta/2} \left|+\right\rangle)^\dagger - (e^{-i\delta/2} \left|-\right\rangle)(e^{-i\delta/2} \left|-\right\rangle)^\dagger, \tag{4.3.136}$$
$$\sigma^+ = 2(e^{i\delta/2} \left|+\right\rangle)(e^{-i\delta/2} \left|-\right\rangle)^\dagger, \qquad \sigma^- = 2(e^{-i\delta/2} \left|-\right\rangle)(e^{i\delta/2} \left|+\right\rangle)^\dagger. \tag{4.3.137}$$

That is, if we re-define $\left|\pm'\right\rangle \equiv e^{\pm i\delta/2} \left|\pm\right\rangle$, followed by dropping the primes, we would have

$$\sigma^3 = \left|+\right\rangle \left\langle+\right| - \left|-\right\rangle \left\langle-\right|, \tag{4.3.138}$$
$$\sigma^+ = 2 \left|+\right\rangle \left\langle-\right|, \qquad \sigma^- = 2 \left|-\right\rangle \left\langle+\right|, \tag{4.3.139}$$

and again using $\sigma^1 = (\sigma^1 + \sigma^2)/2$ and $\sigma^2 = -i(\sigma^1 - \sigma^2)/2$,

$$\sigma^1 = \left|+\right\rangle \left\langle-\right| + \left|-\right\rangle \left\langle+\right|, \tag{4.3.140}$$
$$\sigma^2 = -i \left|+\right\rangle \left\langle-\right| + i \left|-\right\rangle \left\langle+\right|, \qquad \delta \in \mathbb{R}. \tag{4.3.141}$$

We see that the Pauli matrices in eq. (3.2.18) correspond to the matrix representations of σ^i in the basis built out of the unit norm eigenkets of σ^3, with an appropriate choice of phase.

Note that there is nothing special about choosing our basis as the eigenkets of σ^3 – we could have chosen the eigenkets of σ^1 or σ^2 as well. The analogous raising and lower operators can then be constructed from the remaining σ^is.

Finally, for $\widehat{U}$ unitary we have already noted that $\det(\widehat{U}\widehat{\sigma}^i\widehat{U}^\dagger) = \det\widehat{\sigma}^i$ and $\mathrm{Tr}\left[\widehat{U}\widehat{\sigma}^i\widehat{U}^\dagger\right] = \mathrm{Tr}\left[\widehat{\sigma}^i\right]$. Therefore, if we choose $\widehat{U}$ such that $\widehat{U}\widehat{\sigma}^i\widehat{U}^\dagger = \mathrm{diag}(1,-1)$ – since we now know the eigenvalues of each $\widehat{\sigma}^i$ are ± 1 – we readily deduce that

$$\det\widehat{\sigma}^i = -1, \qquad \mathrm{Tr}\left[\widehat{\sigma}^i\right] = 0. \tag{4.3.142}$$

(However, $\widehat{\sigma}^2\widehat{\sigma}^i\widehat{\sigma}^2 = -(\widehat{\sigma}^i)^*$ does not hold unless $\delta = 0$.)

4.3.3 Unitary Operation as Change of Orthonormal Basis

A unitary operator U is one whose inverse is its adjoint, i.e.,

$$U^\dagger U = UU^\dagger = \mathbb{I}. \tag{4.3.143}$$

Like their Hermitian counterparts, unitary operators play a special role in quantum theory. At a somewhat mundane level, they describe the change from one set of basis vectors to another. The analog in Euclidean space is the rotation matrix. But when the quantum dynamics is invariant under a particular change of basis – i.e., there is a symmetry enjoyed by the system at hand – then the eigenvectors of these unitary operators play a special role in classifying the dynamics itself. Also, in order to conserve probabilities, the time evolution operator, which takes an initial wave function(nal) of the quantum system and evolves it forward in time, is in fact a unitary operator itself.

Let us begin by understanding the action of a unitary operator as a change of basis vectors. Up till now we have assumed we can always find an orthonormal set of basis vectors $\{|i\rangle\,|i = 1,2,\dots,D\}$, for a D dimensional vector space. But just as in Euclidean space, this choice of basis vectors is not unique – in 3-space, for instance, we can rotate $\{\widehat{x},\widehat{y},\widehat{z}\}$ to some other $\{\widehat{x}',\widehat{y}',\widehat{z}'\}$ (i.e., redefine what we mean by the x, y and z axes). Hence, let us suppose we have found two such sets of orthonormal basis vectors

$$\{|1\rangle,\dots,|D\rangle\} \qquad \text{and} \qquad \{|1'\rangle,\dots,|D'\rangle\}. \tag{4.3.144}$$

(For concreteness the dimension of the vector space is D.) Remember a linear operator is defined by its action on every element of the vector space; equivalently, by linearity and completeness, it is defined by how it acts on each basis vector. We may thus define our unitary operator U via

$$U|i\rangle = |i'\rangle, \qquad i \in \{1,2,\dots,D\}. \tag{4.3.145}$$

Its matrix representation in the unprimed basis $\{|i\rangle\}$ is gotten by projecting both sides along $|j\rangle$.

$$\langle j|U|i\rangle = \langle j|i'\rangle, \qquad i,j \in \{1,2,\dots,D\}. \tag{4.3.146}$$

Is U really unitary? One way to verify this is through its matrix representation. We have

$$\langle j|U^\dagger|i\rangle = \langle i|U|j\rangle^* = \langle j'|i\rangle. \tag{4.3.147}$$

Whereas $U^\dagger U$ in matrix form is

$$\sum_k \langle j| U^\dagger |k\rangle \langle k| U |i\rangle = \sum_k \langle k| U |j\rangle^* \langle k| U |i\rangle \tag{4.3.148}$$

$$= \sum_k \langle k| i'\rangle \langle k| j'\rangle^* = \sum_k \langle j'| k\rangle \langle k| i'\rangle . \tag{4.3.149}$$

Because both $\{|k\rangle\}$ and $\{|k'\rangle\}$ form an orthonormal basis, we may invoke the completeness relation eq. (4.3.23) to deduce

$$\sum_k \langle j| U^\dagger |k\rangle \langle k| U |i\rangle = \langle j'| i'\rangle = \delta^j_i. \tag{4.3.150}$$

That is, we recover the unit matrix when we multiply the matrix representation of $U^\dagger$ to that of U.[16] Since we have not made any additional assumptions about the two arbitrary sets of orthonormal basis vectors, this verification of the unitary nature of U is itself independent of the choice of basis.

Alternatively, let us observe that the U defined in eq. (4.3.145) can be expressed as

$$U = \sum_j |j'\rangle \langle j| . \tag{4.3.151}$$

All we have to verify is $U|i\rangle = |i'\rangle$ for any $i \in \{1, 2, 3, \dots, D\}$.

$$U|i\rangle = \sum_j |j'\rangle \langle j| i\rangle = \sum_j |j'\rangle \delta^j_i = |i'\rangle . \tag{4.3.152}$$

The unitary nature of U can also be checked explicitly. Remember $(|\alpha\rangle\langle\beta|)^\dagger = |\beta\rangle\langle\alpha|$.

$$\begin{aligned} U^\dagger U &= \sum_j |j\rangle \langle j'| \sum_k |k'\rangle \langle k| \\ &= \sum_{j,k} |j\rangle \langle j'| k'\rangle \langle k| \\ &= \sum_{j,k} |j\rangle \delta^j_k \langle k| = \sum_j |j\rangle \langle j| = \mathbb{I}. \end{aligned} \tag{4.3.153}$$

The very last equality is just the completeness relation in eq. (4.3.23).

Starting from U defined in eq. (4.3.145) as a change-of-basis operator, we have shown U is unitary whenever the old $\{|i\rangle\}$ and new $\{|i'\rangle\}$ basis are given. Turning this around – suppose U is some arbitrary unitary linear operator, given some orthonormal basis $\{|i\rangle\}$ we can construct a new orthonormal basis $\{|j'\rangle\}$ by *defining*

$$|i'\rangle \equiv U|i\rangle . \tag{4.3.154}$$

All we have to show is that $\{|i'\rangle\}$ form an orthonormal set.

$$\langle j'| i'\rangle = (U|j\rangle)^\dagger (U|i\rangle) = \langle j \left| U^\dagger U \right| i\rangle = \langle j| i\rangle = \delta^j_i. \tag{4.3.155}$$

We may therefore pause to summarize our findings as follows.

[16] Strictly speaking we have only verified that the left inverse of U is $U^\dagger$, but for finite dimensional matrices, the left inverse is also the right inverse.

A linear operator U implements a change-of-basis from the orthonormal set $\{|i\rangle\}$ to some other (appropriately defined) orthonormal set $\{|i'\rangle\}$ if and only if U is unitary.

Change-of-basis of $\langle\alpha|\, i\rangle$ Given a bra $\langle\alpha|$, we may expand it either in the new $\{\langle i'|\}$ or old $\{\langle i|\}$ basis bras,

$$\langle\alpha| = \sum_i \langle\alpha|\, i\rangle \langle i| = \sum_i \langle\alpha|\, i'\rangle \langle i'| \,. \tag{4.3.156}$$

We can relate the components of expansions using $\langle i\,|U|\, k\rangle = \langle i|\, k'\rangle$ (cf. eq. (4.3.146)),

$$\begin{aligned}\sum_k \langle\alpha|\, k'\rangle \langle k'| &= \sum_i \langle\alpha|\, i\rangle \langle i| \\ &= \sum_{i,k} \langle\alpha|\, i\rangle \langle i|\, k'\rangle \langle k'| = \sum_k \left(\sum_i \langle\alpha|\, i\rangle \langle i\,|U|\, k\rangle\right) \langle k'| \,.\end{aligned} \tag{4.3.157}$$

Equating the coefficients of $\langle k'|$ on the left and (far-most) right hand sides, we see the components of the bra in the new basis can be gotten from that in the old basis using $\widehat{U}$,

$$\langle\alpha|\, k'\rangle = \sum_i \langle\alpha|\, i\rangle \langle i\,|U|\, k\rangle \,. \tag{4.3.158}$$

In words: the $\langle\alpha|$ row vector in the basis $\{\langle i'|\}$ is equal to U, written in the basis $\{\langle j\,|U|\, i\rangle\}$, acting (from the right) on the $\langle\alpha|\, i\rangle$ row vector, the $\langle\alpha|$ in the basis $\{\langle i|\}$. Moreover, in index notation,

$$\widehat{\alpha}_{k'} = \widehat{\alpha}_i \widehat{U}^i{}_k. \tag{4.3.159}$$

Problem 4.20. Change-of-basis of $\langle i|\, \alpha\rangle$ Given a vector $|\alpha\rangle$, and the orthonormal basis vectors $\{|i\rangle\}$, we can represent it as a column vector, where the ith component is $\langle i|\, \alpha\rangle$. What does this column vector look like in the basis $\{|i'\rangle\}$? Show that it is given by the matrix multiplication

$$\langle i'|\, \alpha\rangle = \sum_k \left\langle i \left|U^\dagger\right| k\right\rangle \langle k|\, \alpha\rangle \,, \qquad U\,|i\rangle = |i'\rangle \,. \tag{4.3.160}$$

In words: the $|\alpha\rangle$ column vector in the basis $\{|i'\rangle\}$ is equal to $U^\dagger$, written in the basis $\{\left\langle j \left|U^\dagger\right| i\right\rangle\}$, acting (from the left) on the $\langle i|\, \alpha\rangle$ column vector, the $|\alpha\rangle$ in the basis $\{|i\rangle\}$.

Furthermore, in index notation,

$$\widehat{\alpha}^{i'} = (\widehat{U}^\dagger)^i{}_k \widehat{\alpha}^k. \tag{4.3.161}$$

From the discussion on how components of bra(s) transform under a change-of-basis, together the analogous discussion of linear operators below, you will begin to see why in index notation, there is a need to distinguish between upper and lower indices – *they transform oppositely from each other.* □

Problem 4.21. 2D rotation in 3D. Let's rotate the basis vectors of the 2D plane, spanned by the x- and z-axis, by an angle θ. If $|1\rangle$, $|2\rangle$, and $|3\rangle$ respectively denote the unit vectors along the x, y, and z axes, how should the operator $U(\theta)$ act to rotate them? For example, since we are rotating the 13-plane, $U|2\rangle = |2\rangle$. (Drawing a picture may help.) Can you then write down the matrix representation $\langle j|U(\theta)|i\rangle$? □

Problem 4.22. Consider a 2 dimension vector space with the orthonormal basis $\{|1\rangle, |2\rangle\}$. The operator U is defined through its actions:

$$U|1\rangle = \frac{1}{\sqrt{2}}|1\rangle + \frac{i}{\sqrt{2}}|2\rangle, \tag{4.3.162}$$

$$U|2\rangle = \frac{i}{\sqrt{2}}|1\rangle + \frac{1}{\sqrt{2}}|2\rangle. \tag{4.3.163}$$

Is U unitary? Solve for its eigenvectors and eigenvalues. □

Change-of-basis of $\langle i|X|j\rangle$ Now we shall proceed to ask, how do we use U to change the matrix representation of some linear operator X written in the basis $\{|i\rangle\}$ to one in the basis $\{|i\rangle'\}$? Starting from $\langle i'|X|j'\rangle$ we insert the completeness relation eq. (4.3.23) in the basis $\{|i\rangle\}$, on both the left and the right,

$$\begin{aligned}\langle i'|X|j'\rangle &= \sum_{k,l}\langle i'|k\rangle\langle k|X|l\rangle\langle l|j'\rangle \\ &= \sum_{k,l}\left\langle i\left|U^\dagger\right|k\right\rangle\langle k|X|l\rangle\langle l|U|j\rangle = \left\langle i\left|U^\dagger XU\right|j\right\rangle,\end{aligned} \tag{4.3.164}$$

where we have recognized (from equations (4.3.146) and (4.3.147)) $\langle i'|k\rangle = \left\langle i\left|U^\dagger\right|k\right\rangle$ and $\langle l|j'\rangle = \langle l|U|j\rangle$. If we denote $\widehat{X}'$ as the matrix representation of X with respect to the primed basis; and $\widehat{X}$ and $\widehat{U}$ as their corresponding operators with respect to the unprimed basis, we recover the similarity transformation

$$\widehat{X}' = \widehat{U}^\dagger\widehat{X}\widehat{U}. \tag{4.3.165}$$

In index notation, with primes on the indices reminding us that the matrix is written in the primed basis $\{|i'\rangle\}$ and the unprimed indices in the unprimed basis $\{|i\rangle\}$,

$$\widehat{X}^{i'}{}_{j'} = (\widehat{U}^\dagger)^i{}_k\widehat{X}^k{}_l\widehat{U}^l{}_j. \tag{4.3.166}$$

As already alluded to, we see here the i and j indices transform "oppositely" from each other – so that, even in matrix algebra, if we view square matrices as (representations of) linear operators acting on some vector space, then the row index i should have a different position from the column index j so as to distinguish their transformation properties. This will allow us to readily implement that fact, when upper and lower indices are repeated, the pair transform as a scalar – for example, $X^{i'}{}_{i'} = X^i{}_i$.[17]

[17] This issue of upper versus lower indices will also appear in differential geometry. Given a pair of indices that transform oppositely from each other, we want them to be placed differently (upper vs. lower), so that when we set their labels equal – with Einstein summation in force – they automatically transforms as a scalar, since the pair of transformations will undo each other.

On the other hand, from the last equality of eq. (4.3.164), we may also view $\widehat{X}'$ as the matrix representation of the operator

$$X' \equiv U^\dagger X U \tag{4.3.167}$$

written in the old basis $\{|i\rangle\}$. To reiterate,

$$\langle i' |X| j'\rangle = \left\langle i \left|U^\dagger X U\right| j\right\rangle . \tag{4.3.168}$$

The next two theorems can be interpreted as telling us that the Hermitian/unitary nature of operators and their spectra are really basis-independent constructs.

> **Theorem** Let $X' \equiv U^\dagger X U$. If U is a unitary operator, X and X' shares the same spectrum.

Proof Let $|\lambda\rangle$ be the eigenvector and λ be the corresponding eigenvalue of X.

$$X|\lambda\rangle = \lambda|\lambda\rangle \tag{4.3.169}$$

By inserting a $\mathbb{I} = UU^\dagger$ between X and $|\lambda\rangle$; and multiplying both sides on the left by $U^\dagger$,

$$U^\dagger X U U^\dagger |\lambda\rangle = \lambda U^\dagger |\lambda\rangle , \tag{4.3.170}$$

$$X'(U^\dagger |\lambda\rangle) = \lambda(U^\dagger |\lambda\rangle). \tag{4.3.171}$$

That is, given the eigenvector $|\lambda\rangle$ of X with eigenvalue λ, the corresponding eigenvector of X' is $U^\dagger|\lambda\rangle$ with precisely the same eigenvalue λ.

> **Theorem.** Let $X' \equiv U^\dagger X U$. Then X is Hermitian iff X' is Hermitian. Moreover, X is unitary iff X' is unitary.

Proof If X is Hermitian, we consider $X'^\dagger$.

$$X'^\dagger = \left(U^\dagger X U\right)^\dagger = U^\dagger X^\dagger (U^\dagger)^\dagger = U^\dagger X U = X'. \tag{4.3.172}$$

If X is unitary we consider $X'^\dagger X'$.

$$X'^\dagger X' = \left(U^\dagger X U\right)^\dagger \left(U^\dagger X U\right) = U^\dagger X^\dagger U U^\dagger X U = U^\dagger X^\dagger X U = U^\dagger U = \mathbb{I}. \tag{4.3.173}$$

Remark We won't prove it here, but it is possible to find a unitary operator U, related to rotation in $\mathbb{R}^3$, that relates any one of the Pauli operators to the other

$$U^\dagger \sigma^i U = \sigma^j, \qquad i \neq j. \tag{4.3.174}$$

This is consistent with what we have already seen earlier, that all the $\{\sigma^k\}$ have the same spectrum $\{-1, +1\}$.

Physical Significance To put the significance of these statements in a physical context, recall the eigenvalues of an observable are possible outcomes of a physical experiment, while U describes a change of basis. Just as classical observables such as lengths, velocity, etc. should not depend on the coordinate system we use to compute the predictions of the underlying theory –

in the discussion of curved space(time)s we will see the analogy there is called *general covariance* – we see here that the possible experimental outcomes from a quantum system is independent of the choice of basis vectors we use to predict them. Also notice the very Hermitian and Unitary nature of a linear operator is invariant under a change of basis.

Diagonalization of observable Diagonalization of a matrix is nothing but the change-of-basis, expressing a linear operator X in some orthonormal basis $\{|i\rangle\}$ to one where it becomes a diagonal matrix with respect to the orthonormal eigenket basis $\{|\lambda\rangle\}$. That is, suppose you started with

$$X = \sum_k \lambda_k |\lambda_k\rangle \langle\lambda_k| \tag{4.3.175}$$

and defined the unitary operator

$$U|k\rangle = |\lambda_k\rangle \qquad \Leftrightarrow \qquad \langle i|U|k\rangle = \langle i|\lambda_k\rangle . \tag{4.3.176}$$

Notice the kth column of $\widehat{U}^i{}_k \equiv \langle i|U|k\rangle$ are the components of the kth unit norm eigenvector $|\lambda_k\rangle$ written in the $\{|i\rangle\}$ basis. This implies, via two insertions of the completeness relation in eq. (4.3.23),

$$X = \sum_{i,j,k} \lambda_k |i\rangle \langle i|\lambda_k\rangle \langle\lambda_k|j\rangle \langle j| . \tag{4.3.177}$$

Taking matrix elements,

$$\langle i|X|j\rangle = \widehat{X}^i{}_j = \sum_{k,l} \langle i|\lambda_k\rangle \lambda_k \delta^k_l \langle\lambda_l|j\rangle = \sum_{k,l} \widehat{U}^i{}_k \lambda_k \delta^k_l (\widehat{U}^\dagger)^l{}_j. \tag{4.3.178}$$

Multiplying both sides by $\widehat{U}^\dagger$ on the left and $\widehat{U}$ on the right, we have

$$\widehat{U}^\dagger \widehat{X} \widehat{U} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_D). \tag{4.3.179}$$

Schur decomposition. Not all linear operators are diagonalizable. However, we already know that any square matrix $\widehat{X}$ can be brought to an upper triangular form

$$\widehat{U}^\dagger \widehat{X} \widehat{U} = \widehat{\Gamma} + \widehat{N}, \qquad \widehat{\Gamma} \equiv \text{diag}\left(\lambda_1, \dots, \lambda_D\right), \tag{4.3.180}$$

where the $\{\lambda_i\}$ are the eigenvalues of X and $\widehat{N}$ is strictly upper triangular. We may now phrase the Schur decomposition as a change-of-basis from $\widehat{X}$ to its upper triangular form.

> Given a linear operator X, it is always possible to find an orthonormal basis such that its matrix representation is upper triangular, with its eigenvalues forming its diagonal elements.

Trace Earlier, we have already defined the trace of a linear operator X as

$$\text{Tr}\left[X\right] = \sum_i \langle i|X|i\rangle , \qquad \langle i|j\rangle = \delta^i_j. \tag{4.3.181}$$

The Trace yields a complex number.[18] Let us now see that this definition is independent of the orthonormal basis $\{|i\rangle\}$. Suppose we found a different set of orthonormal basis $\{|i'\rangle\}$, with $\langle i'| j'\rangle = \delta^i_j$. Now consider

$$\begin{aligned}\sum_i \langle i' |X| i'\rangle &= \sum_{i,j,k} \langle i'| j\rangle \langle j |X| k\rangle \langle k| i'\rangle = \sum_{i,j,k} \langle k| i'\rangle \langle i'| j\rangle \langle j |X| k\rangle \\ &= \sum_{j,k} \langle k| j\rangle \langle j |X| k\rangle = \sum_k \langle k |X| k\rangle .\end{aligned} \tag{4.3.182}$$

Because Tr is invariant under a change of basis, we can view the trace operation that turns an operator into a genuine scalar. This notion of a scalar is analogous to the quantities (pressure of a gas, temperature, etc.) that do not change no matter what coordinates one uses to compute/measure them.

Problem 4.23. Prove the following statements. For linear operators X and Y, and unitary operator U,

$$\mathrm{Tr}\,[XY] = \mathrm{Tr}\,[YX] \tag{4.3.183}$$

$$\mathrm{Tr}\left[U^\dagger X U\right] = \mathrm{Tr}\,[X] \tag{4.3.184}$$

The second identity tells you Tr is a basis-independent operation. □

Problem 4.24. Commutation Relations and Unitary Transformations The commutation relations between linear operators underlie much of the algebraic analysis of quantum systems exhibiting continuous symmetries.

Prove that commutation relations remain invariant under a change-of-basis. Specifically, suppose a set of operators $\{A^i | i = 1, 2, \dots, N\}$ obeys

$$\left[A^i, A^j\right] = i f^{ijk} A^k \tag{4.3.185}$$

for some constants $\{f^{ijk}\}$; then under

$$A'^i \equiv U^\dagger A^i U, \tag{4.3.186}$$

one obtains

$$\left[A'^i, A'^j\right] = i f^{ijk} A'^k \tag{4.3.187}$$

for the same f^{ijk}s. In actuality, U does not need to be unitary but merely invertible: namely, if $A'^i \equiv U^{-1} A^i U$, then eq. (4.3.187) still holds. □

4.3.4 Additional Problems

Problem 4.25. If $\{|i\rangle\, | i = 1, 2, 3 \dots, D\}$ is a set of orthonormal basis vectors, what is $\mathrm{Tr}\,[|j\rangle \langle k|]$, where $j, k \in \{1, 2, \dots, D\}$? □

[18] Be aware that the trace may not make sense in an infinite dimensional continuous vector space.

Problem 4.26. Verify the following Jacobi identity. For linear operators X, Y and Z,

$$[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]=0. \tag{4.3.188}$$

Furthermore, verify that

$$[X,Y]=-[Y,X], \qquad [X,Y+Z]=[X,Y]+[X,Z], \tag{4.3.189}$$

$$[X,YZ]=[X,Y]Z+Y[X,Z]. \tag{4.3.190}$$

The Jacobi identity appears not only within the context of Linear Algebra (the generators of continuous symmetries obey it, for e.g.); it also appears in differential geometry, leading to one of the Bianchi identities obeyed by the Riemann curvature tensor. $\square$

Problem 4.27. Find the unit norm eigenvectors that can be expressed as a linear combination of $|1\rangle$ and $|2\rangle$, and their corresponding eigenvalues, of the operator

$$X \equiv a\left(|1\rangle\langle 1|-|2\rangle\langle 2|+|1\rangle\langle 2|+|2\rangle\langle 1|\right). \tag{4.3.191}$$

Assume that $|1\rangle$ and $|2\rangle$ are orthogonal and of unit norm. (Hint: First calculate the matrix $\langle j|X|i\rangle$.)

Now consider the operators built out of the orthonormal basis vectors $\{|i\rangle\,|i=1,2,3\}$.

$$\begin{aligned} Y &\equiv a\left(|1\rangle\langle 1|-|2\rangle\langle 2|-|3\rangle\langle 3|\right), \\ Z &\equiv b\,|1\rangle\langle 1|-ib\,|2\rangle\langle 3|+ib\,|3\rangle\langle 2|. \end{aligned} \tag{4.3.192}$$

(In equations (4.3.191) and (4.3.192), a and b are real numbers.) Are Y and Z hermitian? Write down their matrix representations. Verify $[Y,Z]=0$ and proceed to simultaneously diagonalize Y and Z. $\square$

Problem 4.28. Pauli matrices re-visited. Refer to the Pauli matrices $\{\sigma^\mu\}$ defined in eq. (3.2.18). Let p_μ be a 4-component collection of real numbers. We may then view $p_\mu\sigma^\mu$ (where μ sums over 0 through 3) as a Hermitian operator acting on a 2 dimensional vector space.

1. Show that the eigenvalues $\lambda_\pm$ and corresponding unit norm eigenvectors $\xi^\pm$ of $p_i\sigma^i$ (where i sums over 1 through 3) are

$$\begin{aligned} {\xi^+}_{\mathrm{A}} &= \left(e^{-i\phi_p}\cos\left[\frac{\theta_p}{2}\right], \sin\left[\frac{\theta_p}{2}\right]\right)^{\mathrm{T}} \\ &= \frac{1}{\sqrt{2}}\sqrt{1-\frac{p_3}{|\vec{p}|}}\left(\frac{|\vec{p}|+p_3}{p_1+ip_2}, 1\right)^{\mathrm{T}} \end{aligned} \tag{4.3.193}$$

$$\begin{aligned} {\xi^-}_{\mathrm{A}} &= \left(-e^{-i\phi_p}\sin\left[\frac{\theta_p}{2}\right], \cos\left[\frac{\theta_p}{2}\right]\right)^{\mathrm{T}} \\ &= \frac{1}{\sqrt{2}}\sqrt{1+\frac{p_3}{|\vec{p}|}}\left(-\frac{|\vec{p}|-p_3}{p_1+ip_2}, 1\right)^{\mathrm{T}}. \end{aligned} \tag{4.3.194}$$

where we have employed spherical coordinates

$$p_i \equiv p\left(\sin\theta_p\cos\phi_p, \sin\theta_p\sin\phi_p, \cos\theta_p\right). \tag{4.3.195}$$

These are called the *helicity eigenstates* – eigenstates of spin along the 'momentum' p_i. Are they also eigenstates of $p_\mu\sigma^\mu$? (Hint: consider $[p_i\sigma^i, p_\mu\sigma^\mu]$.)

2. Explain why

$$p_i\widehat{\sigma}^i = \lambda_+\xi^+(\xi^+)^\dagger + \lambda_-\xi^-(\xi^-)^\dagger. \tag{4.3.196}$$

Can you write down the analogous expansion for $p_\mu\widehat{\sigma}^\mu$?

3. If we define the square root of an operator or matrix $\sqrt{A}$ as the solution to $\sqrt{A}\sqrt{A} = A$, write down the expansion for $\sqrt{p_\mu\widehat{\sigma}^\mu}$.

4. These 2 component spinors $\xi^\pm$ play a key role in the study of Lorentz symmetry in 4 spacetime dimensions. Consider applying an invertible transformation $L_{\text{A}}{}^B$ on these spinors, i.e., replace

$$(\xi^\pm)_{\text{A}} \to L_{\text{A}}{}^B(\xi^\pm)_{\text{B}}. \tag{4.3.197}$$

(The A and B indices run from 1 to 2, the components of $\xi^\pm$.) How does $p_\mu\widehat{\sigma}^\mu$ change under such a transformation? And, how does its determinant change?

□

Problem 4.29. Change-of-non-orthonormal-basis Not all change-of-basis involves a switch from one orthonormal set to another. Let us begin with the orthonormal basis $\{|i\rangle\}$ but switch to a non-orthonormal one $\{|i'\rangle\}$ and define the change-of-basis operator S by specifying the expansion coefficients $\{\widehat{S}^i{}_j\}$ in

$$S\,|i\rangle = \sum_j |j\rangle\,\widehat{S}^j{}_i \equiv |i'\rangle\,. \tag{4.3.198}$$

Explain why $\widehat{S}^i{}_j = \langle i\,|S|\,j\rangle$ is still $\langle i|\,j'\rangle$. (Compare with eq. (4.3.147).) On the other hand, since S is no longer unitary, its matrix representation $\widehat{S}$ is no longer a unitary matrix. Show, however, that the inverse transformation is directly related to the inverse matrix, which obeys $\widehat{S}^{-1}\widehat{S} = \mathbb{I}$:

$$|i\rangle = \sum_j |j'\rangle\left(\widehat{S}^{-1}\right)^j{}_i. \tag{4.3.199}$$

[19]As a concrete problem, let us perform the following change-of-basis, for $\theta \neq \phi$:

$$S\,|1\rangle = \cos(\theta)\,|1\rangle + \sin(\theta)\,|2\rangle \equiv |1'\rangle\,, \tag{4.3.200}$$
$$S\,|2\rangle = -\sin(\phi)\,|1\rangle + \cos(\phi)\,|2\rangle \equiv |2'\rangle\,. \tag{4.3.201}$$

Solve $\widehat{S}^{-1}$ and find $|1\rangle$ and $|2\rangle$ in terms of $\{|1'\rangle\,, |2'\rangle\}$. □

[19]Note that our discussion implicitly assumes $\widehat{S}^{-1}$ exists, for otherwise we are not performing a faithful coordinate transformation but discarding information about the vector space. As a simple 2D example, we could define $S\,|1\rangle = |1\rangle \equiv |1'\rangle$ and $S\,|2\rangle = |1\rangle \equiv |2'\rangle$ but this basically collapses the 2-dimensional vector space to a 1-dimensional one – i.e., we 'lose information' and $\widehat{S}^{-1}$ most certainly does not exist.

Problem 4.30. Schrödinger's Equation and Dyson Series The primary equation in quantum mechanics (and quantum field theory), governing how states evolve in time, is

$$i\hbar\partial_t \left|\psi(t)\right\rangle = H \left|\psi(t)\right\rangle, \tag{4.3.202}$$

where $\hbar \approx 1.054572 \times 10^{-34}$ J s is the reduced Planck's constant, and H is the Hamiltonian ($\equiv$ Hermitian total energy linear operator) of the system. The physics of a particular system is encoded within H.

Suppose H is independent of time, and suppose its orthonormal eigenkets $\{|E_i; n_j\rangle\}$ are known (n_j being the degeneracy label, running over all eigenkets with the same energy E_j), with $H|E_i; n_i\rangle = E_i |E_i; n_i\rangle$ and $\{E_i \in \mathbb{R}\}$, where we will assume the energies are discrete. Show that the solution to Schrödinger's equation in (4.3.202) is

$$|\psi(t)\rangle = \sum_{j,n_j} e^{-(i/\hbar)E_j t} |E_j; n_j\rangle \langle E_j; n_j| \psi(t=0)\rangle, \tag{4.3.203}$$

where $|\psi(t=0)\rangle$ is the initial condition, i.e., the state $|\psi(t)\rangle$ at $t = 0$. (Hint: Check that eq. (4.3.202) and the initial condition are satisfied.) Since the initial state was arbitrary, what you have verified is that the operator

$$U(t,t') \equiv \sum_{j,n_j} e^{-(i/\hbar)E_j(t-t')} |E_j; n_j\rangle \langle E_j; n_j| \tag{4.3.204}$$

obeys Schrödinger's equation,

$$i\hbar\partial_t U(t,t') = HU(t,t'). \tag{4.3.205}$$

Is $U(t,t')$ unitary? Explain what is the operator $U(t = t')$?

Express the expectation value $\langle\psi(t)|H|\psi(t)\rangle$ in terms of the energy eigenkets and eigenvalues. Compare it with the expectation value $\langle\psi(t=0)|H|\psi(t=0)\rangle$.

Time-Dependent Hamiltonian What if the Hamiltonian in Schrödinger's equation depends on time – what is the corresponding U? Consider the following (somewhat formal) solution for U.

$$\begin{aligned} U(t,t') &\equiv \mathbb{I} - \frac{i}{\hbar}\int_{t'}^{t} \mathrm{d}\tau_1 H(\tau_1) + \left(-\frac{i}{\hbar}\right)^2 \int_{t'}^{t} \mathrm{d}\tau_2 \int_{t'}^{\tau_2} \mathrm{d}\tau_1 H(\tau_2)H(\tau_1) + \dots && (4.3.206)\\ &= \mathbb{I} + \sum_{\ell=1}^{\infty} \mathcal{I}_\ell(t,t'), && (4.3.207)\end{aligned}$$

where the ℓ-nested integral $\mathcal{I}_\ell(t,t')$ is

$$\mathcal{I}_\ell(t,t') \equiv \left(-\frac{i}{\hbar}\right)^\ell \int_{t'}^{t} \mathrm{d}\tau_\ell \int_{t'}^{\tau_\ell} \mathrm{d}\tau_{\ell-1} \cdots \int_{t'}^{\tau_3} \mathrm{d}\tau_2 \int_{t'}^{\tau_2} \mathrm{d}\tau_1 H(\tau_\ell)H(\tau_{\ell-1})\dots H(\tau_2)H(\tau_1). \tag{4.3.208}$$

Remark Be aware that, if the Hamiltonian $H(t)$ depends on time, it may not commute with itself at different times, namely one *cannot* assume $[H(\tau_1), H(\tau_2)] = 0$ if $\tau_1 \neq \tau_2$; hence, the order of H in eq. (4.3.208) is important. Furthermore, this time-dependent Dyson series is often phrased in the following 'time-ordered' form:

$$U(t,t') = \mathsf{T}\exp\left(-\frac{i}{\hbar}\int_{t'}^{t} H(\tau)\mathrm{d}\tau\right), \tag{4.3.209}$$

where, upon Taylor expansion, the $(-i/\hbar)^\ell$ term (for $\ell \geq 0$) in the ensuing $\mathsf{T}-$ordered product is given by eq. (4.3.208).

Verify that, for $t > t'$,

$$i\hbar\partial_t U(t,t') = H(t)U(t,t'). \tag{4.3.210}$$

What is $U(t=t')$? You should be able to conclude that $|\psi(t)\rangle = U(t,t')\,|\psi(t')\rangle$. Hint: Start with $i\hbar\partial_t \mathcal{I}_\ell(t,t')$ and employ Leibniz's rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\alpha(t)}^{\beta(t)} F(t,z)\mathrm{d}z\right) = \int_{\alpha(t)}^{\beta(t)} \frac{\partial F(t,z)}{\partial t}\mathrm{d}z + F\left(t,\beta(t)\right)\beta'(t) - F\left(t,\alpha(t)\right)\alpha'(t). \tag{4.3.211}$$

Bonus Question 1: Can you prove Leibniz's rule, by say, using the limit definition of the derivative?
Bonus Question 2: Can you prove that $U(t,t')$ associated with such a time-dependent H is still unitary? $\square$

Problem 4.31. Dyson Series: Matrix Version A very similar problem to Problem (4.30) is given by the first order matrix equation

$$\partial_t \widehat{A}(t) = \widehat{B}(t), \tag{4.3.212}$$

with the initial condition

$$\widehat{A}(t=t') = \widehat{A}_0. \tag{4.3.213}$$

Here, $\widehat{A}$, $\widehat{B}$ and $\widehat{A}_0$ are $D\times D$ matrices. Argue that the solution is

$$\widehat{A}(t \geq t') = \mathsf{T}\exp\left(\int_{t'}^{t}\widehat{B}(\tau)\mathrm{d}\tau\right)\widehat{A}_0 \tag{4.3.214}$$

$$= \widehat{A}_0 + \sum_{\ell=1}^{+\infty}\int_{t'}^{t}\mathrm{d}\tau_\ell\int_{t'}^{\tau_\ell}\mathrm{d}\tau_{\ell-1}\cdots\int_{t'}^{\tau_3}\mathrm{d}\tau_2\int_{t'}^{\tau_2}\mathrm{d}\tau_1\widehat{B}(\tau_\ell)\widehat{B}(\tau_{\ell-1})\dots\widehat{B}(\tau_2)\widehat{B}(\tau_1)\widehat{A}_0. \tag{4.3.215}$$

In the second line, the ordering of the matrices is important: from the earliest $\widehat{B}$ on the rightmost, to the latest $\widehat{B}$ at the leftmost – i.e., $\tau_\ell > \tau_{\ell-1} > \dots > \tau_2 > \tau_1$. $\square$

Problem 4.32. If an operator A is simultaneously unitary and Hermitian, what is A? Hint: Diagonalize it first. $\square$

4.4 Tensor Products of Vector Spaces

In this section we will introduce the concept of a tensor product. It is a way to "multiply" vector spaces, through the product $\otimes$, to form a larger vector space. Tensor products not only arise in quantum theory but is present even in classical electrodynamics, gravitation and field theories of non-Abelian gauge fields interacting with spin$-1/2$ matter. In particular, tensor products arise in quantum theory when you need to, for example, simultaneously describe both the spatial wave-function and the spin of a particle.

Definition To set our notation, let us consider multiplying $N \geq 2$ distinct vector spaces, i.e., $V_1 \otimes V_2 \otimes \dots \otimes V_N$ to form a $V_{\rm L}$. We write the tensor product of a vector $|\alpha_1; 1\rangle$ from V_1, $|\alpha_2; 2\rangle$ from V_2 and so on through $|\alpha_N; N\rangle$ from V_N as

$$|\mathfrak{A}; L\rangle \equiv |\alpha_1; 1\rangle \otimes |\alpha_2; 2\rangle \otimes \dots \otimes |\alpha_N; N\rangle \,, \tag{4.4.1}$$

where it is understood the vector $|\alpha_i; i\rangle$ in the ith slot (from the left) is an element of the ith vector space V_i. As we now see, the tensor product is multi-linear because it obeys the following algebraic rules.

1. The tensor product is distributive over addition. For example,

$$|\alpha\rangle \otimes (|\alpha'\rangle + |\beta'\rangle) \otimes |\alpha''\rangle = |\alpha\rangle \otimes |\alpha'\rangle \otimes |\alpha''\rangle + |\alpha\rangle \otimes |\beta'\rangle \otimes |\alpha''\rangle \,. \tag{4.4.2}$$

2. Scalar multiplication can be factored out. For example,

$$c\,(|\alpha\rangle \otimes |\alpha'\rangle) = (c\,|\alpha\rangle) \otimes |\alpha'\rangle = |\alpha\rangle \otimes (c\,|\alpha'\rangle). \tag{4.4.3}$$

Our larger vector space $V_{\rm L}$ is spanned by all vectors of the form in eq. (4.4.1), meaning every vector in $V_{\rm L}$ can be expressed as a linear combination:

$$|\mathfrak{A}'; L\rangle \equiv \sum_{\alpha_1,\dots,\alpha_N} C^{\alpha_1,\dots,\alpha_N} \, |\alpha_1; 1\rangle \otimes |\alpha_2; 2\rangle \otimes \dots \otimes |\alpha_N; N\rangle \in V_{\rm L}. \tag{4.4.4}$$

(The $C^{\alpha_1,\dots,\alpha_N}$ is just a collection complex numbers.) In fact, if we let $\{|i; j\rangle \,|i = 1, 2, \dots, D_j\}$ be the basis vectors of the jth vector space V_j,

$$\begin{aligned} |\mathfrak{A}'; L\rangle = \sum_{\alpha_1,\dots,\alpha_N} \sum_{i_1,\dots,i_N} & C^{\alpha_1,\dots,\alpha_N} \, \langle i_1; 1|\, \alpha_1\rangle \, \langle i_2; 2|\, \alpha_2\rangle \dots \langle i_N; N|\, \alpha_N\rangle \\ & \times |i_1; 1\rangle \otimes |i_2; 2\rangle \otimes \dots \otimes |i_N; N\rangle \,. \end{aligned} \tag{4.4.5}$$

In other words, the basis vectors of this tensor product space $V_{\rm L}$ are formed from products of the basis vectors from each and every vector space $\{V_i\}$.

Dimension If the ith vector space V_i has dimension D_i, then the dimension of V_L itself is $D_1 D_2 \dots D_{N-1} D_N$. The reason is, for a given tensor product $|i_1; 1\rangle \otimes |i_2; 2\rangle \otimes \dots \otimes |i_N; N\rangle$, there are D_1 choices for $|i_1; 1\rangle$, D_2 choices for $|i_2; 2\rangle$, and so on.

Example Suppose we tensor two copies of the 2-dimensional vector space that the Pauli operators $\{\sigma^i\}$ act on. Each space is spanned by $|\pm\rangle$. The tensor product space is then spanned by the following 4 vectors

$$|1; L\rangle = |+\rangle \otimes |+\rangle \,, \qquad |2; L\rangle = |+\rangle \otimes |-\rangle \,, \tag{4.4.6}$$

$$|3;L\rangle = |-\rangle \otimes |+\rangle\,, \qquad |4;L\rangle = |-\rangle \otimes |-\rangle\,. \tag{4.4.7}$$

(Note that this ordering of the vectors is of course *not* unique.) In particular, an arbitrary state takes the form

$$|\mathfrak{A};L\rangle = C^{++}\,|+\rangle \otimes |+\rangle + C^{+-}\,|+\rangle \otimes |-\rangle + C^{-+}\,|-\rangle \otimes |+\rangle + C^{--}\,|-\rangle \otimes |-\rangle\,. \tag{4.4.8}$$

Adjoint and Inner Product Just as we can form tensor products of kets, we can do so for bras. We have

$$(|\alpha_1\rangle \otimes |\alpha_2\rangle \otimes \cdots \otimes |\alpha_N\rangle)^\dagger = \langle\alpha_1| \otimes \langle\alpha_2| \otimes \cdots \otimes \langle\alpha_N|\,, \tag{4.4.9}$$

where the ith slot from the left is a bra from the ith vector space V_i. We also have the inner product

$$\begin{aligned}&(\langle\alpha_1| \otimes \langle\alpha_2| \otimes \cdots \otimes \langle\alpha_N|)\,(c\,|\beta_1\rangle \otimes |\beta_2\rangle \otimes \cdots \otimes |\beta_N\rangle + d\,|\gamma_1\rangle \otimes |\gamma_2\rangle \otimes \cdots \otimes |\gamma_N\rangle)\\ &= c\,\langle\alpha_1|\,\beta_1\rangle\,\langle\alpha_2|\,\beta_2\rangle \ldots \langle\alpha_N|\,\beta_N\rangle + d\,\langle\alpha_1|\,\gamma_1\rangle\,\langle\alpha_2|\,\gamma_2\rangle \ldots \langle\alpha_N|\,\gamma_N\rangle\,,\end{aligned} \tag{4.4.10}$$

where c and d are complex numbers. For example, the orthonormal nature of the $\{|i_1;1\rangle \otimes \cdots \otimes |i_N;N\rangle\}$ follow from

$$\begin{aligned}(\langle j_1;1| \otimes \cdots \otimes \langle j_N;N|)\,(|i_1;1\rangle \otimes \cdots \otimes |i_N;N\rangle) &= \langle j_1;1|\,i_1;1\rangle\,\langle j_2;2|\,i_2;2\rangle \ldots \langle j_N;N|\,i_N;N\rangle\\ &= \delta^{j_1}_{i_1} \ldots \delta^{j_N}_{i_N}.\end{aligned} \tag{4.4.11}$$

Linear Operators If X_i is a linear operator acting on the ith vector space V_i, we can form a tensor product of them. Their operation is defined as

$$\begin{aligned}&(X_1 \otimes X_2 \otimes \cdots \otimes X_N)\,(c\,|\beta_1\rangle \otimes |\beta_2\rangle \otimes \cdots \otimes |\beta_N\rangle + d\,|\gamma_1\rangle \otimes |\gamma_2\rangle \otimes \cdots \otimes |\gamma_N\rangle)\\ &= c(X_1\,|\beta_1\rangle) \otimes (X_2\,|\beta_2\rangle) \otimes \cdots \otimes (X_N\,|\beta_N\rangle) + d(X_1\,|\gamma_1\rangle) \otimes (X_2\,|\gamma_2\rangle) \otimes \cdots \otimes (X_N\,|\gamma_N\rangle),\end{aligned} \tag{4.4.12}$$

where c and d are complex numbers.

The most general linear operator Y acting on our tensor product space V_{L} can be built out of the basis ket-bra operators.

$$Y = \sum_{\substack{i_1,\ldots,i_N\\ j_1,\ldots,j_N}} |i_1;1\rangle \otimes \cdots \otimes |i_N;N\rangle \widehat{Y}^{i_1\ldots i_N}{}_{j_1\ldots j_N}\,\langle j_1;1| \otimes \cdots \otimes \langle j_N;N|\,, \tag{4.4.13}$$

$$\widehat{Y}^{i_1\ldots i_N}{}_{j_1\ldots j_N} \in \mathbb{C}. \tag{4.4.14}$$

Due to the orthonormality condition in eq. (4.4.11), the action of Y on an arbitrary state

$$|\mathfrak{B}\rangle = \sum_{i_1\ldots i_N} \widehat{B}^{i_1\ldots i_N}\,|i_1;1\rangle \otimes \cdots \otimes |i_N;N\rangle \tag{4.4.15}$$

reads

$$Y\,|\mathfrak{B}\rangle = \sum_{\substack{i_1,\ldots,i_N\\ j_1,\ldots,j_N}} |i_1;1\rangle \otimes \cdots \otimes |i_N;N\rangle\,\widehat{Y}^{i_1\ldots i_N}{}_{j_1\ldots j_N}\widehat{B}^{j_1\ldots j_N}. \tag{4.4.16}$$

Problem 4.33. Tensor transformations Consider the state

$$|\mathfrak{A}'; L\rangle = \sum_{1 \le i_1 \le D_1} \sum_{1 \le i_2 \le D_2} \cdots \sum_{1 \le i_N \le D_N} T^{i_1 i_2 \dots i_{N-1} i_N} |i_1; 1\rangle \otimes |i_2; 2\rangle \otimes \cdots \otimes |i_N; N\rangle , \tag{4.4.17}$$

where $\{|i_j; j\rangle\}$ are the D_j orthonormal basis vectors spanning the jth vector space V_j, and $T^{i_1 i_2 \dots i_{N-1} i_N}$ are complex numbers. Consider a change of basis for each vector space, i.e., $|i; j\rangle \to |i'; j\rangle$. By defining the unitary operator that implements this change-of-basis

$$U \equiv {}_{(1)}U \otimes {}_{(2)}U \otimes \cdots \otimes {}_{(N)}U, \tag{4.4.18}$$

$${}_{(i)}U \equiv \sum_{1 \le j \le D_i} |j'; i\rangle \langle j; i| , \tag{4.4.19}$$

expand $|\mathfrak{A}'; L\rangle$ in the new basis $\{|j_1'; 1\rangle \otimes \cdots \otimes |j_N'; N\rangle\}$; this will necessarily involve the $U^\dagger$'s. Define the coefficients of this new basis via

$$|\mathfrak{A}'; L\rangle = \sum_{1 \le i_1' \le D_1} \sum_{1 \le i_2' \le D_2} \cdots \sum_{1 \le i_N' \le D_N} T'^{i_1' i_2' \dots i_{N-1}' i_N'} |i_1'; 1\rangle \otimes |i_2'; 2\rangle \otimes \cdots \otimes |i_N'; N\rangle . \tag{4.4.20}$$

Now relate $T'^{i_1' i_2' \dots i_{N-1}' i_N'}$ to the coefficients in the old basis $T^{i_1 i_2 \dots i_{N-1} i_N}$ using the matrix elements

$$\left({}_{(i)}\widehat{U}^\dagger \right)^j{}_k \equiv \left\langle j; i \left| \left({}_{(i)}U \right)^\dagger \right| k; i \right\rangle . \tag{4.4.21}$$

Can you perform a similar change-of-basis for the following dual vector?

$$\langle \mathfrak{A}'; L| = \sum_{1 \le i_1 \le D_1} \sum_{1 \le i_2 \le D_2} \cdots \sum_{1 \le i_N \le D_N} T_{i_1 i_2 \dots i_{N-1} i_N} \langle i_1; 1| \otimes \langle i_2; 2| \otimes \cdots \otimes \langle i_N; N| \tag{4.4.22}$$

In differential geometry, tensors will transform in analogous ways. □

Problem 4.34. Product Rule Suppose the collection of states $\{|\psi_i(t)\rangle \,|i = 1, 2, \dots, N\}$ depend on the real parameter t. Explain why the product rule of differentiation holds for their tensor product.

$$\begin{aligned}
&\partial_t \Big(|\psi_1(t)\rangle \otimes |\psi_2(t)\rangle \otimes \cdots \otimes |\psi_N(t)\rangle \Big) \\
&\quad = (\partial_t |\psi_1(t)\rangle) \otimes |\psi_2(t)\rangle \otimes \cdots \otimes |\psi_N(t)\rangle + |\psi_1(t)\rangle \otimes (\partial_t |\psi_2(t)\rangle) \otimes \cdots \otimes |\psi_N(t)\rangle \\
&\qquad + \cdots + |\psi_1(t)\rangle \otimes |\psi_2(t)\rangle \otimes \cdots \otimes (\partial_t |\psi_N(t)\rangle).
\end{aligned} \tag{4.4.23}$$

□

4.5 *Gram-Schmidt: Matrix Version

In this section, we shall summarize the Gram-Schmidt procedure of constructing $N \ge 2$ orthonormal vectors $\{\widehat{e}_{\mathrm{I}}\}$ starting from N linearly independent ones $\{v^i_{\mathrm{I}} | \mathrm{I} = 1, 2, \dots, N\}$. The starting point is the projector

$$P^{ij}(\widehat{n}) \equiv \delta^{ij} - \widehat{n}^i \widehat{n}^j = P^{ji}(\widehat{n}), \tag{4.5.1}$$

where $\widehat{n}$ is a unit vector; i.e., its ordinary dot product with itself is one, $\widehat{n}\cdot\widehat{n} = 1$. The key property of this projector, as its name suggests, is it extracts the component of a vector *perpendicular* to $\widehat{n}$. Namely, since

$$\widehat{n}^i P^{ij}(\widehat{n}) = 0 \tag{4.5.2}$$

the object

$$v_\perp^i \equiv P^{ij}(\widehat{n}) v^j \tag{4.5.3}$$

must necessarily obey

$$\widehat{n} \cdot \vec{v}_\perp = \widehat{n}^i P^{ij}(\widehat{n}) v^j = 0. \tag{4.5.4}$$

Starting from the first vector, we define the first unit vector as $\widehat{e}_1^i \equiv v_1^i/|\vec{v}_1|$; and proceed to construct the object

$$v_{\perp,2}^i \equiv P^{ij}(\widehat{e}_1) v_2^j \tag{4.5.5}$$

perpendicular to $\widehat{e}_1$. The unit length version of it is then $\widehat{e}_2 \equiv v_{\perp,2}^i/|\vec{v}_{\perp,2}|$.

To continue the process, we first extend the definition of the projector in eq. (4.5.1) to include N' *mutually perpendicular* unit vectors $\{\widehat{n}_\mathrm{I}|\mathrm{I} = 1, 2, \dots, N'\}$, where $\widehat{n}_\mathrm{I} \cdot \widehat{n}_\mathrm{J} = \delta_\mathrm{IJ}$, via the relation

$$P^{ij}(\widehat{n}_1, \widehat{n}_2, \dots, \widehat{n}_{N'}) \equiv \delta^{ij} - \widehat{n}_1^i \widehat{n}_1^j - \widehat{n}_2^i \widehat{n}_2^j \dots - \widehat{n}_{N'}^i \widehat{n}_{N'}^j, \tag{4.5.6}$$

where, for $1 \leq \mathrm{I} \leq N'$,

$$\widehat{n}_\mathrm{I}^i P^{ij}(\widehat{n}_1, \widehat{n}_2, \dots, \widehat{n}_{N'}) = 0. \tag{4.5.7}$$

Problem 4.35. Verify eq. (4.5.7). □

The third orthonormal vector is

$$v_{\perp,3}^i = P^{ij}(\widehat{e}_1, \widehat{e}_2) v_3^j, \tag{4.5.8}$$

$$0 = \widehat{e}_1 \cdot \vec{v}_{\perp,3} = \widehat{e}_2 \cdot \vec{v}_{\perp,3}; \tag{4.5.9}$$

and its unit version is $\widehat{e}_3 \equiv v_{\perp,3}^i/|\vec{v}_{\perp,3}|$. The fourth orthonormal vector is

$$v_{\perp,4}^i = P^{ij}(\widehat{e}_1, \widehat{e}_2, \widehat{e}_3) v_4^j, \tag{4.5.10}$$

$$0 = \widehat{e}_1 \cdot \vec{v}_{\perp,4} = \widehat{e}_2 \cdot \vec{v}_{\perp,4} = \widehat{e}_3 \cdot \vec{v}_{\perp,4}; \tag{4.5.11}$$

and its unit version is $\widehat{e}_4 \equiv v_{\perp,4}^i/|\vec{v}_{\perp,4}|$. And so on.

Problem 4.36. Completeness Relation When $N' = D$, i.e., when the set $\{\widehat{e}_\mathrm{I}\}$ form an orthonormal basis, argue from eq. (4.5.6) and (4.5.7) that $P^{ij}(\widehat{e}_1, \dots, \widehat{e}_D) = 0$ and arrive at the completeness relation

$$\delta^\mathrm{IJ} \widehat{e}_\mathrm{I}^i \widehat{e}_\mathrm{J}^j = \delta^{ij}. \tag{4.5.12}$$

Hint: For an arbitrary vector W^i, first explain why $\widehat{e}_\mathrm{I}^i P^{ij}(\widehat{e}_1, \dots, \widehat{e}_D) W^j = 0$ for $\mathrm{I} \in \{1, \dots, D\}$. Can a non-zero vector be perpendicular to the entire set $\{\widehat{e}_\mathrm{I}\}$? □

4.6 *Wedge Products, Determinants as Volumes of N Parallelepipeds

In this section, we shall understand why the volume of an arbitrary N dimensional parallelepiped in a $(D \geq 2)-$dimensional flat space, is intimately connected to both the notion of a wedge product and to matrix determinants. The material of this section also serves as a warm up to the discussion of infinitesimal volumes and differential forms in Chapters (9) and (11) below.

$N-$Parallelepipeds: Definition A $N-$parallelepiped is defined by N vectors. Let us define it iteratively, starting in 2D. There, a parallelepiped is simply a parallelogram, which in turn is defined by specifying two vectors $\vec{v}_1$ and $\vec{v}_2$ – i.e., one pair of parallel lines are specified by $\vec{v}_1$; and another pair by $\vec{v}_2$. By geometry, we note that the volume (aka area) of such a 2D parallelogram is simply the length of the first vector $|\vec{v}_1|$ multiplied by the perpendicular height of the other $|\vec{v}_2^\perp|$; where $\vec{v}_2^\perp$ is $\vec{v}_2$ with its component parallel to $\vec{v}_1$ subtracted out – recall eq. (4.2.32) –

$$\vec{v}_2^\perp \equiv \vec{v}_2 - \left(\vec{v}_2 \cdot \frac{\vec{v}_1}{|\vec{v}_1|}\right) \frac{\vec{v}_1}{|\vec{v}_1|}. \tag{4.6.1}$$

A quick calculation verifies $\vec{v}_2^\perp$ is perpendicular to $\vec{v}_1$:

$$\vec{v}_2^\perp \cdot \vec{v}_1 = \vec{v}_2 \cdot \vec{v}_1 - \left(\vec{v}_2 \cdot \frac{\vec{v}_1}{|\vec{v}_1|}\right) \frac{\vec{v}_1^2}{|\vec{v}_1|} = 0; \tag{4.6.2}$$

and therefore, via a direct calculation,

$$\begin{aligned} |\vec{v}_1|^2 |\vec{v}_2^\perp|^2 &= |\vec{v}_1|^2 \left(|\vec{v}_2|^2 - \frac{(\vec{v}_2 \cdot \vec{v}_1)^2}{|\vec{v}_1|^2}\right) = |\vec{v}_1|^2 |\vec{v}_2|^2 \sin(\theta)^2 & (4.6.3)\\ &= \left(v_1^{[1} v_2^{2]}\right)^2 & (4.6.4)\\ &= \left(\det \left[\begin{array}{cc} \vec{v}_1 & \vec{v}_2 \end{array}\right]_{2\times 2}\right)^2 = \left| \begin{pmatrix} \vec{v}_1 \\ 0 \end{pmatrix} \times \begin{pmatrix} \vec{v}_2 \\ 0 \end{pmatrix} \right|^2. & (4.6.5) \end{aligned}$$

The θ is the angle between $\vec{v}_1$ and $\vec{v}_2$; namely

$$\cos\theta \equiv \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1||\vec{v}_2|}. \tag{4.6.6}$$

We could also have defined the 2D volume to be the length of the second vector $|\vec{v}_2|$ multiplied by the perpendicular height of the first vector $|\vec{v}_1^\perp|$, where

$$\vec{v}_1^\perp \equiv \vec{v}_1 - \left(\vec{v}_1 \cdot \frac{\vec{v}_2}{|\vec{v}_2|}\right) \frac{\vec{v}_2}{|\vec{v}_2|}, \qquad \vec{v}_1^\perp \cdot \vec{v}_2 = 0. \tag{4.6.7}$$

(Drawing a figure helps, if such a projection process is unfamiliar.) One may readily that that these two definitions yield the same answer.

A 3D parallelepiped is a 'volume-of-translation', generated by a 2D parallelepiped (formed by, say, $\vec{v}_1$ and $\vec{v}_2$) translated along a third vector $\vec{v}_3$ *not* lying within the plane containing $\vec{v}_1$ and $\vec{v}_2$. (If $\vec{v}_3$ does lie in the plane, the 'volume-of-translation' remains a 2D plane, and the associated 3D volume would be zero.) By geometry, this 3D parallelepiped has volume given by the 2D one built from $\vec{v}_1$ and $\vec{v}_2$, multiplied by the perpendicular height of $\vec{v}_3$ from the 2D parallelepiped. From vector calculus, since $\vec{v}_1 \times \vec{v}_2$ not only has length equal to the 2D parallelogram spanned by $\vec{v}_1$ and $\vec{v}_2$, it is also perpendicular to it. This in turn tells us the 3D parallelepiped has volume that can be expressed through

$$|(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3| = (\text{Area of 2D parallelogram spanned by } \vec{v}_1 \text{ and } \vec{v}_2)\, |\vec{v}_3^{\perp}|, \tag{4.6.8}$$

$$= \left|\det \left[\begin{array}{ccc} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{array} \right]_{3\times 3}\right| \tag{4.6.9}$$

$$|\vec{v}_3^{\perp}| \equiv \frac{|(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3|}{|\vec{v}_1 \times \vec{v}_2|}. \tag{4.6.10}$$

The generalization of this 2D $\to$ 3D parallelepiped process may be iterated as many times as desired. Suppose an $N \geq 2$ dimensional parallelepiped has already been defined by the vectors $\vec{v}_1$ through $\vec{v}_N$. Then an $N + 1$ dimensional parallelepiped may be created as a 'volume-of-translation' by translating this N dimensional parallelepiped along some vector $\vec{v}_{N+1}$ *not* lying within the N dimensional space spanned by the $\vec{v}_1$ through $\vec{v}_N$.

The primary issue is: how do we compute the volume of an arbitrary parallelepiped, when its dimension is greater than 3? Our method of choice is by invoking the wedge product.

Wedge Product As we shall see, the wedge product $\wedge$ allows us to compute the volume of the $N-$parallelepiped, regardless of the dimension of the ambient space; this is in contrast to the formulas in equations (4.6.5) and (4.6.9), where the vectors have the same dimension as the ambient space.

The $N-$wedge product is a multi-linear fully anti-symmetric object, and exists in any number of dimensions higher than 1. It has the following defining properties, which will turn out to be very similar in spirit to those of the Levi-Civita symbol.

- The first key property of $\wedge$ is its anti-symmetry; for any two vectors $\vec{v}_1$ and $\vec{v}_2$ in $D \geq 2$ dimensions, we define

  $$\vec{v}_1 \wedge \vec{v}_2 = -\vec{v}_2 \wedge \vec{v}_1. \tag{4.6.11}$$

  Extending this definition to arbitrary number of vectors $\{\vec{v}_{\text{L}} | \text{L} = 1, 2, \dots, N | N \leq D\}$, we simply demand the wedge product be fully anti-symmetric:

  $$\vec{v}_{\text{L}_1} \wedge \dots \vec{v}_{\text{L}_i} \wedge \dots \wedge \vec{v}_{\text{L}_j} \wedge \dots \wedge \vec{v}_{\text{L}_N} = -\vec{v}_{\text{L}_1} \wedge \dots \vec{v}_{\text{L}_j} \wedge \dots \wedge \vec{v}_{\text{L}_i} \wedge \dots \wedge \vec{v}_{\text{L}_N}, \tag{4.6.12}$$

  where the $\{\text{L}_1, \dots, \text{L}_N\}$ are a permutation of $\{1, \dots, N\}$. If $N > D$, note that this wedge product is automatically zero – can you explain why?

- The second property is linearity in every 'slot'. If $\vec{w} = \alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2$ for scalars $\alpha_{1,2}$ and vectors $\vec{w}_{1,2}$, then for an arbitrary wedge product,

  $$\begin{aligned} \vec{v}_1 \wedge \dots \wedge \vec{w} \wedge \dots \wedge \vec{v}_N &= \vec{v}_1 \wedge \dots \wedge (\alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2) \wedge \dots \wedge \vec{v}_N \\ &= \alpha_1 \vec{v}_1 \wedge \dots \wedge \vec{w}_1 \wedge \dots \wedge \vec{v}_N + \alpha_2 \vec{v}_1 \wedge \dots \wedge \vec{w}_2 \wedge \dots \wedge \vec{v}_N. \end{aligned} \tag{4.6.13}$$

- The wedge product is associative; for example

  $$\vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3 = (\vec{v}_1 \wedge \vec{v}_2) \wedge \vec{v}_3 = \vec{v}_1 \wedge (\vec{v}_2 \wedge \vec{v}_3). \tag{4.6.14}$$

- The final property is, whenever we are dealing with orthonormal vectors $\{\widehat{e}_{\rm L}|{\rm L} = 1, 2, \dots, N\}$ with N less than or equal to D, their wedge product will be identified – up to a $\pm$ sign – with the volume of a unit cube in the N dimensions spanned by these vectors. For instance, in 2D, if $\widehat{e}_1 \equiv (1, 0)$ and $\widehat{e}_2 \equiv (0, 1)$,

$$\begin{aligned} \widehat{e}_1 \wedge \widehat{e}_2 \equiv &\text{ `right-handed' square defined by} \\ &\text{the vertices } (0, 0), (1, 0), (1, 1), (1, 0). \end{aligned} \tag{4.6.15}$$

Whereas

$$\begin{aligned} &\widehat{e}_2 \wedge \widehat{e}_1 = -\widehat{e}_1 \wedge \widehat{e}_2 \\ &\equiv \text{`left-handed' square defined by the vertices } (0, 0), (0, 1), (1, 1), (0, 1). \end{aligned} \tag{4.6.16}$$

These $\widehat{e}_1$ and $\widehat{e}_2$ may actually reside in higher dimensions too – the definitions do not really change, except we now have to insert additional zeroes to their components. For example, in 5D, $\widehat{e}_1$ and $\widehat{e}_2$ may read instead $(0, 0, 1, 0, 0)$ and $(0, 0, 0, 0, 1)$; then $\widehat{e}_1 \wedge \widehat{e}_2$ is the 'right-handed' unit square defined by $(0, 0, 0, 0, 0)$, $(0, 0, 1, 0, 0)$, $(0, 0, 1, 0, 1)$, $(0, 0, 0, 0, 1)$; and $\widehat{e}_2 \wedge \widehat{e}_1 = -\widehat{e}_1 \wedge \widehat{e}_2$ is the 'left-handed' one.

More generally, $D \geq 2$ dimensions, we may have $N \leq D$ orthonormal vectors $\{\widehat{e}_{\rm L}|{\rm L} = 1, 2, \dots, N\}$ such that

$$\widehat{e}_1 \wedge \dots \wedge \widehat{e}_N \tag{4.6.17}$$

describes a unit $N-$dimensional cube. Swapping any pair of these vectors, say $\widehat{e}_i \leftrightarrow \widehat{e}_j$ for $i \neq j$ and $i, j = 1, \dots, N$, amounts to switching orientations; i.e., from a left-handed to a right-handed cube or vice-versa.

Problem 4.37. As a start, show that in 2D space spanned by the orthonormal $\widehat{e}_{1,2}$,

$$\vec{v}_1 \wedge \vec{v}_2 = (v_1^1 v_2^2 - v_1^2 v_2^1)\widehat{e}_1 \wedge \widehat{e}_2. \tag{4.6.18}$$

Then show that in 3D space spanned by the orthonormal $\widehat{e}_{1,2,3}$,

$$\vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3 = \{(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3\}\, \widehat{e}_1 \wedge \widehat{e}_2 \wedge \widehat{e}_3. \tag{4.6.19}$$

The coefficients of $\widehat{e}_1 \wedge \widehat{e}_2$ and $\widehat{e}_1 \wedge \widehat{e}_2 \wedge \widehat{e}_3$ tell us, the wedge product does in fact recover the volume of the 2D parallelogram and 3D parallelepiped that we worked out using geometric considerations. □

Gram-Schmidt and Volume-as-Wedge Products We will now prove that $\vec{v}_1 \wedge \dots \wedge \vec{v}_N$ computes the volume of the N-parallelepiped defined by the $\vec{v}_{\rm L}$s, even when $N < D$, where D is the dimension of space. Firstly, the Gram-Schmidt process discussed earlier in this Chapter informs us, we may use the $\vec{v}_{\rm L}$s to derive N orthonormal basis vectors $\{\widehat{e}_{\rm L}|{\rm L} = 1, 2, \dots, N\}$ that spans the same $N-$dimensional space; i.e., there must be an invertible $N \times N$ transformation V such that

$$\vec{v}_{\rm L} = V^{\rm M}{}_{\rm L}\widehat{e}_{\rm M}, \qquad \widehat{e}_{\rm A} \cdot \widehat{e}_{\rm B} = \delta_{\rm AB}; \tag{4.6.20}$$

i.e., the L-th column of V are the components of the L-th vector written in the $\{\widehat{e}_{\rm A}\}$ basis. Suppose the $N-$wedge product does produce the volume of the associated $N-$parallelepiped:

$$\vec{v}_1 \wedge \dots \wedge \vec{v}_N = (\pm)(\text{vol. of } N- \text{ parallelepiped})\widehat{e}_1 \wedge \dots \wedge \widehat{e}_N. \tag{4.6.21}$$

If we now generate an $(N+1)$-parallelepiped by translating the N-parallelepiped along $\vec{v}_{N+1}$, the resulting volume must be – by the rules of Euclidean geometry –

$$(\text{volume of } N\text{-parallelepiped}) \times (\text{perpendicular height of } \vec{v}_{N+1} \text{ from } N\text{-parallelepiped}).$$

But the anti-symmetric character of the wedge product automatically projects out any component of $\vec{v}_{N+1}$ lying within the N-space spanned by the $\vec{v}_1, \dots, \vec{v}_N$. For, we may express

$$\vec{v}_{\rm N+1} = \sum_{\rm L=1}^{N} \chi^{\rm L} \vec{v}_{\rm L} + \vec{v}^{\perp}_{N+1} \tag{4.6.22}$$

– for appropriate scalars $\{\chi^{\rm L}\}$ and $\vec{v}^{\perp}_{N+1} \cdot \vec{v}_1 = \dots = \vec{v}^{\perp}_{N+1} \cdot \vec{v}_N = 0$ – and consider

$$\vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{N+1} = \sum_{\rm L=1}^{N} \chi^{\rm L} \vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{\rm L} + \vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}^{\perp}_{N+1}. \tag{4.6.23}$$

For a fixed L in the summation, $\vec{v}_{\rm L}$ must occur within the first N slots of the first wedge product on the right hand side. That means this first term must be zero, as, by anti-symmetry,

$$\vec{v}_1 \wedge \dots \wedge \vec{v}_{\rm L} \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{\rm L} = -\vec{v}_1 \wedge \dots \wedge \vec{v}_{\rm L} \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{\rm L}. \tag{4.6.24}$$

Hence, as claimed, only the components perpendicular to the $\{\vec{v}_{\rm L} | {\rm L} = 1, \dots, N\}$ survive.

$$\vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{N+1} = \vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}^{\perp}_{N+1} \tag{4.6.25}$$

where the $\{\widehat{e}_1, \dots, \widehat{e}_N\}$ form the orthonormal basis vectors, then by defining the new unit vector $\widehat{e}_{N+1} \equiv \vec{v}^{\perp}_{N+1}/|\vec{v}^{\perp}_{N+1}|$, we deduce

$$\begin{aligned} \vec{v}_1 \wedge \dots \wedge \vec{v}_N \wedge \vec{v}_{N+1} &= (\pm)(\text{vol. of } N-\text{parallelepiped})|\vec{v}^{\perp}_{N+1}|\widehat{e}_1 \wedge \dots \wedge \widehat{e}_N \wedge \widehat{e}_{N+1} && (4.6.26) \\ &= (\pm)(\text{vol. of } (N+1)-\text{parallelepiped})\widehat{e}_1 \wedge \dots \wedge \widehat{e}_N \wedge \widehat{e}_{N+1}, && (4.6.27) \end{aligned}$$

since $|\vec{v}^{\perp}_{N+1}|$ *is* the perpendicular height of $\vec{v}_{N+1}$ from the N-parallelepiped. By induction on the number of vectors N, we have proven that the wedge product of N vectors does in fact yield the volume of their associated parallelepiped.

Matrix Determinants as Volumes Now that we have proven eq. (4.6.21), let us also observe that

$$\begin{aligned} &(\pm)(\text{Volume of } N-\text{parallelepiped defined by } \vec{v}_1 \dots \vec{v}_N)(\widehat{e}_1 \wedge \dots \wedge \widehat{e}_N) \\ &= \vec{v}_1 \wedge \dots \wedge \vec{v}_N \\ &= V^{\rm M_1}{}_1 \dots V^{\rm M_N}{}_N \widehat{e}_{\rm M_1} \wedge \dots \wedge \widehat{e}_{\rm M_N} \\ &= \left(\epsilon_{\rm M_1 \dots M_N} V^{\rm M_1}{}_1 \dots V^{\rm M_N}{}_N\right) \widehat{e}_1 \wedge \dots \wedge \widehat{e}_N \end{aligned}$$

$$= \left(\det V^{\mathrm{A}}{}_{\mathrm{B}}\right) \widehat{e}_1 \wedge \dots \wedge \widehat{e}_N = \det \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_N \end{bmatrix} \widehat{e}_1 \wedge \dots \wedge \widehat{e}_N. \tag{4.6.28}$$

That is, the volume of the $N-$parallelepiped *is* a determinant, but that of the $N \times N$ matrix formed from the components of the $D-$dimensional $\vec{v}_{\mathrm{L}}$s expressed in the orthonormal basis $\{\widehat{e}_{\mathrm{L}} | \mathrm{L} = 1, \dots, N\}$. When $N = D = 2$, this recovers eq. (4.6.5); and when $N = D = 3$, this recovers eq. (4.6.9).

Basis-Independence Observe that the choice of orthonormal basis $\{\widehat{e}_{\mathrm{L}} | \mathrm{L} = 1, 2, \dots, N\}$ does not affect the result of the volume in eq. (4.6.21), because under an orthogonal transformation $\widehat{e}_{\mathrm{L}} \equiv O \cdot \widehat{e}'_{\mathrm{L}}$ for all $O \in \mathrm{O}_N$,

$$\begin{aligned} \widehat{e}_1 \wedge \dots \wedge \widehat{e}_N &= O^{\mathrm{L}_1}{}_1 \dots O^{\mathrm{L}_N}{}_N \widehat{e}'_{\mathrm{L}_1} \wedge \dots \wedge \widehat{e}'_{\mathrm{L}_N} && (4.6.29) \\ &= \epsilon_{\mathrm{L}_1 \dots \mathrm{L}_N} O^{\mathrm{L}_1}{}_1 \dots O^{\mathrm{L}_N}{}_N \widehat{e}'_1 \wedge \dots \wedge \widehat{e}'_N && (4.6.30) \\ &= (\det O) \widehat{e}'_1 \wedge \dots \wedge \widehat{e}'_N && (4.6.31) \\ &= (\pm) \widehat{e}'_1 \wedge \dots \wedge \widehat{e}'_N. && (4.6.32) \end{aligned}$$

The Levi-Civita symbol $\epsilon_{\mathrm{L}_1 \dots \mathrm{L}_N}$ above lives in the N dimensions spanned by $\{\vec{v}_{\mathrm{L}}\}$; with $\epsilon_{1 \dots N} \equiv 1$.

Order-Independence Note too, although we have built up the $N-$parallelepiped iteratively; the result in eq. (4.6.21) tells us the order in which we did so does not actually matter – for, all that would change is the order of the $\widehat{e}_{\mathrm{L}}$s within the wedge product. For example, to construct a 3−parallelepiped, we may first choose the 2D one defined by $(\vec{v}_1, \vec{v}_2)$, then translate it along $\vec{v}_2$; or, $(\vec{v}_1, \vec{v}_3)$, then translate it along $\vec{v}_3$; or, $(\vec{v}_2, \vec{v}_3)$, then translate it along $\vec{v}_1$. These will yield, respectively, $\vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3$, $\vec{v}_1 \wedge \vec{v}_3 \wedge \vec{v}_2$, and $\vec{v}_2 \wedge \vec{v}_3 \wedge \vec{v}_1$; which only differ by an overall $\pm$ sign.

Example: 3−parallelepiped in 5D Suppose we have the following 3 vectors:

$$\begin{aligned} \vec{v}_1 &\doteq (3, 0, 1, 2, 0), && (4.6.33) \\ \vec{v}_2 &\doteq (1, 0, 1, 1, 0), && (4.6.34) \\ \vec{v}_3 &\doteq (5, 0, 3, 7, 0). && (4.6.35) \end{aligned}$$

By a direct calculation,

$$\begin{aligned} \vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3 &= (3\widehat{e}_1 + \widehat{e}_3 + 2\widehat{e}_4) \wedge (\widehat{e}_1 + \widehat{e}_3 + \widehat{e}_4) \wedge (5\widehat{e}_1 + 3\widehat{e}_3 + 7\widehat{e}_4) && (4.6.36) \\ &= 6\widehat{e}_1 \wedge \widehat{e}_3 \wedge \widehat{e}_4. && (4.6.37) \end{aligned}$$

(The second equality requires repeated use of the linearity and anti-symmetric properties of $\wedge$.) On the other hand, since the second and fifth components of the $\vec{v}$s are zero, we may simply focus on the non-zero components; namely,

$$\begin{aligned} \vec{v}'_1 &\doteq (3, 1, 2), && (4.6.38) \\ \vec{v}'_2 &\doteq (1, 1, 1), && (4.6.39) \\ \vec{v}'_3 &\doteq (5, 3, 7). && (4.6.40) \end{aligned}$$

This allows us to check the above wedge product calculation by re-evaluating the volume as

$$(\vec{v}'_1 \times \vec{v}'_2) \cdot \vec{v}'_3 = 6. \tag{4.6.41}$$

Yet another approach is to employ Gram-Schmidt. The length of the first vector is $|\vec{v}_1| = \sqrt{9+1+4} = \sqrt{14}$, and

$$\vec{v}_1 \equiv \sqrt{14} \cdot \widehat{e}_1. \tag{4.6.42}$$

Next, we project out from $\vec{v}_2$ the component parallel to $\vec{v}_1$ to obtain the second orthonormal basis vector:

$$\begin{aligned} \vec{v}_2^{\perp} &= \vec{v}_2 - (\vec{v}_2 \cdot \vec{e}_1)\widehat{e}_1 &(4.6.43)\\ &\equiv \sqrt{\frac{3}{7}}\widehat{e}_2. &(4.6.44)\end{aligned}$$

Thus

$$\vec{v}_2 = 3\sqrt{\frac{2}{7}}\widehat{e}_1 + \sqrt{\frac{3}{7}}\widehat{e}_2. \tag{4.6.45}$$

Similar considerations will lead us to

$$\begin{aligned} \vec{v}_3^{\perp} &= \vec{v}_3 - (\vec{v}_3 \cdot \vec{e}_1)\widehat{e}_1 - (\vec{v}_3 \cdot \widehat{e}_2)\widehat{e}_2 &(4.6.46)\\ &\equiv \sqrt{6}\widehat{e}_3. &(4.6.47)\end{aligned}$$

and

$$\vec{v}_3 = 16\sqrt{\frac{2}{7}}\widehat{e}_1 + 3\sqrt{\frac{3}{7}}\widehat{e}_2 + \sqrt{6}\widehat{e}_3. \tag{4.6.48}$$

The wedge product now produces

$$\vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3 = 6\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3. \tag{4.6.49}$$

Problem 4.38. Volume of parallelogram Use the wedge product (and Gram-Schmidt) to find the volume of the 3−parallelepiped in 5D (flat) space defined by

$$\begin{aligned} \vec{v}_1 &= (1, 1, 1, 1, 1), &(4.6.50)\\ \vec{v}_2 &= (1, 2, 3, 4, 5), &(4.6.51)\\ \vec{v}_3 &= (-1, 1, -1, 1, -4). &(4.6.52)\end{aligned}$$

Answer: $\vec{v}_1 \wedge \vec{v}_2 \wedge \vec{v}_3 = \pm 2\sqrt{165} \cdot \widehat{e}_1 \wedge \widehat{e}_2 \wedge \widehat{e}_3$, for the appropriate definitions of these $\{\widehat{e}_{1,2,3}\}$. □

4.7 *Normal Operators

Definition A normal operator N is one that commutes with its own adjoint.

$$\left[N, N^{\dagger}\right] = NN^{\dagger} - N^{\dagger}N = 0 \tag{4.7.1}$$

Hermitian operators are a special case.

An important fact is:

Diagonalizability An operator is diagonalizable if and only if it is normal.

Proof If N is diagonalizable, we may use its eigenvectors $\{|\lambda_k\rangle\}$ which obey $N|\lambda_k\rangle = \lambda_k|\lambda_k\rangle$, $k = 1, 2, \dots$, to express

$$N = \sum_k |\lambda_k\rangle \lambda_k \langle\lambda_k| . \tag{4.7.2}$$

A direct calculation would show

$$NN^\dagger - N^\dagger N = \sum_k |\lambda_k\rangle |\lambda_k|^2 \langle\lambda_k| - \sum_k |\lambda_k\rangle |\lambda_k|^2 \langle\lambda_k| = 0. \tag{4.7.3}$$

Now, suppose N is normal. Let us recall: If N acts on a $D-$dimensional space, its characteristic equation would hand us a polynomial of degree D. By the fundamental theorem of algebra we are guaranteed D solutions for its eigenvalues. Hence, whether or not an operator is diagonalizable amounts to asking if its eigenvectors can form an orthonormal basis.

Then if $|\lambda_k\rangle$ is its kth eigenvector,

$$(N - \lambda_k)|\lambda_k\rangle = 0. \tag{4.7.4}$$

Taking the inner product of $(N - \lambda_k)|\lambda_k\rangle$ with itself, and employing $[N^\dagger, N] = 0$,

$$\langle\lambda_k| \left(N^\dagger - \lambda_k^*\right)(N - \lambda_k)|\lambda_k\rangle = 0, \tag{4.7.5}$$

$$\langle\lambda_k| (N - \lambda_k)\left(N^\dagger - \lambda_k^*\right)|\lambda_k\rangle = 0. \tag{4.7.6}$$

The second equality tells us, the inner product of the vector $(N^\dagger - \lambda_k^*)|\lambda_k\rangle$ with itself is zero – i.e., it must be the zero vector. That in turn implies,

$$N^\dagger|\lambda_k\rangle = \lambda_k^*|\lambda_k\rangle \tag{4.7.7}$$

$$\langle\lambda_k| N = \langle\lambda_k| \lambda_k. \tag{4.7.8}$$

For $\lambda_k \neq \lambda_l$, we may act $\langle\lambda_l| N = \langle\lambda_l| \lambda_l$ on $|\lambda_k\rangle$.

$$\langle\lambda_l| N |\lambda_k\rangle = \lambda_l \langle\lambda_l| \lambda_k\rangle \tag{4.7.9}$$

Whereas acting $\langle\lambda_l|$ on the eigenvector equation $N|\lambda_k\rangle = \lambda_k|\lambda_k\rangle$ yields instead

$$\langle\lambda_l| N |\lambda_k\rangle = \lambda_k \langle\lambda_l| \lambda_k\rangle . \tag{4.7.10}$$

Subtracting equations (4.7.9) and (4.7.10),

$$(\lambda_l - \lambda_k)\langle\lambda_l| \lambda_k\rangle = 0. \tag{4.7.11}$$

By assumption, $\lambda_l - \lambda_k \neq 0$. Therefore $\langle\lambda_l| \lambda_k\rangle = 0$: eigenvectors of distinct eigenvalues are orthogonal. For eigenvectors belonging to a degenerate subspace, we may use Gram-Schmidt to construct an orthonormal set. ☐

Problem 4.39. Compatible operators Prove that two normal operators N and M are simultaneously diagonalizable if and only if they are compatible; i.e., iff $[N, M] = 0$. ☐

5 Continuous Vector Spaces and Infinite $D-$Space

This Chapter deals with vector spaces with infinite dimensionality and continuous spectra. To make this topic rigorous is beyond the scope of these notes; but the interested reader should consult the functional analysis portion of the math literature. Our goal here is a practical one: we want to be comfortable enough with continuous spaces to solve problems in quantum mechanics and (quantum and classical) field theory.

5.1 Dirac's δ-"function", Eigenket integrals, and Continuous (Lie Group) Operators

Dirac's δ-"function" and its representations We will see that transitioning from discrete, finite dimensional vector spaces to continuous ones means summations become integrals; while Kronecker-δs will be replaced with Dirac-δ functions. In case the latter is not familiar, the Dirac-δ function of one variable is to be viewed as an object that occurs within an integral, and is defined via

$$\int_a^b f(x')\delta(x'-x)\mathrm{d}x' = f(x), \tag{5.1.1}$$

for all a less than x and all b greater than x, i.e., $a < x < b$. This indicates $\delta(x'-x)$ has to be sharply peaked at $x' = x$ and zero everywhere, since the result of integral picks out the value of f solely at x.

The Dirac δ-function[20] is often loosely viewed as $\delta(x) = 0$ when $x \neq 0$ and $\delta(x) = \infty$ when $x = 0$. An alternate approach is to define $\delta(x)$ as a sequence of functions more and more sharply peaked at $x = 0$, whose integral over the real line is unity. Three examples are

$$\delta(x) = \lim_{\epsilon\to 0^+} \Theta\left(\frac{\epsilon}{2} - |x|\right)\frac{1}{\epsilon} \tag{5.1.2}$$

$$= \lim_{\epsilon\to 0^+} \frac{e^{-\frac{|x|}{\epsilon}}}{2\epsilon} \tag{5.1.3}$$

$$= \lim_{\epsilon\to 0^+} \frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2} \tag{5.1.4}$$

For the first equality, $\Theta(z)$ is the step function, defined to be

$$\begin{aligned}\Theta(z) &= 1, \qquad \text{for } z > 0 \\ &= 0, \qquad \text{for } z < 0.\end{aligned} \tag{5.1.5}$$

Problem 5.1. Justify these three definitions of $\delta(x)$. What happens, for finite $x \neq 0$, when $\epsilon \to 0^+$? Then, by holding ϵ fixed, integrate them over the real line, before proceeding to set $\epsilon \to 0^+$. □

[20] In the rigorous mathematical literature, Dirac's δ is not a function but a *distribution*, whose theory is due to Laurent Schwartz.

For later use, we record the following integral representation of the Dirac δ-function.

$$\int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} e^{i\omega(z-z')} = \delta(z-z') \tag{5.1.6}$$

Finally, for functions defined within the interval $-1 < x < +1$, the following is yet another representation of the Dirac delta function:

$$\delta(x) = \lim_{n\to+\infty} \delta_n(x), \tag{5.1.7}$$

$$\delta_n(x) \equiv (1-x^2)^n \frac{(2n+1)!}{2^{2n+1}(n!)^2}; \tag{5.1.8}$$

where $n \geq 0$ is to be viewed as a non-negative integer. We may understand eq. (5.1.7) heuristically as follows. Because the even function $P(x) \equiv 1-x^2$ peaks at $P(x=0) = 1$ and falls to zero as $x \to \pm 1$, that means the non-zero portion of $P(x)^n$, for some large $n \gg 1$, will be increasingly localized around $x \approx 0$; namely, any number with magnitude less than unity, when raised to a large positive power, will yield a yet smaller number. The factorials multiplying $(1-x^2)^n$ in eq. (5.1.7) ensure the total area of the right-hand-side is still unity. This representation in eq. (5.1.7) plays a central role in the Weierstrass approximation theorem,[21] which states that any continuous function $f(x)$ defined within a finite interval on the real line, say $a \leq x \leq b$, may be approximated by a polynomial $P_n(x)$ of degree n, by – cf. eq. (5.1.8) – arguing that

$$f(x) = \lim_{n\to+\infty} P_n(x), \tag{5.1.9}$$

$$P_n(x) \equiv \int_a^b \delta_n\left(\frac{x-x'}{b-a}\right) f(x') \frac{\mathrm{d}x'}{b-a}. \tag{5.1.10}$$

That is, if such an argument may be carried out, we would have justified the relation in eq. (5.1.7), that for any $a < x < b$,

$$f(x) = \int_a^b f(x')\delta\left(\frac{x-x'}{b-a}\right) \frac{\mathrm{d}x'}{b-a}. \tag{5.1.11}$$

Problem 5.2. Dirac as the derivative of Heaviside Can you justify the following?

$$\Theta(z-z') = \int_{z_0}^{z} \mathrm{d}z''\delta(z''-z'), \qquad z' > z_0. \tag{5.1.12}$$

We may therefore assert the derivative of the step function is the δ-function,

$$\Theta'(z-z') = \delta(z-z'). \tag{5.1.13}$$

Somewhat more rigorously, we may refer to the integral representation of the step function in eq. (6.3.41) below; and thereby justify its counterpart for the Dirac $\delta-$function in eq. (5.1.6). □

[21] A discourse more detailed than the one here may be found in Byron and Fuller [14].

Problem 5.3. Weierstrass' polynomial approximation In this problem we shall work out an explicit example of the Weierstrass' polynomial approximation using the δ_n in eq. (5.1.8). Specifically, let us obtain a polynomial to approximate the power law

$$f(x) = x^p \tag{5.1.14}$$

for $0 < p < 1$ near $x \gtrapprox 0$. The reason for choosing such a power law, is that while the derivatives of $f(x)$ do not exist at $x = 0$ and hence there is no Taylor expansion available about the origin, Weierstrass tells us we may nonetheless still produce a polynomial $P_n(x)$ that describes $f(x)$ arbitrarily accurately as $n \to \infty$.

Firstly, show that $\delta_n(x)$ is normalized such that

$$\int_{-1}^{+1} \delta_n(x') \mathrm{d}x' = 1. \tag{5.1.15}$$

Then, choose a p and some large $n \gg 1$, and proceed to work out $P_n(x)$ in eq. (5.1.10). Assume $a = 0$ and pick $b \sim \mathcal{O}(\text{few})$. Compare $P_n(x)$ with x^p by plotting the two on the same axes. Try to vary n and the interval $[0, b]$ too.

Remark Unlike the Taylor series approximation $f(x) \approx \sum_{\ell=0}^{n} c_\ell x^\ell$ around $x = 0$, where $c_\ell = (1/\ell!) f^{(\ell)}(0)$, notice all the coefficients $\{c_\ell\}$ of the Weierstrass approximation change as the highest power n is altered. For instance, for $p = 1/3$ and $[a, b] = [0, 1]$, eq. (5.1.10) would yield

$$P_2(x) = \frac{45x^4}{64} - \frac{45x^3}{28} + \frac{9x^2}{32} + \frac{135x}{182} + \frac{81}{256}, \tag{5.1.16}$$

$$P_3(x) = -\frac{105x^6}{128} + \frac{45x^5}{16} - \frac{315x^4}{128} - \frac{15x^3}{26} + \frac{189x^2}{512} + \frac{405x}{494} + \frac{1701}{5632}. \tag{5.1.17}$$

These $P_2(x)$ and $P_3(x)$ are very poor approximations to $x^{1/3}$ because 2 and 3 are not much greater than unity; but they illustrate clearly the n dependence of the polynomial coefficients. – for e.g., $c_0(n = 2) = 81/256$ whereas $c_0(n = 3) = 1701/5632$. □

Dirac's δ-"function": Properties Returning to the general discussion of Dirac $\delta-$functions, several of its properties are worth highlighting.

- From eq. (5.1.13) – that a $\delta(z - z')$ follows from taking the derivative of a discontinuous function (in this case, $\Theta(z - z')$ the Heaviside step function) – will be important for the study of Green's functions of ordinary and partial differential operators. Heuristically: the abrupt 'jump' $\Delta f \equiv f(a + 0^+) - f(a - 0^+)$ of a discontinuous function f at $x = a$ is accounted for by the $\delta-$function via $f'(x = a)\mathrm{d}x = \delta(x - a)(\Delta f)\mathrm{d}x$.

- If the argument of the δ-function is a function f of some variable z, then as long as $f'(z) \neq 0$ whenever $f(z) = 0$, it may be re-written as

$$\delta\left(f(z)\right) = \sum_{z_i \equiv i\text{th zero of } f(z)} \frac{\delta(z - z_i)}{|f'(z_i)|}. \tag{5.1.18}$$

To justify this we recall the fact that, the δ-function itself is non-zero only when its argument is zero. This explains why we sum over the zeros of $f(z)$. Now we need to fix the coefficient of the δ-function near each zero. That is, what are the φ_i's in

$$\delta\left(f(z)\right) = \sum_{z_i \equiv i\text{th zero of } f(z)} \frac{\delta(z - z_i)}{\varphi_i}? \tag{5.1.19}$$

We now use the fact that integrating a δ-function around the small neighborhood of the ith zero of $f(z)$ *with respect to* f has to yield unity. It makes sense to treat f as an integration variable near its zero because we have assumed its slope is non-zero, and therefore near its ith zero,

$$f(z) = f'(z_i)(z - z_i) + \mathcal{O}((z - z_i)^2), \tag{5.1.20}$$

$$\Rightarrow \qquad \mathrm{d}f = f'(z_i)\mathrm{d}z + \mathcal{O}((z - z_i)^1)\mathrm{d}z. \tag{5.1.21}$$

The integration around the ith zero reads, for $0 < \epsilon \ll 1$,

$$1 = \int_{z=z_i-\epsilon}^{z=z_i+\epsilon} \mathrm{d}f \delta\left(f\right) = \int_{z=z_i-\epsilon}^{z=z_i+\epsilon} \mathrm{d}z \left|\left(f'(z_i) + \mathcal{O}((z - z_i)^1)\right)\right| \frac{\delta\left(z - z_i\right)}{\varphi_i} \tag{5.1.22}$$

$$\overset{\epsilon \to 0}{\to} \frac{|f'(z_i)|}{\varphi_i}. \tag{5.1.23}$$

(When you change variables within an integral, remember to include the absolute value of the Jacobian, which is essentially $|f'(z_i)|$ in this case.) The $\mathcal{O}(z^p)$ means "the next term in the series has a dependence on the variable z that goes as z^p"; this first correction can be multiplied by other stuff, but has to be proportional to z^p.

A simple application of eq. (5.1.18) is, for $a \in \mathbb{R}$,

$$\delta(az) = \frac{\delta(z)}{|a|}. \tag{5.1.24}$$

- Since $\delta(z)$ is non-zero only when $z = 0$, it must be that $\delta(-z) = \delta(z)$ and more generally

$$\delta(z - z') = \delta(z' - z). \tag{5.1.25}$$

- We may also take the derivative of a δ-function. Under an integral sign, we may apply integration-by-parts as follows:

$$\int_a^b \delta'(x - x')f(x)\mathrm{d}x = [\delta(x - x')f(x)]_{x=a}^{x=b} - \int_a^b \delta(x - x')f'(x)\mathrm{d}x = -f'(x') \tag{5.1.26}$$

as long as x' lies strictly between a and b, $a < x' < b$, where a and b are both real.

- *Dimension* What is the dimension of the δ-function? Turns out $\delta(\xi)$ has dimensions of $1/[\xi]$, i.e., the reciprocal of the dimension of its argument. The reason is

$$\int \mathrm{d}\xi \delta(\xi) = 1 \qquad \Rightarrow \qquad [\xi]\left[\delta(\xi)\right] = 1. \tag{5.1.27}$$

- More on distributional calculus can be found in §(7.8.1).

Problem 5.4. We may generalize the identities in equations (5.1.18) and (5.1.24) in the following manner. Show that, whenever some function $g(z)$ is strictly positive within the range of $z-$integration, it may be 'pulled out' of the delta function as though it were a constant:

$$\delta\left(g(z)f(z)\right) = \frac{\delta(f(z))}{g(z)} = \sum_{z_i \equiv i\text{th zero of } f(z)} \frac{\delta(z-z_i)}{g(z_i)|f'(z_i)|}. \tag{5.1.28}$$

Hint: Simply apply eq. (5.1.18). □

Problem 5.5. Justify the following results:

$$\int_{\mathbb{R}} \delta(a-z)\delta(z-b)\mathrm{d}z = \delta(a-b), \tag{5.1.29}$$

$$\int_a^b \delta(z-c)\mathrm{d}z = \Theta(b-c)\Theta(c-a); \tag{5.1.30}$$

where, in the second line, $a < b$ and $\Theta(b-c)\Theta(c-a)$ is unity when $a < c < b$ and zero when $c < a$ or $c > b$. □

Continuous spectrum Let Ω be a Hermitian operator whose spectrum is continuous; i.e., $\Omega\left|\omega\right\rangle = \omega\left|\omega\right\rangle$ with ω being a continuous parameter. If $\left|\omega\right\rangle$ and $\left|\omega'\right\rangle$ are both "unit norm" eigenvectors of different eigenvalues ω and ω', we have for example

$$\left\langle\omega\middle|\,\omega'\right\rangle = \delta(\omega-\omega'). \tag{5.1.31}$$

(This assumes a "translation symmetry" in this ω-space; we will see later how to modify this inner product when the translation symmetry is lost.) The completeness relation in eq. (4.3.23) is given by

$$\int \mathrm{d}\omega \left|\omega\right\rangle\left\langle\omega\right| = \mathbb{I}; \tag{5.1.32}$$

because for an arbitrary ket $\left|f\right\rangle$,

$$\left\langle\omega'\middle|\,f\right\rangle = \left\langle\omega'\right|\mathbb{I}\left|f\right\rangle = \int \mathrm{d}\omega \left\langle\omega'\middle|\,\omega\right\rangle\left\langle\omega\middle|\,f\right\rangle \tag{5.1.33}$$

$$= \int \mathrm{d}\omega\delta(\omega'-\omega)\left\langle\omega\middle|\,f\right\rangle. \tag{5.1.34}$$

An arbitrary vector $\left|\alpha\right\rangle$ can thus be expressed as

$$\left|\alpha\right\rangle = \int \mathrm{d}\omega \left|\omega\right\rangle\left\langle\omega\middle|\,\alpha\right\rangle. \tag{5.1.35}$$

When the state is normalized to unity, we say

$$\left\langle\alpha\middle|\,\alpha\right\rangle = \int \mathrm{d}\omega \left\langle\alpha\middle|\,\omega\right\rangle\left\langle\omega\middle|\,\alpha\right\rangle = \int \mathrm{d}\omega|\left\langle\omega\middle|\,\alpha\right\rangle|^2 = 1. \tag{5.1.36}$$

The inner product between arbitrary vectors $|\alpha\rangle$ and $|\beta\rangle$ now reads

$$\langle\alpha|\,\beta\rangle = \int \mathrm{d}\omega\, \langle\alpha|\,\omega\rangle\,\langle\omega|\,\beta\rangle\,. \tag{5.1.37}$$

Since by assumption Ω is diagonal, i.e.,

$$\Omega = \int \mathrm{d}\omega\omega\, |\omega\rangle\,\langle\omega|\,, \tag{5.1.38}$$

the matrix elements of Ω are

$$\langle\omega\,|\Omega|\,\omega'\rangle = \omega\delta(\omega-\omega') = \omega'\delta(\omega-\omega'). \tag{5.1.39}$$

Because of the δ-function, which enforces $\omega = \omega'$, it does not matter if we write ω or ω' on the right hand side.

Continuous operators connected to the identity In the following, we will deal with continuous operators. By a continuous operator A, we mean one that depends on some continuous parameter(s) $\vec{\xi}$. For example, spatial translations would involve a displacement vector; for rotations, the associated angles; for Lorentz boosts, the unit direction vector and rapidity; etc. Furthermore, if these continuous parameters may be tuned such that $A(\vec{\xi})$ becomes the identity, then we say that this operator is continuously connected to the identity. When such a continuous operator is 'close enough' to the identity operator $\mathbb{I}$, we would expect it may be phrased as an exponential of another operator $-iZ(\vec{\xi})$; namely,

$$A(\vec{\xi}) = e^{-iZ(\vec{\xi})}. \tag{5.1.40}$$

[22]The exponential of an operator Y is itself defined through the Taylor series

$$e^Y \equiv \mathbb{I} + Y + \frac{Y^2}{2!} + \frac{Y^3}{3!} + \dots = \sum_{\ell=0}^{\infty}\frac{Y^\ell}{\ell!}. \tag{5.1.41}$$

For later use, note that

$$\left(e^Y\right)^\dagger = \sum_{\ell=0}^{+\infty}\frac{(Y^\ell)^\dagger}{\ell!} = \sum_{\ell=0}^{+\infty}\frac{(Y^\dagger)^\ell}{\ell!} = e^{Y^\dagger}. \tag{5.1.42}$$

It may also be usually argued that Z (dubbed the 'generator'), is in fact linear in the continuous parameters; so that it is a superposition of some basis generators $\{T^a\}$ that induce infinitesimal versions of the transformations under consideration.

$$Z = \vec{\xi}\cdot\vec{T} \equiv \xi_a T^a. \tag{5.1.43}$$

[22]If A and $Y \equiv -iZ$ were complex numbers, then the $A = e^Y$ in eq. (5.1.40) is always true in that, for a given $A = \exp\ln A \equiv \exp Y$; where $Y \equiv \ln A$. For operators A and Y, if we assume Y is 'close enough' to zero (and, therefore, A is 'close enough' to the identity) we may define $Y \equiv \ln(\mathbb{I} + (A-\mathbb{I})) \equiv -\sum_{\ell=1}^{\infty}(\mathbb{I}-A)^\ell/\ell$. Whenever the series make sense, then $A = e^Y$. Furthermore, that the Taylor series for the natural logarithm involves powers of the deviation of the operator from the identity, namely $A - \mathbb{I}$, is why there is a need to demand that A is continuously connected to $\mathbb{I}$ – i.e., $\ln A$ would cease to be valid if the operator norm $||A-\mathbb{I}||$ is too large.

That these $\{T^a\}$ form a vector space in turn follows from the multiplication rules that these operators need to obey. Specifically, operators belonging to the same *group* must take the same form $\{A = \exp(-ia_i T^i)\}$, for appropriate (basis) generators $\{T^i\}$; then since two consecutive operations (parametrized, say, by $\vec{a}$ and $\vec{b}$) must yield another operator of the same group, that means there must be some other $\vec{c}$ such that

$$\exp\left(-i\vec{a}\cdot\vec{T}\right)\exp\left(-i\vec{b}\cdot\vec{T}\right) = \exp\left(-i\vec{c}\cdot\vec{T}\right). \tag{5.1.44}$$

Lie Groups and Lie Algebras The framework we are describing here is a *Lie group*, a group with continuous parameters. (See §(B) for the axioms defining a group.) Because the operators here are already linear operators acting on some Hilbert space, the closure assumption in eq. (5.1.44) – that products of group elements yield another group element – is all we need to ensure these they indeed form a group.[23]

The crucial property ensuring eq. (5.1.44) holds, is that the basis generators $\{T^a\}$ themselves obey a *Lie algebra*:

$$\left[T^a, T^b\right] = if^{abc}T^c \equiv i\sum_c f^{abc}T^c. \tag{5.1.45}$$

These $\{f^{abc}\}$ are called *structure constants*. As we shall witness shortly, this implies the $\vec{c}$ may be solved in terms of $\vec{a}$, $\vec{b}$, and the structure constants.

Problem 5.6. Prove that the set of linear operators $\{T^a\}$ in eq. (5.1.45) that are closed under commutation forms a vector space. Hint: Remember, we have already proven that linear operators themselves form a vector space. What's the only property you need to verify? □

Baker-Campbell-Hausdorff The Baker-Campbell-Hausdorff formula tells us, for generic operators X and Y, the product $e^X e^Y$ would produce an exponential e^Z where the exponent Z only involves $X+Y$ and their commutators $[X,Y]$ and nested commutators; for e.g., $[X,[X,Y]]$, $[Y,[Y,[X,Y]]]$, $[X,[X,[Y[Y,X]]]]$, etc. Because these are operators, note that $e^X e^Y \neq e^{X+Y} \neq e^Y e^X$. In detail, the first few terms in the exponent read

$$e^X e^Y = \exp\left(X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \dots\right). \tag{5.1.46}$$

(Parenthetically, this informs us that, when multiplying exponentials of operators, the exponents add if and only if they commute.) Returning to the discussion around equations (5.1.43) and (5.1.44), if

$$X = -ia_i T^i \qquad \text{and} \qquad Y = -ib_i T^i; \tag{5.1.47}$$

then eq. (5.1.45) inserted into the right hand side of (5.1.46) reads

$$-iZ = -i\left(a_i + b_i\right)T^i + \frac{1}{2}(-i)^2 a_i b_j \left[T^i, T^j\right]$$

[23]Lie groups are analogous to curved space(time)s, where each space(time) point corresponds to a group element; and the superposition of the generators $\vec{\xi}\cdot\vec{T}$ are 'tangent vectors' based at the identity operator.

$$+\frac{1}{12}(-i)^3 a_i a_j b_k \left[T^i, \left[T^j, T^k\right]\right] - \frac{1}{12}(-i)^3 b_i a_j b_k \left[T^i, \left[T^j, T^k\right]\right] + \dots$$
$$= -i\left\{a_l + b_l + \frac{1}{2}a_i b_j f^{ijl} + \frac{1}{12}a_i a_j b_k f^{jks} f^{isl} - \frac{1}{12}b_i a_j b_k f^{jks} f^{isl} + \dots\right\} T^l. \tag{5.1.48}$$

From this, we may now read off – the exponent on the right hand side of eq. (5.1.44) is

$$Z = \vec{c} \cdot \vec{T}, \tag{5.1.49}$$
$$c_l = a_l + b_l + \frac{1}{2}a_i b_j f^{ijl} + \frac{1}{12}a_i a_j b_k f^{jks} f^{isl} - \frac{1}{12}b_i a_j b_k f^{jks} f^{isl} + \dots. \tag{5.1.50}$$

To sum: because the generators of the Lie group are closed under commutation, the Baker-Campbell-Hausdorff formula tells us, upon multiplying two operators (both continuously connected to the identity), the exponent of the result is necessarily again a linear combination of the same generators.

Continuous unitary operators Continuous unitary operators form a special subclass of the Lie groups we have just discussed. When the underlying flat space is space-translation and rotation symmetric, for example, there is no distinguished origin nor special direction. As we shall discuss below, this will also lead to the lack of distinguished basis kets spanning the corresponding Hilbert space. When such a situation arises, the action of these operators amount to a change-of-basis, and hence are unitary. In other words, these operators become unitary due to the underlying symmetries of the flat space.

For now, let us note that

> An operator continuously connected to the identity, namely $U = \exp(-iZ)$, is unitary if and only if its generator Z is Hermitian.

If Z is Hermitian, then we may take the dagger of the Taylor series of $\exp(-iZ)$ term-by-term, and recognize

$$U^\dagger = \left(e^{-iZ}\right)^\dagger = e^{+iZ^\dagger} = e^{iZ}. \tag{5.1.51}$$

Therefore, since iZ certainly commutes with $-iZ$,

$$U^\dagger U = e^{iZ} e^{-iZ} = e^{i(Z-Z)} = \mathbb{I}. \tag{5.1.52}$$

On the other hand, if U is unitary, we may introduce a fictitious real parameter ϵ and expand

$$(e^{-i\epsilon Z})^\dagger e^{-i\epsilon Z} = \mathbb{I}, \tag{5.1.53}$$
$$\left(\mathbb{I} + i\epsilon Z^\dagger + \mathcal{O}(\epsilon^2)\right)\left(\mathbb{I} - i\epsilon Z + \mathcal{O}(\epsilon^2)\right) = \mathbb{I}, \tag{5.1.54}$$
$$\mathbb{I} + i\epsilon(Z^\dagger - Z) + \mathcal{O}(\epsilon^2) = \mathbb{I}. \tag{5.1.55}$$

The presence of the parameter ϵ allows us to see that each order in Z is independent, as we may view the product as a Taylor series in ϵ. At first order, in particular, we have – as advertised –

$$Z^\dagger = Z. \tag{5.1.56}$$

At this juncture, we gather the following:

Symmetry and Observables In quantum mechanics unitary operators $\{U = e^{-iZ}\}$ play an important role, not only because they implement symmetry transformations – the inner product $\langle\alpha|\,\beta\rangle = \langle\alpha'|\,\beta'\rangle$ is preserved whenever both $|\alpha'\rangle \equiv U\,|\alpha\rangle$ and $|\beta'\rangle \equiv U\,|\beta\rangle$ – their *generators* $\{Z\}$ often correspond to physical observables since they are Hermitian.

Unitary Operators and Conservation of Probability An elementary example of a continuous unitary operator is provided by the following example, which occurs in quantum mechanics. Let H be a time-independent Hermitian operator, and suppose $U(t)$ is an operator that satisfies

$$i\partial_t U = HU; \tag{5.1.57}$$

and the boundary condition

$$U(t=0) = \mathbb{I}. \tag{5.1.58}$$

We see the solution is provided by

$$U(t) = \exp(-itH). \tag{5.1.59}$$

We may readily verify $U^\dagger = \exp(+itH)$. Since itH and $-itH$ commute, we have

$$U^\dagger U = e^{itH}e^{-itH} = e^{itH-itH} = e^0 = \mathbb{I}. \tag{5.1.60}$$

Let $|\psi(t_0)\rangle$ be the initial state at time t_0. The state at any later time $t > t_0$ is given by

$$|\psi(t>t_0)\rangle = U(t-t_0)\,|\psi(t_0)\rangle = \exp\left(-i(t-t_0)H\right)|\psi(t_0)\rangle \tag{5.1.61}$$

This $|\psi(t>t_0)\rangle$ automatically satisfies $i\partial_t\,|\psi(t)\rangle = H\,|\psi(t)\rangle$ because $i\partial_t\,|\psi(t)\rangle = i(\partial_t U)\,|\psi(t_0)\rangle$. We may also check that the initial conditions are recovered: $|\psi(t\to t_0)\rangle = U(0)\,|\psi(t_0)\rangle = |\psi(t_0)\rangle$. Moreover, we may write the evolution operator as a sum over the energy eigenstates $\{|E\rangle\,|H\,|E\rangle = E\,|E\rangle\}$ in the following manner.

$$U(t) = e^{-itH} = \sum_E e^{-itE}\,|E\rangle\,\langle E|\,. \tag{5.1.62}$$

Problem 5.7. The physical importance of having time evolution of states governed by a unitary operator, is that it guarantees conservation of probability: if the particle can be found *somewhere* at time t_0, it must be found *somewhere* at any later time $t > t_0$. This is summed up in the statement that $\langle\psi(t)|\,\psi(t)\rangle = 1$ for all t.

Show that

$$\partial_t\left(\langle\psi(t)|\,\psi(t)\rangle\right) = 0 \tag{5.1.63}$$

and therefore if it is $\langle\psi(t_0)|\,\psi(t_0)\rangle = 1$, the constancy of the amplitude implies $\langle\psi(t)|\,\psi(t)\rangle = 1$ for all t. □

Symmetry and Degeneracy Since unitary operators may be associated with symmetry transformations, we may now understand the connection between symmetry and degeneracy. In particular, if A is some Hermitian operator, and it forms mutually compatible observables with the Hermitian generators $\{T^a\}$ of some unitary symmetry operator $U(\vec{\xi}) = \exp(-i\vec{\xi} \cdot \vec{T})$, then A must commute with U as well.

$$\left[A, U(\vec{\xi})\right] = 0. \tag{5.1.64}$$

But that implies, if $|\alpha\rangle$ is an eigenket of A with eigenvalue α, namely

$$A|\alpha\rangle = \alpha|\alpha\rangle, \tag{5.1.65}$$

so must $U|\alpha\rangle$ be. For, $[A, U] = 0$ leads us to consider

$$[A, U]|\alpha\rangle = 0, \tag{5.1.66}$$

$$A(U|\alpha\rangle) = UA|\alpha\rangle = \alpha(U|\alpha\rangle). \tag{5.1.67}$$

If $U|\alpha\rangle$ is not the same ket as $|\alpha\rangle$ (up to an overall phase), then this corresponds to a degeneracy: the physically distinct states $U(\vec{\xi})|\alpha\rangle$ and $|\alpha\rangle$ both correspond to eigenkets of A with the same eigenvalue α. To sum:

Symmetry implies degeneracy.

5.2 Spatial translations and the Fourier transform

In this section, we shall discuss in detail the Hilbert space spanned by the eigenkets of the position operator $\vec{X}$, where we assume there is some underlying infinite (flat/Euclidean) D-space $\mathbb{R}^D$. The arrow indicates the position operator itself has D components, each one corresponding to a distinct axis of the D-dimensional Euclidean space. $|\vec{x}\rangle$ would describe the state that is (infinitely) sharply localized at the position $\vec{x}$; namely, it obeys the D-component equation

$$\vec{X}|\vec{x}\rangle = \vec{x}|\vec{x}\rangle. \tag{5.2.1}$$

Or, in index notation,

$$X^k|\vec{x}\rangle = x^k|\vec{x}\rangle, \qquad k \in \{1, 2, \dots, D\}. \tag{5.2.2}$$

The position eigenkets are normalized as, in Cartesian coordinates,

$$\langle\vec{x}|\,\vec{x}'\rangle = \delta^{(D)}(\vec{x} - \vec{x}') \equiv \prod_{i=1}^{D} \delta(x^i - x'^i) = \delta(x^1 - x'^1)\delta(x^2 - x'^2)\dots\delta(x^D - x'^D). \tag{5.2.3}$$

As an important aside, the generalization of the 1D transformation law in eq. (5.1.18) involving the δ-function has the following higher dimensional generalization. If we are given a transformation $\vec{x} \equiv \vec{x}(\vec{y})$ and $\vec{x}' \equiv \vec{x}'(\vec{y}')$, then

$$\delta^{(D)}(\vec{x} - \vec{x}') = \frac{\delta^{(D)}(\vec{y} - \vec{y}')}{|\det \partial x^a(\vec{y})/\partial y^b|} = \frac{\delta^{(D)}(\vec{y} - \vec{y}')}{|\det \partial x'^a(\vec{y}')/\partial y'^b|}, \tag{5.2.4}$$

where $\delta^{(D)}(\vec{x}-\vec{x}') \equiv \prod_{i=1}^{D} \delta(x^i - x'^i)$, $\delta^{(D)}(\vec{y}-\vec{y}') \equiv \prod_{i=1}^{D} \delta(y^i - y'^i)$, and the Jacobian inside the absolute value occurring in the denominator on the right hand side is the usual determinant of the matrix whose ath row and bth column is given by $\partial x^a(\vec{y})/\partial y^b$. (The second and third equalities follow from each other because the δ-functions allow us to assume $\vec{y} = \vec{y}'$.) Equation (5.2.4) can be justified by demanding that its integral around the point $\vec{x} = \vec{x}'$ gives one. For $0 < \epsilon \ll 1$, and denoting $\delta^{(D)}(\vec{x}-\vec{x}') = \delta^{(D)}(\vec{y}-\vec{y}')/\varphi(\vec{y}')$,

$$1 = \int_{|\vec{x}-\vec{x}'|\leq\epsilon} \mathrm{d}^D\vec{x}\,\delta^{(D)}(\vec{x}-\vec{x}') = \int_{|\vec{x}-\vec{x}'|\leq\epsilon} \mathrm{d}^D\vec{y} \left|\det \frac{\partial x^a(\vec{y})}{\partial y^b}\right| \frac{\delta^{(D)}(\vec{y}-\vec{y}')}{\varphi(\vec{y}')} = \frac{\left|\det \frac{\partial x'^a(\vec{y}')}{\partial y'^b}\right|}{\varphi(\vec{y}')}. \tag{5.2.5}$$

Now, any vector $|\alpha\rangle$ in the Hilbert space can be expanded in terms of the position eigenkets.

$$|\alpha\rangle = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\, |\vec{x}\rangle \langle\vec{x}|\,\alpha\rangle\,. \tag{5.2.6}$$

Notice $\langle\vec{x}|\,\alpha\rangle$ is an ordinary (possibly complex) function of the spatial coordinates $\vec{x}$. We see that the space of functions emerges from the vector space spanned by the position eigenkets. Just as we can view $\langle i|\,\alpha\rangle$ in $|\alpha\rangle = \sum_i |i\rangle\langle i|\,\alpha\rangle$ as a column vector, the function $f(\vec{x}) \equiv \langle\vec{x}|\,f\rangle$ is in some sense a continuous (infinite dimensional) "vector" in this position representation.

In the context of *quantum mechanics* $\langle\vec{x}|\,\alpha\rangle$ would be identified as a wave function, more commonly denoted as $\psi(\vec{x})$; in particular, $|\langle\vec{x}|\,\alpha\rangle|^2$ is interpreted as the probability density that the system is localized around $\vec{x}$ when its position is measured. This is in turn related to the demand that the wave function obey $\int \mathrm{d}^D\vec{x} |\langle\vec{x}|\,\alpha\rangle|^2 = 1$. However, it is worth highlighting here that our discussion regarding the Hilbert spaces spanned by the position eigenkets $\{|\vec{x}\rangle\}$ (and later below, by their momentum counterparts $\{|\vec{k}\rangle\}$) does not necessarily have to involve quantum theory.[24] We will provide concrete examples below, such as how the concept of Fourier transform emerges and how classical field theory problems – the derivation of the Green's function of the Laplacian in eq. (12.3.56), for instance – can be tackled using the methods/formalism delineated here.

Matrix elements Suppose we wish to calculate the matrix element $\langle\alpha\,|Y|\,\beta\rangle$ in the position representation. It is

$$\begin{aligned}\langle\alpha\,|Y|\,\beta\rangle &= \int \mathrm{d}^D\vec{x} \int \mathrm{d}^D\vec{x}'\, \langle\alpha|\,\vec{x}\rangle \langle\vec{x}\,|Y|\,\vec{x}'\rangle \langle\vec{x}'|\,\beta\rangle \\ &= \int \mathrm{d}^D\vec{x} \int \mathrm{d}^D\vec{x}'\, \langle\vec{x}|\,\alpha\rangle^* \langle\vec{x}\,|Y|\,\vec{x}'\rangle \langle\vec{x}'|\,\beta\rangle\,. \end{aligned} \tag{5.2.7}$$

If the operator $Y(\vec{X})$ were built solely from the position operator $\vec{X}$, then

$$\left\langle\vec{x}\left|Y(\vec{X})\right|\vec{x}'\right\rangle = Y(\vec{x})\delta^{(D)}(\vec{x}-\vec{x}') = Y(\vec{x}')\delta^{(D)}(\vec{x}-\vec{x}'); \tag{5.2.8}$$

and the double integral collapses into one,

$$\left\langle\alpha\left|Y(\vec{X})\right|\beta\right\rangle = \int \mathrm{d}^D\vec{x}\, \langle\vec{x}|\,\alpha\rangle^* \langle\vec{x}'|\,\beta\rangle\, Y(\vec{x}). \tag{5.2.9}$$

[24]This is especially pertinent for those whose first contact with continuous Hilbert spaces was in the context of a quantum mechanics course.

Problem 5.8. Show that if U is a unitary operator and $|\alpha\rangle$ is an arbitrary vector, then $|\alpha\rangle$, $U|\alpha\rangle$ and $U^\dagger|\alpha\rangle$ have the same norm. □

Translations in $\mathbb{R}^D$ To make these ideas regarding continuous operators more concrete, we will now study the case of translation in some detail, realized on a Hilbert space spanned by the position eigenkets $\{|\vec{x}\rangle\}$. To be specific, let $\mathcal{T}(\vec{d})$ denote the translation operator parameterized by the displacement vector $\vec{d}$. We shall work in D space dimensions. We define the translation operator by its action

$$\mathcal{T}(\vec{d})\,|\vec{x}\rangle = \left|\vec{x}+\vec{d}\right\rangle. \tag{5.2.10}$$

Since $|\vec{x}\rangle$ and $|\vec{x}+\vec{d}\rangle$ can be viewed as distinct elements of the set of basis vectors, we shall see that the translation operator can be viewed as a unitary operator, changing basis from $\{|\vec{x}\rangle\,|\vec{x}\in\mathbb{R}^D\}$ to $\{|\vec{x}+\vec{d}\rangle|\vec{x}\in\mathbb{R}^D\}$. Let us in fact first show that the translation operator is unitary. Taking the dagger of eq. (5.2.10),

$$\langle\vec{y}|\,\mathcal{T}(\vec{d})^\dagger = \left\langle\vec{y}+\vec{d}\right|. \tag{5.2.11}$$

Therefore, recalling eq. (5.2.3),

$$\langle\vec{y}|\,\mathcal{T}(\vec{d})^\dagger\mathcal{T}(\vec{d})\,|\vec{x}\rangle = \left\langle\vec{y}+\vec{d}\middle|\,\vec{x}+\vec{d}\right\rangle = \delta^{(D)}(\vec{y}-\vec{x}) = \langle\vec{y}\,|\mathbb{I}|\,\vec{x}\rangle\,; \tag{5.2.12}$$

and since this is true for arbitrary states $|\vec{x}\rangle$ and $|\vec{y}\rangle$,

$$\mathcal{T}(\vec{d})^\dagger\mathcal{T}(\vec{d}) = \mathbb{I}. \tag{5.2.13}$$

The inverse transformation of the translation operator is

$$\mathcal{T}(\vec{d})^\dagger\,|\vec{x}\rangle = \left|\vec{x}-\vec{d}\right\rangle \tag{5.2.14}$$

since

$$\mathcal{T}(\vec{d})^\dagger\mathcal{T}(\vec{d})\,|\vec{x}\rangle = \mathcal{T}(\vec{d})^\dagger\left|\vec{x}+\vec{d}\right\rangle = \left|\vec{x}+\vec{d}-\vec{d}\right\rangle = |\vec{x}\rangle\,. \tag{5.2.15}$$

Of course we have the identity operator $\mathbb{I}$ when $\vec{d}=\vec{0}$,

$$\mathcal{T}(\vec{0})\,|\vec{x}\rangle = |\vec{x}\rangle \qquad \Rightarrow \qquad \mathcal{T}(\vec{0}) = \mathbb{I}. \tag{5.2.16}$$

The following composition law has to hold

$$\mathcal{T}(\vec{d}_1)\mathcal{T}(\vec{d}_2) = \mathcal{T}(\vec{d}_1+\vec{d}_2), \tag{5.2.17}$$

because translation is commutative

$$\mathcal{T}(\vec{d}_1)\mathcal{T}(\vec{d}_2)\,|\vec{x}\rangle = \mathcal{T}(\vec{d}_1)\left|\vec{x}+\vec{d}_2\right\rangle = \left|\vec{x}+\vec{d}_2+\vec{d}_1\right\rangle = \left|\vec{x}+\vec{d}_1+\vec{d}_2\right\rangle = \mathcal{T}(\vec{d}_1+\vec{d}_2)\,|\vec{x}\rangle\,. \tag{5.2.18}$$

Problem 5.9. Translation operator is unitary. Show that

$$\mathcal{T}(\vec{d}) = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}' \left|\vec{d}+\vec{x}'\right\rangle \langle \vec{x}'| \tag{5.2.19}$$

satisfies eq. (5.2.10) and therefore is the correct ket-bra operator representation of the translation operator. Check explicitly that $\mathcal{T}(\vec{d})$ is unitary. Remember an operator U is unitary iff it implements a change from one orthonormal basis to another – compare eq. (5.2.19) with eq. (4.3.145). $\square$

Momentum operator We now turn to demonstrate that eq. (5.1.40) can be expressed as

$$\mathcal{T}(\vec{d}) = \exp\left(-i\vec{d}\cdot\vec{P}\right) = \exp\left(-id^k P_k\right). \tag{5.2.20}$$

This because translation in the position representation may simply be implemented as a Taylor series,

$$\left\langle \vec{x}+\vec{d}\right| f\Big\rangle = f(\vec{x}+\vec{d}) = \sum_{\ell=0}^{+\infty} \frac{\left(\vec{d}\cdot\vec{\nabla}_{\vec{x}}\right)^\ell}{\ell!} f(\vec{x}) \tag{5.2.21}$$

$$= \sum_{\ell=0}^{+\infty} \frac{\left(i\vec{d}\cdot(-i)\vec{\nabla}_{\vec{x}}\right)^\ell}{\ell!} \langle \vec{x}|\, f\rangle \tag{5.2.22}$$

$$\equiv \sum_{\ell=0}^{+\infty} \frac{i^\ell}{\ell!} \left\langle \vec{x}\left|\left(\vec{d}\cdot\vec{P}\right)^\ell\right| f\right\rangle; \tag{5.2.23}$$

where we have identified

$$\langle \vec{x}\,|P_a|\, f\rangle = -i\partial_{x^a} \langle \vec{x}|\, f\rangle. \tag{5.2.24}$$

This then allows us to re-collapse the series back into the exponential of the momentum operator:

$$\left\langle \vec{x}+\vec{d}\right| f\Big\rangle = \left\langle \vec{x}\left|e^{i\vec{d}\cdot\vec{P}}\right| f\right\rangle = \left(e^{-i\vec{d}\cdot\vec{P}^\dagger}|\vec{x}\rangle\right)^\dagger |f\rangle. \tag{5.2.25}$$

Since $\mathcal{T}(\vec{d})$ is unitary in infinite space, the $\vec{P}$ is Hermitian. We will call this Hermitian operator $\vec{P}$ the momentum operator.[25] To summarize:

$$\left|\vec{x}+\vec{d}\right\rangle = \mathcal{T}(\vec{d})\,|\vec{x}\rangle = e^{-i\vec{d}\cdot\vec{P}}\,|\vec{x}\rangle. \tag{5.2.26}$$

In this exp form, eq. (5.2.17) reads

$$\exp\left(-i\vec{d}_1\cdot\vec{P}\right)\exp\left(-i\vec{d}_2\cdot\vec{P}\right) = \exp\left(-i(\vec{d}_1+\vec{d}_2)\cdot\vec{P}\right). \tag{5.2.27}$$

[25] Strictly speaking P_j here has dimensions of 1/[length], whereas the momentum you might be familiar with has units of [mass $\times$ length/time^2] = [angular momentum]/[length]. The reason for such nomenclature is because of its application in Quantum Mechanics.

Problem 5.10. Commutation relations between momentum operators Because translation is commutative, $\vec{d}_1 + \vec{d}_2 = \vec{d}_2 + \vec{d}_1$, argue that the translation operators commute:

$$\left[\mathcal{T}(\vec{d}_1), \mathcal{T}(\vec{d}_2)\right] = 0. \tag{5.2.28}$$

By considering infinitesimal displacements $\vec{d}_1 = \mathrm{d}\vec{\xi}_1$ and $\vec{d}_2 = \mathrm{d}\vec{\xi}_2$, show that eq. (5.2.20) leads to us to conclude that momentum operators commute among themselves,

$$[P_i, P_j] = 0, \qquad i, j \in \{1, 2, 3, \dots, D\}. \tag{5.2.29}$$

Comparing against eq. (5.1.45), we may conclude the structure constants occurring within the Lie algebra obeyed by the translation generators are all zero. □

Note that, for any operator A and arbitrary state $|\psi\rangle$,

$$A\,|\psi\rangle = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\,|\vec{x}\rangle\,\langle\vec{x}|\,A\,|\psi\rangle\,. \tag{5.2.30}$$

Therefore to find $\mathcal{T}(\vec{d})$ (and hence $\vec{P}$) of eq. (5.2.20) in any other representation, we merely need to find the change-of-basis from the position eigenkets to the new basis.

Translation invariance Infinite (flat) D-space $\mathbb{R}^D$ is the same everywhere and in every direction. This intuitive fact is intimately tied to the property that $\mathcal{T}(\vec{d})$ is a unitary operator: it just changes one orthonormal basis to another, and physically speaking, there is no privileged set of basis vectors. In particular, the norm of vectors is position independent:

$$\left\langle \vec{x} + \vec{d} \middle| \vec{x}' + \vec{d} \right\rangle = \delta^{(D)}\left(\vec{x} - \vec{x}'\right) = \langle\vec{x}|\,\vec{x}'\rangle\,. \tag{5.2.31}$$

This observation played a crucial role in the proof of the unitary character of $\mathcal{T}$ in eq. (5.2.12). In turn, the unitary $\mathcal{T}(\vec{d}) = \exp(-i\vec{d}\cdot\vec{P})$ implies its generators $\{P_j\}$ must be Hermitian. To reiterate:

> **Symmetry, Unitarity, and Hermicity** The unitary nature of the translation operator $\mathcal{T}(\vec{d}) = \exp(-i\vec{d}\cdot\vec{P})$ and the Hermitian character of the momentum $\vec{P}$ are both direct consequences of the space-translation symmetry of infinite flat space. The transformation between the position eigenbasis and the eigenbasis of $\vec{P}$, as we shall see below, leads to the Fourier transform pairs. □

As we will see below, if we confine our attention to some finite domain in $\mathbb{R}^D$ or if space is no longer flat, then global translation symmetry is lost and the translation operator still exists but is no longer unitary.

Commutation relations between X^i and P_j We have seen, just from postulating a Hermitian position operator X^i, and considering the translation operator acting on the space spanned by its eigenkets $\{|\vec{x}\rangle\}$, that there exists a Hermitian momentum operator P_j that occurs in the exponent of said translation operator. This implies the continuous space at hand can be spanned by either the position eigenkets $\{|\vec{x}\rangle\}$ or the momentum eigenkets, which obey

$$P_j|\vec{k}\rangle = k_j|\vec{k}\rangle. \tag{5.2.32}$$

Are the position and momentum operators simultaneously diagonalizable? Can we label a state with both position and momentum? The answer is no.

To see this, we now consider an infinitesimal displacement operator $\mathcal{T}(\mathrm{d}\vec{\xi})$.

$$\vec{X}\mathcal{T}(\mathrm{d}\vec{\xi})\left|\vec{x}\right\rangle = \vec{X}\left|\vec{x}+\mathrm{d}\vec{\xi}\right\rangle = (\vec{x}+\mathrm{d}\vec{\xi})\left|\vec{x}+\mathrm{d}\vec{\xi}\right\rangle, \tag{5.2.33}$$

and

$$\mathcal{T}(\mathrm{d}\vec{\xi})\vec{X}\left|\vec{x}\right\rangle = \vec{x}\left|\vec{x}+\mathrm{d}\vec{\xi}\right\rangle. \tag{5.2.34}$$

Since $\left|\vec{x}\right\rangle$ was an arbitrary vector, we may subtract the two equations

$$\left[\vec{X},\mathcal{T}(\mathrm{d}\vec{\xi})\right]\left|\vec{x}\right\rangle = \mathrm{d}\vec{\xi}\left|\vec{x}+\mathrm{d}\vec{\xi}\right\rangle = \mathrm{d}\vec{\xi}\left|\vec{x}\right\rangle + \mathcal{O}\left(\mathrm{d}\vec{\xi}^2\right). \tag{5.2.35}$$

At first order in $\mathrm{d}\vec{\xi}$, we have the operator identity

$$\left[\vec{X},\mathcal{T}(\mathrm{d}\vec{\xi})\right] = \mathrm{d}\vec{\xi}. \tag{5.2.36}$$

The left hand side involves operators, but the right hand side only real numbers. At this point we invoke eq. (5.2.20), and deduce, for infinitesimal displacements,

$$\mathcal{T}(\mathrm{d}\vec{\xi}) = 1 - i\mathrm{d}\vec{\xi}\cdot\vec{P} + \mathcal{O}(\mathrm{d}\vec{\xi}^2) \tag{5.2.37}$$

which in turn means eq. (5.2.36) now reads, as $\mathrm{d}\vec{\xi}\to\vec{0}$,

$$\left[\vec{X},-i\mathrm{d}\vec{\xi}\cdot\vec{P}\right] = \mathrm{d}\vec{\xi} \tag{5.2.38}$$

$$\left[X^l,P_j\right]\mathrm{d}\xi^j = i\delta^l_j\mathrm{d}\xi^j \qquad \text{(the lth component)} \tag{5.2.39}$$

Since the $\{\mathrm{d}\xi^j\}$ are independent, the coefficient of $\mathrm{d}\xi^j$ on both sides must be equal. This leads us to the fundamental commutation relation between kth component of the position operator with the j component of the momentum operator:

$$\left[X^k,P_j\right] = i\delta^k_j, \qquad j,k\in\{1,2,\dots,D\}. \tag{5.2.40}$$

To sum: although X^k and P_j are both Hermitian operators in infinite flat $\mathbb{R}^D$, we see they are incompatible and thus, to span the continuous vector space at hand we can use either the eigenkets of X^i or that of P_j but not both. We will, in fact, witness below how changing from the position to momentum eigenket basis gives rise to the Fourier transform and its inverse.

$$\left|f\right\rangle = \int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}'\left|\vec{x}'\right\rangle\left\langle\vec{x}'\middle|f\right\rangle, \qquad X^i\left|\vec{x}'\right\rangle = x'^i\left|\vec{x}'\right\rangle \tag{5.2.41}$$

$$\left|f\right\rangle = \int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}'}{(2\pi)^D}\left|\vec{k}'\right\rangle\left\langle\vec{k}'\middle|f\right\rangle, \qquad P_j\left|\vec{k}'\right\rangle = k'_j\left|\vec{k}'\right\rangle. \tag{5.2.42}$$

For those already familiar with quantum theory, notice there is no $\hbar$ on the right hand side; nor will there be any throughout this section. This is not because we have "set $\hbar = 1$" as is commonly done in theoretical physics literature. Rather, it is because we wish to reiterate that the linear algebra of continuous operators, just like its discrete finite dimension counterparts, is really an independent structure on its own. Quantum theory is merely one of its application, albeit a very important one.

Problem 5.11. If $\left|\vec{k}\right\rangle$ is a momentum eigenket, explain why

$$\mathcal{T}(\vec{d}) \left|\vec{k}\right\rangle = e^{-i\vec{k}\cdot\vec{d}} \left|\vec{k}\right\rangle. \tag{5.2.43}$$

Then explain why the translation operator itself can be therefore be written as

$$\mathcal{T}(\vec{d}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{-i\vec{k}\cdot\vec{d}} \left|\vec{k}\right\rangle \left\langle\vec{k}\right|. \tag{5.2.44}$$

□

Problem 5.12. Employ the position representation of the momentum operator in eq. (5.2.24) to verify the relation

$$\left\langle \vec{x} \left| \left[X^k, P_j\right] \right| \alpha \right\rangle = i\delta^k_j \left\langle \vec{x} \right| \alpha \rangle. \tag{5.2.45}$$

This is, of course, consistent with eq. (5.2.40). □

Problem 5.13. Explain the following relation in the position space representation

$$\left\langle \alpha \left| \vec{P} \right| \beta \right\rangle = \int \mathrm{d}^D\vec{x} \left\langle \alpha \right| \vec{x} \rangle \left(-i\vec{\nabla}_{\vec{x}} \left\langle \vec{x} \right| \beta \rangle\right), \tag{5.2.46}$$

where $-i\vec{\nabla}_{\vec{x}} = -i\left(\partial_{x^1}, \dots, \partial_{x^D}\right)$. Hint: Use the completeness relation. □

Problem 5.14. Show that the negative of the Laplacian, namely

$$-\vec{\nabla}^2 \equiv -\sum_{i=1}^{D} \frac{\partial}{\partial x^i}\frac{\partial}{\partial x^i} \qquad \text{(in Cartesian coordinates } \{x^i\}), \tag{5.2.47}$$

is the square of the momentum operator. That is, for an arbitrary state $|\alpha\rangle$, show that

$$\left\langle \vec{x} \left| \vec{P}^2 \right| \alpha \right\rangle = -\delta^{ij} \frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j} \left\langle \vec{x} \right| \alpha \rangle \equiv -\vec{\nabla}^2 \left\langle \vec{x} \right| \alpha \rangle. \tag{5.2.48}$$

As we shall witness below, even when $\vec{P}$ ceases to be Hermitian – for instance, within a finite domain – the Laplacian itself may still continue to be Hermitian. Furthermore, in the differential geometry chapter below, we shall even learn how to generalize $\vec{\nabla}^2$ to *any* geometry and coordinate system, as well as to include time. □

Problem 5.15. Prove the Campbell-Baker-Hausdorff lemma. For linear operators A and B, and complex number α,

$$e^{i\alpha A} B e^{-i\alpha A} = B + \sum_{\ell=1}^{\infty} \frac{(i\alpha)^\ell}{\ell!} \underbrace{[A, [A, \dots [A}_{\ell \text{ of these}}, B]]]. \tag{5.2.49}$$

Hint: Taylor expand the left-hand-side and use mathematical induction.

Next, consider the expectation values of the position $\vec{X}$ and momentum $\vec{P}$ operator with respect to a general state $|\psi\rangle$:

$$\left\langle \psi \left| \vec{X} \right| \psi \right\rangle \qquad \text{and} \qquad \left\langle \psi \left| \vec{P} \right| \psi \right\rangle. \tag{5.2.50}$$

What happens to these expectation values when we replace $|\psi\rangle \to \mathcal{T}(\vec{d})\,|\psi\rangle$? □

Fourier Analysis We will now show how the concept of a Fourier transform readily arises from the study of the translation group we have developed so far. To initiate the discussion we start with eq. (5.2.24), with $|\alpha\rangle$ replaced with a momentum eigenket $|\vec{k}\rangle$. This yields the eigenvalue/vector equation for the momentum operator in the position representation.

$$\left\langle \vec{x} \left| \vec{P} \right| \vec{k} \right\rangle = \vec{k}\langle\vec{x}|\vec{k}\rangle = -i\frac{\partial}{\partial\vec{x}}\langle\vec{x}|\vec{k}\rangle, \qquad \Leftrightarrow \qquad k_j\langle\vec{x}|\vec{k}\rangle = -i\frac{\partial\langle\vec{x}|\vec{k}\rangle}{\partial x^j}. \tag{5.2.51}$$

In D-space, this is a set of D first order differential equations for the function $\langle\vec{x}|\vec{k}\rangle$. Via a direct calculation you can verify that the solution to eq. (5.2.51) is simply the plane wave

$$\langle\vec{x}|\vec{k}\rangle = \chi \exp\left(i\vec{k}\cdot\vec{x}\right). \tag{5.2.52}$$

where χ is complex constant to be fixed in the following way. We want

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D}\langle\vec{x}|\vec{k}\rangle\langle\vec{k}|\vec{x}'\rangle = \langle\vec{x}|\,\vec{x}'\rangle = \delta^{(D)}(\vec{x}-\vec{x}'). \tag{5.2.53}$$

Using the plane wave solution,

$$(2\pi)^D|\chi|^2\int\frac{\mathrm{d}^D k}{(2\pi)^D}e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} = \delta^{(D)}(\vec{x}-\vec{x}'). \tag{5.2.54}$$

Now, recall the representation of the D-dimensional δ-function

$$\int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} = \delta^{(D)}(\vec{x}-\vec{x}'). \tag{5.2.55}$$

Therefore, up to an overall multiplicative phase $e^{i\delta}$, which we will choose to be unity, $\chi = 1$ and eq. (5.2.52) becomes

$$\langle\vec{x}|\vec{k}\rangle = \exp\left(i\vec{k}\cdot\vec{x}\right). \tag{5.2.56}$$

By comparing eq. (5.2.56) with eq. (4.3.146), we see that the plane wave in eq. (5.2.56) can be viewed as the matrix element of the unitary operator implementing the change-of-basis from position to momentum space, and vice versa.

We may now examine how the position representation of an arbitrary state $\langle\vec{x}|\,f\rangle$ can be expanded in the momentum eigenbasis.

$$\langle\vec{x}|\,f\rangle = \int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\langle\vec{x}|\vec{k}\rangle\left\langle\vec{k}\middle|\,f\right\rangle = \int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}e^{i\vec{k}\cdot\vec{x}}\left\langle\vec{k}\middle|\,f\right\rangle \tag{5.2.57}$$

Similarly, we may expand the momentum representation of an arbitrary state $\left\langle\vec{k}\middle|\,f\right\rangle$ in the position eigenbasis.

$$\left\langle\vec{k}\middle|\,f\right\rangle = \int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}\left\langle\vec{k}\middle|\,\vec{x}\right\rangle\langle\vec{x}|\,f\rangle = \int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}e^{-i\vec{k}\cdot\vec{x}}\langle\vec{x}|\,f\rangle \tag{5.2.58}$$

Equations (5.2.57) and (5.2.58) are nothing but the Fourier expansion of some function $f(\vec{x})$ and its inverse transform.[26] We may sum up the discussion here with the following expansions:

$$|\vec{x}\rangle = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{-i\vec{k}\cdot\vec{x}} \left|\vec{k}\right\rangle, \tag{5.2.59}$$

$$\left|\vec{k}\right\rangle = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} e^{i\vec{k}\cdot\vec{x}} |\vec{x}\rangle. \tag{5.2.60}$$

Plane waves as orthonormal basis vectors For practical calculations, it is of course cumbersome to carry around the position $\{|\vec{x}\rangle\}$ or momentum eigenkets $\{|\vec{k}\rangle\}$. As far as the space of functions in $\mathbb{R}^D$ is concerned, i.e., if one works solely in terms of the components $f(\vec{x}) \equiv \langle\vec{x}| f\rangle$, as opposed to the space spanned by $|\vec{x}\rangle$, then one can view the plane waves $\{\exp(i\vec{k}\cdot\vec{x})/(2\pi)^{D/2}\}$ in the Fourier expansion of eq. (5.2.57) as the orthonormal basis vectors. The coefficients of the expansion are then the $\widetilde{f}(\vec{k}) \equiv \langle\vec{k}|f\rangle$.

$$f(\vec{x}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{i\vec{k}\cdot\vec{x}} \widetilde{f}(\vec{k}) \tag{5.2.61}$$

By multiplying both sides by $\exp(-i\vec{k}'\cdot\vec{x})$, integrating over all space, using the integral representation of the δ-function in eq. (5.1.6), and finally replacing $\vec{k}' \to \vec{k}$,

$$\widetilde{f}(\vec{k}) = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} e^{-i\vec{k}\cdot\vec{x}} f(\vec{x}). \tag{5.2.62}$$

Problem 5.16. Prove that, for the eigenstate of momentum $|\vec{k}\rangle$, arbitrary states $|\alpha\rangle$ and $|\beta\rangle$,

$$\left\langle \vec{k} \left| \vec{X} \right| \alpha \right\rangle = i\frac{\partial}{\partial\vec{k}} \left\langle \vec{k} \middle| \alpha \right\rangle \tag{5.2.63}$$

$$\left\langle \beta \left| \vec{X} \right| \alpha \right\rangle = \int \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left\langle \vec{k} \middle| \beta \right\rangle^* i\frac{\partial}{\partial\vec{k}} \left\langle \vec{k} \middle| \alpha \right\rangle. \tag{5.2.64}$$

The $\vec{X}$ is the position operator. □

Uncertainly relation According to (4.3.101) and (5.2.40), if we work in 1 dimension for now,

$$\left\langle \psi \left| \Delta X^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta P^2 \right| \psi \right\rangle \geq \frac{1}{4}. \tag{5.2.65}$$

In $D-$spatial dimensions, we may consider

$$\left\langle \psi \left| \Delta\vec{X}^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta\vec{P}^2 \right| \psi \right\rangle = \sum_{1\leq i,j\leq D} \left\langle \psi \left| (\Delta X^i)^2 \right| \psi \right\rangle \left\langle \psi \left| (\Delta P_j)^2 \right| \psi \right\rangle. \tag{5.2.66}$$

[26] A warning on conventions: our Fourier transform conventions will be $\int \mathrm{d}^D k/(2\pi)^D$ for the momentum integrals and $\int \mathrm{d}^D x$ for the position space integrals; these conventions can be traced back to equations (5.2.41) and (5.2.42). This is just a matter of where the (2π)s are allocated, and no math/physics content is altered.

We may apply eq. (4.3.101) on each term in the sum – i.e., identify ΔX^i(here) $\leftrightarrow \Delta X$(4.3.101) and ΔP_j(here) $\leftrightarrow \Delta Y$(4.3.101) – to deduce,

$$\left\langle \psi \left| \Delta \vec{X}^2 \right| \psi \right\rangle \left\langle \psi \left| \Delta \vec{P}^2 \right| \psi \right\rangle \geq \frac{1}{4} \sum_{1 \leq i,j \leq D} |\left\langle \psi \left| [X^i, P_j] \right| \psi \right\rangle|^2 = \sum_{i,j} \frac{\delta^i_j}{4} = \frac{D}{4}. \tag{5.2.67}$$

This is the generalization of eq. (5.2.65) to $D-$dimensions.

Problem 5.17. Planck's constant: From inverse length to momentum Notice that eq. (5.2.24) tells us the "momentum operator" has dimension

$$[P_j] = 1/\text{Length}. \tag{5.2.68}$$

Explain why, to re-scale $P_j \to \kappa P_j$ to an object κP_j that truly has dimension of momentum, the κ must have dimension of Planck's (reduced) constant $\hbar$. □

From Linear Algebra to Quantum Mechanics What you have discovered in Problem (5.17) is that, upon rescaling $\vec{P}_{\text{new}} \equiv \hbar \vec{P}_{\text{old}}$, it is the 'new' translation generator $\vec{P}_{\text{new}}$ that has the correction dimensions of momentum. Therefore, the eigenvalues are now $\vec{p} \equiv \hbar \vec{k}$ with corresponding eigenstate $\left| \vec{p} \equiv \hbar \vec{k} \right\rangle$; the position representation in eq. (5.2.24) is now

$$\langle \vec{x} | P_j | \psi \rangle = -i\hbar \partial_j \langle \vec{x} | \psi \rangle ; \tag{5.2.69}$$

while the commutation relations of eq. (5.2.40) becomes

$$\left[X^i, P_j \right] = i\hbar \delta^i_j. \tag{5.2.70}$$

– i.e., each P on the left hand side now has a $\hbar$. Furthermore, the change-of-basis matrix element in eq. (5.2.56), i.e., the plane wave, may now be re-expressed as

$$\langle \vec{x} | \vec{p} \rangle = \exp\left(\frac{i}{\hbar} \vec{p} \cdot \vec{x} \right), \qquad \vec{p} \equiv \hbar \vec{k}. \tag{5.2.71}$$

Problem 5.18. Gaussian States and Uncertainty Relations Consider the function, with $d > 0$,

$$\langle \vec{x} | \psi \rangle = \frac{e^{i\vec{k}\cdot\vec{x}}}{(\sqrt{\pi} d)^{D/2}} \exp\left(-\frac{\vec{x}^2}{2d^2} \right). \tag{5.2.72}$$

Compute $\langle \vec{k}' | \psi \rangle$, the state $|\psi\rangle$ in the momentum eigenbasis. Let $\vec{X}$ and $\vec{P}$ denote the position and momentum operators. Calculate the following expectation values:

$$\left\langle \psi \left| \vec{X} \right| \psi \right\rangle, \qquad \left\langle \psi \left| \vec{X}^2 \right| \psi \right\rangle, \qquad \left\langle \psi \left| \vec{P} \right| \psi \right\rangle, \qquad \left\langle \psi \left| \vec{P}^2 \right| \psi \right\rangle. \tag{5.2.73}$$

What is the value of

$$\left(\left\langle \psi \left| \vec{X}^2 \right| \psi \right\rangle - \left\langle \psi \left| \vec{X} \right| \psi \right\rangle^2 \right) \left(\left\langle \psi \left| \vec{P}^2 \right| \psi \right\rangle - \left\langle \psi \left| \vec{P} \right| \psi \right\rangle^2 \right)? \tag{5.2.74}$$

Hint: In this problem you will need the following results

$$\int_{-\infty}^{+\infty} \mathrm{d}x e^{-a(x+iy)^2} = \int_{-\infty}^{+\infty} \mathrm{d}x e^{-ax^2} = \sqrt{\frac{\pi}{a}}, \qquad a > 0, y \in \mathbb{R}. \tag{5.2.75}$$

If you encounter an integral of the form

$$\int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' e^{-\alpha \vec{x}^2} e^{i\vec{x}\cdot(\vec{q}-\vec{q}')}, \qquad \alpha > 0, \tag{5.2.76}$$

you should try to combine the exponents and "complete the square". Do you find that the uncertainty relation in eq. (5.2.67) to be saturated? □

Problem 5.19. Free Particle in Quantum Mechanics: Wave Packets The Schrödinger equation governing the time evolution of quantum states is

$$i\hbar\partial_t |\psi\rangle = H |\psi\rangle . \tag{5.2.77}$$

The free particle is described by the 'pure kinetic energy' Hamiltonian

$$H = \frac{\vec{P}^2}{2m}, \tag{5.2.78}$$

where $m > 0$ is the particle's mass. As we have discussed in eq. (5.1.59), given an initial state $|\psi(t_0)\rangle$, the solution for $t > t_0$ is given by

$$|\psi(t)\rangle = U |\psi(t_0)\rangle = \exp\left(-i(t-t_0)H\right) |\psi(t_0)\rangle . \tag{5.2.79}$$

Explain why the following solutions for $\widetilde{\psi}(t > t_0, \vec{k}) \equiv \langle \vec{k}|\psi(t > t_0)\rangle$ are equivalent.

$$\widetilde{\psi}(t > t_0, \vec{k}) = \exp\left(-i(t-t_0)\frac{\vec{k}^2}{2m}\right) \widetilde{\psi}_0(\vec{k}), \qquad \widetilde{\psi}_0(\vec{k}) \equiv \left\langle \vec{k}\middle| \psi(t_0)\right\rangle \tag{5.2.80}$$

$$= \int \mathrm{d}^D \vec{x}' \exp\left(-i(t-t_0)\frac{\vec{k}^2}{2m} - i\vec{k}\cdot\vec{x}'\right) \psi(\vec{x}'), \qquad \psi_0(\vec{x}) \equiv \langle \vec{x}|\, \psi(t_0)\rangle . \tag{5.2.81}$$

Explain why the following solutions for $\psi(t > t_0, \vec{x}) \equiv \langle \vec{x}|\, \psi(t)\rangle$ are equivalent.

$$\psi(t > t_0, \vec{x}) = \int \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \exp\left(-i(t-t_0)\frac{\vec{k}^2}{2m} + i\vec{k}\cdot\vec{x}\right) \widetilde{\psi}_0(\vec{k}) \tag{5.2.82}$$

$$= \int \mathrm{d}^D \vec{x}' K(t-t_0; \vec{x}-\vec{x}')\psi(\vec{x}') \tag{5.2.83}$$

Here, the $K(\tau; \vec{z})$ is known as the quantum mechanical propagator of the free particle; you should find that it reads

$$K(\tau; \vec{z}) = \int \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \exp\left(-i\tau\frac{\vec{k}^2}{2m} + i\vec{k}\cdot\vec{z}\right). \tag{5.2.84}$$

Can you evaluate the integral? Answer: with the square root denoting the positive one,

$$K(\tau;\vec{z}) = e^{-iD\frac{\pi}{4}} \left(\frac{m}{2\pi\tau}\right)^{\frac{D}{2}} \exp\left[i\frac{m\vec{z}^2}{2\tau}\right]. \tag{5.2.85}$$

Hint: You may wish to start on this problem by inserting a complete set of momentum states between the time evolution operator U and $|\psi(t_0)\rangle$. $\square$

Space Translation of Momentum Eigenket Let $|\vec{k}\rangle$ be an eigenket of the momentum operator $\vec{P}$. Notice that

$$\mathcal{T}(\vec{d})\left|\vec{k}\right\rangle = \exp(-i\vec{d}\cdot\vec{P})\left|\vec{k}\right\rangle = \exp(-i\vec{d}\cdot\vec{k})\left|\vec{k}\right\rangle. \tag{5.2.86}$$

In words: the momentum eigenstate $|\vec{k}\rangle$ is an eigenvector of $\mathcal{T}(\vec{d})$ with eigenvalue $\exp(-i\vec{d}\cdot\vec{k})$. Since this is merely a phase, in quantum mechanics, we would regard $\mathcal{T}(\vec{d})|\vec{k}\rangle$ and $|\vec{k}\rangle$ as the same physical ket: i.e., space-translation merely shifts the momentum eigenket by a phase.

However, if we were to translate a superposition over different $\vec{k}$ modes,

$$\mathcal{T}(\vec{d})\left|f\right\rangle = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left|\vec{k}\right\rangle\left\langle\vec{k}\right| e^{-i\vec{d}\cdot\vec{P}}\left|f\right\rangle \tag{5.2.87}$$

$$= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left|\vec{k}\right\rangle\left(e^{+i\vec{d}\cdot\vec{P}}\left|\vec{k}\right\rangle\right)^\dagger\left|f\right\rangle = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left|\vec{k}\right\rangle e^{-i\vec{k}\cdot\vec{d}}\widetilde{f}(\vec{k}). \tag{5.2.88}$$

Under a translation $|f\rangle \to \mathcal{T}(\vec{d})\,|f\rangle$ we see the Fourier coefficients transform as

$$\widetilde{f}(\vec{k}) \to e^{-i\vec{k}\cdot\vec{d}}\widetilde{f}(\vec{k}). \tag{5.2.89}$$

Momentum Translation We have discussed how to implement translation in position space using the momentum operator $\vec{P}$, namely $\mathcal{T}(\vec{d}) = \exp(-i\vec{d}\cdot\vec{P})$. What would be the corresponding translation operator in momentum space?[27] That is, what is $\widetilde{\mathcal{T}}$ such that

$$\widetilde{\mathcal{T}}(\vec{d})|\vec{k}\rangle = \left|\vec{k}+\vec{d}\right\rangle, \qquad P_j|\vec{k}\rangle = k_j|\vec{k}\rangle? \tag{5.2.90}$$

Of course, one representation would be the analog of eq. (5.2.19).

$$\widetilde{\mathcal{T}}(\vec{d}) = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{k}'\left|\vec{k}'+\vec{d}\right\rangle\left\langle\vec{k}'\right| \tag{5.2.91}$$

But is there an exponential form, like there is one for the translation in position space (eq. (5.2.20))? We start with the observation that the momentum eigenstate $|\vec{k}\rangle$ can be written as a superposition of the position eigenkets using eq. (5.2.56),

$$|\vec{k}\rangle = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}'\,|\vec{x}'\rangle\left\langle\vec{x}'\middle|\vec{k}\right\rangle = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{x}'}{(2\pi)^{D/2}} e^{i\vec{k}\cdot\vec{x}'}\,|\vec{x}'\rangle. \tag{5.2.92}$$

[27] This question was suggested by Jake Leistico, who also correctly guessed the essential form of eq. (5.2.94).

Now consider

$$\exp(+i\vec{d}\cdot\vec{X})|\vec{k}\rangle = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{x}'}{(2\pi)^{D/2}} e^{i\vec{k}\cdot\vec{x}'} e^{i\vec{d}\cdot\vec{x}'} |\vec{x}'\rangle$$
$$= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{x}'}{(2\pi)^{D/2}} e^{i(\vec{k}+\vec{d})\cdot\vec{x}'} |\vec{x}'\rangle = \left|\vec{k}+\vec{d}\right\rangle. \tag{5.2.93}$$

That means

$$\widetilde{\mathcal{T}}(\vec{d}) = \exp\left(i\vec{d}\cdot\vec{X}\right). \tag{5.2.94}$$

Spectra of $\vec{P}$ and $\vec{P}^2$ in infinite $\mathbb{R}^D$ We conclude this section by summarizing the several interpretations of the plane waves $\{\langle\vec{x}|\vec{k}\rangle \equiv \exp(i\vec{k}\cdot\vec{x})\}$.

1. They can be viewed as the orthonormal basis vectors (in the δ-function sense) spanning the space of complex functions on $\mathbb{R}^D$.

2. They can be viewed as the matrix element of the unitary operator U that performs a change-of-basis between the position and momentum eigenbasis, namely $U|\vec{x}\rangle = |\vec{k}\rangle$.

3. They are simultaneous eigenstates of the momentum operators $\{-i\partial_j \equiv -i\partial/\partial x^j | j = 1, 2, \dots, D\}$ and the negative Laplacian $-\vec{\nabla}^2$ in the position representation.

$$-\vec{\nabla}^2_{\vec{x}}\langle\vec{x}|\vec{k}\rangle = \vec{k}^2\langle\vec{x}|\vec{k}\rangle, \qquad -i\partial_j\langle\vec{x}|\vec{k}\rangle = k_j\langle\vec{x}|\vec{k}\rangle, \qquad \vec{k}^2 \equiv \delta^{ij}k_ik_j. \tag{5.2.95}$$

The eigenvector/value equation for the momentum operators had been solved previously in equations (5.2.51) and (5.2.52). For the negative Laplacian, we may check

$$-\vec{\nabla}^2_{\vec{x}}\langle\vec{x}|\vec{k}\rangle = \left\langle\vec{x}\left|\vec{P}^2\right|\vec{k}\right\rangle = \vec{k}^2\langle\vec{x}|\vec{k}\rangle. \tag{5.2.96}$$

That the plane waves are simultaneous eigenvectors of P_j and $\vec{P}^2 = -\vec{\nabla}^2$ is because these operators commute amongst themselves: $[P_j, \vec{P}^2] = [P_i, P_j] = 0$. This is therefore an example of *degeneracy*. For a fixed eigenvalue k^2 of the negative Laplacian, there is a continuous infinity of eigenvalues of the momentum operators, only constrained by

$$\vec{k}^2 \equiv \sum_{j=1}^{D}(k_j)^2 = k^2, \qquad \vec{P}^2\left|k^2; k_1 \dots k_D\right\rangle = k^2\left|k^2; k_1 \dots k_D\right\rangle. \tag{5.2.97}$$

Physically speaking we may associate this degeneracy with the presence of rotational symmetry of the underlying infinite flat $\mathbb{R}^D$: the eigenvalue of $\vec{P}^2$, namely $\vec{k}^2$, is the same no matter where $\vec{k}/|\vec{k}|$ is pointing.

Additionally, eq. (5.2.86) tells us that, both $\mathcal{T}(\vec{d})|\vec{k}\rangle$ and $|\vec{k}\rangle$ are eigenkets of $\vec{P}^2$ with eigenvalue k^2. This is of course because $[\mathcal{T}(\vec{d}), \vec{P}] = 0$ and is, in turn, a consequence of translation symmetry of the underlying flat space.

5.3 Boundary Conditions, Finite Box, Periodic functions and the Fourier Series

Up to now we have not been terribly precise about the boundary conditions obeyed by our states $\langle \vec{x} | f \rangle$, except to say they are functions residing in an infinite space $\mathbb{R}^D$. Let us now rectify this glaring omission – drop the assumption of infinite space $\mathbb{R}^D$ – and study how, in particular, the Hermitian nature of the $\vec{P}^2 \equiv -\vec{\nabla}^2$ operator now depends crucially on the boundary conditions obeyed by its eigenstates. If $\vec{P}^2$ is Hermitian,

$$\left\langle \psi_1 \left| \vec{P}^2 \right| \psi_2 \right\rangle = \left\langle \psi_1 \left| \left(\vec{P}^2\right)^\dagger \right| \psi_2 \right\rangle = \left\langle \psi_2 \left| \vec{P}^2 \right| \psi_1 \right\rangle^*, \tag{5.3.1}$$

for any states $|\psi_{1,2}\rangle$. Inserting a complete set of position eigenkets, and using

$$\left\langle \vec{x} \left| \vec{P}^2 \right| \psi_{1,2} \right\rangle = -\vec{\nabla}^2_{\vec{x}} \langle \vec{x} | \psi_{1,2} \rangle, \tag{5.3.2}$$

we arrive at the condition that, if $\vec{P}^2$ is Hermitian then the negative Laplacian can be "integrated-by-parts" to act on either ψ_1 or ψ_2.

$$\begin{aligned}
\int_{\mathfrak{D}} \mathrm{d}^D x \, \langle \psi_1 | \vec{x} \rangle \left\langle \vec{x} \left| \vec{P}^2 \right| \psi_2 \right\rangle &\stackrel{?}{=} \int_{\mathfrak{D}} \mathrm{d}^D x \, \langle \psi_2 | \vec{x} \rangle^* \left\langle \vec{x} \left| \vec{P}^2 \right| \psi_1 \right\rangle^*, \\
\int_{\mathfrak{D}} \mathrm{d}^D x \psi_1(\vec{x})^* \left(-\vec{\nabla}^2_{\vec{x}} \psi_2(\vec{x})\right) &\stackrel{?}{=} \int_{\mathfrak{D}} \mathrm{d}^D x \left(-\vec{\nabla}^2_{\vec{x}} \psi_1(\vec{x})^*\right) \psi_2(\vec{x}), \qquad \psi_{1,2}(\vec{x}) \equiv \langle \vec{x} | \psi_{1,2} \rangle.
\end{aligned} \tag{5.3.3}$$

Notice we have to specify a domain $\mathfrak{D}$ to perform the integral. If we now proceed to work from the left hand side, and use Gauss' theorem from vector calculus,

$$\begin{aligned}
\int_{\mathfrak{D}} \mathrm{d}^D x \psi_1(\vec{x})^* \left(-\vec{\nabla}^2_{\vec{x}} \psi_2(\vec{x})\right) &= \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1} \vec{\Sigma} \cdot \left(-\vec{\nabla} \psi_1(\vec{x})^*\right) \psi_2(\vec{x}) + \int_{\mathfrak{D}} \mathrm{d}^D x \vec{\nabla} \psi_1(\vec{x})^* \cdot \vec{\nabla} \psi_2(\vec{x}) \\
&= \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1} \vec{\Sigma} \cdot \left\{ \left(-\vec{\nabla} \psi_1(\vec{x})^*\right) \psi_2(\vec{x}) + \psi_1(\vec{x})^* \vec{\nabla} \psi_2(\vec{x}) \right\} \\
&\quad + \int_{\mathfrak{D}} \mathrm{d}^D x \psi_1(\vec{x})^* \left(-\vec{\nabla}^2 \psi_2(\vec{x})\right)
\end{aligned} \tag{5.3.4}$$

Here, $\mathrm{d}^{D-1}\vec{\Sigma}$ is the $(D-1)$-dimensional analog of the 2D infinitesimal area element $\mathrm{d}\vec{A}$ in vector calculus, and is proportional to the unit (outward) normal $\vec{n}$ to the boundary of the domain $\partial\mathfrak{D}$. We see that integrating-by-parts the $\vec{P}^2$ from ψ_1 onto ψ_2 can be done, but would incur the two surface integrals. To get rid of them, we may demand the eigenfunctions $\{\psi_\lambda\}$ of $\vec{P}^2$ or their normal derivatives $\{\vec{n} \cdot \vec{\nabla} \psi_\lambda\}$ to be zero:

$$\psi_\lambda(\partial\mathfrak{D}) = 0 \text{ (Dirichlet)} \qquad \text{or} \qquad \vec{n} \cdot \vec{\nabla} \psi_\lambda(\partial\mathfrak{D}) = 0 \text{ (Neumann)}. \tag{5.3.5}$$

[28]**No boundaries** The exception to the requirement for boundary conditions, is when the domain $\mathfrak{D}$ itself has *no boundaries* – there will then be no "surface terms" to speak of, and the

[28]We may also allow the eigenfunctions to obey a mixed boundary condition, but we will stick to either Dirichlet or Neumann for simplicity.
Moreover, in a non-relativistic quantum mechanical system with Hamiltonian equals to kinetic $(2m)^{-1}\vec{P}^2$ plus potential $V(\vec{X})$; when $\psi_1 = \psi_2 \equiv \psi$ the integrand $\vec{J} = \psi^* \vec{\nabla} \psi - \psi \vec{\nabla} \psi^*$ within the surface integral of eq. (5.3.4) is proportional to the probability current. Choosing the right boundary conditions to set $\vec{J} = 0$, so as to guarantee the hermicity of $\vec{P}^2$, then amounts to, in this limit, ensuring there is zero flow of probability outside the domain $\mathfrak{D}$ under consideration.

Laplacian is hence automatically Hermitian. In this case, the eigenfunctions often obey periodic boundary conditions; we will see examples below.

> **Boundary Conditions** The abstract bra-ket notation $\langle\psi_1|\vec{P}^2|\psi_2\rangle$ obscures the fact that boundary conditions are required to ensure the Hermitian nature of $\vec{P}^2$ in a finite domain. Not only do we have to specify what the domain $\mathfrak{D}$ of the underlying space actually is; to ensure $\vec{P}^2$ remains Hermitian, we may demand the eigenfunctions or their normal derivatives (expressed in the position representation) to vanish on the boundary $\partial\mathfrak{D}$.

In the discussion of partial differential equations below, we will generalize this analysis to curved spaces.

Example: Finite box The first illustrative example is as follows. Suppose our system is defined only in a finite box. For the ith Cartesian axis, the box is of length L^i. If we demand that the eigenfunctions of $-\vec{\nabla}^2$ vanish at the boundary of the box, we find the eigensystem

$$-\vec{\nabla}^2_{\vec{x}} \langle\vec{x}|\,\vec{n}\rangle = \lambda(\vec{n})\,\langle\vec{x}|\,\vec{n}\rangle\,, \qquad \left\langle\vec{x}; x^i = 0\middle|\,\vec{n}\right\rangle = \left\langle\vec{x}; x^i = L^i\middle|\,\vec{n}\right\rangle = 0, \tag{5.3.6}$$

$$i = 1, 2, 3, \dots, D, \tag{5.3.7}$$

admits the solution

$$\langle\vec{x}|\,\vec{n}\rangle \propto \prod_{i=1}^{D} \sin\left(\frac{\pi n^i}{L^i}x^i\right), \qquad \lambda(\vec{n}) = \sum_{i=1}^{D}\left(\frac{\pi n^i}{L^i}\right)^2. \tag{5.3.8}$$

These $\{n^i\}$ runs over the positive integers only; because sine is an odd function, the negative integers do not yield new solutions.

Problem 5.20. Verify that the basis eigenkets in eq. (5.3.8) do solve eq. (5.3.6). What is the correct normalization for $\langle\vec{x}|\,\vec{n}\rangle$? Also verify that the basis plane waves in eq. (5.3.25) satisfy the normalization condition in eq. (5.3.24).

Use these $\{|\vec{n}\rangle\}$ to solve the free particle Schrödinger equation:

$$\langle\vec{n}\,|i\hbar\partial_t|\,\psi(t)\rangle = \left\langle\vec{n}\,\middle|\vec{P}^2/(2m)\middle|\,\psi(t)\right\rangle \tag{5.3.9}$$

with the initial conditions $\langle\vec{n}|\,\psi(t = t_0)\rangle = 1$. Then solve for $\langle\vec{x}|\,\psi(t)\rangle$. □

Finite Domains & Translation Symmetry Let us recall that, in infinite flat space, the translation operator was unitary because of spatial-translation symmetry. In a finite domain $\mathfrak{D}$; we expect this symmetry to be broken due to the presence of the boundaries, which *does* select a privileged set of position eigenkets. More specifically, the domain is 'here' and not 'there': translating a position eigenket $|\vec{x}\rangle \to |\vec{x}+\vec{d}\rangle$ may in fact place it completely outside the domain, rendering it non-existent.

To be sure, the Taylor expansion of a function,

$$f(\vec{x}+\vec{d}) = \exp\left(d^j\partial_j\right)\,\langle\vec{x}|\,f\rangle \tag{5.3.10}$$

still holds, as long as both $\vec{x}$ and $\vec{x}+\vec{d}$ lie within $\mathfrak{D}$. This means equations (5.2.20), (5.2.24), and (5.2.26) are still valid – the $\mathcal{T}(\vec{d}) = \exp(-i\vec{d}\cdot\vec{P})$ form of the translation operator itself may still be employed – as long as the associated displacement is not too large.

On the other hand, let us study this breaking of translation symmetry in a simple example, by working in a 1D 'box' of size L parametrized by z; restricted to $0 \leq z \leq L$. This means the position eigenket $|z\rangle$ cannot be translated further that $L-z$ to the right or further than z to the left, because it will be outside the box. Moreover, we may attempt to construct the analogue of eq. (5.2.19):

$$\mathcal{T}(d>0) \stackrel{?}{=} \int_0^L \mathrm{d}z' \,|z'+d\rangle \langle z'| \,. \tag{5.3.11}$$

In fact, this would not work because of the reasons already alluded to above. When $z'=L$, for example, the $|L+d\rangle$ contribution to eq. (5.3.11) would not make sense. Likewise, for $d<0$ and $z'=0$, the $\langle d|$ in eq. (5.3.11) would, too, be non-existent. More generally, for $d>0$, the bras $\{\langle z'>L-d|\}$ and ket $\{|z'>L-d\rangle\}$ when translated by d,

$$\langle z'| \to \langle z'+d| \qquad \text{and} \qquad |z'\rangle \to |z'+d\rangle \,; \tag{5.3.12}$$

would place them entirely out of the box.

> The translation operator in a finite 1D box cannot be a change-of-basis operator because some of the position eigenkets will be moved outside the box by the translation operation. Hence, the translation operator cannot be unitary.

Kinetic Energy vs Momentum Notice that, even though P^2 is Hermitian in this finite domain $0 \leq z \leq L$ (if we, say, impose Dirichlet boundary conditions), P itself is no longer Hermitian. For, if it were Hermitian, the translation operator $\mathcal{T}(\xi) = \exp(-i\xi\cdot P)$ would be unitary, contradicting what we have just uncovered. This is a subtle point: even though $[P, P^2]=0$, because P itself is no longer Hermitian, P^2 and P are no longer simultaneously diagonalizable. Specifically, from eq. (5.3.6) and (5.3.8), we recall the eigensystem relation

$$\left\langle z \left| P^2 \right| n \right\rangle = \left(\frac{\pi n}{L}\right)^2 \langle z | n\rangle \,; \tag{5.3.13}$$

but on the other hand,

$$\langle z \,| P |\, n\rangle \propto \frac{\pi n}{L} \cdot \cos\left(\frac{\pi n}{L} z\right), \tag{5.3.14}$$

which is not proportional to $\langle z | n\rangle \propto \sin[(\pi n/L)x]$ – namely, P^2 and P do not share eigensystems, because the latter is simply not Hermitian.

Local vs Global Symmetry What we have described is the breaking of *global* symmetry (and its consequences): translating the entire box does not work, because it would render part of the box non-existent, due to the presence of the boundaries. However, when we restrict the domain $\mathfrak{D}$ to a finite one embedded within flat $\mathbb{R}^D$, there is still *local* translation symmetry in that, performing the same experiment at $\vec{x}$ and at $\vec{x}'$ should not lead to any physical differences as long as both $\vec{x}$ and $\vec{x}'$ lie within the said domain. For instance, we have already noted that the exponential form of the translation operator in eq. (5.2.20) still properly implements local translations, so long as the displacement is not too large.

To further quantify local translation symmetry, let us remain in the 1D box example. We may construct – instead of eq. (5.3.11) – a *local* translation operator in the ket-bra form, in the following manner. Suppose we wish to translate the region $0 < a < z < b < L$ by $\varepsilon > 0$ either to the left or to the right. As long as $\varepsilon < \min(a, L - b)$, we will not run intro trouble: the entire region will still remain in the box. Moreover, the region $a + \varepsilon < z < b$ will remain within the original region $a < z < b$ if it were a left-translation; while $a < z < b - \varepsilon$ remains within the original region if it were a right-translation. These considerations suggest that we consider

$$\mathcal{T}(\varepsilon|a,b) \equiv \int_a^b \mathrm{d}z' \left|z' + \varepsilon\right\rangle \left\langle z'\right| . \tag{5.3.15}$$

For an arbitrary position eigenket $|z\rangle$, we may compute

$$\mathcal{T}(\varepsilon|a,b) \left|z\right\rangle = \int_a^b \mathrm{d}z' \left|z' + \varepsilon\right\rangle \delta(z' - z) \tag{5.3.16}$$

$$= \Theta(z - a)\Theta(b - z) \left|z + \varepsilon\right\rangle . \tag{5.3.17}$$

The $\Theta(z-a)\Theta(b-z)$ is the 'top-hat' function, which is unity within the interval $a < z < b$ and zero outside. The reason for its appearance is, the δ-function within the integral of eq. (5.3.16) is zero unless $z' = z$ and $a < z' < b$ simultaneously. Therefore, as expected, eq. (5.3.16) provides a well-defined ket-bra form of the translation operator; but restricted to acting upon kets lying within (a, b) and for small enough ε.

Problem 5.21. Local 'Identity' and 'Unitary'-Translation Operators Explain why, for $0 < a < b < L$,

$$\mathbb{I}(a,b) \equiv \int_a^b \mathrm{d}z' \left|z'\right\rangle \left\langle z'\right| \tag{5.3.18}$$

is the identity operator when acting on position eigenkets lying within the interval (a, b) in the 1D box example above. Next, verify that $\mathcal{T}(\varepsilon|a,b)$ in eq. (5.3.16) obeys

$$\mathcal{T}(\varepsilon|a,b)^\dagger \mathcal{T}(\varepsilon|a,b) = \mathbb{I}(a+\varepsilon, b+\varepsilon). \tag{5.3.19}$$

Since $\mathbb{I}(a+\varepsilon, b+\varepsilon)$ is the identity on the interval $(a+\varepsilon, b+\varepsilon)$, eq. (5.3.19) may be regarded as a restricted form of the unitary condition $U^\dagger U = \mathbb{I}$. This, in turn, may be interpreted as a consequence of *local* translation symmetry. □

Periodic Domains: the Fourier Series. If we stayed within the infinite space, but instead imposed periodic boundary conditions,

$$\left\langle \vec{x}; x^i \to x^i + L^i \middle| f \right\rangle = \left\langle \vec{x}; x^i \middle| f \right\rangle , \tag{5.3.20}$$

$$f(x^1, \dots, x^i + L^i, \dots, x^D) = f(x^1, \dots, x^i, \dots, x^D) = f(\vec{x}), \tag{5.3.21}$$

this would mean, not all the basis plane waves from eq. (5.2.56) remains in the Hilbert space. Instead, periodicity means

$$\langle \vec{x}; x^j = x^j + L^j|\vec{k}\rangle = \langle \vec{x}; x^j = x^j|\vec{k}\rangle$$

$$e^{ik_j(x^j+L^j)} = e^{ik_j x^j}, \qquad \text{(No sum over } j.\text{)} \tag{5.3.22}$$

(The rest of the plane waves, $e^{ik_l x^l}$ for $l \neq j$, cancel out of the equation.) This further implies the eigenvalue k_j becomes discrete:

$$e^{ik_j L^j} = 1 \text{ (No sum over } j.\text{)} \qquad \Rightarrow \qquad k_j L^j = 2\pi n \qquad \Rightarrow \qquad k_j = \frac{2\pi n^j}{L^j}, \\ n^j = 0, \pm 1, \pm 2, \pm 3, \dots . \tag{5.3.23}$$

We need to re-normalize our basis plane waves. In particular, since space is now periodic, we ought to only need to integrate over one typical volume.

$$\int_{\{0 \leq x^i \leq L^i | i=1,2,\dots,D\}} \mathrm{d}^D\vec{x} \left\langle \vec{n}' \middle| \vec{x} \right\rangle \left\langle \vec{x} \middle| \vec{n} \right\rangle = \delta^{\vec{n}}_{\vec{n}'} \equiv \prod_{i=1}^{D} \delta^{n'^i}_{n^i}. \tag{5.3.24}$$

The set of orthonormal eigenvectors of the negative Laplacian may be taken as

$$\left\langle \vec{x} \middle| \vec{n} \right\rangle \equiv \prod_{j=1}^{D} \frac{\exp\left(i\frac{2\pi n^j}{L^j} x^j\right)}{\sqrt{L^j}}, \tag{5.3.25}$$

$$-\vec{\nabla}^2 \left\langle \vec{x} \middle| \vec{n} \right\rangle = \lambda(\vec{n}) \left\langle \vec{x} \middle| \vec{n} \right\rangle, \qquad \lambda(\vec{n}) = \sum_i \left(\frac{2\pi n^i}{L^i}\right)^2. \tag{5.3.26}$$

Notice the basis vectors in eq. (5.3.25) are momentum eigenkets too:

$$\left\langle \vec{x} \middle| P_j \middle| \vec{n} \right\rangle = -i\partial_j \left\langle \vec{x} \middle| \vec{n} \right\rangle = k_i \left\langle \vec{x} \middle| \vec{n} \right\rangle, \qquad k_i(n^i) = \frac{2\pi n^i}{L^i}. \tag{5.3.27}$$

Even though sines and cosines are also eigenfunctions of $\vec{\nabla}^2$, they are no longer eigenfunctions of $-i\partial_j$. We may use these simultaneous eigenkets of P_j and $\vec{P}^2$ to write the identity operator – i.e., the completeness relation:

$$\left\langle \vec{x} \middle| \vec{x}' \right\rangle = \delta^{(D)}(\vec{x} - \vec{x}') = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \left\langle \vec{x} \middle| \vec{n} \right\rangle \left\langle \vec{n} \middle| \vec{x}' \right\rangle, \tag{5.3.28}$$

$$\mathbb{I} = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \left| \vec{n} \right\rangle \left\langle \vec{n} \right|. \tag{5.3.29}$$

We may also use the position eigenkets themselves to write $\mathbb{I}$, but instead of integrating over all space, we only integrate over one domain (since space is now periodic):

$$\mathbb{I} = \int_{\{0 \leq x^i \leq L^i | i=1,2,\dots,D\}} \mathrm{d}^D\vec{x} \left| \vec{x} \right\rangle \left\langle \vec{x} \right|. \tag{5.3.30}$$

To summarize our discussion here: any periodic function f, subject to eq. (5.3.21), can be expanded as a superposition of periodic plane waves in eq. (5.3.25),

$$f(\vec{x}) = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \widetilde{f}(n^1, \dots, n^D) \prod_{j=1}^{D} (L^j)^{-1/2} \exp\left(i\frac{2\pi n^j}{L^j} x^j\right). \tag{5.3.31}$$

This is known as the *Fourier series*. By using the inner product in eq. (5.3.24), or equivalently, multiplying both sides of eq. (5.3.31) by $\prod_j (L^j)^{-1/2} \exp(-i(2\pi n'^j/L^j)x^j)$ and integrating over a typical volume, we obtain the coefficients of the Fourier series expansion

$$\widetilde{f}(n^1, n^2, \dots, n^D) = \int_{0 \le x^j \le L^j} \mathrm{d}^D \vec{x} f(\vec{x}) \prod_{j=1}^{D} (L^j)^{-1/2} \exp\left(-i\frac{2\pi n^j}{L^j} x^j\right). \tag{5.3.32}$$

Remark I The exp in eq. (5.3.25) are not a unique set of basis vectors, of course. One could use sines and cosines instead, for example.
Remark II Even though we are explicitly integrating the ith Cartesian coordinate from 0 to L^i in eq. (5.3.32), since the function is periodic, we really just need only to integrate over a complete period, from κ to $\kappa + L^i$ (for κ real), to achieve the same result. For example, in 1D, and whenever $f(x)$ is periodic (with a period of L),

$$\int_0^L \mathrm{d}x f(x) = \int_\kappa^{\kappa+L} \mathrm{d}x f(x). \tag{5.3.33}$$

(Drawing a plot here may help to understand this statement.)

Problem 5.22. Translation operator in ket-bra form Construct the translation operator in the ket-bra form, analogous to eq. (5.2.19). Verify that the translation operator in a periodic space is unitary. Can you explain why it is so – in words? □

From eq. (5.3.29) we see that any state $|f\rangle$ may be expanded as

$$|f\rangle = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} |\vec{n}\rangle \langle \vec{n}| f\rangle . \tag{5.3.34}$$

If we apply the translation operator directly to this, eq. (5.3.27) tells us $\mathcal{T}(\vec{d})\,|f\rangle$ is

$$e^{-i\vec{d}\cdot\vec{P}}\,|f\rangle = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} e^{-i\vec{d}\cdot\vec{k}(\vec{n})} |\vec{n}\rangle \langle \vec{n}| f\rangle \tag{5.3.35}$$

$$= \left(\sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \exp\left(-i\vec{d}\cdot\vec{k}(\vec{n})\right) |\vec{n}\rangle \langle \vec{n}| \right) |f\rangle . \tag{5.3.36}$$

Since $|f\rangle$ was arbitrary, we have managed to uncover the diagonal form of the translation operator:

$$\mathcal{T}(\vec{d}) = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \exp\left(-i\vec{d}\cdot\vec{k}(\vec{n})\right) |\vec{n}\rangle \langle \vec{n}| , \qquad k_j = \frac{2\pi n^j}{L^j}. \tag{5.3.37}$$

5.4 Rotations and Translations in $D = 2$ Spatial Dimensions

In this and the following sections we will further develop linear operators in continuous vector spaces by extending our discussion in §(5.2) on spatial translations to that of rotations. This is not just a mathematical exercise, but has deep implications for the study of rotational symmetry in quantum systems as well as the meaning of particles in relativistic Quantum Field Theory.

Let us begin in 2D. We will use cylindrical coordinates, defined through the Cartesian ones $\vec{x}$ via

$$\vec{x}(r,\phi) \equiv r\,(\cos\phi, \sin\phi)\,, \qquad r \geq 0,\ 0 \leq \phi < 2\pi, \tag{5.4.1}$$

$$= r\cos\phi\ \widehat{e}_1 + r\sin\phi\ \widehat{e}_2; \tag{5.4.2}$$

$$\widehat{e}^i_{\mathrm{I}} = \delta^i_{\mathrm{I}}, \qquad i, \mathrm{I} \in \{1,2\}. \tag{5.4.3}$$

We may study the rotation of any arbitrary vector by first studying its effect on the basis unit vectors $\widehat{e}_1$ and $\widehat{e}_2$. By geometry – drawing a picture helps – rotating $\widehat{e}_1$ and $\widehat{e}_2$ counterclockwise by an angle ϕ produces

$$\widehat{R}(\phi)\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} = \vec{x}(r,\phi)/r \tag{5.4.4}$$

$$\text{and} \qquad \widehat{R}(\phi)\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix} = -\sin\phi\ \widehat{e}_1 + \cos\phi\ \widehat{e}_2. \tag{5.4.5}$$

These immediately imply

$$\widehat{R}(\phi) \equiv \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}. \tag{5.4.6}$$

Using the purely off-diagonal Levi-Civita symbol ϵ^{ij}, we may also express this general 2D rotation matrix as the sum of its diagonal and off-diagonal components:

$$\widehat{R}(\phi)^{ij} = \cos\phi \cdot \delta^{ij} - \sin\phi \cdot \epsilon^{ij}. \tag{5.4.7}$$

The counter-clockwise rotated position vector

$$\widehat{R}(\phi)^{ij} x^j = \cos\phi \cdot x^i - \sin\phi \cdot \epsilon^{ij} x^j \tag{5.4.8}$$

may be understood as follows. The cosine term is simply the component of the rotated vector along the original $\vec{x}$; whereas the sine term is that perpendicular to $\vec{x}$. For the latter, by defining $x^i_\perp \equiv \epsilon^{ij} x^j$, we may readily verify that $\vec{x} \cdot \vec{x}_\perp = x^i \epsilon^{ij} x^j = 0$ by the anti-symmetry $\epsilon^{ij} = -\epsilon^{ji}$. Moreover, if say $x^i = \delta^i_1$, the orthogonal component is then $\epsilon^{ij}\delta^j_1 = -\epsilon^{1i} = -\delta^i_2$; confirming the rotation is indeed a counter-clockwise one for positive ϕ.

Problem 5.23. Basis-Independent Form For any pair of orthonormal basis $\{\widehat{e}_1, \widehat{e}_2\}$, argue that the rotation matrix can be expressed as

$$\widehat{R}^{ij}(\phi) = \cos\phi \cdot \delta^{ij} - \sin\phi \cdot \epsilon^{\mathrm{IJ}} \widehat{e}^i_{\mathrm{I}} \widehat{e}^j_{\mathrm{J}}, \tag{5.4.9}$$

$$\mathrm{I}, \mathrm{J} \in \{1,2\}, \qquad \epsilon^{12} \equiv 1. \tag{5.4.10}$$

Make sure you argue why this form is invariant under any orthogonal rotation of the basis $\widehat{e}_{\mathrm{I}} \to O_{\mathrm{IJ}} \widehat{e}_{\mathrm{J}}$, where $\delta_{\mathrm{MN}} O_{\mathrm{MI}} O_{\mathrm{NJ}} = \delta_{\mathrm{IJ}}$. □

Problem 5.24. Verify through a direct calculation that

$$\widehat{R}(\phi)\widehat{R}(\phi') = \widehat{R}(\phi + \phi'). \tag{5.4.11}$$

Draw a picture and explain what this result means. $\square$

Relation between SO$_2$ and U$_1$: Group Representations We now consider the following function of rotation matrices:

$$D\left(\widehat{R}(\phi)\right) \equiv e^{i\phi}. \tag{5.4.12}$$

In §(5.5) below, we shall identity rotation matrices with the group SO$_D$; and dub the set of $D \times D$ unitary matrices as U$_D$. Hence, the above is a map from SO$_2$ (2D rotations) to U$_1$ (1×1 unitary 'matrices').

Let us consider replacing each rotation matrix of the multiplication rule in eq. (5.4.11) with its corresponding $D(\widehat{R})$. The left hand side is

$$D(\widehat{R}(\phi))D(\widehat{R}(\phi')) = e^{i\phi}e^{i\phi'} \tag{5.4.13}$$

whereas the right hand side is simply

$$D(\widehat{R}(\phi + \phi')) = e^{i(\phi+\phi')}. \tag{5.4.14}$$

That is, we see that the group multiplication rule of eq. (5.4.11) is preserved:

$$D(\widehat{R}(\phi)\widehat{R}(\phi')) = D(\widehat{R}(\phi + \phi')). \tag{5.4.15}$$

This map from rotation matrices in 2D $\{\widehat{R}(\phi)\}$ to the unit circle on the complex plane $\{e^{i\phi}\}$ is related to the fact that, multiplication of one complex number by another is a rotation (plus a strech in the radial direction). Below, we shall build more involved *group representations* – functions of group elements that yield as output linear operators, in such a manner that the latter preserves the product rules of the original group (e.g., eq. (5.4.15)).

Effect on position eigenkets We will now consider position eigenstates $\{|\psi\rangle\,|0 \leq \psi < 2\pi\}$ on a circle of fixed radius r; and denote $D(\phi)$ to be the rotation operator that acts in the following manner:

$$D(\phi)\,|\psi\rangle = |\psi + \phi\rangle\,; \tag{5.4.16}$$

with the identity

$$\mathbb{I}\,|\psi\rangle \equiv D(\phi = 0)\,|\psi\rangle = |\psi\rangle\,. \tag{5.4.17}$$

We will assume the periodic boundary condition

$$|\psi + 2\pi\rangle = |\psi\rangle\,. \tag{5.4.18}$$

We will normalize these position eigenstates such that

$$\langle\phi|\,\phi'\rangle = \delta(\phi - \phi'). \tag{5.4.19}$$

Starting from the definition in eq. (5.4.16), we may first check that it obeys the same product rule as the rotation matrix in eq. (5.4.11).

$$D(\phi)D(\phi') \left|\psi\right\rangle = D(\phi) \left|\psi + \phi'\right\rangle = \left|\psi + \phi' + \phi\right\rangle \tag{5.4.20}$$

$$= D(\phi + \phi') \left|\psi\right\rangle \tag{5.4.21}$$

Since addition is associative and commutative, i.e., $\left|\psi + \phi\right\rangle = \left|\phi + \psi\right\rangle$ and $\left|\psi + \phi + \phi'\right\rangle = \left|\psi + \phi' + \phi\right\rangle$, we see that rotation on 2D is associative and commutative – in accordance to our intuition. To sum:

$$D(\phi)D(\phi') = D(\phi')D(\phi) = D(\phi + \phi'). \tag{5.4.22}$$

Unitary Acting on this abstract vector space of position eigenstates, the rotation operator in eq. (5.4.16) is unitary. To see this, we take the adjoint of $D(\phi) \left|\psi'\right\rangle = \left|\psi' + \phi\right\rangle$,

$$\left\langle\psi'\right| D(\phi)^\dagger = \left\langle\psi' + \phi\right|. \tag{5.4.23}$$

Combining equations (5.4.16) and (5.4.23) hands us

$$\left\langle \psi' \left| D(\phi)^\dagger D(\phi) \right| \psi \right\rangle = \left\langle \psi' + \phi \middle| \psi + \phi \right\rangle = \delta(\psi' - \psi) = \left\langle \psi' \middle| \psi \right\rangle. \tag{5.4.24}$$

But since $\left\langle\psi'\right|$ and $\left|\psi\right\rangle$ are arbitrary, we must have

$$D(\phi)^\dagger D(\phi) = \mathbb{I}. \tag{5.4.25}$$

Problem 5.25. Can you argue that

$$D(\phi)^\dagger = D(\phi)^{-1} = D(-\phi)? \tag{5.4.26}$$

Hint: We just proved the second equality. The third can be gotten by acting on an arbitrary state. □

Problem 5.26. Parity Parity is the operation where all the vectors in a given space is reversed in direction. If $\widehat{P}$ denotes the parity operator, we have

$$\widehat{P}\vec{v} = -\vec{v} \tag{5.4.27}$$

for arbitrary vector $\vec{v}$. In 2D, this parity operator is simply

$$\widehat{P} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{5.4.28}$$

What is the angle $\phi_{\rm P}$ in the rotation matrix of eq. (5.4.6) such that $\widehat{R}(\phi_{\rm P}) = \widehat{P}$? The existence of $\phi_{\rm P}$ tells us:

In 2D, the parity operation can be implemented as a rotation.

As we shall see below, this is *not* true in 3D – there is no rotation matrix that can implement parity simultaneously on all 3 axes. □

Exponential Form We have alluded to earlier that any operator, such as translation or rotations, that is continuously connected to the identity may be written as e^X. The exponent X would depend on the continuous parameter(s) concerned, such as the amount of translation or angle subtended by the rotation. For the 2D rotation case at hand, let us postulate

$$D(\phi) = e^{-i\phi J} \tag{5.4.29}$$

where the J is the *generator* of rotation (aka angular momentum operator). We may readily verify the multiplication rule in eq. (5.4.22),

$$D(\phi)D(\phi') = e^{-i\phi J}e^{-i\phi' J} = e^{-i(\phi+\phi')J} = D(\phi+\phi') \tag{5.4.30}$$

since $-i\phi J$ and $-i\phi' J$ commutes; i.e., $[-i\phi J, -i\phi' J] = 0$ and refer to eq. (5.1.46). Since we just proved that $D(\phi)$ is unitary (cf. eq. (5.4.25)) it must be that J is Hermitian. Firstly, note that

$$\left(e^X\right)^\dagger = \sum_{\ell=0}^{+\infty} \frac{(X^\ell)^\dagger}{\ell!} = \sum_{\ell=0}^{+\infty} \frac{(X^\dagger)^\ell}{\ell!} = e^{X^\dagger}. \tag{5.4.31}$$

Utilizing eq. (5.1.42), we see by Taylor expanding eq. (5.4.25) up to first order in ϕ bring us

$$\left(e^{-i\phi J}\right)^\dagger e^{-i\phi J} = e^{i\phi J^\dagger} e^{-i\phi J} = \left(\mathbb{I} + i\phi J^\dagger + \dots\right)(\mathbb{I} - i\phi J + \dots) \tag{5.4.32}$$

$$= \mathbb{I} + i\phi\left(J^\dagger - J\right) + \dots = \mathbb{I}. \tag{5.4.33}$$

The coefficient of ϕ must therefore vanish and we have

$$J^\dagger = J. \tag{5.4.34}$$

Problem 5.27. Pauli Matrices and 2D Rotations Express the rotation matrix $\widehat{R}(\phi)$ in eq. (5.4.6) in exponential form by identifying the appropriate matrix representation of the generator J. That is, what is $\widehat{J}$ such that the $\widehat{R}(\phi)$ of eq. (5.4.6) can be written as

$$\widehat{R}(\phi) = \exp(-i\phi\widehat{J})? \tag{5.4.35}$$

Hint: Use the Pauli matrices in eq. (3.2.18) and the result in eq. (3.2.26) to construct $\widehat{J}$. □

Problem 5.28. Taylor Expansion as Rotation To confirm eq. (5.4.29) is the right form of the rotation operator, argue, for an arbitrary state $|f\rangle$ and $|\phi\rangle$ a position eigenstate, that by denoting $f(\phi) \equiv \langle\phi| f\rangle$,

$$f(\phi - \phi') = e^{-\phi'\partial_\phi} f(\phi) = \left\langle \phi \left| e^{-i\phi' J} \right| f \right\rangle. \tag{5.4.36}$$

Can you also prove that

$$\langle \phi |J| f\rangle = -i\partial_\phi \langle\phi| f\rangle? \tag{5.4.37}$$

(Hint: Taylor expansion.) Compare these results to the one in eq. (5.2.24). □

Problem 5.29. Ket-Bra form of rotation operator Argue that an alternate representation for the rotation operator is

$$D(\phi) = \int_0^{2\pi} \mathrm{d}\varphi \, |\varphi + \phi\rangle \langle\varphi| \,. \tag{5.4.38}$$

Show that it is unitary. (Hint: Recall the normalization in eq. (5.4.19).) Explain why the rotation will no longer unitary if the position eigenstates in eq. (5.4.38) are restricted to a wedge $\{|\psi\rangle \, |0 \leq \psi \leq \phi_0 < 2\pi\}$ instead of a full circle. $\square$

'Orbital' Angular Momentum According to eq. (5.4.6), if we rotate the Cartesian coordinates $\vec{x} \equiv (x^1, x^2)$ via $\vec{x} \to \widehat{R}\vec{x}$, and employ the Taylor expansions $\cos\phi = 1-(1/2)\phi^2+\dots$ and $\sin\phi = \phi - (1/3!)\phi^3 + \dots$,

$$\widehat{R}(\phi) \cdot \vec{x} = \left(\mathbb{I}_{2\times 2} + \phi \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \mathcal{O}(\phi^2)\right) \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \tag{5.4.39}$$

$$= \exp\left(-\phi \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right) \cdot \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \tag{5.4.40}$$

$$= \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} + \phi \begin{bmatrix} -x^2 \\ x^1 \end{bmatrix} + \mathcal{O}(\phi^2). \tag{5.4.41}$$

We may re-write this by first defining

$$i\widehat{J} \equiv \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \doteq \epsilon^{ij}; \tag{5.4.42}$$

where in the second equality we have identified the 2D Levi-Civita symbol ϵ^{ij}, with $\epsilon^{12} \equiv 1$. Then eq. (5.4.39) reads

$$\widehat{R}(\phi)\vec{x} = \left(\mathbb{I} - i\phi\widehat{J} + \mathcal{O}(\phi^2)\right)\vec{x}, \tag{5.4.43}$$

$$\widehat{R}(\phi)^i{}_j x^j = \left(\delta^{ij} - \phi\epsilon^{ij} + \mathcal{O}(\phi^2)\right) x^j. \tag{5.4.44}$$

As we shall see below, that the $-i\widehat{J}$ is anti-symmetric – and therefore the 'generator' of rotations $\widehat{J}$ is Hermitian – is a feature that holds in general; not just in the 2D case here.

Next, we may consider the Taylor expansion resulting from the infinitesimal rotation carried out in eq. (5.4.41):

$$\langle\vec{x}|\, f\rangle \to \left\langle x^i - \phi\epsilon^{ij}x^j + \dots \middle|\, f\right\rangle$$

$$= \langle\vec{x}|\, f\rangle - \phi(-)\epsilon^{ji}x^j\partial_i \langle\vec{x}|\, f\rangle + \mathcal{O}(\phi^2) \tag{5.4.45}$$

$$= \langle\vec{x}|\, f\rangle + \phi(x^1\partial_2 - x^2\partial_1) \langle\vec{x}|\, f\rangle + \mathcal{O}(\phi^2). \tag{5.4.46}$$

On the other hand, we must have $|x^i - \phi\epsilon^{ij}x^j + \dots\rangle = \exp(-i\phi J)\,|\vec{x}\rangle$.

$$\left\langle x^i - \phi\epsilon^{ij}x^j + \dots \middle|\, f\right\rangle = \left(e^{-i\phi J}\,|\vec{x}\rangle\right)^\dagger |f\rangle = \langle\vec{x}|\,(1 + i\phi J + \dots)\,|f\rangle \tag{5.4.47}$$

$$= \langle\vec{x}|\, f\rangle + i\phi\,\langle\vec{x}\,|J|\, f\rangle + \dots. \tag{5.4.48}$$

Comparing equations (5.4.46) and (5.4.48) hands us the position representation of the generator of rotations – aka 'orbital' angular momentum – in 2D:

$$\langle \vec{x} | J | f \rangle = -i\epsilon^{ij} x^i \partial_j \langle \vec{x} | f \rangle = -i \left(x^1 \partial_2 - x^2 \partial_1 \right) \langle \vec{x} | f \rangle \tag{5.4.49}$$

$$= \left\langle \vec{x} \left| X^1 P_2 - X^2 P_1 \right| f \right\rangle ; \tag{5.4.50}$$

where, in the final equality, we have recalled eq. (5.2.24). Below, we will generalize the identification

$$J = X^1 P_2 - X^2 P_1 \tag{5.4.51}$$

to higher dimensions.

Problem 5.30. 'Flow' of J In differential geometry, §(9) below, we will learn that directional derivatives $v^i \partial_i$ may be viewed as tangent vectors spanning a vector space at a given point in space. Their geometric meaning is they 'generate' flow along the direction v^i, not unlike how J in eq. (5.4.29) generates rotations on the coordinates $\vec{x}$.

Verify that the first order term $-\phi\epsilon^{ij} x^j \partial_i f$ in eq. (5.4.45) is consistent with polar coordinates version of J in eq. (5.4.37). Explain your results in words. □

Eigenstates and Topology Since J is Hermitian, we are guaranteed its eigenstates form a complete basis $\{|m\rangle\}$. Let us now witness how the choice of boundary conditions in eq. (5.4.18) will allow us to fix the eigenvalues. Consider, for $|\psi\rangle$ some position eigenbra and $|m\rangle$ an eigenstate of J,

$$\left\langle \psi \left| e^{i(2\pi)J} \right| m \right\rangle = e^{i(2\pi)m} \langle \psi | m \rangle \tag{5.4.52}$$

$$= \left(e^{-i(2\pi)J} |\psi\rangle \right)^\dagger |m\rangle = \langle \psi + 2\pi | m \rangle = \langle \psi | m \rangle . \tag{5.4.53}$$

Comparing the rightmost terms on the first and second line,

$$e^{i(2\pi)m} = 1 \qquad \Leftrightarrow \qquad m = 0, \pm 1, \pm 2, \pm 3, \dots . \tag{5.4.54}$$

Choosing eq. (5.4.18) as our boundary condition implies any (bosonic) function $f(\phi) \equiv \langle \phi | f \rangle$ is periodic on a circle. It may be 'intuitively obvious' but is not actually always the case: fermionic states describing fundamental matter – electrons, muons, taus, quarks, etc. – in fact obey instead

$$|\psi + 2\pi\rangle = - |\psi\rangle . \tag{5.4.55}$$

In such a case, we have, from the above analysis, $e^{+i(2\pi)m} \langle \psi | m \rangle = - \langle \psi | m \rangle$. In turn, we see the eigenvalues $\{m\}$ of J are now

$$m = \frac{1}{2} + n, \qquad n = 0, \pm 1, \pm 2, \pm 3, \dots \tag{5.4.56}$$

$$= \pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}, \dots . \tag{5.4.57}$$

Completeness We may construct the identity operator

$$\mathbb{I} = \int_0^{2\pi} \mathrm{d}\varphi \, |\varphi\rangle \langle \varphi | . \tag{5.4.58}$$

As a check, we recall eq. (5.4.19) and calculate

$$\langle \phi' | \mathbb{I} | \phi \rangle = \int_0^{2\pi} \mathrm{d}\varphi \, \langle \phi' | \varphi \rangle \langle \varphi | \phi \rangle \tag{5.4.59}$$

$$= \int_0^{2\pi} \mathrm{d}\varphi \delta(\phi' - \varphi)\delta(\varphi - \phi) = \delta(\phi' - \phi). \tag{5.4.60}$$

Problem 5.31. Change-of-basis Using eq. (5.4.37), can you show that

$$\langle \phi | m \rangle \propto e^{im\phi}? \tag{5.4.61}$$

If we agree to normalize the eigenstates as

$$\langle m | n \rangle = \delta^m_n \tag{5.4.62}$$

explain why, up to an overall phase factor,

$$\langle \phi | m \rangle = \frac{\exp(im\phi)}{\sqrt{2\pi}}. \tag{5.4.63}$$

□

If we assume the eigenstates of J, namely $\{|m\rangle\}$, are normalized to unity, we may also write

$$\mathbb{I} = \sum_{m=-\infty}^{+\infty} |m\rangle \langle m| \,. \tag{5.4.64}$$

For, we may check that we have the proper normalization; invoking eq. (5.4.63),

$$\int_0^{2\pi} \mathrm{d}\phi \, \langle \phi | \mathbb{I} | \phi' \rangle = 1 = \int_0^{2\pi} \mathrm{d}\phi \sum_{m=-\infty}^{+\infty} \langle \phi | m \rangle \langle m | \phi' \rangle \tag{5.4.65}$$

$$= \int_0^{2\pi} \mathrm{d}\phi \sum_{m=-\infty}^{+\infty} \frac{e^{im(\phi-\phi')}}{2\pi} = \sum_{m=-\infty}^{+\infty} \delta^0_m. \tag{5.4.66}$$

Observe that eq. (5.4.64) is essentially the 1D version of (5.3.29). This is not a coincidence: the circle $\phi \in [0, 2\pi)$ can, of course, be thought of as a periodic space with period 2π.

Fourier Series Any state $|f\rangle$ may be decomposed into modes by inserting the completeness relation in eq. (5.4.64):

$$|f\rangle = \sum_{m=-\infty}^{+\infty} |m\rangle \langle m | f \rangle \,. \tag{5.4.67}$$

Multiplying both sides by a position eigenbra $\langle \phi |$ and using employing eq. (5.4.63), we obtain the Fourier series expansion

$$\langle \phi | f \rangle = \sum_{m=-\infty}^{+\infty} \frac{e^{im\phi}}{\sqrt{2\pi}} \langle m | f \rangle \,. \tag{5.4.68}$$

Rotation Operator At this point, recalling (4.3.65) and the eigenvector equations

$$J|m\rangle = m|m\rangle \qquad \text{and} \qquad e^{-i\phi J}|m\rangle = e^{-im\phi}|m\rangle \tag{5.4.69}$$

hands us the following representation of the rotation operator:

$$D(\phi) = \sum_{m=-\infty}^{+\infty} e^{-im\phi}|m\rangle\langle m| \tag{5.4.70}$$

$$= |0\rangle\langle 0| + \sum_{m=1}^{+\infty}\left(e^{-im\phi}|m\rangle\langle m| + e^{im\phi}|-m\rangle\langle -m|\right). \tag{5.4.71}$$

This result is the 1D analog of eq. (5.3.37).

Problem 5.32. Recover eq. (5.4.71) from eq. (5.4.38). Hint: Employ equations (5.4.63) and (5.4.64). □

Problem 5.33. If we view the rotation matrix in eq. (5.4.6) as the matrix element of some operator,

$$\langle i|R|j\rangle = \widehat{R}^i{}_j, \qquad i,j \in \{1,2\}; \tag{5.4.72}$$

show that this R is in fact related to the $m=1$ term in eq. (5.4.71) via a change-of-basis. In other words, the 2D rotation in real space is a 'sub-operator' of the $D(\phi)$ of this section. Hint: Consider the subspace spanned by the two states $|\pm\rangle \equiv (\pm i/\sqrt{2})|\phi=0\rangle + (1/\sqrt{2})|\phi=\pi/2\rangle$. □

Problem 5.34. Rotating Momentum Show that, if $\left|\vec{k}\right\rangle$ is a momentum eigenket,

$$D(\phi)\left|\vec{k}\right\rangle = \left|\widehat{R}(\phi)\cdot\vec{k}\right\rangle. \tag{5.4.73}$$

(Hint: Insert a complete set of position eigenstates.) If $|\vec{x}\rangle$ is the position eigenket, argue that $D(2\pi)|\vec{x}\rangle = |\vec{x}\rangle$ iff $D(2\pi)|\vec{k}\rangle = |\vec{k}\rangle$. □

Invariant subspaces We close this section by making the observation that, due to the abelian (or, commutative) nature of 2D rotations in eq. (5.4.22),

$$D(\varphi)^\dagger D(\phi) D(\varphi) = D(\varphi)^\dagger D(\varphi) D(\phi) = D(\phi) \tag{5.4.74}$$

since $D(\varphi)^\dagger D(\varphi) = \mathbb{I}$ (cf. (5.4.25)). Recall from the discussion in §(4.3.3) that $U^\dagger D(\phi) U$, for any unitary U, may be regarded as $D(\phi)$ but written in a different basis. Eq. (5.4.74) informs us that the 2D rotation operator in fact remains invariant under all change-of-basis transformations.

5.4.1 Including Spatial Translations: Euclidean Group $\mathbb{E}_2$

We now consider combining both rotations and translations in 2D (flat) space. In Cartesian components, this amounts to replacing the position vector $\vec{x}$ as follows:

$$\vec{x} \to \widehat{R}(\phi)\cdot\vec{x} + \vec{a}, \tag{5.4.75}$$

$$x^i \to \widehat{R}(\phi)^i{}_j x^j + a^i. \tag{5.4.76}$$

The $\widehat{R}$ is the rotation matrix and $\vec{a}$ is the constant displacement vector. We may in fact package this transformation rule using a 3×3 matrix Π by first promoting $\vec{x}$ to a 3-component object with 1 as its 3rd entry:

$$x^{\mathrm{A}} \equiv (x^i, 1)^{\mathrm{T}} = (\vec{x}, 1)^{\mathrm{T}}; \tag{5.4.77}$$

followed by defining

$$\widehat{\Pi}(\phi, \vec{a}) \equiv \left[\begin{array}{ccc} \cos\phi & -\sin\phi & a^1 \\ \sin\phi & \cos\phi & a^2 \\ 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{cc} \widehat{R}(\phi)^i{}_j & a^i \\ 0 \quad\ 0 & 1 \end{array}\right]. \tag{5.4.78}$$

By expanding in small angle ϕ and displacement $\vec{a}$,

$$\widehat{\Pi}(\phi, \vec{a}) = \mathbb{I}_{3\times 3} - i\phi\widehat{J} - ia^j\widehat{P}_j + \dots, \tag{5.4.79}$$

we may read off from eq. (5.4.78) that the 3×3 matrix representation of the rotation and translation generators in 2D space are

$$\widehat{J} = i\left[\begin{array}{ccc} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \qquad \text{and} \qquad \widehat{P}_a = i\left[\begin{array}{ccc} 0 & 0 & \delta^1{}_a \\ 0 & 0 & \delta^2{}_a \\ 0 & 0 & 0 \end{array}\right]. \tag{5.4.80}$$

The action in eq. (5.4.75) may then be read off the top two components of

$$x^{\mathrm{A}} \to \Pi(\phi, \vec{a})^{\mathrm{A}}{}_{\mathrm{B}} x^{\mathrm{B}}; \tag{5.4.81}$$

$$(\vec{x}, 1)^{\mathrm{T}} \to \left(\widehat{R} \cdot \vec{x} + \vec{a}, 1\right)^{\mathrm{T}}, \tag{5.4.82}$$

with the indices A and B running over $\{1, 2, 3\}$.

Problem 5.35. Why does the set $\{(\vec{x}, 1)^{\mathrm{T}}\}$ *not* form a vector space? Is $\Pi(\phi, \vec{a})$ unitary? Orthogonal? □

Problem 5.36. 2D Euclidean Group Verify that

$$\Pi(0, \vec{0}) = \mathbb{I}_{3\times 3}, \tag{5.4.83}$$

$$\Pi(\phi, \vec{a})^{-1} = \Pi\left(-\phi, -\widehat{R}(-\phi) \cdot \vec{a}\right), \tag{5.4.84}$$

$$\Pi(\phi_1, \vec{a}_1)\Pi(\phi_2, \vec{a}_2) = \Pi\left(\phi_1 + \phi_2, \widehat{R}(\phi_1) \cdot \vec{a}_2 + \vec{a}_1\right). \tag{5.4.85}$$

Explain why the set of $\{\Pi\}$ forms a group. The collection of rotations and translations is known as the Euclidean group $\mathbb{E}_D$, where $D \geq 2$ is the dimension of space. □

Note that, for a general $\widehat{\Pi}(\phi, \vec{a})$, a zero rotation angle corresponds to a pure displacement by $\vec{a}$; whereas a zero displacement $\vec{a} = \vec{0}$ corresponds to a pure rotation by ϕ. Hence,

$$\Pi(\phi = 0, \vec{a}) = \mathcal{T}(\vec{a}) \qquad \text{and} \qquad \Pi(\phi, \vec{a} = \vec{0}) = D(\phi). \tag{5.4.86}$$

Let us also recognize the most general 2D Euclidean transformation to be a rotation $\widehat{R}(\phi)$ followed by a translation $\vec{a}$; i.e., $\vec{x} \to \widehat{R} \cdot \vec{x} \to \widehat{R} \cdot \vec{x} + \vec{a}$. Or, from eq. (5.4.85):

$$\Pi(0, \vec{a})\Pi(\phi, \vec{0}) = \Pi(\phi, \vec{a}) = \mathcal{T}(\vec{a}) \cdot D(\phi) = \exp\left(-i\vec{a} \cdot \vec{P}\right) \exp\left(-i\phi \cdot J\right). \tag{5.4.87}$$

From this, we may observe that

$$\Pi(\phi, \vec{a})^{-1} = D(-\phi) \cdot \mathcal{T}(-\vec{a}). \tag{5.4.88}$$

Moreover, from eq. (5.4.85),

$$\begin{aligned} D(\phi)\mathcal{T}(\vec{a})D(\phi)^\dagger &= \Pi(\phi, \vec{0}) \cdot \Pi(-\phi, \vec{a}) \\ &= \Pi\left(0, \widehat{R}(\phi) \cdot \vec{a}\right) = \mathcal{T}\left(\widehat{R}(\phi) \cdot \vec{a}\right). \end{aligned} \tag{5.4.89}$$

On the other hand, the same computation may be carried out via

$$D(\phi)\mathcal{T}(\vec{a})D(\phi)^\dagger = D(\phi)e^{-i\vec{a}\cdot\vec{P}}D(\phi)^\dagger \tag{5.4.90}$$

$$= \exp\left(-i\vec{a} \cdot \left(D(\phi)\vec{P}D(\phi)^\dagger\right)\right). \tag{5.4.91}$$

Comparing the two routes leads us to the recognition that

$$D(\phi)(\vec{a} \cdot \vec{P})D(\phi)^\dagger = (\widehat{R}(\phi) \cdot \vec{a}) \cdot \vec{P} \tag{5.4.92}$$

$$D(\phi)P_i D(\phi)^\dagger = P_j \widehat{R}(\phi)^j{}_i. \tag{5.4.93}$$

In other words: the generator of translations $\vec{P}$ transforms as a vector under rotations.

Previously, we have derived in the position representation that P_a may be identified with the derivative operator $-i\partial_{x^a}$ in Cartesian coordinates. Let us now invoke eq. (5.4.49) to compute the commutators of J and P_a in the position representation:

$$\langle \vec{x} \left| [J, P_a] \right| f \rangle = (-i)^2 \epsilon^{ij} x^i \partial_j \partial_a f(\vec{x}) - (-i)^2 \partial_a \left(\epsilon^{ij} x^i \partial_j f(\vec{x})\right) \tag{5.4.94}$$

$$= +\epsilon^{ij} \delta^i_a \partial_j f = i(-i)\epsilon^{ab}\partial_b f \tag{5.4.95}$$

$$= i\epsilon^{ab} \langle \vec{x} \left| P_b \right| f \rangle. \tag{5.4.96}$$

Since $\langle \vec{x}|$ was arbitrary, we may now record the Lie algebra of $\mathbb{E}_2$:

$$[J, P_a] = i\epsilon^{ab} P_b \qquad \text{and} \qquad [P_a, P_b] = 0. \tag{5.4.97}$$

The anti-symmetric character of ϵ^{ab} also indicates

$$\left[J, \vec{P}^2\right] = [J, P_i] P_i + P_i [J, P_i] = 0. \tag{5.4.98}$$

Problem 5.37. Rotating Momentum: Infinitesimal Version Work out eq. (5.4.93) to first order in ϕ and verify that it reproduces the commutation relation (on the left) in eq. (5.4.97). Since the Lie algebra determines group multiplication for all elements continuously connected to the identity, both eq. (5.4.93) and the infinitesimal version in eq. (5.4.97) are equivalent. More generally, we would say that an operator V^i (in a Cartesian basis) transforms as a vector iff $[J, V^i] = i\epsilon^{ij} V^j$. $\square$

Problem 5.38. Rotational Invariance of $\vec{P}^2$ Eq. (5.4.98) is, in fact, a statement of rotational invariance of $\vec{P}^2$. Can you explain why? Hint: Relate it to the change-of-basis $D(\phi)\vec{P}^2 D(\phi)^\dagger$. $\square$

Problem 5.39. Verify the commutation relation between J and $\vec{P}$ using (1) the matrix representation in eq. (5.4.80); and (2) the relation in eq. (5.4.93). $\square$

Problem 5.40. One may be tempted to write

$$\Pi(\phi, \vec{a}) = \exp\left(-i\vec{a} \cdot \vec{P} - i\phi J\right). \tag{5.4.99}$$

This cannot hold because of eq. (5.4.87) and the fact that $[J, P_a] \neq 0$.

$$\Pi(\phi, \vec{a}) = \exp\left(-i\vec{a} \cdot \vec{P}\right) \exp\left(-i\phi J\right) \neq \exp\left(-i\vec{a} \cdot \vec{P} - i\phi J\right). \tag{5.4.100}$$

By exploiting eq. (5.1.46), show that

$$\mathcal{T}(\vec{a}) D(\phi) \mathcal{T}(\vec{a})^\dagger = \exp\left[-i\phi J + i\phi\epsilon^{mn} a^m P_n\right]. \tag{5.4.101}$$

Why does this imply

$$\exp\left[-i\phi J + i\phi\epsilon^{mn} a^m P_n\right] = \exp\left(-i\left(\vec{a} - \widehat{R}(\phi) \cdot \vec{a}\right) \cdot \vec{P}\right) \exp\left(-i\phi J\right)? \tag{5.4.102}$$

Use this to deduce, the general $\mathbb{E}_2$ group element is in fact

$$\Pi\left(\phi, \vec{a}\right) = e^{-i\vec{a}\cdot\vec{P}} e^{-i\phi J} = \exp\left(-\frac{i\phi}{2} \begin{bmatrix} \cot(\phi/2) & 1 \\ -1 & \cot(\phi/2) \end{bmatrix}^i{}_j a^j P_i - i\phi J\right). \tag{5.4.103}$$

Does this result recover $\Pi\left(\phi = 0, \vec{a}\right) = \mathcal{T}(\vec{a})$? $\square$

Compatible Observables The Lie algebra of $\mathbb{E}_2$ tells us not all three generators are mutually compatible. If J is excluded, we have already seen that picking the eigenstates of $\{P_1, P_2\}$ yields a basis that complements the position eigenkets $\{|\vec{x}\rangle\}$. The eigenbasis of $\vec{P}$ is of course the momentum eigenkets obeying

$$P_i \left|\vec{k}\right\rangle = k_i \left|\vec{k}\right\rangle. \tag{5.4.104}$$

On the other hand, if we *do* include the rotation generator J in the set of compatible observables, then neither P_1 nor P_2 may be included; only $\{J, \vec{P}^2\}$ are compatible. Their simultaneous eigenket is $\{|m, k\rangle\}$, where

$$J\left|m, k\right\rangle = m\left|m, k\right\rangle \qquad \text{and} \qquad \vec{P}^2\left|m, k\right\rangle = k^2\left|m, k\right\rangle; \tag{5.4.105}$$

and m is an integer if we demand periodic boundary conditions and half integer if we impose anti-periodic ones.

$$D(\phi = 2\pi)\,|m,k\rangle = e^{-i(2\pi)m}\,|m,k\rangle = \pm\,|m,k\rangle \tag{5.4.106}$$

For the rest of this section, we will assume $m \in \mathbb{Z}$.

Problem 5.41. Laplacian in Polar Coordinates If $\vec{x} = r(\cos\phi, \sin\phi)$, verify that

$$\begin{bmatrix} \partial_{x^1} \\ \partial_{x^2} \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \partial_r \\ \frac{1}{r}\partial_\phi \end{bmatrix}. \tag{5.4.107}$$

From this, show that

$$-\left\langle \vec{x} \left| \vec{P}^2 \right| f \right\rangle = \vec{\nabla}^2 \langle \vec{x}|\, f\rangle \equiv \vec{\nabla}^2 f(\vec{x}) \tag{5.4.108}$$

$$\equiv (\partial_{x^1}^2 + \partial_{x^2}^2) f(\vec{x}) \tag{5.4.109}$$

$$= \frac{1}{r}\partial_r\left(r\partial_r f\right) + \frac{1}{r^2}\partial_\phi^2 f. \tag{5.4.110}$$

for any arbitrary state $|f\rangle$. Hint: First start from $(\partial_r, \partial_\phi)$, but relate them to their Cartesian counterparts. $\square$

Change-Of-Basis and the Hankel Transform Previously, we discovered the change-of-basis from momentum to position space eigenkets gave us the plane wave

$$\langle \vec{x}|\vec{k}\rangle = e^{i\vec{k}\cdot\vec{x}} \tag{5.4.111}$$

and led us to the Fourier transform. Let us now examine the change of basis between the position, momentum, and the $\{|m,k\rangle\}$ basis. Firstly, we recognize that, for

$$\vec{x} = r(\cos\phi, \sin\phi), \tag{5.4.112}$$

the eigenequations for the compatible observables $\vec{P}^2$ and J read

$$k^2 \langle \vec{x}|\, m,k\rangle = \left\langle \vec{x} \left| \vec{P}^2 \right| m,k \right\rangle = -\vec{\nabla}^2 \langle \vec{x}|\, m,k\rangle \tag{5.4.113}$$

$$m \langle \vec{x}|\, m,k\rangle = \langle \vec{x}\,|J|\, m,k\rangle = -i\partial_\phi \langle \vec{x}|\, m,k\rangle\,. \tag{5.4.114}$$

The first equation is also obeyed by the plane wave; i.e., $\vec{\nabla}^2\langle\vec{x}|\vec{k}\rangle = -k^2\langle\vec{x}|\vec{k}\rangle$. This indicates there ought to be an intimate relation between $|\vec{k}\rangle$ and $|m,k\rangle$; in fact, we shall see that the latter is a Fourier mode of the former. On the other hand, the second line tells us immediately that

$$\langle r,\phi|\, m,k\rangle = \langle r|\, m,k\rangle\, e^{im\phi}, \tag{5.4.115}$$

where $\langle r|\, m,k\rangle$ is now shorthand for a function of r, m, and k and is not necessarily an inner product. Additionally, if we now employ eq. (5.4.110),

$$\frac{1}{r}\partial_r\left(r\partial_r \langle r|\, m,k\rangle\, e^{im\phi}\right) + \frac{1}{r^2}\partial_\phi^2\left(\langle r|\, m,k\rangle\, e^{im\phi}\right) = -k^2 \langle r|\, m,k\rangle\, e^{im\phi}. \tag{5.4.116}$$

Dividing throughout by $k^2 e^{im\phi}$ and moving everything to the left hand side, we obtain Bessel's equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \frac{1}{\xi}\frac{\mathrm{d}}{\mathrm{d}\xi} + 1 - \frac{m^2}{\xi^2}\right)\langle r|\, m, k\rangle = 0; \tag{5.4.117}$$

where we have defined $\xi \equiv kr$. (Even though m is integer here, note that this equation holds for the Bessel function even when m is complex.) A pair of linearly independent solutions are $\langle r|\, m, k\rangle = J_m(kr)$ and $\langle r|\, m, k\rangle = Y_m(kr)$, with the former given by the series[29]

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!\Gamma(\ell+\nu+1)}\left(\frac{z}{2}\right)^{2\ell}; \tag{5.4.118}$$

whereas for non-negative integer $m \in \{0, 1, 2, 3, \dots\}$,

$$Y_m(z) = \frac{2}{\pi}\ln\left(\frac{z}{2}\right) J_m(z) - \frac{1}{\pi\cdot(z/2)^m}\sum_{\ell=0}^{m-1}\frac{(m-\ell-1)!}{\ell!}\left(\frac{z}{2}\right)^{2\ell} \tag{5.4.119}$$

$$- \frac{(z/2)^m}{\pi}\sum_{\ell=0}^{\ell}\left(\frac{\Gamma'(\ell+1)}{\Gamma(\ell+1)} + \frac{\Gamma'(m+\ell+1)}{\Gamma(m+\ell+1)}\right)\frac{(-)^\ell (z/2)^{2\ell}}{\ell!(m+\ell)!}. \tag{5.4.120}$$

To obtain $J_{-m}(z)$ and $Y_{-m}(z)$ – namely, negative integer indices $\{-m\}$ – we record the formulas

$$J_{-m}(z) = (-)^m J_m(z) \qquad \text{and} \qquad Y_{-m}(z) = (-)^m Y_m(z). \tag{5.4.121}$$

For small $kr \ll 1$, $Y_m(kr)$ therefore diverges as $\ln(kr)$ for $m = 0$ and as $1/(kr)^{|m|}$ for non-zero integers $\{m\}$. Whereas, $J_m(kr)$ for small kr goes as $(kr)^m$. Since $r = 0$ is not a distinguished point, and since we are transforming from one complete basis to another, we expect $\langle r = 0, \phi|\, m, k\rangle$ to remain finite for any choice of origin. Therefore, up to an overall constant χ_m,

$$\langle r, \phi|\, m, k\rangle = \chi_m J_m(kr) e^{im\phi}. \tag{5.4.122}$$

Remember that the infinitesimal volume in 2D polar coordinates (r, ϕ) is given by $\mathrm{d}r(r\mathrm{d}\phi)$. Hence, the coordinate representation of the operator identity must be

$$\langle \vec{x}\,|\mathbb{I}|\,\vec{x}'\rangle = \langle r, \phi\,|\mathbb{I}|\, r', \phi'\rangle = \frac{\delta(r-r')}{\sqrt{r\cdot r'}}\delta(\phi-\phi'); \tag{5.4.123}$$

so that the volume integral with respect to (r, ϕ) or (r', ϕ') is one. Likewise, in 2D momentum space, we may define

$$\vec{k} \equiv k\,(\cos\varphi, \sin\varphi)\,; \tag{5.4.124}$$

[29]In equations (5.4.118) and (5.4.119) below, the $\Gamma(z)$ is known as the Gamma function; which is the generalization of the factorial. In particular, $\Gamma(\ell+1) = \ell!$ whenever ℓ is a non-negative integer. We will study its properties in §(6) below.

and deduce its associated identity operator to be

$$\left\langle \vec{k} \left| \mathbb{I} \right| \vec{k}' \right\rangle = \left\langle k, \varphi \left| \mathbb{I} \right| k', \varphi' \right\rangle = (2\pi)^2 \frac{\delta(k-k')}{\sqrt{k \cdot k'}} \delta(\varphi - \varphi'). \tag{5.4.125}$$

For instance, we must have – via completeness –

$$\int_{\mathbb{R}^2} \mathrm{d}^2\vec{x}' \left\langle \vec{k} \middle| \vec{x}' \right\rangle \left\langle \vec{x}' \middle| \vec{k}' \right\rangle = (2\pi)^2 \frac{\delta(k-k')}{\sqrt{k \cdot k'}} \delta(\varphi - \varphi'). \tag{5.4.126}$$

Let us now turn to computing

$$\begin{aligned}
\langle m, k | m', k' \rangle &= \int_{\mathbb{R}^2} \mathrm{d}^2\vec{x}' \left\langle m, k \middle| \vec{x}' \right\rangle \left\langle \vec{x}' \middle| m', k' \right\rangle &(5.4.127)\\
&= \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \mathrm{d}r' r' e^{i(m'-m)\phi'} \chi_m^* \chi_{m'} J_m(kr') J_{m'}(k'r') &(5.4.128)\\
&= 2\pi \cdot |\chi_m|^2 \delta^m_{m'} \int_0^\infty \mathrm{d}r' r' J_m(kr') J_{m'}(k'r'). &(5.4.129)
\end{aligned}$$

We will show below that, for $r, r', k, k' > 0$ and $\nu > -1$,

$$\begin{aligned}
\int_0^\infty \mathrm{d}r' r' J_\nu(kr') J_\nu(k'r') &= \frac{\delta(k-k')}{\sqrt{k \cdot k'}}, &(5.4.130)\\
\int_0^\infty \mathrm{d}k' k' J_\nu(k'r) J_\nu(k'r') &= \frac{\delta(r-r')}{\sqrt{r \cdot r'}}. &(5.4.131)
\end{aligned}$$

This leads us to choose $\chi_m \equiv 1$ and therefore

$$\langle r, \phi | m, k \rangle = J_m(kr) e^{im\phi}, \tag{5.4.132}$$

so that

$$\langle m, k | m', k' \rangle = 2\pi \cdot \delta^m{}_{m'} \frac{\delta(k-k')}{\sqrt{k \cdot k'}}. \tag{5.4.133}$$

Furthermore,

$$\begin{aligned}
\sum_{m'=-\infty}^{+\infty} \int_0^{+\infty} \frac{\mathrm{d}k' k'}{2\pi} \left\langle r, \phi \middle| m', k' \right\rangle \left\langle m', k' \middle| r', \phi' \right\rangle &= \sum_{m'=-\infty}^{+\infty} \frac{e^{im'(\phi-\phi')}}{2\pi} \int_0^{+\infty} \mathrm{d}k' k' \cdot J_m(k'r) J_m(k'r') \\
&= \delta(\phi - \phi') \frac{\delta(r-r')}{\sqrt{r \cdot r'}} = \left\langle \vec{x} \middle| \vec{x}' \right\rangle. \qquad (5.4.134)
\end{aligned}$$

Since the position eigenstates are arbitrary, this teaches us that the identity written in momentum space polar coordinates is

$$\mathbb{I} = \sum_{m'=-\infty}^{+\infty} \int_0^{+\infty} \frac{\mathrm{d}k' k'}{2\pi} \left| m', k' \right\rangle \left\langle m', k' \right|. \tag{5.4.135}$$

Hankel Transform Let us observe that eq. (5.4.130) and (5.4.131) imply, for some fixed ν, the $\{J_\nu(kr)\}$ may be regarded as a complete set of functions on the positive real line $r \in \mathbb{R}^+$. This forms the underpinnings of the *Hankel transform*; which in turn is, loosely speaking, the 'radial component' of the 2D Fourier transform when $\nu = m$ is an integer. If, for an arbitrary function f, we define

$$\widetilde{f}_\nu(k) \equiv \int_0^\infty \mathrm{d}r' \cdot r' \cdot J_\nu(kr') f(r'); \tag{5.4.136}$$

then f itself may be recovered by multiplying both sides by $kJ_\nu(kr)$ and integrating over $k \in [0, \infty)$ to obtain

$$f(r) = \int_0^\infty \mathrm{d}k' \cdot k' \cdot J_\nu(k'r) \widetilde{f}_\nu(k). \tag{5.4.137}$$

$\{\vec{P}\}$ versus $\{\vec{P}^2, J\}$ eigenbasis Next, we turn to examine the change-of-basis

$$\begin{aligned}
\left\langle \vec{k}' \middle| m, k \right\rangle = \langle k', \varphi' | m, k\rangle &= \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \mathrm{d}r' r' \left\langle \vec{k}' \middle| r', \phi' \right\rangle \langle r', \phi' | m, k\rangle & (5.4.138)\\
&= \int_0^{2\pi} \mathrm{d}\phi' \int_0^\infty \mathrm{d}r' r' e^{-ik'r'\cos(\varphi' - \phi')} J_m(kr') e^{im\phi'} & (5.4.139)\\
&= \int_0^\infty \mathrm{d}r' r' \left(\int_0^{2\pi} \mathrm{d}\theta e^{-ik'r'\cos\theta} e^{im\theta} \right) J_m(kr') e^{im\varphi'}. & (5.4.140)
\end{aligned}$$

The Bessel function $J_n(z)$ for integer $n = 0, \pm 1, \pm 2, \dots,$ admits the integral representations

$$J_n(z) = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi} e^{iz\sin(\theta) - in\theta} = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi i^n} e^{iz\cos(\theta) - in\theta}. \tag{5.4.141}$$

Since the integration domain runs over a full period of the integrand, we may replace the limits $\int_0^{2\pi}$ with $\int_{-\pi}^{\pi}$ and recognize $J_n(z)$ is real whenever z is real. For,

$$\begin{aligned}
J_n(z)^* &= \int_{-\pi}^{+\pi} \frac{\mathrm{d}\theta}{2\pi} e^{-iz^*\sin(\theta) + in\theta} & (5.4.142)\\
&= -\int_{-\theta = +\pi}^{-\theta = -\pi} \frac{\mathrm{d}(-\theta)}{2\pi} e^{iz^*\sin(-\theta) - in(-\theta)} & (5.4.143)\\
&= J_n(z^*). & (5.4.144)
\end{aligned}$$

Another way to arrive at the same conclusion is to derive its series representation, which we shall do so in eq. (5.4.118) below. In any case, this result now informs us, the change-of-basis is therefore

$$\left\langle \vec{k}' \middle| m, k \right\rangle = \langle k', \varphi' | m, k\rangle = 2\pi(-i)^m \frac{\delta(k - k')}{\sqrt{kk'}} e^{im\varphi'}. \tag{5.4.145}$$

As already alluded to, this tells us $|m, k\rangle$ is a single Fourier mode of $|\vec{k} = k(\cos\varphi, \sin\varphi)\rangle$.

$$|\vec{k}(k, \varphi)\rangle = \sum_{m'=-\infty}^{+\infty} \int_0^\infty \frac{\mathrm{d}k' \cdot k'}{2\pi} |m', k'\rangle \langle m', k' | \vec{k}\rangle \tag{5.4.146}$$

$$= \sum_{m=-\infty}^{+\infty} i^m e^{-im\varphi} |m, k\rangle \tag{5.4.147}$$

At this point, we may summarize: For any state $|f\rangle$,

$$f(r, \phi) \equiv \langle r, \phi | f\rangle = \int_0^{2\pi} \frac{\mathrm{d}\varphi}{2\pi} \int_0^\infty \frac{\mathrm{d}k \cdot k}{2\pi} e^{ikr\cos(\varphi-\phi)} \langle k, \varphi | f\rangle \tag{5.4.148}$$

$$\widetilde{f}(k, \varphi) = \langle k, \varphi | f\rangle = \int_0^{2\pi} \mathrm{d}\phi \int_0^\infty \mathrm{d}r \cdot r \cdot e^{-ikr\cos(\varphi-\phi)} \langle r, \phi | f\rangle \, ; \tag{5.4.149}$$

and

$$f(r, \phi) \equiv \langle r, \phi | f\rangle = \sum_{m'=-\infty}^{+\infty} \int_0^\infty \frac{\mathrm{d}k \cdot k}{2\pi} J_m(kr) e^{im\phi} \langle m, k | f\rangle \tag{5.4.150}$$

$$\widetilde{f}_m(k) \equiv \langle m, k | f\rangle = \int_0^{2\pi} \mathrm{d}\phi \int_0^\infty \mathrm{d}r \cdot r \cdot J_m(kr) e^{-im\phi} \langle r, \phi | f\rangle \, . \tag{5.4.151}$$

Problem 5.42. Verify that the integral representation of $J_n(z)$ in eq. (5.4.141) in fact satisfies Bessel's equation: $J_n''(z) + (1/z)J_n'(z) + (1 - (n/z)^2)J_n(z) = 0$. □

Problem 5.43. Bessel Function Identity If n is an integer, show from the integral representation in eq. (5.4.141) that $J_{-n}(z) = (-)^n J_n(z)$. □

Problem 5.44. Bessel J_n as Fourier Series Coefficient Show that

$$\langle \vec{x} | \vec{k}\rangle = e^{i\vec{k}\cdot\vec{x}} = \sum_{\substack{n=-\infty \\ n\in\mathbb{Z}}}^{+\infty} i^n J_n(kr) e^{in\varphi}; \tag{5.4.152}$$

where $k \equiv |\vec{k}|$, $r \equiv |\vec{x}|$, and

$$\cos\varphi \equiv (\vec{k} \cdot \vec{x})/(kr). \tag{5.4.153}$$

Why does this justify the statement that the Hankel transform is the 'radial component' of the Fourier transform in 2D? □

Problem 5.45. Spherical Bessel Functions For non-negative integer $n = 0, 1, 2, \dots$, the spherical Bessel function is defined in terms of its Bessel J_ν counterpart as

$$j_n(z) \equiv \sqrt{\frac{\pi}{2z}} J_{n+\frac{1}{2}}(z). \tag{5.4.154}$$

From the completeness relations in equations (5.4.130) and (5.4.131), derive the following completeness relations:

$$\int_0^\infty \mathrm{d}r' r'^2 j_n(kr') j_n(k'r') = \frac{\pi}{2} \cdot \frac{\delta(k-k')}{k \cdot k'}, \tag{5.4.155}$$

$$\int_0^\infty \mathrm{d}k' k'^2 j_n(k'r) j_n(k'r') = \frac{\pi}{2} \cdot \frac{\delta(r-r')}{r \cdot r'}. \tag{5.4.156}$$

These relations will be useful in the discussion of spherical symmetry in 3D space. □

Matrix Elements of $\Pi(\phi, \vec{a})$ Let us compute the matrix elements of $\mathcal{T}(\vec{a})$ and $\Pi(\phi, \vec{a})$ in the $\{|m, k\rangle\}$ eigenbasis. By inserting a complete set of momentum eigenstates,

$$
\begin{aligned}
&\langle m', k' \left| \mathcal{T}(\vec{a}(\theta)) \right| m, k\rangle \\
&= \int_0^{2\pi} \frac{\mathrm{d}\varphi''}{2\pi} \int_0^{\infty} \frac{\mathrm{d}k'' \cdot k''}{2\pi} \left\langle m', k' \left| e^{-i\vec{a}\cdot\vec{P}} \right| k'', \varphi'' \right\rangle \langle k'', \varphi'' | \, m, k\rangle \\
&= \int_0^{2\pi} \frac{\mathrm{d}\varphi''}{2\pi} \int_0^{\infty} \frac{\mathrm{d}k'' \cdot k''}{2\pi} e^{-ik'' a \cos(\varphi''-\theta)} (2\pi)^2 i^{m'} \frac{\delta(k'-k'')}{\sqrt{k' k''}} e^{-im'\varphi''} (-i)^m \frac{\delta(k''-k)}{\sqrt{k'' k}} e^{im\varphi''} \\
&= 2\pi(-)^{m-m'} \int_0^{2\pi} \frac{\mathrm{d}\varphi''}{2\pi(-i)^{m-m'}} e^{-ik'' a \cos(\varphi''-\theta)} e^{i(m-m')\varphi''} \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \\
&= 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot J_{m'-m}(ka) e^{-i(m'-m)\theta}.
\end{aligned}
\tag{5.4.157}
$$

Raising and Lowering Operators Alternatively, we shall now witness that raising and lowering operators may be employed to arrive at a series solution to these matrix elements. If we define

$$
P_\pm \equiv P_1 \pm i P_2, \tag{5.4.158}
$$

we may deduce from eq. (5.4.97) and linearity of the commutator that $[J, P_\pm] = i(\epsilon^{12} P_2 \pm i \epsilon^{21} P_1)$.

$$
[J, P_\pm] = \pm P_\pm \tag{5.4.159}
$$

Now, if $|m, k\rangle$ is the simultaneous eigenket of J and $\vec{P}^2$, we may consider

$$
J \left(P_\pm |m, k\rangle\right) = \left(J P_\pm - P_\pm J + P_\pm J\right) |m\rangle = \left([J, P_\pm] + P_\pm J\right) |m, k\rangle \tag{5.4.160}
$$

$$
= (m \pm 1) P_\pm |m, k\rangle . \tag{5.4.161}
$$

Moreover,

$$
(P_\pm)^\dagger P_\pm = (P_1 \mp i P_2)(P_1 \pm i P_2) = \vec{P}^2, \tag{5.4.162}
$$

which means

$$
\left(P_\pm |m, k\rangle\right)^\dagger \left(P_\pm |m, k\rangle\right) = k^2 \langle m, k | \, m, k\rangle . \tag{5.4.163}
$$

If we assume these $\{|m, k\rangle\}$ are normalized to unity, then we shall verify below that the consistent choice of phases is provided by the relation

$$
P_\pm |m, k\rangle = \pm i k |m \pm 1, k\rangle . \tag{5.4.164}
$$

Returning to the matrix element of $\Pi(\phi, \vec{a})$, we first employ eq. (5.4.93) to express the translation along $\vec{a} \equiv a \cdot \widehat{a}(\theta)$, with $\widehat{a}^2 = 1$, as one rotated from the 1−axis:

$$
\mathcal{T}\left(\vec{a}(\theta)\right) = D(\theta) \mathcal{T}\left(a \widehat{e}_1\right) D(\theta)^\dagger. \tag{5.4.165}
$$

Then, exploiting this decomposition, and denoting $a \equiv |\vec{a}|$,

$$\langle m', k | \mathcal{T}(\vec{a}) | m, k \rangle = \left\langle m', k \left| D(\theta) \exp\left[-iaP_1\right] D(\theta)^\dagger \right| m, k \right\rangle \tag{5.4.166}$$
$$= e^{i(m-m')\theta} \langle m', k | \exp\left[-iaP_1\right] | m, k \rangle . \tag{5.4.167}$$

If we recognize that P_1 can be expressed as the average of $P_\pm$,

$$\langle m', k' | \exp\left[-iaP_1\right] | m, k \rangle$$
$$= \sum_{\ell=0}^{+\infty} \frac{(-i)^\ell}{\ell!} \left\langle m', k' \left| a^\ell \left(\frac{P_+ + P_-}{2}\right)^\ell \right| m, k \right\rangle \tag{5.4.168}$$
$$= \sum_{\ell=0}^{+\infty} \frac{(-i)^\ell}{\ell!} \sum_{b=0}^{\ell} \frac{\ell!}{b!(\ell-b)!} \frac{a^\ell}{2^\ell} \left\langle m', k' \left| (P_+)^b (P_-)^{\ell-b} \right| m, k \right\rangle \tag{5.4.169}$$
$$= \sum_{\ell=0}^{+\infty} (-i)^\ell \sum_{b=0}^{\ell} \frac{(a/2)^\ell}{b!(\ell-b)!} (+ik)^b (-ik)^{\ell-b} \langle m', k' | m - \ell + 2b, k \rangle . \tag{5.4.170}$$

The final inner product is zero unless $\ell = m - m' + 2b$. Note that b runs to infinity since ℓ does. Hence, if we allow the ℓ−summation to be collapsed by $\langle m', k' | m - \ell + 2b, k \rangle \propto \delta^\ell_{2b+m-m'}$,

$$\langle m', k' | \exp\left[-iaP_1\right] | m, k \rangle = 2\pi (-)^{m-m'} \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \left(\frac{ka}{2}\right)^{m-m'} \sum_{b=0}^{+\infty} \frac{(-)^b}{b!(b+m-m')!} \left(\frac{ka}{2}\right)^{2b}$$
$$= 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot J_{m'-m}(ka), \tag{5.4.171}$$

where we have employed $(-)^n J_n(z) = J_{-n}(z)$ for integer n; and recognized the series representation of the Bessel function J in eq. (5.4.118).

Note that eq. (5.4.118) defines $J_\nu(z)$ even when ν is complex. In particular, the factorial of α when α is a complex number is defined as $\alpha! \equiv \Gamma(\alpha+1)$, where the right hand side is the Gamma function. Furthermore, one may check that eq. (5.4.118) converges for all $z \in \mathbb{C}$; the even power series multiplying $(z/2)^\nu$ is in fact analytic in *both* z and ν.

Problem 5.46. Series from Integral Representation Derive the series in eq. (5.4.118) directly from the integral representation in eq. (5.4.141). Hint: Taylor expand the $e^{iz\sin\theta}$ or $e^{iz\cos\theta}$ within the integrand. □

To summarize: with the choice of phases in eq. (5.4.164), the translation operator matrix elements takes the form

$$\langle m', k' | \mathcal{T}(\vec{a}) | m, k \rangle = 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot J_{m'-m}(ka) e^{-i(m'-m)\theta}, \tag{5.4.172}$$
$$\vec{a} = a\left(\cos\theta, \sin\theta\right) . \tag{5.4.173}$$

Finally, we compute $\Pi(\phi, \vec{a}) = \mathcal{T}(\vec{a}) \cdot D(\phi)$ in the same basis.

$$\langle m', k' | \Pi(\phi, \vec{a}) | m, k \rangle = 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot J_{m'-m}(ka) e^{-i(m'-m)\theta} e^{-im\phi} . \tag{5.4.174}$$

Addition Theorem of Bessel Let us define

$$\vec{a} \equiv a\,(\cos\theta, \sin\theta) \tag{5.4.175}$$

$$\vec{a}' \equiv a'\,(\cos\theta', \sin\theta')\,; \tag{5.4.176}$$

and study the implication of

$$\mathcal{T}(\vec{a})\mathcal{T}(\vec{a}') = \mathcal{T}\left(\vec{R}(\psi) \equiv \vec{a}(\theta) + \vec{a}'(\theta')\right). \tag{5.4.177}$$

(Drawing a figure helps.) Computing the matrix elements of $\mathcal{T}(\vec{a})\mathcal{T}(\vec{a}')$,

$$\begin{aligned}
&\langle m', k' \,|\mathcal{T}(\vec{a})\mathcal{T}(\vec{a}')|\, m, k\rangle \\
&= \sum_{m''=-\infty}^{+\infty} \int_0^\infty \frac{\mathrm{d}k'' k''}{2\pi} \langle m', k' \,|\mathcal{T}(\vec{a})|\, m'', k''\rangle \langle m'', k'' \,|\mathcal{T}(\vec{a}')|\, m, k\rangle \\
&= 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot \sum_{m''=-\infty}^{+\infty} J_{m'-m''}(ka) J_{m''-m}(ka') e^{im''(\theta-\theta')} e^{-im'\theta} e^{im\theta'}.
\end{aligned} \tag{5.4.178}$$

This is supposed to be equivalent to the same matrix element of $\mathcal{T}(\vec{R})$. Denoting $R \equiv |\vec{R}|$,

$$\left\langle m', k' \left| \mathcal{T}(\vec{R}(\psi)) \right| m, k \right\rangle = 2\pi \frac{\delta(k'-k)}{\sqrt{k' \cdot k}} \cdot J_{m'-m}(kR) e^{-i(m'-m)\psi}. \tag{5.4.179}$$

Reading off the coefficients of the $\delta-$functions on both sides, we arrive at the following additional formula for Bessel $J_n(z)$:

$$J_{m'-m}(kR) e^{-i(m'-m)\psi} = \sum_{m''=-\infty}^{+\infty} J_{m'-m''}(ka) J_{m''-m}(ka') e^{im''(\theta-\theta')} e^{-im'\theta} e^{im\theta'}. \tag{5.4.180}$$

If we set $m' = n \in \mathbb{Z}$ and $m = 0$,

$$J_n(kR) e^{-in(\psi-\theta)} = \sum_{m''=-\infty}^{+\infty} J_{n-m''}(ka) J_{m''}(ka') e^{im''(\theta-\theta')}; \tag{5.4.181}$$

$$\vec{R} \equiv R\,(\cos\psi, \sin\psi)\,, \qquad R \equiv |\vec{a} + \vec{a}'|. \tag{5.4.182}$$

Note that $\cos(\psi - \theta) = (\vec{R} \cdot \vec{a})/(R \cdot a)$ and $\cos(\theta - \theta') = (\vec{a} \cdot \vec{a}')/(a \cdot a')$.

Problem 5.47. Demonstrate that the translation operator $\mathcal{T}(\vec{a})$ admits the following Fourier series representation in the momentum eigenbasis.

$$\left\langle \vec{k} \,|\mathcal{T}(\vec{a})|\, \vec{k}' \right\rangle = (2\pi)^2 \delta^{(2)}(\vec{k} - \vec{k}') \sum_{\substack{n=-\infty \\ n \in \mathbb{Z}}}^{+\infty} (-i)^n J_n(ka) e^{-in\varphi}; \tag{5.4.183}$$

Hint: The involves the decomposition of $\exp(-i\vec{k} \cdot \vec{a})$. □

Problem 5.48. Alternate Derivation of Bessel Addition Formula By recalling eq. (5.4.132), explain why

$$\left\langle \vec{0} \left| \mathcal{T}(\vec{a}(\theta))^\dagger \right| m, k \right\rangle = J_m(ka) e^{im\theta}. \tag{5.4.184}$$

Then derive eq. (5.4.181) by considering $\left\langle \vec{0} \left| \left(\mathcal{T}(\vec{a}(\theta)) \mathcal{T}(\vec{a}'(\theta')) \right)^\dagger \right| m, k \right\rangle$. □

Problem 5.49. Multiple Translations If we define

$$\vec{a}_{\rm I} \equiv a_{\rm I} \left(\cos\theta_{\rm I}, \sin\theta_{\rm I} \right), \tag{5.4.185}$$

for $\mathrm{I} = 1, \dots, N$; and

$$\vec{R} = \vec{a}_1 + \dots + \vec{a}_N \equiv R \left(\cos\psi, \sin\psi \right); \tag{5.4.186}$$

show that

$$\begin{aligned} & J_{m'-m}(kR) e^{-i(m'-m)\psi} \\ & = \sum_{m_1 \in \mathbb{Z}} \sum_{m_2 \in \mathbb{Z}} \dots \sum_{m_{N-2} \in \mathbb{Z}} \sum_{m_{N-1} \in \mathbb{Z}} J_{m'-m_1}(ka_1) J_{m_1-m_2}(ka_2) \dots J_{m_{N-2}-m_{N-1}}(ka_{N-1}) J_{m_{N-1}-m}(ka_N) \\ & \qquad \times e^{-i(m'-m_1)\theta_1} e^{-i(m_1-m_2)\theta_2} \dots e^{-i(m_{N-2}-m_{N-1})\theta_{N-1}} e^{-i(m_{N-1}-m)\theta_N}. \end{aligned} \tag{5.4.187}$$

Explain why the right hand side remains the same upon swapping any of the two $\vec{a}$s. □

Problem 5.50. If $\vec{a} \cdot \vec{P} \equiv a^1 P_1 + a^2 P_2$ in Cartesian coordinates (a^1, a^2), show that the dot product can be expressed as

$$\vec{a} \cdot \vec{P} = a^- P_+ + a^+ P_-, \tag{5.4.188}$$

where $a^\mp \equiv (1/2)(a^1 \mp i a^2)$. □

Problem 5.51. Derivatives If $\vec{x} = r(\cos\phi, \sin\phi)$, show that

$$\frac{i}{2} \left\langle r, \phi \left| e^{-i\phi} P_+ + e^{+i\phi} P_- \right| f \right\rangle = \partial_r \left\langle r, \phi \right| f \rangle \tag{5.4.189}$$

$$\frac{1}{2} \left\langle r, \phi \left| e^{-i\phi} P_+ - e^{+i\phi} P_- \right| f \right\rangle = \frac{1}{r} \partial_\phi \left\langle r, \phi \right| f \rangle; \tag{5.4.190}$$

for any arbitrary state $|f\rangle$. Also verify the following commutators.

$$\left\langle r, \phi \left| \left[P_\pm, e^{\pm i\phi} \right] \right| f \right\rangle = \frac{i}{r} e^{\pm 2i\phi} \left\langle r, \phi \right| f \rangle \tag{5.4.191}$$

$$\left\langle r, \phi \left| \left[P_\pm, e^{\mp i\phi} \right] \right| f \right\rangle = -\frac{i}{r} \left\langle r, \phi \right| f \rangle \tag{5.4.192}$$

$$\left\langle r, \phi \left| \left[P_\pm, r^{-1} \right] \right| f \right\rangle = \frac{i}{r^2} e^{\pm i\phi} \left\langle r, \phi \right| f \rangle. \tag{5.4.193}$$

Hint: First start by computing $\langle r, \phi | P_\pm | f \rangle$ in polar coordinates. □

Problem 5.52. Bessel Equation Via a direct calculation, show that

$$\left\{ \left(\frac{i}{2k} \left(e^{-i\phi} P_+ + e^{+i\phi} P_- \right) \right)^2 + \frac{i}{2k^2 r} \left(e^{-i\phi} P_+ + e^{+i\phi} P_- \right) + 1 - \frac{m^2}{(kr)^2} \right\} |m, k\rangle = 0. \tag{5.4.194}$$

By acting $\langle r, \phi|$ from the left, explain why this yields the Bessel equation,

$$J_n''(\xi) + \frac{1}{\xi} J_n'(\xi) + \left(1 - \frac{m^2}{\xi^2} \right) J_n(\xi) = 0; \tag{5.4.195}$$

where $\xi \equiv kr$. Remember that the $P_\pm$ do not commute with $1/r$ nor $e^{\pm i\phi}$. $\square$

Problem 5.53. Bessel Recursion Relations Use eq. (5.4.164) and the results in Problem (5.51) to show that

$$J_m'(\xi) = -J_{m+1}(\xi) + \frac{m}{\xi} J_m(\xi) \tag{5.4.196}$$

$$= J_{m-1}(\xi) - \frac{m}{\xi} J_m(\xi). \tag{5.4.197}$$

Explain why these immediately imply

$$2 J_m'(\xi) = J_{m-1}(\xi) - J_{m+1}(\xi) \tag{5.4.198}$$

$$2 \frac{m}{\xi} J_m(\xi) = J_{m-1}(\xi) + J_{m+1}(\xi); \tag{5.4.199}$$

as well as, for $n, m \in \mathbb{Z}^+$,

$$\left(\frac{1}{\xi} \frac{\mathrm{d}}{\mathrm{d}\xi} \right)^n \left(\xi^m J_m(\xi) \right) = \xi^{m-n} J_{m-n}(\xi) \tag{5.4.200}$$

$$\left(\frac{1}{\xi} \frac{\mathrm{d}}{\mathrm{d}\xi} \right)^n \left(\xi^{-m} J_m(\xi) \right) = (-)^n \xi^{-m-n} J_{m+n}(\xi). \tag{5.4.201}$$

Hint: Work in the position representation. $\square$

Invariant Subgroup of $\mathbb{E}_2$ Next, we compute the following change-of-basis of the translation operator, keeping in mind equations (5.4.87) and (5.4.89) as well as the fact that translations commute:

$$\Pi(\phi, \vec{a}) \cdot \mathcal{T}(\vec{b}) \cdot \Pi(\phi, \vec{a})^{-1} = \mathcal{T}(\vec{a}) \cdot D(\phi) \cdot \mathcal{T}(\vec{b}) \cdot D(-\phi) \cdot \mathcal{T}(-\vec{a}) \tag{5.4.202}$$

$$= \mathcal{T}(\vec{a}) \cdot \mathcal{T}(\widehat{R}(\phi) \cdot \vec{b}) \cdot \mathcal{T}(-\vec{a}) \tag{5.4.203}$$

$$= \mathcal{T}(\vec{a}) \cdot \mathcal{T}(-\vec{a}) \cdot \mathcal{T}(\widehat{R}(\phi) \cdot \vec{b}) \tag{5.4.204}$$

$$= \mathcal{T}\left(\widehat{R}(\phi) \cdot \vec{b} \right). \tag{5.4.205}$$

Since $\Pi(\phi, \vec{a})$ is a completely general $\mathbb{E}_2$ group element, we see that the subgroup of translations remain so under arbitrary change-of-basis. In group theory lingo, we say that translations in 2D form an *invariant subgroup* of $\mathbb{E}_2$.

We may consider forming the *left coset* of the group of translations, viewed as an invariant subgroup of $\mathbb{E}_2$ translations by multiplying some *fixed* $\Pi(\phi, \vec{a})$ from the left to the set of all translation operators. (Multiplying it from the right would yield the *right coset.*) Specifically, for fixed ϕ and $\vec{a}$, we may construct the left coset

$$\Pi(\phi, \vec{a}) \left\{ \mathcal{T}(\vec{b}) \middle| \vec{b} \in \mathbb{R}^2 \right\} \equiv \left\{ \Pi\left(\phi, \widehat{R}(\phi) \cdot \vec{b} + \vec{a}\right) \middle| \vec{b} \in \mathbb{R}^2 \right\}. \tag{5.4.206}$$

Since $\vec{b}$ runs over all possible vectors in $\mathbb{R}^2$, we may always find a $\vec{b}'$ such that $\widehat{R}(\phi) \cdot \vec{b} + \vec{a} = \vec{b}'$. Likewise, given $\vec{b}'$, we may always find a $\vec{b}$ such that $\vec{b} = \widehat{R}(-\phi) \cdot (\vec{b}' - \vec{a})$. Our left coset is therefore the set of all Πs in $\mathbb{E}_2$ with a fixed ϕ.

$$\Pi(\phi, \vec{a}) \left\{ \mathcal{T}(\vec{b}) \middle| \vec{b} \in \mathbb{R}^2 \right\} = \left\{ \Pi\left(\phi, \vec{b}'\right) \middle| \vec{b}' \in \mathbb{R}^2 \right\} \tag{5.4.207}$$

If H is an invariant subgroup and if we define group multiplication of (left) cosets by $(g_1 \cdot H)(g_2 \cdot H) = (g_1 g_2) \cdot H$, where g_1 and g_2 are group elements, then in the $H = \mathbb{T}^2$ case at hand,

$$\begin{aligned}
\left\{ \Pi\left(\phi, \vec{b}\right) \middle| \vec{b} \in \mathbb{R}^2 \right\} \cdot \left\{ \Pi\left(\phi', \vec{b}'\right) \middle| \vec{b}' \in \mathbb{R}^2 \right\} &= (\Pi(\phi, \vec{a}) \cdot \Pi(\phi', \vec{a}')) \left\{ \mathcal{T}(\vec{b}) \middle| \vec{b} \in \mathbb{R}^2 \right\} \\
&= \Pi\left(\phi + \phi', \vec{a} + \widehat{R}(\phi) \cdot \vec{a}'\right) \left\{ \mathcal{T}(\vec{b}) \middle| \vec{b} \in \mathbb{R}^2 \right\} \\
&= \left\{ \Pi\left(\phi + \phi', \vec{b}''\right) \middle| \vec{b}'' \in \mathbb{R}^2 \right\}. \qquad (5.4.208)
\end{aligned}$$

The two cosets on the leftmost and the coset on the rightmost portions of the above calculation are the same except for the angles of rotation – the latter's is the sum of the former two. By viewing the entire set into one group element of the coset group $gH \in \mathbb{E}_2/\mathbb{T}_2$, we are modding out the translation group from the Euclidean group to obtain SO_2. This tells us the factor group $\mathbb{E}_2/\mathbb{T}_2$ defined via the rule $(g_1 \cdot H)(g_2 \cdot H) = (g_1 g_2) \cdot H$ is in fact equivalent to that of the rotation group in 2D.

5.5 Rotations in $D \geq 3$ Spatial Dimensions

We now move on to study rotations in spatial dimensions D higher than 2. In 2D we were able to write down the general rotation matrix $\widehat{R}(\phi)$ in eq. (5.4.6) because the associated geometric considerations were simple enough. In general $D \geq 3$ dimensions, however, we would need a more abstract starting point. For the $D = 3$ case, we have an intuitive understanding that rotations are a continuous operation, parametrized by some appropriate angles, that do not change the length of vectors nor their dot products. For general D,[30] we thus begin by *defining* rotations as the linear operator that preserves the lengths of vectors and their dot products. If $\widehat{R}$ is a rotation matrix and $\vec{u}$ and $\vec{v}$ are arbitrary but distinct vectors, for their dot product to be preserved under rotations means

$$\vec{u} \cdot \vec{v} = \vec{u}^{\mathrm{T}} \vec{v} = (\widehat{R} \vec{u}) \cdot (\widehat{R} \vec{v}) \tag{5.5.1}$$

$$= \vec{u}^{\mathrm{T}} (\widehat{R}^{\mathrm{T}} \widehat{R}) \vec{v}. \tag{5.5.2}$$

[30] Do not let the general-D character of the discussion intimidate you: a good portion of what follows would be identical even if we had put $D = 3$.

Since this holds for any pair of distinct $\vec{u}$ and $\vec{v}$, we see that rotation matrices are orthogonal ones:

$$\mathbb{I} = \widehat{R}^{\mathrm{T}}\widehat{R}. \tag{5.5.3}$$

In index notation,

$$\delta_{ij} = \delta_{ab}\widehat{R}^{a}{}_{i}\widehat{R}^{b}{}_{j}. \tag{5.5.4}$$

Because the inverse is unique, and the left- and right-inverses are the same for finite dimensional matrices, we may in fact deduce from eq. (5.5.3) that

$$\widehat{R}^{\mathrm{T}} = \widehat{R}^{-1} \qquad \text{and} \qquad \widehat{R}\widehat{R}^{\mathrm{T}} = \mathbb{I} = \widehat{R}^{\mathrm{T}}\widehat{R}. \tag{5.5.5}$$

We will soon see that these $D \times D$ rotation matrices $\widehat{R}$ are parametrized by $D(D-1)/2$ angles $\{\vec{\theta}\}$. Moreover, it must be possible to tune these angles $\{\vec{\theta} \to \vec{\theta}_0\}$ so that the identity matrix is recovered – i.e., no rotation at all:

$$\widehat{R}(\vec{\theta}_0) = \mathbb{I}_{D\times D} \tag{5.5.6}$$

Let us now take the determinant of both sides of eq. (5.5.3). Using $\det \widehat{A}\widehat{B} = (\det \widehat{A})(\det \widehat{B})$ and $\det \widehat{A}^{\mathrm{T}} = \det \widehat{A}$,

$$1 = \det \mathbb{I} = \det \widehat{R}^{\mathrm{T}}\widehat{R} = (\det \widehat{R})^2. \tag{5.5.7}$$

In other words, orthogonal matrices satisfying eq. (5.5.3) have either ± 1 determinant:

$$\det \widehat{R} = \pm 1. \tag{5.5.8}$$

If $\widehat{R}$ is a rotation matrix, however, since $\vec{\theta}$ are continuous parameters, the $\det \widehat{R}(\vec{\theta})$ must be a number that is also a continuous function of these $\vec{\theta}$. Suppose we start from the angles $\vec{\theta}_0$ in eq. (5.5.6); where, $\det \widehat{R}(\vec{\theta}_0) = 1$. By tuning $\vec{\theta}$ away from $\vec{\theta}_0$ in the number $\det \widehat{R}(\vec{\theta})$, one might think the determinant would vary continuously as a function of these angles. But we have just seen it can only take on 2 discrete values ± 1; therefore $\det \widehat{R}(\vec{\theta})$ has to remain $+1$ for all $\vec{\theta}$ for otherwise it would violate continuity.

$$\text{Rotations: } \det \widehat{R}(\vec{\theta}) = +1 \tag{5.5.9}$$

To summarize, for arbitrary $D \geq 2$ dimensional space, the set of rotation matrices are thus *defined* to be the set of $D \times D$ matrices that are simultaneously orthogonal (eq. (5.5.3)) and have unit determinant (eq. (5.5.9)).

Problem 5.54. Explain why eq. (5.5.4) implies eq. (5.5.3). That is, convert the index notation to a matrix notation. Also explain why, eq. (5.5.4) implies

$$\delta^{ij}\widehat{R}^{a}{}_{i}\widehat{R}^{b}{}_{j} = \delta^{ab}. \tag{5.5.10}$$

Hint: Use $\widehat{R}^{\mathrm{T}} = \widehat{R}^{-1}$. ☐

Problem 5.55. Levi-Civita: (Pseudo-)Invariant Tensor If $\widehat{R}$ obeys $\widehat{R}^{\rm T}\widehat{R} = \mathbb{I}$ and $\epsilon_{i_1 \dots i_D} \equiv \epsilon^{i_1 \dots i_D}$ is the Levi-Civita symbol in D dimensions, with $\epsilon_{1\ 2\ \dots\ D} \equiv 1$, explain why

$$\widehat{R}^{i_1}{}_{j_1} \dots \widehat{R}^{i_D}{}_{j_D} \epsilon_{i_1 \dots i_D} = (\det \widehat{R}) \epsilon_{j_1 \dots j_D}, \tag{5.5.11}$$

$$\widehat{R}^{i_1}{}_{j_1} \dots \widehat{R}^{i_D}{}_{j_D} \epsilon^{j_1 \dots j_D} = (\det \widehat{R}) \epsilon^{i_1 \dots i_D}. \tag{5.5.12}$$

(Hint: Recall the definition of the matrix determinant in eq. (3.2.1).) Therefore, when $\widehat{R}$ is a rotation matrix, with $\det \widehat{R} = 1$; we see that the Levi-Civita is an invariant tensor. However, when we include parity flips, where $\det \widehat{R} = -1$, and the statement no longer holds.

Next, show that

$$\widehat{R}^{i_{N+1}}{}_{j_{N+1}} \dots \widehat{R}^{i_D}{}_{j_D} \epsilon_{i_1 \dots i_N i_{N+1} \dots i_D} = (\det \widehat{R}) \widehat{R}^{i_1}{}_{j_1} \dots \widehat{R}^{i_N}{}_{j_N} \epsilon_{j_1 \dots j_N j_{N+1} \dots j_D}, \tag{5.5.13}$$

$$\widehat{R}^{i_{N+1}}{}_{j_{N+1}} \dots \widehat{R}^{i_D}{}_{j_D} \epsilon^{j_1 \dots j_N j_{N+1} \dots j_D} = (\det \widehat{R}) \widehat{R}^{i_1}{}_{j_1} \dots \widehat{R}^{i_N}{}_{j_N} \epsilon^{i_1 \dots i_N i_{N+1} \dots i_D}. \tag{5.5.14}$$

Hints: Multiply both sides of equations (5.5.11) and (5.5.12) with the appropriate factors of $\widehat{R}$'s and utilize $\widehat{R}^{ia}\widehat{R}^{ib} = \delta^{ab} = \widehat{R}^{ai}\widehat{R}^{bi}$. □

Group Structure Note that the product of two rotations $\widehat{R}_1$ and $\widehat{R}_2$ is another rotation because, as long as eq. (5.5.3) is obeyed – namely $\widehat{R}_1^{\rm T}\widehat{R}_1 = \mathbb{I}$ and $\widehat{R}_2^{\rm T}\widehat{R}_2 = \mathbb{I}$ – then we have

$$\left(\widehat{R}_1\widehat{R}_2\right)^{\rm T}\left(\widehat{R}_1\widehat{R}_2\right) = \widehat{R}_2^{\rm T}\widehat{R}_1^{\rm T}\widehat{R}_1\widehat{R}_2 \tag{5.5.15}$$

$$= \widehat{R}_2^{\rm T}\widehat{R}_2 = \mathbb{I}. \tag{5.5.16}$$

That is, $\widehat{R}_3 \equiv \widehat{R}_1\widehat{R}_2$ satisfies $\widehat{R}_3^{\rm T}\widehat{R}_3 = \mathbb{I}$. Moreover, as long as as eq. (5.5.9) is respected – namely, $\det \widehat{R}_1 = 1 = \det \widehat{R}_2$ – then $\det \widehat{R}_3 = 1$ too, because

$$\det\left(\widehat{R}_1\widehat{R}_2\right) = (\det \widehat{R}_1)(\det \widehat{R}_2) = 1. \tag{5.5.17}$$

(Special) Orthogonal Group The set of matrices $\{\widehat{R}\}$ obeying eq. (5.5.3) form the orthogonal group $\mathrm{O}_D (\equiv \mathrm{O}(D))$; if they are further restricted to be of unit determinant, i.e., obeying eq. (5.5.9) (aka 'special'), they form the $\mathrm{SO}_D (\equiv \mathrm{SO}(D))$ group. Hence, rotations in $D-$space correspond to the study of the SO_D group.

Problem 5.56. U_D and SU_D Groups Prove that the set of $D \times D$ unitary matrices $\{\widehat{U} | \widehat{U}^\dagger \widehat{U} = \mathbb{I} = \widehat{U}\widehat{U}^\dagger\}$ forms a group. For every matrix in the U_D group, explain how to obtain a matrix in the SU_D group. Hint: If $\widehat{U}$ is in U_D, consider multiplying $\widehat{U}$ by an appropriate phase $e^{i\varphi}\mathbb{I}_{D\times D}$.

How is the U_1 group related to the complex plane? Can you write down the general 2×2 matrix of the U_2 and SU_2 groups using the Pauli matrices $\{\sigma^\mu | \mu = 0, 1, 2, 3\}$ in eq. (3.2.18)? Hints: Remember these Pauli matrices span the space of 2×2 matrices. Consider the following complex $4-$component object:

$$\vec{a} = \Big(e^{i\varphi_1}\sin(\theta_1)\sin(\theta_2)\cos(\theta_3), e^{i\varphi_2}\sin(\theta_1)\sin(\theta_2)\sin(\theta_3),$$
$$e^{i\varphi_3}\sin(\theta_1)\cos(\theta_2), e^{i\varphi_4}\cos(\theta_1)\Big), \qquad \varphi_{1,2,3,4}, \theta_{1,2,3} \in \mathbb{R}. \tag{5.5.18}$$

Compute $\vec{a}^* \cdot \vec{a}$. □

Geometry Let us turn to the geometry of flat space itself, and witness how it enjoys rotational and spatial translation symmetries. Firstly, by Pythagoras's theorem – which holds for arbitrary $D \geq 2$ dimensions – the distance $\mathrm{d}\ell$ between $\vec{x}$ and $\vec{x} + \mathrm{d}\vec{x}$ obeys

$$\mathrm{d}\ell^2 = \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x} = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \equiv g_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{5.5.19}$$

In differential geometry, the metric g_{ij} is defined as the coefficient of $\mathrm{d}x^i\mathrm{d}x^j$ in the square of the infinitesimal distance $\mathrm{d}\ell^2$. Here, $g_{ij} = \delta_{ij}$. Moreover, under an orthogonal transformation of the Cartesian coordinates $\vec{x} \to \widehat{R}\vec{x}$; followed by a spatial translation $\vec{a}$, namely $\widehat{R}\vec{x} \to \widehat{R}\vec{x} + \vec{a}$; we gather

$$\vec{x} \to \widehat{R} \cdot \vec{x} + \vec{a}, \tag{5.5.20}$$

$$x^i \to \widehat{R}^i{}_j x^j + a^i. \tag{5.5.21}$$

For constant rotation angles $\vec{\theta}$ and translation vector $\vec{a}$, eq. (5.5.20) tells us the infinitesimal displacement is rotated as

$$\mathrm{d}x^i \to \widehat{R}^i{}_j \mathrm{d}x^j. \tag{5.5.22}$$

The square of the infinitesimal distance $\mathrm{d}\ell$ is, in turn, replaced as

$$\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \to \delta_{ij}(\widehat{R}^i{}_a\mathrm{d}x^a)(\widehat{R}^j{}_b\mathrm{d}x^b) \tag{5.5.23}$$

$$= (\delta_{ab}\widehat{R}^a{}_i\widehat{R}^b{}_j)\mathrm{d}x^i\mathrm{d}x^j = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{5.5.24}$$

where eq. (5.5.4) was used in the final equality. In matrix notation, and invoking the matrix version of orthogonality in eq. (5.5.3),

$$\mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x} \to (\widehat{R}\mathrm{d}\vec{x}) \cdot (\widehat{R}\mathrm{d}\vec{x})\delta_{ab}\frac{\partial x^a}{\partial x'^i}\frac{\partial x^b}{\partial x'^j} = \mathrm{d}\vec{x} \cdot (\widehat{R}^{\mathrm{T}}\widehat{R})\mathrm{d}\vec{x} = \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}. \tag{5.5.25}$$

Conversely, if we perform a coordinate transformation $\vec{x} = \vec{x}(\vec{x}')$, the element transforms into

$$\delta_{ab}\mathrm{d}x^a\mathrm{d}x^b = \delta_{ab}\frac{\partial x^a}{\partial x'^i}\frac{\partial x^b}{\partial x'^j}\mathrm{d}x'^i\mathrm{d}x'^j. \tag{5.5.26}$$

We won't show it here – but see §(10.1) below for the Poincaré symmetry case – but one can show that the most general transformation that preserves the metric, namely

$$g_{ij}(\vec{x}) = \delta_{ij} \to \delta_{ab}\frac{\partial x^a}{\partial x'^i}\frac{\partial x^b}{\partial x'^j} = \delta_{ij} = g_{i'j'}(\vec{x}'), \tag{5.5.27}$$

is

$$\vec{x} = \widehat{R}\vec{x}' + \vec{a}, \tag{5.5.28}$$

$$x^i = \widehat{R}^i{}_j x'^j + a^i. \tag{5.5.29}$$

To sum:

The rotation and translation operation(s), namely $\vec{x} \to \widehat{R}\vec{x} + \vec{a}$ – with orthogonal transformations implemented by $\{\widehat{R}(\vec{\theta})|\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I} = \widehat{R}\widehat{R}^{\mathrm{T}}\}$ and translations implemented by constant vectors $\{\vec{a}\}$ – together preserve the infinitesimal line element $\mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x} \to \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}$. We say that flat space is rotationally and translationally invariant: $\delta_{ij} \to \delta_{ij}$.

Problem 5.57. Euclidean Group Explain how you would generalize eq. (5.4.78) to all spatial dimensions $D \geq 2$. $\square$

Change of orthonormal basis If we have two sets of orthonormal basis $\{\widehat{e}_a|a = 1, 2, \dots, D\}$ and $\{\widehat{e}'_a|a = 1, 2, \dots, D\}$, each obeying $\widehat{e}_a \cdot \widehat{e}_b = \delta_{ab} = \widehat{e}'_a \cdot \widehat{e}'_b$, the change-of-basis matrix that would transform one basis to another is in fact a rotation matrix. For, suppose we seek $\widehat{R}_{ma}$ such that it would implement such a change-of-basis,

$$\widehat{e}'_a \equiv \widehat{e}_m \widehat{R}_{ma}. \tag{5.5.30}$$

Then, dotting both sides with $\widehat{e}_b$ would inform us, the components of the rotation matrix are in fact the cosine of the angle between $\widehat{e}'_a$ and $\widehat{e}_b$:

$$\widehat{e}_b \cdot \widehat{e}'_a = \widehat{R}_{ba}. \tag{5.5.31}$$

Keeping in mind the completeness relation $\sum_\ell \widehat{e}^a_\ell \widehat{e}^b_\ell = \delta^{ab}$, we may check directly that

$$(\widehat{e}_m \widehat{R}_{ma})^i = \widehat{e}^i_m \widehat{e}^\ell_m \widehat{e}'^\ell_a = \delta^{i\ell} \widehat{e}'^\ell_a = \widehat{e}'^i_a. \tag{5.5.32}$$

Moreover,

$$\delta_{ab} = \widehat{e}'_a \cdot \widehat{e}'_b = \widehat{e}_m \cdot \widehat{e}_n \widehat{R}_{ma} \widehat{R}_{nb} = \delta_{mn} \widehat{R}_{ma} \widehat{R}_{nb} = (\widehat{R}^{\mathrm{T}} \widehat{R})_{ab}; \tag{5.5.33}$$

which is nothing but the definition of the O_D group.

On the other hand, suppose $\{\widehat{e}_a\}$ were orthonormal but we do not assume $\{\widehat{e}'_a\}$ to be from the outset. If $\widehat{R}$ were in the O_D group, defining $\widehat{e}'_a \equiv \widehat{e}_m \widehat{R}_{ma}$ would automatically lead to the conclusion that $\{\widehat{e}'_a\}$ are in fact unit length and mutually perpendicular:

$$\widehat{e}'_a \cdot \widehat{e}'_b = \widehat{R}_{ma} \widehat{R}_{nb} \widehat{e}_m \cdot \widehat{e}_n = (\widehat{R}^{\mathrm{T}} \widehat{R})_{ab} = \delta_{ab}. \tag{5.5.34}$$

Our discussion here is the real orthogonal version of the more general one in §(4.3.3).

Anti-Symmetric Generators We now turn to the construction of $\widehat{R}$ itself. As we have argued previously, any operator continuously connected to the identity can be expressed as an exponential:

$$\widehat{R}(\vec{\theta}) = e^{\varepsilon \cdot \widehat{\Omega}} = \mathbb{I} + \varepsilon \cdot \widehat{\Omega} + \mathcal{O}\left(\varepsilon^2\right), \tag{5.5.35}$$

where the matrix $\widehat{\Omega}$ is known as the *generator* of rotations, which we will take to be real since $\widehat{R}$ is real. Furthermore, we have inserted a parameter ε so that eq. (5.5.3) may be now regarded as a Taylor series in ε.

$$\left(\mathbb{I} + \varepsilon \cdot \widehat{\Omega}^{\mathrm{T}} + \mathcal{O}\left(\varepsilon^2\right)\right)\left(\mathbb{I} + \varepsilon \cdot \widehat{\Omega} + \mathcal{O}\left(\varepsilon^2\right)\right) = \mathbb{I} \tag{5.5.36}$$

$$\mathbb{I}+\varepsilon\left(\widehat{\Omega}^{\mathrm{T}}+\widehat{\Omega}\right)+\mathcal{O}\left(\varepsilon^{2}\right)=\mathbb{I} \tag{5.5.37}$$

The identity cancels out from both sides; leaving us with the conclusion that each order in ε must cancel. In particular, at first oder,

$$\widehat{\Omega}^{\mathrm{T}}=-\widehat{\Omega}. \tag{5.5.38}$$

Now, if eq. (5.5.38) were true then from eq. (5.5.35), we may verify eq. (5.5.3). By Taylor expanding the exponential, one may readily verify that $(\exp(\widehat{\Omega}))^{\mathrm{T}}=\exp(\widehat{\Omega}^{\mathrm{T}})=\exp(-\widehat{\Omega})$. Since $-\widehat{\Omega}$ and $\widehat{\Omega}$ commutes, we may combine the exponents in eq. (5.5.3)

$$\widehat{R}^{\mathrm{T}}\widehat{R}=e^{-\widehat{\Omega}}e^{\widehat{\Omega}}=e^{\widehat{\Omega}-\widehat{\Omega}}=\mathbb{I}, \tag{5.5.39}$$

where we have now absorbed ε into the generator $\widehat{\Omega}$.

Rotation angles Moreover, note that antisymmetric matrices (with a total of D^2 entries) have zeros on the diagonal (since $\widehat{\Omega}_{ii}=-\widehat{\Omega}_{ii}$, with no sum over i) and are hence fully specified by either its strictly upper or lower triangular components (since its off diagonal counterparts may be obtained via $\widehat{\Omega}_{ij}=-\widehat{\Omega}_{ji}$). Thus, the space of antisymmetric matrices is $(D^2-D)/2=D(D-1)/2$ dimensional.

On the other hand, there are $\binom{D}{2}=D!/(2!(D-2)!)=D(D-1)/2$ ways to choose a 2D plane spanned by 2 of the D axes in a Cartesian coordinate system. As we shall see below, each of the $D(D-1)/2$ basis anti-symmetric matrices $\widehat{J}^{ij}$ that span the space of $\{\widehat{\Omega}=-i\omega_{ij}\widehat{J}^{ij}|\widehat{\Omega}^{\mathrm{T}}=-\widehat{\Omega}\}$ in fact generate rotations about these 2D planes, with rotation angle $\theta^{\mathrm{I}}\leftrightarrow\omega_{ij}$. Hence, rotations in D spatial dimensions are parametrized by a total of $D(D-1)/2$ rotation angles $\{\theta^{\mathrm{I}}|\mathrm{I}=1,2,\dots,D(D-1)/2\}$.

Basis generators One such basis of anti-symmetric generators $\widehat{\Omega}$ is as follows. First recall all diagonal components are zero. For the first generator basis matrix, set the $(1,2)$ component to -1, the $(2,1)$ to $+1$, and the rest to 0. For the next, set the $(1,3)$ component to -1, the $(3,1)$ to $+1$, and the rest to 0. And so on, until all the upper triangular components have been covered. For example, in 2D, the generator is proportional to the 2D Levi-Civita symbol (with $\epsilon^{12}\equiv1$):

$$i\widehat{J}\doteq\epsilon^{ij}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}; \tag{5.5.40}$$

whereaas in 3D, we have the following three generators

$$i\widehat{J}^{12}=\begin{bmatrix}0&1&0\\-1&0&0\\0&0&0\end{bmatrix},\qquad i\widehat{J}^{13}=\begin{bmatrix}0&0&1\\0&0&0\\-1&0&0\end{bmatrix},\qquad i\widehat{J}^{23}=\begin{bmatrix}0&0&0\\0&0&1\\0&-1&0\end{bmatrix}. \tag{5.5.41}$$

And so on, for all $D\geq2$. Altogether, this amounts to writing eq. (5.5.35) as

$$\widehat{R}=\exp\left(-\frac{i}{2}\omega_{ij}\widehat{J}^{ij}\right), \tag{5.5.42}$$

where $\omega_{ij} = -\omega_{ji}$ is a collection of $D(D-1)/2$ parameters (i.e., rotation angles) expressing the superposition of the anti-symmetric basis generators $\{iJ^{ij}\}$ as described above:

$$i\left(\widehat{J}^{ij}\right)_{ab} = \delta^i_a\delta^j_b - \delta^i_b\delta^j_a \equiv \delta^i_{[a}\delta^j_{b]}. \tag{5.5.43}$$

[31]The reason for writing the generators in the form in eq. (5.5.42) is that, since $\widehat{\Omega} = -(i/2)\omega_{ij}\widehat{J}^{ij}$ is real and antisymmetric, $\widehat{\Omega}^{\mathrm{T}} = \widehat{\Omega}^\dagger = -\widehat{\Omega}$ (cf. eq. (5.5.38)),

$$\frac{1}{2}\omega^*_{ij}(-i\widehat{J}^{ij})^\dagger = \frac{i}{2}\omega_{ij}\widehat{J}^{ij} = -\frac{1}{2}\omega_{ij}(-i\widehat{J}^{ij}). \tag{5.5.44}$$

But eq. (5.5.43) tells us $-i\widehat{J}^{ij}$ is real and anti-symmetric, i.e., $(-i\widehat{J}^{ij})^\dagger = +i\widehat{J}^{ij}$; so not only is $\widehat{J}^{ij}$ therefore Hermitian

$$+i\widehat{J}^{ij} = (-i\widehat{J}^{ij})^\dagger = i(\widehat{J}^{ij})^\dagger \tag{5.5.45}$$

eq. (5.5.44) informs us the parameters $\{\omega_{ij}\}$ in eq. (5.5.42) must therefore be real.

Problem 5.58. Rotation on (i,j)-plane By regarding J^{ij} as an operator acting on the $D-$Euclidean space, whose Cartesian basis we shall denote by $\{|i\rangle\}$, explain why

$$-iJ^{ij}|j\rangle = -|i\rangle \qquad \text{and} \qquad -iJ^{ij}|i\rangle = +|j\rangle; \tag{5.5.46}$$

while

$$-iJ^{ij}|k\rangle = 0, \qquad \forall k \neq i,j. \tag{5.5.47}$$

In other words,

$$-iJ^{ij} = |j\rangle\langle i| - |i\rangle\langle j|. \tag{5.5.48}$$

Can you compute $(-iJ^{ij})^n$ for odd and even n? Then show that

$$e^{-i\theta J^{ij}} = \cos(\theta)|i\rangle\langle i| - \sin(\theta)|i\rangle\langle j| + \sin(\theta)|j\rangle\langle i| + \cos(\theta)|j\rangle\langle j| + \sum_{k\neq i,j}|k\rangle\langle k|. \tag{5.5.49}$$

In other words, the basis anti-symmetric rotation generators in eq. (5.5.43) produce counter-clockwise rotations on the (i,j) 2D plane, leaving the rest of the $D-$space untouched. □

Problem 5.59. Rotation on $(i,j)-$plane: Matrix Form Let $\widehat{e}_i$ and $\widehat{e}_j$ be the orthonormal basis vectors spanning the $(i,j)-$plane. Show that the matrix rotating the latter by an angle θ can be expressed as

$$\widehat{R}^{ab}(\theta) = P^{ab}(\widehat{e}_i,\widehat{e}_j) + \cos\theta\left(\widehat{e}^a_i\widehat{e}^b_i + \widehat{e}^a_j\widehat{e}^b_j\right) - \sin\theta\left(\widehat{e}^a_i\widehat{e}^b_j - \widehat{e}^a_j\widehat{e}^b_i\right), \tag{5.5.50}$$

[31]One may check that $\widehat{J}^{ij} = -\widehat{J}^{ji}$, and therefore the sum in eq. (5.5.42) over the upper triangular indices are not independent from those over the lower triangular ones; this accounts for the factor of 1/2. In other words: $\exp(-(i/2)\omega_{ab}J^{ab}) = \exp(-i\sum_{a<b}\omega_{ab}J^{ab}) = \exp(-i\sum_{a>b}\omega_{ab}J^{ab})$. Furthermore, whenever $i=j$ we see eq. (5.5.43) vanishes; whereas for a fixed pair $i\neq j$, the kronecker deltas on the right hand side tell us $-i(J^{ij})_{ab} = -1$ (coming solely from the first term on the left) when $i=a$ and $j=b$ while $-i(J^{ij})_{ab} = +1$ (coming solely from the second term) when $i=b$ and $j=a$.

where the projector P^{ab} was defined in eq. (4.5.6). Also explain the completeness-like relations

$$P^{ab}\left(\widehat{e}_i, \widehat{e}_j\right) = \sum_{\mathrm{A,B}\neq\{i,j\}}^{D} \delta^{\mathrm{AB}} \widehat{e}^a_{\mathrm{A}} \widehat{e}^b_{\mathrm{B}}, \tag{5.5.51}$$

$$P^{ab}(\{\widehat{e}_{\mathrm{I}} | \mathrm{I} \neq \{i,j\}\}) = \sum_{\mathrm{A,B}\in\{i,j\}} \delta^{\mathrm{AB}} \widehat{e}^a_{\mathrm{A}} \widehat{e}^b_{\mathrm{B}}. \tag{5.5.52}$$

Hint: Recall the completeness relation in eq. (4.5.12). □

Change-of-Basis and Rotating the ωs We may now show that

$$\widehat{R} \exp\left(-\frac{i}{2}\omega_{ab}\widehat{J}^{ab}\right)\widehat{R}^{\mathrm{T}} = \exp\left(-\frac{i}{2}\omega'_{ab}\widehat{J}^{ab}\right), \tag{5.5.53}$$

with

$$\omega'_{ab} = \widehat{R}_{am}\widehat{R}_{bn}\omega_{mn}; \tag{5.5.54}$$

or, if we view ω as a matrix,

$$\widehat{\omega}' = \widehat{R} \cdot \widehat{\omega} \cdot \widehat{R}^{\mathrm{T}}. \tag{5.5.55}$$

Equivalently,

$$\widehat{R} \exp\left(-\frac{i}{2}\omega_{ab}\widehat{J}^{ab}\right)\widehat{R}^{\mathrm{T}} = \exp\left(-\frac{i}{2}\omega_{ab}\widehat{J}'^{ab}\right), \tag{5.5.56}$$

with

$$\begin{aligned} \widehat{J}'^{mn} &= \widehat{J}^{ab}\widehat{R}^{am}\widehat{R}^{bn}, \\ \widehat{J}' &= \widehat{R}^{\mathrm{T}} \cdot \widehat{J} \cdot \widehat{R}. \end{aligned} \tag{5.5.57}$$

To see this, we first note from Taylor expansion and $\widehat{R}\widehat{R}^{\mathrm{T}} = \mathbb{I}$ that

$$\widehat{R}\exp(\widehat{X})\widehat{R}^{\mathrm{T}} = \exp\left(\widehat{R} \cdot \widehat{X} \cdot \widehat{R}^{\mathrm{T}}\right). \tag{5.5.58}$$

Therefore we may employ eq. (5.5.43) to evaluate

$$-\frac{i}{2}\omega'_{ij}\left(\widehat{J}^{ij}\right)_{ab} \equiv -\frac{i}{2}\omega_{ij}\left(\widehat{R}\widehat{J}^{ij}\widehat{R}^{\mathrm{T}}\right)_{ab} = -\frac{1}{2}\omega_{ij}\widehat{R}_{am}\delta^i_{[m}\delta^j_{n]}\widehat{R}_{bn} \tag{5.5.59}$$

$$= -\widehat{R}_{am}\widehat{R}_{bn}\omega_{mn} \tag{5.5.60}$$

On the other hand,

$$-\frac{i}{2}\omega'_{ij}\left(\widehat{J}^{ij}\right)_{ab} = -\frac{1}{2}\omega'_{ij}\delta^i_{[a}\delta^j_{b]} \tag{5.5.61}$$

$$= -\omega'_{ab}. \tag{5.5.62}$$

Comparing equations (5.5.60) and (5.5.62), we arrive at eq. (5.5.54).

As an example of this transformation property, we may in fact extract the generator that rotates the $(i,j)-$plane from eq. (5.5.50). That is, since $\widehat{R}(\theta) = \exp(-i\theta\widehat{J}^{ij})$,

$$i\partial_\theta \widehat{R}^{ab}(\theta = 0) = (\widehat{J}^{ij})^{ab} = -i\widehat{e}^a_{[i}\widehat{e}^b_{j]}; \tag{5.5.63}$$

which implies the matrix describing the counterclockwise rotation of the $(i,j)-$plane by angle θ is the exponential

$$\widehat{R}_{(i,j)}(\theta) = \exp\left(-\theta\widehat{e}_{[i}\widehat{e}^{\mathrm{T}}_{j]}\right). \tag{5.5.64}$$

(This discussion is, of course, the matrix version of equations (5.5.49) and (5.5.48).) Had we chosen the Cartesian axes such that $\widehat{e}^a_i = \delta^a_i$ and $\widehat{e}^a_j = \delta^a_j$, we would recover eq. (5.5.43). Had we chosen a different set of Cartesian axes, on the other hand, the two may be related by a rotation:

$$\delta^a_i = \widehat{R}^{am}\widehat{e}^m_i, \tag{5.5.65}$$

$$\delta^a_j = \widehat{R}^{am}\widehat{e}^m_j; \tag{5.5.66}$$

so that, when eq. (5.5.57) is employed on eq. (5.5.63),

$$i(\widehat{J}'^{ij})^{ab} = (\widehat{R}\cdot\widehat{J}^{ij}\cdot\widehat{R}^{\mathrm{T}})^{ab} = \widehat{R}^{am}\widehat{e}^m_{[i}\widehat{R}^{bn}\widehat{e}^n_{j]} = \delta^a_{[i}\delta^b_{j]} \tag{5.5.67}$$

– yielding eq. (5.5.43).

Summary Rotation matrices in $D-$space – defined to be length-preserving linear transformations continuously connected to the identity – are the exponential of anti-symmetric $D \times D$ matrices. These anti-symmetric matrices may be chosen in such a manner that they 'generate' an infinitesimal rotation on the (x^a, x^b)-plane, for a fixed and distinct pair $1 \leq a, b \leq D$. There are $D(D-1)/2$ such basis anti-symmetric matrices, corresponding to the $\binom{D}{2}$ ways of choosing a 2D plane formed by a pair of Cartesian axes.

Rotation Operators Acting on Position Eigenkets We now turn to the eigenkets of the position operators $\{|\vec{x}\rangle\}$. Let us implement rotation via

$$D(\widehat{R})\,|\vec{x}\rangle \equiv \exp\left(-\frac{i}{2}\omega_{ab}J^{ab}\right)|\vec{x}\rangle = \left|\widehat{R}\vec{x}\right\rangle, \tag{5.5.68}$$

where $D(\widehat{R})$ is now the linear operator associated with the rotation matrix $\widehat{R}$ in eq. (5.5.42). Now, we must have, for two rotation matrices $\widehat{R}_1$ and $\widehat{R}_2$,

$$D(\widehat{R}_1)D(\widehat{R}_2)\,|\vec{x}\rangle = D(\widehat{R}_1)\left|\widehat{R}_2\vec{x}\right\rangle = \left|\widehat{R}_1\widehat{R}_2\vec{x}\right\rangle = D(\widehat{R}_1\widehat{R}_2)\,|\vec{x}\rangle. \tag{5.5.69}$$

Since $|\vec{x}\rangle$ was arbitrary, we have the product rule

$$D(\widehat{R}_1)D(\widehat{R}_2) = D(\widehat{R}_1\widehat{R}_2). \tag{5.5.70}$$

Now, according to the discourse enveloping equations (5.1.44) through (5.1.49), the product of linear operators continuously connected to the identity is determined by the (nested) commutators of its generators. The latter, in turn, is completely determined by the Lie Algebra of the basis generators (cf. eq. (5.1.45)). On the other hand, eq. (5.5.70) tells us

$$\exp\left(-\frac{i}{2}\omega_{ab}J^{ab}\right)\exp\left(-\frac{i}{2}\omega'_{ab}J^{ab}\right) = \exp\left(-\frac{i}{2}\omega''_{ab}J^{ab}\right); \tag{5.5.71}$$

where ω_{ab} are the rotation angles describing $\widehat{R}_1(\omega)$, ω'_{ab} are those describing $\widehat{R}_2(\omega')$, and ω''_{ab} are those describing their product $(\widehat{R}_1\widehat{R}_2)(\omega'')$. One way to guarantee eq. (5.5.71) holds, is therefore to ensure the operators $\{J^{ab}\}$ obey the *same* Lie algebra as their matrix counterparts $\{\widehat{J}^{ab}\}$.

Problem 5.60. Rotating Momenta If $\left|\vec{k}\right\rangle$ is the eigenket of the momentum operator $\vec{P}$ in Cartesian coordinates, show that

$$D(\widehat{R})\left|\vec{k}\right\rangle = \left|\widehat{R}\cdot\vec{k}\right\rangle. \tag{5.5.72}$$

Hint: Insert a complete set of position eigenkets. □

Problem 5.61. Lie Algebra of SO$_D$ Use the choice of basis $\{\widehat{J}^{ab}\}$ in eq. (5.5.43) to argue there must a basis $\{J^{ab}\}$ such that

$$\left[J^{kl}, J^{mn}\right] = -i\left(\delta^{k[m}J^{n]l} - \delta^{l[m}J^{n]k}\right). \tag{5.5.73}$$

That the generators do not commute indicates rotations for $D > 2$ do not, in general, commute: $\widehat{R}_1\widehat{R}_2 \neq \widehat{R}_2\widehat{R}_1$. (The anti-symmetrization symbol means, for e.g., $T^{[ij]} = T^{ij} - T^{ji}$.) □

Group Representations This is a good place to highlight, the product rule in eq. (5.5.70) does not only apply to rotations nor to linear operators acting only on position eigenkets. More generally, we may motivate the notation of *group representations* as follows. To apply the notion of translations, rotations, Lorentz boosts, parity flips, and more general group operations to quantum states $\{|\psi\rangle\}$, there needs to be a function $D(\cdot)$ of these translation/rotation matrices/Lorentz boost matrices/group elements (which we will simply denote here as $g_1, g_2, \dots$) that converts them into linear operators acting on these $\{|\psi\rangle\}$. For these $D(g_1), D(g_2), \dots$ to preserve the notion of translations, rotations, Lorentz boosts, parity flips, etc., they must preserve the group multiplication rules of the original group. If $g_1 \cdot g_2 = g_3$, then we must have for all such group products

$$D(g_1)D(g_2) = D(g_3) = D(g_1 g_2). \tag{5.5.74}$$

Such a map that preserves group multiplication rules is known as a *group homomorphism*. For instance, in the case of 2D rotations, we defined $D(\widehat{R}(\phi)) \equiv e^{i\phi}$. One may readily check

$$e^{i\phi}e^{i\phi'} = D(\widehat{R}(\phi))\cdot D(\widehat{R}(\phi')) = D(\widehat{R}(\phi)\cdot\widehat{R}(\phi')) = D(\widehat{R}(\phi+\phi')) = e^{i(\phi+\phi')}. \tag{5.5.75}$$

Furthermore, a map that converts these original group elements $\{g_1, g_2, \dots\}$ into linear operators $\{D(g_1), D(g_2), \dots\}$ acting on some vector space, is known as a *group representation*. Note that, if e is the identity group element (obeying $e \cdot g = g$ for arbitrary group element g), it must map into the identity linear operator $\mathbb{I}$,

$$D(e) = \mathbb{I}. \tag{5.5.76}$$

This guarantees that, by eq. (5.5.74), $\mathbb{I} \cdot D(g) = D(e)D(g) = D(e \cdot g) = D(g)$ for arbitrary group element g.

Additionally, since the $\{g_1, g_2, \dots\}$ are assumed to have inverses, the linear operators themselves must be invertible as well. If g is a group element and g^{-1} is its inverse, we have

$$D(g^{-1}) = D(g)^{-1}; \tag{5.5.77}$$

so that, by equations (5.5.74) and (5.5.76), $D(g^{-1})D(g) = D(g^{-1}g) = D(e) = \mathbb{I}$.

For group homomorphisms of elements continuously connected to the identity operator, for them to be *faithful* representations – i.e., with no loss of information on the original group – we see that their corresponding generators must obey the same Lie Algebra as that of original group generators.

Finally, if $\{D(g)\}$ are matrices, we may ask if $\{D(g)^*\}$ are also a valid representation. The answer is yes: $D(g_1)^* D(g_2)^* = (D(g_1)D(g_2))^* = D(g_1 g_2)^*$. If $D(g)$ and $D(g)^*$ can be related via the same change-of-basis for all g, $D(g) = U D(g)^* U^{-1}$, however, we regard them to be equivalent.

Unitary $D(\widehat{R})$ and Hermitian J^{ab} We will next see that these $\{J^{ab}\}$ are Hermitian because $D(\widehat{R}) = \exp(-(i/2)\omega_{ab}J^{ab})$ is unitary, since the ω_{ab} are rotation angles and hence always real. To this end, we may discover from eq. (5.5.68) that

$$\left\langle \widehat{R}\vec{x}' \right| = \langle \vec{x}' | \, D(\widehat{R})^\dagger. \tag{5.5.78}$$

Together, we deduce

$$\begin{aligned} \langle \vec{x}' | \, D(\widehat{R})^\dagger D(\widehat{R}) \, | \vec{x} \rangle &= \left\langle \widehat{R}\vec{x}' \middle| \widehat{R}\vec{x} \right\rangle = \delta^{(D)}(\widehat{R}(\vec{x} - \vec{x}')) \qquad (5.5.79) \\ &= \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{|\det \partial(\widehat{R}x)^i/\partial x^a|^{\frac{1}{2}} |\det \partial(\widehat{R}x')^i/\partial x'^a|^{\frac{1}{2}}} = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{|\det \partial(\widehat{R}x)^i/\partial x^a|} = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{|\det \partial(\widehat{R}x')^i/\partial x'^a|}; \end{aligned}$$

where in the second line we have appealed to eq. (5.2.4). Moreover, we may compute

$$\det \frac{\partial(\widehat{R}^i{}_j x^j)}{\partial x^a} = \det \frac{\partial(\widehat{R}^i{}_j x'^j)}{\partial x'^a} = \det \widehat{R}^i{}_j \delta^j_a = \det \widehat{R}^i{}_a = 1, \tag{5.5.80}$$

if we recall eq. (5.5.9). At this point, we gather

$$\langle \vec{x}' | \, D(\widehat{R})^\dagger D(\widehat{R}) \, | \vec{x} \rangle = \delta^{(D)}(\vec{x}' - \vec{x}) = \langle \vec{x}' | \, \vec{x} \rangle \, . \tag{5.5.81}$$

Since $|\vec{x}\rangle$ and $|\vec{x}'\rangle$ are arbitrary, we have proven

$$D(\widehat{R})^\dagger D(\widehat{R}) = \mathbb{I}. \tag{5.5.82}$$

Problem 5.62. Can you argue that, when acting on the position eigenkets $\{|\vec{x}\rangle\}$,

$$D(\widehat{R})^\dagger = D(\widehat{R}^{\mathrm{T}})? \tag{5.5.83}$$

This is the generalization of eq. (5.4.26) to $D \geq 3$. Hint: Remember $\widehat{R}^{\mathrm{T}} = \widehat{R}^{-1}$. □

Problem 5.63. Rotating the Generators: Operator Version If $\exp(-(i/2)\omega_{ab}J^{ab})\,|\vec{x}\rangle = |\widehat{R}\cdot\vec{x}\rangle$, explain why equations (5.5.56) and (5.5.57) imply

$$\exp\left(-\frac{i}{2}\omega_{ab}J^{ab}\right)\exp\left(-\frac{i}{2}\omega'_{ab}J^{ab}\right)\exp\left(+\frac{i}{2}\omega_{ab}J^{ab}\right) \tag{5.5.84}$$

$$= \exp\left(-\frac{i}{2}\omega''_{ab}J^{ab}\right) = \exp\left(-\frac{i}{2}\omega_{ab}J''^{ab}\right); \tag{5.5.85}$$

where

$$\omega''_{ab} = \widehat{R}_{am}\widehat{R}_{bn}\omega_{mn}, \tag{5.5.86}$$

$$J''^{mn} = J^{ab}\widehat{R}^{am}\widehat{R}^{bn}. \tag{5.5.87}$$

In words: The generators of rotations are themselves rank-2 tensors under rotations. □

Problem 5.64. 'Orbital' Angular Momentum: Position Representation In this problem, we shall work out J^{ab} within the position representation.

According to eq. (5.5.35), an infinitesimal rotation may be implemented via the replacement

$$x^i \to \widehat{R}^i{}_j x^j = \left(\delta^i_j + \widehat{\Omega}^i{}_j + \dots\right)x^j, \tag{5.5.88}$$

where Ω is anti-symmetric; cf. eq. (5.5.38). (Here, the placement of indices on $\widehat{\Omega}$, i.e., up versus down, is unimportant.) In eq. (5.5.68), take

$$\widehat{\Omega} = -i\theta\widehat{J}^{ij}, \tag{5.5.89}$$

where J^{ij} is one of the basis anti-symmetric matrices in eq. (5.5.43); and let $f(\vec{x}) = \langle\vec{x}|\,f\rangle$ be an arbitrary function. Explain why the replacement in eq. (5.5.88) induces

$$f(\vec{x}) \to f(\vec{x}) + \theta\cdot(x^i\partial_j - x^j\partial_i)f(\vec{x}). \tag{5.5.90}$$

Next show that, upon an infinitesimal rotation generated by J^{ij} – now acting on the $\{|\vec{x}\rangle\}$ –

$$\langle\vec{x}|\,f\rangle \to \left(D(\widehat{R})\,|\vec{x}\rangle\right)^\dagger|f\rangle \tag{5.5.91}$$

$$= \langle\vec{x}|\,f\rangle + i\theta\left\langle\vec{x}\left|J^{ij}\right|f\right\rangle + \mathcal{O}(\theta^2). \tag{5.5.92}$$

We may therefore identify

$$\langle\vec{x}|\,J^{ij}\,|f\rangle = -i\left(x^i\partial_j - x^j\partial_i\right)\langle\vec{x}|\,f\rangle, \qquad \partial_j \equiv \frac{\partial}{\partial x^j}. \tag{5.5.93}$$

That is, these J^{ij} are the $D-$dimensional analogs of the 'orbital angular-momentum' operators in 3D space. Employing eq. (5.2.24), we may deduce from (5.5.93) that

$$J^{ij} = X^i P_j - X^j P_i, \tag{5.5.94}$$

where $\vec{X}$ and $\vec{P}$ are now, respectively, the position and momentum operators.

Can you verify that eq. (5.5.94) satisfies the Lie Algebra in eq. (5.5.73) through a direct calculation? Recall that the same Lie Algebra has to be satisfied for all representations of the group elements continuously connected to the identity, because it is the Lie Algebra that completely determines the product rule between any two such elements. □

Relation between $\vec{J}^2 \equiv (1/2)J^{ab}J^{ab}$ and $\vec{P}^2$ In §(9) below, we explain how to express $\vec{P}^2$, which in the position representation is the negative Laplacian, in any coordinate system. Practically speaking, the key is to first write down the Euclidean metric in eq. (5.5.19) in the desired coordinate system. For our case, we will focus on the $D-$dimensional spherical coordinate system $(r, \vec{\theta})$, which yields

$$\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}r^2 + r^2 H_{\mathrm{IJ}}\mathrm{d}\theta^{\mathrm{I}}\mathrm{d}\theta^{\mathrm{J}} \tag{5.5.95}$$

if we set the Cartesian coordinate vector to be equal to the radial distance r times an appropriately defined unit radial vector parameterized by $\vec{\theta}$: $x^i = r\widehat{r}(\vec{\theta})$. In particular,

$$r^2 H_{\mathrm{IJ}} = r^2\delta_{ij}\frac{\partial\widehat{n}^i}{\partial\theta^{\mathrm{I}}}\frac{\partial\widehat{n}^j}{\partial\theta^{\mathrm{J}}}. \tag{5.5.96}$$

Generically, given a metric

$$\mathrm{d}\ell^2 = g_{ij}\mathrm{d}x^i\mathrm{d}x^j \tag{5.5.97}$$

we may first compute its determinant $g \equiv \det g_{ij}$, its inverse g^{ij} (which satisfies $g^{ij}g_{jk} = \delta^i_k$), and its scalar Laplacian

$$\vec{\nabla}^2\psi = \frac{1}{\sqrt{g}}\partial_i\left(\sqrt{g}g^{ij}\partial_j\psi\right). \tag{5.5.98}$$

For instance, the metric in Cartesian coordinates is simply δ_{ij}, whose determinant is unity, inverse δ^{ij}, and Laplacian

$$\vec{\nabla}^2\psi = \delta^{ij}\partial_i\partial_j\psi. \tag{5.5.99}$$

Problem 5.65. If H denotes the determinant of H_{IJ} in eq. (5.5.95), show that the $D-$space Laplacian is

$$\vec{\nabla}^2\psi = \frac{1}{r^{D-1}}\partial_r\left(r^{D-1}\partial_r\psi\right) + \frac{1}{r^2}\vec{\nabla}^2_{\mathbb{S}^{D-1}}\psi; \tag{5.5.100}$$

where $\vec{\nabla}^2_{\mathbb{S}^{D-1}}$ is the Laplacian on the unit $(D-1)$-sphere (i.e., $r = 1$), namely

$$\vec{\nabla}^2_{\mathbb{S}^{D-1}}\psi = \frac{1}{\sqrt{H}}\partial_{\mathrm{I}}\left(\sqrt{H}H^{\mathrm{IJ}}\partial_{\mathrm{J}}\psi\right), \tag{5.5.101}$$

where the I and J indices run only over the angular coordinates $\{\theta^{\mathrm{I}}\}$. This result will be used in the next problem. □

Problem 5.66. $(2r^2)^{-1}J^{ab}J^{ab} = \vec{J}^2/r^2$ **and Laplacian on Sphere** Use eq. (5.5.93) to show that

$$\frac{1}{2}\left\langle \vec{x}\left|J^{ab}J^{ab}\right|\psi\right\rangle = \Big((D-1)x^a\partial_a + x^a x^b\partial_a\partial_b - \vec{x}^2\vec{\nabla}^2\Big)\left\langle\vec{x}\right|\psi\rangle. \tag{5.5.102}$$

Next, recall eq. (5.2.48) and show that

$$-\left\langle \vec{x}\left|\vec{P}^2 - \frac{1}{2r^2}J^{ab}J^{ab}\right|\psi\right\rangle = \frac{1}{r^{D-1}}\partial_r\left(r^{D-1}\partial_r\left\langle\vec{x}\right|\psi\rangle\right). \tag{5.5.103}$$

From this, identify $(1/2)J^{ab}J^{ab}$ as the negative Laplacian on the $(D-1)-$sphere:

$$\left\langle \vec{x}\left|\vec{J}^2\right|\psi\right\rangle \equiv \frac{1}{2}\left\langle\vec{x}\left|J^{ab}J^{ab}\right|\psi\right\rangle = -\vec{\nabla}^2_{\mathbb{S}^{D-1}}\left\langle\vec{x}\right|\psi\rangle. \tag{5.5.104}$$

Hints: You may need the result from the previous problem. Recognize too, from $x^i = r\widehat{r}^i(\vec{\theta})$,

$$r\partial_r = r\frac{\partial x^i}{\partial r}\partial_i = r\widehat{r}^i\partial_i = x^i\partial_i; \tag{5.5.105}$$

as well as (keeping in mind $\widehat{r}^a\partial_a\widehat{r}^b = 0$ – can you see why?)

$$x^i x^j\partial_i\partial_j\psi = r^2\partial_r^2\psi. \tag{5.5.106}$$

To reiterate: just as $-\vec{P}^2$ is the $D-$space Laplacian in Euclidean space, the $-(2r^2)^{-1}J^{ab}J^{ab}$ is its counterpart on the $(D-1)$-sphere of radius r. □

Generator of rotations in terms of angular derivatives Since J^{ab} 'generates' rotation, in the position representation they must correspond to strictly angular derivatives, for any radial ones would imply a moving off the surface of some constant radius – thereby violating the notation of rotation as length-preserving. To see this, we first assume it is possible to find angular coordinates $\vec{\theta}$ such that not only does the Cartesian coordinate vector take the form

$$x^i = r\widehat{r}^i(\vec{\theta}), \qquad \widehat{r}^i\widehat{r}^i = 1 \tag{5.5.107}$$

these angles are orthogonal in the sense that

$$\partial_{\mathrm{I}}\widehat{r}\cdot\partial_{\mathrm{J}}\widehat{r} = \delta_{ij}\partial_{\mathrm{I}}\widehat{r}^i\cdot\partial_{\mathrm{J}}\widehat{r}^j = H_{\mathrm{IJ}} \equiv \mathrm{diag}\left[H_{22}, H_{33}, \dots, H_{DD}\right]. \tag{5.5.108}$$

In other words, we assume the angular metric in eq. (5.5.96) is diagonal.

Another consequence of eq. (5.5.107) follows from differentiating $\widehat{r}^i\widehat{r}^i = 1$ with respect to any of one of the angles is

$$\widehat{r}^i\partial_{\mathrm{I}}\widehat{r}^i = 0. \tag{5.5.109}$$

The Jacobian $\partial x^i/\partial(r,\theta)^a$ takes the form

$$\frac{\partial x^i}{\partial r} = \widehat{r}^i \qquad \text{and} \qquad \frac{\partial x^i}{\partial\theta^{\mathrm{I}}} = r\frac{\partial\widehat{r}^i}{\partial\theta^{\mathrm{I}}}. \tag{5.5.110}$$

Let us now observe, through the chain rule

$$\frac{\partial x^i}{\partial(r,\vec{\theta})^a}\frac{\partial(r,\vec{\theta})^a}{\partial x^j}=\frac{\partial x^i}{\partial x^j}=\delta^i_j, \tag{5.5.111}$$

the matrix $\partial(r,\vec{\theta})^a/\partial x^j$ is simply the inverse of $\partial x^i/\partial(r,\vec{\theta})^a$; namely,

$$\frac{\partial(r,\vec{\theta})^a}{\partial x^j}=\left(\left(\frac{\partial\vec{x}}{\partial(r,\vec{\theta})}\right)^{-1}\right)^a{}_j. \tag{5.5.112}$$

It has components

$$\frac{\partial r}{\partial x^i}=\widehat{r}^i \qquad \text{and} \qquad \frac{\partial\theta^{\rm I}}{\partial x^i}=\frac{H^{\rm IJ}}{r}\frac{\partial\widehat{r}^i}{\partial\theta^{\rm J}}; \tag{5.5.113}$$

where the inverse angular metric is defined through the relation

$$H^{\rm IK}H_{\rm KJ}=\delta^{\rm I}{}_{\rm J}. \tag{5.5.114}$$

To see this, we simply check that our expressions for $\partial r/\partial x^i$ and $\partial\theta^{\rm I}/\partial x^i$ do indeed yield the components of the inverse of $\partial x^i/\partial(r,\theta)^a$; namely,

$$\frac{\partial r}{\partial r}=\frac{\partial r}{\partial x^i}\frac{\partial x^i}{\partial r}=\widehat{r}^i\widehat{r}^i=1; \tag{5.5.115}$$

$$\frac{\partial r}{\partial\theta^{\rm I}}=\frac{\partial r}{\partial x^i}\frac{\partial x^i}{\partial\theta^{\rm I}}=r\widehat{r}^i\partial_{\rm I}\widehat{r}^i=0; \tag{5.5.116}$$

$$\frac{\partial\theta^{\rm I}}{\partial r}=\frac{\partial\theta^{\rm I}}{\partial x^i}\frac{\partial x^i}{\partial r}=\frac{H^{\rm IJ}}{r}\frac{\partial\widehat{r}^i}{\partial\theta^{\rm J}}\widehat{r}^i=0; \tag{5.5.117}$$

and

$$\frac{\partial\theta^{\rm I}}{\partial\theta^{\rm J}}=\frac{\partial\theta^{\rm I}}{\partial x^i}\frac{\partial x^i}{\partial\theta^{\rm J}}=\frac{H^{\rm IK}}{r}\partial_{\rm K}\widehat{r}^i\cdot r\partial_{\rm J}\widehat{r}^i=H^{\rm IK}H_{\rm KJ}=\delta^{\rm I}{}_{\rm J}. \tag{5.5.118}$$

Hence, from eq. (5.5.113),

$$\left\langle r,\vec{\theta}\middle|J^{ab}\middle|\psi\right\rangle=-ix^{[a}\partial_{b]}\left\langle\vec{x}\middle|\psi\right\rangle=-i\left(x^{[a}\frac{\partial r}{\partial x^{b]}}\partial_r+x^{[a}\frac{\partial\theta^{\rm I}}{\partial x^{b]}}\partial_{\rm I}\right)\left\langle\vec{x}\middle|\psi\right\rangle \tag{5.5.119}$$

$$=-i\left(x^{[a}\widehat{r}^{b]}\partial_r+x^{[a}\frac{\partial\theta^{\rm I}}{\partial x^{b]}}\partial_{\rm I}\right)\left\langle\vec{x}\middle|\psi\right\rangle. \tag{5.5.120}$$

Because $r\widehat{r}=\vec{x}$ (cf. eq. (5.5.107)), the left term in the last line is zero because $r\widehat{r}^{[a}\widehat{r}^{b]}=0$. From eq. (5.5.113), we now arrive at the spherical coordinates analog of eq. (5.5.93):

$$\left\langle r,\vec{\theta}\middle|J^{ab}\middle|\psi\right\rangle=-iH^{\rm IJ}\left(\widehat{r}^a\frac{\partial\widehat{r}^b}{\partial\theta^{\rm I}}-\widehat{r}^b\frac{\partial\widehat{r}^a}{\partial\theta^{\rm I}}\right)\frac{\partial}{\partial\theta^{\rm J}}\left\langle r,\vec{\theta}\middle|\psi\right\rangle. \tag{5.5.121}$$

Problem 5.67. Casimir Operators Are Invariants Casimir operators are operators that commute with all the generators of a given group. Here, we see that $(1/2)J^{ij}J^{ij}$ is a Casimir operator. Show that J^{ab} commutes with $(1/2)J^{ij}J^{ij}$, i.e.

$$\left[J^{ab}, \frac{1}{2}J^{ij}J^{ij}\right] = 0. \tag{5.5.122}$$

Hint: Remember eq. (4.3.86).

Eq. (5.5.122) is, in fact, a statement that $(1/2)J^{ab}J^{ab}$ is a rotation scalar; i.e., invariant under SO_D. By referring to Problem (5.63), can you explain why? Hint: Consider the infinitesimal version of eq. (5.5.84). $\square$

Spherical Harmonics in $D-$dimensions The Poisson equation of Newtonian gravity or Coulomb's law reads

$$\vec{\nabla}^2\psi = 4\pi\rho, \tag{5.5.123}$$

where ρ is either mass or charge density. Suppose we were solving ψ away from the source at $\vec{x}$, where $\rho(\vec{x}) = 0$. If we choose the origin to be located nearby, so that $\rho(\vec{x} = \vec{0}) = 0$ too, then we may perform a Taylor expansion

$$\psi(\vec{x}) = \sum_{\ell=0}^{+\infty} \frac{x^{i_1}\dots x^{i_\ell}}{\ell!}\psi_{i_1\dots i_\ell}, \tag{5.5.124}$$

$$\psi_{i_1\dots i_\ell} \equiv \partial_{i_1}\dots\partial_{i_\ell}\psi(\vec{x} = \vec{0}). \tag{5.5.125}$$

Since $\rho(\vec{x}) = 0$ in this region, eq. (5.5.123) reduces to $\vec{\nabla}^2\psi = 0$. Eq. (5.5.124) inserted into eq. (5.5.123) must yield the statement that, for a fixed ℓ but with summation over the ℓ indices $\{i_1, \dots, i_\ell\}$ still in force,

$$\vec{\nabla}^2_{\vec{x}}\left(x^{i_1}x^{i_2}\dots x^{i_{\ell-1}}x^{i_\ell}\right)\psi_{i_1\dots i_\ell} = 0. \tag{5.5.126}$$

Notice $x^{i_1}\dots x^{i_\ell}\psi_{i_1\dots i_\ell}$ is a homogeneous polynomial of degree ℓ. (Here, a homogeneous polynomial of degree ℓ, P_ℓ, is a polynomial built of out the Cartesian components of $\vec{x}$ such that, under the re-scaling $\vec{x} \to \lambda\vec{x}$, the polynomial scales as $P_\ell \to \lambda^\ell P_\ell$.) Therefore, for each ℓ, the solution of the vacuum Poisson equation in $D-$dimensions involves eq. (5.5.126): homogeneous polynomials of degree ℓ annihilated by the Laplacian – this is often the starting definition of the *spherical harmonics*.

Problem 5.68. Recall the space of polynomials of degree less than or equal to ℓ forms a vector space. Is the space of homogeneous polynomials of degree $\leq \ell$ a vector space? What about the space of polynomials of degree $\leq \ell$ satisfying eq. (5.5.126)? Hint: Remember the discussion at the end of §(4.1). $\square$

If we employ spherical coordinates in $D-$dimensions,

$$x^i = r\widehat{r}^i(\vec{\theta}), \qquad \vec{\theta} = (\theta^1, \dots, \theta^{D-1}); \tag{5.5.127}$$

then eq. (5.5.126) takes the form

$$\vec{\nabla}^2\left(r^\ell Y(\vec{\theta})\right) = 0; \tag{5.5.128}$$

where the angular portion arises from

$$Y(\vec{\theta}) = \widehat{r}^{i_1}\dots\widehat{r}^{i_\ell}\psi_{i_1\dots i_\ell}. \tag{5.5.129}$$

Problem 5.69. Eigensystem on the $(D-1)$-sphere Show that eq. (5.5.128) leads to the eigenvector/value equation

$$\vec{\nabla}^2_{\mathbb{S}^{D-1}}Y(\vec{\theta}) = -\ell(\ell+D-2)Y(\vec{\theta}). \tag{5.5.130}$$

These angular spherical harmonics, for $D=3$, are usually denoted as $Y_\ell^m(\theta,\phi)$, where $\ell = 0,1,2,\dots$ and $-\ell \leq m \leq +\ell$. We shall examine them in the next section. $\square$

Problem 5.70. Second Homogeneous Solution Apply eq. (5.5.130) to show that the second radial solution to $\vec{\nabla}^2(R(r)Y(\vec{\theta})) = 0$ for $r \neq 0$ is $R(r) = 1/r^{\ell+D-2}$; namely,

$$\vec{\nabla}^2\left(\frac{Y(\vec{\theta})}{r^{\ell+D-2}}\right) = 0. \tag{5.5.131}$$

Because these $\{Y(\vec{\theta})\}$ are eigenvectors of a Hermitian operator (i.e., the Laplacian on the sphere), they form a complete basis for functions on $\mathbb{S}^{D-1}$. Together with eq. (5.5.128), the general homogeneous solution to Laplace's equation $\vec{\nabla}^2\psi = 0$ away from the origin is therefore

$$\psi(\vec{x}; r\neq 0) = \sum_{\ell,m_1,\dots,m_{D-2}}\left(A_\ell^{\{m_{\mathrm{I}}\}}r^\ell + \frac{B_\ell^{\{m_{\mathrm{I}}\}}}{r^{\ell+D-2}}\right)Y_\ell^{\{m_{\mathrm{I}}\}}(\vec{\theta}), \tag{5.5.132}$$

for constants $\{A_\ell^{\{m_{\mathrm{I}}\}}\}$ and $\{B_\ell^{\{m_{\mathrm{I}}\}}\}$ and I running over the $\{1,\dots,D-2\}$ labels of the azimuthal eigenvalues $\{m_{\mathrm{I}}\}$. $\square$

5.6 Rotations in 3 Spatial Dimensions

5.6.1 Lie Algebra of Generators

We will now focus on rotations in 3D, where rotating the (i,j) plane is equivalent to rotating space about the k-axis perpendicular to it. Such a 'dual' perspective is unique to 3D; because there are more than one axes perpendicular to (i,j) in higher dimensions. More quantitatively, this statement may be captured by utilizing the fully anti-symmetric 3D Levi-Civita symbol $\epsilon_{ijk} = \epsilon^{ijk}$, with $\epsilon_{123} = \epsilon^{123} \equiv 1$. Specifically, let us define the Hermitian operator

$$J^i \equiv \frac{1}{2}\epsilon^{ijk}J^{jk}, \tag{5.6.1}$$

which, by multiplying both sides with ϵ^{mni} and using the result

$$\epsilon^{aij}\epsilon^{amn} = \delta^i_{[m}\delta^j_{n]} = \delta^i_m\delta^j_n - \delta^i_n\delta^j_m, \tag{5.6.2}$$

[32]is equivalent to

$$J^{ij} = \epsilon^{ijm} J^m; \tag{5.6.3}$$

so that

$$-\frac{i}{2}\omega_{ij} J^{ij} = -i\theta^i J^i \qquad \Leftrightarrow \qquad \theta^i = \frac{1}{2}\omega_{ab}\epsilon^{abi}. \tag{5.6.4}$$

Hence,

$$\theta^1 = \frac{1}{2}\epsilon^{123}\omega_{23} + \frac{1}{2}\epsilon^{132}\omega_{32} = \omega_{23}, \tag{5.6.5}$$

$$\theta^2 = \frac{1}{2}\epsilon^{213}\omega_{13} + \frac{1}{2}\epsilon^{231}\omega_{31} = -\omega_{13}, \tag{5.6.6}$$

$$\theta^3 = \frac{1}{2}\epsilon^{312}\omega_{12} + \frac{1}{2}\epsilon^{321}\omega_{21} = \omega_{12}. \tag{5.6.7}$$

Recall that J^{23} generates rotations of the $(2,3)$ plane, and ω_{23} is the corresponding angle (for e.g., eq. (5.5.49)); we see that $-i\theta^1 J^1$ can be thought of as generating a rotation around the $1-$axis because it actually generates rotations around the $(2,3)$ plane. Keeping in mind equations (5.6.1) and (5.6.4), when $D=3$, we thus specialize the $D-$dimensional result in eq. (5.5.42) as

$$\text{3D}: \widehat{R}(\vec{\theta}) = \exp\left(-\frac{i}{2}\omega_{ab}J^{ab}\right) = \exp\left(-i\vec{\theta}\cdot\vec{J}\right). \tag{5.6.8}$$

It is worth reiterating, that this rotation operator can be written either in terms of J^i or J^{ab} is unique to 3D. Using eq. (5.5.49), the rotations around each Cartesian axes can be constructed explicitly:

$$e^{-i\theta J^{12}} = e^{-i\theta J^3} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.6.9}$$

$$e^{-i\theta J^{13}} = e^{+i\theta J^2} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{5.6.10}$$

$$e^{-i\theta J^{23}} = e^{-i\theta J^1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}. \tag{5.6.11}$$

Euler Angles and 3D Rotations What is the most general rotation matrix? Imagine a perfectly rigid body in 3D space, defined as one where each point within it always lie at a fixed distance from every other point. **YZ: Incomplete.**

Problem 5.71. Rotating the rotation axis Show that the 3D version of eq. (5.5.54) is:

$$\widehat{R}\exp\left(-i\vec{\theta}\cdot\vec{J}\right)\widehat{R}^{\mathrm{T}} = \exp\left(-i\vec{\theta}'\cdot\vec{J}\right), \qquad \vec{\theta}\cdot\vec{J} \equiv \theta_a J^a, \tag{5.6.12}$$

[32]The proof of eq. (5.6.2) can be found in the discussion following eq. (9.6.34) below.

where

$$\theta'_a \equiv \widehat{R}_{ab}\theta_b. \tag{5.6.13}$$

In the other words, a change-of-basis through a rotation $\widehat{R}$ amounts to rotating the angles $\vec{\theta}$. $\square$

Problem 5.72. Explicit Form of Rotation-Around-Axis Let us employ the insight gained from eq. (5.6.12) to compute the explicit form of the 3×3 rotation matrix

$$\widehat{R}_{\widehat{n}}(\theta) \equiv \exp\left(-i\theta\ \widehat{n}(\alpha,\beta)\cdot\vec{J}\right) \tag{5.6.14}$$

describing counter-clockwise rotation by θ about the unit axis

$$\widehat{n}(\alpha,\beta) \equiv (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)\,. \tag{5.6.15}$$

Show that the answer is

$$\begin{aligned}
&\widehat{R}_{\widehat{n}(\alpha,\beta)}(\theta) \\
&= \left[\begin{array}{ccc}
\frac{1}{4}c_\theta\left(3 - 2s_\alpha^2 c_{2\beta} + c_{2\alpha}\right) + s_\alpha^2 c_\beta^2 & s_\alpha^2 s_{2\beta} s_{\theta/2}^2 - c_\alpha s_\theta & s_\alpha s_\beta s_\theta + s_{2\alpha} c_\beta s_{\theta/2}^2 \\
s_\alpha^2 s_{2\beta} s_{\theta/2}^2 + c_\alpha s_\theta & \frac{1}{4}c_\theta\left(3 + 2s_\alpha^2 c_{2\beta} + c_{2\alpha}\right) + s_\alpha^2 s_\beta^2 & s_{2\alpha} s_\beta s_{\theta/2}^2 - s_\alpha c_\beta s_\theta \\
s_{2\alpha} c_\beta s_{\theta/2}^2 - s_\alpha s_\beta s_\theta & s_{2\alpha} s_\beta s_{\theta/2}^2 + s_\alpha c_\beta s_\theta & s_\alpha^2 c_\theta + c_\alpha^2
\end{array}\right];
\end{aligned} \tag{5.6.16}$$

where s and c are, respectively, sine and cosine.

Hints: First start with $\exp(-i\theta J^3)$; namely, rotation about the $3-$axis by angle θ. Then note that $\widehat{e}_3$, the unit vector parallel to the $3-$axis may be rotated to $\widehat{n}(\alpha,\beta)$ by first rotating it around the $2-$axis by α, followed by rotating the ensuing result around the $3-$axis by β. That is, write down the appropriate $\widehat{R}$ in eq. (5.6.12) by decomposing it into $\widehat{R} = \widehat{R}_{\widehat{e}_3}(\alpha)\cdot\widehat{R}_{\widehat{e}_2}(\beta)$. $\square$

There is an alternate means to arrive at eq. (5.6.16) but without specializing to a particular coordinate system. The key geometric insight is that rotating a position vector $\vec{x}$ around some axis $\widehat{n}$ means preserving the former's components along $\widehat{n}$ while rotating its components perpendicular to $\widehat{n}$. But this latter operation is really a 2D one, since the space orthogonal to $\widehat{n}$ is a 2D plane. Borrowing the discussion following eq. (5.4.8), we may thus gather, $\vec{x}$ rotated counter-clockwise around the axis $\widehat{n}$ by angle θ is given by

$$\widehat{R}_{\widehat{n}}(\theta)^{ij} x^j = \widehat{n}^i(\widehat{n}\cdot\vec{x}) + \left(\cos\theta\cdot\delta^{ij} - \sin\theta\cdot\epsilon^{ijl}\widehat{n}^l\right)\left(\delta^{jk} - \widehat{n}^j\widehat{n}^k\right)x^k, \tag{5.6.17}$$

$$= \widehat{n}^i(\widehat{n}^j\cdot x^j) + \cos\theta\cdot\left(\delta^{ij} - \widehat{n}^i\widehat{n}^j\right)x^j - \sin\theta\cdot\epsilon^{ijl}\widehat{n}^l x^j; \tag{5.6.18}$$

where the $\mathbb{I}_{3\times3} - \widehat{n}\widehat{n}$ is the projector that extracts the components of $\vec{x}$ orthogonal to $\widehat{n}$ – recall eq. (4.2.33). In vector notation,

$$\widehat{R}_{\widehat{n}}(\theta)\cdot\vec{x} = \widehat{n}(\widehat{n}\cdot\vec{x}) + \cos\theta\cdot(\vec{x} - \widehat{n}(\widehat{n}\cdot\vec{x})) + \sin\theta\cdot(\widehat{n}\times\vec{x})\,. \tag{5.6.19}$$

Since $\vec{x}$ is arbitrary, we may read off the rotation matrix as

$$\widehat{R}_{\widehat{n}}(\theta)^{ij} = \widehat{n}^i\widehat{n}^j + \cos\theta\cdot\left(\delta^{ij} - \widehat{n}^i\widehat{n}^j\right) - \sin\theta\cdot\epsilon^{ijk}\widehat{n}^k. \tag{5.6.20}$$

Problem 5.73. Explain why eq. (5.6.20) is the 3D version of eq. (5.5.50). Hint: One key step is to explain why $\epsilon^{ijk}\widehat{n}^k$ in 3D is to be identified with the $\epsilon^{\text{AB}}\widehat{e}^i_{\text{A}}\widehat{e}^j_{\text{B}}$ in the general case. $\square$

Problem 5.74. Let $\widehat{a}$ and $\widehat{b}$ be linearly independent unit vectors. If $\widehat{n} \equiv (\widehat{a}\times\widehat{b})/|\widehat{a}\times\widehat{b}|$, demonstrate that

$$\widehat{R}^{ij}_{\widehat{n}} \equiv \widehat{n}^i\widehat{n}^j + (\widehat{a}\cdot\widehat{b})\cdot\left(\delta^{ij} - \widehat{n}^i\widehat{n}^j\right) - |\widehat{a}\times\widehat{b}|\cdot\epsilon^{ijk}\widehat{n}^k \tag{5.6.21}$$

is not only a rotation matrix, it rotates $\widehat{a}$ to $\widehat{b}$; namely, verify $\widehat{R}\cdot\widehat{a} = \widehat{b}$. What is the matrix that rotates $\widehat{b}$ to $\widehat{a}$? $\square$

Global Structure of SO$_3$

Algebra of 3D Rotation Generators Next, we invoke eq. (5.6.2) to deduce

$$\vec{J}^2 \equiv J^a J^a = \frac{1}{4}\epsilon^{amn}\epsilon^{aij}J^{mn}J^{ij} \tag{5.6.22}$$

$$= \frac{1}{4}\left(\delta^m_i\delta^n_j - \delta^m_j\delta^n_i\right)J^{mn}J^{ij} = \frac{1}{2}J^{mn}J^{mn}. \tag{5.6.23}$$

In the previous section, we have already demonstrated that $D(\widehat{R})$ is unitary, and hence $\{J^{ab}\}$ and $\{J^a\}$ are Hermitian operators, with real eigenvalues and a complete set of eigenkets. We will now attempt to perform a systematic analysis of the eigensystem of the $\{J^a\}$ in 3D. The following problem will provide the key ingredient.

Problem 5.75. Lie Algebra of Rotation Generators in 3D Show that

$$\left[J^a, J^b\right] = i\epsilon^{abc}J^c \tag{5.6.24}$$

and

$$\left[J^a, \vec{J}^2\right] = 0, \qquad \vec{J}^2 \equiv J^i J^i. \tag{5.6.25}$$

Hint: Recall equations (5.5.73) and (5.5.122). Eq. (5.6.24) may also be tackled by first utilizing eq. (5.5.43) to prove that the matrix generator is

$$(\widehat{J}^i)_{ab} = -i\epsilon^{iab} \qquad \Leftrightarrow \qquad i(\widehat{J}^i)_{ab} = \epsilon^{iab}. \tag{5.6.26}$$

Compare equations (5.4.42) and (5.6.26). $\square$

Eigenvalues of $\vec{J}^2$ and J^3 from Ladder Operators in 3D According to eq. (5.6.24), the $\{J^a\}$ do not commute among themselves. However, eq. (5.6.25) tells us we may choose $\vec{J}^2$ and one of the $\{J^a\}$ as a pair of mutually compatible observables. As it is customary to do so, we shall choose to simultaneously diagonalize $\vec{J}^2$ and J^3. Denote the simultaneous eigenket of $\vec{J}^2$ and J^3 as $|\lambda_J, m\rangle$.

$$\vec{J}^2|\lambda_J, m\rangle = \lambda_J|\lambda_J, m\rangle \qquad \text{and} \qquad J^3|\lambda_J, m\rangle = m|\lambda_J, m\rangle \tag{5.6.27}$$

To this end, let us define the raising J^+ and lowering J^- operators

$$J^{\pm} \equiv J^1 \pm iJ^2; \tag{5.6.28}$$

and compute, using the linearity of the commutator and the Lie Algebra of eq. (5.6.24),

$$\left[J^3, J^\pm\right] = \left[J^3, J^1\right] \pm i\left[J^3, J^2\right] \tag{5.6.29}$$
$$= -i\epsilon^{132}J^2 \mp i^2\epsilon^{231}J^1 = \pm\left(J^1 \pm iJ^2\right). \tag{5.6.30}$$

In other words,

$$\left[J^3, J^\pm\right] = \pm J^\pm. \tag{5.6.31}$$

These are dubbed ladder or raising/lower operators because the $J^\pm$ acting on $|\lambda_J, m\rangle$ will raise or lower the m by unity.

$$J^\pm|\lambda_J, m\rangle = c_{m\pm1}|\lambda_J, m\pm1\rangle \tag{5.6.32}$$

To see this, we employ eq. (5.6.31),

$$\begin{aligned} J^3J^\pm|\lambda_J, m\rangle &= \left(J^3J^\pm - J^\pm J^3 + J^\pm J^3\right)|\lambda_J, m\rangle \\ &= \left(\left[J^3, J^\pm\right] + J^\pm J^3\right)|\lambda_J, m\rangle \\ &= (m\pm1)J^\pm|\lambda_J, m\rangle. \end{aligned} \tag{5.6.33}$$

Problem 5.76. Show that

$$\left[J^+, J^-\right] = 2J^3. \tag{5.6.34}$$

□

Next, let us prove that, for a fixed λ_J, there is a maximum and minimum eigenvalue of J^3. We shall use the non-negative character of the norm to do so. Specifically,

$$\left(J^\pm|\lambda_J, m\rangle\right)^\dagger J^\pm|\lambda_J, m\rangle \geq 0. \tag{5.6.35}$$

Now,

$$(J^\pm)^\dagger J^\pm = (J^1 \mp iJ^2)(J^1 \pm iJ^2) \tag{5.6.36}$$
$$= (J^1)^2 + (J^2)^2 \pm i(J^1J^2 - J^2J^1) \tag{5.6.37}$$
$$= (J^1)^2 + (J^2)^2 \pm i^2J^3 \tag{5.6.38}$$
$$= (J^1)^2 + (J^2)^2 + (J^3)^2 - (J^3)^2 \mp J^3 = \vec{J}^2 - (J^3)^2 \mp J^3. \tag{5.6.39}$$

Therefore, their average is

$$\frac{1}{2}(J^+)^\dagger J^+ + \frac{1}{2}(J^-)^\dagger J^- = \vec{J}^2 - (J^3)^2; \tag{5.6.40}$$

and we have

$$\frac{1}{2}\left(J^+|\lambda_J, m\rangle\right)^\dagger J^+|\lambda_J, m\rangle + \frac{1}{2}\left(J^-|\lambda_J, m\rangle\right)^\dagger J^-|\lambda_J, m\rangle \geq 0 \tag{5.6.41}$$
$$\langle\lambda_J, m|\,\vec{J}^2 - (J^3)^2\,|\lambda_J, m\rangle \geq 0 \tag{5.6.42}$$

$$\lambda_J \geq m^2. \tag{5.6.43}$$

If there were no m_{max}, eq. (5.6.33) tells us we may keep applying more and more powers of J^+ to obtain an ever increasing m^2 – but that would certainly be greater than λ_J at some point, contradicting eq. (5.6.43). By applying more and more powers of J^-, we may similarly argue there has to be a m_{min}, otherwise m^2 will eventually violate eq. (5.6.43) again. These considerations also tell us,

$$J^+ \left|\lambda_J, m_{\text{max}}\right\rangle = 0; \tag{5.6.44}$$

for otherwise eq. (5.6.33) would imply there is no m_{max}; likewise,

$$J^- \left|\lambda_J, m_{\text{min}}\right\rangle = 0. \tag{5.6.45}$$

Let us in fact consider the former; this implies

$$\left\langle\lambda_J, m_{\text{max}}\right| (J^+)^\dagger J^+ \left|\lambda_J, m_{\text{max}}\right\rangle = 0 \tag{5.6.46}$$

$$\left\langle\lambda_J, m_{\text{max}}\right| \vec{J}^2 - (J^3)^2 - J^3 \left|\lambda_J, m_{\text{max}}\right\rangle = 0 \tag{5.6.47}$$

$$\lambda_J = m_{\text{max}}(m_{\text{max}} + 1); \tag{5.6.48}$$

where eq. (5.6.39) was employed in the second line. If we instead considered $J^- \left|\lambda_J, m_{\text{min}}\right\rangle = 0$,

$$\left\langle\lambda_J, m_{\text{min}}\right| (J^-)^\dagger J^- \left|\lambda_J, m_{\text{min}}\right\rangle = 0 \tag{5.6.49}$$

$$\left\langle\lambda_J, m_{\text{min}}\right| \vec{J}^2 - (J^3)^2 + J^3 \left|\lambda_J, m_{\text{min}}\right\rangle = 0 \tag{5.6.50}$$

$$\lambda_J = m_{\text{min}}(m_{\text{min}} - 1); \tag{5.6.51}$$

where once again eq. (5.6.39) was employed in the second line. Equating the right hand sides of equations (5.6.48) and (5.6.51),

$$m_{\text{max}} = \frac{-1 \pm \sqrt{1 - 4(1)(-)m_{\text{min}}(m_{\text{min}} - 1)}}{2} \tag{5.6.52}$$

$$= -\frac{1}{2} \pm \left(m_{\text{min}} - \frac{1}{2}\right). \tag{5.6.53}$$

This indicates, either $m_{\text{max}} = m_{\text{min}} - 1$ or $m_{\text{max}} = -m_{\text{min}}$. But the former is a contradiction, since the maximum should never be smaller than the minimum. Moreover, there must be some positive integer n such that $(J^+)^n \left|\lambda_J, m_{\text{min}}\right\rangle \propto \left|\lambda_J, m_{\text{max}}\right\rangle$; namely, there are n integer steps between m_{min} and m_{max}. At this point we gather

$$m_{\text{min}} + n = -m_{\text{max}} + n = m_{\text{max}}; \tag{5.6.54}$$

which in turn implies

$$m_{\text{max}} = \frac{n}{2}. \tag{5.6.55}$$

Since we have no further constraints on the integer n, we now search the cases where m_{max} is integer (i.e., when n is even) and when it is half-integer (i.e., when n is odd). Cleaning up our notation somewhat, $m_{\text{max}} = -m_{\text{min}} \equiv \ell$, and recalling eq. (5.6.48):

Spin and 3D Rotations Starting solely from the commutation relations between the Hermitian angular momentum operators $\{J^i\}$ in eq. (5.6.24), we surmise: the simultaneous eigensystem of $\vec{J}^2$ and J^3 is encoded within

$$\vec{J}^2 \left|\ell, m\right\rangle = \ell(\ell+1) \left|\ell, m\right\rangle \qquad \text{and} \qquad J^3 \left|\ell, m\right\rangle = m \left|\ell, m\right\rangle . \tag{5.6.56}$$

Here, the *spin* ℓ can be a non-negative integer ($\ell = 0, 1, 2, \dots$) or positive half-integer ($\ell = 1/2, 3/2, 5/2, \dots$); whereas the *azimuthal* eigenvalue runs between $-\ell$ to ℓ in integer steps:

$$m \in \{-\ell, -\ell+1, -\ell+2, \dots, \ell-2, \ell-1, \ell\}. \tag{5.6.57}$$

We may also compute, up to an overall phase, the normalization constant in eq. (5.6.32). We have, from eq. (5.6.39),

$$0 \le |c_{m\pm1}|^2 \left\langle \ell, m\pm1 \middle| \ell, m\pm1 \right\rangle = \left\langle \ell, m\right| (J^\pm)^\dagger J^\pm \left|\ell, m\right\rangle \tag{5.6.58}$$

$$= \left\langle \ell, m\right| \vec{J}^2 - (J^3)^2 \mp J^3 \left|\ell, m\right\rangle \tag{5.6.59}$$

$$= \ell(\ell+1) - m(m\pm1) = (\ell\mp m)(\ell \pm m + 1). \tag{5.6.60}$$

Since eigenvectors are only defined up to a phase, we shall *choose* to simply take the positive square root on both sides.

$$\begin{aligned} J^\pm \left|\ell, m\right\rangle &= \sqrt{\ell(\ell+1) - m(m\pm1)} \left|\ell, m\pm1\right\rangle \\ &= \sqrt{(\ell\mp m)(\ell\pm m+1)} \left|\ell, m\pm1\right\rangle . \end{aligned} \tag{5.6.61}$$

Invariant Subspaces in 3D, Degeneracy & Symmetry Because $[\vec{J}^2, J^a] = 0$, we must have, for $D(\widehat{R}) = \exp(-i\vec{\theta}\cdot\vec{J})$,

$$\left[D(\widehat{R}), \vec{J}^2\right] = 0, \tag{5.6.62}$$

$$\vec{J}^2 D(\widehat{R}) \left|\ell, m\right\rangle = D(\widehat{R})\vec{J}^2 \left|\ell, m\right\rangle = \ell(\ell+1)\cdot D(\widehat{R}) \left|\ell, m\right\rangle . \tag{5.6.63}$$

In words, $D(\widehat{R})\left|\ell, m\right\rangle$ is an eigenvector of $\vec{J}^2$ with eigenvalue $\ell(\ell+1)$. Hence, we see that rotations do not 'mix' the eigenvectors $\{\left|\ell, m\right\rangle\}$ of $\vec{J}^2$ with different ℓ. That is,

$$\begin{aligned} D(\widehat{R}) \left|\ell, m\right\rangle &= \sum_{m'=-\ell}^{+\ell} \left|\ell, m'\right\rangle \widehat{D}_{(\ell)}{}^{m'}{}_{m}(\widehat{R}), \\ \widehat{D}_{(\ell)}{}^{m'}{}_{m}(\widehat{R}) &\equiv \left\langle \ell, m'\right| D(\widehat{R}) \left|\ell, m\right\rangle . \end{aligned} \tag{5.6.64}$$

While the completeness relation should involve a sum over all ℓ', namely

$$\mathbb{I} = \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{+\ell'} \left|\ell', m'\right\rangle \left\langle \ell', m'\right| ; \tag{5.6.65}$$

only the $\ell' = \ell$ terms will survive when employed in eq. (5.6.64) – due to the result in eq. (5.6.63).

Now, suppose a Hermitian operator A remains invariant under rotations. That means it should be invariant under all change-of-basis induced by rotations – for e.g., in the position representation,

$$\langle \vec{x} | A | \vec{x}' \rangle = \langle \widehat{R}\vec{x} | A | \widehat{R}\vec{x}' \rangle \tag{5.6.66}$$

$$= \left\langle \vec{x} \left| D(\widehat{R})^\dagger A D(\widehat{R}) \right| \vec{x}' \right\rangle \tag{5.6.67}$$

for all rotations $\widehat{R}$. This motivates the following: A is rotationally invariant iff it obeys

$$D(\widehat{R})^\dagger A D(\widehat{R}) = A. \tag{5.6.68}$$

By Taylor expanding $D(\widehat{R})$, we see an equivalent definition is: A is rotationally invariant iff it commutes with the generators $\{J^i\}$.

$$\left[J^i, A\right] = 0 \tag{5.6.69}$$

We must therefore be able to simultaneously diagonalize $\{A, \vec{J}^2, J^3\}$. Furthermore, let us observe that the eigenstates $|\lambda; \ell, m\rangle$ of A – obeying $A|\lambda; \ell, m\rangle = \lambda |\lambda; \ell, m\rangle$ – must in fact be degenerate with respect to the eigenvalues of J^3. This is an explicit example of the "symmetry implies degeneracy" discussion at the end of §(5.1). To see this, we first compute

$$A\left(D(\widehat{R})|\lambda; \ell, m\rangle\right) = D(\widehat{R}) A |\lambda; \ell, m\rangle \tag{5.6.70}$$

$$= \lambda\left(D(\widehat{R})|\lambda; \ell, m\rangle\right). \tag{5.6.71}$$

Inserting a complete set of eigenstates, and exploiting the fact that eigenstates with distinct eigenvalues are necessarily orthogonal,

$$A \sum_{\lambda', \ell', m'} |\lambda'; \ell', m'\rangle \left\langle \lambda'; \ell', m' \left| D(\widehat{R}) \right| \lambda; \ell, m \right\rangle \tag{5.6.72}$$

$$= A \sum_{m'} |\lambda; \ell, m'\rangle \left\langle \lambda; \ell, m' \left| D(\widehat{R}) \right| \lambda; \ell, m \right\rangle \tag{5.6.73}$$

$$= \lambda \sum_{m'} |\lambda; \ell, m'\rangle \left\langle \lambda; \ell, m' \left| D(\widehat{R}) \right| \lambda; \ell, m \right\rangle. \tag{5.6.74}$$

Since we have made no assumption of the rotation $\widehat{R}$ here, we see that an arbitrary superposition of eigenstates of different $m-$values remain an eigenstate of A. That is, all $m-$values must belong the same degenerate subspace of a given λ: there must be at least be a $2\ell + 1$ degeneracy if A is rotationally invariant. Note, however, this says nothing about states of distinct $\vec{J}^2$ eigenvalues – i.e., eigenvectors with different ℓs can have either the same or different eigenvalues of A, depending on what A itself actually is.

Vector Operators Suppose $D(\widehat{R})$ is a rotation operator. Consider the following operation involving the position operator X^i and its eigenkets $\{|\vec{x}\rangle\}$:

$$D(\widehat{R})^\dagger X^i D(\widehat{R}) |\vec{x}\rangle = D(\widehat{R})^\dagger X^i \left| \widehat{R}\vec{x} \right\rangle \tag{5.6.75}$$

$$= (\widehat{R}\vec{x})^i D(\widehat{R})^\dagger \left|\widehat{R}\vec{x}\right\rangle = (\widehat{R}\vec{x})^i \left|\widehat{R}^{\mathrm{T}}\widehat{R}\vec{x}\right\rangle \tag{5.6.76}$$

$$= \widehat{R}^i{}_j x^j \left|\vec{x}\right\rangle . \tag{5.6.77}$$

(We have employed eq. (5.5.83) in the third equality.) Since this holds for arbitrary position eigenkets, we must have the operator identity

$$D(\widehat{R})^\dagger X^i D(\widehat{R}) = \widehat{R}^i{}_j X^j. \tag{5.6.78}$$

Problem 5.77. Using eq. (5.2.92), first explain why the rotation operator applied to $\left|\vec{k}\right\rangle$, the eigenket of the momentum operator, behaves similarly as its position cousin:

$$D(\widehat{R})\left|\vec{k}\right\rangle = \left|\widehat{R}\vec{k}\right\rangle . \tag{5.6.79}$$

Then show that the analog to eq. (5.6.78) for the momentum operator holds; namely,

$$D(\widehat{R})^\dagger P_i D(\widehat{R}) = \widehat{R}_i{}^j P_j, \tag{5.6.80}$$

where we have defined $\widehat{R}_i{}^j \equiv \widehat{R}^i{}_j$. $\square$

Equations (5.6.78) and (5.6.80) motivate the following definition:

> **Vector Operator: Definition** A vector operator V^i is one whose components transforms like those of an ordinary 3−vector in flat space, upon a change-of-basis induced by a rotation operator $D(\widehat{R})$:
>
> $$D(\widehat{R})^\dagger V^i D(\widehat{R}) = \widehat{R}^i{}_j V^j. \tag{5.6.81}$$
>
> Although we shall focus on the $D = 3$ case here, note that this definition holds in arbitrary dimensions $D \geq 3$.

Recall that, since $D(\widehat{R})$ is unitary, that means $D^\dagger V^i D$ may be thought of V^i computed in a rotated orthonormal basis. In particular, if V^i is a vector operator, the matrix element

$$\left\langle \psi_1 \left| D(\widehat{R})^\dagger V^i D(\widehat{R}) \right| \psi_2 \right\rangle = \left\langle \psi_1' \left| V^i \right| \psi_2' \right\rangle , \tag{5.6.82}$$

where $\left|\psi_{1,2}'\right\rangle \equiv D(\widehat{R})\left|\psi_{1,2}\right\rangle$ are the rotated kets, transforms as – according to eq. (5.6.80) –

$$\left\langle \psi_1' \left| V^i \right| \psi_2' \right\rangle = \widehat{R}^i{}_j \left\langle \psi_1 \left| V^j \right| \psi_2 \right\rangle . \tag{5.6.83}$$

In words: the matrix element of V^i with respect to the rotated kets amounts to that with respect to the 'old' kets, but rotated with the matrix $\widehat{R}$.

Problem 5.78. Vector Operator in 3D: Infinitesimal Version In 3D, show that if V^i is a vector operator obeying eq. (5.6.81), then it also obeys

$$\left[J^a, V^b\right] = i\epsilon^{abc}V^c. \tag{5.6.84}$$

Can you argue, if eq. (5.6.84) holds, then so does eq. (5.6.81) – i.e., they are equivalent definitions of a vector operator? Hint: Recall equations (5.2.49) and (5.6.26).

Remark Notice, from eq. (5.6.84), that the angular momentum generators $\{J^a\}$ are themselves vector operators. $\square$

Problem 5.79. Scalars from Dot Product Show that the 'dot product' of vector operators V^i and W^i, namely $\vec{V} \cdot \vec{W} \equiv V^a W^a$, transforms as a scalar:

$$\left[J^a, \vec{V} \cdot \vec{W}\right] = 0. \tag{5.6.85}$$

Through eq. (5.2.49), this means $D(\widehat{R})^\dagger(\vec{V} \cdot \vec{W})D(\widehat{R}) = \vec{V} \cdot \vec{W}$. □

Problem 5.80. Parity In 3D, the parity operator $\widehat{P}$ acting on 3−vectors $\vec{v}$, namely $\widehat{P}\vec{v} = -\vec{v}$ for arbitrary $\vec{v}$, may be readily identified as

$$\widehat{P} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{5.6.86}$$

What is the determinant of this 3×3 parity-implementing $\widehat{P}$? Use this to argue, there is no rotation matrix $\widehat{R}$ in 3D that can implement $\widehat{P}$. If we do not flip all 3 directions, but only 1, namely

$$_{(1)}\widehat{P} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad _{(2)}\widehat{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad _{(3)}\widehat{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{5.6.87}$$

explain why these $\{{}_{(i)}\widehat{P}\}$ cannot be implemented by a continuous rotation operator. Hint: Remember eq. (5.5.9). On the other hand, explain why flipping 2 out of the 3 axes, namely

$$_{(1,2)}\widehat{P} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad _{(1,3)}\widehat{P} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad _{(2,3)}\widehat{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{5.6.88}$$

can in fact be implemented by rotations – find the appropriate rotation matrices and their associated angles.

Parity in general $D \geq 2$ dimensions More generally, can you show that reversing the directions of even number of Cartesian coordinate axes may always be implemented by an appropriate rotation? (Write down the rotation matrix and the associated rotation angles.) Whereas, show that reversing the direction of an odd number of Cartesian coordinate axes cannot be implemented by a rotation. □

5.6.2 Integer Spin and Spherical Harmonics

[33]In this section, we shall witness how the angular spherical harmonics introduced in equations (5.5.128) and (5.5.130) are in fact the position representation of the integer spin case ($\ell = 0, 1, 2, 3, \dots$) in eq. (5.6.56) for 3D rotations. Specifically, if we apply the position eigenket $\langle r, \theta, \phi|$ – written in spherical coordinates

$$(x^1, x^2, x^3) = r\left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)\right) = r\widehat{r}(\theta, \phi) \tag{5.6.89}$$

[33]This and the next 2 sections are under heavy construction.

– on both sides of eq. (5.6.56):

$$\langle r,\theta,\phi| \vec{J}^2 |\ell,m\rangle = \ell(\ell+1) \langle r,\theta,\phi| \ell,m\rangle . \tag{5.6.90}$$

Recalling the result in eq. (5.5.104),

$$-\vec{\nabla}_{\mathbb{S}^2} \langle r,\theta,\phi| \ell,m\rangle = \ell(\ell+1) \langle r,\theta,\phi| \ell,m\rangle . \tag{5.6.91}$$

Notice, when $D=3$, eq. (5.5.130) reads

$$-\vec{\nabla}_{\mathbb{S}^2} Y_\ell^m(\theta,\phi) = \ell(\ell+1) Y_\ell^m(\theta,\phi). \tag{5.6.92}$$

Below, we shall identify

$$Y_\ell^m(\theta,\phi) = \langle \theta,\phi| \ell,m\rangle . \tag{5.6.93}$$

Firstly, if we convert Cartesian to Spherical coordinates via eq. (5.6.89), the metric in 3D flat space becomes

$$\mathrm{d}\ell^2 = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j = \delta_{ij}\frac{\partial x^i}{\partial(r,\theta,\phi)^a}\frac{\partial x^j}{\partial(r,\theta,\phi)^b}(\mathrm{d}r,\mathrm{d}\theta,\mathrm{d}\phi)^a(\mathrm{d}r,\mathrm{d}\theta,\mathrm{d}\phi)^b \tag{5.6.94}$$

$$= \mathrm{d}r^2 + r^2 H_{\mathrm{IJ}}\mathrm{d}\theta^{\mathrm{I}}\mathrm{d}\theta^{\mathrm{J}}, \tag{5.6.95}$$

$$H_{\mathrm{IJ}}\mathrm{d}\theta^{\mathrm{I}}\mathrm{d}\theta^{\mathrm{J}} = \mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2. \tag{5.6.96}$$

[34]The square root of the determinant is

$$\sqrt{g} = r^2\sqrt{H} = r^2\sin\theta; \tag{5.6.97}$$

the non-zero components of the inverse metric are

$$g^{rr} = 1, \qquad g^{\theta\theta} = r^{-2}H^{\theta\theta} = r^{-2}, \qquad g^{\phi\phi} = r^{-2}H^{\phi\phi} = (r\sin(\theta))^{-2}. \tag{5.6.98}$$

Therefore, the Laplacian is

$$\vec{\nabla}^2\psi = \frac{1}{r^2 s_\theta}\left(\partial_r\left(r^2 s_\theta\partial_r\psi\right) + \partial_\theta\left(r^2 s_\theta r^{-2}\partial_\theta\psi\right) + \partial_\phi\left(r^2 s_\theta (r s_\theta)^{-2}\partial_\phi\psi\right)\right) \tag{5.6.99}$$

$$= \frac{1}{r^2}\partial_r\left(r^2\partial_r\psi\right) + \frac{1}{r^2}\vec{\nabla}^2_{\mathbb{S}^2}\psi, \tag{5.6.100}$$

where $\vec{\nabla}^2_{\mathbb{S}^2}$ is the Laplacian on the 2−sphere of unit radius,

$$-\left\langle \vec{x}\left|\vec{J}^2\right|\psi\right\rangle = \vec{\nabla}^2_{\mathbb{S}^2}\psi = \frac{1}{\sin(\theta)}\left(\partial_\theta\left(\sin(\theta)\partial_\theta\psi\right) + \frac{1}{\sin(\theta)}\partial_\phi^2\psi\right). \tag{5.6.101}$$

We may directly infer from equations (5.5.93) and (5.6.1) that in 3D, the position representation of the generators of rotations (aka "angular momentum operators") are

$$\left\langle \vec{x}\left|\vec{J}\right| f\right\rangle = -i\vec{x}\times\vec{\nabla}\left\langle \vec{x}\right| f\rangle , \tag{5.6.102}$$

$$\left(\vec{x}\times\vec{\nabla}\right)^k = \epsilon^{ijk}x^i\partial_j; \tag{5.6.103}$$

whereas the operator form is

$$\vec{J} = \vec{X}\times\vec{P} \qquad \Leftrightarrow \qquad J^k = \epsilon^{ijk}X^iP_j = \frac{1}{2}\epsilon^{ijk}J^{ij}. \tag{5.6.104}$$

[34]The above calculation is described in more detail in §(9.1).

Problem 5.81. Cross Product & Levi-Civita By working out the components explicitly, show that the cross product can indeed be written in terms of the Levi-Civita symbol:

$$(\vec{A} \times \vec{B})^i = \epsilon^{ijk} A^j B^k. \tag{5.6.105}$$

For instance, $(\vec{A} \times \vec{B})^1 = \epsilon^{1jk} A^j B^k = \epsilon^{123} A^2 B^3 + \epsilon^{132} A^3 B^2 = A^2 B^3 - A^3 B^2$. $\square$

Problem 5.82. Orbital Angular Momentum Operators In the spherical coordinate system defined in eq. (5.6.89), show that the angular momentum operators, i.e., the generators of rotation in 3D, are

$$\left\langle r, \theta, \phi \left| J^1 \right| \psi \right\rangle = i \left(\sin(\phi) \partial_\theta + \cos(\phi) \cot(\theta) \partial_\phi \right) \left\langle r, \theta, \phi \right| \psi \rangle , \tag{5.6.106}$$

$$\left\langle r, \theta, \phi \left| J^2 \right| \psi \right\rangle = i \left(-\cos(\phi) \partial_\theta + \sin(\phi) \cot(\theta) \partial_\phi \right) \left\langle r, \theta, \phi \right| \psi \rangle , \tag{5.6.107}$$

$$\left\langle r, \theta, \phi \left| J^3 \right| \psi \right\rangle = -i \partial_\phi \left\langle r, \theta, \phi \right| \psi \rangle . \tag{5.6.108}$$

In turn, deduce that the position representations of the ladder operators in eq. (5.6.28) are

$$\left\langle r, \theta, \phi \left| J^\pm \right| \psi \right\rangle = e^{\pm i\phi} \left(\pm \partial_\theta + i \cot(\theta) \partial_\phi \right) \left\langle r, \theta, \phi \right| \psi \rangle . \tag{5.6.109}$$

Hint: Recall equations (5.5.121) and (5.6.1). $\square$

Spherical Harmonics in 3D Let us now turn to solving the spherical harmonics in 3D, and the associated eigenfunctions of $\vec{J}^2$ – recall equations (5.5.128) and (5.5.130). Remember, since $[J^3, \vec{J}^2] = 0$, we must be able to simultaneously diagonalize J^3 and $\vec{J}^2$. In fact, since $\langle r, \theta, \phi | J^3 | \psi \rangle = -i\partial_\phi \langle r, \theta, \phi | \psi \rangle$, we must have

$$\left\langle \theta, \phi \left| J^3 \right| \ell, m \right\rangle = m \left\langle \theta, \phi \right| \ell, m \rangle , \tag{5.6.110}$$

$$-i\partial_\phi \left\langle \theta, \phi \right| \ell, m \rangle = m \left\langle \theta, \phi \right| \ell, m \rangle . \tag{5.6.111}$$

The solution to the second line is the solution to $-i\partial_\phi f(\phi) = m f(\phi) \Rightarrow f(\phi) = f_0 \exp(im\phi)$, except in our case f_0 can still depend on θ and other parameters in the problem. This implies the angular spherical harmonics takes the form

$$Y_\ell^m(\theta, \phi) = \langle \theta, \phi | \ell, m \rangle = \langle \theta | \ell, m \rangle \exp(im\phi). \tag{5.6.112}$$

Next, we recall the discussions around equations (5.6.44) and (5.6.45), that the raising operator applied to the state with maximum azimuthal eigenvalue $m_{\text{max}} \equiv \ell$ must be a null vector (otherwise there would not be a maximum value in the first place). Similarly the lowering operator applied to the state with minimum azimuthal eigenvalue $m_{\text{min}} = -m_{\text{max}} = -\ell$ must also be a null vector. Using the results in eq. (5.6.109) and (5.6.112), we may write the position representation of eq. (5.6.44) as

$$e^{i\phi} \left(\partial_\theta + i \cot(\theta) \partial_\phi \right) Y_\ell^\ell(\theta, \phi) = e^{i(\ell+1)\phi} \left(\partial_\theta - \ell \cdot \cot(\theta) \right) \langle \theta | \ell, \ell \rangle = 0. \tag{5.6.113}$$

Using the results in eq. (5.6.109) and (5.6.112), we may write the position representation of eq. (5.6.45) as

$$e^{-i\phi} \left(-\partial_\theta + i \cot(\theta) \partial_\phi \right) Y_\ell^{-\ell}(\theta, \phi) = e^{-i(\ell+1)\phi} \left(-\partial_\theta + \ell \cdot \cot(\theta) \right) \langle \theta | \ell, -\ell \rangle = 0. \tag{5.6.114}$$

Problem 5.83. Solve equations (5.6.113) and (5.6.114) and proceed to normalize

$$\langle \ell, \pm\ell | \, \ell, \pm\ell \rangle = \int_{\mathbb{S}^2} \mathrm{d}^2\Omega |Y_\ell^{\pm\ell}|^2 \tag{5.6.115}$$

$$= \int_{-1}^{+1} \mathrm{d}(\cos\theta) \int_0^{2\pi} \mathrm{d}\phi |Y_\ell^{\pm\ell}(\theta,\phi)|^2 = 1; \tag{5.6.116}$$

to arrive at – up to an overall multiplicative phase $e^{i\delta_\pm}$ –

$$Y_\ell^{\pm\ell}(\theta,\phi) = \frac{e^{i\delta_\pm}}{2^\ell \ell!}\sqrt{\frac{2\ell+1}{4\pi}(2\ell)!}\sin^\ell(\theta) e^{\pm i\ell\phi}. \tag{5.6.117}$$

Hint: The integrand may be binomial expanded in powers of $e^{\pm i\theta}$. □

For $m \geq 0$, it is consistent to define

$$Y_\ell^m(\theta,\phi) = \frac{(-)^\ell}{2^\ell \ell!}\sqrt{\frac{2\ell+1}{4\pi}\cdot\frac{(\ell+m)!}{(\ell-m)!}}\frac{e^{im\phi}}{\sin^m(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\cos(\theta)}\right)^{\ell-m}\left(1-\cos^2(\theta)\right)^\ell; \tag{5.6.118}$$

whereas for define negative m, we may obtain it via the definition

$$Y_\ell^{-m} = (-)^m \overline{Y_\ell^m(\theta,\phi)}. \tag{5.6.119}$$

The validity of eq. (5.6.118) may be demonstrated via induction on m. For, one may assume the mth case is true. Then, the Y_ℓ^{m-1} must be gotten by applying the lowering operator once. Keeping in mind equations (5.6.61) and (5.6.109),

$$\left\langle \theta,\phi \left| J^- \right| \ell, m \right\rangle = \sqrt{(\ell+m)(\ell-m+1)}\,\langle \theta,\phi | \, \ell, m-1 \rangle \tag{5.6.120}$$

$$Y_\ell^{m-1}(\theta,\phi) = \frac{e^{i\delta'} e^{-i\phi}}{\sqrt{(\ell+m)(\ell-m+1)}}\left(-\partial_\theta + i\cot(\theta)\partial_\phi\right) Y_\ell^m(\theta,\phi), \tag{5.6.121}$$

where $e^{i\delta'}$ is an arbitrary phase. The proof is established via a direct calculation. Additionally, since everything except the $e^{im\phi}$ in eq. (5.6.118) is real, eq. (5.6.119) says the negative m value spherical harmonics may simply obtained by flipping the sign of m; i.e., $e^{im\phi} \to e^{-im\phi}$. This certainly yields the correct eigenvalue of $J^3 \leftrightarrow -i\partial_\phi$ in eq. (5.6.108). But what about the $\theta-$dependent portion? To this end, we recall the position space eigenvector/value equation (5.6.92). If we insert into eq. (5.6.92) the definition

$$Y_\ell^m(\theta,\phi) \equiv P_\ell^m(\cos\theta) e^{im\phi}, \tag{5.6.122}$$

we would obtain the ordinary differential equation for the associated Legendre function:

$$\partial_c\left((1-c^2)\partial_c P_\ell^m(c)\right) + \left(\ell(\ell+1) - \frac{m^2}{1-c^2}\right) P_\ell^m(c) = 0, \qquad c \equiv \cos\theta. \tag{5.6.123}$$

The presence of m^2 in eq. (5.6.123) tells us, $P_\ell^m(\cos\theta)$ and $P_\ell^{-m}(\cos\theta)$ must both refer to the same solution, since this ODE is insensitive to the sign of m.

Problem 5.84. Verify eq. (5.6.123) from the $D = 3$ version of eq. (5.5.130). $\square$

Spherical Harmonics as Homogeneous Polynomials[35] We already know from the discussion in §(5.5) that $r^\ell Y(\theta, \phi)$ is a homogeneous polynomial of degree $\ell \geq 0$, involving the Cartesian components (x^1, x^2, x^3), satisfying the (homogeneous) Laplace equation (5.5.126) – or, equivalently eq. (5.5.128):

$$\vec{\nabla}^2 \left(r^\ell Y_\ell^m(\theta, \phi) \right) = 0 \tag{5.6.124}$$

The $\{r^\ell Y_\ell^m = r^\ell P_\ell^m(\cos\theta) e^{im\phi}\}$ may be constructed by first defining

$$x^\pm = x^1 \pm i x^2 = r \sin(\theta) e^{\pm i\phi}. \tag{5.6.125}$$

Let us consider homogeneous polynomials of degree ℓ by superposing the products of positive powers of $x^\pm$ and x^3, namely

$$r^\ell Y_\ell^m = \psi_{a_+ a_- b} (x^+)^{a_+} (x^-)^{a_-} (x^3)^b \tag{5.6.126}$$

$$= \psi_{a_+ a_- b} \cdot r^{a_+ + a_- + b} (\sin(\theta))^{a_+ + a_-} \cos^b(\theta) \exp(i(a_+ - a_-)\phi). \tag{5.6.127}$$

In order to obtain a degree ℓ polynomial, the sum of the powers must yield ℓ.

$$a_+ + a_- + b = \ell \tag{5.6.128}$$

To achieve this, for a fixed a_+, we may choose $a_- \in \{0, 1, 2, \dots, \ell - a_+\}$; followed by putting $b = \ell - a_+ - a_-$. Therefore, the total number of independent terms in eq. (5.6.126) is

$$N_\ell = \sum_{a_+=0}^{\ell} \sum_{a_-=0}^{\ell - a_+} 1 = \sum_{a_+=0}^{\ell} (\ell - a_+ + 1) = (\ell+1)^2 - \frac{0+\ell}{2}(\ell+1) = \frac{(\ell+1)(\ell+2)}{2}. \tag{5.6.129}$$

Problem 5.85. Show that the Laplacian acting on an arbitrary function $\psi(x^+, x^-, x^3)$ is

$$\delta^{ij} \partial_i \partial_j \psi = (4 \partial_+ \partial_- + \partial_3^2) \psi, \tag{5.6.130}$$

where $\partial_\pm$ is the derivative with respect to $x^\pm \equiv x^1 \pm i x^2$. $\square$

Inserting eq. (5.6.126) into eq. (5.5.126), one would find

$$\begin{aligned} &\vec{\nabla}^2 \left(\psi_{a_+ a_- b} (x^+)^{a_+} (x^-)^{a_-} (x^3)^b \right) \\ &= \psi_{a_+ a_- b} (x^+)^{a_+} (x^-)^{a_-} (x^3)^{b-2} + 4 \psi_{a_+ a_- b} (x^+)^{a_+ - 1} (x^-)^{a_- - 1} (x^3)^b = 0. \end{aligned} \tag{5.6.131}$$

This explicitly demonstrates that the Laplacian acting on a homogeneous polynomial of degree ℓ is a homogeneous polynomial of degree $\ell - 2$. Since the latter has $N_{\ell-2}$ independent terms (by eq. (5.6.129)), that means eq. (5.6.131) provides us $N_{\ell-2}$ constraints to be obeyed by the N_ℓ independent terms of eq. (5.6.126). Therefore, there must actually be

$$N_\ell - N_{\ell-2} = \frac{(\ell+1)(\ell+2)}{2} - \frac{(\ell-1)\ell}{2} = 2\ell + 1 \tag{5.6.132}$$

[35] Part of the discussion here is modeled after the one in Weinberg [12].

independent terms in the most general homogeneous polynomial of degree ℓ that solves eq. (5.5.126) in 3D.

But, as we have already discovered, $2\ell + 1$ is exactly the number of linearly independent spherical harmonics $\{Y_\ell^m | m = -\ell, \dots, +\ell\}$ for a fixed ℓ. This indicates the solutions of eq. (5.5.126) in 3D must, up to an overall multiplicative constant, be the Y_ℓ^m themselves. In fact, let us define

$$m \equiv a_+ - a_-. \tag{5.6.133}$$

By superposing the $(a_+ - a_-)/2$ and $(a_+ + a_-)/2$ axes on the (a_+, a_-) plane – drawing a figure here would help – we may readily observe that

$$\max(a_+ - a_-) = \ell \qquad \text{and} \qquad \min(a_- - a_-) = -\ell. \tag{5.6.134}$$

In the other words,

$$-\ell \leq m \leq +\ell. \tag{5.6.135}$$

By taking into account equations (5.6.128) and (5.6.133), eq. (5.6.126) now reads

$$Y'^m_\ell(\theta, \phi) = \frac{1}{r^\ell} \sum_{\substack{a_+ + a_- + b = \ell \\ a_+ - a_- = m}} \psi'_b (x^+)^{a_+} (x^-)^{a_-} (x^3)^b \qquad \left(a_\pm = \frac{\ell - b \pm m}{2} \right) \tag{5.6.136}$$

$$= \sum_b \psi'_b \cdot \sin^{\ell - b}(\theta) \cos^b(\theta) e^{im\phi}; \tag{5.6.137}$$

for appropriate coefficients $\{\psi'_b\}$. The $\exp(im\phi)$ indicates it obeys the equivalent of eq. (5.6.111), namely

$$-i\partial_\phi Y'^m_\ell = m Y'^m_\ell. \tag{5.6.138}$$

Furthermore, from our analysis, these $\{Y'^m_\ell(\theta, \phi)\}$ must be proportional to the corresponding $\{Y_\ell^m(\theta, \phi)\}$; since they correspond to the same number of independent solutions to

$$-\vec{\nabla}^2_{\mathbb{S}^2} Y'^m_\ell = \ell(\ell + 1) Y'^m_\ell. \tag{5.6.139}$$

To sum: in 3D, the $r^\ell Y_\ell^m(\theta, \phi)$, when expressed in Cartesian coordinates $\vec{x}$, are homogeneous polynomials of degree ℓ satisfying equations (5.6.138) and (5.6.139).

Example: $\ell = 0$ For $\ell = 0$, this corresponds to having zero powers of the coordinates – i.e., a constant: i.e., $Y_0^0 =$constant.

Example: $\ell = 1$ A polynomial linear in either x^+, x^-, or x^3 is automatically a solution of the Laplace equation $\vec{\nabla}^2 \psi = 0$ since the Laplacian has two derivatives. Hence, we must have

$$r Y_1^{\pm 1} \propto x^\pm = r \sin(\theta) e^{\pm i\phi}, \qquad r Y_1^0 \propto x^3 = r\cos(\theta). \tag{5.6.140}$$

Example: $\ell = 2$ For $\ell = 2$, we have the possibilities

$$(a_+, a_-, b) = (2, 0, 0) \Rightarrow m = 2 \tag{5.6.141}$$

$$(a_+, a_-, b) = (0, 2, 0) \Rightarrow m = -2 \tag{5.6.142}$$
$$(a_+, a_-, b) = (0, 0, 2) \Rightarrow m = 0 \tag{5.6.143}$$
$$(a_+, a_-, b) = (1, 1, 0) \Rightarrow m = 0 \tag{5.6.144}$$
$$(a_+, a_-, b) = (1, 0, 1) \Rightarrow m = 1 \tag{5.6.145}$$
$$(a_+, a_-, b) = (0, 1, 1) \Rightarrow m = -1. \tag{5.6.146}$$

Here, the "$\Rightarrow$" means the term $(x^+)^{a_+}(x^-)^{a_-}(x^3)^b$ under consideration (given by the (a_+, a_-, b) on its left hand side) contributes to the corresponding azimuthal eigenvalue (on its right hand side).

Problem 5.86. Normalize the spherical harmonics to unity on the sphere, i.e.,

$$\langle \ell, m | \, \ell, m \rangle = \int_{-1}^{+1} \mathrm{d}(\cos\theta) \int_0^{2\pi} \mathrm{d}\phi |Y_\ell^m(\theta, \phi)|^2 = 1. \tag{5.6.147}$$

Proceed to compute Y_ℓ^m (up to a multiplicative phase) for $\ell = 0, 1, 2$ by demanding they satisfy the homogeneous equations (5.5.126) and (5.5.128). Hint: The answers can be found in equations (12.2.67), (12.2.68) and (12.2.69) below. □

Problem 5.87. Parity If $\vec{x} \equiv r \cdot \widehat{r}(\theta, \phi)$, where $\widehat{r}$ is the unit radial vector, explain the following parity transformation of the spherical harmonics:

$$Y_\ell^m(-\widehat{r}) = (-)^\ell Y_\ell^m(\widehat{r}). \tag{5.6.148}$$

If P is the parity operator, defined by $P \, |\vec{x}\rangle = |-\vec{x}\rangle$, explain why this implies

$$P \, |\ell, m\rangle = (-)^\ell \, |\ell, m\rangle \, . \tag{5.6.149}$$

Hint: Use the fact that $r^\ell Y_\ell^m(\widehat{r})$ is a homogeneous polynomial of degree ℓ. □

5.6.3 Half Integer Spin and SU_2

In this section, we shall witness how the Special Unitary group of 2×2 matrices, or SU_2 for short, implements rotations on spin$-1/2$ systems, the smallest of the half integer spin solutions we obtained in §(5.6.1). Let us construct its group elements explicitly, using the Pauli matrices in eq. (3.2.18). If $\widehat{U}$ denotes an arbitrary element, it obeys

$$\widehat{U}^\dagger \widehat{U} = \mathbb{I} \qquad \text{and} \qquad \det \widehat{U} = 1. \tag{5.6.150}$$

(The 'Special' in the SU_2 refers to the $\det \widehat{U} = 1$ condition.) The $\widehat{U}^\dagger \widehat{U} = \mathbb{I}$ is a matrix equation and therefore provides 4 constraints; whereas $\det \widehat{U} = 1$ provides another – altogether 5 algebraic equations for the 4 complex matrix entires of $\widehat{U}$. This leaves 3 real parameters. On the other hand, if we assume SU_2 matrices are continuously connected to the identity, we may write $\widehat{U} = \exp(-i\widehat{X})$. In particular, since the Pauli matrices $\{\widehat{\sigma}^\mu\}$ in eq. (3.2.18) are a complete set, we may express

$$\widehat{U} = \exp\left(-i\xi_\mu \widehat{\sigma}^\mu\right), \qquad \xi_\mu \in \mathbb{R}. \tag{5.6.151}$$

Next, let us use the matrix identity

$$\det e^{\widehat{X}} = e^{\text{Tr}\left[\widehat{X}\right]} \tag{5.6.152}$$

[36]as well as the traceless propery of $\{\widehat{\sigma}^i\}$ to infer

$$\det \widehat{U} = e^{\text{Tr}\left[\xi_0 \widehat{\sigma}^0\right]} = e^{-i2\xi_0}. \tag{5.6.153}$$

Hence, $\xi_0 = \pi n$ (for integer n) and we have $\widehat{U} = (-)^n \exp(-i\vec{\xi} \cdot \vec{\sigma})$. We will later see that the odd n case, i.e., $-\mathbb{I}$, can gotten by an appropriate choice of $\vec{\xi} \cdot \vec{\sigma}$. Hence, the most general SU_2 group element must take the form

$$\widehat{U} = \exp\left(-i\vec{\xi} \cdot \vec{\sigma}\right) \qquad \in \qquad \text{SU}_2. \tag{5.6.154}$$

Spin-1/2 From the algebra in eq. (4.3.114), we see that

$$\left[\sigma^i, \sigma^j\right] = i(\epsilon^{ijk} - \epsilon^{jik})\sigma^k = 2i\epsilon^{ijk}\sigma^k. \tag{5.6.155}$$

Dividing throughout by 4,

$$\left[\frac{\sigma^i}{2}, \frac{\sigma^j}{2}\right] = i\epsilon^{ijk}\frac{\sigma^k}{2}, \tag{5.6.156}$$

allows us to recover the SO_3 algebra in eq. (5.6.24), provided we identify

$$J^i \equiv \frac{\sigma^i}{2}, \qquad\qquad i \in \{1, 2, 3\}. \tag{5.6.157}$$

At this point, we may re-express eq. (5.6.154) as

$$\widehat{U} = \exp\left(-i\vec{\theta} \cdot \vec{\sigma}/2\right) \tag{5.6.158}$$

$$= \cos\left(\frac{1}{2}|\vec{\theta}|\right) - i\frac{\vec{\theta} \cdot \vec{\sigma}}{|\vec{\theta}|}\sin\left(\frac{1}{2}|\vec{\theta}|\right), \qquad |\vec{\theta}| = \sqrt{\theta_i \theta_i} \equiv \sqrt{\vec{\theta} \cdot \vec{\theta}}, \tag{5.6.159}$$

where $\vec{\theta}$ are the same rotation angles as in the SO_3 element $\exp(-i\vec{\theta} \cdot \vec{J})$ and in the second equality we have recalled eq. (3.2.26). Moreover, as we have already derived in §(4.3.2), the $\{\sigma^i\}$ obeying eq. (4.3.114) have eigenvalues ± 1; therefore, upon diagonalization,

$$J^3 \left|\pm\right\rangle = \pm\frac{1}{2}\left|\pm\right\rangle. \tag{5.6.160}$$

These SU_2 group elements $\{\widehat{U}\}$ in eq. (5.6.158) are acting on spin$-1/2$ systems.

[36]See, for e.g., Theorem 3.10 of arXiv: math-ph/0005032.

Problem 5.88. SU_2 and the Unit Sphere $\mathbb{S}^3$ We have learned that the identity matrix $\widehat{\sigma}^0$ and the Pauli matrices $\{\widehat{\sigma}^i\}$ form a complete basis of the vector space of 2×2 complex matrices. Therefore, it must be possible to write any SU_2 matrix as a superposition $p_\mu \widehat{\sigma}^\mu$. If we first redefine

$$p_0 \equiv q_0 \qquad \text{and} \qquad p_i \equiv i q_i, \tag{5.6.161}$$

for $i = 1, 2, 3$ and for *real* $\{q_\mu | \mu = 0, 1, 2, 3\}$; then show that the superposition

$$\widehat{U} \equiv p_\mu \widehat{\sigma}^\mu \tag{5.6.162}$$

yields the relation

$$\widehat{U}^\dagger \widehat{U} = \widehat{U}\widehat{U}^\dagger = \vec{q}^2 \cdot \mathbb{I}_{2\times 2}, \tag{5.6.163}$$

where

$$\vec{q}^2 \equiv \delta^{\mu\nu} q_\mu q_\nu. \tag{5.6.164}$$

Also explain why this parametrization is independent of the choice of basis for the $\{\widehat{\sigma}^\mu\}$; i.e., if $\widehat{U} \equiv p_\mu \widehat{\sigma}^\mu$ is an SU_2 matrix then so is $\widehat{U}' \equiv p_\mu \widehat{\sigma}'^\mu$, as long as

$$\widehat{\sigma}'^\mu = \widehat{S}^\dagger \widehat{\sigma}^\mu \widehat{S} \tag{5.6.165}$$

for some unitary $\widehat{S}$ change-of-basis transformation.

To sum, the expansion $q_0 + i q_i \widehat{\sigma}^i$ yields an SU_2 matrix if the vector $\vec{q} \equiv (q_0, q_1, q_2, q_3)$ lies on the unit sphere $\mathbb{S}^3$ in 4 spatial dimensions. If we now compare equations (5.6.161) and (5.6.162) with the general SU_2 matrix in eq. (3.5.20), we see that the two forms coincides.

> An SU_2 matrix may be identified with a unique point on $\mathbb{S}^3$. Since the surface of the unit sphere is simply connected, so is the group SU_2 – in contrast to SO_3, which is *not* simply connected.

$\square$

Pseudo-Real representation We may also obtain the SO_3 Lie Algebra in eq. (5.6.24) using $-(\sigma^i)^*/2$ by simply taking the complex conjugate of eq. (5.6.156).

$$\left[\frac{-(\sigma^i)^*}{2}, \frac{-(\sigma^j)^*}{2}\right] = i\epsilon^{ijk}\frac{-(\sigma^k)^*}{2} \tag{5.6.166}$$

Since the Lie Algebra determines the group multiplication rules (for elements continuously connected to the identity) we have shown that

$$\widehat{U}^* = \exp\left(-i\vec{\theta} \cdot (-\vec{\sigma})^*/2\right) \tag{5.6.167}$$

are also SU_2 group elements. However, according to eq. (3.2.23), $\epsilon(\sigma^i)\epsilon^\dagger = (-\sigma^i)^*$, where ϵ is the 2D Levi-Civita symbol with non-zero entries

$$\epsilon_{12} = 1 = -\epsilon_{21}. \tag{5.6.168}$$

Moreover, a direct calculation would demonstrate it is unitary and anti-symmetric,

$$\epsilon^{-1} = \epsilon^\dagger = -\epsilon. \tag{5.6.169}$$

Therefore, $\widehat{U}^*$ is actually related to $\widehat{U}$ via a similarity transformation – i.e., a change-of-basis –

$$\epsilon \cdot \exp\left(-i\vec{\theta} \cdot (-\vec{\sigma})^*/2\right) \cdot \epsilon^\dagger = \exp\left(-(i/2)\theta^j \cdot \left(\epsilon(-\sigma^j)^*\epsilon^\dagger\right)\right) \tag{5.6.170}$$

$$= \exp\left(-i\vec{\theta} \cdot \vec{\sigma}/2\right), \tag{5.6.171}$$

$$\epsilon \cdot \widehat{U}^*(\vec{\theta}) \cdot \epsilon^\dagger = \widehat{U}(\vec{\theta}); \tag{5.6.172}$$

and we therefore consider $\widehat{U}$ and $\widehat{U}^*$ to be equivalent. In the literature, the $\widehat{U}$ in eq. (5.6.158) is said to be a *pseudo-real* representation: for, it is equivalent to its complex conjugate via a change-of-basis matrix ϵ that is anti-symmetric.[37]

SU_2 As Double Cover of SO_3 Even though the Lie Algebra of SU_2 and SO_3 are the same, we now show that group elements in eq. (5.6.158) versus the 3×3 rotation matrices

$$\widehat{R} = \exp\left(-i\vec{\theta} \cdot \vec{J}\right) \tag{5.6.173}$$

cannot be mapped into each other in a $1-$to-1 manner. Instead, there is a $2-$to-1 map from SU_2 to SO_3; and this is why the former is often said to be a 'double cover' of the latter. For this purpose, without loss of generality, we may choose $\vec{\theta} = \theta\widehat{e}_3$ to point along the $3-$axis. Choosing a diagonal basis, $\sigma^3 = \text{diag}[1, -1]$, we have

$$\widehat{U}(\theta) = e^{-i\theta\widehat{\sigma}^3/2} = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}; \tag{5.6.174}$$

while ordinary $3D$ rotation along the $3-$axis yields

$$\widehat{R}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.6.175}$$

Borrowing the result from Problem (5.33), and replacing ϕ there with the $\theta/2$ here, we see that $\widehat{U}(\theta)$ can in fact be readily mapped to

$$\widehat{R}(\theta/2) = \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) & 0 \\ \sin(\theta/2) & \cos(\theta/2) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.6.176}$$

Specifically, it is not difficult to find a $\widehat{S}$ that diagonalizes the 2×2 block

$$\begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix} = \widehat{S}e^{-i\theta\widehat{\sigma}^3/2}\widehat{S}^\dagger. \tag{5.6.177}$$

[37] It can be shown, if group elements $\{\widehat{U}\}$ were equivalent to their complex conjugates $\{\widehat{U}^*\}$, the associated change-of-basis matrix $\widehat{A}$ can only be symmetric or anti-symmetric. If it were symmetric, the representation will be dubbed 'real'.

This means, while it takes two revolutions on the complex circle $e^{i\phi}$ to return to the same SU_2 group element, namely

$$\widehat{U}(\theta + 4\pi) = \widehat{U}(\theta), \tag{5.6.178}$$

this journey would have returned the corresponding rotation matrix $\widehat{R}(\theta)$ to itself *twice*:

$$\widehat{R}(\theta) = \widehat{R}(\theta + 2\pi) = \widehat{R}(\theta + 4\pi). \tag{5.6.179}$$

Moreover, if θ only went through 1 revolution, the SU_2 element $\widehat{U}(2\pi)$ in eq. (5.6.174) would actually yield the negative identity operator:

$$\widehat{U}(2\pi) = \exp\left(-i\pi\widehat{\sigma}^3\right) = -\mathbb{I}. \tag{5.6.180}$$

This is how a spin$-1/2$ *spinor* transforms – for e.g., the electron's wavefunction ψ transforms into $(-1)\cdot\psi$ upon a 2π rotation, and only returns to ψ after 2 full rotations.

Rotating the rotation axis According to equations (5.6.12) and (5.6.13), if $\widehat{R}(\vec{\vartheta})$ is a 3D rotation matrix, then

$$\widehat{U}(\vec{\vartheta}) \exp\left(-\frac{i}{2}\vec{\theta}\cdot\vec{\sigma}\right) \widehat{U}(\vec{\vartheta})^\dagger = \exp\left(-\frac{i}{2}(\widehat{R}\cdot\vec{\theta})\cdot\vec{\sigma}\right). \tag{5.6.181}$$

This also implies

$$\widehat{U}(\vec{\vartheta})\vec{\sigma}\widehat{U}(\vec{\vartheta})^\dagger = \widehat{R}^{\mathrm{T}}\cdot\vec{\sigma} \tag{5.6.182}$$

$$\widehat{U}(\vec{\vartheta})^\dagger\vec{\sigma}\widehat{U}(\vec{\vartheta}) = \widehat{R}\cdot\vec{\sigma}. \tag{5.6.183}$$

If $\widehat{R}(\vec{\vartheta})\widehat{p} = \widehat{p}'$, where both $\widehat{p}$ and $\widehat{p}'$ are unit vectors, then if $\xi^\pm$ are the eigenvectors of $\widehat{p}\cdot\vec{\sigma}$ – namely

$$(\widehat{p}\cdot\vec{\sigma})\,\xi^\pm = \pm\xi^\pm \tag{5.6.184}$$

– the eigenvector of $\widehat{p}'\cdot\vec{\sigma}$ is simply

$$\xi'^\pm \equiv \widehat{U}(\vartheta)\xi^\pm \tag{5.6.185}$$

because

$$(\widehat{p}'\cdot\vec{\sigma})\,\xi'^\pm = (\widehat{R}\cdot\widehat{p})^i{}_j\sigma^j\widehat{U}(\vec{\vartheta})\xi^\pm \tag{5.6.186}$$

$$= (\widehat{R}\cdot\widehat{p})^i{}_j\widehat{U}(\vec{\vartheta})\widehat{U}(\vec{\vartheta})^\dagger\sigma^j\widehat{U}(\vec{\vartheta})\xi^\pm \tag{5.6.187}$$

$$= (\widehat{R}\cdot\widehat{p})^i{}_j(\widehat{R}\cdot\vec{\sigma})^j\xi^\pm = \widehat{U}(\vec{\vartheta})\widehat{p}\cdot\widehat{R}^{\mathrm{T}}\widehat{R}\cdot\vec{\sigma}\xi^\pm \tag{5.6.188}$$

$$= \widehat{U}(\vec{\vartheta})(\widehat{p}\cdot\vec{\sigma})\xi^\pm = \pm\widehat{U}(\vec{\vartheta})\xi^\pm. \tag{5.6.189}$$

In words: since the SU_2 group element $\widehat{U}(\vec{\vartheta})$ is supposed to 'represent' the 3D rotation $\widehat{R}(\vec{\vartheta})$ (but acting on spinors instead), the eigenvectors $\xi'^\pm$ of the helicity operator $\widehat{p}'\cdot\vec{\sigma}$, where $\widehat{R}\cdot\widehat{p} = \widehat{p}'$, is gotten by simply 'rotating' the eigenvectors $\xi^\pm$ of $\widehat{p}\cdot\vec{\sigma}$ – namely $\xi'^\pm = \widehat{U}(\vec{\vartheta})\xi^\pm$.

Problem 5.89. Rotating 'spin-up/down' If $\widehat{U}$ in eq. (5.6.158) does in fact implement rotations on 2 component spinors, then we should be able to obtain the eigenvectors of $\widehat{p} \cdot \vec{\sigma}$, where $\widehat{p} \equiv (\sin(\theta_p)\cos(\phi_p), \sin(\theta_p)\sin(\phi_p), \cos(\theta_p))$ is a generic unit radial vector, by rotating the eigenvectors of $\widehat{\sigma}^3 \equiv \widehat{e}_3 \cdot \vec{\sigma}$. In particular, we know that

$$\widehat{\sigma}^3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \widehat{\sigma}^3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = - \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{5.6.190}$$

To rotate the unit vector $\widehat{e}_3$ along the 3−axis to the general $\widehat{r}$ direction, we may first rotate it by θ around the unit vector $\widehat{e}_2$ parallel to the 2−axis. Then rotate the result by ϕ around the unit vector $\widehat{e}_3$. Show that the result yields equations (4.3.193) and (4.3.194), up to possibly an overall multiplicative phase factor. □

Position representation? One may wonder if a position representation $\langle \theta, \phi | \frac{1}{2} \ \pm \frac{1}{2} \rangle$ exists. As Sakurai [11] explains, this turns out to be impossible. Using raising/lowering operators,

$$\left\langle \theta, \phi \left| J^{\pm} \right| \frac{1}{2}, \pm\frac{1}{2} \right\rangle = 0 \tag{5.6.191}$$

we may obtain

$$\left\langle \theta, \phi \left| \frac{1}{2}, \ \pm\frac{1}{2} \right. \right\rangle \propto \sqrt{\sin(\theta)} \exp(\pm i\phi/2). \tag{5.6.192}$$

But if we, say, start from the $\langle \theta, \phi | 1/2, +1/2 \rangle \propto \sqrt{\sin(\theta)} \exp(i\phi/2)$ solution above and apply the lowering operator, we would not obtain the above $\langle \theta, \phi | 1/2, -1/2 \rangle \propto \sqrt{\sin(\theta)} \exp(-i\phi/2)$. Instead, from eq. (5.6.109),

$$\left\langle \theta, \phi \left| \frac{1}{2}, -\frac{1}{2} \right. \right\rangle \propto \left\langle \theta, \phi \left| J^{-} \right| \frac{1}{2}, \ \frac{1}{2} \right\rangle = e^{-i\phi} \left(-\partial_\theta + i\cot(\theta)\partial_\phi \right) \left\langle \theta, \phi \left| \frac{1}{2}, +\frac{1}{2} \right. \right\rangle \tag{5.6.193}$$

$$\propto e^{-\frac{i}{2}\phi} \cos(\theta)/\sqrt{\sin(\theta)}. \tag{5.6.194}$$

5.6.4 'Adding' Angular Momentum, Tensor Operators, Wigner-Eckart Theorem

In this section, we will consider how to 'add' angular momentum. By that, we really mean the study of the vector space spanned by the orthonormal basis $\{|\ell_1 \ m_1, \ell_2 \ m_2\rangle\}$ formed from the tensor product of two separate angular momentum spaces:

$$|\ell_1 \ m_1, \ell_2 \ m_2\rangle \equiv |\ell_1 \ m_1\rangle \otimes |\ell_2 \ m_2\rangle ; \tag{5.6.195}$$

where the $|\ell_1 \ m_1\rangle$ and $|\ell_2 \ m_2\rangle$ are the eigenvectors of two separate sets of angular momentum operators $\{\vec{J}'^2 = J'^i J'^i, J'^3\}$ and $\{\vec{J}''^2 = J''^i J''^i, J''^3\}$. Namely, we have 2 distinct sets of eq. (5.6.56):

$$\vec{J}'^2 \, |\ell_1 \ m_1\rangle = \ell_1(\ell_1 + 1) \, |\ell_1 \ m_1\rangle \tag{5.6.196}$$

$$J'^3 \, |\ell_1 \ m_1\rangle = m_1 \, |\ell_1 \ m_1\rangle \tag{5.6.197}$$

and

$$\vec{J}''^2 \left|\ell_2\ m_2\right\rangle = \ell_2(\ell_2+1)\left|\ell_2\ m_2\right\rangle \tag{5.6.198}$$

$$J''^3 \left|\ell_2\ m_2\right\rangle = m_2 \left|\ell_2\ m_2\right\rangle. \tag{5.6.199}$$

The situation represented by eq. (5.6.195) has widespread applications in physics. For instance, it occurs frequently in atomic and nuclear physics, where one of the kets on the right hand side represent orbital angular momentum and the other intrinsic spin.

The 'addition' of angular momentum comes from defining the 'total' angular momentum operator

$$J^i \equiv J'^i \otimes \mathbb{I} + \mathbb{I} \otimes J''^i. \tag{5.6.200}$$

Oftentimes, the $\otimes$ is dropped for notational convenience.

$$J^i \equiv J'^i + J''^i. \tag{5.6.201}$$

By assumption, the 2 separate sets of angular momentum operators commute

$$\left[J'^a, J''^b\right] = 0. \tag{5.6.202}$$

This allows us to see that, the exponential of the total angular momentum operator yields the rotation operator that implements rotations on the states $\{\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle\}$.

$$\exp\left(-i\theta^a J^a\right)\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle = \exp\left(-i\theta^a(J'^a + J''^a)\right)\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle \tag{5.6.203}$$

$$= \left(\exp\left(-i\vec{\theta}\cdot\vec{J}'\right)\left|\ell_1\ m_1\right\rangle\right) \otimes \left(\exp\left(-i\vec{\theta}\cdot\vec{J}''\right)\left|\ell_2\ m_2\right\rangle\right) \tag{5.6.204}$$

$$= \left(D\left(\widehat{R}(\vec{\theta})\right)\left|\ell_1\ m_1\right\rangle\right) \otimes \left(D\left(\widehat{R}(\vec{\theta})\right)\left|\ell_2\ m_2\right\rangle\right). \tag{5.6.205}$$

In words: rotating the state $\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle$ means simultaneously rotating the $\left|\ell_1\ m_1\right\rangle$ and $\left|\ell_2\ m_2\right\rangle$; this is precisely what exponential of the total angular momentum operator does. Note that $\exp(-i\theta^a(J'^a + J''^a)) = \exp(-i\theta^a J'^a)\exp(-i\theta^a J''^a)$ because J'^a commutes with J''^b.

Eigensystems We may see from equations (5.6.196) through (5.6.199) that the tensor product state $\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle$ are, too, eigenstates of

$$\{\vec{J}'^2, J'^3, \vec{J}''^2, J''^3\}. \tag{5.6.206}$$

For example,

$$\begin{aligned}\vec{J}'^2 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle &= \vec{J}'^2 \left|\ell_1\ m_1\right\rangle \otimes \left|\ell_2 m_2\right\rangle \\ &= \ell_1(\ell_1+1)\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle\end{aligned} \tag{5.6.207}$$

and

$$J'^3 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle = m_1 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle. \tag{5.6.208}$$

Likewise

$$\vec{J}''^2 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle = \ell_2(\ell_2+1)\left|\ell_1\ m_1, \ell_2\ m_2\right\rangle \tag{5.6.209}$$

$$J''^3 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle = m_2 \left|\ell_1\ m_1, \ell_2\ m_2\right\rangle . \tag{5.6.210}$$

We will now proceed to argue that, instead of the mutually compatible observables in eq. (5.6.206), one may also pick the set

$$\left\{\vec{J}^2, J^3, \vec{J}'^2, \vec{J}''^2\right\} . \tag{5.6.211}$$

Their simultaneous eigenstates will be denoted as $\{|j\ m; \ell_1\ \ell_2\rangle\}$, obeying the relations

$$\vec{J}^2 \left|j\ m; \ell_1\ell_2\right\rangle = j(j+1) \left|j\ m; \ell_1\ell_2\right\rangle , \tag{5.6.212}$$

$$J^3 \left|j\ m; \ell_1\ell_2\right\rangle = m \left|j\ m; \ell_1\ell_2\right\rangle , \tag{5.6.213}$$

$$\vec{J}'^2 \left|j\ m; \ell_1\ell_2\right\rangle = \ell_1(\ell_1+1) \left|j\ m; \ell_1\ell_2\right\rangle \tag{5.6.214}$$

$$\vec{J}''^2 \left|j\ m; \ell_1\ell_2\right\rangle = \ell_2(\ell_2+1) \left|j\ m; \ell_1\ell_2\right\rangle . \tag{5.6.215}$$

The total angular momentum j will turn out to be restricted within the range

$$j \in \{|\ell_1 - \ell_2|, |\ell_1 - \ell_2| + 1, \dots, \ell_1 + \ell_2 - 1, \ell_1 + \ell_2\} . \tag{5.6.216}$$

and, of course,

$$m \in \{-j, -j+1, \dots, j-1, j\} . \tag{5.6.217}$$

Problem 5.90. Explain why the total angular momentum generators still obey the Lie Algebra in eq. (5.6.24). That is, verify

$$\left[J^a, J^b\right] = i\epsilon^{abc} J^c. \tag{5.6.218}$$

From the discussions in the previous sections, we see that upon diagonalization, equations (5.6.212), (5.6.213) and (5.6.217) follow. □

Eq. (5.6.218) tells us the total angular momentum operators $\{J^a\}$ are vector operators – recall eq. (5.6.84). Therefore, referring to eq. (5.6.85), $\vec{J}^2$ must be a scalar.

$$\left[\vec{J}^2, J^a\right] = 0 \tag{5.6.219}$$

Moreover, since $[J'^a, J''^b] = 0$, that means the angular momentum operators acting on the individual ℓ_1- and ℓ_2-spaces are also vector operators:

$$\left[J^a, J'^b\right] = i\epsilon^{abc} J'^c, \tag{5.6.220}$$

$$\left[J^a, J''^b\right] = i\epsilon^{abc} J''^c. \tag{5.6.221}$$

These relations in turn informs us, again via eq. (5.6.85), the $\vec{J}'^2$ and $\vec{J}''^2$ are scalars.

$$\left[J^a, \vec{J}'^2\right] = 0 = \left[J^a, \vec{J}''^2\right] \tag{5.6.222}$$

Thus,

$$\left[\vec{J}^2, \vec{J}'^2\right] = 0 = \left[\vec{J}^2, \vec{J}''^2\right]. \tag{5.6.223}$$

Of course, $[J'^a, J''^b] = 0$ also implies

$$[\vec{J}'^2, \vec{J}''^2] = 0. \tag{5.6.224}$$

At this point, we have checked that the following $\binom{4}{2} = 6$ commutators are zero:

$$\left[\vec{J}^2, J^3\right], \ \left[\vec{J}^2, \vec{J}'^2\right], \ \left[\vec{J}^2, \vec{J}''^2\right], \ \left[J^3, \vec{J}'^2\right], \ \left[J^3, \vec{J}''^2\right] \text{ and } \left[\vec{J}'^2, \vec{J}''^2\right]. \tag{5.6.225}$$

We have verified that eq. (5.6.211) consists of a set of mutually compatible observables.

Problem 5.91. Note, however, that *none* of the individual components of J'^i or J''^i commute with $\vec{J}^2$. Show that

$$\left[\vec{J}^2, J'^i\right] = -2i\left(\vec{J}'' \times \vec{J}'\right)^i, \tag{5.6.226}$$

$$\left[\vec{J}^2, J''^i\right] = -2i\left(\vec{J}' \times \vec{J}''\right)^i; \tag{5.6.227}$$

where, for vector operators $\vec{A}$ and $\vec{B}$, we have defined

$$(\vec{A} \times \vec{B})^i \equiv \epsilon^{iab} A^a B^b. \tag{5.6.228}$$

Recalling the discussion in Problem (5.6.85), we see these commutators are non-zero because J'^i generates rotation only on the $|\ell_1, m_1\rangle$ space; and J''^i only the $|\ell_2, m_2\rangle$ space. Hence, only the $\vec{J}'$ operators in $\vec{J}^2$ are altered for the former; and only the $\vec{J}''$ operators are transformed for the latter. $\square$

Change-of-basis & Clebsch-Gordan Coefficients How does one switch between the basis $\{|\ell_1 m_1, \ell_2 m_2\rangle\}$ and $\{|j\ m; \ell_1\ell_2\rangle\}$? Here, we will attempt to do so by computing the Clebsch-Gordan coefficients $\{\langle \ell_1 m_1, \ell_2 m_2 |\, j\ m; \ell_1\ell_2\rangle\}$ occurring within the change-of-basis expansion

$$|j\ m; \ell_1\ell_2\rangle = \sum_{-\ell_1 \leq m_1 \leq \ell_1} \sum_{-\ell_2 \leq m_2 \leq \ell_2} |\ell_1\ m_1, \ell_2\ m_2\rangle \langle \ell_1\ m_1, \ell_2\ m_2 |\, j\ m; \ell_1\ell_2\rangle. \tag{5.6.229}$$

There is no sum over the ℓs, because $|\ell_1' m_1', \ell_2' m_2'\rangle$ would be a simultaneous eigenvector of $\vec{J}'^2$ (or $\vec{J}''^2$) but different eigenvalues from $|j\ m; \ell_1\ell_2\rangle$, whenever $\ell_1 \neq \ell_1'$ (or $\ell_2 \neq \ell_2'$). In such a situation, remember $\langle \ell_1' m_1', \ell_2' m_2' |\, j\ m; \ell_1\ell_2\rangle = 0$. Within this (ℓ_1, ℓ_2) subspace, we therefore have

$$\sum_{-\ell_1 \leq m_1 \leq \ell_1} \sum_{-\ell_2 \leq m_2 \leq \ell_2} |\ell_1\ m_1, \ell_2\ m_2\rangle \langle \ell_1\ m_1, \ell_2\ m_2| = \mathbb{I}. \tag{5.6.230}$$

To begin, let us first notice that

$$J^3 |j\ m; \ell_1\ell_2\rangle = m\, |j\ m; \ell_1\ell_2\rangle \tag{5.6.231}$$

$$= \sum_{m_1',m_2'} (J'^3 + J''^3) |\ell_1 m_1', \ell_2 m_2'\rangle \langle \ell_1 m_1', \ell_2 m_2'| j\ m; \ell_1\ell_2\rangle \tag{5.6.232}$$

$$= \sum_{m_1',m_2'} (m_1' + m_2') |\ell_1 m_1', \ell_2 m_2'\rangle \langle \ell_1 m_1', \ell_2 m_2'| j\ m; \ell_1\ell_2\rangle . \tag{5.6.233}$$

Applying $\langle \ell_1 m_1, \ell_2 m_2|$ on both sides, and employing the orthonormality of these eigen states, we deduce that the superposition over $\{|\ell_1 m_1, \ell_2 m_2\rangle\}$ in eq. (5.6.229) must be constrained by

$$m = m_1 + m_2. \tag{5.6.234}$$

Now, the largest possible m, which is also the maximum j (cf. (5.6.217)), is gotten from $\max(m_1 + m_2) = \max m_1 + \max m_2 = \ell_1 + \ell_2$.

$$\max m = \ell_1 + \ell_2 = \max j. \tag{5.6.235}$$

A similar argument informs us, $\min m = \min m_1 + \min m_2 = -(\ell_1 + \ell_2)$. Altogether,

$$\begin{aligned} |j = \ell_1 + \ell_2 \quad m = \pm(\ell_1 + \ell_2); \ell_1\ell_2\rangle &= |\ell_1 \quad \pm \ell_1, \ell_2 \quad \pm \ell_2\rangle \\ &= |\ell_1, \pm\ell_1\rangle \otimes |\ell_2, \pm\ell_2\rangle . \end{aligned} \tag{5.6.236}$$

Problem 5.92. Checking eq. (5.6.236) Defining the total raising (+) and lowering (−) operator as

$$J^{\pm} \equiv J^1 \pm i J^2 = J'^{\pm} + J''^{\pm}, \tag{5.6.237}$$

verify the relation

$$\vec{J}^2 = \vec{J}'^2 + \vec{J}''^2 + 2(J'^3)(J''^3) + J'^+ J''^- + J'^- J''^+. \tag{5.6.238}$$

Use it to directly calculate the result of acting $\vec{J}^2$ on both sides of eq. (5.6.236). □

We may now follow the procedure we used to relate the $|\ell, m\rangle$ with $|\ell, \pm\ell\rangle$, using the raising/lowering operators in eq. (5.6.237). We recall eq. (5.6.61):

$$J^{\pm} |j\ m; \ell_1\ell_2\rangle = \sqrt{j(j+1) - m(m \pm 1)}\, |j, m \pm 1; \ell_1\ell_2\rangle \tag{5.6.239}$$

$$= \sqrt{(j \mp m)(j \pm m + 1)}\, |j, m \pm 1; \ell_1\ell_2\rangle . \tag{5.6.240}$$

On the other hand, using $J^{\pm} = J'^{\pm} + J''^{\pm}$,

$$\begin{aligned} &\sqrt{(j \mp m)(j \pm m + 1)}\, |j\ m \pm 1; \ell_1\ell_2\rangle \\ &= \sum_{m_1,m_2} (J'^{\pm} + J''^{\pm}) |\ell_1 m_1, \ell_2 m_2\rangle \langle \ell_1 m_1, \ell_2 m_2| j\ m; \ell_1\ell_2\rangle \qquad (5.6.241) \\ &= \sum_{m_1,m_2} \Big(\sqrt{(\ell_1 \mp m_1)(\ell_1 \pm m_1 + 1)}\, |\ell_1\ m_1 \pm 1, \ell_2 m_2\rangle \\ &\qquad + \sqrt{(\ell_2 \mp m_2)(\ell_2 \pm m_2 + 1)}\, |\ell_1\ m_1, \ell_2\ m_2 \pm 1\rangle \Big) \langle \ell_1 m_1, \ell_2 m_2| j\ m; \ell_1\ell_2\rangle . \qquad (5.6.242) \end{aligned}$$

Let us now apply the lowering operator to the maximum possible m state in eq. (5.6.236). For $j = \ell_1 + \ell_2$,

$$\sqrt{(j+j)(j-j+1)} \left| j \ j-1; \ell_1 \ell_2 \right\rangle = \sqrt{(\ell_1+\ell_1)(\ell_1-\ell_1+1)} \left| \ell_1 \ \ell_1 - 1, \ell_2 \ell_2 \right\rangle + \sqrt{(\ell_2+\ell_2)(\ell_2-\ell_2+1)} \left| \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right\rangle . \tag{5.6.243}$$

Because there is only one ket on both sides, we have managed to solve the next-to-highest m state (for the maximum j) in terms of the tensor product ones:

$$\left| j \ j-1; \ell_1 \ell_2 \right\rangle = \frac{1}{\sqrt{j}} \left(\sqrt{\ell_1} \left| \ell_1 \ \ell_1 - 1, \ell_2 \ell_2 \right\rangle + \sqrt{\ell_2} \left| \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right\rangle \right), \tag{5.6.244}$$

$$j \equiv \ell_1 + \ell_2. \tag{5.6.245}$$

We may continue this 'lowering procedure' to obtain all the m states $\left| j = \ell_1 + \ell_2 \ m; \ell_1 \ell_2 \right\rangle$ until we reach $\left| j \ -j; \ell_1 \ell_2 \right\rangle \propto (J^-)^{2j} \left| j \ j; \ell_1 \ell_2 \right\rangle$.[38]

Now that we see how to construct the maximum j states, with $j = \ell_1 + \ell_2$, let us move on to the construction of the next-to-highest j states. Since this next-to-highest j must be equal to its highest m value, according to eq. (5.6.234) it must be an integer step away from the highest j because $m = m_1 + m_2$ are integer steps away from $\ell_1 + \ell_2$. In other words, the next-to-highest j and its associated maximum m value must both be $j = \ell_1 + \ell_2 - 1 = \max m$. Furthermore, according to eq. (5.6.234), we need to superpose all states consistent with $m = \ell_1 + \ell_2 - 1$. Only two such states, $(m_1 = \ell_1 - 1, m_2 = \ell_2)$ and $(m_1 = \ell_1, m_2 = \ell_2 - 1)$, are relevant:

$$\left| j = \ell_1 + \ell_2 - 1 \ j; \ell_1 \ell_2 \right\rangle = \left| \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 \right\rangle \left\langle \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 \right| j \ j; \ell_1 \ell_2 \rangle + \left| \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right\rangle \left\langle \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right| j \ j; \ell_1 \ell_2 \rangle . \tag{5.6.246}$$

Above in eq. (5.6.244), we have already constructed the highest-j state with the same m value as the next-to-highest-j state in eq. (5.6.246), namely $\left| j = \ell_1 + \ell_2 \ j-1; \ell_1 \ell_2 \right\rangle$, which has the same m value as $\left| j = \ell_1 + \ell_2 - 1 \ j; \ell_1 \ell_2 \right\rangle$. These two states must be orthogonal because they have different $\vec{J}^2$ eigenvalues. Taking their inner product and setting it to zero,

$$\sqrt{\ell_1} \left\langle \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 \right| j \ j; \ell_1 \ell_2 \rangle + \sqrt{\ell_2} \left\langle \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right| j \ j; \ell_1 \ell_2 \rangle = 0. \tag{5.6.247}$$

Inserting it back into eq. (5.6.246),

$$\left| j = \ell_1 + \ell_2 - 1 \ j; \ell_1 \ell_2 \right\rangle = \left(-\sqrt{\frac{\ell_2}{\ell_1}} \left| \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 \right\rangle + \left| \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right\rangle \right) \times \left\langle \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right| j \ j; \ell_1 \ell_2 \rangle . \tag{5.6.248}$$

Since this state needs to be normalized to unity, we have up to an arbitrary phase $e^{i\delta_{\ell_1+\ell_2-1}}$,

$$\left| j = \ell_1 + \ell_2 - 1 \ j; \ell_1 \ell_2 \right\rangle = \frac{e^{i\delta_{\ell_1+\ell_2-1}}}{\sqrt{\ell_1+\ell_2}} \left(-\sqrt{\ell_2} \left| \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 \right\rangle + \sqrt{\ell_1} \left| \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 1 \right\rangle \right). \tag{5.6.249}$$

[38] Actually, we know from the preceding arguments that $\left| j = \ell_1 + \ell_2 \ -j; \ell_1 \ell_2 \right\rangle = \left| \ell_1 \ -\ell_1, \ell_2 \ -\ell_2 \right\rangle$. But if you do push the analysis all the way till $(J^-)^{2j} \left| j \ j; \ell_1 \ell_2 \right\rangle$, this would serve as a consistency check.

As before, we may then apply the lowering operator repeatedly to obtain all the $j = \ell_1 + \ell_2 - 1$ states, namely

$$|j = \ell_1 + \ell_2 - 1 \ j - s; \ell_1 \ell_2\rangle \propto (J^-)^s \, |j = \ell_1 + \ell_2 - 1 \ j; \ell_1 \ell_2\rangle \, . \tag{5.6.250}$$

Moving on to the states $\{|j = \ell_1 + \ell_2 - 2 \ m; \ell_1 \ell_2\rangle\}$, we may again begin with the highest m value. This may be expressed as a superposition of tensor product states involving – by distributing -2 among the the (m_1, m_2)s –

$$(m_1 = \ell_1 - 2, m_2 = \ell_2), \qquad (m_1 = \ell_1 - 1, m_2 = \ell_2 - 1), \qquad (m_1 = \ell_1, m_2 = \ell_2 - 2). \tag{5.6.251}$$

The $|j = \ell_1 + \ell_2 - 2 \ j; \ell_1 \ell_2\rangle$ must be perpendicular to both

$$|j = \ell_1 + \ell_2 \ m = j - 2\rangle \ \text{ and } \ |j = \ell_1 + \ell_2 - 1 \ m = j - 1\rangle \tag{5.6.252}$$

because they have different $\vec{J}^2$ eigenvalues. Setting to zero

$$\begin{aligned} &\langle j = \ell_1 + \ell_2 \ m = j - 2 | \, j' = \ell_1 + \ell_2 - 2 \ m' = j' \rangle \\ \text{and } \ &\langle j = \ell_1 + \ell_2 - 1 \ m = j - 1 | \, j' = \ell_1 + \ell_2 - 2 \ m' = j' \rangle \end{aligned} \tag{5.6.253}$$

yields 2 equations for 3 unknown Clebsch-Gordan coefficients

$$\langle \ell_1 \ \ell_1 - 2, \ell_2 \ \ell_2 | \, j = \ell_1 + \ell_2 - 2 \ j; \ell_1 \ell_2 \rangle \, , \tag{5.6.254}$$

$$\langle \ell_1 \ \ell_1 - 1, \ell_2 \ \ell_2 - 1 | \, j = \ell_1 + \ell_2 - 2 \ j; \ell_1 \ell_2 \rangle \, , \tag{5.6.255}$$

$$\langle \ell_1 \ \ell_1, \ell_2 \ \ell_2 - 2 | \, j = \ell_1 + \ell_2 - 2 \ j; \ell_1 \ell_2 \rangle \, . \tag{5.6.256}$$

This allows us to solve 2 of them in terms of a third. This remaining coefficient can then be fixed, up to an overall phase, by demanding the state has unit norm. Once this is done, all the $j = \ell_1 + \ell_2 - 2$ and $m < j$ states may be obtained by applying the lowering operator repeatedly.

This process can continue for the $j = \ell_1 + \ell_2 - 3, \ \ell_1 + \ell_2 - 4$ states, and so on. But it will have to terminate, since we know from the tensor product

$$|\ell_1, m_1\rangle \otimes |\ell_2, m_2\rangle \equiv |\ell_1 \ m_1, \ell_2 \ m_2\rangle \tag{5.6.257}$$

there are $N \equiv (2\ell_1 + 1)(2\ell_2 + 1)$ such orthonormal basis vectors; i.e., the dimension of the vector space, for fixed $\ell_{1,2}$, is N. On the other hand, we know the $j = \ell_1 + \ell_2$ states have $(2\ell_1 + 2\ell_2 + 1)$ distinct m values; the $j = \ell_1 + \ell_2 - 1$ ones have $(2\ell_1 + 2\ell_2 - 2 + 1)$ distinct m values; and so on. Let's suppose our procedure terminates at $j = \ell_1 + \ell_2 - s$, for some non-negative integer s. Then we may count the total number of orthonormal states as

$$N = (2\ell_1 + 1)(2\ell_2 + 1) = \sum_{i=0}^{s} (2\ell_1 + 2\ell_2 - 2i + 1) \tag{5.6.258}$$

$$= (2\ell_1 + 2\ell_2 + 1)(s + 1) - 2\frac{s + 0}{2}(s + 1) = (2\ell_1 + 2\ell_2 + 1)(s + 1) - s(s + 1). \tag{5.6.259}$$

This quadratic equation for s has two solutions

$$s = 2\ell_1 \qquad \Rightarrow \qquad j = \ell_2 - \ell_1, \tag{5.6.260}$$

$$\text{or} \qquad s = 2\ell_2 \qquad \Rightarrow \qquad j = \ell_1 - \ell_2. \tag{5.6.261}$$

Since $j \geq 0$, the correct solution is $j = \ell_1 - \ell_2$ whenever $\ell_1 > \ell_2$ and $j = \ell_2 - \ell_1$ whenever $\ell_2 > \ell_1$. As already alluded to earlier,

$$\min j = |\ell_1 - \ell_2|. \tag{5.6.262}$$

Problem 5.93. From $|j\ j-1\rangle$ to $|j-1\ j-1\rangle$ YZ: Ignore this problem for now. Use equations (5.6.226) and (5.6.227) to show that

$$\left[\vec{J}^2, J'^{\pm}\right] = \mp 2J'^3 J''^{+} \pm 2J'^{+} J''^{3}, \tag{5.6.263}$$

$$\left[\vec{J}^2, J''^{\pm}\right] = \mp 2J''^3 J'^{+} \pm 2J''^{+} J'^{3}. \tag{5.6.264}$$

Consider

$$\vec{J}^2 J'^{\pm} = \left[\vec{J}^2, J'^{\pm}\right] + J'^{\pm}\vec{J}^2 \tag{5.6.265}$$

$$= \mp 2J'^3 J''^{\pm} \pm 2J'^{\pm} J''^{3} + J'^{\pm}\vec{J}^2, \tag{5.6.266}$$

$$\vec{J}^2 J''^{\pm} = \left[\vec{J}^2, J''^{\pm}\right] + J''^{\pm}\vec{J}^2 \tag{5.6.267}$$

$$= \mp 2J''^3 J'^{\pm} \pm 2J''^{\pm} J'^{3} + J''^{\pm}\vec{J}^2. \tag{5.6.268}$$

Example: Tensor product of spin-1/2 systems Consider the tensor product of two spin-1/2 systems. Eq. (5.6.216) informs us, the total angular momentum j runs from $|1/2 - 1/2| = 0$ to $1/2 + 1/2 = 1$.

$$j \in \{0, 1\}. \tag{5.6.269}$$

To save notational baggage, let us denote

$$|{+}{+}\rangle \equiv \left|\frac{1}{2}, \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle, \tag{5.6.270}$$

$$|{+}{-}\rangle \equiv \left|\frac{1}{2}, \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle, \tag{5.6.271}$$

$$|{-}{+}\rangle \equiv \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle, \tag{5.6.272}$$

$$|{-}{-}\rangle \equiv \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle. \tag{5.6.273}$$

We will also suppress the (1/2)s in the total j states; i.e.,

$$\left|j\ m; \frac{1}{2}\ \frac{1}{2}\right\rangle \equiv |j\ m\rangle. \tag{5.6.274}$$

$j = 1$ Let us start with $j = 1$. The highest m state is

$$|j = 1\ m = 1\rangle = |{+}{+}\rangle. \tag{5.6.275}$$

Applying the lowering operator gives

$$\sqrt{2}\,|1\ 0\rangle = |-+\rangle + |+-\rangle \tag{5.6.276}$$

$$|1\ 0\rangle = \frac{|-+\rangle + |+-\rangle}{\sqrt{2}}. \tag{5.6.277}$$

Applying the lowering operator once more,

$$\begin{aligned}
&\sqrt{1(1+1)}\,|1\ -1\rangle \\
&= \frac{1}{\sqrt{2}}(J'^- + J''^-)\,|-+\rangle + \frac{1}{\sqrt{2}}(J'^- + J''^-)\,|+-\rangle \qquad (5.6.278)\\
&= \frac{1}{\sqrt{2}}J''^-\,|-+\rangle + \frac{1}{\sqrt{2}}J'^-\,|+-\rangle \qquad (5.6.279)\\
&= \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}-1\right)}\,|--\rangle + \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}-1\right)}\,|--\rangle. \\
&= \sqrt{2}\,|--\rangle. \qquad (5.6.280)
\end{aligned}$$

This final calculation is really a consistency check: we already know, from the previous discussion, that the minimum m is given by $\min m_1 = -1/2$ and $\min m_2 = -1/2$. We gather the results thus far.

$$\left|1\ 1; \frac{1}{2}\ \frac{1}{2}\right\rangle = \left|\frac{1}{2}\ \frac{1}{2}, \frac{1}{2}\ \frac{1}{2}\right\rangle = |++\rangle, \tag{5.6.281}$$

$$\left|1\ 0; \frac{1}{2}\ \frac{1}{2}\right\rangle = \frac{1}{\sqrt{2}}\left|\frac{1}{2}\ -\frac{1}{2}, \frac{1}{2}\ \frac{1}{2}\right\rangle + \frac{1}{\sqrt{2}}\left|\frac{1}{2}\ \frac{1}{2}, \frac{1}{2}\ -\frac{1}{2}\right\rangle = \frac{|-+\rangle + |+-\rangle}{\sqrt{2}}, \tag{5.6.282}$$

$$\left|1\ -1; \frac{1}{2}\ \frac{1}{2}\right\rangle = \left|\frac{1}{2}\ -\frac{1}{2}, \frac{1}{2}\ -\frac{1}{2}\right\rangle = |--\rangle. \tag{5.6.283}$$

$j = 0$ For the $|j = 0\ m = 0\rangle$ state, we need to superpose m_1 and m_2 such that $m_1 + m_2 = 0$. There are only two choices

$$\left(m_1 = \pm\frac{1}{2}, m_2 = \mp\frac{1}{2}\right). \tag{5.6.284}$$

Hence,

$$|0\ 0\rangle = |-+\rangle\,\langle -+|\,0\ 0\rangle + |+-\rangle\,\langle +-|\,0\ 0\rangle. \tag{5.6.285}$$

This state must be perpendicular to $|1\ 0\rangle$ in eq. (5.6.282), because they have distinct $\vec{J}^2$ eigenvalues ($1(1+1)$ vs. 0). Taking their inner product,

$$\langle -+|\,0\ 0\rangle + \langle +-|\,0\ 0\rangle = 0. \tag{5.6.286}$$

At this point,

$$|0\ 0\rangle = (|-+\rangle - |+-\rangle)\,\langle -+|\,0\ 0\rangle. \tag{5.6.287}$$

Because the state has to be normalized to unity, we have now determined it up to a phase $e^{i\delta_0}$:

$$\left|0\ 0; \frac{1}{2}\ \frac{1}{2}\right\rangle = \frac{e^{i\delta_0}}{\sqrt{2}}\left(\left|\frac{1}{2}\ \ -\frac{1}{2}, \frac{1}{2}\ \frac{1}{2}\right\rangle - \left|\frac{1}{2}\ \frac{1}{2}, \frac{1}{2}\ \ -\frac{1}{2}\right\rangle\right) \tag{5.6.288}$$

$$= \frac{e^{i\delta_0}}{\sqrt{2}}\left(|-+\rangle - |+-\rangle\right). \tag{5.6.289}$$

Example: 'Orbital' angular momentum and spin-half Let us now consider taking the tensor product

$$|\ell, m\rangle \otimes \left|\frac{1}{2}, \pm\frac{1}{2}\right\rangle; \tag{5.6.290}$$

for integer $\ell = 0, 1, 2, \dots$ and $-\ell \leq m \leq \ell$. This can be viewed as simultaneously describing the orbital and intrinsic spin of a single electron bound to a central nucleus.
$\ell = 0$ For $\ell = 0$, the only possible total j is $1/2$. Hence,

$$\left|j = \frac{1}{2}\ m = \pm\frac{1}{2}; 0\ \frac{1}{2}\right\rangle = |0, 0\rangle \otimes \left|\frac{1}{2}\ \ \pm\frac{1}{2}\right\rangle. \tag{5.6.291}$$

$\ell \geq 1$ For non-zero ℓ, eq. (5.6.216) says we must have j running from $\ell - 1/2$ to $\ell + 1/2$:

$$j = \ell \pm \frac{1}{2}. \tag{5.6.292}$$

We start from the highest possible m value.

$$\left|j = \ell + \frac{1}{2}\ m = j; \ell\ \frac{1}{2}\right\rangle = |\ell, \ell\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle. \tag{5.6.293}$$

Applying the lowering operator s times, we have on the left hand side

$$(J^-)^s \left|j = \ell + \frac{1}{2}\ m = j; \ell\ \frac{1}{2}\right\rangle = A_s^{\ell+\frac{1}{2}} \left|j = \ell + \frac{1}{2}\ m = j - s\right\rangle, \tag{5.6.294}$$

where the constant $A_s^{\ell+\frac{1}{2}}$ follows from repeated application of eq. (5.6.61)

$$A_s^{\ell+\frac{1}{2}} = \prod_{i=0}^{s-1} \sqrt{(2\ell + 1 - i)(i + 1)}. \tag{5.6.295}$$

Whereas on the right hand side, $(J^-)^s = (J'^- + J''^-)^s$ may be expanded using the binomial theorem since $[J'^-, J''^-] = 0$. Altogether,

$$A_s^{\ell+\frac{1}{2}} \left|j = \ell + \frac{1}{2}\ m = j - s; \ell\ \frac{1}{2}\right\rangle = \sum_{i=0}^{s} \binom{s}{i} (J'^-)^{s-i} |\ell, \ell\rangle \otimes (J''^-)^i \left|\frac{1}{2}, \frac{1}{2}\right\rangle. \tag{5.6.296}$$

But $(J''^-)^i \left|\frac{1}{2}\ \frac{1}{2}\right\rangle = 0$ whenever $i \geq 2$. This means there are only two terms in the sum, which can of course be inferred from the fact that – since the azimuthal number for the spin-half sector can only take 2 values ($\pm 1/2$) – for a fixed total azimuthal number m, there can only be two possible solutions for the $\ell-$sector azimuthal number.

$$A_s^{\ell+\frac{1}{2}} \left| j = \ell + \frac{1}{2} \ m = j - s; \ell \ \frac{1}{2} \right\rangle \tag{5.6.297}$$
$$= (J'^-)^s \left|\ell, \ell\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle + \frac{s!}{(s-1)!}\sqrt{\left(\frac{1}{2}+\frac{1}{2}\right)\left(\frac{1}{2}-\frac{1}{2}+1\right)}(J'^-)^{s-1}\left|\ell, \ell\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle$$
$$= A_s^\ell \left|\ell, \ell - s\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle + s \cdot A_{s-1}^\ell \left|\ell, \ell - s + 1\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle.$$

Here, the constants are

$$A_s^\ell = \prod_{i=0}^{s-1} \sqrt{(2\ell - i)(i+1)}, \tag{5.6.298}$$
$$A_{s-1}^\ell = \prod_{i=0}^{s-2} \sqrt{(2\ell - i)(i+1)}. \tag{5.6.299}$$

Writing them out more explicitly,

$$\sqrt{2\ell+1}\sqrt{1}\sqrt{2\ell}\sqrt{2}\sqrt{2\ell-1}\sqrt{3}\dots\sqrt{2\ell-(s-2)}\sqrt{s}\left| j = \ell + \frac{1}{2} \ m = j - s; \ell \ \frac{1}{2} \right\rangle \tag{5.6.300}$$
$$= \sqrt{2\ell}\sqrt{1}\sqrt{2\ell-1}\sqrt{2}\sqrt{2\ell-2}\sqrt{3}\dots\sqrt{2\ell-(s-1)}\sqrt{s}\left|\ell, \ell - s\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle$$
$$+ (\sqrt{s})^2\sqrt{2\ell}\sqrt{1}\sqrt{2\ell-1}\sqrt{2}\sqrt{2\ell-2}\sqrt{3}\dots\sqrt{2\ell-(s-2)}\sqrt{s-1}\left|\ell, \ell - s + 1\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle.$$

The factors $\sqrt{2\ell}\dots\sqrt{2\ell-(s-2)}$ and $\sqrt{1}\dots\sqrt{s}$ are common throughout.

$$\sqrt{2\ell+1}\left| j = \ell + \frac{1}{2} \ m = j - s; \ell \ \frac{1}{2} \right\rangle$$
$$= \sqrt{2\ell-(s-1)}\left|\ell, \ell - s\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle + \sqrt{s}\left|\ell, \ell - s + 1\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle$$

We use the definition $j - s = \ell + (1/2) - s \equiv m$ to re-express s in terms of m.

$$\left| j = \ell + \frac{1}{2} \ \ m; \ell \ \frac{1}{2} \right\rangle \tag{5.6.301}$$
$$= \frac{1}{\sqrt{2}\sqrt{2\ell+1}}\left(\sqrt{2\ell+2m+1}\left|\ell, m - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle + \sqrt{2\ell-2m+1}\left|\ell, m + \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle\right).$$

(Remember $\ell \pm 1/2$ is half-integer, since ℓ is integer; so the azimuthal number $m \pm 1/2$ itself is an integer.) For the states $|j = \ell - (1/2) \ m\rangle$, we will again see that there are only two terms in the superposition over the tensor product states. For a fixed m, $|j = \ell - (1/2) \ m\rangle$ must be perpendicular to $|j = \ell + (1/2)\ m\rangle$. This allows us to write down its solution (up to an arbitrary phase) by inspecting eq. (5.6.301):

$$\left|j = \ell - \frac{1}{2} \quad m; \ell\ \frac{1}{2}\right\rangle \tag{5.6.302}$$
$$= \frac{e^{i\delta_{\ell-\frac{1}{2}}}}{\sqrt{2}\sqrt{2\ell+1}} \left(\sqrt{2\ell - 2m + 1} \left|\ell, m - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, \frac{1}{2}\right\rangle - \sqrt{2\ell + 2m + 1} \left|\ell, m + \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}, -\frac{1}{2}\right\rangle\right).$$

Problem 5.94. Use the form of $\vec{J}^2$ in eq. (5.6.238) to confirm the right hand sides of equations (5.6.301) and (5.6.302) are indeed its eigenvectors, with respective eigenvalues of $(\ell + 1/2)(\ell + 1/2 + 1)$ and $(\ell - 1/2)(\ell - 1/2 + 1)$. □

Problem 5.95. Express the states $\left|j\ m; \frac{3}{2}\ 1\right\rangle$ in terms of the basis $\{\left|\frac{3}{2}, m_1\right\rangle \otimes |1, m_2\rangle\}$. □

Invariant Subspaces & Clebsch-Gordan Unitarity Among the mutually compatible observables in eq. (5.6.211), we highlight

$$\left[\vec{J}^2, J^3\right] = 0. \tag{5.6.303}$$

Because the rotation operator involves the exponential of the generators $\vec{J}$, that means it also commutes with $\vec{J}^2$.

$$\left[\vec{J}^2, D(\widehat{R})\right] = \left[\vec{J}^2, \exp\left(-i\vec{\theta} \cdot \vec{J}\right)\right] = 0 \tag{5.6.304}$$

This in turn allows us to point out, it is the total angular momentum basis $\{|j\ m; \ell_1 \ell_2\rangle\}$ that spans – for a fixed triplet of (j, ℓ_1, ℓ_2) – an invariant subspace under rotations. For, we may utilize eq. (5.6.304) to compute

$$\vec{J}^2 \left(D(\widehat{R}) |j\ m; \ell_1 \ell_2\rangle\right) = D(\widehat{R}) \vec{J}^2 |j\ m; \ell_1 \ell_2\rangle \tag{5.6.305}$$
$$= j(j+1) D(\widehat{R}) |j\ m; \ell_1 \ell_2\rangle. \tag{5.6.306}$$

In words: both $|j\ m; \ell_1 \ell_2\rangle$ and $D(\widehat{R}) |j\ m; \ell_1 \ell_2\rangle$ are eigenvectors of $\vec{J}^2$, with the same eigenvalue $j(j+1)$. Therefore, under an arbitrary rotation $D(\widehat{R})$, the vector space spanned by $\{|j\ m; \ell_1 \ell_2\rangle\}$ gets rotated into itself – the matrix element

$$\left\langle j'\ m'; \ell_1 \ell_2 \left| D(\widehat{R}) \right| j\ m; \ell_1 \ell_2 \right\rangle \propto \delta^{j'}{}_{j} \tag{5.6.307}$$

is zero unless $j' = j$. In fact, for fixed (ℓ_1, ℓ_2) the matrix of eq. (5.6.307) is a $(2\ell_1 + 1)(2\ell_2 + 1)$-dimensional square one; taking a block-diagonal form, with a unitary matrix comprising each block. If the basis vectors are arranged in the following order,

$$\{|j = \ell_1 + \ell_2 \quad -j \leq m \leq j; \ell_1 \ell_2\rangle\} \tag{5.6.308}$$
$$\{|j = \ell_1 + \ell_2 - 1 \quad -j \leq m \leq j; \ell_1 \ell_2\rangle\} \tag{5.6.309}$$
$$\dots \tag{5.6.310}$$
$$\{|j = |\ell_1 - \ell_2| + 1 \quad -j \leq m \leq j; \ell_1 \ell_2\rangle\} \tag{5.6.311}$$

$$\{|j = |\ell_1 - \ell_2| \quad -j \le m \le j; \ell_1\ell_2\rangle\}\,; \tag{5.6.312}$$

then the uppermost block would be a $(2(\ell_1 + \ell_2) + 1)$-dimensional unitary square matrix; the second (to its lower right) would be a $(2(\ell_1 + \ell_2 - 1) + 1)$-dimensional one; and so on, until the lowest block on the bottom right, which would be a $(2|\ell_1 - \ell_2| + 1)$-dimensional square unitary transformation.

Moreover, note that the Clebsch-Gordan coefficients themselves form a unitary matrix, since they implement a change-of-orthonormal basis – i.e., from $\{|\ell_1\ m_1, \ell_2\ m_2\rangle\}$ to $\{|j\ m; \ell_1\ell_2\rangle\}$ and vice versa. In particular, the inverse relation to eq. (5.6.229) is

$$|\ell_1\ m_1, \ell_2\ m_2\rangle = \sum_{j=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m=-j}^{+j} |j\ m; \ell_1\ell_2\rangle \langle j\ m; \ell_1\ell_2|\ \ell_1\ m_1, \ell_2\ m_2\rangle\,, \tag{5.6.313}$$

with the associated completeness relation

$$\sum_{j=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m=-j}^{+j} |j\ m; \ell_1\ell_2\rangle \langle j\ m; \ell_1\ell_2| = \mathbb{I}. \tag{5.6.314}$$

The unitary character of these Clebsch-Gordan coefficients follow from the completeness relation in equations (5.6.230) and (5.6.314).

$$\sum_{m_1=-\ell_1}^{+\ell_1} \sum_{m_2=-\ell_2}^{+\ell_2} \langle j'\ m'; \ell_1\ell_2|\ \ell_1\ m_1, \ell_2\ m_2\rangle \langle \ell_1\ m_1, \ell_2\ m_2|\ j\ m; \ell_1\ell_2\rangle = \delta^{j'}{}_{j}\delta^{m'}{}_{m} \tag{5.6.315}$$

$$\sum_{j=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m=-j}^{+j} \langle \ell_1\ m_1', \ell_2\ m_2'|\ j\ m; \ell_1\ell_2\rangle \langle j\ m; \ell_1\ell_2|\ \ell_1\ m_1, \ell_2\ m_2\rangle = \delta^{m_1'}{}_{m_1}\delta^{m_2'}{}_{m_2}. \tag{5.6.316}$$

(Irreducible) Spherical Vector & Tensor Operators We may generalize the definition of a vector operator in eq. (5.6.81) to a higher rank tensor $T_{i_1 i_2 \dots i_N}$.

$$D(\widehat{R})^\dagger T_{i_1 i_2 \dots i_N} D(\widehat{R}) = \widehat{R}_{i_1}{}^{j_1} \widehat{R}_{i_2}{}^{j_2} \dots \widehat{R}_{i_N}{}^{j_N} T_{j_1 j_2 \dots j_N} \tag{5.6.317}$$

But we may also remember eq. (5.6.63), where we found a different representation for the rotation operation, one based on the angular momentum eigenkets themselves. Specifically, because the rotation operator $D(\widehat{R})$ leaves invariant the space spanned by $\{|\ell, m\rangle\}$ for a fixed ℓ and this is the smallest such space (i.e., $D(\widehat{R})$ mixes all the m values in general), this basis is said to provide an *irreducible* representation for the rotation operator in eq. (5.6.64). In many physical applications, moreover, it is these angular momentum eigenstates $\{|\ell, m\rangle\}$ that play an important role.

To motivate the definition of irreducible tensors, we shall follow an analogous path that led to eq. (5.6.63); but one that would involve the angular spherical harmonics $Y_\ell^m(\widehat{r})$. We first define the spherical harmonic of the 3D position operator X^i by

$$Y_\ell^m(\vec{X})\,|\vec{x}\rangle \equiv r^\ell Y_\ell^m(\widehat{r})\,|\vec{x} \equiv r\widehat{r}\rangle\,; \tag{5.6.318}$$

where $r\widehat{r}$ is simply the Cartesian coordinates $\vec{x}$ expressed in spherical coordinates, with $r \equiv |\vec{x}|$ and $\widehat{r} = \vec{x}/r$. Remember that $r^\ell Y_\ell^m$ is a homogeneous polynomial of degree ℓ. This means, for instance:

$$Y_0^0(\vec{X}) = \frac{1}{\sqrt{4\pi}}, \tag{5.6.319}$$

and

$$\sqrt{\frac{3}{4\pi}} X^\pm \equiv Y_1^\pm(\vec{X}) = \mp\sqrt{\frac{3}{4\pi}} \frac{X^1 \pm iX^2}{\sqrt{2}}, \tag{5.6.320}$$

$$\sqrt{\frac{3}{4\pi}} X^0 \equiv Y_1^0(\vec{X}) = \sqrt{\frac{3}{4\pi}} X^3. \tag{5.6.321}$$

Now, on the one hand

$$\begin{aligned} D(\widehat{R})^\dagger Y_\ell^m(\vec{X}) D(\widehat{R}) |\vec{x}\rangle &= D(\widehat{R})^\dagger Y_\ell^m(\vec{X}) \left|\widehat{R}\vec{x}\right\rangle \\ &= D(\widehat{R})^\dagger Y_\ell^m(\widehat{R}\vec{x}) \left|\widehat{R}\vec{x}\right\rangle = r^\ell Y_\ell^m(\widehat{R}\widehat{r}) |\vec{x}\rangle ; \end{aligned} \tag{5.6.322}$$

which holds for arbitrary $|\vec{x}\rangle$ and hence

$$D(\widehat{R})^\dagger Y_\ell^m(\vec{X}) D(\widehat{R}) = Y_\ell^m(\widehat{R}\vec{X}). \tag{5.6.323}$$

On the other hand,

$$D(\widehat{R}^{\mathrm{T}}) |\ell, m\rangle = D(\widehat{R})^\dagger |\ell, m\rangle = \sum_{m'} |\ell, m'\rangle \overline{D_{(\ell)}{}^m{}_{m'}}. \tag{5.6.324}$$

Upon acting $\langle\theta, \phi| \equiv \langle\widehat{r}|$ from the left on both sides, and recognizing $\langle\widehat{r}| D(\widehat{R})^\dagger = \left\langle\widehat{R}\widehat{r}\right|$ and $\langle\widehat{r}|\, \ell, m\rangle = Y_\ell^m(\theta, \phi)$,

$$Y_\ell^m(\widehat{R}\widehat{r}) = \sum_{m'} \overline{D_{(\ell)}{}^m{}_{m'}}(\widehat{R}) Y_\ell^{m'}(\widehat{r}). \tag{5.6.325}$$

Comparing equations (5.6.323) and (5.6.325),

$$D(\widehat{R})^\dagger Y_\ell^m(\vec{X}) D(\widehat{R}) = \sum_{m'=-\ell}^{+\ell} Y_\ell^{m'}(\vec{X}) \overline{D_{(\ell)}{}^m{}_{m'}}(\widehat{R}). \tag{5.6.326}$$

Swapping $\widehat{R} \leftrightarrow \widehat{R}^{\mathrm{T}}$ and recalling $D(\widehat{R}^{\mathrm{T}}) = D(\widehat{R})^\dagger$, we see that

$$\begin{aligned} D(\widehat{R}) Y_\ell^m(\vec{X}) D(\widehat{R})^\dagger = Y_\ell^m(\widehat{R}^{\mathrm{T}}\vec{X}) &= \sum_{m'=-\ell}^{+\ell} Y_\ell^{m'}(\vec{X}) D_{(\ell)}{}^{m'}{}_m(\widehat{R}) \\ &= \sum_{m'=-\ell}^{+\ell} Y_\ell^{m'}(\vec{X}) \left\langle \ell, m' \left| D(\widehat{R}) \right| \ell, m \right\rangle. \end{aligned} \tag{5.6.327}$$

In words: we have found an explicit example of a linear operator operator – namely the spherical harmonics of position operators – where a change-of-basis induced by a rotation transforms it in a manner as though it were an angular momentum eigenket $|\ell, m\rangle$; i.e., as if we were doing the right hand side of

$$D(\widehat{R})\,|\ell, m\rangle = \sum_{m'} |\ell, m'\rangle \langle \ell, m'|\, D(\widehat{R})\,|\ell, m\rangle\,. \tag{5.6.328}$$

The relation in eq. (5.6.327) really does not depend on $\vec{X}$ being the position operator; rather, it is really due to $\vec{X}$ being a vector operator. That is, eq. (5.6.327) would still hold if we replaced $\vec{X}$ with any vector operator V^i obeying $D(\widehat{R})^\dagger V^i D(\widehat{R}) = \widehat{R}^i{}_j V^j$.

$$D(\widehat{R}) Y_\ell^m(\vec{V}) D(\widehat{R})^\dagger = Y_\ell^m(\widehat{R}^{\mathrm{T}}\vec{V}) = \sum_{m'=-\ell}^{+\ell} Y_\ell^{m'}(\vec{V}) \left\langle \ell, m' \left| D(\widehat{R}) \right| \ell, m \right\rangle \tag{5.6.329}$$

These considerations motivates the following generalization to arbitrary linear operators O_j^m.

Spherical Tensor: Definition A spherical tensor O_j^m of rank j with $2j+1$ components is defined as a linear operator obeying

$$D(\widehat{R}) O_j^m D(\widehat{R})^\dagger = \sum_{m'=-j}^{+j} O_j^{m'} D_{(j)}{}^{m'}{}_m(\widehat{R}). \tag{5.6.330}$$

The equivalent infinitesimal version is provided by the equations

$$\left[J^i, O_j^m\right] = \sum_{m'=-j}^{+j} O_j^{m'} \left\langle j, m' \left| J^i \right| j, m \right\rangle; \tag{5.6.331}$$

where $J^i = (J^1, J^2, J^3)$ refers to the Cartesian components of the rotation generators.

Problem 5.96. Derive eq. (5.6.331) from eq. (5.6.330). Then explain why

$$\left[J^3, O_j^m\right] = m O_j^m, \tag{5.6.332}$$

$$\left[J^\pm, O_j^m\right] = \sqrt{(j \mp m)(j \pm m + 1)} O_j^{m\pm 1}; \tag{5.6.333}$$

where $J^\pm \equiv J^1 \pm i J^2$ are the raising/lowering angular momentum operators. □

Example We may immediately generalize the results in equations (5.6.320) and (5.6.321) to an arbitrary vector operator V^i. We define

$$V_1^{\pm 1} \equiv \mp \frac{V^1 \pm i V^2}{\sqrt{2}}, \tag{5.6.334}$$

$$V_1^0 \equiv V^3. \tag{5.6.335}$$

In other words, once a 3−axis has been chosen, a Cartesian vector V^i is a spin-1 object; with V^1 and V^2 contributing to its $m = \pm 1$ azimuthal modes and V^0 to its $m = 0$ component. The inverse relations can be summed up by writing the Cartesian components $\vec{V}$ as

$$\vec{V} = \frac{V_1^{-1} - V_1^{+1}}{\sqrt{2}} \widehat{e}_1 + \frac{i}{\sqrt{2}} \left(V_1^{-1} + V_1^{+1}\right) \widehat{e}_2 + V_1^0 \widehat{e}_3, \tag{5.6.336}$$

where $\widehat{e}_i$ is the unit vector along the ith axis.

In particular, the angular momentum operators themselves can be expressed as

$$J_1^{\pm 1} = \mp \frac{J^{\pm}}{\sqrt{2}} \qquad \text{and} \qquad J_1^0 = J^3. \tag{5.6.337}$$

Problem 5.97. Generating spherical tensors from products If $A_{j_1}^{m_1}$ and $B_{j_2}^{m_2}$ are spherical tensors of ranks j_1 and j_2 respectively, explain why the construction

$$Q_j^m = \sum_{|j_1 - j_2| \leq j \leq j_1 + j_2} \sum_{m = m_1 + m_2} A_{j_1}^{m_1} B_{j_2}^{m_2} \langle j_1\ m_1, j_2\ m_2 | j\ m; j_1 j_2 \rangle \tag{5.6.338}$$

produces a spherical tensor Q_j^m. This teaches us, we may superpose the products of spherical tensors to produce another spherical tensor, in the same way we superpose the tensor product of angular momentum eigenstates to produce a 'total' angular momentum state. □

Problem 5.98. Irreducible decomposition of Vector ⊗ Vector Via a direct calculation, show that the following trace, antisymmetric, and symmetric-trace-free decomposition of a product of two vectors, namely

$$V^i W^j = \frac{1}{3}\delta^{ij}\vec{V}\cdot\vec{W} + \frac{V^i W^j - V^j W^i}{2} + \left(\frac{V^i W^j + V^j W^i}{2} - \frac{1}{3}\delta^{ij}\vec{V}\cdot\vec{W}\right); \tag{5.6.339}$$

admits the following irreducible decomposition. First define

$$\widehat{e}_{\pm} \equiv \frac{\widehat{e}_1 \pm i\widehat{e}_2}{\sqrt{2}}. \tag{5.6.340}$$

The trace portion is

$$\vec{V}\cdot\vec{W} = V_1^0 W_1^0 - V_1^{+1} W_1^{-1} - V_1^{-1} W_1^{+1}; \tag{5.6.341}$$

the anti-symmetric sector is

$$\begin{aligned}
\frac{V^i W^j - V^j W^i}{2} &= \frac{i}{2}\left(V_1^{-1} W_1^{+1} - V_1^{+1} W_1^{-1}\right)\left(\widehat{e}_1^i \widehat{e}_2^j - \widehat{e}_1^j \widehat{e}_2^i\right) \\
&\qquad + \frac{1}{2}\frac{V_1^{-1} W_1^0 - V_1^0 W_1^{-1}}{\sqrt{2}}\left\{\left(\widehat{e}_1^i \widehat{e}_3^j - \widehat{e}_1^j \widehat{e}_3^i\right) + i\left(\widehat{e}_2^i \widehat{e}_3^j - \widehat{e}_2^j \widehat{e}_3^i\right)\right\} \\
&\qquad - \frac{1}{2}\frac{V_1^{+1} W_1^0 - V_1^0 W_1^{+1}}{\sqrt{2}}\left\{\left(\widehat{e}_1^i \widehat{e}_3^j - \widehat{e}_1^j \widehat{e}_3^i\right) - i\left(\widehat{e}_2^i \widehat{e}_3^j - \widehat{e}_2^j \widehat{e}_3^i\right)\right\} \\
&= -\frac{1}{2}\left(V_1^{-1} W_1^{+1} - V_1^{+1} W_1^{-1}\right)\left(\widehat{e}_+^i \widehat{e}_-^j - \widehat{e}_+^j \widehat{e}_-^i\right) \\
&\qquad + \frac{1}{2}\left(V_1^{-1} W_1^0 - V_1^0 W_1^{-1}\right)\left(\widehat{e}_+^i \widehat{e}_3^j - \widehat{e}_+^j \widehat{e}_3^i\right) \\
&\qquad - \frac{1}{2}\left(V_1^{+1} W_1^0 - V_1^0 W_1^{+1}\right)\left(\widehat{e}_-^i \widehat{e}_3^j - \widehat{e}_-^j \widehat{e}_3^i\right);
\end{aligned} \tag{5.6.342}$$

and the symmetric and trace-less part is

$$\frac{V^i W^j + V^i W^j}{2} - \frac{1}{3}\delta^{ij}\vec{V}\cdot\vec{W}$$

$$
\begin{aligned}
&= \frac{V_1^{-1}W_1^{-1}}{2}\left\{\left(\widehat{e}_1^i\widehat{e}_1^j - \widehat{e}_2^i\widehat{e}_2^j\right) + i\left(\widehat{e}_1^i\widehat{e}_2^j + \widehat{e}_2^i\widehat{e}_1^j\right)\right\} \\
&\qquad + \frac{V_1^{+1}W_1^{+1}}{2}\left\{\left(\widehat{e}_1^i\widehat{e}_1^j - \widehat{e}_2^i\widehat{e}_2^j\right) - i\left(\widehat{e}_1^i\widehat{e}_2^j + \widehat{e}_2^i\widehat{e}_1^j\right)\right\} \\
&\qquad + \frac{1}{2}\frac{V_1^{-1}W_1^0 + V_1^0W_1^{-1}}{\sqrt{2}}\left\{\left(\widehat{e}_1^i\widehat{e}_3^j + \widehat{e}_1^j\widehat{e}_3^i\right) + i\left(\widehat{e}_2^i\widehat{e}_3^j + \widehat{e}_2^j\widehat{e}_3^i\right)\right\} \\
&\qquad - \frac{1}{2}\frac{V_1^{+1}W_1^0 + V_1^0W_1^{+1}}{\sqrt{2}}\left\{\left(\widehat{e}_1^i\widehat{e}_3^j + \widehat{e}_1^j\widehat{e}_3^i\right) - i\left(\widehat{e}_2^i\widehat{e}_3^j + \widehat{e}_2^j\widehat{e}_3^i\right)\right\} \\
&\qquad - \frac{1}{6}\left(2V_1^0W_1^0 + V_1^{+1}W_1^{-1} + V_1^{-1}W_1^{+1}\right)\left(\widehat{e}_1^i\widehat{e}_1^j + \widehat{e}_2^i\widehat{e}_2^j - 2\widehat{e}_3^i\widehat{e}_3^j\right) \\
&= V_1^{-1}W_1^{-1}\widehat{e}_+^i\widehat{e}_+^j + V_1^{+1}W_1^{+1}\widehat{e}_-^i\widehat{e}_-^j \\
&\qquad + \frac{V_1^{-1}W_1^0 + V_1^0W_1^{-1}}{2}\left(\widehat{e}_+^i\widehat{e}_3^j + \widehat{e}_+^j\widehat{e}_3^i\right) - \frac{V_1^{+1}W_1^0 + V_1^0W_1^{+1}}{2}\left(\widehat{e}_-^i\widehat{e}_3^j + \widehat{e}_-^j\widehat{e}_3^i\right) \\
&\qquad - \frac{1}{6}\left(2V_1^0W_1^0 + V_1^{+1}W_1^{-1} + V_1^{-1}W_1^{+1}\right)\left(\widehat{e}_+^i\widehat{e}_-^j + \widehat{e}_+^i\widehat{e}_-^j - 2\widehat{e}_3^i\widehat{e}_3^j\right).
\end{aligned} \tag{5.6.343}
$$

Identity all the *irreducible* spherical tensors in these expressions; there are $1+3+5=9$ of them; with the "1" coming from the scalar dot product, "3" from the anti-symmetric sector, and "5" from the symmetric and traceless portion. Hint: Taking the product of two vectors is like taking the tensor product of two spin-1 objects – what are the possible outcomes? Also note that, since the $\{\widehat{e}_i\}$ are orthonormal vectors, $\delta^{ab}\widehat{e}_a^i\widehat{e}_b^j = \delta^{ij}$. (Can you explain why?) □

Problem 5.99. Helicity 0, 1 and 2 Suppose linear momentum $\vec{p} = p\widehat{e}_3$ is parallel to the 3$-$axis. Let us consider performing a rotation that leaves it invariant; i.e., rotate the $(1,2)-$plane:

$$\widehat{e}_1 \to \widehat{R}(\phi)\cdot\widehat{e}_1 \equiv \cos(\phi)\widehat{e}_1 - \sin(\phi)\widehat{e}_2, \tag{5.6.344}$$

$$\widehat{e}_2 \to \widehat{R}(\phi)\cdot\widehat{e}_2 \equiv \sin(\phi)\widehat{e}_1 + \cos(\phi)\widehat{e}_2, \tag{5.6.345}$$

$$\widehat{e}_3 \to \widehat{R}(\phi)\cdot\widehat{e}_3 \equiv \widehat{e}_3. \tag{5.6.346}$$

Show that the $\widehat{e}_\pm$ in eq. (5.6.340) rotates as

$$\widehat{e}_\pm \to \widehat{R}(\phi)\cdot\widehat{e}_\pm = e^{\pm i\phi}\widehat{e}_\pm. \tag{5.6.347}$$

We may view $\widehat{e}_\pm$ as helicity eigenstates, with eigenvalues ±1. Explain why the trace part of a rank-2 tensor, eq. (5.6.341), is a zero helicity state. Whereas, identify the terms in the anti-symmetric part in eq. (5.6.342) that has helicities 0 and ±1; and those in the symmetric-traceless ones in (5.6.343) carrying helicities 0, ±1 and ±2. □

Problem 5.100. 'Flow' Induced By Irreducible Decomposition Physical problems – from elasticity to gravitation – often lead to the following matrix equation in 3D (flat) space

$$\left(\frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} \text{ or } \frac{\mathrm{d}x^i}{\mathrm{d}t}\right) = \Sigma^{ij}(t)x^j(t), \tag{5.6.348}$$

where x^i is the Cartesian coordinate displacement vector, and Σ^{ij} is an arbitrary 3×3 matrix. Typically, there is a preferred direction – that of linear momentum, velocity of some flow, etc.

– which we shall choose to be parallel to the 3−axis. Following Problem (5.98), we may then decompose

$$\Sigma^{ij} = \frac{\delta^{ij}}{3}\delta_{ab}\Sigma^{ab} + \frac{1}{2}\Sigma^{[ij]} + \left(\frac{1}{2}\Sigma^{\{ij\}} - \frac{\delta^{ij}}{3}\delta_{ab}\Sigma^{ab}\right). \tag{5.6.349}$$

Use a computer program to plot the vector field representation of its irreducible parts:

$$\delta^{ij}x^j; \tag{5.6.350}$$
$$\left(\widehat{e}^i_+\widehat{e}^j_- - \widehat{e}^j_+\widehat{e}^i_-\right)x^j, \qquad \left(\widehat{e}^i_+\widehat{e}^j_3 - \widehat{e}^j_+\widehat{e}^i_3\right)x^j, \qquad \left(\widehat{e}^i_-\widehat{e}^j_3 - \widehat{e}^j_-\widehat{e}^i_3\right)x^j; \tag{5.6.351}$$
$$\widehat{e}^i_+\widehat{e}^j_+x^j, \qquad \widehat{e}^i_-\widehat{e}^j_-x^j, \qquad \left(\widehat{e}^i_+\widehat{e}^j_3 + \widehat{e}^j_+\widehat{e}^i_3\right)x^j, \qquad \left(\widehat{e}^i_-\widehat{e}^j_3 + \widehat{e}^j_-\widehat{e}^i_3\right)x^j, \tag{5.6.352}$$
$$\left(\widehat{e}^i_+\widehat{e}^j_- + \widehat{e}^i_+\widehat{e}^j_- - 2\widehat{e}^i_3\widehat{e}^j_3\right)x^j. \tag{5.6.353}$$

□

The notion of irreducible spherical tensors allows the effective classification and calculation of matrix elements by exploiting the transformation properties of the operators at hand under rotations. To this end, we first prove the following result.

Lemma If the states $|j,m;\Psi\rangle$ and $|j,m;\Phi\rangle$ obey

$$J^i|j,m;\Psi\rangle = \sum_{m'}|j,m';\Psi\rangle\left\langle j,m'\left|J^i\right|j,m\right\rangle, \tag{5.6.354}$$
$$J^i|j,m;\Phi\rangle = \sum_{m'}|j,m';\Phi\rangle\left\langle j,m'\left|J^i\right|j,m\right\rangle; \tag{5.6.355}$$

then the matrix element

$$\langle j,m;\Psi|Q|j,m;\Phi\rangle \tag{5.6.356}$$

is in fact independent of m, as long as the operator Q commutes with $\{J^i\}$.

We will use the raising/lowering operators. Consider

$$\left\langle j,m\pm1;\Psi\left|QJ^\pm\right|j,m;\Phi\right\rangle = \sqrt{(j\mp m)(j\pm m+1)}\langle j,m\pm1;\Psi|Q|j,m\pm1;\Phi\rangle. \tag{5.6.357}$$

By assumption, the $J^\pm$ may also be moved to the left of Q,

$$\begin{aligned}\left\langle j,m\pm1;\Psi\left|QJ^\pm\right|j,m;\Phi\right\rangle &= \left\langle j,m\pm1;\Psi\left|(J^\mp)^\dagger Q\right|j,m;\Phi\right\rangle \\ &= \sqrt{(j\pm(m\pm1))(j\mp(m\pm1)+1)}\langle j,m;\Psi|Q|j,m;\Phi\rangle.\end{aligned} \tag{5.6.358}$$

Comparing the two results tell us

$$\langle j,m\pm1;\Psi|Q|j,m\pm1;\Phi\rangle = \langle j,m;\Psi|Q|j,m;\Phi\rangle. \tag{5.6.359}$$

This relation may be iterated to show that, since all nearest-neighbors in m-values yield the same matrix element, the $\langle j,m;\Psi|Q|j,m;\Phi\rangle$ must thus yield the same answer regardless of m. We will employ this lemma to prove the Wigner-Eckart theorem.

Wigner-Eckart The matrix element of a tensor operator O_j^m with respect to angular momentum states, namely $\langle j''\ m''; \Phi | O_j^m | j'm'; \Psi\rangle$, is proportional to a matrix element $\langle j''; \Phi || O_j || j'; \Psi\rangle$ that does not depend on the azimuthal numbers m, m', m''.

$$\left\langle j''\ m''; \Phi \left| O_j^m \right| j'\ m'; \Psi \right\rangle = \langle j''; \Phi || O_j || j'; \Psi\rangle \left\langle j''\ m''; j\ j' \right| j\ m, j'\ m'\rangle \tag{5.6.360}$$

The proportionality constant that depends on m, m', m'' is simply the Clebsch-Gordan coefficient obtained from projecting the 'total' angular momentum j'' with azimuthal number m'' onto the tensor product state $|j'\ m'\rangle \otimes |j\ m\rangle$. □

Proof of Wigner-Eckart theorem Let O_j^m be a spherical tensor operator and $|j'm'; \Psi\rangle$ be an angular momentum eigenstate that could also depend on other variables (which we collectively denote as Ψ).

$$\vec{J}^2 |j'm'; \Psi\rangle = j'(j'+1) |j'm'; \Psi\rangle, \tag{5.6.361}$$
$$J^3 |j'm'; \Psi\rangle = m' |j'm'; \Psi\rangle. \tag{5.6.362}$$

We see that

$$\begin{aligned} D(\widehat{R}) \left(O_j^m |j'm'; \Psi\rangle\right) &= D(\widehat{R}) O_j^m D(\widehat{R})^\dagger D(\widehat{R}) |j'm'; \Psi\rangle \\ &= \sum_{n,n'} O_j^n |j'n'; \Psi\rangle \left\langle j, n \left| D(\widehat{R}) \right| j, m \right\rangle \left\langle j', n' \left| D(\widehat{R}) \right| j', m' \right\rangle. \end{aligned} \tag{5.6.363}$$

In other words, this $O_j^m |j'm'; \Psi\rangle$ transforms in the same manner under rotations as the tensor product state $|j\ m\rangle \otimes |j'm'\rangle$.

$$O_j^m |j'm'; \Psi\rangle \leftrightarrow |j\ m\rangle \otimes |j'm'\rangle \tag{5.6.364}$$

Hence it must be possible to use the Clebsch-Gordan coefficients to construct the analog of the 'total angular momentum' state

$$\left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle \equiv \sum_{m+m'=\underline{m}''} O_j^m |j'\ m'; \Psi\rangle \left\langle j\ m, j'\ m' \right| \underline{j}''\underline{m}''; j\ j'\rangle. \tag{5.6.365}$$

By construction, this state obeys

$$\vec{J}^2 \left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle = \underline{j}''(\underline{j}''+1) \left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle, \tag{5.6.366}$$
$$J^3 \left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle = \underline{m}'' \left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle. \tag{5.6.367}$$

This also implies we should be able to invert this relation and solve for

$$O_j^m |j'm'; \Psi\rangle = \sum_{\substack{\underline{j}'' \in \{|j-j'|, |j-j'|+1, \dots, j+j'\} \\ \underline{m}''=m+m'}} \left|\underline{j}''\underline{m}''; j\ j'; O, \Psi\right\rangle \left\langle \underline{j}''\underline{m}''; j\ j' \right| j\ m, j'\ m'\rangle. \tag{5.6.368}$$

If $|j''\ m''; \Phi\rangle$ is another eigenstate of angular momentum (which may depend on other variables, collectively denoted as Φ); then we may project both sides of eq. (5.6.368) with it.

$$\langle j''\ m''; \Phi | O_j^m |j'm'; \Psi\rangle = \langle j''m''; j\ j'; \Phi | j''m''; j\ j'; O, \Psi\rangle \left\langle j''\ m''; j\ j' \right| j\ m, j'\ m'\rangle \tag{5.6.369}$$

But $\langle j''m''; j\ j'; \Phi |\, j''m''; j\ j'; O, \Psi\rangle$ is independent of m''. We have thus arrived at the primary statement.

Example If V^i and W^i are vector operators, we may exploit the Wigner-Eckart theorem to examine their matrix elements between states that transform like angular momentum eigenstates under rotation. We have three distinct ones:

$$\langle \ell, m; \alpha \,| V_1^n |\, \ell', m'; \beta\rangle = \langle \ell\ m; 1\ \ell' |\, 1\ n, \ell' m'\rangle \, \langle \ell; \alpha |\, |V_1|\, |\ell'; \beta\rangle \tag{5.6.370}$$

$$n \in \{\pm 1, 0\}. \tag{5.6.371}$$

Likewise for W^i,

$$\langle \ell, m; \alpha \,| W_1^n |\, \ell', m'; \beta\rangle = \langle \ell\ m; 1\ \ell' |\, 1\ n, \ell' m'\rangle \, \langle \ell; \alpha |\, |W_1|\, |\ell'; \beta\rangle \tag{5.6.372}$$

$$n \in \{\pm 1, 0\}. \tag{5.6.373}$$

Since the Clebsch-Gordan coefficients are common between the two, this means the ratio of the matrix elements in equations (5.6.370) and (5.6.372) only depends on the m-independent matrix elements.

$$\frac{\langle \ell, m; \alpha \,| V_1^n |\, \ell', m'; \beta\rangle}{\langle \ell, m; \alpha \,| W_1^n |\, \ell', m'; \beta\rangle} = \frac{\langle \ell; \alpha |\, |V_1|\, |\ell'; \beta\rangle}{\langle \ell; \alpha |\, |W_1|\, |\ell'; \beta\rangle}. \tag{5.6.374}$$

This must hold for the ratio of the Cartesian components $\{V^i, W^i\}$ too, provided it is the same component in both the numerator and denominator.

$$\frac{\langle \ell, m; \alpha \,| V^i |\, \ell', m'; \beta\rangle}{\langle \ell, m; \alpha \,| W^i |\, \ell', m'; \beta\rangle} = \frac{\langle \ell; \alpha |\, |V_1|\, |\ell'; \beta\rangle}{\langle \ell; \alpha |\, |W_1|\, |\ell'; \beta\rangle}. \tag{5.6.375}$$

Selection rules from Angular Momentum Addition We see that such matrix elements in equations (5.6.370) and (5.6.372) are non-zero only when the following selection rules are satisfied, as dictated by the Clebsch-Gordan coefficient $\langle \ell\ m; 1\ \ell' |\, 1\ n, \ell' m'\rangle$.

$$\ell \in \{|\ell' - 1|, \dots, \ell' + 1\}, \qquad m' + n = m \tag{5.6.376}$$

In other words, ℓ cannot differ from ℓ' by more than one; and similarly for m and m' since $n = \pm 1, 0$.

$$|\ell - \ell'| \leq 1, \qquad |m - m'| \leq 1 \tag{5.6.377}$$

When $\ell' = 0$ then $\ell > 0$; i.e., $\ell = 0 = \ell'$ is forbidden. Therefore, for integer ℓ and ℓ', note that

$$\ell + \ell' \geq 1, \tag{5.6.378}$$

Selection Rules from Parity If V^i is also a vector under parity, namely

$$P^\dagger V^i P = -V^i, \tag{5.6.379}$$

then

$$\left\langle \ell, m; \alpha \,\middle| P^\dagger V_1^n P \middle|\, \ell', m'; \beta \right\rangle = -\left\langle \ell, m; \alpha \,\middle| P V_1^n P^{-1} \middle|\, \ell', m'; \beta \right\rangle. \tag{5.6.380}$$

On the other hand, for integer ℓ and ℓ',

$$\langle \ell, m; \alpha | P^\dagger = (-)^\ell \langle \ell, m; \alpha | \tag{5.6.381}$$

$$P |\ell', m'; \beta\rangle = (-)^{\ell'} |\ell', m'; \beta\rangle , \tag{5.6.382}$$

which implies

$$\left\langle \ell, m; \alpha \left| P^\dagger V_1^n P \right| \ell', m'; \beta \right\rangle = (-)^{\ell+\ell'} \langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle . \tag{5.6.383}$$

Altogether,

$$\langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle = (-)^{\ell+\ell'+1} \langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle . \tag{5.6.384}$$

We conclude: if V^i is a vector under parity and if ℓ and ℓ' are integers, only when the difference of the latter is an odd number, namely $|\ell-\ell'| = 2n+1$ for $n = 0, 1, 2, 3, \dots$, is the corresponding matrix element $\langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle$ non-zero. But we have already found $|\ell - \ell'| \leq 1$; hence,

$$|\ell - \ell'| = 1 \tag{5.6.385}$$

Electric dipole transitions in quantum mechanics are described by replacing $\vec{V}$ with the position operator $\vec{X}$ in eq. (5.6.370).

If, instead, V^i were a pseudo-vector under parity,

$$P^\dagger V^i P = V^i, \tag{5.6.386}$$

we'd discover that

$$\langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle = (-)^{\ell+\ell'} \langle \ell, m; \alpha | V_1^n | \ell', m'; \beta\rangle . \tag{5.6.387}$$

For the matrix element to be non-zero, the difference between ℓ and ℓ' must now be even. To be consistent with $|\ell - \ell'| \leq 1$, therefore, the non-trivial matrix element involving a pseudo-vector must have

$$\ell = \ell'. \tag{5.6.388}$$

Projection theorem We now turn to using the lemma enveloping eq. (5.6.356) and the Wigner-Eckart theorem itself to deduce, the matrix element within a spin j subspace of the rank 1 spherical tensor V_1^n is related to its angular momentum counterpart J_1^n via the relation

$$\begin{aligned} \langle j, m; \Psi | V_1^n | j, m'; \Phi\rangle &= \frac{\left\langle j, m; \Psi \left| \vec{J} \cdot \vec{V} \right| j, m; \Phi \right\rangle}{j(j+1)} \langle j, m | J_1^n | j, m'\rangle \\ &= \frac{\left\langle j, m'; \Psi \left| \vec{J} \cdot \vec{V} \right| j, m'; \Phi \right\rangle}{j(j+1)} \langle j, m | J_1^n | j, m'\rangle , \end{aligned} \tag{5.6.389}$$

where $\vec{J} \cdot \vec{V}$ is the Cartesian dot product; and the $\langle j, m | J_1^n | j, m'\rangle$ is simply the matrix element of J_1^n in the angular momentum eigenstate basis, and does not involve Ψ nor Φ.

By the Wigner-Eckart theorem, we know that

$$\langle j, m; \Psi | V_1^n | j, m'; \Phi \rangle = \langle j\ m; 1\ j | 1\ n; j\ m' \rangle \langle j; \Psi || V_1 || j; \Phi \rangle, \tag{5.6.390}$$

$$\langle j, m | J_1^n | j, m' \rangle = \langle j\ m; 1\ j | 1\ n; j\ m' \rangle \langle j || J_1 || j \rangle. \tag{5.6.391}$$

Therefore, we may establish eq. (5.6.389) once we can show

$$\frac{\left\langle j, m; \Psi \left| \vec{J} \cdot \vec{V} \right| j, m; \Phi \right\rangle}{j(j+1)} = \frac{\left\langle j, m; \Psi \left| \vec{J} \cdot \vec{V} \right| j, m; \Phi \right\rangle}{\left\langle j, m \left| \vec{J}^2 \right| j, m \right\rangle} = \frac{\langle j; \Psi || V_1 || j; \Phi \rangle}{\langle j || J_1 || j \rangle}; \tag{5.6.392}$$

since eq. (5.6.389) will read

$$\langle j\ m; 1\ j | 1\ n; j\ m \rangle \langle j; \Psi || V_1 || j; \Phi \rangle = \frac{\langle j; \Psi || V_1 || j; \Phi \rangle}{\langle j || J_1 || j \rangle} \langle j\ m; 1\ j | 1\ n; j\ m \rangle \langle j || J_1 || j \rangle. \tag{5.6.393}$$

The key point is that, since the Cartesian versions of $\vec{J}$ and $\vec{V}$ are vector operators, both $\vec{J} \cdot \vec{V}$ and $\vec{J}^2$ are scalar operators and therefore commute with J^i itself. By the lemma surrounding eq. (5.6.356), we see that both the numerator and denominator after the first equality of eq. (5.6.392) are m independent. In particular, we may exploit the decomposition in eq. (5.6.341) (and eq. (5.6.336)),

$$\begin{aligned}
&\left\langle j, m; \Psi \left| \vec{J} \cdot \vec{V} \right| j, m; \Phi \right\rangle \\
&= \left\langle j, m; \Psi \left| J_1^0 V_1^0 + 2^{-1/2} J^+ V_1^{-1} - 2^{-1/2} J^- V_1^{+1} \right| j, m; \Phi \right\rangle &(5.6.394)\\
&= m \left\langle j, m; \Psi \left| V_1^0 \right| j, m; \Phi \right\rangle + \sqrt{(j+m)(j-m+1)/2} \left\langle j, m-1; \Psi \left| V_1^{-1} \right| j, m; \Phi \right\rangle \\
&\qquad - \sqrt{(j-m)(j+m+1)/2} \left\langle j, m+1; \Psi \left| V_1^{+1} \right| j, m; \Phi \right\rangle &(5.6.395)\\
&= \Big(m \langle j\ m; 1\ j | 1\ 0, j\ m \rangle + \sqrt{(j+m)(j-m+1)/2} \langle j\ m-1; 1\ j | 1\ -1, j\ m \rangle \\
&\qquad - \sqrt{(j-m)(j+m+1)/2} \langle j\ m+1; 1\ j | 1\ +1, j\ m \rangle \Big) \langle j; \Psi || V_1 || j; \Phi \rangle &(5.6.396)\\
&\equiv \chi_j \langle j; \Psi || V_1 || j; \Phi \rangle. &(5.6.397)
\end{aligned}$$

(This χ_j actually does not depend on m – can you explain why?) By replacing $\vec{V} \to \vec{J}$ we may immediately write down

$$\begin{aligned}
j(j+1) &= \left\langle j, m \left| \vec{J}^2 \right| j, m \right\rangle &(5.6.398)\\
&= \Big(m \langle j\ m; 1\ j | 1\ 0, j\ m \rangle + \sqrt{(j+m)(j-m+1)/2} \langle j\ m-1; 1\ j | 1\ -1, j\ m \rangle \\
&\qquad - \sqrt{(j-m)(j+m+1)/2} \langle j\ m+1; 1\ j | 1\ +1, j\ m \rangle \Big) \langle j || J_1 || j \rangle &(5.6.399)\\
&= \chi_j \langle j || J_1 || j \rangle. &(5.6.400)
\end{aligned}$$

Dividing equations (5.6.397) by (5.6.400) lead us to eq. (5.6.392). This proves the projection theorem.

5.7 *Rotations in 4 Spatial Dimensions

In this section, let us briefly examine 4D rotations – i.e., SO$_4$. Because the generators $\{J^{ab}\}$ in the general SO$_D$ algebra are anti-symmetric, $J^{ab} = -J^{ba}$, recall that means there are $(4^2-4)/2 = 6$ independent ones. More geometrically, in 4D, there are $\binom{4}{2} = 4!/(2^2) = 6$ independent 2D planes that may be rotated. When a and b of J^{ab} are both not equal to 4, the generators are simply the set of 3 generators $\{J^i = \epsilon^{ijk} J^k\}$ of the 3D case above. To avoid confusion, we will now use capital letters to denote an index that runs between 1 and 3; so, for e.g., we have

$$J^{\mathrm{I}} = \frac{1}{2}\epsilon^{\mathrm{IJK}} J^{\mathrm{JK}} \qquad \Leftrightarrow \qquad \epsilon^{\mathrm{IJK}} J^{\mathrm{I}} = J^{\mathrm{JK}}. \tag{5.7.1}$$

The remaining 3 generators of SO$_4$ are then $\{J^{\mathrm{I4}}\}$. Like the preceding 3D case, we need to compute the Lie Algebra of these angular momentum operators. We already know from eq. (5.6.24) that

$$\left[J^{\mathrm{A}}, J^{\mathrm{B}}\right] = i\epsilon^{\mathrm{ABC}} J^{\mathrm{C}}. \tag{5.7.2}$$

We therefore only need to figure out the commutation relations among the $\{J^{\mathrm{I4}}\}$ and between them and the $\{J^{\mathrm{I}}\}$. From eq. (5.5.73), we have

$$\left[J^{\mathrm{A4}}, J^{\mathrm{B4}}\right] = -i\left(\delta^{\mathrm{A[B}} J^{4]4} - \delta^{4[\mathrm{B}} J^{4]\mathrm{A}}\right). \tag{5.7.3}$$

Keeping in mind A, B $\neq 4$ and $J^{44} = 0$ because of anti-symmetry,

$$\left[J^{\mathrm{A4}}, J^{\mathrm{B4}}\right] = -iJ^{\mathrm{BA}} \tag{5.7.4}$$
$$\left[J^{\mathrm{A4}}, J^{\mathrm{B4}}\right] = i\epsilon^{\mathrm{ABC}} J^{\mathrm{C}}. \tag{5.7.5}$$

Next, we do

$$\left[J^{\mathrm{A4}}, J^{\mathrm{BC}}\right] = -i\left(\delta^{\mathrm{A[B}} J^{\mathrm{C]4}} - \delta^{4[\mathrm{B}} J^{\mathrm{C]}A}\right) \tag{5.7.6}$$
$$\left[J^{\mathrm{A4}}, J^{\mathrm{K}}\right] = -\frac{i}{2}\epsilon^{\mathrm{KBC}}\left(\delta^{\mathrm{A[B}} J^{\mathrm{C]4}} - \delta^{4[\mathrm{B}} J^{\mathrm{C]}A}\right). \tag{5.7.7}$$

This leads us to

$$\left[J^{\mathrm{A4}}, J^{\mathrm{B}}\right] = i\epsilon^{\mathrm{ABC}} J^{\mathrm{C4}}. \tag{5.7.8}$$

SO$_4$ Lie Algebra If A, B, C runs from 1 through 3 only, and if we remember the definition in eq. (5.7.1), the angular momentum operators in 4D obey the Lie Algebra in equations (5.7.2), (5.7.5) and (5.7.8).

Problem 5.101. so$_4$: Two copies of SO$_3$ Lie Algebra Define

$$M^{\mathrm{I}}_{\pm} \equiv \frac{J^{\mathrm{I}} \pm J^{\mathrm{I4}}}{2}. \tag{5.7.9}$$

Show that

$$\left[M^{\mathrm{I}}_{+}, M^{\mathrm{J}}_{-}\right] = 0 \qquad \text{and} \qquad \left[M^{\mathrm{I}}_{\pm}, M^{\mathrm{J}}_{\pm}\right] = i\epsilon^{\mathrm{IJK}} M^{\mathrm{K}}_{\pm}. \tag{5.7.10}$$

That is, the SO_4 Lie Algebra can be re-written into two independent copies of the SO_3 ones. Borrowing the 3D discussion, we may deduce that the eigenstates of the angular momentum operators in 4D may be described by two independent pairs of numbers $(\ell_\pm, m_\pm)$; with $\ell_\pm$ non-negative integer/half-integer,

$$\vec{M}_\pm^2 \left|\substack{\ell_+,m_+\\ \ell_-,m_-}\right\rangle = \ell_\pm(\ell_\pm+1)\left|\substack{\ell_+,m_+\\ \ell_-,m_-}\right\rangle, \qquad \vec{M}_\pm^2 \equiv M_\pm^{\mathrm{I}} M_\pm^{\mathrm{I}} \tag{5.7.11}$$

$$M_\pm^3 \left|\substack{\ell_+,m_+\\ \ell_-,m_-}\right\rangle = m_\pm \left|\substack{\ell_+,m_+\\ \ell_-,m_-}\right\rangle \tag{5.7.12}$$

and $m_\pm \in \{-\ell_\pm, -\ell_\pm+1, \dots, \ell_\pm-1, \ell_\pm\}$. $\square$

5.8 *Dilatations and the Mellin Transform

In this section, we shall study the group of dilatations, where the positive real line $\{r|r \geq 0\}$ is re-scaled by a non-negative number $\lambda \geq 0$. If $|r\rangle$ denotes the position eigenket, we define the dilatation operator as

$$D_s(\lambda)\,|r\rangle \equiv \lambda^{s+1}\,|\lambda\cdot r\rangle\,. \tag{5.8.1}$$

We may readily verify,

$$\begin{aligned} D_s(\lambda)D_s(\lambda')\,|r\rangle &= (\lambda\cdot\lambda')^{s+1}\,|\lambda\cdot\lambda'\cdot r\rangle \\ &= D_s(\lambda\cdot\lambda')\,|r\rangle\,; \end{aligned} \tag{5.8.2}$$

$$\begin{aligned} D_s(\lambda)D_s(\lambda')D_s(\lambda'')\,|r\rangle &= (\lambda\cdot\lambda'\cdot\lambda'')^{s+1}\,|\lambda\cdot\lambda'\cdot\lambda''\cdot r\rangle \\ &= (D_s(\lambda)D_s(\lambda'))D_s(\lambda'')\,|r\rangle \\ &= D_s(\lambda)(D_s(\lambda')D_s(\lambda''))\,|r\rangle\,; \end{aligned} \tag{5.8.3}$$

$$D_s(1)\,|r\rangle = |r\rangle = \mathbb{I}\,|r\rangle\,; \tag{5.8.4}$$

and

$$D_s(\lambda)^{-1} = D_s(\lambda^{-1}). \tag{5.8.5}$$

The inverse in eq. (5.8.5) may be readily checked, since

$$D_s(\lambda)D_s(\lambda^{-1})\,|r\rangle = (\lambda\cdot\lambda^{-1})^{s+1}\left|\lambda\cdot\lambda^{-1}\cdot r\right\rangle. \tag{5.8.6}$$

If we define the inner product between two arbitrary states $|f\rangle$ and $|g\rangle$ as

$$\langle f|\,g\rangle\,(s) \equiv \int_0^\infty \langle f|\,r\rangle\,\langle r|\,g\rangle\,r^{2s+1}\mathrm{d}r \tag{5.8.7}$$

$$= \int_0^\infty f(r)^* g(r) r^{2s+1}\mathrm{d}r, \tag{5.8.8}$$

the identity operator is then given by the following sum-over-eigenkets

$$\mathbb{I} = \int_0^\infty |r\rangle\,\langle r|\,r^{2s+1}\mathrm{d}r. \tag{5.8.9}$$

This allows us to, im turn, deduce the inner product between two position eigenkets. Since $\mathbb{I}\left|r'\right\rangle = \left|r'\right\rangle$, eq. (5.8.9) leads us to

$$\left|r'\right\rangle = \int_0^\infty \left|r\right\rangle \left\langle r\right|\left.r'\right\rangle r^{2s+1}\mathrm{d}r. \tag{5.8.10}$$

Because $\left|r'\right\rangle$ is arbitrary,

$$\left\langle r\right|\left.r'\right\rangle = \frac{\delta(r-r')}{(r\cdot r')^{s+\frac{1}{2}}}. \tag{5.8.11}$$

Caution Note that we are *not* assuming $\left|r\right\rangle$ resides in some flat (Euclidean) space, where the associated inner product in spherical or polar coordinates would have s here dependent on the number of spatial dimensions.

Translation Non-Invariance Additionally, notice eq. (5.8.7) is an example of an inner product that is *not* invariant under translations $r \to r+a$ (for constant a); due to both the lower limit 0 as well as the integration measure $r^{2s+1}\mathrm{d}r$. In other words: moving to the left of $r=0$ is illegal; and the integration measure r^{2s+1}, at least when $s>1/2$, gives a heavier weight for larger values of r.

Problem 5.102. $D_s(\lambda)$ Is Unitary Show that $D_s(\lambda)$ is unitary. □

Without loss of generality, if we express

$$\lambda = e^\epsilon \tag{5.8.12}$$

for $\epsilon \in \mathbb{R}$, we can check that eq. now becomes

$$D_s(e^\epsilon)D_s(e^{\epsilon'}) = D_s(e^{\epsilon+\epsilon'}); \tag{5.8.13}$$

i.e., the exponents of λ and λ' add under group multiplication. This and the unitary character of $D_s(\lambda)$ informs us,

$$D_s(e^\epsilon) = \exp(-i\epsilon\mathcal{E}_s) \tag{5.8.14}$$

for some Hermitian generator $\mathcal{E}_s$.

Problem 5.103. Dilatation Generator in Position $r-$Space If we write $\lambda = e^\epsilon$ and assume $|\epsilon| \ll 1$, $D_s(\lambda)$ will be very 'close' to the identity. Derive the following representation for the dilatation generator:

$$\left\langle r\left|\mathcal{E}_s\right|f\right\rangle = i\left(r\partial_r + s + 1\right)\left\langle r\right|\left.f\right\rangle. \tag{5.8.15}$$

Solve its eigen-equation $\mathcal{E}_s\left|\nu\right\rangle = \nu\left|\nu\right\rangle$ and show that up to a multiplicative phase factor,

$$\left\langle r\right|\left.\nu\right\rangle = r^{-i\nu-(s+1)}, \tag{5.8.16}$$

$$\left\langle \nu\right|\left.\nu'\right\rangle = 2\pi\delta(\nu-\nu'). \tag{5.8.17}$$

Hint: Start by considering $\left\langle r\left|D_s(e^\epsilon)\right|f\right\rangle = \left\langle r\left|e^{-i\epsilon\cdot\mathcal{E}_s}\right|f\right\rangle$. □

The eigenvalue ν in $\mathcal{E}_s |\nu\rangle = \nu |\nu\rangle$ runs over the entire real line. To see this, we note that the inner product of eq. (5.8.7) as well as the measure of the $\delta-$functions on the right hand side of eq. (5.8.17) strongly indicates

$$\mathbb{I} = \int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} |\nu\rangle \langle\nu| . \tag{5.8.18}$$

An arbitrary state $|f\rangle$ may therefore be expanded as

$$\begin{aligned} f(r \geq 0) \equiv \langle r| f\rangle &= \int_{-\infty}^{\infty} \langle r| \nu\rangle \langle \nu| f\rangle \frac{\mathrm{d}\nu}{2\pi} \\ &= r^{-(s+1)} \int_{\mathbb{R}} e^{-i\nu \ln r} \widetilde{f}(\nu) \frac{\mathrm{d}\nu}{2\pi}. \end{aligned} \tag{5.8.19}$$

To prove its validity, we simply compute its matrix element in the position basis:

$$\int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} \langle r| \nu\rangle \langle \nu| r'\rangle = \int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} r^{-i\nu-(s+1)} \cdot r'^{+i\nu-(s+1)} \tag{5.8.20}$$

$$= (r \cdot r')^{-(s+1)} \int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} e^{-i(\nu-\nu')\ln(t/r')} \tag{5.8.21}$$

$$= (r \cdot r')^{-(s+1)} \delta\left(\ln(r/r')\right) \tag{5.8.22}$$

$$= \frac{\delta\left(r - r'\right)}{(r \cdot r')^{s+\frac{1}{2}}}. \tag{5.8.23}$$

The identity operator in the position basis, i.e., eq. (5.8.9), tells us the state $|f\rangle$ may also be expressed as

$$\begin{aligned} \widetilde{f}(\nu \in \mathbb{R}) \equiv \langle \nu| f\rangle &= \int_0^{\infty} \langle \nu| r\rangle \langle r| f\rangle r^{2s+1} \mathrm{d}r \\ &= \int_0^{\infty} r^{s+i\nu} f(r) \mathrm{d}r. \end{aligned} \tag{5.8.24}$$

Mellin Transform For $s, \nu \in \mathbb{R}$, the Mellin transform of some function $f(r \geq 0)$ defined on the positive real line is

$$\widetilde{f}(z) \equiv \int_0^{\infty} r^{z-1} f(r) \mathrm{d}r, \tag{5.8.25}$$

$$z \equiv s + 1 + i\nu. \tag{5.8.26}$$

Comparing equations (5.8.19) and (5.8.25) demonstrates the intimate relationship between the Mellin transform and the dilatation group. In particular, the imaginary part of the exponent z in the Mellin transform is the eigenvalue of the generator of the dilatation operator. Recall the analogous situation for the Fourier transform, which we discovered while studying the translation group – for e.g., the momentum vector $\vec{k}$ occurring in the plane wave expansion $e^{i\vec{k}\cdot\vec{x}}$ was the eigenvalue of the generator of translations.

The absolute value of eq. (5.8.25) is bounded as

$$|\widetilde{f}(z)| \leq \int_0^{\infty} r^s |f(r)| \mathrm{d}r. \tag{5.8.27}$$

Hence, whenever the integral on the right hand side of eq. (5.8.27) convergences within some open interval $s+1 \in (a,b)$ – then so does the Mellin transform in eq. (5.8.25). For instance, if $f(r)$ is continuous and

$$f(r \to 0) \sim r^{-a+i\alpha}, \qquad f(r \to +\infty) \sim r^{-b+i\beta}; \tag{5.8.28}$$

then

$$\lim_{r \to 0} |f(r) r^{z-1}| \sim r^{\mathrm{Re}(z)-a-1}, \tag{5.8.29}$$

$$\lim_{r \to \infty} |f(r) r^{z-1}| \sim r^{\mathrm{Re}(z)-b-1}; \tag{5.8.30}$$

and therefore the Mellin transform in eq. (5.8.25) converges because the contributions from the lower and upper limits converge.

Moreover, the *Riemann-Lebesgue lemma* tells us

$$\lim_{\nu \to \infty} \widetilde{f}(z = s+1+i\nu) = 0. \tag{5.8.31}$$

As we shall explore further in §(6) below, these properties allow us to view the Mellin transform of $f(r)$ as an (infinitely) differentiable function of the complex variable z in the (open) vertical strip $a < \mathrm{Re}(z) < b$.

> **Mellin Transform Pair** Given an open interval $\mathrm{Re}(z) = s+1 \in (a,b)$ where the Mellin transform in eq. (5.8.25) is well defined; its inverse can be found in eq. (5.8.19).

A simple example of the Mellin transform is that of the radial integral in $D-$dimensional infinite flat space. The volume measure is $\mathrm{d}\Omega \mathrm{d}r \cdot r^{D-1}$, where $\mathrm{d}\Omega$ refers to the infinitesimal solid angle and r is the radial coordinate. The radial portion of a volume integral involving some function $f(r)$ (with all angular coordinates suppressed) would read

$$\widetilde{f}(z = D) = \int_0^\infty f(r) r^{D-1} \mathrm{d}r; \tag{5.8.32}$$

We shall explore further applications of the Mellin transform in §(6) below.

Problem 5.104. Inverse Mellin as Fourier Explain why the inverse Mellin transform in eq. (5.8.19) is really a Fourier decomposition.

Hint: Multiply both sides of eq. (5.8.25) with $e^{-i\nu \ln r'}$, and integrate over ν using the integral representation of the Dirac $\delta-$function in eq. (5.1.6). □

5.9 *Clebsch-Gordan Coefficients

In this section, we compile for the reader's reference, Clebsch-Gordan coefficients

$$\{\langle \ell_1 \; m_1, \ell_2 \; m_2 | \, j \; m; \ell_1 \ell_2 \rangle\} \tag{5.9.1}$$

for adding the spins ℓ_1 and ℓ_2, with m_1 and m_2 being their respectively azimuthal numbers.The j demotes the total angular momentum label, and m its corresponding azimuthal number. We use the notation

$$|\ell_1, m_1\rangle \otimes |\ell_2, m_2\rangle \equiv |\ell_1 \ m_1, \ell_2 \ m_2\rangle \, . \tag{5.9.2}$$

for the tensor product states; and

$$|j \ m; \ell_1 \ell_2\rangle \tag{5.9.3}$$

for the total angular momentum states arising from adding spin ℓ_1 and ℓ_2.

Remember the constraint

$$m_1 + m_2 = m \tag{5.9.4}$$

and admissible range of j:

$$j \in \{|\ell_1 - \ell_2|, |\ell_1 - \ell_2| + 1, |\ell_1 - \ell_2| + 2, \ldots, \ell_1 + \ell_2 - 2, \ell_1 + \ell_2 - 1, \ell_1 + \ell_2\} \, . \tag{5.9.5}$$

Adding $\left|\frac{1}{2}, m_1\right\rangle \otimes \left|\frac{1}{2}, m_2\right\rangle$ to yield spin 0

$$\left\langle \frac{1}{2} \ \frac{1}{2}, \frac{1}{2} \ -\frac{1}{2} \right| 0 \ 0; \frac{1}{2} \ \frac{1}{2} \Big\rangle = \frac{1}{\sqrt{2}}, \tag{5.9.6}$$

$$\left\langle \frac{1}{2} \ -\frac{1}{2}, \frac{1}{2} \ \frac{1}{2} \right| 0 \ 0; \frac{1}{2} \ \frac{1}{2} \Big\rangle = -\frac{1}{\sqrt{2}}. \tag{5.9.7}$$

Adding $\left|\frac{1}{2}, m_1\right\rangle \otimes \left|\frac{1}{2}, m_2\right\rangle$ to yield spin 1

$$\left\langle \frac{1}{2} \ \frac{1}{2}, \frac{1}{2} \ \frac{1}{2} \right| 1 \ 1; \frac{1}{2} \ \frac{1}{2} \Big\rangle = 1, \tag{5.9.8}$$

$$\left\langle \frac{1}{2} \ \frac{1}{2}, \frac{1}{2} \ -\frac{1}{2} \right| 1 \ 0; \frac{1}{2} \ \frac{1}{2} \Big\rangle = \frac{1}{\sqrt{2}}, \tag{5.9.9}$$

$$\left\langle \frac{1}{2} \ -\frac{1}{2}, \frac{1}{2} \ \frac{1}{2} \right| 1 \ 0; \frac{1}{2} \ \frac{1}{2} \Big\rangle = \frac{1}{\sqrt{2}}, \tag{5.9.10}$$

$$\left\langle \frac{1}{2} \ -\frac{1}{2}, \frac{1}{2} \ -\frac{1}{2} \right| 1 \ -1; \frac{1}{2} \ \frac{1}{2} \Big\rangle = 1. \tag{5.9.11}$$

Adding $\left|\frac{1}{2}, m_1\right\rangle \otimes |1, m_2\rangle$ or $|1, m_1\rangle \otimes \left|\frac{1}{2}, m_2\right\rangle$ to yield spin 1/2

$$\left\langle \frac{1}{2} \ -\frac{1}{2}, 1 \ 1 \right| \frac{1}{2} \ \frac{1}{2}; \frac{1}{2} \ 1 \Big\rangle = -\sqrt{\frac{2}{3}}, \tag{5.9.12}$$

$$\left\langle \frac{1}{2} \ \frac{1}{2}, 1 \ 0 \right| \frac{1}{2} \ \frac{1}{2}; \frac{1}{2} \ 1 \Big\rangle = \sqrt{\frac{1}{3}}, \tag{5.9.13}$$

$$\left\langle \frac{1}{2} \ -\frac{1}{2}, 1 \ 0 \right| \frac{1}{2} \ \frac{1}{2}; -\frac{1}{2} \ 1 \Big\rangle = -\sqrt{\frac{1}{3}}, \tag{5.9.14}$$

$$\left\langle \frac{1}{2}\;\frac{1}{2}, 1\;\;-1 \middle| \frac{1}{2}\;\frac{1}{2}; \frac{1}{2}\;\;1 \right\rangle = \sqrt{\frac{2}{3}}; \tag{5.9.15}$$

and

$$\left\langle 1\;1, \frac{1}{2}\;\;-\frac{1}{2} \middle| \frac{1}{2}\;\frac{1}{2}; \frac{1}{2}\;\;1 \right\rangle = \sqrt{\frac{2}{3}}, \tag{5.9.16}$$

$$\left\langle 1\;0, \frac{1}{2}\;\frac{1}{2} \middle| \frac{1}{2}\;\frac{1}{2}; \frac{1}{2}\;\;1 \right\rangle = -\sqrt{\frac{1}{3}}, \tag{5.9.17}$$

$$\left\langle 1\;0, \frac{1}{2}\;\;-\frac{1}{2} \middle| \frac{1}{2}\;\frac{1}{2}; -\frac{1}{2}\;\;1 \right\rangle = \sqrt{\frac{1}{3}}, \tag{5.9.18}$$

$$\left\langle 1\;\;-1, \frac{1}{2}\;\frac{1}{2} \middle| \frac{1}{2}\;\frac{1}{2}; \frac{1}{2}\;\;1 \right\rangle = -\sqrt{\frac{2}{3}}. \tag{5.9.19}$$

Adding $\left|\frac{1}{2}, m_1\right\rangle \otimes |1, m_2\rangle$ or $|1, m_1\rangle \otimes \left|\frac{1}{2}, m_2\right\rangle$ to yield spin 3/2 YZ: This section is woefully incomplete.

5.10 ⋆Approximation Methods for Eigensystems

5.10.1 Rayleigh-Schrödinger Perturbation Theory

Suppose we know how to diagonalize some Hermitian operator H_0 exactly.

$$H_0 \left|\bar{E}\right\rangle = \bar{E} \left|\bar{E}\right\rangle \tag{5.10.1}$$

In this section[39] we will address how to diagonalize a H, namely

$$H \left|E\right\rangle = E \left|E\right\rangle; \tag{5.10.2}$$

in the situation where it is a small perturbation of the H_0, in the following sense:

$$H \equiv H_0 + \epsilon H_1 + \epsilon^2 H_2 + \mathcal{O}\left(\epsilon^3\right) \tag{5.10.3}$$

$$= H_0 + \sum_{\ell=1}^{+\infty} \epsilon^\ell H_\ell. \tag{5.10.4}$$

Here, $0 < \epsilon \ll 1$ is oftentimes fictitious parameter indicating the 'smallness' of the H_is; so the ϵ in ϵH_1 reminds us H_1 is to be considered first order in perturbation, $\epsilon^2 H_2$ second order, etc. and the $\{\delta_\ell H | \ell = 1, 2, 3, \dots\}$ are assumed to be Hermitian operators. Such a perturbed operator H appears in many physical situations, such as atomic physics – where the H_0 is the Hamiltonian of the nucleus-electron(s) atomic system itself; and the $\delta_\ell H$ are perturbations induced, say, spin-orbit interactions; relativistic corrections; or by immersing the atom in an electric (Stark effect) and/or magnetic field (Zeeman effect); etc. In physical problems, the 'smallness' parameter ϵ of an operator may often be identified with ratios of important dimensionful quantities of

[39] The discussion here is inspired by §4.11 of Byron and Fuller [14], Sakurai and Napolitano [11], and Weinberg [12] Chapter 5.

the setup at hand. Moreover, we have implicitly assumed a single parameter ϵ here; while in physical problems where there are more than 2 dimension-ful quantities there will generically be multiple independent ϵs. In such a scenario the perturbation theory delineated here will have to be extended appropriately.

Non-degenerate Case The solution strategy is to postulate that the eigensystems of H are themselves a power series in ϵ, where the zeroth order (i.e., ϵ-independent) piece is simply a given, exact, eigensystem of H_0. The corrections to the eigenstates induced by the $H_{\ell\geq 1}$s will in turn be expressed in terms of the unperturbed eigensystems of H_0. Specifically, let $\left|\bar{E}_a\right\rangle$ be the ath eigenstate of H_0 and $\left|E_a\right\rangle$ be that of the full H. For now, we shall assume that $\bar{E}_a$ is a non-degenerate eigenvalue; and the eigenstates are orthonormal,

$$\left\langle\bar{E}_b\middle|\,\bar{E}_a\right\rangle = \delta^b_a, \tag{5.10.5}$$

$$\sum_a \left|\bar{E}_a\right\rangle\left\langle\bar{E}_a\right| = \mathbb{I}. \tag{5.10.6}$$

[40]Then, we assert

$$\begin{aligned}\left|E_a\right\rangle &= \left|\bar{E}_a\right\rangle + \sum_{\ell=1}^{+\infty}\epsilon^\ell\left|{}_\ell E_a\right\rangle &(5.10.7)\\ &= \left|\bar{E}_a\right\rangle + \sum_{\ell=1}^{+\infty}\sum_s \epsilon^\ell\left|\bar{E}_s\right\rangle\left\langle\bar{E}_s\middle|\,{}_\ell E_a\right\rangle. &(5.10.8)\end{aligned}$$

where $\left|{}_\ell E_a\right\rangle$ is the ℓth correction to the ath energy eigenstate. The ath energy eigenvalue is, itself, a power series

$$E_a = \bar{E}_a + \sum_{\ell=1}^{+\infty}\epsilon^\ell\delta_\ell E_a. \tag{5.10.9}$$

The goal is therefore to compute the perturbations of the eigenstate along the unperturbed ones $\{\left\langle\bar{E}_s\middle|\,{}_\ell E_a\right\rangle\}$ and of the eigenvalues $\{\delta_\ell E_a\}$ in terms of the unperturbed ones $\{\bar{E}_a\}$.

Now, the eigenvalue problem is given by $H\left|E_a\right\rangle = E_a\left|E_a\right\rangle$. Expanding in powers of ϵ,

$$\begin{aligned}&\left(H_0+\epsilon H_1+\epsilon^2 H_2+\dots\right)\left(\left|\bar{E}_a\right\rangle+\epsilon\left|{}_1E_a\right\rangle+\epsilon^2\left|{}_2E_a\right\rangle+\dots\right)\\ &\qquad=\left(\bar{E}_a+\epsilon\delta_1E_a+\epsilon^2\delta_2E_a+\dots\right)\left(\left|\bar{E}_a\right\rangle+\epsilon\left|{}_1E_a\right\rangle+\epsilon^2\left|{}_2E_a\right\rangle+\dots\right);\end{aligned} \tag{5.10.10}$$

we may collect powers of ϵ in the following manner:

$$\begin{array}{llll} H_0\left|\bar{E}_a\right\rangle+ & \epsilon H_0\left|{}_1E_a\right\rangle+ & \epsilon^2 H_0\left|{}_2E_a\right\rangle+ & \dots \quad (5.10.11)\\ + & \epsilon H_1\left|\bar{E}_a\right\rangle+ & \epsilon^2 H_1\left|{}_1E_a\right\rangle+ & \dots \quad (5.10.12)\\ & + & \epsilon^2 H_2\left|\bar{E}_a\right\rangle+ & \dots \quad (5.10.13)\end{array}$$

[40]The completeness relation of eq. (5.10.6) involves the sum over all states – both degenerate and non-degenerate ones. While we are assuming $\bar{E}_a$ is non-degenerate for now; the other eigenstates $\{\left|\bar{E}_{b\neq a}\right\rangle\}$ are allowed to be degenerate. Strictly speaking, in such a situation we ought to introduce a degeneracy label, for e.g., $\left|\bar{E}_b;i\right\rangle$, but prefer not to do so to avoid notation overload.

$$
\begin{aligned}
= \bar{E}_a \left|\bar{E}_a\right\rangle + \quad & \epsilon \bar{E}_a \left|{}_1 E_a\right\rangle + \quad & \epsilon^2 \bar{E}_a \left|{}_2 E_a\right\rangle + \quad & \dots && (5.10.14)\\
+ \quad & \epsilon \delta_1 E_a \left|\bar{E}_a\right\rangle + \quad & \epsilon^2 \delta_1 E_a \left|{}_1 E_a\right\rangle + \quad & \dots && (5.10.15)\\
& + \quad & \epsilon^2 \delta_2 E_a \left|\bar{E}_a\right\rangle + \quad & \dots . && (5.10.16)
\end{aligned}
$$

The $\mathcal{O}(\epsilon^0)$ terms on both sides cancel out because they simply amount to eq. (5.10.1). The $\mathcal{O}(\epsilon, \epsilon^2)$ terms are

$$
\left(H_0 - \bar{E}_a\right) \left|{}_1 E_a\right\rangle = -\left(H_1 - \delta_1 E_a\right) \left|\bar{E}_a\right\rangle, \tag{5.10.17}
$$

$$
\left(H_0 - \bar{E}_a\right) \left|{}_2 E_a\right\rangle = -\left(H_1 - \delta_1 E_a\right) \left|{}_1 E_a\right\rangle - \left(H_2 - \delta_2 E_a\right) \left|\bar{E}_a\right\rangle. \tag{5.10.18}
$$

More generally, at the $\mathcal{O}(\epsilon^{\ell \geq 1})$ level,

$$
\begin{aligned}
&\left(H_0 - \bar{E}_a\right) \left|{}_\ell E_a\right\rangle \\
&= -\left(H_1 - \delta_1 E_a\right) \left|{}_{\ell-1} E_a\right\rangle - \left(H_2 - \delta_2 E_a\right) \left|{}_{\ell-2} E_a\right\rangle \\
&\qquad \cdots - \left(H_{\ell-2} - \delta_{\ell-2} E_a\right) \left|{}_2 E_a\right\rangle - \left(H_{\ell-1} - \delta_{\ell-1} E_a\right) \left|{}_1 E_a\right\rangle - \left(H_\ell - \delta_\ell E_a\right) \left|\bar{E}_a\right\rangle \qquad (5.10.19)\\
&= -\sum_{s=1}^{\ell-1} \left(H_s - \delta_s E_a\right) \left|{}_{\ell-s} E_a\right\rangle - \left(H_\ell - \delta_\ell E_a\right) \left|\bar{E}_a\right\rangle.
\end{aligned}
$$

Due to the hermitian character of H_0, eq. (5.10.1) may be expressed as

$$
\left\langle \bar{E}_a \right| \left(H_0 - \bar{E}_a\right) = 0. \tag{5.10.20}
$$

Therefore, $\left\langle \bar{E}_a \right|$ acting on both sides of equations (5.10.17), (5.10.18), (5.10.19), etc. would yield zero on their left hand sides and in turn lead to

$$
0 = -\left\langle \bar{E}_a \right| H_1 \left|\bar{E}_a\right\rangle + \delta_1 E_a, \tag{5.10.21}
$$

$$
0 = -\left\langle \bar{E}_a \right| H_1 - \delta_1 E_a \left|{}_1 E_a\right\rangle - \left\langle \bar{E}_a \right| H_2 \left|\bar{E}_a\right\rangle + \delta_2 E_a, \tag{5.10.22}
$$

$\dots\dots$

$$
\begin{aligned}
0 = &-\left\langle \bar{E}_a \right| H_1 - \delta_1 E_a \left|{}_{\ell-1} E_a\right\rangle - \left\langle \bar{E}_a \right| H_2 - \delta_2 E_a \left|{}_{\ell-2} E_a\right\rangle \qquad (5.10.23)\\
&\cdots - \left\langle \bar{E}_a \right| H_{\ell-2} - \delta_{\ell-2} E_a \left|{}_2 E_a\right\rangle - \left\langle \bar{E}_a \right| H_{\ell-1} - \delta_{\ell-1} E_a \left|{}_1 E_a\right\rangle - \left\langle \bar{E}_a \right| H_\ell \left|\bar{E}_a\right\rangle + \delta_\ell E_a.
\end{aligned}
$$

At this juncture, let us observe it is always possible to render

$$
\left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle = 0 \tag{5.10.24}
$$

for all $\ell \geq 1$ simply by choosing to normalize our eigenstates as

$$
\left\langle \bar{E}_a \middle| E_a \right\rangle = 1. \tag{5.10.25}
$$

To this end, let us first recall that an eigenvector $|\lambda\rangle$ is only defined up to an overall multiplicative complex amplitude χ; i.e., if $A|\lambda\rangle = \lambda|\lambda\rangle$ so does $A(\chi|\lambda\rangle) = \lambda(\chi|\lambda\rangle)$. Therefore, since χ multiplies every coefficient when we expand $\chi|\lambda\rangle$ as a superposition over basis vectors $\{|j\rangle\}$, as long as the overlap between $|\lambda\rangle = \sum_j |j\rangle \langle j|\lambda\rangle$ and a given basis vector $|i\rangle$ is non-zero; we may choose to normalize $|\lambda\rangle$ by specifying $\langle i|\lambda\rangle$ – since, under re-scaling $|\lambda\rangle \to \chi|\lambda\rangle$, $\langle i|\lambda\rangle \to \chi\langle i|\lambda\rangle$.

This is precisely the case in eq. (5.10.25), where we know $\left|\bar{E}_a\right\rangle$ must have significant overlap with the exact eigenstate $|E_a\rangle$. Expanding eq. (5.10.25),

$$\left\langle \bar{E}_a \middle| \bar{E}_a \right\rangle + \sum_{\ell=1}^{+\infty} \epsilon^\ell \left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle = 1 \tag{5.10.26}$$

$$\sum_{\ell=1}^{+\infty} \epsilon^\ell \left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle = 0; \tag{5.10.27}$$

followed by setting to zero the coefficient of each $\epsilon^{\ell \geq 1}$, we arrive at eq. (5.10.24).

Additionally, starting with $|{}_1E_a\rangle$, notice eq. (5.10.17) is invariant under the replacement $|{}_1E_a\rangle \to |{}_1E_a\rangle + \chi_1 \left|\bar{E}_a\right\rangle$ – for arbitrary complex number χ_1 – because of the eigen-equation (5.10.1). In other words, if we found a solution $|{}_1E_a\rangle = |\psi_1\rangle$; then so is $|{}_1E_a\rangle = |\psi_1\rangle + \chi_1 \left|\bar{E}_a\right\rangle$. Hence, if $\left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle = \left\langle \bar{E}_a \middle| \psi_1 \right\rangle \neq 0$, we may simply choose χ_1 such that the new solution $|{}_1E_a\rangle_{\text{new}} \equiv |\psi_1\rangle + \chi \left|\bar{E}_a\right\rangle$ satisfies $\left\langle \bar{E}_a \middle| {}_1E_a \right\rangle_{\text{new}} = \left\langle \bar{E}_a \middle| \psi_1 \right\rangle + \chi = 0$. Now, suppose we have solved $|{}_iE_a\rangle$ from $i = 1$ up to $i = \ell - 1$. Then we see, just like the $\ell = 1$ case, both $|{}_\ell E_a\rangle = |\psi_\ell\rangle$ and $|{}_\ell E_a\rangle = |\psi_\ell\rangle + \chi_\ell \left|\bar{E}_a\right\rangle$ solve eq. (5.10.19) as long as $|\psi_\ell\rangle$ is a solution. Therefore if $\left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle$ were not zero, then the 'new' solution $|{}_\ell E_a\rangle_{\text{new}} \equiv |\psi_\ell\rangle + \chi_\ell \left|\bar{E}_a\right\rangle$ can be made to satify $0 = \left\langle \bar{E}_a \middle| {}_\ell E_a \right\rangle_{\text{new}} = \left\langle \bar{E}_a \middle| \psi_\ell \right\rangle + \chi$ simply by choosing $\chi = -\left\langle \bar{E}_a \middle| \psi_\ell \right\rangle$.

The freedom to shift the perturbation by a constant multiple of $\left|\bar{E}_a\right\rangle$ at each step of the construction is related to the freedom to re-scale the eigenket $|E_a\rangle$ itself. For, at the ℓth step, when we perform $|{}_\ell E_a\rangle \to |{}_\ell E_a\rangle + \chi_\ell \left|\bar{E}_a\right\rangle$, the normalization condition in eq. (5.10.25) is altered into

$$\begin{aligned} \left\langle \bar{E}_a \middle| E_a \right\rangle &= \left\langle \bar{E}_a \right| \left(\left|\bar{E}_a\right\rangle + \sum_{k=1}^{\ell-1} \epsilon^k |{}_kE_a\rangle + \epsilon^\ell (|{}_\ell E_a\rangle + \chi_\ell \left|\bar{E}_a\right\rangle) + \mathcal{O}(\epsilon^{\ell+1}) \right) \\ &= 1 + \epsilon^\ell \chi_\ell + \mathcal{O}(\epsilon^{\ell+1}) \end{aligned} \tag{5.10.28}$$

As you will witness in Problem (5.106) below, at least up to the second order in perturbations, this freedom to shift the solution may in fact be used to construct a unit norm eigenket.

Returning to the equations (5.10.21)–(5.10.23), we may thus employ eq. (5.10.24) to gather:

$$0 = -\left\langle \bar{E}_a \right| H_1 \left|\bar{E}_a\right\rangle + \delta_1 E_a, \tag{5.10.29}$$

$$0 = -\left\langle \bar{E}_a \right| H_1 |{}_1E_a\rangle - \left\langle \bar{E}_a \right| H_2 \left|\bar{E}_a\right\rangle + \delta_2 E_a, \tag{5.10.30}$$

$$\dots\dots$$

$$\begin{aligned} 0 = &-\left\langle \bar{E}_a \right| H_1 |{}_{\ell-1}E_a\rangle - \left\langle \bar{E}_a \right| H_2 |{}_{\ell-2}E_a\rangle \\ &\dots - \left\langle \bar{E}_a \right| H_{\ell-2} |{}_2E_a\rangle - \left\langle \bar{E}_a \right| H_{\ell-1} |{}_1E_a\rangle - \left\langle \bar{E}_a \right| H_\ell \left|\bar{E}_a\right\rangle + \delta_\ell E_a. \end{aligned} \tag{5.10.31}$$

The eigensystem of eq. (5.10.1) also tells us, for $b \neq a$,

$$\left\langle \bar{E}_b \right| \left(H_0 - \bar{E}_a\right) = (\bar{E}_b - \bar{E}_a) \left\langle \bar{E}_b \right|. \tag{5.10.32}$$

Therefore, $\left\langle \bar{E}_{b\neq a} \right|$ acting on both sides of equations (5.10.17), (5.10.18), (5.10.19), etc. would produce

$$\left(\bar{E}_b - \bar{E}_a\right) \left\langle \bar{E}_b \middle| {}_1E_a \right\rangle = -\left\langle \bar{E}_b \right| H_1 \left|\bar{E}_a\right\rangle, \tag{5.10.33}$$

$$\left(\bar{E}_b - \bar{E}_a\right) \left\langle \bar{E}_b \right| {}_2E_a \big\rangle = - \left\langle \bar{E}_b \right| H_1 - \delta_1 E_a \left| {}_1E_a \right\rangle - \left\langle \bar{E}_b \right| H_2 \left| \bar{E}_a \right\rangle, \tag{5.10.34}$$

$$\cdots\cdots$$

$$\begin{aligned}\left(\bar{E}_b - \bar{E}_a\right) \left\langle \bar{E}_b \right| {}_\ell E_a \big\rangle = &- \left\langle \bar{E}_b \right| H_1 - \delta_1 E_a \left| {}_{\ell-1}E_a \right\rangle - \left\langle \bar{E}_b \right| H_2 - \delta_2 E_a \left| {}_{\ell-2}E_a \right\rangle \\ &\cdots - \left\langle \bar{E}_b \right| H_{\ell-2} - \delta_{\ell-2} E_a \left| {}_2E_a \right\rangle - \left\langle \bar{E}_b \right| H_{\ell-1} - \delta_{\ell-1} E_a \left| {}_1E_a \right\rangle \\ &- \left\langle \bar{E}_b \right| H_\ell \left| \bar{E}_a \right\rangle.\end{aligned} \tag{5.10.35}$$

From eq. (5.10.29), we see that the first order correction to the energy is the expectation value of the first correction to the Hamiltonian:

$$\delta_1 E_a = \left\langle \bar{E}_a \right| H_1 \left| \bar{E}_a \right\rangle. \tag{5.10.36}$$

The off-diagonal term in eq. (5.10.33) allows us to infer, the first-order correction to the ath eigenstate along the $\left|\bar{E}_{b\neq a}\right\rangle$-direction is

$$\left\langle \bar{E}_{b\neq a} \right| {}_1E_a \big\rangle = \frac{\left\langle \bar{E}_b \left| H_1 \right| \bar{E}_a \right\rangle}{\bar{E}_a - \bar{E}_b}. \tag{5.10.37}$$

Turning to the second order corrections, the completeness relation $\mathbb{I} = \sum_c \left|\bar{E}_c\right\rangle \left\langle \bar{E}_c\right|$, together with equations (5.10.5), (5.10.24), (5.10.36), and (5.10.37), allow us to massage equations (5.10.30) and (5.10.34):

$$0 = -\sum_b \left\langle \bar{E}_a \right| H_1 \left| \bar{E}_b \right\rangle \left\langle \bar{E}_b \right| {}_1E_a \big\rangle - \left\langle \bar{E}_a \right| H_2 \left| \bar{E}_a \right\rangle + \delta_2 E_a \tag{5.10.38}$$

$$\delta_2 E_a = \sum_{b\neq a} \left\langle \bar{E}_a \right| H_1 \left| \bar{E}_b \right\rangle \left\langle \bar{E}_b \right| {}_1E_a \big\rangle + \left\langle \bar{E}_a \right| H_2 \left| \bar{E}_a \right\rangle \tag{5.10.39}$$

$$= \sum_{b\neq a} \left\langle \bar{E}_a \right| H_1 \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \left| H_1 \right| \bar{E}_a \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a \right| H_2 \left| \bar{E}_a \right\rangle. \tag{5.10.40}$$

$$\begin{aligned}\left(\bar{E}_b - \bar{E}_a\right) \left\langle \bar{E}_b \right| {}_2E_a \big\rangle &= -\sum_c \left\langle \bar{E}_b \right| H_1 - \delta_1 E_a \left| \bar{E}_c \right\rangle \left\langle \bar{E}_c \right| {}_1E_a \big\rangle - \left\langle \bar{E}_b \right| H_2 \left| \bar{E}_a \right\rangle \\ &= -\sum_{c\neq a} \left\langle \bar{E}_b \right| H_1 - \left\langle \bar{E}_a \left| H_1 \right| \bar{E}_a \right\rangle \left| \bar{E}_c \right\rangle \frac{\left\langle \bar{E}_c \left| H_1 \right| \bar{E}_a \right\rangle}{\bar{E}_a - \bar{E}_c} - \left\langle \bar{E}_b \right| H_2 \left| \bar{E}_a \right\rangle\end{aligned} \tag{5.10.41}$$

$$\begin{aligned}\left\langle \bar{E}_b \right| {}_2E_a \big\rangle = \left(\bar{E}_a - \bar{E}_b\right)^{-1} \Bigg(&\sum_{c\neq a} \left\langle \bar{E}_b \left| H_1 \right| \bar{E}_c \right\rangle \frac{\left\langle \bar{E}_c \left| H_1 \right| \bar{E}_a \right\rangle}{\bar{E}_a - \bar{E}_c} \\ &- \left\langle \bar{E}_a \left| H_1 \right| \bar{E}_a \right\rangle \frac{\left\langle \bar{E}_b \left| H_1 \right| \bar{E}_a \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_b \right| H_2 \left| \bar{E}_a \right\rangle \Bigg).\end{aligned} \tag{5.10.42}$$

We now have the necessary ingredients to construct both the eigenvalue and eigenstate perturbations $\left| {}_\ell E_a \right\rangle = \sum_b \left|\bar{E}_b\right\rangle \left\langle \bar{E}_b \right| {}_\ell E_a \big\rangle$ up to second order. Collecting the first order results from equations (5.10.36) and (5.10.37); the second order ones from equations (5.10.40) and (5.10.42); and remembering we have chosen to satisfy the constraint in eq. (5.10.24) that eigenket perturbations are orthogonal to $\left|\bar{E}_a\right\rangle$:

Non-Degenerate PT The eigensystem of

$$H = H_0 + \epsilon H_1 + \epsilon^2 H_2 + \mathcal{O}(\epsilon^3), \tag{5.10.43}$$

expressed in terms of the unperturbed ones – i.e., $\{|\bar{E}_a\rangle\}$ obeying $H_0 |\bar{E}_a\rangle = \bar{E}_a |\bar{E}_a\rangle$ – are given by

$$H |E_a\rangle = E_a |E_a\rangle, \tag{5.10.44}$$

$$E_a = \bar{E}_a + \epsilon \langle \bar{E}_a |H_1| \bar{E}_a\rangle + \epsilon^2 \left(\sum_{b\neq a} \frac{|\langle \bar{E}_b |H_1| \bar{E}_a\rangle|^2}{\bar{E}_a - \bar{E}_b} + \langle \bar{E}_a| H_2 |\bar{E}_a\rangle \right) + \mathcal{O}(\epsilon^3) \tag{5.10.45}$$

$$|E_a\rangle = |\bar{E}_a\rangle + \epsilon \sum_{b\neq a} |\bar{E}_b\rangle \frac{\langle \bar{E}_b |H_1| \bar{E}_a\rangle}{\bar{E}_a - \bar{E}_b} + \epsilon^2 \sum_{b\neq a} \frac{|\bar{E}_b\rangle}{\bar{E}_a - \bar{E}_b} \left(\sum_{c\neq a} \langle \bar{E}_b |H_1| \bar{E}_c\rangle \frac{\langle \bar{E}_c |H_1| \bar{E}_a\rangle}{\bar{E}_a - \bar{E}_c} - \langle \bar{E}_a |H_1| \bar{E}_a\rangle \frac{\langle \bar{E}_b |H_1| \bar{E}_a\rangle}{\bar{E}_a - \bar{E}_b} + \langle \bar{E}_b| H_2 |\bar{E}_a\rangle \right) + \mathcal{O}(\epsilon^3).$$

Note that, for $\bar{E}_b \neq \bar{E}_a$, the eigenvalue $\bar{E}_b$ can of course be degenerate. The sums in the above formula must include all the orthonormal basis states within such a degenerate subspace.

Problem 5.105. 'Inverse' of $H_0 - \bar{E}_a$ The operator $H_0 - \bar{E}_a$ on the left hand sides of equations (5.10.17), (5.10.18) and (5.10.19), etc. has no inverse because it has a null eigenvector $|\bar{E}_a\rangle$; i.e., $(H_0 - \bar{E}_a)|\bar{E}_a\rangle = 0$. However, if we restrict our attention to the portion of the vector space perpendicular to $|\bar{E}_a\rangle$, then we may write down a pseudo-inverse of sorts:

$$(H_0 - \bar{E}_a)_\perp^{-1} \equiv \sum_{b\neq a} \frac{|\bar{E}_b\rangle \langle \bar{E}_b|}{\bar{E}_b - \bar{E}_a}. \tag{5.10.46}$$

Verify that

$$(H_0 - \bar{E}_a)_\perp^{-1}(H_0 - \bar{E}_a) = (H_0 - \bar{E}_a)(H_0 - \bar{E}_a)_\perp^{-1} \tag{5.10.47}$$

$$= \sum_{b\neq a} |\bar{E}_b\rangle \langle \bar{E}_b| = \mathbb{I} - |\bar{E}_a\rangle \langle \bar{E}_a|. \tag{5.10.48}$$

Let $|\psi\rangle$ be orthogonal to $|\bar{E}_a\rangle$ but otherwise arbitrary. Explain why

$$(H_0 - \bar{E}_a)_\perp^{-1}(H_0 - \bar{E}_a)|\psi\rangle = (H_0 - \bar{E}_a)(H_0 - \bar{E}_a)_\perp^{-1}|\psi\rangle = |\psi\rangle. \tag{5.10.49}$$

Now explain how we may solve for $\{|{}_iE_a\rangle | i = 1, 2, 3, \dots\}$ from (5.10.17), (5.10.18) and (5.10.19), etc. through the pseudo-inverse $(H_0 - \bar{E}_a)_\perp^{-1}$. □

Problem 5.106. Unit Norm Eigenket Above, we have argued that, if we had already solved $|{}_iE_a\rangle$ from $i = 1$ up to $i = \ell - 1$, then if $|{}_\ell E_a\rangle$ solves eq. (5.10.19) – so does $|{}_\ell E_a\rangle + \chi_\ell |\bar{E}_a\rangle$. For example, if $|{}_1E_a\rangle$ is given by equations (5.10.24) and (5.10.37); then both $|\bar{E}_a\rangle + \epsilon |{}_1E_a\rangle +$

$\epsilon^2 \left| {}_2E_a \right\rangle$ and $(1+\epsilon^2\chi_2)\left|\bar{E}_a\right\rangle + \epsilon\left|{}_1E_a\right\rangle + \epsilon^2\left|{}_2E_a\right\rangle$ are eigenkets of $H = H_0 + \epsilon H_1 + \epsilon^2 H_2$ up to quadratic order in ϵ; as long as $\left|{}_2E_a\right\rangle$ solves eq. (5.10.18).

Demonstrate that we may normalize eq. (5.10.46) to unity, up to $\mathcal{O}(\epsilon^2)$, by shifting the second order correction by

$$\left|{}_2E_a\right\rangle \to \left|{}_2E_a\right\rangle - \frac{\left|\bar{E}_a\right\rangle}{2}\sum_{c\neq a}\left|\frac{\left\langle \bar{E}_c\left|H_1\right|\bar{E}_a\right\rangle}{\bar{E}_a - \bar{E}_c}\right|^2. \tag{5.10.50}$$

Hence, up to second order in perturbation theory,

$$\begin{aligned}\left|E_a\right\rangle = \left|\bar{E}_a\right\rangle + \epsilon\sum_{b\neq a}\left|\bar{E}_b\right\rangle\frac{\left\langle\bar{E}_b\left|H_1\right|\bar{E}_a\right\rangle}{\bar{E}_a-\bar{E}_b} + \epsilon^2\Bigg\{\sum_{b\neq a}\frac{\left|\bar{E}_b\right\rangle}{\bar{E}_a-\bar{E}_b}\Bigg(\sum_{c\neq a}\left\langle\bar{E}_b\left|H_1\right|\bar{E}_c\right\rangle\frac{\left\langle\bar{E}_c\left|H_1\right|\bar{E}_a\right\rangle}{\bar{E}_a-\bar{E}_c} \\ - \left\langle\bar{E}_a\left|H_1\right|\bar{E}_a\right\rangle\frac{\left\langle\bar{E}_b\left|H_1\right|\bar{E}_a\right\rangle}{\bar{E}_a-\bar{E}_b} + \left\langle\bar{E}_b\right|H_2\left|\bar{E}_a\right\rangle\Bigg) - \frac{\left|\bar{E}_a\right\rangle}{2}\sum_{c\neq a}\left|\frac{\left\langle\bar{E}_c\left|H_1\right|\bar{E}_a\right\rangle}{\bar{E}_a-\bar{E}_c}\right|^2\Bigg\}\end{aligned} \tag{5.10.51}$$

is not only an eigenket of $H = H_0 + \epsilon H_1 + \epsilon^2 H_2$ it is also unit norm. □

Degenerate Case If the eigenvalue $\bar{E}_a$ in eq. (5.10.1) (and (5.10.20)) is degenerate, the preceding discussion goes through, up to equations (5.10.17), (5.10.18), and (5.10.19); but we now need to add an enumeration label j to the eigenstate – namely, $\left|\bar{E}_a; j\right\rangle$ – that runs from 1 through N, the dimension of this degenerate subspace. Beginning with eq. (5.10.17), we have

$$\left(H_0 - \bar{E}_a\right)\left|{}_1E_a; j\right\rangle = -\left(H_1 - \delta_1 E_a\right)\left|\bar{E}_a; j\right\rangle. \tag{5.10.52}$$

Let us now act $\left\langle\bar{E}_a; i\right|$ on both sides of this equations, keeping in mind eq. (5.10.20).

$$0 = -\left\langle\bar{E}_a; i\right| H_1\left|\bar{E}_a; j\right\rangle + \delta_1 E_a\left\langle\bar{E}_a; i\middle|\bar{E}_a; j\right\rangle. \tag{5.10.53}$$

Within the degenerate subspace, we may of course choose the $\{\left|\bar{E}_a; i\right\rangle\}$ to be orthonormal,

$$\left\langle\bar{E}_a; i\middle|\bar{E}_a; j\right\rangle = \delta^i_j. \tag{5.10.54}$$

Eq. (5.10.53) then reads

$$\delta_1 E_a \cdot \delta^i_j = \left\langle\bar{E}_a; i\right| H_1\left|\bar{E}_a; j\right\rangle. \tag{5.10.55}$$

This equation teaches us why there is a need to divide our analysis into non-degenerate and degenerate cases. For the non-degenerate case, we have eq. (5.10.36). But for the degenerate case, we appear instead to arrive at a potential inconsistency. For, while the diagonal $i = j$ equations appear to return us to eq. (5.10.36); the off-diagonal $i \neq j$ equations appear to tell us H_1 must have trivial off-diagonal components,

$$\left\langle\bar{E}_a; i\right| H_1\left|\bar{E}_a; j\right\rangle = 0, \qquad (i\neq j). \tag{5.10.56}$$

But H_1 has not been specified at all; i.e., this cannot possibly be true for all possible H_1. Instead, we should view eq. (5.10.55) as an instruction to choose the basis of this degenerate subspace such that H_1 is diagonal within it:

$$\delta_{1,j} E_a \cdot \delta^i_j = \left\langle \bar{E}_a; i \right| H_1 \left| \bar{E}_a; j \right\rangle . \tag{5.10.57}$$

Note, however, that eq. (5.10.55) does not imply $\{ \left| \bar{E}_a; i \right\rangle | i = 1, 2, \dots, N \}$ are eigenvectors of H_1; because the diagonal $N \times N$ matrix equation does not say anything about $H_1 \left| \bar{E}_a; j \right\rangle$ along the directions perpendicular to these $\left| \bar{E}_a; i \right\rangle$s; namely, $\left\langle \bar{E}_{b \neq a} \right| H_1 \left| \bar{E}_a; j \right\rangle$ are not yet fixed.

We have also appended an additional subscript j to the first eigenvalue correction because, if N is the dimension of the degenerate subspace, eq. (5.10.57) now informs us the first correction to $\bar{E}_a$ could take up to N distinct values $\{ \delta_{1,j} E_a = \left\langle \bar{E}_a; j \left| H_1 \right| \bar{E}_a; j \right\rangle | j = 1, \dots, N \}$. Note that not all the $\{ \delta_{1,j} E_a \}$ may be distinct – the breaking of degeneracy is often intimately tied to the amount of symmetries enjoyed by H_1 relative to H_0. Equations (5.10.7) and (5.10.9) with this updated notation now read

$$|E_a; j\rangle = \left| \bar{E}_a; j \right\rangle + \sum_{\ell=1}^{+\infty} \sum_{s \neq a} \epsilon^\ell \left| \bar{E}_s \right\rangle \left\langle \bar{E}_s \right| {}_\ell E_a; j \rangle + \sum_{\ell=1}^{+\infty} \sum_{i=1}^{N} \epsilon^\ell \left| \bar{E}_a; i \right\rangle \left\langle \bar{E}_a; i \right| {}_\ell E_a; j \rangle , \tag{5.10.58}$$

$$E_{a,j} = \bar{E}_a + \sum_{\ell=1}^{+\infty} \epsilon^\ell \delta_{\ell,j} E_a; \tag{5.10.59}$$

and eq. (5.10.52) takes the form

$$\left(H_0 - \bar{E}_a \right) | {}_1 E_a; j \rangle = - \left(H_1 - \delta_{1,j} E_a \right) \left| \bar{E}_a; j \right\rangle . \tag{5.10.60}$$

Like the non-degenerate case, we now require that the eigenket be normalized as

$$\left\langle \bar{E}_a; j \right| E_a; j \rangle = 1. \tag{5.10.61}$$

Then, the expansion in eq. (5.10.58) tells us

$$\left\langle \bar{E}_a; j \right| \bar{E}_a; j \rangle + \sum_{\ell=1}^{\infty} \epsilon^\ell \left\langle \bar{E}_a; j \right| {}_\ell E_a; j \rangle = 1 \tag{5.10.62}$$

$$\sum_{\ell=1}^{\infty} \epsilon^\ell \left\langle \bar{E}_a; j \right| {}_\ell E_a; j \rangle = 0. \tag{5.10.63}$$

Setting the coefficient of ϵ^ℓ to zero,

$$\left\langle \bar{E}_a; j \right| {}_{\ell \geq 1} E_a; j \rangle = 0. \tag{5.10.64}$$

Even though the left hand side of eq. (5.10.60) appears to be invariant under the replacement $| {}_1 E_a; j \rangle \to | {}_1 E_a; j \rangle + \chi_1 \left| \bar{E}_a; i \right\rangle$, we will discover from the $\mathcal{O}(\epsilon^2)$ equations below that, it is inconsistent to set $\left\langle \bar{E}_a; i \neq j \right| {}_\ell E_a; j \rangle = 0$. However, eq. (5.10.64) will remain consistent.

For now, let us apply $\left\langle \bar{E}_b \right|$, for $\bar{E}_b \neq \bar{E}_a$, on both sides of eq. (5.10.60).

$$\left(\bar{E}_b - \bar{E}_a \right) \left\langle \bar{E}_{b \neq a} \right| {}_1 E_a; j \rangle = - \left\langle \bar{E}_{b \neq a} \right| H_1 \left| \bar{E}_a; j \right\rangle \tag{5.10.65}$$

Since $\bar{E}_b \neq \bar{E}_a$, we have eliminated the $\delta_{1,j}E_a$ term in eq. (5.10.60) via the orthogonality condition

$$\left\langle \bar{E}_{b\neq a}\right| \bar{E}_a; j\left.\right\rangle = 0. \tag{5.10.66}$$

Eq. (5.10.65) returns us the component of the first order eigenstate correction along $\left|\bar{E}_{b\neq a}\right\rangle$.

$$\left\langle \bar{E}_{b\neq a}\right| {}_1E_a; j\left.\right\rangle = \frac{\left\langle \bar{E}_{b\neq a}\right| H_1 \left|\bar{E}_a; j\right\rangle}{\bar{E}_a - \bar{E}_b} \tag{5.10.67}$$

As already alluded to, in order to determine $\left\langle \bar{E}_a; i\right| {}_1E_a; j\left.\right\rangle$, we need to turn to the $\mathcal{O}(\epsilon^2)$ eq. (5.10.18).

$$\left(H_0 - \bar{E}_a\right) \left| {}_2E_a; j\right\rangle = -\left(H_1 - \delta_{1,j}E_a\right) \left| {}_1E_a; j\right\rangle - \left(H_2 - \delta_{2,j}E_a\right) \left|\bar{E}_a; j\right\rangle. \tag{5.10.68}$$

Keeping in mind eq. (5.10.57), applying $\left\langle \bar{E}_a; i\right|$ and $\left\langle \bar{E}_{b\neq a}\right|$ on both sides now yield, respectively

$$0 = -\left\langle \bar{E}_a; i\right| H_1 \left| {}_1E_a; j\right\rangle + \delta_{1,j}E_a \left\langle \bar{E}_a; i\right| {}_1E_a; j\left.\right\rangle - \left\langle \bar{E}_a; i\right| H_2 \left|\bar{E}_a; j\right\rangle + \delta_{2,j}E_a\delta^i_j \tag{5.10.69}$$

$$\begin{aligned} 0 = &-\sum_{b\neq a} \left\langle \bar{E}_a; i\right| H_1 \left|\bar{E}_b\right\rangle \frac{\left\langle \bar{E}_b\right| H_1 \left|\bar{E}_a; j\right\rangle}{\bar{E}_a - \bar{E}_b} - \left(\delta_{1,i}E_a - \delta_{1,j}E_a\right) \left\langle \bar{E}_a; i\right| {}_1E_a; j\left.\right\rangle \\ &- \left\langle \bar{E}_a; i\right| H_2 \left|\bar{E}_a; j\right\rangle + \delta_{2,j}E_a\delta^i_j \end{aligned} \tag{5.10.70}$$

and

$$\begin{aligned} \left(\bar{E}_b - \bar{E}_a\right) \left\langle \bar{E}_{b\neq a}\right| {}_2E_a; j\left.\right\rangle = &-\left\langle \bar{E}_{b\neq a}\right| H_1 \left| {}_1E_a; j\right\rangle + \delta_{1,j}E_a \left\langle \bar{E}_{b\neq a}\right| {}_1E_a; j\left.\right\rangle \\ &- \left\langle \bar{E}_{b\neq a}\right| H_2 \left|\bar{E}_a; j\right\rangle + \delta_{2,j}E_a \left\langle \bar{E}_{b\neq a}\right| \bar{E}_a; j\left.\right\rangle \end{aligned} \tag{5.10.71}$$

$$\begin{aligned} \left\langle \bar{E}_b\right| {}_2E_a; j\left.\right\rangle = (\bar{E}_a - \bar{E}_b)^{-1} \Bigg(&\sum_{c\neq a} \left\langle \bar{E}_b\right| H_1 \left|\bar{E}_c\right\rangle \frac{\left\langle \bar{E}_c\right| H_1 \left|\bar{E}_a; j\right\rangle}{\bar{E}_a - \bar{E}_c} \\ &+ \sum_{\substack{k=1\\k\neq j}}^{N} \left\langle \bar{E}_{b\neq a}\right| H_1 \left|\bar{E}_a; k\right\rangle \left\langle \bar{E}_a; k\right| {}_1E_a; j\left.\right\rangle \\ &- \delta_{1,j}E_a \frac{\left\langle \bar{E}_b\right| H_1 \left|\bar{E}_a; j\right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_b\right| H_2 \left|\bar{E}_a; j\right\rangle \Bigg). \end{aligned} \tag{5.10.72}$$

By setting $i = j$ in eq. (5.10.70) we immediately obtain the second order corrections to the eigenvalue

$$\delta_{2,j}E_a = \sum_{b\neq a} \left\langle \bar{E}_a; j\right| H_1 \left|\bar{E}_b\right\rangle \frac{\left\langle \bar{E}_b\right| H_1 \left|\bar{E}_a; j\right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; j\right| H_2 \left|\bar{E}_a; j\right\rangle. \tag{5.10.73}$$

whereas the $i \neq j$ equations hand us the components of $\left| {}_1E_a; j\right\rangle$ along $\left|\bar{E}_a; i \neq j\right\rangle$,

$$\left\langle \bar{E}_a; i\right| {}_1E_a; j\left.\right\rangle = -(\delta_{1,i}E_a - \delta_{1,j}E_a)^{-1}$$

$$\times\left(\sum_{b\neq a}\left\langle\bar{E}_a;i\right|H_1\left|\bar{E}_b\right\rangle\frac{\left\langle\bar{E}_b\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_b}+\left\langle\bar{E}_a;i\right|H_2\left|\bar{E}_a;j\right\rangle\right). \tag{5.10.74}$$

This explicitly demonstrates, as long as $i\neq j$, it is inconsistent to set $\left\langle\bar{E}_a;i\right|\,{}_1E_a;j\rangle=0$. To obtain eq. (5.10.74), note that we have assumed the degeneracy has been completely lifted; so $\delta_{1,i}E_a\neq\delta_{1,j}E_a$ for all $i\neq j$.

Inserting eq. (5.10.74) into eq. (5.10.72),

$$\begin{aligned}\left\langle\bar{E}_{b\neq a}\right|\,{}_2E_a;j\rangle=(\bar{E}_a-\bar{E}_b)^{-1}&\Bigg(\sum_{c\neq a}\left\langle\bar{E}_b\right|H_1\left|\bar{E}_c\right\rangle\frac{\left\langle\bar{E}_c\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_c}\\&-\sum_{\substack{k=1\\k\neq j}}^{N}\frac{\left\langle\bar{E}_b\right|H_1\left|\bar{E}_a;k\right\rangle}{\delta_{1,k}E_a-\delta_{1,j}E_a}\left(\sum_{c\neq a}\left\langle\bar{E}_a;k\right|H_1\left|\bar{E}_c\right\rangle\frac{\left\langle\bar{E}_c\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_c}+\left\langle\bar{E}_a;k\right|H_2\left|\bar{E}_a;j\right\rangle\right)\\&-\left\langle\bar{E}_a;j\left|H_1\right|\bar{E}_a;j\right\rangle\frac{\left\langle\bar{E}_b\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_b}+\left\langle\bar{E}_b\right|H_2\left|\bar{E}_a;j\right\rangle\Bigg).\end{aligned} \tag{5.10.75}$$

To derive $\left\langle\bar{E}_a;i\right|\,{}_2E_a;j\rangle$ we need to move on to the $\mathcal{O}(\epsilon^3)$ equations. Applying $\left\langle\bar{E}_a;j\right|$ on both sides of $\ell=3$ version of eq. (5.10.35).

$$\left(H_0-\bar{E}_a\right)|\,{}_3E_a;j\rangle=-\sum_{s=1}^{2}\left(H_s-\delta_{s,j}E_a\right)|\,{}_{\ell-s}E_a;j\rangle-\left(H_3-\delta_{3,j}E_a\right)\left|\bar{E}_a;j\right\rangle. \tag{5.10.76}$$

Because we are only after $\left\langle\bar{E}_a;i\right|\,{}_2E_a;j\rangle$, let us apply $\left\langle\bar{E}_a;i\right|$ on eq. (5.10.76) to eliminate $|\,{}_3E_a;j\rangle$ (cf. eq. (5.10.20)).

$$\begin{aligned}0=&-\left(\left\langle\bar{E}_a;i\right|H_1|\,{}_2E_a;j\rangle-\delta_{1,j}E_a\left\langle\bar{E}_a;i\right|\,{}_2E_a;j\rangle\right)\\&-\left(\left\langle\bar{E}_a;i\right|H_2|\,{}_1E_a;j\rangle-\delta_{2,j}E_a\left\langle\bar{E}_a;i\right|\,{}_1E_a;j\rangle\right)\\&-\left\langle\bar{E}_a;i\right|H_3\left|\bar{E}_a;j\right\rangle\end{aligned} \tag{5.10.77}$$

Inserting the zeroth order completeness relations,

$$\begin{aligned}&\sum_{c\neq a}\left\langle\bar{E}_a;i\right|H_1\left|\bar{E}_c\right\rangle\left\langle\bar{E}_c\right|\,{}_2E_a;j\rangle+\sum_{k=1}^{N}\left\langle\bar{E}_a;i\right|H_1\left|\bar{E}_a;k\right\rangle\left\langle\bar{E}_a;k\right|\,{}_2E_a;j\rangle-\delta_{1,j}E_a\left\langle\bar{E}_a;i\right|\,{}_2E_a;j\rangle\\&=-\sum_{c\neq a}\left\langle\bar{E}_a;i\right|H_2\left|\bar{E}_c\right\rangle\left\langle\bar{E}_c\right|\,{}_1E_a;j\rangle-\sum_{k=1}^{N}\left\langle\bar{E}_a;i\right|H_2\left|\bar{E}_a;k\right\rangle\left\langle\bar{E}_a;k\right|\,{}_1E_a;j\rangle+\delta_{2,j}E_a\left\langle\bar{E}_a;i\right|\,{}_1E_a;j\rangle\\&\quad-\left\langle\bar{E}_a;i\right|H_3\left|\bar{E}_a;j\right\rangle\end{aligned} \tag{5.10.78}$$

$$\begin{aligned}&\left(\delta_{1,i}E_a-\delta_{1,j}E_a\right)\left\langle\bar{E}_a;i\right|\,{}_2E_a;j\rangle \qquad (5.10.79)\\&=-\sum_{f\neq a}\frac{\left\langle\bar{E}_a;i\right|H_1\left|\bar{E}_f\right\rangle}{\bar{E}_a-\bar{E}_f}\Bigg(\sum_{c\neq a}\left\langle\bar{E}_f\right|H_1\left|\bar{E}_c\right\rangle\frac{\left\langle\bar{E}_c\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_c}\\&\qquad-\sum_{\substack{k=1\\k\neq j}}^{N}\frac{\left\langle\bar{E}_f\right|H_1\left|\bar{E}_a;k\right\rangle}{\delta_{1,i}E_a-\delta_{1,j}E_a}\left(\sum_{c\neq a}\left\langle\bar{E}_a;k\right|H_1\left|\bar{E}_c\right\rangle\frac{\left\langle\bar{E}_c\right|H_1\left|\bar{E}_a;j\right\rangle}{\bar{E}_a-\bar{E}_c}+\left\langle\bar{E}_a;k\right|H_2\left|\bar{E}_a;j\right\rangle\right)\end{aligned}$$

$$- \left\langle \bar{E}_a; j \left| H_1 \right| \bar{E}_a; j \right\rangle \frac{\left\langle \bar{E}_f \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_f \right| H_2 \left| \bar{E}_a; j \right\rangle \Bigg)$$

$$- \sum_{c \neq a} \left\langle \bar{E}_a; i \right| H_2 \left| \bar{E}_c \right\rangle \frac{\left\langle \bar{E}_c \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_c}$$

$$+ \sum_{k=1}^{N} \frac{\left\langle \bar{E}_a; i \right| H_2 \left| \bar{E}_a; k \right\rangle}{\delta_{1,k} E_a - \delta_{1,i} E_a} \left(\sum_{b \neq a} \left\langle \bar{E}_a; k \right| H_1 \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; k \right| H_2 \left| \bar{E}_a; j \right\rangle \right)$$

$$- \left(\sum_{b \neq a} \left\langle \bar{E}_a; j \right| H_1 \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; j \right| H_2 \left| \bar{E}_a; j \right\rangle \right) (\delta_{1,i} E_a - \delta_{1,j} E_a)^{-1}$$

$$\times \left(\sum_{b \neq a} \left\langle \bar{E}_a; i \right| H_1 \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; i \right| H_2 \left| \bar{E}_a; j \right\rangle \right) - \left\langle \bar{E}_a; i \right| H_3 \left| \bar{E}_a; j \right\rangle .$$

Let us summarize the situation thus far.

- When a given set of eigenkets $\{ \left| \bar{E}_a; j \right\rangle | i = 1, 2, \dots, N \}$ of H_0 is degenerate, to find the corresponding eigenvectors of the perturbed operator

$$H = H_0 + \epsilon H_1 + \epsilon^2 H_2 + \dots, \tag{5.10.80}$$

  first ensure these $\{ \left| \bar{E}_a; j \right\rangle \}$ have been chosen such that H_1 is diagonal within this subspace (cf. eq. (5.10.57)):

$$\left\langle \bar{E}_a; i \left| H_1 \right| \bar{E}_a; j \right\rangle = \delta_{1,j} E_a \; \delta^i{}_j. \tag{5.10.81}$$

- With respect to such a basis, the perturbed eigenvalue up to second order then reads (cf. equations (5.10.57) and (5.10.73)):

$$E_{a,j} = \bar{E}_a + \epsilon \left\langle \bar{E}_a; j \left| H_1 \right| \bar{E}_a; j \right\rangle + \epsilon^2 \left(\sum_{\bar{E}_b \neq \bar{E}_a} \frac{\left| \left\langle \bar{E}_a; j \right| H_1 \left| \bar{E}_b \right\rangle \right|^2}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; j \right| H_2 \left| \bar{E}_a; j \right\rangle \right) + \mathcal{O}(\epsilon^3). \tag{5.10.82}$$

- If the first order corrections completely break the degeneracy, then the eigenkets of H up to second order, namely

$$\begin{aligned} \left| E_a; j \right\rangle = \left| \bar{E}_a; j \right\rangle + \epsilon \Bigg(&\sum_{\bar{E}_b \neq \bar{E}_a} \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} \\ &- \sum_{\substack{i=1 \\ i \neq j}}^{N} \frac{\left| \bar{E}_a; i \right\rangle}{\delta_{1,i} E_a - \delta_{1,j} E_a} \left(\sum_{b \neq a} \left\langle \bar{E}_a; i \right| H_1 \left| \bar{E}_b \right\rangle \frac{\left\langle \bar{E}_b \right| H_1 \left| \bar{E}_a; j \right\rangle}{\bar{E}_a - \bar{E}_b} + \left\langle \bar{E}_a; i \right| H_2 \left| \bar{E}_a; j \right\rangle \right) \Bigg) \\ + \epsilon^2 &\left(\sum_{\substack{i=1 \\ i \neq j}}^{N} \left| \bar{E}_a; i \right\rangle \left\langle \bar{E}_a; i \right| {}_2 E_a; j \rangle + \sum_{\bar{E}_b \neq \bar{E}_a} \left| \bar{E}_b \right\rangle \left\langle \bar{E}_b \right| {}_2 E_a; j \rangle \right) + \mathcal{O}(\epsilon^3), \end{aligned} \tag{5.10.83}$$

may be constructed using the coefficients in equations (5.10.67), (5.10.74), (5.10.75), and (5.10.79).

Examples

Problem 5.107. Auxiliary Observable for First Order Degenerate PT Suppose Q is a Hermitian operator that commutes with the first order H_1, and suppose the spectrum of Q restricted to some degenerate subspace of the zeroth order H_0 is completely distinct; that is, for the subspace spanned by $\{|\bar{E};i\rangle \, |H_0 \, |\bar{E};i\rangle = \bar{E} \, |\bar{E};i\rangle\}$, we have

$$\left\langle \bar{E};a \left| Q \right| \bar{E};b \right\rangle = q_a \cdot \delta^a{}_b. \tag{5.10.84}$$

Prove that

$$\left\langle \bar{E};a \left| H_1 \right| \bar{E};b \right\rangle = \delta_{1,a} E \cdot \delta^a{}_b. \tag{5.10.85}$$

In other words, if we already know how to diagonalize an observable Q within the degenerate subspace of the zeroth order H_0, its eigenstates necessarily also diagonalize the first order H_1 – and can therefore be used to compute the first order shift in energies – as long as $[Q, H_1] = 0$. □

5.10.2 Variational Method

The variational method is usually used to estimate the lowest eigenvalue of a given Hermitian operator. It is based on the following upper bound statement.

An Upper Bound The lowest eigenvalue E_0 of a Hermitian operator H is less than or equal to its expectation value with respect to *any* state $|\psi\rangle$ within the Hilbert space it is acting upon.

$$E_0 \leq \frac{\langle \psi \left| H \right| \psi \rangle}{\langle \psi | \psi \rangle}, \qquad \forall \left| \psi \right\rangle. \tag{5.10.86}$$

To see this, we exploit the fact that H is Hermitian to insert a complete set of its eigenstates $\{|E_a\rangle\}$ on the right hand side.

$$\frac{\langle \psi \left| H \right| \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\sum_n E_n \langle \psi | E_n \rangle \langle E_n | \psi \rangle}{\langle \psi | \psi \rangle}. \tag{5.10.87}$$

Denote the lowest eigenvalue as E_0 – i.e., $E_0 \leq E_n$ for all $n \neq 0$; we have $\leq$ instead of $<$ because the lowest eigenvalue might be degenerate. Then each term in the sum is greater than itself, but with E_a replaced with E_0; namely,

$$E_0 \langle \psi | E_n \rangle \langle E_n | \psi \rangle = E_0 | \langle \psi | E_n \rangle |^2 \leq E_n | \langle \psi | E_n \rangle |^2 = E_n \langle \psi | E_n \rangle \langle E_n | \psi \rangle \tag{5.10.88}$$

for all $n \neq 0$. Therefore

$$\frac{\sum_n E_n \langle \psi | E_n \rangle \langle E_n | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\sum_0 E_0 \langle \psi | E_0 \rangle \langle E_0 | \psi \rangle + \sum_{n \neq 0} E_n | \langle \psi | E_n \rangle |^2}{\langle \psi | \psi \rangle} \tag{5.10.89}$$

$$\geq E_0 \frac{\sum_n \langle\psi|\, E_n\rangle \, \langle E_n|\, \psi\rangle}{\langle\psi|\, \psi\rangle} = E_0 \frac{\langle\psi|\, \psi\rangle}{\langle\psi|\, \psi\rangle}. \tag{5.10.90}$$

Variational Method Since the lowest eigenvalue of H is bounded from above by any of its expectation values, one could attempt to get as close to E_0 as possible by choosing an appropriate state $|\psi\rangle$. In particular, if $|\psi\rangle$ depends on a host of parameters $\{\alpha_{\mathrm{I}}|\mathrm{I} = 1, 2, 3, \dots\}$, then

$$\mathcal{E}\left(\alpha_{\mathrm{I}}\right) \equiv \frac{\langle\psi; \{\alpha_{\mathrm{I}}\}\, |H|\, \psi; \{\alpha_{\mathrm{I}}\}\rangle}{\langle\psi; \{\alpha_{\mathrm{I}}\}|\, \psi; \{\alpha_{\mathrm{I}}\}\rangle} \tag{5.10.91}$$

is necessarily a function of these αs. We may then search for its minimum – i.e., evaluate it at the $\bar{\alpha}_{\mathrm{I}}$ obeying $\partial\mathcal{E}/\partial\bar{\alpha}_{\mathrm{I}} = 0$. This must yield, at least within this class of states $|\psi; \{\alpha_{\mathrm{I}}\}\rangle$, a number closest to E_0 (from above): $E_0 \leq \mathcal{E}(\bar{\alpha}_{\mathrm{I}})$. There is a certain art here, to cook up the right family of states, with parameters $\{\alpha_{\mathrm{I}}\}$ introduced, so that one may obtain a good enough estimate for the application at hand.

Higher Eigenstates Suppose we know how to construct states $\{|\psi\rangle\}$ that are orthogonal to the ground state, namely $\langle\psi|\, E_0\rangle = 0$; either because we somehow know the ground state $|E_0\rangle$, or using symmetry arguments (without knowing $|E_0\rangle$), or by some other means. Then the expectation value of H with respect to $|\psi\rangle$ must be greater or equal to the first excited state $E_1 > E_0$:

$$\frac{\langle\psi\,|H|\,\psi\rangle}{\langle\psi|\,\psi\rangle} \geq E_1, \qquad \forall\, \langle\psi|\, E_0\rangle = 0. \tag{5.10.92}$$

To see this, we simply insert a complete set of eigenstates.

$$\frac{\langle\psi\,|H|\,\psi\rangle}{\langle\psi|\,\psi\rangle} = \frac{\sum_{E_\ell} |\, \langle\psi|\, E_\ell\rangle\, |^2 E_\ell}{\langle\psi|\,\psi\rangle} = \frac{\sum_{E_\ell > E_0} |\, \langle\psi|\, E_\ell\rangle\, |^2 E_\ell}{\langle\psi|\,\psi\rangle} \tag{5.10.93}$$

$$\geq E_1 \frac{\sum_{E_\ell > E_0} \langle\psi|\, E_\ell\rangle \, \langle E_\ell|\, \psi\rangle}{\langle\psi|\,\psi\rangle} = E_1 \frac{\sum_{E_\ell} \langle\psi|\, E_\ell\rangle \, \langle E_\ell|\, \psi\rangle}{\langle\psi|\,\psi\rangle} = E_1. \tag{5.10.94}$$

We may continue this argument. Suppose we know how to construct $\{|\psi\rangle\}$ such that it is orthogonal to the first n eigenstates $\{|E_0\rangle, \dots, |E_{n-1}\rangle\}$, where $E_{n-1} > \dots > E_0$. Then

$$\frac{\langle\psi\,|H|\,\psi\rangle}{\langle\psi|\,\psi\rangle} \geq E_n, \tag{5.10.95}$$

$$\forall\, \langle\psi|\, E_{n-1}\rangle = \langle\psi|\, E_{n-2}\rangle = \dots = \langle\psi|\, E_1\rangle = \langle\psi|\, E_0\rangle = 0. \tag{5.10.96}$$

For,

$$\frac{\langle\psi\,|H|\,\psi\rangle}{\langle\psi|\,\psi\rangle} = \frac{\sum_{E_\ell} |\, \langle\psi|\, E_\ell\rangle\, |^2 E_\ell}{\langle\psi|\,\psi\rangle} = \frac{\sum_{E_\ell > E_{n-1}} |\, \langle\psi|\, E_\ell\rangle\, |^2 E_\ell}{\langle\psi|\,\psi\rangle} \tag{5.10.97}$$

$$\geq E_n \frac{\sum_{E_\ell > E_{n-1}} \langle\psi|\, E_\ell\rangle \, \langle E_\ell|\, \psi\rangle}{\langle\psi|\,\psi\rangle} = E_n \frac{\sum_{E_\ell} \langle\psi|\, E_\ell\rangle \, \langle E_\ell|\, \psi\rangle}{\langle\psi|\,\psi\rangle} = E_n. \tag{5.10.98}$$

Extremization We also note that, the first order perturbation of the normalized expectation value obtained by varying the state $|\psi\rangle \to |\psi\rangle + |\delta\psi\rangle$ is

$$\delta\left(\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}\right) = \frac{\langle\delta\psi|H|\psi\rangle + \langle\psi|H|\delta\psi\rangle}{\langle\psi|\psi\rangle} - \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle^2}\left(\langle\delta\psi|\psi\rangle + \langle\psi|\delta\psi\rangle\right) \tag{5.10.99}$$

$$= \frac{\langle\delta\psi|}{\langle\psi|\psi\rangle}\left(H|\psi\rangle - |\psi\rangle\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}\right) + \text{h.c.}. \tag{5.10.100}$$

Hence, if $|\psi\rangle$ is an eigenvector of H, namely $H|\psi\rangle = \lambda|\psi\rangle$ and hence $\langle\psi|H|\psi\rangle / \langle\psi|\psi\rangle = \lambda$, we in turn have

$$\delta\left(\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}\right) = 0. \tag{5.10.101}$$

On the other hand, if the first order variation of this normalized expectation value is zero for all variations $|\delta\psi\rangle$, then

$$\begin{aligned}0 &= \frac{\langle\delta\psi|}{\langle\psi|\psi\rangle}\left(H|\psi\rangle - |\psi\rangle\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}\right) + \text{h.c.} \\ &= \frac{2}{\langle\psi|\psi\rangle}\text{Re}\left[\langle\delta\psi|\left(H|\psi\rangle - |\psi\rangle\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}\right)\right].\end{aligned} \tag{5.10.102}$$

Since $|\delta\psi\rangle$ is arbitrary, we may rotate its phase $\langle\delta\psi| \to e^{-i\vartheta}\langle\delta\psi|$ (for real ϑ) and render the argument inside the square bracket real, if it were not real to begin with. Then, again by the arbitrariness of $|\delta\psi\rangle$, we may conclude that the eigensystem equation

$$H|\psi\rangle = |\psi\rangle\frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} \tag{5.10.103}$$

must hold for the expectation value to be extremized. To sum:

> **Eigenvectors extremize** The averaged expectation value of the Hermitian operator H, namely $\lambda \equiv \langle\psi|H|\psi\rangle / \langle\psi|\psi\rangle$, is extremized iff $|\psi\rangle$ is its eigenvector with corresponding eigenvalue λ.

Example Let us consider the Hermitian operator consisting of the unit radial vector $\widehat{r}$ dotted into the Pauli matrices in eq. (3.2.18):

$$\widehat{r}\cdot\vec{\sigma} \equiv \delta_{ij}\widehat{r}^i\sigma^j \tag{5.10.104}$$

$$\widehat{r}(0 \leq \theta \leq \pi, 0 < \phi \leq 2\pi) \equiv (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{5.10.105}$$

Now, up to an overall (irrelevant) multiplicative phase $e^{i\delta}$, the most general unit norm $\xi^\dagger\xi = 1$ 2-component object ξ can be parametrized as

$$\xi(\alpha,\beta) = \left(e^{i\beta}\sin\alpha, \cos\alpha\right)^{\mathrm{T}}, \tag{5.10.106}$$

where α, β are real angles. We are going to extremize the expectation value

$$\mathcal{E}(\alpha,\beta) \equiv \frac{\langle\xi|\widehat{r}\cdot\vec{\sigma}|\xi\rangle}{\langle\xi|\xi\rangle} = \xi^\dagger(\widehat{r}\cdot\vec{\sigma})\xi \tag{5.10.107}$$

$$= \sin(2\alpha)\sin(\theta)\cos(\beta+\phi) - \cos(2\alpha)\cos(\theta). \tag{5.10.108}$$

Because eq. (5.10.106) is the most general 2-component object, the extremum of $\mathcal{E}$ through the variation of ξ should not only provide an estimate of the lowest eigenvalue; it should in fact provide both the exact eigenvalues and their corresponding eigenvectors.

Differentiation $\mathcal{E}$ with respect to α and β, and setting the results to zero yield the following relations.

$$\begin{aligned} 0 &= \frac{\partial \mathcal{E}}{\partial \alpha} \\ &= 2\cos(2\alpha)\sin(\theta)\cos(\beta+\phi) + 2\sin(2\alpha)\cos(\theta) \end{aligned} \tag{5.10.109}$$

$$\begin{aligned} 0 &= \frac{\partial \mathcal{E}}{\partial \beta} \\ &= -\sin(2\alpha)\sin(\theta)\sin(\beta+\phi). \end{aligned} \tag{5.10.110}$$

Suppose $\sin\theta = 0$, then eq. (5.10.110) becomes trivial; while equations (5.10.108) and (5.10.109) become instead

$$\mathcal{E} = -(2\cos(\alpha)^2 - 1)\cos(\theta) \tag{5.10.111}$$

$$0 = \frac{\partial \mathcal{E}}{\partial \alpha} = \sin(\alpha)\cos(\alpha)\cos(\theta). \tag{5.10.112}$$

If $\sin(2\alpha) = 0 = \sin\theta$, the possible solutions are

$$(\alpha, \cos\alpha, \theta, \cos\theta, \mathcal{E}) = (0, 1, 0, 1, -1), \tag{5.10.113}$$

$$(\alpha, \cos\alpha, \theta, \cos\theta, \mathcal{E}) = (0, 1, \pi, -1, +1), \tag{5.10.114}$$

$$(\alpha, \cos\alpha, \theta, \cos\theta, \mathcal{E}) = (\pi, -1, 0, 1, -1), \tag{5.10.115}$$

$$(\alpha, \cos\alpha, \theta, \cos\theta, \mathcal{E}) = (\pi, -1, \pi, -1, +1). \tag{5.10.116}$$

If $\theta = 0$, the $\sin\theta = 0$ and $\cos\theta = +1$. Then $\partial\mathcal{E}/\partial\beta$ is trivially zero; whereas $\partial\mathcal{E}/\partial\alpha = 0 = 2\sin(2\alpha)$. This in turn implies β can be anything; while $\alpha = 0, \pm\pi/2, \pm\pi, \pm(3/2)\pi, \dots = (n/2)\pi$.

$$\mathcal{E}(\alpha = (n/2)\pi, \beta) = -(-)^n \tag{5.10.117}$$

$$\xi(\alpha = n\pi, \beta) = (0, 1)^{\mathrm{T}} \tag{5.10.118}$$

Problem 5.108. Variation Method for Excited States via Symmetry S

5.11 *Ordinary Differential Operators and Their Inverses

In this section we are going to consider the vector space of single-variable functions defined on some interval $a \leq z \leq b$ on the real line, endowed with the following inner product. Given two arbitrary functions $f_{1,2}$,

$$\langle f_1 | f_2 \rangle = \int_a^b \langle f_1 | z \rangle \langle z | f_2 \rangle w(z) \mathrm{d}z \tag{5.11.1}$$

$$= \int_a^b f_1(z)^* f_2(z) w(z) \mathrm{d}z. \tag{5.11.2}$$

The $w(z)$ is some appropriately chosen real function such that the inner product of any state $|f\rangle$ with itself is always non-negative:

$$\langle f|\, f\rangle = \int_a^b |f(z)|^2 w(z) \mathrm{d}z \geq 0. \tag{5.11.3}$$

Decomposing the inner product in eq. (5.11.1) as

$$\langle f_1|\, f_2\rangle = \langle f_1| \left(\int_a^b |z\rangle \langle z|\, w(z)\mathrm{d}z\right) |f_2\rangle \tag{5.11.4}$$

in turn tells us the identity operator is expressed as

$$\int_a^b |z\rangle \langle z|\, w(z)\mathrm{d}z = \mathbb{I}, \tag{5.11.5}$$

since $f_{1,2}$ are arbitrary. We may view $|z\rangle$ as the position eigenket restricted to the range $a \leq z \leq b$, with the inner product

$$\langle z|\, z'\rangle = \frac{\delta(z-z')}{\sqrt{w(z)w(z')}}. \tag{5.11.6}$$

Note that, because of the $\delta(z-z')$ enforcing $z=z'$, the denominator $\sqrt{w(z)w(z')}$ can also be written as either $w(z)$ or as $w(z')$.

5.11.1 Second Order Linear ODEs and Wronksians

[41]The key object of study in this section is the general second order linear ordinary differential equation (ODE) – both the homogeneous version

$$D_z\psi(z) \equiv p_2(z)\frac{\mathrm{d}^2\psi(z)}{\mathrm{d}z^2} + p_1(z)\frac{\mathrm{d}\psi(z)}{\mathrm{d}z} + p_0(z)\psi(z) = 0; \tag{5.11.7}$$

as well as its inhomogeneous one

$$D_z\psi(z) = J(z); \tag{5.11.8}$$

for some real domain $a \leq z \leq b$. Here, we shall assume the coefficients $p_{0,1,2}$ and the 'source' $J(z)$ are given and real; except for p_2, which we shall assume is positive. (If p_2 were negative, we may simply multiply -1 throughout to render it positive.) An equivalent way of writing the differential operator D is

$$D_z\psi(z) \equiv \frac{1}{w(z)}\frac{\mathrm{d}}{\mathrm{d}z}\left(w(z)p_2(z)\frac{\mathrm{d}\psi(z)}{\mathrm{d}z}\right) + p_0(z)\psi(z). \tag{5.11.9}$$

Observe that the sign of w is immaterial since they cancel out; hence, it is often useful to assume $w \geq 0$. As we shall discover in this subsection, the set of homogeneous solutions obeying eq. (5.11.7) spans a two dimensional vector space. The key tool we shall need to understand this result is known as the Wronskian. In the subsequent subsections, we shall further encounter the inverse of the differential operator D, and witness how it plays a key role in constructing solutions to both equations (5.11.7) and (5.11.8).

[41]This subsection was inspired by part by Simmons [34].

Problem 5.109. Equivalence of D in equations (5.11.7) and (5.11.9) Show that the Ds in equations (5.11.7) and (5.11.9) are equivalent by relating p_2, w and p_1.

Hint: Explain why the key identification is

$$p_1(z) = \frac{1}{w(z)}\frac{\mathrm{d}}{\mathrm{d}z}(w(z)p_2(z)). \tag{5.11.10}$$

In the next subsection, we shall exploit the form of D in eq. (5.11.9) to argue that *any* linear second order differential operator can be viewed as a Hermitian one provided the appropriate boundary conditions are imposed. $\square$

Wronskian The Wronskian of n functions $\{f_1(z), f_2(z), \dots, f_n(z)\}$ of some variable z is defined as the following determinant (cf. eq. (3.2.1)):

$$\mathrm{Wr}_z\left(f_1, \dots, f_n\right) \equiv \det \begin{bmatrix} f_1(z) & f_2(z) & \dots & f_n(z) \\ f_1'(z) & f_2'(z) & \dots & f_n'(z) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(z) & f_2^{(n-1)}(z) & \dots & f_n^{(n-1)}(z) \end{bmatrix} \tag{5.11.11}$$

$$= \epsilon^{\mathrm{I}_1\mathrm{I}_2\dots\mathrm{I}_n} f_{\mathrm{I}_1}(z) f_{\mathrm{I}_2}'(z) f_{\mathrm{I}_3}^{(2)}(z) \dots f_{\mathrm{I}_n}^{(n-1)}(z), \tag{5.11.12}$$

where $\epsilon^{\mathrm{I}_1\mathrm{I}_2\dots\mathrm{I}_n}$ is the Levi-Civita symbol with $\epsilon^{(1)(2)\dots(n)} \equiv 1$. Let us now prove

> If the set of functions $\{f_1, \dots, f_n\}$ are linearly dependent over some domain $a \leq z \leq b$, their Wronskian $\mathrm{Wr}_z\left(f_1, \dots, f_n\right)$ is zero there. Equivalently, if their Wronskian is non-zero over the same domain, they are linearly independent.

Suppose $\{f_1, \dots, f_n\}$ are linearly dependent over some domain $a \leq z \leq b$. That means there are a set of constants $\{c_\mathrm{I}\}$, of which two or more are non-zero, such that

$$c_1 f_1(z) + \dots + c_n f_n(z) = 0. \tag{5.11.13}$$

If the only solution was $c_1 = c_2 = \dots = c_n = 0$, the set is linearly independent. Hence, a set is either linearly independent or linearly dependent. In any case, since eq. (5.11.13) holds over a range of z, we may differentiate it n times to obtain the matrix equation

$$\begin{bmatrix} f_1(z) & f_2(z) & \dots & f_n(z) \\ f_1'(z) & f_2'(z) & \dots & f_n'(z) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(z) & f_2^{(n-1)}(z) & \dots & f_n^{(n-1)}(z) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{5.11.14}$$

where the first row corresponds to eq. (5.11.13), second row its first derivative, and so on until the last row is its nth derivative. Since the c's are non-zero, the determinant of the $n \times n$ matrix – i.e., the $\mathrm{Wr}_z\left(f_1, \dots, f_n\right)$ – must be zero for $a \leq z \leq b$; otherwise the matrix is invertible and the only solution is $c_1 = \dots = c_n = 0$.

Problem 5.110. Two Properties of Wronskian Show that the Wronskian $\mathrm{Wr}_z(f_1, f_2, \dots, f_n)$ is anti-symmetric under the interchange of any pair of functions $f_{\mathrm{I}} \leftrightarrow f_{\mathrm{J}}$ for I, J $\in \{1, 2, \dots, n\}$:

$$\mathrm{Wr}_z\left(f_1, \dots, f_{\mathrm{I}}, \dots, f_{\mathrm{J}}, \dots, f_n\right) = -\mathrm{Wr}_z\left(f_1, \dots, f_{\mathrm{J}}, \dots, f_{\mathrm{I}}, \dots, f_n\right). \tag{5.11.15}$$

Next, consider 'rotating' the set of functions $\{f_1, \dots, f_n\}$,

$$g_{\mathrm{I}}(z) \equiv M_{\mathrm{I}}{}^{\mathrm{J}} f_{\mathrm{J}}(z), \tag{5.11.16}$$

where the $n \times n$ matrix $M_{\mathrm{I}}{}^{\mathrm{J}}$ is constant and invertible; show that

$$\mathrm{Wr}_z(g_1, \dots, g_n) = \left(\det M_{\mathrm{I}}{}^{\mathrm{J}}\right) \mathrm{Wr}_z(f_1, \dots, f_n). \tag{5.11.17}$$

Hint: Equations (3.2.1) and (3.2.10) may be useful here. □

Second Order Case If two functions $f_{1,2}$ are linearly dependent, $f_1(z) = A \cdot f_2(z)$ for some non-zero constant A, and therefore their Wronskian must be zero due to the latter's anti-symmetric property:

$$\mathrm{Wr}_z(f_1, f_2) = A \cdot \mathrm{Wr}_z(f_1, f_1) = 0. \tag{5.11.18}$$

Therefore, if the $\mathrm{Wr}_z(f_1, f_2) \neq 0$, the two functions $f_{1,2}$ must be linearly independent. On the other hand, in Problem (5.131) below, we shall see there are linearly independent functions whose Wronskian is zero – for example z and $z \cdot |z|$ in the neighborhood of $z \sim 0$ – and therefore zero Wronskian as a standalone criterion *does not* in general imply linear dependence. However, if we know that at least one of the two non-trivial functions $\{f_1, f_2\}$ is not only continuously differentiable, but is strictly non-zero within $a \leq z \leq b$, then the pair are in fact linearly dependent if their Wronskian is zero. That is, suppose

$$\mathrm{Wr}_z(f_1, f_2) = f_1(z) f_2'(z) - f_1'(z) f_2(z) = 0; \tag{5.11.19}$$

and further suppose $f_1(a \leq z \leq b) \neq 0$. We may then re-express

$$\mathrm{Wr}_z(f_1, f_2) = f_1^2 (f_2/f_1)' = 0. \tag{5.11.20}$$

Since $f_1 \neq 0$ we may divide throughout by $f_1(z)^2$ and integrate once to obtain $f_2 = C \cdot f_1$ for some non-zero constant C. In fact, notice in the above example that $f_1(z) = z^2$ and $f_2(z) = z \cdot |z|$ are linearly dependent if z is restricted to be either negative or positive, where they are strictly non-zero. To sum:

> The pair of functions $\{f_1, f_2\}$ have zero Wronskian iff they are linearly dependent, provided at least one of them is strictly non-zero within the relevant domain $a \leq z \leq b$. Equivalently, they are linearly independent iff their Wronskian is non-zero, provided one of them is strictly non-zero within the same domain.

Problem 5.111. Wronskian: Solution If $\psi \in \{f_1, \dots, f_n\}$ obeys the nth order homogeneous ODE

$$p_n(z) \frac{\mathrm{d}^n \psi}{\mathrm{d}z^n} + \dots + p_1(z) \frac{\mathrm{d}\psi}{\mathrm{d}z} + p_0(z)\psi(z) = 0 \tag{5.11.21}$$

show that the Wronskian obeys the first order differential equation

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Wr}_z(f_1,\dots,f_n) = -\frac{p_{n-1}(z)}{p_n(z)}\mathrm{Wr}_z(f_1,\dots,f_n), \tag{5.11.22}$$

provided p_{n-1}/p_n is non-singular. Integrate this equation from some fixed z_0 to obtain the Wronskian up to an overall multiplicative constant – without the need to know explicitly the solutions $\{f_1,\dots,f_n\}$:

$$\mathrm{Wr}_z(f_1,\dots,f_n) = W_0 \exp\left(-\int_{z_0}^{z}\frac{p_{n-1}(z'')}{p_n(z'')}\mathrm{d}z''\right), \tag{5.11.23}$$

$$W_0 = \mathrm{Wr}_{z=z_0}(f_1,\dots,f_n) = \text{constant}. \tag{5.11.24}$$

Since exp(something real and finite) is strictly positive, that means the sign of the Wronskian is fixed: it is determined solely by the constant W_0. $\square$

At this juncture, let us record an important theorem:

> **Uniqueness of solution to initial value problem** If Q_1 and Q_0 are continuous within $a \leq z \leq b$, then the solution to
>
> $$\psi''(z) + Q_1(z)\psi'(z) + Q_0(z)\psi(z) = 0 \tag{5.11.25}$$
>
> subject to the 'initial conditions'
>
> $$\psi(z_0) = \psi_0 \qquad \text{and} \qquad \psi'(z_0) = \psi_0' \tag{5.11.26}$$
>
> – where z_0 is some fixed location lying within $[a,b]$ whereas ψ_0 and ψ_0' are arbitrary but fixed numbers – is unique.

Note that, if ψ and $Q_{0,1}$ were all smooth, so that each of them admit a Taylor series around the neighborhood of $z \sim z_0$, then one may simply compute the derivatives of ψ at z_0 recursively. The second derivative may simply be obtained from eq. (5.11.25) since

$$\psi''(z_0) = -Q_1(z_0)\psi_0' - Q_0(z_0)\psi_0. \tag{5.11.27}$$

Taking one derivative of eq. (5.11.25) then allows $\psi'''(z_0)$ to be solved in terms of ψ_0, ψ_0' and ψ_0''. Taking n derivatives,

$$\psi^{(n+2)}(z_0) = -\sum_{\ell=0}^{n}\binom{n}{\ell}\left(p_1^{(n-\ell)}(z_0)\psi^{(\ell+1)}(z_0) + p_0^{(n-\ell)}(z_0)\psi^{(\ell)}(z_0)\right), \tag{5.11.28}$$

allowing us to determine the $(n+2)$th derivative of ψ in terms of the zeroth through $(n+1)$th derivatives. Since the Taylor series is unique for a smooth function, namely $\psi(z) = \sum_{\ell=0}^{\infty}(z-z_0)^\ell\psi^{(\ell)}(z_0)/\ell!$, this determines the function ψ uniquely.

2D Vector Space That the initial value problem has a unique solution now informs us that the second order homogeneous ODE in equations (5.11.7) and (5.11.25) admits a two dimensional vector space of solutions, at least within the region $a \leq z \leq b$ where $Q_1 = p_1/p_2$ and $Q_0 = p_0/p_2$ are non-singular – namely, any solution $\psi(a \leq z \leq b)$ can be written as

$$\psi(z) = c_1 f_1(z) + c_2 f_2(z) \tag{5.11.29}$$

for suitably chosen constants $c_{1,2}$, provided $\{f_1, f_2\}$ themselves solve eq. (5.11.7) or (5.11.25) and are linearly independent. To understand why, simply choose a fixed point z_0 lying within $[a, b]$ and compute $\psi(z_0)$ and its first derivative $\psi'(z_0)$. By linearity we already know $c_1 f_1 + c_2 f_2$ solves eq. (5.11.7) or (5.11.25) as long as $f_{1,2}$ do. By the above uniqueness theorem, what remains is to verify that $c_{1,2}$ may chosen so that the right hand sides also obeys the same 'initial conditions':

$$\psi(z_0) = c_1 f_1(z_0) + c_2 f_2(z_0), \tag{5.11.30}$$

$$\psi'(z_0) = c_1 f_1'(z_0) + c_2 f_2'(z_0); \tag{5.11.31}$$

which can be packaged into the matrix form

$$\begin{bmatrix} \psi(z_0) \\ \psi'(z_0) \end{bmatrix} = \begin{bmatrix} f_1(z_0) & f_2(z_0) \\ f_1'(z_0) & f_2'(z_0) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}. \tag{5.11.32}$$

This matrix equation renders it manifest that – as long as the 2×2 matrix is invertible – the solution to $c_{1,2}$ is unique and

$$\psi(z) = \begin{bmatrix} f_1(z) \\ f_2(z) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} f_1(z_0) & f_2(z_0) \\ f_1'(z_0) & f_2'(z_0) \end{bmatrix}^{-1} \begin{bmatrix} \psi(z_0) \\ \psi'(z_0) \end{bmatrix} \tag{5.11.33}$$

$$= f_1(z) \frac{\mathrm{Wr}_{z_0}(\psi, f_2)}{\mathrm{Wr}_{z_0}(f_1, f_2)} - f_2(z) \frac{\mathrm{Wr}_{z_0}(\psi, f_1)}{\mathrm{Wr}_{z_0}(f_1, f_2)}. \tag{5.11.34}$$

The inverse 2×2 matrix in the first line exists whenever the Wronskian of $\{f_1, f_2\}$ is non-zero. In turn, within the region where f_1 or f_2 are strictly non-zero, the $\mathrm{Wr}_{z_0}(f_1, f_2)$ is non-zero iff the pair $\{f_1, f_2\}$ are linearly independent.

Problem 5.112. Verify that eq. (5.11.34) follows from eq. (5.11.33). □

Problem 5.113. Generating One Independent Solution From Another It turns out, if one of the two linearly independent functions $\psi \in \{f_1, f_2\}$ obeying eq. (5.11.7) are known – say f_1 – the Wronskian solution in eq. (5.11.23), namely

$$\mathrm{Wr}_z(f_1, f_2) = W_0 \exp\left(- \int^z \frac{p_1(z')}{p_2(z')} \mathrm{d}z' \right), \tag{5.11.35}$$

may be exploited to generate f_2. Show by exploiting $\mathrm{Wr}_z(f_1, f_2) = f_1^2 (f_2/f_1)'$ (cf. eq. (5.11.20)) or otherwise, that

$$f_2(z) = f_1(z) \left(W_0 \int^z \frac{\mathrm{d}z''}{f_1(z'')^2} \exp\left(- \int^{z''} \frac{p_1(z')}{p_2(z')} \mathrm{d}z' \right) + W_1 \right), \tag{5.11.36}$$

where $W_{0,1}$ are arbitrary constants. Of course, the $W_1 \cdot f_1(z)$ portion is linearly dependent on f_1 itself; but, explain why, as long as W_0 is non-zero, we are guaranteed that $f_{1,2}$ are linearly independent. □

Problem 5.114. Damped Harmonic Oscillator The damped harmonic oscillator obeys the differential equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}}{\mathrm{d}t} + \Omega^2\right) x(t) = 0, \tag{5.11.37}$$

where $x(t)$ is its Cartesian position as a function of time t, $2\gamma > 0$ is the friction parameter, and Ω the un-damped oscillation angular frequency. We may solve for its linearly independent solutions via ansatz. Starting from

$$x(t) = x_0 e^{\Gamma \cdot t}, \tag{5.11.38}$$

where both x_0 and Γ are constants, show that the two solutions are $\Gamma = -\gamma \pm \sqrt{\gamma^2 - \Omega^2}$. Next, use eq. (5.11.36) to generate (up to multiplicative constant factors) $x_1(z) \propto \exp((\sqrt{\Omega^2-\gamma^2} - \gamma)t)$ from $x_2(z) \propto \exp((-\sqrt{\Omega^2-\gamma^2} - \gamma)t)$ or vice versa.

Notice this solution-by-ansatz fails when $\gamma = \Omega$, in that it only generates a single solution $x_1(t) \propto e^{-\gamma \cdot t}$. Use eq. (5.11.36) to derive the other linearly independent solution $x_2(t) \propto t \cdot e^{-\gamma t}$.

Hint: You may set $W_1 = 0$ because it is $W_0 \neq 0$ that ensures linear independence of the second solution. □

Problem 5.115. First derivative may be transformed away To facilitate the analysis of second order linear differential equations, it is often useful to first transform away its first derivative term. Show that this is always possible. Specifically, for D taking the form in eq. (5.11.7), show that

$$\begin{aligned} D_z & \left(\exp\left(-\frac{1}{2}\int^z \frac{p_1(z')}{p_2(z')}\mathrm{d}z'\right) F(z)\right) \\ & = \exp\left(-\frac{1}{2}\int^z \frac{p_1(z')}{p_2(z')}\mathrm{d}z'\right)\left(p_2(z)F''(z) + Q_0(z)F(z)\right), \end{aligned} \tag{5.11.39}$$

where

$$Q_0(z) = p_0(z) - \frac{p_1(z)^2}{4p_2(z)} + \frac{p_1(z)p_2'(z) - p_1'(z)p_2(z)}{2p_2(z)}. \tag{5.11.40}$$

For D taking the form in eq. (5.11.9), on the other hand, verify

$$D_z\left(\frac{F(z)}{\sqrt{w(z)\cdot p_2(z)}}\right) = \frac{p_2(z)F''(z) + Q_0(z)F(z)}{\sqrt{w(z)\cdot p_2(z)}}, \tag{5.11.41}$$

where, instead,

$$Q_0(z) = p_0(z) + \frac{\left(w(z)p_2'(z) - p_2(z)w'(z)\right)^2}{4p_2(z)w(z)^2} - \frac{p_2(z)w''(z)}{2w(z)} - \frac{p_2''(z)}{2}. \tag{5.11.42}$$

An alternate approach for the D in eq. (5.11.9) is to re-express $w \cdot p_2 \cdot (\mathrm{d}/\mathrm{d}z) \equiv \mathrm{d}/\mathrm{d}y$, so that

$$D_z\psi(z) = \frac{1}{w(y)^2 p_2(y)}\left(\frac{\mathrm{d}^2\psi(y)}{\mathrm{d}y^2} + w(y)^2 p_2(y)p_0(y)\psi(y)\right). \tag{5.11.43}$$

Show that is may be accomplished via the 'coordinate redefinition'

$$y(z) \equiv y(z=a) + \int_a^z \frac{\mathrm{d}z'}{w(z')p_2(z')}. \tag{5.11.44}$$

□

Problem 5.116. Constant Wr(F_1, F_2) One reason why we transform away the first derivative term is following. If we were solving the homogeneous equation $D\psi = 0$, explain why the Wronskian of two linearly independent solutions to the above Fs in equations (5.11.39) and (5.11.41) must be constant. (Hint: How does the Wronskian depend on the first derivative term?) This constancy of the Wronskian will in turn aid us in the study of the qualitative behavior of homogeneous solutions. □

Problem 5.117. Wronskian of two different ODEs On the other hand, suppose F_1 and F_2 obey *different* equations,

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + R_1(z)\right) F_1(z) = 0 = \left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + R_2(z)\right) F_2(z), \tag{5.11.45}$$

where $R_1 \neq R_2$, show that

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Wr}_z\left(F_1, F_2\right) = \left(R_1(z) - R_2(z)\right) F_1(z) F_2(z). \tag{5.11.46}$$

We shall use this result to compare zeroes of eigenfunctions corresponding to different eigenvalues. □

Problem 5.118. Two Examples For the damped harmonic oscillator, show that

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}}{\mathrm{d}t} + \Omega^2\right)\left(e^{-\gamma \cdot t}X(t)\right) = e^{-\gamma \cdot t}\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \Omega^2 - \gamma^2\right) X(t); \tag{5.11.47}$$

while for the differential operator occurring in Bessel's equation,

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + \frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z} + 1 - \frac{\nu^2}{z^2}\right)\frac{F(z)}{\sqrt{z}} = \frac{1}{\sqrt{z}}\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + 1 + \frac{1-(2\nu)^2}{(2z)^2}\right) F(z). \tag{5.11.48}$$

These results can help us understand the zeroes of $X(t)$ and $F(z)$. □

Zeroes of Homogeneous Solutions We shall conclude this subsection by developing some qualitative understanding of the homogeneous solutions to the general second order linear differential equation with four theorems regarding their zeroes.

The first theorem relates the zeroes of linearly independent solutions:

> **ZI** Whenever p_1/p_2 is non-singular, and the Wronskian of a pair of functions $\{f_1, f_2\}$ is non-zero – i.e., $\{f_1, f_2\}$ are hence linearly independent – the zeroes of f_1 and f_2 alternate: the zero of f_1 is followed by that of f_2, then by the zero of f_1, and so on.

For, at some zero of f_1, which we shall dub z_{I}, we have

$$\mathrm{Wr}_{z_{\mathrm{I}}}(f_1, f_2) = f_1(z_{\mathrm{I}})f_2'(z_{\mathrm{I}}) - f_1'(z_{\mathrm{I}})f_2(z_{\mathrm{I}}) = -f_1'(z_{\mathrm{I}})f_2(z_{\mathrm{I}}); \tag{5.11.49}$$

and the next zero to the right, which we shall dub z_{II}, we have

$$\mathrm{Wr}_{z_{\mathrm{II}}}(f_1, f_2) = f_1(z_{\mathrm{II}})f_2'(z_{\mathrm{II}}) - f_1'(z_{\mathrm{II}})f_2(z_{\mathrm{II}}) = -f_1'(z_{\mathrm{II}})f_2(z_{\mathrm{II}}). \tag{5.11.50}$$

Because the sign of the Wronskian stays fixed, both of these quantities must be of the same sign. However, the derivatives at alternate zeroes, namely $f_1'(z_{\mathrm{I}})$ and $f_1'(z_{\mathrm{II}})$, must be of opposite sign. (Drawing a picture helps.) Hence, the only way this can be consistent is for $f_2(z_{\mathrm{I}})$ and $f_2(z_{\mathrm{II}})$ to have opposite signs – by continuity, f_2 must have passed through odd number of zeroes between the two zeroes z_{I} and z_{II} of f_1. We could have applied this argument using two consecutive zeroes of f_2 instead of those of f_1; and conclude that f_1 must have gone through odd number of zeroes in between. But that means between the two consecutive zeroes of f_1 must lie precisely one zero of f_2, and vice versa, because if there were more than one zero of f_2 between two consecutive zeroes of f_1, then by the preceding argument between these zeroes of f_2 must lie more zeroes of f_1 – thereby contradicting the assumption that they lie within two consecutive zeroes of f_1.

Both $x = \sin t$ and $x = \cos t$ both satisfy $\ddot{x} + x = 0$, with $\mathrm{Wr}(\sin, \cos) = -1$; and, indeed, their zeroes do alternate: $\sin(0) = 0$, $\cos(\pi/2) = 0$, $\sin(\pi) = 0$, etc. More non-trivial are the Bessel functions $J_\nu(z)$ and $Y_\nu(z)$, whose Wronskian is given by

$$\mathrm{Wr}_z\left(J_\nu(z), Y_\nu(z)\right) = \frac{2}{\pi z}, \tag{5.11.51}$$

and by the above theorem must have alternating zeroes at least for $z > 0$.

Next, let us prove that:

> **ZII** If $F''(z) = -R_0(z)F(z)$ and R_0 is strictly negative over some domain $z \in [a, b]$; then $F(z)$ has at most one zero over the same domain.

If we had found such a zero at $z = z_0$, then $F'' = -R_0 \cdot F = |R_0 \cdot F|\mathrm{sgn}(F)$ tells us, if $F'(z_0) > 0$ then as z moves away from z_0 the derivative will only grow steeper on either side; similar arguments would apply if $F'(z_0) < 0$. (Drawing a picture helps.) But that means it is not possible for the function to turn back around to re-intersect the horizontal axis. For that to occur R_0 must become positive for some domain of z.

A simple example of this general observation is the equation $\ddot{x} = \Omega^2 x$, where $\Omega > 0$. The general solution is $x(t) = c_- e^{-\Omega \cdot t} + c_+ e^{+\Omega \cdot t}$, for constants $c_\pm$. If both $c_\pm$ are of the same signs there are no zeroes at all, since the exponentials are positive. Only when the $c_\pm$ are of opposite signs is there exactly one zero at $t = \Omega^{-1} \cdot \ln\sqrt{-c_-/c_+}$.

Problem 5.119. Explain why $X(t)$ in eq. (5.11.47) has at most one zero for all real t when $\gamma > \Omega$. Similarly, argue why $F(z)$ in eq. (5.11.48) has at most one zero for $0 < z < (1/2)\sqrt{(2\nu)^2 - 1}$ at least for $\nu > 1/2$.

Another example is the modified Bessel equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + \frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z} - \left(1 + \frac{\nu^2}{z^2}\right)\right)\psi(z) = 0. \tag{5.11.52}$$

Argue that the solution has at most one zero when $\nu > 1/2$ and $z \geq 0$. $\square$

Next, we prove that:

ZIII If $F''(z) = -R_0(z)F(z)$, with $R_0(z \geq z_\star) > 0$ for some fixed $z_\star$ and the integral of R_0 to infinity is unbounded – namely

$$\int_{\geq z_\star}^{+\infty} R_0(z)\mathrm{d}z = +\infty \tag{5.11.53}$$

– then $\psi(z)$ has infinite number of zeroes between $z_\star$ and $+\infty$; in other words, it is oscillatory there. This un-boundness holds for all lower limits greater or equal to some fixed $z_\star$.[42]

For, suppose there are only finite number of zeroes. Then, if z_0 denote the rightmost zero and if $\psi'(z_0) > 0$ – which implies $\psi(z > z_0) > 0$ – let us, following [34], define $v \equiv -\psi'/\psi$, which must begin as a negative quantity at $z = z_0$. Its first derivative must obey

$$v' = -\psi''/\psi + (\psi'/\psi)^2 = R_0 + v^2. \tag{5.11.54}$$

Integrating it once,

$$v(z) - v(z_0) = \int_{z_0}^{z} \mathrm{d}z' \left(R_0(z') + v(z')^2\right). \tag{5.11.55}$$

Since the integral of R_0 is unbounded, that means $v(z) = -\psi'(z)/\psi(z)$ must become positive at some sufficiently large z, indicating $\psi'(z)$ must have flipped sign since we had assumed $\psi'(z_0) > 0$. Moreover, since $R_0 > 0$ for $z \geq z_0$, the derivative ψ' will continue to decrease and ψ *has* to hit the horizontal axis at some point, contradicting the assumption that z_0 was the rightmost zero. Running the argument by assuming $\psi'(z_0) < 0$ will lead to the same contradiction.

A simple application of this theorem is to the simple harmonic oscillator $\ddot{x} = -\Omega^2 x$. Of course, we already know the general solution is the superposition of sines and cosines, and the infinite number of zeroes can be enumerated explicitly. But with this theorem, we do not even need to know the explicit solutions to draw the same conclusion regarding their existence. Another simple application is to the damped harmonic oscillator $\ddot{x} + 2\gamma\dot{x} + \Omega^2 x = 0$. Using eq. (5.11.47), we see that for $\Omega > \gamma > 0$, the position $x(t)$ must be $e^{-\gamma \cdot t}$ multiplied by a function that oscillates indefinitely with increasing t.

Problem 5.120. Bessel functions Prove that the Bessel functions $\psi \in \{J_\nu(z > 0), Y_\nu(z > 0)\}$ have infinite number of zeroes. Hint: The transformation in eq. (5.11.48) may be useful. □

Finally, let us prove:

ZIV If F_1 and F_2 satisfies eq. (5.11.45) and if $R_1 > R_2$ for $z \in [a, b]$, then between two zeroes of F_2 must lie *at least one* zero of F_1.

[42] On the other hand, it is not possible for F to yield infinite number of zeroes in a finite interval $[a, b]$; i.e., where a and b are finite real numbers.

By integrating eq. (5.11.45) from some zero $z_{\rm I}$ of F_2 to the next consecutive zero $z_{\rm II}$ to its right, we obtain

$$F_1(z_{\rm II})F_2'(z_{\rm II}) - F_1(z_{\rm I})F_2'(z_{\rm I}) = \int_{z_{\rm I}}^{z_{\rm II}} \left(R_1(z') - R_2(z')\right) F_1(z')F_2(z')\mathrm{d}z'. \tag{5.11.56}$$

If $F_{1,2}$ solves eq. (5.11.45) so does their negative $-F_{1,2}$; in particular, the location of their zeroes is not affected by this overall sign flip. Hence, let us suppose F_1 is strictly positive and F_2 only vanishes at $z_{\rm I}$ and $z_{\rm II}$ and nowhere in between. The latter informs us that $F_2'(z_{\rm I})$ must be positive and $F_2'(z_{\rm II})$ negative. Hence, the integral is strictly positive and

$$F_1(z_{\rm II})F_2'(z_{\rm II}) > F_1(z_{\rm I})F_2'(z_{\rm I}) \tag{5.11.57}$$

which implies – since $F_2'(z_{\rm II}) < 0$ –

$$F_1(z_{\rm II}) < F_1(z_{\rm I})\frac{F_2'(z_{\rm I})}{F_2'(z_{\rm II})} \leq 0. \tag{5.11.58}$$

But this is a contradiction, since we have assumed F_1 is strictly positive.

The simplest example illustrating this theorem is the pair of SHOs $\ddot{x}+\Omega_1^2 x = 0$ and $\ddot{y}+\Omega_2^2 y = 0$, with $\Omega_1 > \Omega_2$. The SHO with the higher frequency oscillates more rapidly, and for instance when $\Omega_1 \gg \Omega_2$, many zeroes of x would lie between any two zeroes of y.

Problem 5.121. Compare the zeroes of $\sin(\Omega \cdot x)$ and $\cos(\Omega \cdot x)$ with those of $J_\nu(x > 0)$ and $Y_\nu(x > 0)$, for $\Omega > 0$ and $\nu > 0$. □

5.11.2 Second Order Linear Differential Operators: Hermitian Case

The inverse of differential operators is usually known as the *Green's function*,[43] whose physical application spans both classical and quantum physics. We begin the study of Green's functions by considering the inverse of the following second order Hermitian differential operators of a single variable.[44]

$$D_z f(z) \equiv \frac{1}{w(z)}\frac{\mathrm{d}}{\mathrm{d}z}\left(w(z)p_2(z)\frac{\mathrm{d}f(z)}{\mathrm{d}z}\right) + p_0(z)f(z), \tag{5.11.59}$$

where $p_{0,2}(z)$ are assumed to be real. If this operator is defined over the interval $a \leq z \leq b$, then by Hermitian we mean that

$$\langle f_1 | D | f_2 \rangle \equiv \int_a^b f_1(z)^* \left(D_z f_2(z)\right) w(z)\mathrm{d}z \tag{5.11.60}$$

$$= \left[w(z)p_2(z)f_1(z)^* \frac{\mathrm{d}f_2(z)}{\mathrm{d}z}\right]_{z=a}^{z=b} + \int_a^b \left(-\frac{\mathrm{d}f_1(z)^*}{\mathrm{d}z}p_2(z)\frac{\mathrm{d}f_2(z)}{\mathrm{d}z} + p_0(z)f_1(z)^* f_2(z)\right) w(z)\mathrm{d}z \tag{5.11.61}$$

[43]In some books, "*Green function*" is used instead, without the apostrophe.

[44]The inverse of higher order differential operators of a single variable is examined in Byron and Fuller [14] Chapter 7 Problems 7, 8, and 9.

$$= \left[w(z)p_2(z) \left(f_1(z)^* \frac{\mathrm{d}f_2(z)}{\mathrm{d}z} - \frac{\mathrm{d}f_1(z)^*}{\mathrm{d}z} f_2(z) \right) \right]_{z=a}^{z=b}$$
$$+ \int_a^b \left\{ \frac{1}{w(z)} \frac{\mathrm{d}}{\mathrm{d}z} \left(w(z)p_2(z) \frac{\mathrm{d}f_1(z)^*}{\mathrm{d}z} \right) + p_0(z) f_1(z)^* \right\} f_2(z) w(z) \mathrm{d}z$$
$$= \int_a^b \left(D_z f_1(z)^* \right) f_2(z) w(z) \mathrm{d}z = \left(D \left| f_1 \right\rangle \right)^\dagger \left| f_2 \right\rangle; \tag{5.11.62}$$

for all functions $f_{1,2}(z)$ that satisfy

$$\left[w(z)p_2(z) \left(f_1(z)^* \frac{\mathrm{d}f_2(z)}{\mathrm{d}z} - \frac{\mathrm{d}f_1(z)^*}{\mathrm{d}z} f_2(z) \right) \right]_{z=a}^{z=b} = 0. \tag{5.11.63}$$

Two common choices of boundary conditions that would satisfy eq. (5.11.63) are:

- *Dirichlet (aka, homogeneous)* $f_{1,2}(z = a) = 0 = f_{1,2}(z = b)$;
- *Neumann* $(\mathrm{d}/\mathrm{d}z) f_{1,2}(z = a) = 0 = (\mathrm{d}/\mathrm{d}z) f_{1,2}(z = b)$.

Eigensystem of Hermitian D: Dirichlet Case Let us focus on the case of Dirichlet boundary conditions. Since D is Hermitian it must enjoy a complete set of eigenbasis functions $\{\psi_\lambda(z) \equiv \langle z | \lambda \rangle\}$, where

$$D \left| \lambda \right\rangle = \lambda \left| \lambda \right\rangle. \tag{5.11.64}$$

In the position representation

$$\langle z | D | \lambda \rangle = D_z \langle z | \lambda \rangle = \lambda \langle z | \lambda \rangle. \tag{5.11.65}$$

Using the form of D_z in eq. (5.11.9) and defining $\psi_\lambda(z) \equiv \langle z | \lambda \rangle$,

$$\frac{1}{w(z)} \frac{\mathrm{d}}{\mathrm{d}z} \left(w(z)p_2(z) \frac{\mathrm{d}\psi_\lambda(z)}{\mathrm{d}z} \right) + \left(p_0(z) - \lambda \right) \psi_\lambda(z) = 0. \tag{5.11.66}$$

By re-scaling the eigenfunction as

$$\psi_\lambda(z) \equiv \frac{F_\lambda(z)}{\sqrt{w(z) \cdot p_2(z)}}, \tag{5.11.67}$$

followed by employing the result in equations (5.11.41), the eigensystem equation becomes

$$F_\lambda''(z) + \frac{Q_0(z) - \lambda}{p_2(z)} F_\lambda(z) = 0, \tag{5.11.68}$$

where Q_0 is given by eq. (5.11.42). As long as $w \cdot p_2$ is non-singular, we see that the zeroes of $\psi_\lambda(z)$ are encoded in $F(z)$, which in turn is governed by a second order differential equation without first derivative terms. Alternatively, if we invoke eq. (5.11.43) to transform z to y given in eq. (5.11.44),

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}y^2} + w^2 \cdot p_2 \cdot (p_0 - \lambda)\psi = 0; \tag{5.11.69}$$

where this eigensystem equation now holds within $y \in [y(z=a), y(z=b)] \equiv [y_a, y_b]$.

Now, Dirichlet boundary conditions mean we need to solve either eq. (5.11.68) or eq. (5.11.69), subject to $\psi_\lambda(z=a) = 0 = \psi_\lambda(z=b)$ or $\psi_\lambda(y_a) = 0 = \psi_\lambda(y_b)$. In these forms, we see that $(Q_0 - \lambda)/p_2$ or $p_0 - \lambda$ need to be positive and large enough for ψ_λ to develop at least two zeroes; namely, for $\psi_\lambda(z)$ or $\psi_\lambda(y)$ to go to zero at least at the end points a and b or y_a and y_b. (Recall that both w and p_2 may be assumed to be non-negative.) To solve eq. (5.11.68) or (5.11.69), however, we may first phrase it as an initial value problem. Firstly, because eigenfunctions remain eigenfunctions upon multiplication by a constant – the Dirichlet boundary conditions are preserved too – note that the derivative $\psi'_\lambda(a)$ or $\psi'_\lambda(y_a)$ is thus arbitrary. If $\psi'_\lambda(a)$ or $\psi'_\lambda(y_a)$ were not unity, say, we may consider instead $\psi_\lambda^{(\text{new})}(z) \equiv \psi_\lambda(z)/\psi'_\lambda(a)$ or $\psi_\lambda^{(\text{new})}(y) \equiv \psi_\lambda(y)/\psi'_\lambda(y_a)$ and we may thereby always set $\psi'_\lambda(a) = 1$ or $\psi'_\lambda(y_a) = 1$. (Only when the solution is found do we then need to worry about normalization.) By starting from $z = a$ or $y = y_a$ and $\psi_\lambda(a) = 0$ or $\psi_\lambda(y_a) = 0$, we may then try out different values of $-\lambda$ until we find the next zero of $\psi_\lambda(z)$ or $\psi_\lambda(y)$ to lie exactly at $z = b$ or $y = y_b$. Let us label this eigenvalue as λ_0, and continue to increase $-\lambda$ away from $-\lambda_0$. If we apply theorem ZIV in the previous subsection to eq. (5.11.68) or (5.11.69), whenever

$$\frac{-\lambda}{p_2(z)} > \frac{-\lambda_0}{p_2(z)} \qquad \text{or} \qquad w(y)^2 p_2(y)(-\lambda) > w(y)^2 p_2(y)(-\lambda_0) \tag{5.11.70}$$

holds, there must now be at least one zero of ψ_λ lying between a and b or y_a and y_b. Since this is a continuous process, what must be happening is that the zero at $y = y_b$ has moved leftward while the 'new' $\psi_\lambda(y)$ is changing its profile away from $\psi_{\lambda_0}(y)$. Eventually, $(Q_0 - \lambda)/p_2$ or $p_0 - \lambda$ is large enough that another zero develops at $z = b$ or $y = y_b$; satisfying the eigensystem equation with Dirichlet boundary conditions, but now the eigenstate has three zeroes – in addition to the two at the boundaries at a and b or y_a and y_b, there is now an additional one in between, which we shall dub $z_{1,1}$ or $y_{1,1}$. If this second eigenvalue is labeled λ_1, then $\psi_{\lambda_1}(a) = \psi_{\lambda_1}(z_{1,1}) = \psi_{\lambda_1}(b) = 0$ or $\psi_{\lambda_1}(y_a) = \psi_{\lambda_1}(y_{1,1}) = \psi_{\lambda_1}(y_b) = 0$. We may continue to the process, increasing $-\lambda$ while solving eq. (5.11.69) as an initial value problem. The zero at b or y_b would again start moving to the left until another zero appears at y_b. The process can be carried out indefinitely simply by increasing $-\lambda$; such that, the (J+1)-th eigenvalue is $\lambda_{\rm J}$ and its corresponding eigenfunction $\psi_{\lambda_{\rm J}}(z)$ or $\psi_{\lambda_{\rm J}}(y)$ has J zeroes $\{z_{{\rm J},1}, z_{{\rm J},2}, \dots, z_{{\rm J},{\rm J}}\}$ between the end points.

Let us also observe that each eigenfunction $\psi_{\lambda_{\rm J}}$ is unique, up to an overall multiplicative constant. We have already noted that the derivative at the left boundary can always be set to unity without loss of generality; in fact, the choice of the overall normalization constant may be viewed as the choice of the derivative of ψ_λ at the left boundary. But $(\psi_{\lambda_{\rm J}}(a) = 0, \psi'_{\lambda_{\rm J}}(a) = 1)$ or $(\psi_{\lambda_{\rm J}}(y_a) = 0, \psi'_{\lambda_{\rm J}}(y_a) = 1)$, for a *fixed* eigenvalue $\lambda_{\rm J}$, yields a unique solution, since the initial value problem does.

Altogether, we now gather:

- There is a smallest $-\lambda_0$ such that there are exactly two zeroes, one at each of the end points of the domain $z \in [a, b]$. The eigenfunction $\psi_{\lambda_0}(z)$ can therefore be taken as either strictly non-negative or non-positive over this domain.

- The eigenvalues must be discrete, since the λ in the initial value solution process is continuous, and hence ψ_λ cannot jump abruptly from an eigenstate with, say, two zeroes to one with three zeroes over a finite domain $[a, b]$.

- Since the $-\lambda$ in the initial value solution process can be increased indefinitely, the solutions must simply oscillate more and more rapidly, allowing for more zeroes to form. There must therefore be infinite number of eigenvalues $\{\lambda_0, \lambda_1, \lambda_2, \dots\}$, where $-\lambda_0 < -\lambda_1 < -\lambda_2 < \dots$, and corresponding eigenfunctions $\{\psi_{\lambda_0}(z), \psi_{\lambda_1}(z), \psi_{\lambda_2}(z), \dots\}$.

Example: Infinite Square Well As a simple example of the above construction, we may recall the the Dirichlet eigensystem of $D_z \equiv (\mathrm{d}/\mathrm{d}z)^2$, restricted between $z = 0$ and $z = L > 0$:

$$\frac{\mathrm{d}^2\psi_{\lambda_n}(z)}{\mathrm{d}z^2} = -\left(\frac{n\pi}{L}\right)^2 \psi_{\lambda_n}(z), \tag{5.11.71}$$

$$\psi_{\lambda_n}(z) \propto \sin\left(\frac{n\pi}{L} \cdot z\right); \tag{5.11.72}$$

where $n = 1, 2, 3, \dots$. We start with

$$\psi''(z) - \lambda \cdot \psi(z) = 0 \tag{5.11.73}$$

and the initial value problem

$$\psi_\lambda(z = 0) = 0 \qquad \text{and} \qquad \psi'_\lambda(z = 0) = 1. \tag{5.11.74}$$

The solution is

$$\psi_\lambda(z) = \sin\left(\sqrt{-\lambda} \cdot z\right) / \sqrt{-\lambda}. \tag{5.11.75}$$

When $-\lambda$ takes a small positive value the sine has a very long wavelength. As we increase $-\lambda$ its wavelength will decrease until the sine eventually develops a zero at $z = L$, when $\sqrt{-\lambda_0}L = \pi$; or, $\lambda_0 = -(\pi/L)^2$. As we increase $-\lambda$ further, the zero that was at $z = L$ will move to the left; while $\psi_{\lambda>\lambda_0}(z = L) = \sin(\sqrt{-\lambda}L)/\sqrt{-\lambda}$ itself is no longer a zero, at least until $\sqrt{-\lambda}L = 2\pi$; or, $\lambda_1 = -(2\pi/L)^2$. This process can be carried on indefinitely, where more and more zeroes of sine exist between $[0, L]$.

Problem 5.122. Uniqueness of 'Ground State' We have argued that the eigenstate $\psi_{\lambda_0}(z)$ corresponding to the lowest $-\lambda \equiv -\lambda_0$ cannot have any zeroes between $a < z < b$. Use theorem ZI from the preceding subsection to prove that $\psi_{\lambda_0}(z)$ must therefore be unique up to an overall multiplicative constant. □

Problem 5.123. Orthogonality vs Zeroes There is a complimentary route to understand the relative location of the zeroes of different eigenstates, via the orthogonality of the states corresponding to distinct eigenvalues, which is guaranteed by the hermitian character of D.

Consider, for instance, the two lowest consecutive eigenstates $\psi_{\lambda_0}(z)$ and $\psi_{\lambda_1}(z)$. Since $-\lambda_0$ is the lowest (negative) eigenvalue, $\psi_{\lambda_0}(z)$ may be assumed to be strictly non-negative, because it has no zeroes in (a, b). Explain why,

$$\int_a^b \mathrm{d}z' w(z') \psi_{\lambda_0}(z') \psi_{\lambda_1}(z') = 0 \tag{5.11.76}$$

implies $\psi_{\lambda_1}(z)$ *must* have at least one zero between a and b. Hint: Consider what happens when ψ_{λ_1} has no zeroes between a and b. □

Mode Expansion of Green's Function Because of the Hermitian character of D, it must admit a complete set of eigenfunctions.

$$D\left|\lambda\right\rangle = \lambda\left|\lambda\right\rangle \tag{5.11.77}$$

$$D = \sum_{\lambda} \lambda\left|\lambda\right\rangle\left\langle\lambda\right| \tag{5.11.78}$$

We may them use it to construct its inverse, i.e., the Green's function:

$$G_s = \sum_{\lambda} \frac{\left|\lambda\right\rangle\left\langle\lambda\right|}{\lambda}, \tag{5.11.79}$$

$$G_s[z, z'] \equiv \left\langle z\left|G\right|z'\right\rangle = \sum_{\lambda} \frac{\left\langle z\right|\lambda\rangle\left\langle\lambda\right|z'\rangle}{\lambda}. \tag{5.11.80}$$

That is, if $\langle z|\,\lambda\rangle \equiv \psi_\lambda(z)$ is the unit norm eigenfunction of D_z with eigenvalue λ – i.e., $D_z\psi_\lambda(z) = \lambda \cdot \psi_\lambda(z)$ – then we have

$$G_s[z, z'] = \sum_{\lambda} \frac{\psi_\lambda(z)\psi_\lambda(z')^*}{\lambda}; \tag{5.11.81}$$

$$\langle\lambda'|\,\lambda\rangle = \int_a^b \overline{\psi_{\lambda'}(z)}\psi_\lambda(z)w(z)\mathrm{d}z = \delta^{\lambda'}{}_{\lambda}; \tag{5.11.82}$$

$$\sum_{\lambda} \psi_\lambda(z)\psi_\lambda(z')^* = \frac{\delta(z-z')}{\sqrt{w(z)w(z')}}. \tag{5.11.83}$$

This mode expansion immediately teaches us that the Green's function obeys *reciprocity* – or, hermiticity in the operator sense –

$$G_s[z, z']^* = G_s[z', z]; \tag{5.11.84}$$

as well as the ordinary differential equation (ODE)

$$D_z G_s[z, z'] = D_{z'} G_s[z, z'] = \sum_{\lambda} \psi_\lambda(z)\psi_\lambda(z')^* = \frac{\delta(z-z')}{\sqrt{w(z)w(z')}}. \tag{5.11.85}$$

Example Consider the operator $D_z \equiv -(\mathrm{d}/\mathrm{d}z)^2 = P^2$; where $p_2 = -1$, $w = 1$, and $p_0 = 0$. This can be viewed as the square of the momentum operator $P = -i(\mathrm{d}/\mathrm{d}z)$. If z runs over the entire real line, its eigenfunction is

$$\langle z|\,k\rangle = e^{ikz} \tag{5.11.86}$$

for $k \in \mathbb{R}$; with associated eigenvalue $\lambda = k^2$, since

$$\left\langle z\left|D\right|k\right\rangle = -(\mathrm{d}/\mathrm{d}z)^2 e^{ikz} = k^2 e^{ikz}. \tag{5.11.87}$$

The mode sum expansion in eq. (5.11.80) thus reads

$$G_s[z-z'] = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} \frac{e^{ik(z-z')}}{k^2}. \tag{5.11.88}$$

Referring to eq. (7.2.8) below,

$$G_s(z-z') = \left\langle z \left| \frac{1}{P^2} \right| z' \right\rangle = \frac{1}{2}|z-z'|. \tag{5.11.89}$$

Problem 5.124. $(-\mathbf{d}^2/\mathbf{d}z^2)^{-1}$ for a box Solve $G_s(z,z')$ for $D_z = -(\mathrm{d}/\mathrm{d}z)^2$; but now z, z' is confined in the box of length $L > 0$, i.e., $z, z' \in [0, L]$. You should find

$$\begin{aligned} G_s(z,z') = \frac{L}{2\pi^2}\Big(&\mathrm{Li}_2\left(e^{-i\frac{\pi}{L}(z-z')}\right) + \mathrm{Li}_2\left(e^{i\frac{\pi}{L}(z-z')}\right) \\ &- \mathrm{Li}_2\left(e^{-i\frac{\pi}{L}(z+z')}\right) - \mathrm{Li}_2\left(e^{i\frac{\pi}{L}(z+z')}\right)\Big), \end{aligned} \tag{5.11.90}$$

where Li_2 is the dilogarithm. Hint: See DLMF 25.12. □

Problem 5.125. Hermite Polynomials A more sophisticated problem than the momentum-squared one above involves adding a quadratic 'potential term':

$$D_z \equiv -\frac{\mathrm{d}^2}{\mathrm{d}z^2} + z^2, \tag{5.11.91}$$

with $z \in \mathbb{R}$. (In quantum mechanics, this D is twice of the unit mass and frequency simple harmonic oscillator Hamiltonian.) Let us impose Dirichlet boundary conditions at infinity; i.e., for the nth eigenstate

$$D_z \psi_n(z) = \lambda_n \psi_n(z), \tag{5.11.92}$$

it goes to zero as $x \to \pm\infty$:

$$\psi_n(x \to \pm\infty) = 0. \tag{5.11.93}$$

In this problem, we shall diagonlize D and obtain the mode sum representation of its inverse.

- Explain why

$$\langle z | D | \psi \rangle = \left\langle z \left| P^2 + X^2 \right| \Psi \right\rangle, \tag{5.11.94}$$

where $|\Psi\rangle$ is an arbitrary state; $|z\rangle$ is the position eigenket; and X and P are respectively the position and momentum operators.

- Treating $D = P^2 + X^2$ as an abstract linear operator, show that it may be factorized as

$$D = (P - iX)^\dagger (P - iX) + 1. \tag{5.11.95}$$

Hint: Remember $[X, P] = i$.

- By defining

$$a \equiv P - iX, \tag{5.11.96}$$

so that $D = a^\dagger a + 1$, verify that

$$[a, a^\dagger] = 2 \tag{5.11.97}$$

and therefore

$$[D, a] = -2a \qquad \text{and} \qquad [D, a^\dagger] = +2a^\dagger. \tag{5.11.98}$$

- These a and $a^\dagger$ are dubbed, respectively, the lowering and raising operator. To see this, if $D\,|\lambda\rangle = \lambda\,|\lambda\rangle$, use eq. (5.11.98) to show that

$$a^\dagger\,|\lambda\rangle \propto |\lambda+2\rangle \qquad \text{and} \qquad a\,|\lambda\rangle \propto |\lambda-2\rangle\,. \tag{5.11.99}$$

Hint: Consider $D(a^\dagger\,|\lambda\rangle)$ and $D(a\,|\lambda\rangle)$.

- On the other hand, explain why $\langle\Psi\,|D|\,\Psi\rangle \geq 0$, and therefore all eigenvalues of D are non-negative. We have just witnessed: Applying the lowering operator a to an eigenstate with eigenvalue λ yields an eigenstate with eigenvalue $\lambda-2$. After n applications of a, the resulting eigenstate has eigenvalue $\lambda-2n$. At some point, this $\lambda-2n$ is going to become negative for large enough n. This implies there must be a lowest eigenvalue λ_0, such that λ_0-2 is negative, which obeys

$$a\,|\lambda_0\rangle = |\text{zero}\rangle\,; \tag{5.11.100}$$

for, if the right hand side were not $|\text{zero}\rangle$, we would obtain a negative eigenvalue. Solve eq. (5.11.100) in the position representation – i.e., $\langle z\,|a|\,\lambda_0\rangle = 0$ – and, by imposing $\langle\lambda_0|\,\lambda_0\rangle = 1$, show that the unique solution is

$$\langle z|\,\lambda_0\rangle = \pi^{-1/4}\exp\left(-z^2/2\right) \tag{5.11.101}$$

up to an overall multiplicative phase factor; and the lowest eigenvalue is thus

$$\lambda_0 = 1. \tag{5.11.102}$$

The uniqueness of this eigensystem is why there is only *one* tower of eigenstates that can be reached from each other via repeated application of a and $a^\dagger$. Explain why this leads to the complete spectrum

$$\lambda_n = 2n+1, \qquad\qquad n = 0,1,2,3,\dots. \tag{5.11.103}$$

- Let us now build the unit-norm $\{|\lambda_n\rangle\}$ from the 'ground state' $|\lambda_0\rangle$. First show that

$$\left[a,(a^\dagger)^n\right] = 2n(a^\dagger)^{n-1}. \tag{5.11.104}$$

(Hint: Try induction on n.) Then argue that, up to an overall multiplicative phase factor,

$$|\lambda_n\rangle = \frac{(-i)^n(a^\dagger)^n\,|\lambda_0\rangle}{\sqrt{2^n\cdot n!}}, \tag{5.11.105}$$

$$\langle z|\,\lambda_n\rangle = \frac{(-)^n}{\sqrt{2^n n!}}\left(\partial_z - z\right)^n\frac{\exp(-z^2/2)}{\pi^{1/4}}. \tag{5.11.106}$$

(Hint: $\langle\lambda_n|\,\lambda_n\rangle \propto \left\langle\lambda_0\left|a^n(a^\dagger)^n\right|\lambda_0\right\rangle$.) The multiple derivatives acting on the Gaussian would generate a polynomial of degree n. These eigenstates are thus usually packaged as

$$\langle z|\,\lambda_n\rangle \equiv \frac{H_n(z)}{2^{n/2}\sqrt{n!}}\,\langle z|\,\lambda_0\rangle\,, \tag{5.11.107}$$

where the

$$H_n(z) = (-)^n\exp(+z^2/2)\left(\partial_z - z\right)^n\exp(-z^2/2) \tag{5.11.108}$$

are called Hermite polynomials.

- Finally, show that the inverse of $D_z = -(\mathrm{d}/\mathrm{d}z)^2 + z^2$ with Dirichlet boundary conditions, i.e., the Green's function G_s, is

$$G_s(z, z') = \left\langle z \left| \frac{1}{D} \right| z' \right\rangle = \frac{\exp(-\frac{1}{2}(z^2 + z'^2))}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{1}{2n+1} \frac{H_n(z) H_n(z')}{2^n n!}. \tag{5.11.109}$$

□

Problem 5.126. Damped Harmonic Oscillator (DHO) The damped harmonic oscillator, with position $x(t)$, is described by the differential equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\gamma \frac{\mathrm{d}}{\mathrm{d}t} + \Omega^2\right) x(t) \equiv D_t x(t) = 0. \tag{5.11.110}$$

The 2γ and Ω are, respectively, the friction per unit mass and the angular frequency.

Show that D_t here can be cast in the form of eq. (5.11.9) by explaining why

$$p_2 = 1, \qquad w = e^{2\gamma t}, \qquad \text{and} \qquad p_0 = \Omega^2. \tag{5.11.111}$$

Let us focus on the time interval $t \in [t_\mathrm{i}, t_\mathrm{f}]$. Show that the associated Dirichlet Green's function of D_t is given by

$$\begin{aligned} G_s(t, t') = \frac{2 \exp\left(-\gamma(t + t')\right)}{\Delta T} \sum_{n=1}^{\infty} \left(\Omega^2 - \gamma^2 - \left(\frac{n\pi}{\Delta T}\right)^2\right)^{-1} & \\ \times \sin\left(\frac{n\pi}{\Delta T}(t - t_\mathrm{i})\right) \sin\left(\frac{n\pi}{\Delta T}(t' - t_\mathrm{i})\right); & \end{aligned} \tag{5.11.112}$$

where $\Delta T \equiv t_\mathrm{f} - t_\mathrm{i}$. □

Hermitian Differential Operators: Examples Some examples of the Hermitian form of the linear second order differential operator are as follows.

$$\text{Laguerre:} \qquad D_z f = e^z \frac{\mathrm{d}}{\mathrm{d}z}\left(z \cdot e^{-z} \frac{\mathrm{d}f}{\mathrm{d}z}\right), \qquad 0 \leq z < \infty, \tag{5.11.113}$$

$$\text{Hermite:} \qquad D_z f = e^{z^2} \frac{\mathrm{d}}{\mathrm{d}z}\left(e^{-z^2} \frac{\mathrm{d}f}{\mathrm{d}z}\right), \qquad z \in \mathbb{R}, \tag{5.11.114}$$

$$\text{Chebyshev:} \qquad D_z f = \sqrt{1 - z^2} \frac{\mathrm{d}}{\mathrm{d}z}\left(\sqrt{1 - z^2} \frac{\mathrm{d}f}{\mathrm{d}z}\right), \qquad -1 < z < +1, \tag{5.11.115}$$

$$\text{Legendre:} \qquad D_z f = \frac{\mathrm{d}}{\mathrm{d}z}\left((1 - z^2) \frac{\mathrm{d}f}{\mathrm{d}z}\right), \qquad -1 \leq z \leq +1. \tag{5.11.116}$$

Problem 5.127. Identify the w, p_2 and p_0 of equations (5.11.113) through (5.11.116). □

No Boundary Conditions Needed? For the *Legendre* differential operator in eq. (5.11.116) defined on the interval $-1 \leq z \leq +1$ – namely, $a = -1$ and $b = +1$ – we see that, as long as $f(z = \pm 1)$ and $f'(z = \pm 1)$ are finite, the $w = 1$ and $p_2(z) = 1 - z^2$ will set the boundary terms in eq. (5.11.63) to zero, and render D_z Hermitian. In fact, we have already seen the complete set of eigenfunctions of D_z: the Legendre polynomials $\{P_\ell(z)|\ell = 0, 1, 2, 3, \dots\}$, which satisfy

$$\frac{\mathrm{d}}{\mathrm{d}z}\left((1-z^2)\frac{\mathrm{d}P_\ell(z)}{\mathrm{d}z}\right) = -\ell(\ell+1)P_\ell(z); \tag{5.11.117}$$

i.e., the eigenvalues of D_z are $-\ell(\ell+1)$ for $\ell = 0, 1, 2, \dots$.

Discontinuous First Derivatives We now infer from eq. (5.11.85) that the first derivatives of the Green's function near $z \approx z'$ must be discontinuous, *regardless of the boundary conditions it obeys*. By integrating $D_z G = \delta/w$ around $z \approx z'$ and as long as p_2 is continuous, then for $\varepsilon \to 0^+$:

$$\int_{z'-\varepsilon}^{z'+\varepsilon}\left\{\frac{\mathrm{d}}{\mathrm{d}z}\left(w(z)p_2(z)\frac{\mathrm{d}G}{\mathrm{d}z}\right) + w(z)p_0(z)G(z,z')\right\}\mathrm{d}z = \int_{z'-\varepsilon}^{z'+\varepsilon}\delta(z-z')\mathrm{d}z. \tag{5.11.118}$$

Roughly speaking, for a second order differential operator to produce a $\delta-$function singularity, the function it is acting on must be continuous otherwise a second derivative would produce a δ' singularity. On the other hand, if the first derivative were continuous, the second derivative would be at most discontinuous. (These considerations are closely related to the identity $(\mathrm{d}/\mathrm{d}z)\Theta(z) = \delta(z)$, where Θ is the step function.) Therefore, we may assert that G itself must be continuous at $z \approx z'$ and the second term on the left hand side must integrate to zero, leaving us with

$$w(z=z')p_2(z=z')\left(\frac{\mathrm{d}G(z=z'+\varepsilon)}{\mathrm{d}z} - \frac{\mathrm{d}G(z=z'-\varepsilon)}{\mathrm{d}z}\right) = 1. \tag{5.11.119}$$

A similar calculation involving $D_{z'}G = \delta/w$ would hand us

$$w(z=z')p_2(z=z')\left(\frac{\mathrm{d}G(z'=z+\varepsilon)}{\mathrm{d}z'} - \frac{\mathrm{d}G(z'=z-\varepsilon)}{\mathrm{d}z'}\right) = 1. \tag{5.11.120}$$

We may summarize this 'jump condition' using the following notation,

$$w \cdot p_2 \cdot \left[G'\right]_{z=z'} = 1. \tag{5.11.121}$$

where the square bracket denotes the *difference* of the quantity inside across the interface $z = z'$.

Hermitian Green's Function from Homogeneous Solutions Now, for $z \neq z'$ the $\delta(z-z')$ on the right hand side of $DG = \delta/w$ is zero and we need to solve the homogeneous equations

$$D_z G[z,z'] = 0 = D_{z'}G[z,z']. \tag{5.11.122}$$

Since 2nd order ODEs should admit two linearly independent solutions $\psi_{1,2}$, we expect that the most general form of G_s to be

$$\begin{aligned} G[z,z'] &= A_<{}^{\mathrm{IJ}} \cdot \psi_{\mathrm{I}}(z)\psi_{\mathrm{J}}(z'), \qquad z < z' \\ &= A_>{}^{\mathrm{IJ}} \cdot \psi_{\mathrm{I}}(z)\psi_{\mathrm{J}}(z'), \qquad z > z'; \end{aligned} \tag{5.11.123}$$

where the $A_<$s and $A_>$s are constants; and the I and J indices run over $\{1,2\}$. Since the $G(z,z')$ needs to be continuous across $z=z'$,

$$\begin{aligned}
&{A_>}^{11}\cdot\psi_1(z)^2+{A_>}^{22}\cdot\psi_2(z)^2+({A_>}^{12}+{A_>}^{21})\psi_1(z)\psi_2(z)\\
&\qquad={A_<}^{11}\cdot\psi_1(z)^2+{A_<}^{22}\cdot\psi_2(z)^2+({A_<}^{12}+{A_<}^{21})\psi_1(z)\psi_2(z)
\end{aligned}\tag{5.11.124}$$

For this to hold for any $a\leq z\leq b$, the coefficients of $\psi_1(z)^2$, $\psi_2(z)^2$ and $\psi_1(z)\psi_2(z)$ on both sides must be equal.

$${A_>}^{11}={A_<}^{11}\equiv A^{11}\tag{5.11.125}$$

$${A_>}^{22}={A_<}^{22}\equiv A^{22}\tag{5.11.126}$$

$${A_>}^{12}+{A_>}^{21}={A_<}^{12}+{A_<}^{21}\tag{5.11.127}$$

Dirichlet B.C.s The analysis up to this point holds for any boundary condition, since we have not used any. But if D is viewed as a Hermitian operator acting on the space of functions vanishing at the boundaries $z=a$ and $z=b$, then we must also have

$$G_s[z=a,z']=0=G_s[z,z'=a]\tag{5.11.128}$$

$$G_s[z=b,z']=0=G_s[z,z'=b].\tag{5.11.129}$$

Employing the form in eq. (5.11.123), we see that

$$\psi_{\mathrm{M}}(a){A_<}^{\mathrm{MN}}\psi_{\mathrm{N}}(z')=0=\psi_{\mathrm{M}}(z){A_<}^{\mathrm{MN}}\psi_{\mathrm{N}}(b),\tag{5.11.130}$$

$$\psi_{\mathrm{M}}(b){A_>}^{\mathrm{MN}}\psi_{\mathrm{N}}(z')=0=\psi_{\mathrm{M}}(z){A_>}^{\mathrm{MN}}\psi_{\mathrm{N}}(a).\tag{5.11.131}$$

To be consistent with Dirichlet boundary conditions, we may demand that $\psi_{\mathrm{I}}(a)=0=\psi_{\mathrm{I}}(b)$. But the only consistent solution that vanishes at both end points would be zero. Hence, we should only require either ψ_1 to vanish at $z=a$; and ψ_2 to do so at $z=b$ – or, vice versa. (The $\psi_{1,2}$ are linearly independent but otherwise arbitrary at this point, so either choice is equivalent.) For the former,

$$\psi_2(a){A_<}^{2\mathrm{N}}\psi_{\mathrm{N}}(z')=0=\psi_{\mathrm{M}}(z){A_<}^{\mathrm{M}1}\psi_1(b),\tag{5.11.132}$$

$$\psi_1(b){A_>}^{1\mathrm{N}}\psi_{\mathrm{N}}(z')=0=\psi_{\mathrm{M}}(z){A_>}^{\mathrm{M}2}\psi_2(a).\tag{5.11.133}$$

Since $\{\psi_{\mathrm{N}}(z')\}$ (or $\{\psi_{\mathrm{N}}(z)\}$) are linearly independent we conclude that the only non-zero components are ${A_<}^{12}$ and ${A_>}^{21}$. At this point, recalling eq. (5.11.127) to recognize ${A_>}^{21}={A_<}^{12}\equiv A$,

$$G_s[z,z']=\Theta(z-z')A\cdot\psi_2(z)\psi_1(z')+\Theta(z'-z)A\cdot\psi_1(z)\psi_2(z').\tag{5.11.134}$$

We have only fixed $\psi_1(a)$ and $\psi_2(b)$; to fix the solutions $\psi_{1,2}(z)$ uniquely we would have to specify, say, $\psi_1(b)$ and $\psi_2(a)$, which in turn amounts to specifying the overall amplitudes of these solutions. But since $G_s(z,z')$ involves the product $\psi_1\cdot\psi_2$, that means $G_s(z,z')$ at this point is determined once the overall amplitude A is pinned down; i.e., the individual amplitudes of $\psi_{1,2}$ are not needed. To this end, let us apply the junction conditions in equations (5.11.119) and (5.11.120):

$$A\cdot\mathrm{Wr}_z\left(\psi_1,\psi_2\right)=\frac{1}{w(z)p_2(z)};\tag{5.11.135}$$

where the $\mathrm{Wr}_z(\psi_1, \psi_2)$, dubbed the Wronskian of the two solutions $\psi_{1,2}$, is given by

$$\mathrm{Wr}_z(\psi_1, \psi_2) \equiv \psi_1(z)\psi_2'(z) - \psi_1'(z)\psi_2(z) = \det \begin{bmatrix} \psi_1(z) & \psi_2(z) \\ \psi_1'(z) & \psi_2'(z) \end{bmatrix}. \tag{5.11.136}$$

The prime here denotes a derivative with respect to the argument,

In eq. (5.11.23) below, you will show that the Wronskian can be solved without knowing the explicit forms of the two linearly independent solutions. In particular, from eq. (5.11.10), $p_1/p_2 = (\ln(w \cdot p_2))'$

$$\mathrm{Wr}_z(\psi_1, \psi_2) = W_0 \exp\left(-\int^z (\ln w(y)p_2(y))' \mathrm{d}y\right) \tag{5.11.137}$$

$$= W_0/(w(z)p_2(z)). \tag{5.11.138}$$

This verifies the consistency of our 'junction conditions' in equations (5.11.119) and (5.11.120). If we had defined our Green's function equation to be $DG_s = \Gamma(z, z')\delta(z - z')$ instead, all the $1/(w \cdot p_2)$ in the ensuing steps would be replaced with $\Gamma(z = z')/(w \cdot p_2)$. But the Wronskian would not be proportional to it.

Let us summarize the construction of the Green's function as the inverse of a Hermitian 2nd order differential operator D_z in eq. (5.11.9), defined on $a \leq z \leq b$, with Dirichlet boundary conditions.

- Obtain two linearly independent solutions to $D_z\psi_{1,2}(z) = 0$ such that one vanishes at the left boundary and the other at the right: $\psi_1(z = a) = 0 = \psi_2(z = b)$.
- Compute the Wronskian $\mathrm{Wr}(\psi_1, \psi_2)$.
- The solution to G_s is then given by

$$G_s[z, z'] = (w \cdot p_2 \cdot \mathrm{Wr}(\psi_1, \psi_2))^{-1} (\Theta(z - z')\psi_2(z)\psi_1(z') + \Theta(z' - z)\psi_2(z')\psi_1(z))$$

$$\equiv \frac{\psi_1(z_<)\psi_2(z_>)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1, \psi_2)}; \tag{5.11.139}$$

where $z_<$ is the smaller of the two (z, z') while $z_>$ is the larger. Remember, the denominator in the second line is a constant by construction.

Example Let us now solve the G_s of momentum squared $D_z \equiv -(\mathrm{d}/\mathrm{d}z)^2$ but within the interval $a \leq z \leq b$. The general solution to $\psi''(z) = 0$. The solutions that vanish at $z = a$ and $z = b$ are, respectively,

$$\psi_1(z) = z - a \qquad \text{and} \qquad \psi_2(z) = z - b. \tag{5.11.140}$$

The Wronskian of this pair is

$$\mathrm{Wr}(\psi_1, \psi_2) = (z - a)(+1) - (+1)(z - b) \tag{5.11.141}$$

$$= b - a. \tag{5.11.142}$$

The Green's function is therefore

$$G_s(z, z') = \frac{-1}{b-a}\left(\Theta(z-z')(z'-a)(z-b) + \Theta(z-z')(z-a)(z'-b)\right) \tag{5.11.143}$$

$$= \frac{(z_< - a)(b - z_>)}{a-b}. \tag{5.11.144}$$

By setting $a = 0$ and $b = L > 0$, followed by recalling eq. (5.11.90), we may also deduce

$$\frac{z_<(L-z_>)}{L} = \frac{L}{2\pi^2}\Big(\mathrm{Li}_2\left(e^{-i\frac{\pi}{L}(z-z')}\right) + \mathrm{Li}_2\left(e^{i\frac{\pi}{L}(z-z')}\right) - \mathrm{Li}_2\left(e^{-i\frac{\pi}{L}(z+z')}\right) - \mathrm{Li}_2\left(e^{i\frac{\pi}{L}(z+z')}\right)\Big). \tag{5.11.145}$$

Neumann is illegal when $p_0 = 0$ Let us observe that it is in fact illegal to impose Neumann boundary conditions on the Green's function if $p_0 = 0$ in eq. (5.11.9). This can be seen from integrating its equation $DG_s = \delta/w$ on both sides with respect to $\int_a^b(\dots)w(z)\mathrm{d}z$.

$$\int_a^b \partial_z\left(w(z)p_2(z)\partial_z G_s(z,z')\right)\mathrm{d}z = 1 \tag{5.11.146}$$

$$w(b)p_2(b)\partial_z G_s(z=b,z') - w(a)p_2(a)\partial_z G_s(z=a,z') = 1 \tag{5.11.147}$$

If the eigenfunctions of D has zero derivatives at the boundaries $z = a, b$; the mode sum in eq. (5.11.81) indicates the first derivatives of G_s at the boundaries would, too, vanish and we would arrive at a $0 = 1$ contradiction. Thus, for a mathematically consistent Green's function, at least one of the boundaries cannot obey Neumann boundary conditions.

Problem 5.128. Damped Harmonic Oscillator Show that, focusing on the time interval $t \in [t_\mathrm{i}, t_\mathrm{f}]$, the hermitian Green's function of the damped harmonic oscillator in eq. (5.11.110) is

$$G_{s,\mathrm{DHO}}(t,t') = -\exp\left(-\gamma(t+t')\right)\frac{\sin[\Gamma(t_< - t_\mathrm{i})]\sin[\Gamma(t_\mathrm{f} - t_>)]}{\Gamma\sin[\Gamma(t_\mathrm{f} - t_\mathrm{i})]}, \tag{5.11.148}$$

where $\Gamma \equiv \sqrt{\Omega^2 - \gamma^2}$; and $t_<$ is the smaller of (t,t') while $t_>$ is the larger. □

Problem 5.129. Mixed Boundary Conditions Let us turn to considering Neumann or mixed boundary conditions at $z = a$ and $z = b$. Argue that, as long as $p_0 \neq 0$, eq. (5.11.139) is still the correct solution, but the $\psi_{1,2}$ now need to be subject to the following conditions:

- $G_s(z=a,z') = 0 = \partial_{z'}G_s(z,z'=b)$ and $G_s(z,z'=a) = 0 = \partial_z G_s(z=b,z')$ require $\psi_1(z=a) = 0 = \psi_2'(z=b)$;
- $\partial_z G_s(z=a,z') = 0 = G_s(z,z'=b)$ and $\partial_{z'}G_s(z,z'=a) = 0 = G_s(z=b,z')$ require $\psi_1'(z=a) = 0 = \psi_2(z=b)$;
- $\partial_z G_s(z=a,z') = 0 = \partial_{z'}G_s(z,z'=b)$ and $\partial_{z'}G_s(z,z'=a) = 0 = \partial_z G_s(z=b,z')$ require $\psi_1'(z=a) = 0 = \psi_2'(z=b)$.

□

5.11.3 Green's Functions as Signals Due To Point Sources

Let us now recognize that

$$f(a < z < b) = \int_a^b G_s(z,z')J(z')w(z')\mathrm{d}z' \tag{5.11.149}$$

is a (particular) solution to

$$D_z f(z) = J(z); \tag{5.11.150}$$

as can be checked by direct differentiation:

$$D_z \int_a^b G_s(z,z')J(z')w(z')\mathrm{d}z' = \int_a^b \frac{\delta(z-z')}{w(z')}J(z')w(z')\mathrm{d}z' \tag{5.11.151}$$

$$= J(z). \tag{5.11.152}$$

Comparison of the Green's function equation $DG = \delta/w$ with eq. (5.11.150) tells us that G can in fact be viewed as the 'signal' ψ generated by an infinitely-sharply-localized 'point source' described by the Dirac δ−function $\delta(z-z')/\sqrt{w(z)w(z')}$. To this end, the particular solution in eq. (5.11.149) may hence be interpreted as the signal ψ due to the collection of all point sources – i.e., the integral over $\int_a^b(\cdots)w(z')\mathrm{d}z'$ – weighted by the problem-specific source J. Notice, this interpretation does not rely on G being the inverse of a Hermitian operator; and would therefore hold for the retarded/advanced Green's functions we would encounter below.

For the Hermitian Green's function of eq. (5.11.139), a further insight into the nature of the inhomogeneous solution in eq. (5.11.149) may be obtained by using the explicit prescription in eq. (5.11.139). Remember that $w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)$ is constant, and therefore may be pulled out of the integration itself:

$$f(z) = \int_a^b \frac{\psi_1(z_<)\psi_2(z_>)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}J(z')w(z')\mathrm{d}z' \tag{5.11.153}$$

$$= \frac{\psi_1(z)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}\int_z^b \psi_2(z')J(z')w(z')\mathrm{d}z' + \frac{\psi_2(z)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}\int_a^z \psi_1(z')J(z')w(z')\mathrm{d}z'.$$

To simplify the interpretation further, we will suppose $J(z)$ is non-zero only within some range $z_{\mathrm{L}} \leq z \leq z_{\mathrm{R}}$. Whenever the ψ is evaluated to the left of this range,

$$f(z < z_{\mathrm{L}}) = \frac{\psi_1(z)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}\int_{z_{\mathrm{L}}}^{z_{\mathrm{R}}} \psi_2(z')J(z')w(z')\mathrm{d}z'; \tag{5.11.154}$$

and whenever it is instead evaluated to the right of J,

$$f(z > z_{\mathrm{R}}) = \frac{\psi_2(z)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}\int_{z_{\mathrm{L}}}^{z_{\mathrm{R}}} \psi_1(z')J(z')w(z')\mathrm{d}z'. \tag{5.11.155}$$

These expressions teach us, away from the source J, the solution is the homogeneous one (either ψ_1 or ψ_2) but with an amplitude that is built out of J integrated against the other linearly independent solution.

Asymptotic Boundary Conditions Now, as long as $\psi_{1,2}$ are linearly independent the

$$G(z,z') \equiv \frac{\psi_1(z_<)\psi_2(z_>)}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)} \tag{5.11.156}$$

satisfies $DG = \delta/w$ regardless of the boundary conditions they obey – i.e., G here need not be Hermitian. This tells us that, if we wish to solve $Df = J$ subject to asymptotic boundary conditions

$$f(z \to a) \propto L(z) \qquad \text{and} \qquad f(z \to b) \propto R(z); \tag{5.11.157}$$

then, as long as

$$\mathrm{Wr}[L,R] \neq 0 \tag{5.11.158}$$

– i.e., L and R are linearly independent – we may simply choose the $\psi_{1,2}$ such that

$$\psi_1(z \to a) \to L(z) \qquad \text{and} \qquad \psi_2(z \to b) \to R(z); \tag{5.11.159}$$

An example is provided by the following 'simple harmonic oscillator' equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + k^2\right) f(z) = J(z) \tag{5.11.160}$$

defined on the entire real line, so $p_2 = 1 = w$, $a = -\infty$, and $b = +\infty$. A pair of linearly independent solutions are $\{e^{\pm ikz}\}$. If we view $e^{-ikt}f(z)$ is a superposition of traveling waves, then it is reasonable to demand that the wave is left-moving as $z \to -\infty$ and right-moving as $z \to +\infty$,

$$e^{-ikt} f(z \to -\infty) \propto e^{-ik(t+z)} \qquad \text{and} \qquad e^{-ikt} f(z \to +\infty) \propto e^{-ik(t-z)}. \tag{5.11.161}$$

This prompts us to choose

$$\psi_1(z) = e^{-ikz} \qquad \text{and} \qquad \psi_2(z) = e^{ikz}. \tag{5.11.162}$$

Their Wronskian is $e^{-ikz}(e^{ikz})' - (e^{-ikz})'e^{ikz} = 2ik$. The relevant Green's function is therefore

$$G(z,z') = \frac{1}{2ik}\left(\Theta(z'-z)e^{+ik(z'-z)} + \Theta(z-z')e^{+ik(z-z')}\right) \tag{5.11.163}$$

$$= -\frac{i}{2k}\exp\left(+ik|z-z'|\right). \tag{5.11.164}$$

Let us justify why the form of the solution in eq. (5.11.139) supplemented by the boundary conditions in eq. (5.11.159) do in fact provide the (particular) solution to $Df = J$ with the asymptotic conditions in eq. (5.11.159). For fixed z', as $z \to a$, we have from equations (5.11.123), (5.11.125), and (5.11.126),

$$G = A^{11}\psi_1(z)\psi_1(z') + A^{22}\psi_2(z)\psi_2(z') + A_<{}^{12}\psi_1(z)\psi_2(z') + A_<{}^{21}\psi_2(z)\psi_1(z'). \tag{5.11.165}$$

Let us choose $\psi_{1,2}$ such that $\psi_1(z \to a) \to L(z)$ but $\psi_2(z \to a)$ does not approach $L(z)$ – they cannot *both* approach $L(z)$, lest they become linearly dependent – then we must put $A^{22} = 0 = A_<{}^{21}$. Similarly, for fixed z', as $z \to b$, we have again from equations (5.11.123), (5.11.125), and (5.11.126),

$$G = A^{11}\psi_1(z)\psi_1(z') + A_>{}^{12}\psi_1(z)\psi_2(z') + A_>{}^{21}\psi_2(z)\psi_1(z'). \tag{5.11.166}$$

If we choose $\psi_{1,2}$ such that $\psi_2(z \to b) \to R(z)$ while $\psi_1(z \to b)$ does not tend to $R(z)$, then we must put $A^{11} = 0 = A_>{}^{12}$. At this point, eq. (5.11.127) tells us $A_<{}^{12} = A_>{}^{21} \equiv A$ and $G = A\psi_1(z_<)\psi_2(z_>)$. The jump condition of eq. (5.11.119) will then set $A = (w \cdot p_2 \cdot \mathrm{Wr}_z(\psi_1, \psi_2))^{-1}$.

5.11.4 Retarded Green's Functions and Distributional Calculus

We now turn to a different perspective on solving the Green's function equation

$$D_z G^+[z, z'] = \frac{\delta(z - z')}{\sqrt{w(z)w(z')}} = D_{z'} G^+[z, z'] \tag{5.11.167}$$

– *without* assuming D_z in eq. (5.11.9) is Hermitian nor Dirichlet boundary conditions for its homogeneous solutions.

We shall employ the 'retarded' ansatz, defined by the condition $G^+(z < z') = 0$:

$$G^+[z, z'] = \Theta(z - z')\mathcal{G}(z, z'). \tag{5.11.168}$$

The physical interpretation of the inhomogeneous solution to $Df = J$ built out of such a G^+,

$$f(a < z < b) = \int_a^b G^+(z, z')J(z')w(z')\mathrm{d}z', \tag{5.11.169}$$

is that *cause precedes effect* – or, simply, causality is obeyed – if z and z' refers to time. For, if we recall the interpretation that $G^+(z, z')J(z')$ is the signal at z emitted from z', with the integral describing the sum over all relevant point-sources in the problem, we see that such a signal would be zero if $z < z'$; i.e., when signal receipt time occurs *before* the emission time.

The first and second derivatives of G^+ with respect to z and z' are

$$\partial_z G^+[z, z'] = \delta(z - z')\mathcal{G}(z, z') + \Theta(z - z')\partial_z \mathcal{G}(z, z'), \tag{5.11.170}$$
$$\partial_{z'} G^+[z, z'] = -\delta(z - z')\mathcal{G}(z, z') + \Theta(z - z')\partial_{z'} \mathcal{G}(z, z'), \tag{5.11.171}$$
$$\partial_z^2 G^+[z, z'] = \delta'(z - z')\mathcal{G}(z, z') + \delta(z - z')2\partial_z \mathcal{G}(z, z') + \Theta(z - z')\partial_z^2 \mathcal{G}(z, z'), \tag{5.11.172}$$
$$\partial_{z'}^2 G^+[z, z'] = \delta'(z - z')\mathcal{G}(z, z') - \delta(z - z')2\partial_{z'} \mathcal{G}(z, z') + \Theta(z - z')\partial_{z'}^2 \mathcal{G}(z, z'). \tag{5.11.173}$$

For G^+ to satisfy the Green's function equation $DG^+ = \delta/w$ we must therefore have

$$\begin{aligned} \delta'(z - z') \cdot p_2(z) \cdot \mathcal{G} + \delta(z - z')\left(2p_2(z) \cdot \partial_z \mathcal{G} + (w \cdot p_2)'(z)w(z)^{-1} \cdot \mathcal{G}\right) \\ + \Theta[t - t']D_z \mathcal{G} = \frac{\delta(z - z')}{\sqrt{w(z)w(z')}}, \end{aligned} \tag{5.11.174}$$

$$\delta'(z - z') \cdot p_2(z') \cdot \mathcal{G} - \delta(z - z')\left(2p_2(z') \cdot \partial_{z'} \mathcal{G} + (w \cdot p_2)'(z')w(z')^{-1} \cdot \mathcal{G}\right)$$

$$+\Theta[t-t']D_{z'}\mathcal{G}=\frac{\delta(z-z')}{\sqrt{w(z)w(z')}}. \tag{5.11.175}$$

Now, since $\delta(z)$ is non-trivial only when $z=0$, that means we may deduce the distributional identity $z\delta(z)=0\cdot\delta(z)=0$. Differentiating it once,

$$\delta(z)=-z\delta'(z). \tag{5.11.176}$$

By multiplying both sides by z^{n-1}, we see that

$$z^n\delta'(z)=0 \tag{5.11.177}$$

for all higher powers $n=2,3,4,\dots$. Therefore, we may Taylor expand

$$\begin{aligned}\delta'(z)g(z)&=\delta'(z)\left(g(0)+g'(0)\cdot z+\mathcal{O}(z^2)\right) &(5.11.178)\\ &=\delta'(z)\cdot g(0)-\delta(z)\cdot g'(0). &(5.11.179)\end{aligned}$$

At this point, our Green's function equations become

$$\begin{aligned}&\delta'(z-z')\cdot(p_2\cdot\mathcal{G})_{z=z'}+\delta(z-z')\left\{(p_2\cdot\partial_z\mathcal{G})_{z=z'}+((w\cdot p_2)'w^{-1}\mathcal{G})_{z=z'}\right\}\\ &\qquad+\Theta(z-z')D_z\mathcal{G}(z,z')=\delta(z-z')(w(z)w(z'))^{-\frac{1}{2}}, &(5.11.180)\\ &\delta'(z-z')\cdot(p_2\cdot\mathcal{G})_{z=z'}-\delta(z-z')\left\{(p_2\cdot\partial_z\mathcal{G})_{z=z'}+((w\cdot p_2)'w^{-1}\mathcal{G})_{z=z'}\right\}\\ &\qquad+\Theta(z-z')D_{z'}\mathcal{G}(z,z')=\delta(z-z')(w(z)w(z'))^{-\frac{1}{2}}. &(5.11.181)\end{aligned}$$

Since there is nothing special about the location $z=z'$, in order to recover eq. (5.11.167), we must have the coefficients of $\delta'(z-z')$ and $\Theta(z-z')$ vanish and that of the $\delta(z-z')$ equal to $1/w(z)=1/w(z')=1/\sqrt{w(z)w(z')}$. The δ' term implies the $\mathcal{G}$ must be zero at coincidence, since p_2 cannot be zero everywhere:

$$\mathcal{G}(z=z')=0. \tag{5.11.182}$$

The $\mathcal{G}$ term multiplying δ is thus set to zero, provided $(w\cdot p_2)'/w$ is finite; whereas the remaining δ term now informs us

$$\partial_z\mathcal{G}(z=z')=(w(z=z')p_2(z=z'))^{-1}=-\partial_{z'}\mathcal{G}(z=z'). \tag{5.11.183}$$

The Θ term then teaches us $\mathcal{G}$ must be a homogeneous solution to the ODE $D\psi=0$:

$$D_z\mathcal{G}(z,z')=0=D_{z'}\mathcal{G}(z,z'). \tag{5.11.184}$$

It is worth highlighting, from equations (5.11.182) and (5.11.182), because $\mathcal{G}$ vanishes at $z=z'$, $G^+(z,z')$ itself must be continuous at $z=z'$, since one side of the step function yields zero; while the other side is unity. It is in fact the first derivative of G that is necessarily discontinuous, since $\partial_z G^+(z=z'+0^+)=\partial_z\mathcal{G}(z=z'+0^+)$ whereas $\partial_z G^+(z=z'-0^+)=0$; so that the 'jump' in the first derivative across $z=z'$ is $1/(w\cdot p_2)$. In fact, this is nothing but equations (5.11.119) and (5.11.120), which was derived using only the equation $DG=\delta/w$, without assuming G was Hermitian, retarded, etc.

Let us now argue that the solution to the retarded G is

$$G^+(z,z') = \Theta(z-z')\mathcal{G}(z,z'), \tag{5.11.185}$$

$$\mathcal{G}(z,z') = \frac{\psi_1(z')\psi_2(z) - \psi_1(z)\psi_2(z')}{w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)}; \tag{5.11.186}$$

where the $\psi_{1,2}$ are arbitrary but linearly independent solutions to $D\psi_{1,2} = 0$. (As discussed earlier, the combination $w \cdot p_2 \cdot \mathrm{Wr}(\psi_1,\psi_2)$ is constant.) In Problem (5.135) below, you will show that this solution is, in fact, independent of the choice of the 2D basis (ψ_1,ψ_2). Hence, the solution to the retarded Green's function is *unique*.

Firstly, from eq. (5.11.184), we may infer

$$\mathcal{G}(z,z') = A^{\mathrm{IJ}}\psi_{\mathrm{I}}(z)\psi_{\mathrm{J}}(z') \tag{5.11.187}$$

for constants $\{A^{\mathrm{IJ}}\}$. Eq. (5.11.182) then translates to

$$A^{11}\psi_1(z)^2 + A^{22}\psi_2(z)^2 + (A^{12}+A^{21})\psi_1(z)\psi_2(z) = 0. \tag{5.11.188}$$

Since this has to be true for all $a \leq z \leq b$, we must have $A^{11} = 0 = A^{22}$ and $A^{12} = -A^{21} \equiv A$.

$$\mathcal{G}(z,z') = A(\psi_2(z)\psi_1(z') - \psi_2(z')\psi_1(z)) \tag{5.11.189}$$

Imposing eq. (5.11.183),

$$\partial_z\mathcal{G}(z=z') = A \cdot \mathrm{Wr}_z(\psi_1,\psi_2) = \frac{1}{w(z)p_2(z)}. \tag{5.11.190}$$

We have already seen that the Wronskian *is* necessarily proportional to $1/(wp_2)$. This proves our assertion.

Example: SHO For the simple harmonic oscillator, we may take the two linearly independent solutions to be $\psi_1(t) = \cos(\Omega t)$ and $\psi_2(t) = \sin(\Omega t)$. Their Wronskian is

$$\mathrm{Wr}(\psi_1,\psi_2) = \Omega\cos(\Omega t)^2 + \Omega\sin(\Omega t)^2 = \Omega. \tag{5.11.191}$$

We must therefore have

$$\mathcal{G}(t,t') = \frac{\cos(\Omega t')\sin(\Omega t) - \cos(\Omega t)\sin(\Omega t')}{\Omega} \tag{5.11.192}$$

$$= \frac{\sin\left(\Omega(t-t')\right)}{\Omega}; \tag{5.11.193}$$

$$G^+[t-t'] = \Theta(t-t')\frac{\sin[\Omega(t-t')]}{\Omega}. \tag{5.11.194}$$

In §(6) you will find this 'retarded' Green's function by a direct evaluation of

$$G^+(t-t') = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\frac{e^{-ik(t-t')}}{\Omega^2 - k^2} \tag{5.11.195}$$

by choosing the right contour on the complex $k-$plane. Even though this is not the Hermitian Green's function G_s, note the similarity with the mode expansion of eq. (5.11.79) since $(\mathrm{d}/\mathrm{d}t)^2 + \Omega^2$ acting on $\exp(-ikt)$ yields $\Omega^2 - k^2$.

Problem 5.130. DHO Retarded Green's Function Show that the retarded Green's function of the DHO, defined in eq. (5.11.110), is

$$G^{+}(t,t') = \Theta(t-t')\exp\left(-\gamma(t+t')\right)\frac{\sin\left(\Gamma(t-t')\right)}{\Gamma}; \tag{5.11.196}$$

where $\Gamma \equiv \sqrt{\Omega^2-\gamma^2}$. $\square$

Problem 5.131. Wronskian of Independent Functions Show that the Wronskian of

$$f_1(x) \equiv x^2 \qquad \text{and} \qquad f_2(x) = |x|\cdot x \tag{5.11.197}$$

is zero. (Notice this fact is independent of the differential equation they may satisfy.) This shows that, while linear dependence implies zero Wronskian, zero Wronskian does not imply linear dependence. This is the case here for any finite neighborhood of $x=0$. Hint: To use distributional calculus, first explain why

$$|x| = \Theta(x)x - \Theta(-x)x. \tag{5.11.198}$$

Next show that $f_{1,2}(x)$ both satisfy the second order ODE $xf''_{1,2}(x) - f'_{1,2}(x) = 0$. $\square$

Problem 5.132. Check via a direct calculation that eq. (5.11.156) does in fact satisfy $DG = \delta/w$ regardless of the boundary conditions satisfied by the $\psi_{1,2}$. $\square$

Problem 5.133. 'Advanced' Green's Function Consider instead the alternate ansatz

$$G^{-}[z,z'] = \Theta[z'-z]\mathcal{G}^{-}[z,z']. \tag{5.11.199}$$

Show that

$$D_z\mathcal{G}^{-}[z,z'] = 0 = D_{z'}\mathcal{G}^{-}[z,z'], \tag{5.11.200}$$

$$\mathcal{G}^{-}[z=z'] = 0, \tag{5.11.201}$$

$$\partial_z\mathcal{G}^{-}[z=z'] = -\frac{1}{w(z=z')p_2(z=z')} = -\partial_{z'}\mathcal{G}^{-}[z=z']. \tag{5.11.202}$$

Then argue that the solution can be built out of the two linearly independent homogeneous solutions to $D\psi_{1,2}=0$ via the prescription

$$\mathcal{G}^{-}(z,z') = \frac{\psi_1(z)\psi_2(z') - \psi_1(z')\psi_2(z)}{w\cdot p_2\cdot \mathrm{Wr}(\psi_1,\psi_2)}. \tag{5.11.203}$$

(It is not a coincidence, this $\mathcal{G}^{-}$ is the negative of that in eq. (5.11.186).) Moreover, verify that

$$G^{+}(z',z) = G^{-}(z,z'). \tag{5.11.204}$$

That is, G^{+} can be obtained from G^{-}, and vice versa. $\square$

Problem 5.134. Green's Functions of the Second Derivative Show that

$$G^{\pm}(z,z') \equiv \pm\Theta(\pm(z-z'))(z-z') \tag{5.11.205}$$

are the retarded $(+)$ and advanced $(-)$ Green's functions of $-(\mathrm{d}/\mathrm{d}z)^2$. Why does this imply eq. (5.11.89) is, also, a Green's function? Hint: Refer to eq. (5.11.198). $\square$

Problem 5.135. Basis Independence of $\psi_{1,2}$ You may be puzzled by why no boundary conditions appeared necessary for the homogeneous solutions $\psi_{1,2}$ occurring within the 'retarded' and 'advanced' Green's functions.

Prove that equations (5.11.186) and (5.11.203) are, in fact, independent of the choice of basis (ψ_1, ψ_2). Under the transformation $\psi_{\rm I} = M_{\rm I}{}^{\rm J}\varphi_{\rm J}$ for any invertible 2×2 matrix M, show that, for e.g.,

$$\mathcal{G}(z, z') = \frac{\psi_1(z')\psi_2(z) - \psi_1(z)\psi_2(z')}{w \cdot p_2 \cdot \text{Wr}(\psi_1, \psi_2)} = \frac{\varphi_1(z')\varphi_2(z) - \varphi_1(z)\varphi_2(z')}{w \cdot p_2 \cdot \text{Wr}(\varphi_1, \varphi_2)}. \tag{5.11.206}$$

In other words, since the solution for $\mathcal{G}$ is basis independent, there is no need to impose boundary conditions for the $\psi_{1,2}$. This in turn implies the retarded and advanced Green's functions are unique.

Hint: You may find it useful to first recognize

$$\mathcal{G}(z, z') = -\frac{\epsilon^{\rm AB}\psi_{\rm A}(z)\psi_{\rm B}(z')}{w(z'')p_2(z'')\epsilon^{\rm CE}\psi_{\rm C}(z'')\psi'_{\rm E}(z'')}; \tag{5.11.207}$$

where $\epsilon^{\rm AB}$ is the Levi-Civita symbol with $\epsilon^{12} \equiv 1$. □

Problem 5.136. Distributional Calculus Near $z \approx z_0$, let us write an arbitrary function $f(z)$ as

$$f(z) = \Theta(z_0 - z)f_<(z) + \Theta(z - z_0)f_>(z); \tag{5.11.208}$$

i.e., split it into two portions: $f_<(z)$ describes its behavior for $z < z_0$ and $f_>(z)$ for $z > z_0$. (If f is smooth, we will simply have $f(z) = (\Theta(z-z_0)+\Theta(z_0-z))f(z)$, where $\Theta(z-z_0)+\Theta(z_0-z) = 1$.) Use the distributional calculus we developed above to demonstrate

$$f'(z) = \delta(z - z_0)\left(f_>(z_0) - f_<(z_0)\right) + \Theta(z_0 - z)f'_<(z) + \Theta(z - z_0)f'_>(z); \tag{5.11.209}$$

and

$$\begin{aligned} f''(z) = \delta'(z - z_0)\left(f_>(z_0) - f_<(z_0)\right) + \delta(z - z_0)\left(f'_>(z_0) - f'_<(z_0)\right) \\ + \Theta(z_0 - z)f''_<(z) + \Theta(z - z_0)f''_>(z). \end{aligned} \tag{5.11.210}$$

This result teaches us, for instance: the homogeneous solution of a 2nd order linear ODE and its first derivative must be continuous. □

5.11.5 Boundary versus Initial/Final Value Problems

YZ: Check notation is consistent; f vs ψ. We now turn to some basic application of Green's functions: the solution of ODEs. For $f(z)$ sourced by some given $J(z)$:

$$D_z f(z) = J(z); \tag{5.11.211}$$

we wish to solve f subject to one of the following.

- Boundary conditions: $f(z = a)$ and $f(z = b)$ given.

- 'Initial' or 'final' conditions: either $f(z=a)$ and $f'(z=a)$; or $f(z=b)$ and $f'(z=b)$ given.

For the former, we shall see that G_s with Dirichlet boundary conditions can be used to obtain the solution; while for the initial value problem it is the G^+; and for the final value problem it is the G^- that will prove useful.

To begin we consider the following integral, where G can be either one of the three Green's functions.

$$I(a<z<b) \equiv \int_a^b \left(D_{z'}G(z,z')\cdot f(z') - G(z,z')\cdot D_{z'}f(z')\right) w(z')\mathrm{d}z' \tag{5.11.212}$$

On the one hand, we may simply employ $DG = \delta/w$ and eq. (5.11.211).

$$\begin{aligned} I(z\in(a,b)) &= \int_a^b \left(\delta(z-z')f(z') - G(z,z')J(z')w(z')\right)\mathrm{d}z' && (5.11.213)\\ &= f(z) - \int_a^b G(z,z')J(z')w(z')\mathrm{d}z'. && (5.11.214)\end{aligned}$$

(We have excluded the end points a and b from the range of z above, to ensure $\int_a^b \delta(z-z')F(z')\mathrm{d}z' = F(z)$.) On the other hand, we may integrate-by-parts,

$$\begin{aligned} I(z) = &\left[w(z')p_2(z')\left(f(z')\partial_{z'}G(z,z') - G(z,z')\partial_{z'}f(z')\right)\right]_{z'=a}^{z'=b} \\ &- \int_a^b w(z')p_2(z')\left(\partial_{z'}G(z,z')\partial_{z'}f(z') - \partial_{z'}G(z,z')\partial_{z'}f(z')\right)\mathrm{d}z'. \end{aligned} \tag{5.11.215}$$

The p_0 terms cancel right away; whereas we see the p_2 terms cancel after integration-by-parts. We may now equate the two expressions for $I(z)$ to arrive at

$$\begin{aligned} f(a<z<b) = &\int_{a^-}^{b^+} G(z,z')J(z')w(z')\mathrm{d}z' \\ &+ \left[w(z')p_2(z')\left\{f(z')\partial_{z'}G(z,z') - G(z,z')\partial_{z'}f(z')\right\}\right]_{z'=a}^{z'=b}. \end{aligned} \tag{5.11.216}$$

Boundary Value Solution Now, if $f(z=a)$ and $f(z=b)$ are specified, then choosing the hermitian $G_s(z,z')$ obeying Dirichlet boundary conditions would render the $G_s(z,z'=a) = 0 = G_s(z,z'=b)$ and reduce eq. (5.11.216) to

$$\begin{aligned} f(a\leq z\leq b) &= \int_a^b G_s(z,z')J(z')w(z')\mathrm{d}z' + f_{\mathrm{h}}(z), \\ f_{\mathrm{h}}(z) &\equiv w(b)p_2(b)f(b)\partial_{z'}G_s(z,z'=b) - w(a)p_2(a)f(a)\partial_{z'}G_s(z,z'=a). \end{aligned} \tag{5.11.217}$$

The integral term on the right hand side of the first line may be interpreted as the inhomogeneous (source-dependent) solution; while the $f_{\mathrm{h}}(z)$ may be interpreted as the homogeneous solution. Notice, too, we have extended the original range of validity $a<z<b$ to include the end points. Altogether, these lead us to the following problem.

Problem 5.137. Homogeneous solution with Boundary Conditions Show that the homogeneous solution portion of eq. (5.11.217) can be written as

$$f_{\mathrm{h}}(z) = f(a)\frac{\psi_2(z)}{\psi_2(a)} + f(b)\frac{\psi_1(z)}{\psi_1(b)}. \tag{5.11.218}$$

Explain why this observation, together with the Dirichlet boundary conditions obeyed by G_s, allows one to extend the range of validity of $f(z)$ from $z \in (a, b)$ to include the end points; i.e., to $z \in [a, b]$.

Hint: For the first and second terms of eq. (5.11.218), evaluate the Wronskian in G_s at a and b respectively. $\square$

Problem 5.138. SHO: Initial and Final Positions If the initial position $\vec{x}(t_1) = \vec{x}_1$ and final position $\vec{x}(t_2) \equiv \vec{x}_2$ of a SHO, obeying $\ddot{x} + \omega^2 \vec{x} = 0$, are given, show that its trajectory is given by

$$\vec{x}(t) = \frac{\vec{x}_2 \sin(\omega(t - t_1)) + \vec{x}_1 \sin(\omega(t_2 - t))}{\sin(\omega(t_2 - t_1))}. \tag{5.11.219}$$

Use the Green's function method; but also check explicitly that both the SHO equations and the relevant boundary conditions are obeyed. $\square$

Initial Value Solution If $f(z = a)$ and $f'(z = a)$ are specified instead, choosing G^+ in equations (5.11.185) and (5.11.186) would render the $G^+(z, z' = b) = 0 = \partial_{z'} G^+(z, z' = b)$ because of the step function in $G^+(z, z' = b) \propto \Theta(z - b) = 0$, for all $z < b$.

$$\begin{aligned} f(a \leq z \leq b) = &\int_a^z \mathcal{G}(z, z') J(z') w(z') \mathrm{d}z' \\ &- w(a) p_2(a) \left\{ f(a) \partial_{z'} \mathcal{G}(z, z' = a) - \mathcal{G}(z, z' = a) f'(a) \right\}. \end{aligned} \tag{5.11.220}$$

We have discarded the step function within G^+, so that with $G^+(z, z')$ is simply replaced with $\mathcal{G}(z, z') = \psi_{[1}(z')\psi_{2]}(z)/(w \cdot p_2 \cdot \mathrm{Wr}(\psi_1, \psi_2))$ because $\Theta(z - a) = 1$ within the range of z in question. Moreover, note that $\partial_{z'} G^+(z, z' = a) = \Theta(z - a) \partial_{z'} \mathcal{G}(z, z' = a)$. There is no need to differentiate the step function because $\mathcal{G}(z = z') = 0$, for

$$\partial_a \Theta(z - a) \mathcal{G}(z, z' = a) = \delta(z - a) \mathcal{G}(z = a = z') = 0. \tag{5.11.221}$$

Upon making these replacements $G^+ \to \mathcal{G}$, notice eq. (5.11.220) is now valid across the entire interval $a \leq z \leq b$, including the end points.

Once again, the term on the right hand side of the first line in eq. (5.11.220) may be interpreted as the inhomogeneous (source-dependent) solution; while those in the second may be interpreted as the homogeneous solution (because $\mathcal{G}$ is, in fact, one).

Problem 5.139. SHO: Initial Position and Velocity If the initial position $\vec{x}(t') = \vec{x}_0$ and initial velocity $\dot{\vec{x}}(t') = \vec{v}_0$ of a SHO, obeying $\ddot{x} + \omega^2 \vec{x} = 0$, are given, show that its trajectory is given by

$$\vec{x}(t) = \vec{x}_0 \cdot \cos(\omega(t - t')) + \frac{\vec{v}_0}{\omega} \cdot \sin(\omega(t - t')). \tag{5.11.222}$$

Use the Green's function method; but also check explicitly that both the SHO equations and the relevant initial conditions are obeyed. $\square$

Problem 5.140. Damped Harmonic Oscillator For the damped harmonic oscillator defined in eq. (5.11.110), show that eq. (5.11.220) reduces to – with f replaced with x – the following expression:

$$
\begin{aligned}
x(t \geq a) = &\int_{a^-}^{\infty} \Theta(t-t')\mathcal{G}_{\text{DHO}}(t,t')J(t')\text{d}t' \\
&+ \frac{\text{d}x(t'=a)}{\text{d}t'}\mathcal{G}_{\text{DHO}}(t,t'=a) + x(a)\left(2\gamma\mathcal{G}_{\text{DHO}}(t,t'=a) - \partial_{t'}\mathcal{G}_{\text{DHO}}(t,t'=a)\right); \qquad (5.11.223)
\end{aligned}
$$

where the homogeneous solution portion of the effective Green's function now takes a time-translation invariant form:

$$
\mathcal{G}_{\text{DHO}}(t,t') = e^{-\gamma(t-t')}\frac{\sin\left(\Gamma(t-t')\right)}{\Gamma}, \qquad (5.11.224)
$$

with $\Gamma \equiv \sqrt{\Omega^2-\gamma^2}$. The effective Green's function itself is

$$
G^+_{\text{DHO}}(t,t') = \Theta(t-t')\mathcal{G}_{\text{DHO}}(t,t'). \qquad (5.11.225)
$$

Contrast G^+_{DHO} against the Green's function in eq. (5.11.196); we shall also obtain the homogeneous solution portion of this result again in eq. (7.8.57) below, albeit using slightly different arguments. □

SHO Solutions: Divergences and Caustics Notice, the solution to the SHO "initial value problem" in eq. (5.11.222) is well defined for *any* initial and final times t' and t; whereas, the solution to the "boundary value problem" in eq. (5.11.219) is *not* well defined when the time elapsed is an integer multiple of the half-period: $t - t' = \pi/\omega, 2\pi/\omega, 3\pi/\omega, \dots \equiv n\pi/\omega$. In fact, at half-integer periods, eq. (5.11.222) tells us the final position does not depend on the initial velocity, but only on the initial position:

$$
\vec{x}\left(t = t' + n\pi/\omega\right) = (-)^n\vec{x}_0. \qquad (5.11.226)
$$

When n is even (full periods) the SHO arrives back at the initial position $\vec{x}_0$; and when n is odd (half-integer periods) it is at the opposite side of the origin at $-\vec{x}_0$. (This is true even for the *quantum* SHO, because of parity invariance.) This is why the boundary value solution in eq. (5.11.219) breaks down, because at half-integer periods, the end points cannot be arbitrary positions, but rather must be equal or negative to each other. Moreover, since this statement holds for *any* initial velocity $\vec{v}_0$, this indicates there are *infinite* number of solutions that satisfy the boundary value problem defined by $\vec{x}(t = t') = \vec{x}_0$ and $\vec{x}(t - t' = n\pi/\omega) = (-)^n\vec{x}_0$ for fixed integer $n \in \mathbb{Z}^+$, In other words, the actual solution (as opposed to the one in eq. (5.11.219)) to the boundary value problem at hand is no longer unique.

Some jargon: Starting its motion from $(t', \vec{x}_0)$, the SHO at the locations

$$
\left(t = t' + n\pi/\omega, \vec{x}(t) = (-)^n\vec{x}_0\right) \qquad (5.11.227)
$$

– where more than one solution (actually, for the case at hand, an infinity of them) converge – are said to form *caustics*. Moreover, the pair $(t', \vec{x}_0)$ and $(t' + n\pi/\omega, (-)^n\vec{x}_0)$, for some fixed n, are called *conjugate points*.

Problem 5.141. Boundary Value Solution from 'Retarded' Green's Function It is actually possible to re-phrase the boundary value problem solution in eq. (5.11.217) solely in terms of the 'retarded' Green's function of eq. (5.11.185), or at least its homogeneous portion $\mathcal{G}$.

First show that the Dirichlet Green's function can be expressed as

$$G_s(z, z') = \frac{\mathcal{G}(z_<, a) \cdot \mathcal{G}(z_>, b)}{\mathcal{G}(b, a)}. \tag{5.11.228}$$

Then verify by a direct calculation, the homogeneous portion of the boundary value problem is

$$f_{\rm h}(z) = \frac{\mathcal{G}(z, a)}{\mathcal{G}(b, a)} \cdot f(b) + \frac{\mathcal{G}(z, b)}{\mathcal{G}(a, b)} \cdot f(a). \tag{5.11.229}$$

That is, check that $Df_{\rm h} = 0$ and $f(a)$ and $f(b)$ are recovered upon setting z to a and b respectively. Since $\mathcal{G}$ is the homogeneous solution portion of the retarded Green's function, there is a direct relation of this result to eq. (5.11.218). $\square$

Final Value Solution If $f(z = b)$ and $f'(z = b)$ are specified, then choosing G^- in eq. (5.11.199) would render the $G^-(z, z' = a) = 0 = \partial_{z'} G^-(z, z' = a)$ because of the step function in $G^-(z, z' = a) \propto \Theta(a - z) = 0$, for all $z > a$. Moreover, one may check that the following is a solution across the entire interval $z \in [a, b]$, including the boundaries:

$$\begin{aligned} f(a \le z \le b) = &\int_z^b \mathcal{G}^-(z, z') J(z') w(z') \mathrm{d}z' \\ &+ w(b) p_2(b) \left\{ f(b) \partial_{z'} \mathcal{G}^-(z, z' = b) - \mathcal{G}^-(z, z' = b) f'(b) \right\}. \end{aligned} \tag{5.11.230}$$

Like its previous two counterparts, the term on the right hand side of the first line in eq. (5.11.230) may be interpreted as the inhomogeneous (source-dependent) solution; while those in the second may be interpreted as the homogeneous solution.

Problem 5.142. Explain why the limits of integration in equations (5.11.220) and (5.11.230) depend on z. Hints: If we allow the limits to depend on z, the relevant step functions may be dropped – as I have done – otherwise if the limits run from a to b, then the step functions need to be kept.

Then, verify directly that the right hand sides of equations (5.11.217), (5.11.220), and (5.11.230) satisfies $Df = J$; and, moreover, also obey the relevant boundary, initial, or final conditions. $\square$

5.12 *Nested commutators and $e^{tA(x)}(\partial_x e^{-tA(x)} |\psi(x)\rangle)$

Let t and x be continuous variables; $A(x)$ an arbitrary operator; and $|\psi(x)\rangle$ some arbitrary state. In this section we shall prove the following formula:

$$\begin{aligned} &e^{t \cdot A(x)} \partial_x \left(e^{-t \cdot A(x)} |\psi(x)\rangle \right) \\ &= \left(\partial_x - t \cdot \partial_x A + \frac{t^2}{2!} [\partial_x A, A] + \frac{(-t)^3}{3!} [[\partial_x A, A], A] + \dots \right) |\psi(x)\rangle \end{aligned} \tag{5.12.1}$$

$$= \left(\partial_x - t \cdot \partial_x A + \sum_{\ell=2}^{\infty} \frac{(-t)^\ell}{\ell!}[\dots[\partial_x A, \underbrace{A], \dots], A]}_{\ell-1 \text{ As}}\right) |\psi(x)\rangle . \tag{5.12.2}$$

Proof We proceed by Taylor expansion in powers of t.

$$e^{t\cdot A(x)}\partial_x \left(e^{-t\cdot A(x)} |\psi(x)\rangle\right) = \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \frac{\mathrm{d}^\ell}{\mathrm{d}t^\ell} \left(e^{t\cdot A(x)}\partial_x \left(e^{-t\cdot A(x)} |\psi(x)\rangle\right)\right)\Big|_{t=0} \tag{5.12.3}$$

$$= \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \sum_{n=0}^{\ell} \binom{\ell}{n} A^n \partial_x \left((-A)^{\ell-n} |\psi\rangle\right) . \tag{5.12.4}$$

We have exploited the formula

$$\frac{\mathrm{d}^\ell}{\mathrm{d}t^\ell} (\alpha(t) \cdot \beta(t)) = \sum_{n=0}^{\ell} \binom{\ell}{n} \frac{\mathrm{d}^n \alpha}{\mathrm{d}t^n} \frac{\mathrm{d}^{\ell-n}\beta}{\mathrm{d}t^{\ell-n}}, \tag{5.12.5}$$

where α and β are arbitrary functions of t. (Can you prove this?) Let us work out the summation multiplying the $t^\ell/\ell!$ term. For the $\ell = 0$ case, we have

$$\binom{0}{0} A^0 \partial_x \left((-A)^0 |\psi\rangle\right) = \partial_x |\psi\rangle ; \tag{5.12.6}$$

the $\ell = 1$ case reads

$$\binom{1}{0} A^1 \partial_x \left((-A)^0 |\psi\rangle\right) + \binom{1}{1} A^0 \partial_x \left((-A)^1 |\psi\rangle\right) \tag{5.12.7}$$

$$= (A\partial_x - \partial_x A - A\partial_x) |\psi\rangle = -(\partial_x A) |\psi\rangle ; \tag{5.12.8}$$

and the $\ell = 2$ case becomes

$$\binom{2}{0} A^0 \partial_x \left((-A)^2 |\psi\rangle\right) + \binom{2}{1} A\partial_x ((-A) |\psi\rangle) + \binom{2}{2} A^2 \partial_x \left((-A)^0 |\psi\rangle\right) \tag{5.12.9}$$

$$= ((\partial_x A)A + A(\partial_x A) + A^2\partial_x - 2A(\partial_x A) - 2A^2\partial_x + A^2\partial_x) |\psi\rangle \tag{5.12.10}$$

$$= (-)^2[\partial_x A, A] |\psi\rangle . \tag{5.12.11}$$

We are done once we can prove that, for $\ell \geq 2$,

$$\sum_{n=0}^{\ell} \binom{\ell}{n} A^n \partial_x \left((-A)^{\ell-n} |\psi\rangle\right) = (-)^\ell[\dots[\partial_x A, A], A, \dots, A] |\psi\rangle , \tag{5.12.12}$$

with the right hand side containing $\ell - 1$ (nested) commutators. Suppose the $(n \geq 2)-$th case holds. Let us consider

$$\sum_{n=0}^{\ell} \binom{\ell}{n} A^n \partial_x \left((-A)^{\ell-n} \cdot A |\psi\rangle\right) - A \sum_{n=0}^{\ell} \binom{\ell}{n} A^n \partial_x \left((-A)^{\ell-n} |\psi\rangle\right) \tag{5.12.13}$$

$$= (-)^n[\dots[\partial_x A, \underbrace{A], A, \dots, A}_{\ell-1}](A\,|\psi\rangle) - (-)^n A[\dots[\partial_x A, \underbrace{A], A, \dots, A}_{\ell-1}]\,|\psi\rangle \tag{5.12.14}$$

$$= (-)^n[\dots[\partial_x A, \underbrace{A], A, \dots, A}_{\ell}]\,|\psi\rangle\,. \tag{5.12.15}$$

On the other hand,

$$\sum_{n=0}^{\ell}\binom{\ell}{n}A^n\partial_x\left((-A)^{\ell-n}\cdot A\,|\psi\rangle\right) - A\sum_{n=0}^{\ell}\binom{\ell}{n}A^n\partial_x\left((-A)^{\ell-n}\,|\psi\rangle\right) \tag{5.12.16}$$

$$= -\sum_{n=0}^{\ell}\binom{\ell}{n}A^n\partial_x\left((-A)^{\ell+1-n}\,|\psi\rangle\right) - \sum_{n=1}^{\ell+1}\binom{\ell}{n-1}A^n\partial_x\left((-A)^{\ell+1-n}\,|\psi\rangle\right), \tag{5.12.17}$$

Let us note that $\binom{\ell}{0} = \binom{\ell+1}{0} = \binom{\ell}{\ell} = \binom{\ell+1}{\ell+1} = 1$. Therefore, in the second line, the zeroth term of the first sum from the left is simply $\binom{\ell+1}{0}\partial_x((-A)^{\ell+1}\,|\psi\rangle)$ and the $(\ell+1)-$th term of the second sum is simply $\binom{\ell+1}{\ell+1}A^{\ell+1}\partial_x\,|\psi\rangle$. Furthermore, we may combine the first through ℓ terms of both summations by recognizing

$$\binom{\ell}{n} + \binom{\ell}{n-1} = \frac{\ell!}{n!(\ell-n)!} + \frac{\ell!}{(n-1)!(\ell+1-n)!} \tag{5.12.18}$$

$$= \frac{\ell!}{n!(\ell+1-n)!}((\ell+1-n)+(n)) = \binom{\ell+1}{n}. \tag{5.12.19}$$

Altogether,

$$\sum_{n=0}^{\ell}\binom{\ell}{n}A^n\partial_x\left((-A)^{\ell-n}\cdot A\,|\psi\rangle\right) - A\sum_{n=0}^{\ell}\binom{\ell}{n}A^n\partial_x\left((-A)^{\ell-n}\,|\psi\rangle\right) \tag{5.12.20}$$

$$= -\sum_{n=0}^{\ell+1}\binom{\ell+1}{n}A^n\partial_x\left((-A)^{\ell+1-n}\,|\psi\rangle\right). \tag{5.12.21}$$

This completes the proof. $\square$

6 Calculus on the Complex Plane

In this Chapter, we are going to generalize the calculus of one real variable x to that of a complex z. From providing further techniques of ordinary integral calculus, to the integral representations of special functions and the (closely related) solutions to ordinary differential equations encountered in physics, to the Fourier analysis of partial differential equations, to the computation of quasi-normal oscillations of black holes, to the study of scattering amplitudes of Quantum Field Theories – this seemingly basic generalization of introductory real calculus has numerous applications to both mathematics and physics.

6.1 Differentiation

[45]The derivative of a complex function $f(z)$ is defined in a similar way as its real counterpart:

$$f'(z) \equiv \frac{\mathrm{d}f(z)}{\mathrm{d}z} \equiv \lim_{\Delta z \to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z}. \tag{6.1.1}$$

However, the meaning is considerably more subtle because Δz, just like z itself, is now complex. What this means is that, in taking this limit, it has to yield the same answer no matter what direction you approach z on the complex plane. For example, if $z = x + iy$, taking the derivative along the real direction must be equal to that along the imaginary one,

$$\begin{aligned} f'(z) &= \lim_{\Delta x \to 0} \frac{f(x+\Delta x+iy) - f(x+iy)}{\Delta x} = \partial_x f(z) \\ &= \lim_{\Delta y \to 0} \frac{f(x+i(y+\Delta y)) - f(x+iy)}{i\Delta y} = \frac{\partial f(z)}{\partial (iy)} = \frac{1}{i}\partial_y f(z), \end{aligned} \tag{6.1.2}$$

where x, y, Δx and Δy are real. This direction independence imposes very strong constraints on complex differentiable functions: they will turn out to be extremely smooth, in that if you can differentiate them at a given point z, you are guaranteed they are differentiable infinite number of times there. (This is not true of real functions.) If $f(z)$ is differentiable in some region on the complex plane, we say $f(z)$ is analytic there.

Problem 6.1. Derivative in the $\theta-$direction Suppose we take the derivative in the $\theta-$direction, namely $\Delta z = \epsilon e^{i\theta}$ for ϵ infinitesimal, show that

$$f'(z) = e^{-i\theta}\left(\cos\theta \cdot \partial_x f(z) + \sin\theta \cdot \partial_y f(z)\right). \tag{6.1.3}$$

Check this result against the $\theta = 0$ and $\theta = \pi/2$ cases above. □

If the first derivatives of f are continuous, the criteria for determining whether it is differentiable comes in the following pair of partial differential equations.

> **Cauchy-Riemann conditions for analyticity** Let $z = x + iy$ and $f(z) = u(x,y) + iv(x,y)$, where x, y, u and v are real. Let u and v have continuous first partial derivatives in x and y. Then $f(z)$ is an analytic function in the neighborhood

[45]The material in this Chapter is based on Arfken, Weber and Harris [18]; Morse and Feshbach [13]; and Byron and Fuller [14].

of z if and only if the following (Cauchy-Riemann) equations are satisfied by the real and imaginary parts of f:

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v. \tag{6.1.4}$$

To understand these Cauchy-Riemann conditions, we first consider differentiating along the (real) x direction,

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial f}{\partial x} = \partial_x u + i\partial_x v. \tag{6.1.5}$$

If we instead differentiate along the (imaginary) iy direction,

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{1}{i}\frac{\partial f}{\partial y} = \frac{1}{i}\partial_y u + \partial_y v = \partial_y v - i\partial_y u. \tag{6.1.6}$$

Since these two results must be the same, we may equate their real and imaginary parts to obtain eq. (6.1.4). Altogether, eq. (6.1.4) is equivalent to:

$$\left(\frac{\partial}{\partial x} - \frac{1}{i}\frac{\partial}{\partial y}\right) f(x,y) = 0. \tag{6.1.7}$$

Conversely, if u and v have continuous first derivatives then we may consider an arbitrary variation of the function f.[46]

$$\mathrm{d}f = (\partial_x u + i\partial_x v)\mathrm{d}x + (\partial_y u + i\partial_y v)\mathrm{d}y. \tag{6.1.8}$$

For the $(\dots)\mathrm{d}y$ terms, the Cauchy-Riemann conditions in eq. (6.1.4) may now be employed to turn $\partial_y u$ into $-\partial_x v$; and $\partial_y v$ into $\partial_x u$.

$$\begin{aligned}\mathrm{d}f &= (\partial_x u + i\partial_x v)\mathrm{d}x + (\partial_x u + i\partial_x v)i\mathrm{d}y \\ &= \partial_x f \mathrm{d}z \end{aligned} \tag{6.1.9}$$

Along similar lines, it is also possible to write $\mathrm{d}f = \partial_y f \mathrm{d}z$, upon using Cauchy-Riemann. Hence, we may vary f along *any* $\mathrm{d}z$ and if Cauchy-Riemann holds, then $\mathrm{d}f/\mathrm{d}z = \partial_x f = \partial_y f$ is well-defined – i.e., f is thus analytic.

Remark Notice, the Cauchy-Riemann equations of (6.1.4) allow us to solve for the real part of f in terms its imaginary part (or, vice versa) by integrating these first order relations. We will, below, provide such an integral solution known as the Hilbert transform pairs.

Cauchy-Riemann in Polar Coordinates It is also useful to express the Cauchy-Riemann conditions in polar coordinates $(x,y) = r(\cos\theta, \sin\theta)$. We have

$$\partial_r = \frac{\partial x}{\partial r}\partial_x + \frac{\partial y}{\partial r}\partial_y = \cos\theta\partial_x + \sin\theta\partial_y, \tag{6.1.10}$$

$$\partial_\theta = \frac{\partial x}{\partial \theta}\partial_x + \frac{\partial y}{\partial \theta}\partial_y = -r\sin\theta\partial_x + r\cos\theta\partial_y. \tag{6.1.11}$$

[46] It is at this point, if we did not assume u and v have continuous first derivatives, that we see the Cauchy-Riemann conditions in eq. (6.1.4) are necessary but not necessarily sufficient ones for analyticity. For, $\mathrm{d}f$ may no longer be approximated by its first derivatives, and the following steps no longer hold.

By viewing this as a matrix equation $(\partial_r, \partial_\theta)^{\mathrm{T}} = M(\partial_x, \partial_y)^{\mathrm{T}}$, we may multiply M^{-1} on both sides and obtain the (∂_x, ∂_y) in terms of the $(\partial_r, \partial_\theta)$.

$$\partial_x = \cos\theta\partial_r - \frac{\sin\theta}{r}\partial_\theta \tag{6.1.12}$$

$$\partial_y = \sin\theta\partial_r + \frac{\cos\theta}{r}\partial_\theta. \tag{6.1.13}$$

The Cauchy-Riemann conditions in eq. (6.1.4) can now be manipulated by replacing the ∂_x and ∂_y with the right hand sides above. Denoting $c \equiv \cos\theta$ and $s \equiv \sin\theta$,

$$\left(c\partial_r - \frac{s}{r}\partial_\theta\right) u = \left(s\partial_r + \frac{c}{r}\partial_\theta\right) v, \tag{6.1.14}$$

$$\left(s\partial_r + \frac{c}{r}\partial_\theta\right) u = -\left(c\partial_r - \frac{s}{r}\partial_\theta\right) v. \tag{6.1.15}$$

Multiplying these by appropriate factors of sines and cosines, we deduce

$$\left(cs\partial_r - \frac{s^2}{r}\partial_\theta\right) u = \left(s^2\partial_r + \frac{cs}{r}\partial_\theta\right) v, \tag{6.1.16}$$

$$\left(cs\partial_r + \frac{c^2}{r}\partial_\theta\right) u = -\left(c^2\partial_r - \frac{cs}{r}\partial_\theta\right) v; \tag{6.1.17}$$

and

$$\left(c^2\partial_r - \frac{sc}{r}\partial_\theta\right) u = \left(sc\partial_r + \frac{c^2}{r}\partial_\theta\right) v, \tag{6.1.18}$$

$$\left(s^2\partial_r + \frac{sc}{r}\partial_\theta\right) u = -\left(cs\partial_r - \frac{s^2}{r}\partial_\theta\right) v. \tag{6.1.19}$$

Subtracting the first pair and adding the second pair of equations, we arrive at the polar coordinates version of Cauchy-Riemann:

$$\frac{1}{r}\partial_\theta u = -\partial_r v \qquad \text{and} \qquad \partial_r u = \frac{1}{r}\partial_\theta v. \tag{6.1.20}$$

Examples Complex differentiability is much more restrictive than the real case. An example is $f(z) = |z|$. If z is real, then at least for $z \neq 0$, we may differentiate $f(z)$ – the result is $f'(z) = 1$ for $z > 0$ and $f'(z) = -1$ for $z < 0$. But in the complex case we would identify, with $z = x + iy$,

$$f(z) = |z| = \sqrt{x^2 + y^2} = u(x, y) + iv(x, y) \qquad \Rightarrow \qquad v(x, y) = 0. \tag{6.1.21}$$

We see that the Cauchy-Riemann conditions in eq. (6.1.4) cannot be satisfied since v is zero while u is non-zero. Alternatively, one may simply recognize $|z| = \sqrt{z^* z}$ is not independent of $\bar{z}$.

Moreover, any f that remains strictly real across the complex z plane is not differentiable unless it is constant.

$$f(x, y) = u(x, y) \qquad \Rightarrow \qquad \partial_x u = \partial_y v = 0, \quad \partial_y u = -\partial_x v = 0. \tag{6.1.22}$$

Similarly, if f were purely imaginary across the complex z plane, it is not differentiable unless it is constant.

$$f(x,y) = iv(x,y) \qquad \Rightarrow \qquad 0 = \partial_x u = \partial_y v, \quad 0 = -\partial_y u = \partial_x v. \tag{6.1.23}$$

Cauchy-Riemann as $\bar{z}$ independence It is also useful to recast the Cartesian coordinates (x,y) in terms of the complex coordinate $z = x + iy$ and its complex conjugate $z^* = \bar{z} = x - iy$ through the relations

$$x = \frac{z+\bar{z}}{2} \qquad \text{and} \qquad y = \frac{z-\bar{z}}{2i}. \tag{6.1.24}$$

This in turn means any (not necessarily analytic) complex function $f(x,y) = u(x,y) + iv(x,y) = f(z,\bar{z})$ can be viewed as a function of z and $\bar{z}$. Taking into account eq. (6.1.24),

$$\begin{aligned} \mathrm{d}f(z,\bar{z}) = \mathrm{d}u + i\mathrm{d}v &= \partial_x f(z,\bar{z}) \frac{\mathrm{d}z + \mathrm{d}\bar{z}}{2} + \partial_y f(z,\bar{z}) \frac{\mathrm{d}z - \mathrm{d}\bar{z}}{2i} \\ &= \frac{1}{2}\left(\frac{\partial f(z,\bar{z})}{\partial x} + \frac{\partial f(z,\bar{z})}{\partial (iy)}\right)\mathrm{d}z + \frac{1}{2}\left(\frac{\partial f(z,\bar{z})}{\partial x} - \frac{\partial f(z,\bar{z})}{\partial (iy)}\right)\mathrm{d}\bar{z} \qquad (6.1.25) \\ &\equiv \frac{\partial f(z,\bar{z})}{\partial z}\mathrm{d}z + \frac{\partial f(z,\bar{z})}{\partial \bar{z}}\mathrm{d}\bar{z}. \qquad (6.1.26) \end{aligned}$$

[47]Using the version of Cauchy-Riemann relations in eq. (6.1.7), the $\mathrm{d}\bar{z}$ term in eq. (6.1.25) is set to zero and we infer

$$\frac{\mathrm{d}f}{\mathrm{d}z} = \frac{\partial f}{\partial x} = \frac{1}{i}\frac{\partial f}{\partial y} \qquad \text{and} \qquad \frac{\partial f}{\partial \bar{z}} = 0. \tag{6.1.27}$$

To sum:

A complex function $f(x,y) = f(z,\bar{z}) = f(z,z^*)$ with continuous first derivatives is analytic (i.e., complex differentiable) if and only if it is independent of $z^* = \bar{z}$.

For instance, z^*, $\mathrm{Re}\, z = (z+\bar{z})/2$, $|z|^2 = z^* z$ are not analytic functions because they depend on both z and z^*.

Differentiation rules We will prove below – by the principle of *analytic continuation* – that if you know how to differentiate a function $f(z)$ when z is real, then as long as you can show that $f'(z)$ exists, the differentiation formula for the complex case would carry over from the real case. That is, suppose $f'(z) = g(z)$ when f, g and z are real; then this *form* has to hold for complex z. For example, powers are differentiated the same way

$$\frac{\mathrm{d}}{\mathrm{d}z} z^\alpha = \alpha z^{\alpha-1}, \qquad \alpha \in \mathbb{R}; \tag{6.1.28}$$

and

$$\frac{\mathrm{d}\sin(z)}{\mathrm{d}z} = \cos z, \qquad \frac{\mathrm{d}a^z}{\mathrm{d}z} = \frac{\mathrm{d}e^{z\ln a}}{\mathrm{d}z} = a^z \ln a. \tag{6.1.29}$$

[47]In case the assumption of continuous first derivatives is not clear – note that, if $\partial_x f$ and $\partial_y f$ were not continuous, then $\mathrm{d}f$ (the variation of f) in the direction across the discontinuity cannot be computed in terms of the first derivatives. Drawing a plot for a real function $F(x)$ with a discontinuous first derivative (i.e., a "kink") would help.

It is not difficult to check the first derivatives of z^α, $\sin(z)$ and a^z are continuous; and the Cauchy-Riemann conditions are satisfied. For instance, $z^\alpha = r^\alpha e^{i\alpha\theta} = r^\alpha \cos(\alpha\theta) + i r^\alpha \sin(\alpha\theta)$ and eq. (6.1.20) can be verified.

$$r^{\alpha-1}\partial_\theta \cos(\alpha\theta) = -\alpha r^{\alpha-1}\sin(\alpha\theta) \stackrel{?}{=} -\sin(\alpha\theta)\partial_r r^\alpha = -\alpha r^{\alpha-1}\sin(\alpha\theta), \tag{6.1.30}$$

$$\cos(\alpha\theta)\partial_r r^\alpha = \alpha r^{\alpha-1}\cos(\alpha\theta) \stackrel{?}{=} r^{\alpha-1}\partial_\theta \sin(\alpha\theta) = \alpha r^{\alpha-1}\cos(\alpha\theta). \tag{6.1.31}$$

(This proof that z^α is analytic fails at $r = 0$; in fact, for $\alpha < 1$, we see that z^α is not analytic there.) In particular, differentiability is particularly easy to see if $f(z)$ can be *defined* through its power series.

Product and chain rules The product and chain rules apply too. For instance,

$$(fg)' = f'g + fg'. \tag{6.1.32}$$

because

$$\begin{aligned}(fg)' &= \lim_{\Delta z\to 0}\frac{f(z+\Delta z)g(z+\Delta z) - f(z)g(z)}{\Delta z}\\ &= \lim_{\Delta z\to 0}\frac{(f(z) + f'\cdot\Delta z)(g(z) + g'\Delta z) - f(z)g(z)}{\Delta z}\\ &= \lim_{\Delta z\to 0}\frac{fg + fg'\Delta z + f'g\Delta z + \mathcal{O}((\Delta z)^2) - fg}{\Delta z} = f'g + fg'. \end{aligned}\tag{6.1.33}$$

We will have more to say later about carrying over properties of real differentiable functions to their complex counterparts.

Problem 6.2. Simple analytic functions Use the Cauchy-Riemann conditions to verify that e^z, $\ln z$ and z^α are analytic functions. (You may assume α is real.) Can you identify where Cauchy-Riemann would break down for these functions? Then, explain why $(z+5)/(z^*+i)$ is not analytic. □

Problem 6.3. Conformal (angle-preserving) transformations Complex functions $\{f(x,y) = u(x,y) + iv(x,y)\}$ can be thought of as a map from one 2D plane to another; i.e., from (x,y) to (u,v). Prove that analytic ones define angle preserving transformations everywhere their derivatives are not zero, $\{f'(z) \neq 0\}$.

Hints: First, let us recall eq. (2.0.14), which tells us the Re part of $z_1^* z_2$ yields the dot product between the two complex numbers z_1 and z_2, when they are viewed as vectors. For analytic $f(z = x+iy) = u(x,y) + iv(x,y)$, the directions $\mathrm{d}f$ on the $(u,v)-$plane induced by the small displacements $\mathrm{d}z_1 = \mathrm{d}x_1 + i\mathrm{d}y_1$ and $\mathrm{d}z_2 = \mathrm{d}x_2 + i\mathrm{d}y_2$ on the $z-$plane are described by

$$\mathrm{d}f_1 = (\partial_x u + i\partial_x v)(\mathrm{d}x_1 + i\mathrm{d}y_1), \tag{6.1.34}$$

$$\mathrm{d}f_2 = (\partial_x u + i\partial_x v)(\mathrm{d}x_2 + i\mathrm{d}y_2). \tag{6.1.35}$$

Show that Re $\mathrm{d}f_1^*\mathrm{d}f_2$ is proportional Re $\mathrm{d}z_1^*\mathrm{d}z_2$; and explain why that implies, as long as $f'(z) \neq 0$, the angle between $\mathrm{d}f_1$ and $\mathrm{d}f_2$ is the same as that between $\mathrm{d}z_1$ and $\mathrm{d}z_2$. □

2D Laplace's equation Suppose $f(z) = u(x, y) + iv(x, y)$, where $z = x + iy$ and x, y, u and v are real. If $f(z)$ is complex-differentiable then the Cauchy-Riemann relations in eq. (6.1.4) imply that both the real and imaginary parts of a complex function obey Laplace's equation, namely

$$(\partial_x^2 + \partial_y^2)u(x, y) = (\partial_x^2 + \partial_y^2)v(x, y) = 0. \tag{6.1.36}$$

To see this we differentiate eq. (6.1.4) appropriately,

$$\partial_x\partial_y u = \partial_y^2 v, \qquad \partial_x\partial_y u = -\partial_x^2 v \tag{6.1.37}$$

$$\partial_x^2 u = \partial_x\partial_y v, \qquad -\partial_y^2 u = \partial_x\partial_y v. \tag{6.1.38}$$

We now can equate the right hand sides of the first line; and the left hand sides of the second line. This leads to (6.1.36).

Because of eq. (6.1.36), complex analysis can be very useful for 2D electrostatic problems.

Moreover, u and v cannot admit local minimum or maximums, as long as $\partial_x^2 u$ and $\partial_x^2 v$ (or, equivalently, $\partial_y^2 u$ and $\partial_y^2 v$) are non-zero. In particular, the determinants of the 2×2 Hessian matrices $\partial^2 u/\partial(x, y)^i\partial(x, y)^j$ and $\partial^2 v/\partial(x, y)^i\partial(x, y)^j$ – and hence the product of their eigenvalues – are negative. For,

$$\begin{aligned} \det \frac{\partial^2 u}{\partial(x, y)^i\partial(x, y)^j} &= \det \begin{bmatrix} \partial_x^2 u & \partial_x\partial_y u \\ \partial_x\partial_y u & \partial_y^2 u \end{bmatrix} \\ &= \partial_x^2 u \partial_y^2 u - (\partial_x\partial_y u)^2 = -(\partial_y^2 u)^2 - (\partial_y^2 v)^2 \leq 0, \end{aligned} \tag{6.1.39}$$

$$\begin{aligned} \det \frac{\partial^2 v}{\partial(x, y)^i\partial(x, y)^j} &= \det \begin{bmatrix} \partial_x^2 v & \partial_x\partial_y v \\ \partial_x\partial_y v & \partial_y^2 v \end{bmatrix} \\ &= \partial_x^2 v \partial_y^2 v - (\partial_x\partial_y v)^2 = -(\partial_y^2 v)^2 - (\partial_y^2 u)^2 \leq 0, \end{aligned} \tag{6.1.40}$$

where both equations (6.1.36) and (6.1.37) were employed.

All turning points of an analytic $f(z)$ are saddles points.

This will be an important feature when, for instance, we seek approximate expressions for certain class of integrals involving complex analytic functions.

6.2 Integration, Laurent Series, Analytic Continuation

Complex integration is really a line integral $\int \vec{A} \cdot (\mathrm{d}x, \mathrm{d}y)$ on the 2D complex plane. Given some path (aka "contour") C, defined by $z(\lambda_1 \leq \lambda \leq \lambda_2) = x(\lambda) + iy(\lambda)$, with $z(\lambda_1) = z_1$ and $z(\lambda_2) = z_2$,

$$\begin{aligned} \int_C \mathrm{d}z f(z) &= \int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (\mathrm{d}x + i\mathrm{d}y)\,(u(x, y) + iv(x, y)) \\ &= \int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (u\mathrm{d}x - v\mathrm{d}y) + i\int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (v\mathrm{d}x + u\mathrm{d}y) \\ &= \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \left(u\frac{\mathrm{d}x(\lambda)}{\mathrm{d}\lambda} - v\frac{\mathrm{d}y(\lambda)}{\mathrm{d}\lambda}\right) + i\int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \left(v\frac{\mathrm{d}x(\lambda)}{\mathrm{d}\lambda} + u\frac{\mathrm{d}y(\lambda)}{\mathrm{d}\lambda}\right) \end{aligned}$$

$$\equiv \int_{\vec{z}(\lambda_1 \leq \lambda \leq \lambda_2)} \vec{A}_{\mathrm{R}} \cdot \mathrm{d}\vec{z} + i \int_{\vec{z}(\lambda_1 \leq \lambda \leq \lambda_2)} \vec{A}_{\mathrm{I}} \cdot \mathrm{d}\vec{z}; \tag{6.2.1}$$

$$\vec{A}_{\mathrm{R}} \equiv (u(x,y), -v(x,y))^{\mathrm{T}}, \qquad \vec{A}_{\mathrm{I}} \equiv (v(x,y), u(x,y))^{\mathrm{T}}, \qquad \mathrm{d}\vec{z} \equiv (\mathrm{d}x, \mathrm{d}y)^{\mathrm{T}}. \tag{6.2.2}$$

The real part of the line integral involves $\mathrm{Re}\vec{A} \equiv \vec{A}_{\mathrm{R}} = (u, -v)$ and its imaginary part Im $\vec{A} \equiv \vec{A}_{\mathrm{I}} = (v, u)$.
Remark I Because complex integration is a line integral, reversing the direction of contour C (which we denote as $-C$) would yield return negative of the original integral.

$$\int_{-C} \mathrm{d}z f(z) = -\int_C \mathrm{d}z f(z) \tag{6.2.3}$$

Fundamental Theorem of Calculus For analytic f, i.e., where $f'(z)$ exists, the complex version of the fundamental theorem of calculus has to hold:

$$\begin{aligned} \int_C \mathrm{d}z f'(z) &= \int_C \mathrm{d}f = f(\text{“upper” end point of } C) - f(\text{“lower” end point of } C) \\ &= \int_{z_1}^{z_2} \mathrm{d}z f'(z) = f(z_2) - f(z_1). \end{aligned} \tag{6.2.4}$$

In words: the integral of $f'(z)$ with respect to z *does not* depend on the path taken, even though the general line integral does. It turns out this path-independence also applies to the integral of $f(z)$ itself.

Cauchy's integral theorem In introducing the contour integral in eq. (6.2.1), we are not assuming any properties about the integrand $f(z)$. However, if the complex function $f(z)$ is analytic throughout some simply connected region[48] containing the contour C, then we are lead to one of the key results of complex integration theory: the integral of $f(z)$ within any *closed* path C there is zero.

$$\oint_C f(z)\mathrm{d}z = 0 \tag{6.2.5}$$

For a detailed proof, the mathematically minded can consult, say, Brown and Churchill [19].

Problem 6.4. Stokes vs Cauchy If the first derivatives of $f(z)$ are assumed to be continuous, then a proof of this modified Cauchy's theorem can be carried out by starting with the view that $\oint_C f(z)\mathrm{d}z$ is a (complex) line integral around a closed loop. Then apply Stokes' theorem followed by the Cauchy-Riemann conditions in eq. (6.1.4). Can you fill in the details? □

Path Independence Cauchy's theorem has an important implication. Suppose we have a contour integral $\int_C g(z)\mathrm{d}z$, where C is some arbitrary (not necessarily closed) contour. Suppose we have another contour C' whose end points coincide with those of C. If the function $g(z)$ is analytic inside the region bounded by C and C', then it has to be that

$$\int_C g(z)\mathrm{d}z = \int_{C'} g(z)\mathrm{d}z. \tag{6.2.6}$$

[48] A simply connected region is one where every closed loop in it can be shrunk to a point.

The reason is that, by subtracting these two integrals, say $(\int_C - \int_{C'})g(z)\mathrm{d}z$, the $-$ sign can be absorbed by reversing the direction of the C' integral. We then have a closed contour integral $(\int_C - \int_{C'})g(z)\mathrm{d}z = \oint g(z)\mathrm{d}z$ and Cauchy's theorem in eq. (6.2.5) applies. Along similar lines, we may argue the integral of f is itself an analytic function F, defined as

$$F(z) \equiv \int_{z_0}^{z} f(z')\mathrm{d}z'; \tag{6.2.7}$$

because the result is independent of the path taken from z_0 to z.

This path-independence is a very useful observation because it means, for a given contour integral, you can deform the contour itself to a shape that would make the integral easier to evaluate. Below, we will generalize this and show that, even if there are isolated points where the function is not analytic, you can still pass the contour over these points, but at the cost of incurring additional terms resulting from taking the residues there. Another common type of singularity is known as a branch point, which will then require us to introduce a branch cut.

Note that the simply connected requirement can often be circumvented by considering an appropriate cut line. For example, suppose C_1 and C_2 were both counterclockwise (or both clockwise) contours around an annulus region, within which $f(z)$ is analytic. Then

$$\oint_{C_1} f(z)\mathrm{d}z = \oint_{C_2} f(z)\mathrm{d}z. \tag{6.2.8}$$

Example I A simple but important example is the following integral, where the contour C is an arbitrary counterclockwise closed loop that encloses the point $z = 0$.

$$I \equiv \oint_C \frac{\mathrm{d}z}{z} \tag{6.2.9}$$

Cauchy's integral theorem does not apply directly because $1/z$ is not analytic at $z = 0$. By considering a counterclockwise circle C' of radius $R > 0$, however, we may argue

$$\oint_C \frac{\mathrm{d}z}{z} = \oint_{C'} \frac{\mathrm{d}z}{z}. \tag{6.2.10}$$

[49]We may then employ polar coordinates, so that the path C' could be described as $z = Re^{i\theta}$, where θ would run from 0 to 2π.

$$\oint_C \frac{\mathrm{d}z}{z} = \int_0^{2\pi} \frac{\mathrm{d}(Re^{i\theta})}{Re^{i\theta}} = \int_0^{2\pi} i\mathrm{d}\theta = 2\pi i. \tag{6.2.11}$$

Example II Let's evaluate $\oint_C z\mathrm{d}z$ and $\oint_C \mathrm{d}z$ directly and by using Cauchy's integral theorem. Here, C is some closed contour on the complex plane. Directly:

$$\oint_C z\mathrm{d}z = \left.\frac{z^2}{2}\right|_{z=z_0}^{z=z_0} = 0, \qquad \oint_C \mathrm{d}z = z|_{z=z_0}^{z=z_0} = 0. \tag{6.2.12}$$

Using Cauchy's integral theorem – we first note that z and 1 are analytic, since they are powers of z; we thus conclude the integrals are zero.

[49]This is where drawing a picture would help: for simplicity, if C' lies entirely within C, the first portion of the cut lines would begin anywhere from C' to anywhere to C, followed by the reverse trajectory from C to C' that runs infinitesimally close to the first portion. Because they are infinitesimally close, the contributions of these two portions cancel; but we now have a simply connected closed contour integral that amounts to $0 = (\int_C - \int_{C'})\mathrm{d}z/z$.

Problem 6.5. For some contour C, let M be the maximum of $|f(z)|$ along it and $L \equiv \int_C \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$ be the length of the contour itself, where $z = x + iy$ (for x and y real). Argue that

$$\left| \int_C f(z)\mathrm{d}z \right| \leq \int_C |f(z)||\mathrm{d}z| \leq M \cdot L. \tag{6.2.13}$$

Note: $|\mathrm{d}z| = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$. (Why?) Hints: Can you first argue for the triangle inequality, $|z_1 + z_2| \leq |z_1| + |z_2|$, for any two complex numbers $z_{1,2}$? What about $|z_1 + z_2 + \cdots + z_N| \leq |z_1| + |z_2| + \cdots + |z_N|$? Then view the integral as a discrete sum, and apply this generalized triangle inequality to it. □

Problem 6.6. Evaluate

$$\oint_C \frac{\mathrm{d}z}{z(z+1)}, \tag{6.2.14}$$

where C is an arbitrary contour enclosing the points $z = 0$ and $z = -1$. Note that Cauchy's integral theorem is not directly applicable here. Hint: Apply a partial fractions decomposition of the integrand, then for each term, convert this arbitrary contour to an appropriate circle. □

The next major result allows us to deduce $f(z)$, for z lying within some contour C, by knowing its values on C.

> **Cauchy's integral formula** If $f(z)$ is analytic on and within some closed counterclockwise contour C, then
>
> $$\begin{aligned} \oint_C \frac{\mathrm{d}z'}{2\pi i} \frac{f(z')}{z'-z} &= f(z) \qquad \text{if } z \text{ lies inside } C \\ &= 0 \qquad \text{if } z \text{ lies outside } C. \end{aligned} \tag{6.2.15}$$

Proof If z lies outside C then the integrand is analytic within its interior and therefore Cauchy's integral theorem applies. If z lies within C we may then deform the contour such that it becomes an infinitesimal counterclockwise circle around $z' \approx z$,

$$z' \equiv z + \epsilon e^{i\theta}, \qquad 0 < \epsilon \ll 1. \tag{6.2.16}$$

We then have

$$\begin{aligned} \oint_C \frac{\mathrm{d}z'}{2\pi i} \frac{f(z')}{z'-z} &= \frac{1}{2\pi i} \int_0^{2\pi} \epsilon e^{i\theta} i \mathrm{d}\theta \frac{f(z + \epsilon e^{i\theta})}{\epsilon e^{i\theta}} \\ &= \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi} f(z + \epsilon e^{i\theta}). \end{aligned} \tag{6.2.17}$$

By taking the limit $\epsilon \to 0^+$, we get $f(z)$, since $f(z')$ is analytic and thus continuous at $z' = z$.

Problem 6.7. Cauchy's integral as an average Let C be a circular contour of radius R centered at z; i.e., $z' \equiv z + Re^{i\theta}$. Suppose $f(z')$ is analytic on and within C, show that

$$f(z) = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi} f\left(z + Re^{i\theta}\right). \tag{6.2.18}$$

This informs us, the value of an analytic function at z is the average of its values on *any* circle centered at z. Use this fact to explain why neither the Re or Im portion of f can be a local minimum or maximum. This is, in fact, consistent with the fact that they satisfy the homogeneous Laplace equation – recall eq. (6.1.36). $\square$

> **Cauchy's integral formula for derivatives** By applying the limit definition of the derivative, we may obtain an analogous definition for the nth derivative of $f(z)$. For some closed counterclockwise contour C,
>
> $$\oint_C \frac{\mathrm{d}z'}{2\pi i} \frac{f(z')}{(z'-z)^{n+1}} = \frac{f^{(n)}(z)}{n!} \qquad \text{if } z \text{ lies inside } C$$
> $$= 0 \qquad \text{if } z \text{ lies outside } C. \tag{6.2.19}$$

This implies – as already advertised earlier – once $f'(z)$ exists, $f^{(n)}(z)$ also exists for any n. Complex-differentiable functions are infinitely smooth.

The converse of Cauchy's integral formula is known as Morera's theorem, which we will simply state without proof.

> **Morera's theorem** If $f(z)$ is continuous in a simply connected region and $\oint_C f(z)\mathrm{d}z = 0$ for *any* closed contour C within it, then $f(z)$ is analytic throughout this region.

Now, even though $f^{(n>1)}(z)$ exists once $f'(z)$ exists (cf. (6.2.19)), $f(z)$ cannot be infinitely smooth everywhere on the complex $z-$plane.

> **Liouville's theorem** If $f(z)$ is analytic and bounded – i.e., $|f(z)|$ is less than some positive constant M – for all complex z, then $f(z)$ must in fact be a constant. Apart from the constant function, analytic functions must therefore blow up somewhere on the complex plane.

Proof To prove this result we employ eq. (6.2.19). Choose a counterclockwise circular contour C that encloses some arbitrary point z,

$$|f^{(n)}(z)| \leq n! \oint_C \frac{|\mathrm{d}z'|}{2\pi} \frac{|f(z')|}{|(z'-z)^{n+1}|} \tag{6.2.20}$$

$$\leq n! \frac{M}{2\pi r^{n+1}} \oint_C |\mathrm{d}z'| = n! \frac{M}{r^n}. \tag{6.2.21}$$

Here, r is the radius from z to C. But by Cauchy's theorem, the circle can be made arbitrarily large. By sending $r \to \infty$, we see that $|f^{(n)}(z)| = 0$, the nth derivative of the analytic function at an arbitrary point z is zero for any integer $n \geq 1$. This proves the theorem.

Examples The exponential e^z while differentiable everywhere on the complex plane, does in fact blow up at Re $z \to \infty$. Sines and cosines are oscillatory and bounded on the real line; and are differentiable everywhere on the complex plane. However, they blow up as one move towards positive or negative imaginary infinity. Remember $\sin(z) = (e^{iz} - e^{-iz})/(2i)$ and $\cos(z) = (e^{iz} + e^{-iz})/2$. Then, for $R \in \mathbb{R}$,

$$\sin(iR) = \frac{e^{-R} - e^R}{2i}, \qquad \cos(iR) = \frac{e^{-R} + e^R}{2}. \tag{6.2.22}$$

Both $\sin(iR)$ and $\cos(iR)$ blow up as $R \to \pm\infty$.

Problem 6.8. Fundamental theorem of algebra. Let $P(z) = p_0 + p_1 z + \dots p_n z^n$ be an nth degree polynomial, where n is an integer greater or equal to 1. By considering $f(z) = 1/P(z)$ in view of Liouville's theorem, show that $P(z)$ has at least one root.

Once a root has been found, we can divide it out from $P(z)$ and repeat the argument for the remaining $(n-1)$-degree polynomial. By induction, this implies an nth degree polynomial has exactly n roots – this is the fundamental theorem of algebra. □

Taylor series The generalization of the Taylor series of a real differentiable function to the complex case is known as the Laurent series. If the function is completely smooth in some region on the complex plane, then we shall see that it can in fact be Taylor expanded the usual way, except the expressions are now complex. If there are isolated points where the function blows up, then it can be (Laurent) expanded about those points, in powers of the complex variable – except the series begins at some negative integer power, as opposed to the zeroth power in the usual Taylor series.

To begin, let us show that the geometric series still works in the complex case.

Problem 6.9. Complex Geometric Series By starting with the Nth partial sum,

$$S_N \equiv \sum_{\ell=0}^{N} t^\ell, \tag{6.2.23}$$

prove that, as long as $|t| < 1$,

$$\frac{1}{1-t} = \sum_{\ell=0}^{\infty} t^\ell. \tag{6.2.24}$$

Hint: You may need to consider $\lim_{N\to\infty} |S_N - (1-t)^{-1}|$. □

Now pick a point z_0 on the complex plane and identify the nearest point, say z_1, where f is no longer analytic. Consider some closed counterclockwise contour C that lies within the circular region $|z - z_0| < |z_1 - z_0|$. Then we may apply Cauchy's integral formula eq. (6.2.15), and deduce a series expansion about z_0:

$$\begin{aligned} f(z) &= \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{z'-z} \\ &= \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{(z'-z_0)-(z-z_0)} = \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{(z'-z_0)(1-(z-z_0)/(z'-z_0))} \\ &= \sum_{\ell=0}^{\infty} \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{(z'-z_0)^{\ell+1}} (z-z_0)^\ell . \end{aligned} \tag{6.2.25}$$

We have used the geometric series in eq. (6.2.24) and the fact that it converges uniformly to interchange the order of integration and summation. At this point, if we now recall Cauchy's integral formula for the nth derivative of an analytic function, eq. (6.2.19), we have arrived at its Taylor series.

For $f(z)$ complex analytic within the circular region $|z - z_0| < |z_1 - z_0|$, where z_1 is the nearest point to z_0 where f is no longer differentiable,

$$f(z) = \sum_{\ell=0}^{\infty} (z - z_0)^\ell \frac{f^{(\ell)}(z_0)}{\ell!}, \tag{6.2.26}$$

where $f^{(\ell)}(z)/\ell!$ is given by eq. (6.2.19).

Problem 6.10. Complex Binomial Theorem. For p any real number and z any complex number obeying $|z| < 1$, prove the complex binomial theorem using eq. (6.2.26),

$$(1+z)^p = \sum_{\ell=0}^{\infty} \binom{p}{\ell} z^\ell, \tag{6.2.27}$$

$$\binom{p}{0} \equiv 1, \qquad \binom{p}{\ell} = \frac{p(p-1)\dots(p-(\ell-1))}{\ell!}. \tag{6.2.28}$$

Explain why this, in turn, implies the generalized binomial expansion

$$(X+Y)^p = \Theta\left(|X/Y| - 1\right) X^p \sum_{\ell=0}^{\infty} \binom{p}{\ell} \left(\frac{Y}{X}\right)^\ell + \Theta\left(|Y/X| - 1\right) Y^p \sum_{\ell=0}^{\infty} \binom{p}{\ell} \left(\frac{X}{Y}\right)^\ell. \tag{6.2.29}$$

The existence of two different expansion schemes, depending on whether $|X| > |Y|$ or $|X| < |Y|$, leads us to the following discussion of the Laurent series. □

Laurent series We are now ready to derive the Laurent expansion of a function $f(z)$ that is analytic within an annulus, say bounded by the circles $|z - z_0| = r_1$ and $|z - z_0| = r_2 > r_1$. That is, the center of the annulus region is z_0 and the smaller circle has radius r_1 and larger one r_2. To start, we let C_1 be a clockwise circular contour with radius $r_2 > r_1' > r_1$ and let C_2 be a counterclockwise circular contour with radius $r_2 > r_2' > r_1' > r_1$. As long as z lies between these two circular contours, we have

$$f(z) = \left(\int_{C_1} + \int_{C_2} \right) \frac{\mathrm{d}z'}{2\pi i} \frac{f(z')}{z' - z}. \tag{6.2.30}$$

Strictly speaking, we need to integrate along a cut line joining the C_1 and C_2 – and another one infinitesimally close to it, in the opposite direction – so that we can form a closed contour. But by assumption $f(z)$ is analytic and therefore continuous; the integrals along these pair of cut lines must cancel. For the C_1 integral, we may write $z' - z = -(z - z_0)(1 - (z' - z_0)/(z - z_0))$ and apply the geometric series in eq. (6.2.24) because $|(z' - z_0)/(z - z_0)| < 1$. Similarly, for the C_2 integral, we may write $z' - z = (z' - z_0)(1 - (z - z_0)/(z' - z_0))$ and geometric series expand the right factor because $|(z - z_0)/(z' - z_0)| < 1$. These lead us to

$$f(z) = \sum_{\ell=0}^{\infty} (z - z_0)^\ell \int_{C_2} \frac{\mathrm{d}z'}{2\pi i} \frac{f(z')}{(z' - z_0)^{\ell+1}} - \sum_{\ell=0}^{\infty} \frac{1}{(z - z_0)^{\ell+1}} \int_{C_1} \frac{\mathrm{d}z'}{2\pi i} (z' - z_0)^\ell f(z'). \tag{6.2.31}$$

Remember complex integration can be thought of as a line integral, which reverses sign if we reverse the direction of the line integration. Therefore we may absorb the $-$ sign in front of the C_1 integral(s) by turning C_1 from a clockwise circle into $C_1' = -C_1$, a counterclockwise one. Moreover, note that we may now deform the contour C_1' into C_2,

$$\int_{C_1'} \frac{\mathrm{d}z'}{2\pi i}(z'-z_0)^\ell f(z') = \int_{C_2} \frac{\mathrm{d}z'}{2\pi i}(z'-z_0)^\ell f(z'), \tag{6.2.32}$$

because for positive ℓ the integrand $(z'-z_0)^\ell f(z')$ is analytic in the region lying between the circles C_1' and C_2. At this point we have

$$f(z) = \sum_{\ell=0}^{\infty} \int_{C_2} \frac{\mathrm{d}z'}{2\pi i}\left((z-z_0)^\ell \frac{f(z')}{(z'-z_0)^{\ell+1}} + \frac{1}{(z-z_0)^{\ell+1}}(z'-z_0)^\ell f(z')\right). \tag{6.2.33}$$

Proceeding to re-label the second series by replacing $\ell + 1 \to -\ell'$, so that the summation then runs from -1 through $-\infty$, the Laurent series emerges.

> Let $f(z)$ be analytic within the annulus $r_1 < |z - z_0| < r_2 < |z_1 - z_0|$, where z_0 is some complex number such that $f(z)$ may not be analytic within $|z - z_0| < r_1$; z_1 is the nearest point outside of $|z - z_0| \geq r_1$ where $f(z)$ fails to be differentiable; and the radii $r_2 > r_1 > 0$ are real positive numbers. The Laurent expansion of $f(z)$ about z_0, valid throughout the entire annulus, reads
>
> $$f(z) = \sum_{\ell=-\infty}^{\infty} L_\ell(z_0)\cdot(z-z_0)^\ell, \tag{6.2.34}$$
>
> $$L_\ell(z_0) \equiv \oint_C \frac{\mathrm{d}z'}{2\pi i}\frac{f(z')}{(z'-z_0)^{\ell+1}}. \tag{6.2.35}$$
>
> The C is any counterclockwise closed contour containing both z and the inner circle $|z - z_0| = r_1$.

Uniqueness It is worth asserting that the Laurent expansion of a function, in the region where it is analytic, is unique. That means it is not always necessary to perform the integrals in eq. (6.2.34) to obtain the expansion coefficients L_ℓ.

Problem 6.11. For complex z, a and b, obtain the Laurent expansion of

$$f(z) \equiv \frac{1}{(z-a)(z-b)}, \qquad a \neq b, \tag{6.2.36}$$

about $z = a$, in (I) the region $0 < |z - a| < |a - b|$; as well as (II) in the region $|z - a| > |a - b|$ using eq. (6.2.34). Check your result either by writing

$$\frac{1}{z-b} = -\frac{1}{1-(z-a)/(b-a)}\frac{1}{b-a}. \tag{6.2.37}$$

and employing the geometric series in eq. (6.2.24), or directly performing a Taylor expansion of $1/(z - b)$ about $z = a$.

Problem 6.12. Schwarz reflection principle. Proof the following statement using Laurent expansion. If a function $f(z = x + iy) = u(x, y) + iv(x, y)$ can be Laurent expanded (for x, y, u, and v real) about some point on the real line, and if $f(z)$ is real whenever z is real, then

$$f(z)^* = u(x, y) - iv(x, y) = f(z^*) = u(x, -y) + iv(x, -y). \tag{6.2.38}$$

Comment on why this is called the "reflection principle".

There are lots of functions satisfying this $f(z)^* = f(z^*)$ property: $\sin(z)$, $\cos(z)$, $\tan(z)$, Gamma $\Gamma(z)$, Bessel $J_\nu(z)$, etc. □

Analytic continuation We now turn to an important result that allows us to extend the definitions of complex differentiable functions beyond their original range of validity. Suppose the function $f(z = x+iy)$ is analytic within some region on the $z-$plane, and its value is specified on some line segment $(x(\lambda), y(\lambda))$ lying with this region for some real parameter λ. That means we may compute its derivatives $\mathrm{d}^n f(z(\lambda) = x(\lambda) + iy(\lambda))/\mathrm{d}\lambda^n$ for all $n \geq 1$ at any fixed point z_0 on the line segment. This in turn implies, the nth derivative $f^{(n)}(z_0)$ can be computed once the values $f(z(\lambda))$ on a line are known. Since an analytic function is uniquely determined by its infinite set of derivatives at any fixed point within its domain of analyticity, we arrive at the following statement.

> An analytic function $f(z)$ is fixed uniquely throughout a given region Σ on the complex plane, once its value is specified on a line segment lying within Σ.

This in turn means, suppose we have an analytic function $f_1(z)$ defined in a region Σ_1 on the complex plane, and suppose we found another analytic function $f_2(z)$ defined in some region Σ_2 such that $f_2(z)$ agrees with $f_1(z)$ in their common region of intersection. (It is important that Σ_2 does have some overlap with Σ_1.) Then we may view $f_2(z)$ as an analytic continuation of $f_1(z)$, because this extension is unique – it is not possible to find a $f_3(z)$ that agrees with $f_1(z)$ in the common intersection between Σ_1 and Σ_2, yet behave different in the rest of Σ_2.

These results inform us, any real differentiable function we are familiar with can be extended to the complex plane, simply by knowing its Taylor expansion. For example, e^x is infinitely differentiable on the real line, and its definition can be readily extended into the complex plane via its Taylor expansion.

An example of analytic continuation is that of the geometric series. If we define

$$f_1(z) \equiv \sum_{\ell=0}^{\infty} z^\ell, \qquad |z| < 1, \tag{6.2.39}$$

and

$$f_2(z) \equiv \frac{1}{1 - z}, \tag{6.2.40}$$

then we know they agree in the region $|z| < 1$ and therefore any line segment within it. But while $f_1(z)$ is defined only in this region, $f_2(z)$ is valid for any $z \neq 1$. Therefore, we may view $1/(1 - z)$ as the analytic continuation of $f_1(z)$ for the region $|z| > 1$. Also observe that we can now understand why the series is valid only for $|z| < 1$: the series of $f_1(z)$ is really the Taylor expansion of $f_2(z)$ about $z = 0$, and since the nearest singularity is at $z = 1$, the circular region of validity employed in our (constructive) Taylor series proof is in fact $|z| < 1$.

YZ: Example. Coincide on a line. Gamma function.

Exponential and trigonometric functions The exponential $e^z = e^x e^{iy}$, for x and y denoting the real and imaginary parts of z, may be readily checked to be analytic – i.e., it satisfies the Cauchy-Riemann equations and hence

$$(e^z)' = \partial_x e^x e^{iy} = e^z. \tag{6.2.41}$$

On the other hand, if we first restrict z to the real line, so that $e^z = e^x$, then

$$(e^z)' = \partial_x e^x = e^z. \tag{6.2.42}$$

Since the rightmost expression may be extended to the entire complex $z-$plane, and since we know $(e^z)'$ has to exist we may 'analytic continue' the result and deduce

$$(e^z)' = e^z \tag{6.2.43}$$

for all $z \in \mathbb{C}$. Along similar lines, we know that $(\sin z)' = \cos z$, $(\cos z)' = -\sin z$, $(\sinh z)' = \cosh z$, etc. on the real line. Since the right hand sides can be analytically continued to the entire complex plane, these relations must therefore hold for all complex z.

Gamma Function The Gamma function $\Gamma(z)$ – the generalization of the factorial $n!$ for arbitrary complex numbers – is an excellent place to demonstrate the principle of analytic continuation. (It is also a function that shows up frequently in theoretical physics; for example, even in quantum field theory calculations.) We start by *defining* it through the integral, for Re $z > 0$,

$$\Gamma(z) \equiv \int_0^{+\infty} t^{z-1} e^{-t} \mathrm{d}t. \tag{6.2.44}$$

The Re $z > 0$ is needed because of the contribution from the lower end of the integration; for $\varepsilon < 1$,

$$\int_0^{\varepsilon} t^{z-1} e^{-t} \mathrm{d}t = \int_0^{\varepsilon} t^{z-1} \left(1 - t + \frac{t^2}{2} + \dots\right) \mathrm{d}t. \tag{6.2.45}$$

We see that the (first) t^{z-1} term yields a divergence upon integration once Re $z \leq 0$. We will use the principle of analytic continuation to understand how to extend $\Gamma(z)$ to the Re $z \leq 0$ region. Hence, we may ask the question: How do we extend the definition of $\Gamma(z)$ to the Re $z < 0$ portion of the complex plane?

One way to do so is to first derive the recursion relation

$$z\Gamma(z) = \Gamma(z+1). \tag{6.2.46}$$

For the reader's reference we also collect the identity

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \tag{6.2.47}$$

and the 'duplication formula'

$$2^{2z-1}\Gamma(z)\Gamma\left(z+\frac{1}{2}\right)=\sqrt{\pi}\Gamma(2z). \tag{6.2.48}$$

We apply integrate by parts to the definition in eq. (**??**).

$$\Gamma(z+1)=-\int_0^{+\infty}t^{(z+1)-1}\partial_t e^{-t}\mathrm{d}t \tag{6.2.49}$$

$$=\left[-t^z e^{-t}\right]_{t=0}^{t=\infty}+z\int_0^{+\infty}t^{z-1}e^{-t}\mathrm{d}t=z\Gamma(z). \tag{6.2.50}$$

Shifting $z\to z-1$ in eq. (6.2.46) and dividing both sides by $z-1$,

$$\Gamma(z-1)=\frac{\Gamma(z)}{z-1}. \tag{6.2.51}$$

The right hand side is an analytic function for Re $z>0$ except at $z=1$. But for the same domain of z, the $z-1$ argument of the Gamma function on the left hand side now extends to Re $z>-1$. Hence, we may now recognize this relation as an analytic continuation of $\Gamma(z)$ from the Re $z>0$ to the Re $z>-1$ region: since both $\Gamma(z-1)$ and $\Gamma(z)/(z-1)$ certainly agree in the Re $z>0$ region, we may define $\Gamma(z-1)$ uniquely in the $-1<\mathrm{Re}z\leq 0$ region via $\Gamma(z)/(z-1)$. We may then continue this process, by repeated application of eq. (6.2.46):

$$(z-1)(z-2)\dots(z-n+1)(z-n)\Gamma(z-n)=\Gamma(z). \tag{6.2.52}$$

Even though this is derived with the Re $z>0$ constraint, we may write

$$\Gamma(z-n)=\frac{\Gamma(z)}{(z-1)(z-2)\dots(z-(n-1))(z-n)}. \tag{6.2.53}$$

The right hand side is analytic for Re $z>0$ except at $z=1,2,3,\dots,n-1,n$. For the same domain, the $z-n$ argument of the Gamma function of the left hand side lies within the region Re $z>-n$. Hence, this result tells us we may analytic continue $\Gamma(z)$ using the recursion relation in eq. (6.2.46) to the strip $-n<\mathrm{Re}z<0$, for any positive integer $n=1,2,3,\dots$.

Incidentally, we may now verify that $\Gamma(n+1)$ is the factorial $n!$ by using eq. (6.2.46). First, we compute from eq. (6.2.44):

$$\Gamma(1)\equiv\int_0^{+\infty}e^{-t}\mathrm{d}t=1. \tag{6.2.54}$$

Then, using eq. (6.2.46),

$$n!=n!\cdot\Gamma(1)=n\cdot(n-1)\cdot(n-2)\dots(2)\cdot(1)\cdot\Gamma(1) \tag{6.2.55}$$

$$=n\cdot(n-1)\cdot(n-2)\dots(2)\cdot\Gamma(2)=\dots \tag{6.2.56}$$

$$=n\cdot(n-1)\cdot\Gamma(n-1)=n\cdot\Gamma(n)=\Gamma(n+1). \tag{6.2.57}$$

We also observe that $\Gamma(0)$ must be singular. We set $z = \varepsilon$, for $0 < \varepsilon \ll 1$ and compute eq. (6.2.44):

$$\varepsilon\Gamma(\varepsilon) = \Gamma(\varepsilon + 1) = \int_0^\infty t^\varepsilon e^{-t} \mathrm{d}t \tag{6.2.58}$$

$$= \int_0^\infty \left(1 + \varepsilon \cdot \ln t + \mathcal{O}(\varepsilon^2)\right) e^{-t} \mathrm{d}t. \tag{6.2.59}$$

The second integral is $\Gamma'(1)$ because

$$\Gamma'(z) = \int_0^\infty t^{z-1} \ln(t) e^{-t} \mathrm{d}t. \tag{6.2.60}$$

Therefore

$$\Gamma(\varepsilon) = \frac{1}{\varepsilon} + \Gamma'(1) + \mathcal{O}\left(\varepsilon\right). \tag{6.2.61}$$

Even though we have derived this with the assumption that $\varepsilon > 0$, since $1/\varepsilon$ is analytic for $\varepsilon < 0$, this relation must therefore – by analytic continuation – be valid for the region around $\varepsilon \approx 0$. In the following problem, you will uncover the structure of the singularities at $z = 0, -1, -2, \dots$.

Problem 6.13. Poles of $\Gamma(z)$ at non-positive integers Use equations (6.2.46) and (6.2.61) to show that, for $n = 0, 1, 2, \dots$ (non-positive integers) and $|\varepsilon| \ll 1$,

$$\Gamma(\varepsilon - n) = \frac{(-)^n}{n!}\left(\frac{1}{\varepsilon} - \gamma_{\mathrm{E}} + 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \mathcal{O}\left(\varepsilon\right)\right); \tag{6.2.62}$$

where $\gamma_{\mathrm{E}} = -\Gamma'(1) = 0.57721566490\dots$ is the Euler–Mascheroni constant.

This result is used to regulate divergences encountered in quantum field theory calculations, within a scheme known as dimensional regularization, where the dimension of spacetime is $d = 4 - 2\varepsilon$. □

We will derive eq. (6.2.47) below, after discussing branch cuts; but assuming it holds we may now deduce these $z = 0, -1, -2, -3, \dots$ are the *only* singularities of $\Gamma(z)$. First note that the only singularities of the right hand side of eq. (6.2.47) occurs at the zeroes of $\sin(\pi z)$; at $z = 0, \pm 1, \pm 2, \pm 3$. However, at the positive integers $z = n + 1 = 1, 2, 3, \dots$, we have already computed $\Gamma(z) = \Gamma(n + 1) = n!$; and therefore $\Gamma(z)\Gamma(1 - z)$ is singular there because of the $\Gamma(1 - z) = \Gamma(-n)$ factor. Likewise, $\Gamma(z)\Gamma(1 - z)$ is singular at the non-positive integers $z = 0, -1, -2, -3, \dots$ only because of the $\Gamma(z)$ factor, as $\Gamma(1 - z) = \Gamma(n + 1) = n!$.

We may also observe, eq. (6.2.47) tells us $\Gamma(z)$ does not go to zero for any finite $|z|$. The right hand side of eq. (6.2.47) does not go to zero for any finite $|z|$. Hence, suppose $\Gamma(z)$ has a zero at $z_\star$, then $\Gamma(1 - z_\star)$ must have a singularity to 'cancel out' the $\Gamma(z_\star)$ in such a manner so that $\Gamma(z_\star)\Gamma(1 - z_\star)$ is finite. (Similarly, if $\Gamma(1 - z)$ had a zero at $z_\star$, then $\Gamma(z_\star)$ must blow up.) But as we have just shown, the Gamma function can only blow up when its argument $z_\star$ (or, $1 - z_\star$) is a non-positive integer $-n = 0, -1, -2, -\dots$; which means it must have a zero at $1 - z_\star$ (or, $z_\star$) – a positive integer $n + 1 = 1, 2, 3, \dots$, where $\Gamma(n + 1) = n!$. This contradiction proves there is indeed no finite $|z|$ that yields $\Gamma(z) = 0$. As a consequence

The reciprocal of the Gamma function $1/\Gamma(z)$ is an entire function – it is analytic everywhere on the complex plane where $|z|$ is finite – because $\Gamma(z) = 0$ has no solutions there. Moreover, the only singularities of $\Gamma(z)$ are those at $z = -n = 0, -1, -2, \dots$, and the Gamma function behaves as eq. (6.2.62) in the neighborhoods of these non-positive integers.

There is another way we may analytically continue $\Gamma(z)$, as described in Lebedev [5], by splitting the integral in eq. (6.2.44) into two.

$$\Gamma\left(\mathrm{Re}\ z > 0\right) = \left(\int_0^1 + \int_1^\infty\right) t^{z-1} e^{-t} \mathrm{d}t \tag{6.2.63}$$

$$= \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \int_0^1 t^{z-1+\ell} \mathrm{d}t + \int_1^\infty t^{z-1} e^{-t} \mathrm{d}t \tag{6.2.64}$$

$$= \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!(z+\ell)} + \int_1^\infty t^{z-1} e^{-t} \mathrm{d}t. \tag{6.2.65}$$

The remaining integral on the last line converges for all complex $z = x + iy$ and is an analytic function.

$$\left|\int_1^\infty \frac{e^{-t}}{t} t^x t^{iy} \mathrm{d}t\right| \leq \int_1^\infty \left|\frac{e^{-t}}{t} t^x t^{iy}\right| \mathrm{d}t \leq \int_1^\infty e^{-t} t^{x-1} \mathrm{d}t < \infty \tag{6.2.66}$$

$$\left|\partial_z \int_1^\infty t^{z-1} e^{-t} \mathrm{d}t\right| \leq \int_1^\infty \left|(\ln t) t^{z-1} e^{-t}\right| \mathrm{d}t \leq \left|\partial_x \int_1^\infty t^x e^{-t} \mathrm{d}t\right| \tag{6.2.67}$$

Whereas the summation also converges for all $z \neq 0, -1, -2, -3, \dots$, for the following reasons. First note that, by picking the nearest non-negative integer $n_\star$ to z, the $1/(z+n_\star)$ must be such that its length is the largest among all the $1/(z+n)$ occurring within the sum. Then

$$\left|\sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!(z+\ell)}\right| \leq \sum_{\ell=0}^{+\infty} \left|\frac{(-)^\ell}{\ell!(z+\ell)}\right| \leq \frac{1}{z+n_\star} \sum_{\ell=0}^{+\infty} \frac{1}{\ell!} = \frac{e}{z+n_\star}. \tag{6.2.68}$$

Remarkably, decomposing the $\int_0^\infty t^{z-1} e^{-t} \mathrm{d}t$ into an infinite sum plus the same integral with the lower limit lifted from 0 to 1 allow us to analytically continue it to all z except when it is zero or a negative integer.

$$\Gamma\left(z \in \mathbb{C}\right) = \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!(z+\ell)} + \int_1^\infty t^{z-1} e^{-t} \mathrm{d}t. \tag{6.2.69}$$

As $z = \varepsilon - n$ for $n = 0, -1, -2, \dots$ and $|\varepsilon| \ll 1$, the sum tells us the singularity goes as

$$\Gamma\left(z \to \varepsilon - n\right) = \frac{(-)^n}{n!\varepsilon} + \dots; \tag{6.2.70}$$

consistent with eq. (6.2.62).

Problem 6.14. Hypergeometric Function One key application of analytic continuation is that, some special functions in mathematical physics admit a power series expansion that has a finite radius of convergence. This can occur if the differential equations they solve have singular points. Many of these special functions also admit an integral representation, whose range of validity lies beyond that of the power series. This allows the domain of these special functions to be extended.

The hypergeometric function ${}_2F_1(\alpha, \beta; \gamma; z)$ is such an example. For $|z| < 1$ it has a power series expansion

$$\begin{aligned} {}_2F_1(\alpha, \beta; \gamma; z) &= \sum_{\ell=0}^{\infty} C_\ell(\alpha, \beta; \gamma) \frac{z^\ell}{\ell!}, \\ C_0(\alpha, \beta; \gamma) &\equiv 1, \\ C_{\ell \geq 1}(\alpha, \beta; \gamma) &\equiv \frac{\alpha(\alpha+1)\dots(\alpha+(\ell-1)) \cdot \beta(\beta+1)\dots(\beta+(\ell-1))}{\gamma(\gamma+1)\dots(\gamma+(\ell-1))}. \end{aligned} \tag{6.2.71}$$

On the other hand, it also has the following integral representation,

$${}_2F_1(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\gamma-\beta)\Gamma(\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha} \mathrm{d}t, \tag{6.2.72}$$

$$\mathrm{Re}(\gamma) > \mathrm{Re}(\beta) > 0. \tag{6.2.73}$$

(Here, $\Gamma(z)$ is the Gamma function.) Show that eq. (6.2.72) does in fact agree with eq. (6.2.71) for $|z| < 1$. You can apply the binomial expansion in eq. (6.2.27) to $(1-tz)^{-\alpha}$, followed by Beta function integral representation

$$\int_0^1 \mathrm{d}t (1-t)^{\alpha-1} t^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \equiv B(\alpha, \beta), \qquad \mathrm{Re}(\alpha),\ \mathrm{Re}(\beta) > 0. \tag{6.2.74}$$

You may also need to invoke eq. (6.2.46).

To reiterate: eq. (6.2.72) therefore extends eq. (6.2.71) into the region $|z| > 1$. □

Generating Functions One application of equations (6.2.19) and (6.2.34) is to the understanding of generating functions of 'special functions' in mathematical physics. For example, the Bessel function of integer order $J_n(z)$, for $n = 0, 1, 2, 3, \dots$, may be viewed as the coefficient of t^n in the expansion of

$$\exp\left[\frac{z}{2}\left(t - \frac{1}{t}\right)\right] = \sum_{n=-\infty}^{+\infty} t^n J_n(z). \tag{6.2.75}$$

Viewing this as a Laurent series in t, eq. (6.2.34) informs us

$$J_n(z) = \oint_C \frac{\mathrm{d}t'}{2\pi i t'^{n+1}} \exp\left[\frac{z}{2}\left(t' - \frac{1}{t'}\right)\right]. \tag{6.2.76}$$

If we choose the counter-clockwise closed contour C to be the unit circle, $t' = e^{i\theta}$ for $0 \leq \theta < 2\pi$, we arrive an integral representation of the Bessel function:

$$J_n(z) = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi} \exp\left[iz\sin(\theta) - in\theta\right]. \tag{6.2.77}$$

Since $\sin(x)$ is odd, we may also write this as

$$J_n(z) = \int_0^\pi \frac{\mathrm{d}\theta}{\pi} \cos\left[z\sin(\theta) - n\theta\right]. \tag{6.2.78}$$

Problem 6.15. Generating Function: Hermite Polynomials The Hermite polynomials may be defined by the derivatives

$$H_n(x) = (-)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}. \tag{6.2.79}$$

Explain why

$$\frac{H_n(x)}{n!} = (-)^n e^{x^2} \oint_C \frac{\mathrm{d}z'}{2\pi i} \frac{e^{-z'^2}}{(z'-x)^{n+1}}, \tag{6.2.80}$$

for some closed counterclockwise C that contains x. Use this result to prove that

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x) = \exp\left(-t^2 + 2xt\right). \tag{6.2.81}$$

Hint: You should find the sum involves a geometric series. The integral then becomes an application of eq. (6.2.15). □

6.3 Poles and Residues

In this section we will consider the closed counterclockwise contour integral

$$\oint_C \frac{\mathrm{d}z}{2\pi i} f(z), \tag{6.3.1}$$

where $f(z)$ is analytic everywhere on and within C except at isolated singular points of $f(z)$ – which we will denote as $\{z_1, \dots, z_n\}$, for $(n \geq 1)$-integer. That is, we will assume there is no other type of singularities. We will show that the result is the sum of the residues of $f(z)$ at these points. This case will turn out to have a diverse range of physical applications, including the study of the vibrations of black holes.

We begin with some jargon.

Nomenclature If a function $f(z)$ admits a Laurent expansion about $z = z_0$ starting from $1/(z-z_0)^m$, for m some positive integer,

$$f(z) = \sum_{\ell=-m}^{\infty} L_\ell(z_0) \cdot (z-z_0)^\ell, \tag{6.3.2}$$

we say the function has a pole of order m at $z = z_0$. (For example, the $\Gamma(z)$ has poles of order 1, aka simple poles, at $z = 0, -1, -2, -3, \dots$; see eq. (6.2.62).) If $m = \infty$ we say the function has an essential singularity. The residue of a function f at some location z_0 is simply the coefficient L_{-1} of the negative one power ($\ell = -1$ term) of the Laurent series expansion about $z = z_0$.

The key to the result already advertised is the following.

Problem 6.16. If n is an arbitrary integer, show that

$$\oint_C (z'-z)^n \frac{\mathrm{d}z'}{2\pi i} = 1, \qquad \text{when } n = -1,$$
$$= 0, \qquad \text{when } n \neq -1, \tag{6.3.3}$$

where C is any contour (whose interior defines a simply connected domain) that encloses the point $z' = z$. □

By assumption, we may deform our contour C so that they become the collection of closed counterclockwise contours $\{C_i' | i = 1, 2, \dots, n\}$ around each and every isolated point. This means

$$\oint_C f(z') \frac{\mathrm{d}z'}{2\pi i} = \sum_i \oint_{C_i'} f(z') \frac{\mathrm{d}z'}{2\pi i}. \tag{6.3.4}$$

Strictly speaking, to preserve the full closed contour structure of the original C, we need to join these new contours – say C_i' to C_{i+1}', C_{i+1}' to C_{i+2}', and so on – by a pair of contour lines placed infinitesimally apart, for e.g., one from $C_i' \to C_{i+1}'$ and the other $C_{i+1}' \to C_i'$. But by assumption $f(z)$ is analytic and therefore continuous there, and thus the contribution from these pairs will surely cancel. Let us perform a Laurent expansion of $f(z)$ about z_i, the ith singular point, and then proceed to integrate the series term-by-term using eq. (6.3.3).

$$\oint_{C_i'} f(z') \frac{\mathrm{d}z'}{2\pi i} = \int_{C_i'} \sum_{\ell=-m_i}^{\infty} L_\ell(z_i) \cdot (z' - z_i)^\ell \frac{\mathrm{d}z'}{2\pi i} = L_{-1}(z_i). \tag{6.3.5}$$

Residue theorem As advertised, the closed counterclockwise contour integral of a function that is analytic everywhere on and within the contour, except at isolated points $\{z_i | i = 1, \dots, N\}$, yields the sum of the residues at each of these points. In equation form,

$$\oint_C f(z') \frac{\mathrm{d}z'}{2\pi i} = \sum_{i=1}^{N} \oint_{C_i'} f(z') \frac{\mathrm{d}z'}{2\pi i} = \sum_{i=1}^{N} L_{-1}(z_i), \tag{6.3.6}$$

where C_i' is the contour that encircles *only* the ith singularity z_i and $L_{-1}(z_i)$ is the residue at the same z_i.

Example I Let us start with a simple application of this result. Let C be some closed counterclockwise contour containing the points $z = 0, a, b$.

$$I = \oint_C \frac{\mathrm{d}z}{2\pi i} \frac{1}{z(z-a)(z-b)}. \tag{6.3.7}$$

One way to do this is to perform a partial fractions expansion first.

$$I = \oint_C \frac{\mathrm{d}z}{2\pi i} \left(\frac{1}{abz} + \frac{1}{a(a-b)(z-a)} + \frac{1}{b(b-a)(z-b)} \right). \tag{6.3.8}$$

In this form, the residues are apparent, because we can view the first term as some Laurent expansion about $z = 0$ with only the negative one power; the second term as some Laurent expansion about $z = a$; the third about $z = b$. Therefore, the sum of the residues yield

$$I = \frac{1}{ab} + \frac{1}{a(a-b)} + \frac{1}{b(b-a)} = \frac{(a-b)+b-a}{ab(a-b)} = 0. \tag{6.3.9}$$

If you don't do a partial fractions decomposition, you may instead recognize, as long as the 3 points $z = 0, a, b$ are distinct, then near $z = 0$ the factor $1/((z-a)(z-b))$ is analytic and admits an ordinary Taylor series that begins at the zeroth order in z, i.e.,

$$\frac{1}{z(z-a)(z-b)} = \frac{1}{z}\left(\frac{1}{ab} + \mathcal{O}(z)\right). \tag{6.3.10}$$

Because the higher positive powers of the Taylor series cannot contribute to the $1/z$ term of the Laurent expansion, to extract the negative one power of z in the Laurent expansion of the integrand, we simply evaluate this factor at $z = 0$. Likewise, near $z = a$, the factor $1/(z(z-b))$ is analytic and can be Taylor expanded in zero and positive powers of $(z - a)$. To understand the residue of the integrand at $z = a$ we simply evaluate $1/(z(z-b))$ at $z = a$. Ditto for the $z = b$ singularity.

$$\begin{aligned}\oint_C \frac{\mathrm{d}z}{2\pi i}\frac{1}{z(z-a)(z-b)} &= \sum_{z_i=0,a,b}\left(\text{Residue of } \frac{1}{z(z-a)(z-b)} \text{ at } z_i\right) \\ &= \frac{1}{ab} + \frac{1}{a(a-b)} + \frac{1}{b(b-a)} = 0. \end{aligned}\tag{6.3.11}$$

The reason why the result is zero can actually be understood via contour integration as well. If you now consider a closed clockwise contour C_∞ at infinity and view the integral $(\int_C + \int_{C_\infty})f(z)\mathrm{d}z$, you will be able to convert it into a closed contour integral by linking C and C_∞ via two infinitesimally close radial lines which would not actually contribute to the answer. But $(\int_C + \int_{C_\infty})f(z)\mathrm{d}z = \int_{C_\infty} f(z)\mathrm{d}z$ because C_∞ does not contribute either – why? Therefore, since there are no poles in the region enclosed by C_∞ and C, the answer has to be zero.

Example II Let C be a closed counterclockwise contour around the origin $z = 0$. Let us do

$$I \equiv \oint_C \exp(1/z^2)\mathrm{d}z. \tag{6.3.12}$$

We series expand the exp, and notice there is no term that goes as $1/z$; i.e., the residue at $z = 0$ is 0. Hence,

$$I = \sum_{\ell=0}^{\infty}\frac{1}{\ell!}\oint_C \frac{\mathrm{d}z}{z^{2\ell}} = 0. \tag{6.3.13}$$

A major application of contour integration is to that of integrals involving real variables.

6.3.1 Trigonometric integrals

If we have an integral of the form

$$\int_0^{2\pi} \mathrm{d}\theta f(\cos\theta, \sin\theta), \tag{6.3.14}$$

then it may help to change from θ to

$$z \equiv e^{i\theta} \qquad \Rightarrow \qquad \mathrm{d}z = i\mathrm{d}\theta \cdot e^{i\theta} = i\mathrm{d}\theta \cdot z, \tag{6.3.15}$$

and

$$\sin\theta = \frac{z - 1/z}{2i}, \qquad \cos\theta = \frac{z + 1/z}{2}. \tag{6.3.16}$$

The integral is converted into a sum over residues:

$$\begin{aligned} \int_0^{2\pi} \mathrm{d}\theta f(\cos\theta, \sin\theta) &= 2\pi \oint_{|z|=1} \frac{\mathrm{d}z}{2\pi i z} f\left(\frac{z+1/z}{2}, \frac{z-1/z}{2i}\right) \\ &= 2\pi \sum_j \left(j\text{th residue of } \frac{f\left(\frac{z+1/z}{2}, \frac{z-1/z}{2i}\right)}{z} \text{ for } |z| < 1 \right). \end{aligned} \tag{6.3.17}$$

Example For $a \in \mathbb{R}$,

$$\begin{aligned} I = \int_0^{2\pi} \frac{\mathrm{d}\theta}{a + \cos\theta} &= \oint_{|z|=1} \frac{\mathrm{d}z}{iz} \frac{1}{a + (1/2)(z + 1/z)} = \oint_{|z|=1} \frac{\mathrm{d}z}{i} \frac{1}{az + (1/2)(z^2 + 1)} \\ &= 4\pi \oint_{|z|=1} \frac{\mathrm{d}z}{2\pi i} \frac{1}{(z - z_+)(z - z_-)}, \qquad z_\pm \equiv -a \pm \sqrt{a^2 - 1}. \end{aligned} \tag{6.3.18}$$

Assume, for the moment, that $|a| < 1$. Then $|-a \pm \sqrt{a^2-1}|^2 = |-a \pm i\sqrt{1-a^2}|^2 = |a^2 + (1-a^2)|^2 = 1$. Both $z_\pm$ lie on the unit circle, and the contour integral does not make much sense as it stands because the contour C passes through both $z_\pm$. So let us assume that a is real but $|a| > 1$. When a runs from 1 to infinity, $-a - \sqrt{a^2-1}$ runs from -1 to $-\infty$; while $-a + \sqrt{a^2-1} = -(a - \sqrt{a^2-1})$ runs from -1 to 0 because $a > \sqrt{a^2-1}$. When $-a$ runs from 1 to ∞, on the other hand, $-a - \sqrt{a^2-1}$ runs from 1 to 0; while $-a + \sqrt{a^2-1}$ runs from 1 to ∞. In other words, for $a > 1$, $z_+ = -a + \sqrt{a^2-1}$ lies within the unit circle and the relevant residue is $1/(z_+ - z_-) = 1/(2\sqrt{a^2-1}) = \mathrm{sgn}(a)/(2\sqrt{a^2-1})$. For $a < -1$ it is $z_- = -a - \sqrt{a^2-1}$ that lies within the unit circle and the relevant residue is $1/(z_- - z_+) = -1/(2\sqrt{a^2-1}) = \mathrm{sgn}(a)/(2\sqrt{a^2-1})$. Therefore,

$$\int_0^{2\pi} \frac{\mathrm{d}\theta}{a + \cos\theta} = \frac{2\pi \mathrm{sgn}(a)}{\sqrt{a^2-1}}, \qquad a \in \mathbb{R},\ |a| > 1. \tag{6.3.19}$$

6.3.2 Integrals along the real line

If you need to do $\int_{-\infty}^{+\infty} f(z)\mathrm{d}z$, it may help to view it as a complex integral and "close the contour" either in the upper or lower half of the complex plane – thereby converting the integral along the real line into one involving the sum of residues in the upper or lower plane.

An example is the following

$$I \equiv \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{z^2+z+1}. \tag{6.3.20}$$

Let us complexify the integrand and consider its behavior in the limit $z = \lim_{\rho\to\infty} \rho e^{i\theta}$, either for $0 \leq \theta \leq \pi$ (large semi-circle in the upper half plane) or $\pi \leq \theta \leq 2\pi$ (large semi-circle in the lower half plane).

$$\lim_{\rho\to\infty} \left| \frac{i\mathrm{d}\theta \cdot \rho e^{i\theta}}{\rho^2 e^{i2\theta} + \rho e^{i\theta} + 1} \right| \to \lim_{\rho\to\infty} \frac{\mathrm{d}\theta}{\rho} = 0. \tag{6.3.21}$$

This is saying the integral along this large semi-circle either in the upper or lower half complex plane is zero. Therefore I is equal to the integral along the real axis plus the contour integral along the semi-circle, since the latter contributes nothing. But the advantage of this view is that we now have a closed contour integral. Because the roots of the polynomial in the denominator of the integrand are $e^{-i2\pi/3}$ and $e^{i2\pi/3}$, so we may write

$$I = 2\pi i \oint_C \frac{\mathrm{d}z}{2\pi i} \frac{1}{(z-e^{-i2\pi/3})(z-e^{i2\pi/3})}. \tag{6.3.22}$$

Closing the contour in the upper half plane yields a counterclockwise path, which yields

$$I = \frac{2\pi i}{e^{i2\pi/3} - e^{-i2\pi/3}} = \frac{\pi}{\sin(2\pi/3)}. \tag{6.3.23}$$

Closing the contour in the lower half plane yields a clockwise path, which yields

$$I = \frac{-2\pi i}{e^{-i2\pi/3} - e^{i2\pi/3}} = \frac{\pi}{\sin(2\pi/3)}. \tag{6.3.24}$$

Of course, the two answers have to match.

Example: Fourier transform The Fourier transform is in fact a special case of the integral on the real line that can often be converted to a closed contour integral.

$$f(t) = \int_{-\infty}^{\infty} \widetilde{f}(\omega) e^{i\omega t} \frac{\mathrm{d}\omega}{2\pi}, \qquad t \in \mathbb{R}. \tag{6.3.25}$$

We will assume t is real and $\widetilde{f}$ has only isolated singularities.[50] Let C be a large semi-circular path, either in the upper or lower complex plane; consider the following integral along C.

$$I' \equiv \int_C \widetilde{f}(\omega) e^{i\omega t} \frac{\mathrm{d}\omega}{2\pi} = \lim_{\rho\to\infty} \int \widetilde{f}\left(\rho e^{i\theta}\right) e^{i\rho(\cos\theta)t} e^{-\rho(\sin\theta)t} \frac{i\mathrm{d}\theta \cdot \rho e^{i\theta}}{2\pi} \tag{6.3.26}$$

[50]In physical applications $\widetilde{f}$ may have branch cuts; this will be dealt with in the next section.

At this point we see that, for $t < 0$, unless $\widetilde{f}$ goes to zero much faster than the $e^{-\rho(\sin\theta)t}$ for large ρ, the integral blows up in the upper half plane where $(\sin\theta) > 0$. For $t > 0$, unless f goes to zero much faster than the $e^{-\rho(\sin\theta)t}$ for large ρ, the integral blows up in the lower half plane where $(\sin\theta) < 0$. In other words, the sign of t will determine how you should "close the contour" – in the upper or lower half plane.

Let us suppose $|\widetilde{f}| \leq M$ on the semi-circle and consider the magnitude of this integral,

$$|I'| \leq \lim_{\rho\to\infty} \left(\rho M \int e^{-\rho(\sin\theta)t} \frac{\mathrm{d}\theta}{2\pi} \right), \tag{6.3.27}$$

Remember if $t > 0$ we integrate over $\theta \in [0, \pi]$, and if $t < 0$ we do $\theta \in [-\pi, 0]$. Either case reduces to

$$|I'| \leq \lim_{\rho\to\infty} \left(2\rho M \int_0^{\pi/2} e^{-\rho(\sin\theta)|t|} \frac{\mathrm{d}\theta}{2\pi} \right), \tag{6.3.28}$$

because

$$\int_0^{\pi} F(\sin(\theta))\mathrm{d}\theta = 2\int_0^{\pi/2} F(\sin(\theta))\mathrm{d}\theta \tag{6.3.29}$$

for any function F. The next observation is that, over the range $\theta \in [0, \pi/2]$,

$$\frac{2\theta}{\pi} \leq \sin\theta, \tag{6.3.30}$$

because $y = 2\theta/\pi$ is a straight line joining the origin to the maximum of $y = \sin\theta$ at $\theta = \pi/2$. (Making a plot here helps.) This in turn means we can replace $\sin\theta$ with $2\theta/\pi$ in the exponent, i.e., exploit the inequality $e^{-X} < e^{-Y}$ if $X > Y > 0$, and deduce

$$|I'| \leq \lim_{\rho\to\infty} \left(2\rho M \int_0^{\pi/2} e^{-2\rho\theta|t|/\pi} \frac{\mathrm{d}\theta}{2\pi} \right) \tag{6.3.31}$$

$$= \lim_{\rho\to\infty} \left(\frac{\rho M}{\pi} \pi \frac{e^{-\rho\pi|t|/\pi} - 1}{-2\rho|t|} \right) = \frac{1}{2|t|} \lim_{\rho\to\infty} M \tag{6.3.32}$$

As long as $|\widetilde{f}(\omega)|$ goes to zero as $\rho \to \infty$, we see that I' (which is really 0) can be added to the Fourier integral $f(t)$ along the real line, converting $f(t)$ to a closed contour integral. If $\widetilde{f}(\omega)$ is analytic except at isolated points, then I can be evaluated through the sum of residues at these points.

To summarize, when faced with the frequency-transform type integral in eq. (6.3.25),

- If $t > 0$ and if $|\widetilde{f}(\omega)|$ goes to zero as $|\omega| \to \infty$ on the large semi-circle path of radius $|\omega|$ on the upper half complex plane, then we close the contour there and convert the integral $f(t) = \int_{-\infty}^{\infty} \widetilde{f}(\omega) e^{i\omega t} \frac{\mathrm{d}\omega}{2\pi}$ to i times the sum of the residues of $\widetilde{f}(\omega)e^{i\omega t}$ for $\mathrm{Im}(\omega) > 0$ – provided the function $\widetilde{f}(\omega)$ is analytic except at isolated points there.

- If $t < 0$ and if $|\widetilde{f}(\omega)|$ goes to zero as $|\omega| \to \infty$ on the large semi-circle path of radius $|\omega|$ on the lower half complex plane, then we close the contour there and convert the integral $f(t) = \int_{-\infty}^{\infty} \widetilde{f}(\omega)e^{i\omega t}\frac{d\omega}{2\pi}$ to $-i$ times the sum of the residues of $\widetilde{f}(\omega)e^{i\omega t}$ for $\text{Im}(\omega) < 0$ – provided the function $\widetilde{f}(\omega)$ is analytic except at isolated points there.

- A quick guide to how to close the contour is to evaluate the exponential on the imaginary ω axis, and take the infinite radius limit of $|\omega|$, namely $\lim_{|\omega|\to\infty} e^{it(\pm i|\omega|)} = \lim_{|\omega|\to\infty} e^{\mp t|\omega|}$, where the upper sign is for the positive infinity on the imaginary axis and the lower sign for negative infinity. We want the exponential to go to zero, so we have to choose the upper/lower sign based on the sign of t.

If $\widetilde{f}(\omega)$ requires branch cut(s) in either the lower or upper half complex planes – branch cuts will be discussed shortly – we may still use this closing of the contour to tackle the Fourier integral $f(t)$. In such a situation, there will often be additional contributions from the part of the contour hugging the branch cut itself.

An example is the following integral

$$I(t) \equiv \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega t}}{(\omega + i)^2(\omega - 2i)}, \qquad t \in \mathbb{R}. \tag{6.3.33}$$

The denominator $(\omega + i)^2(\omega - 2i)$ has a double root at $\omega = -i$ (in the lower half complex plane) and a single root at $\omega = 2i$ (in the upper half complex plane). You can check readily that $1/((\omega + i)^2(\omega - 2i))$ does go to zero as $|\omega| \to \infty$. If $t > 0$ we close the integral on the upper half complex plane. Since $e^{i\omega t}/(\omega + i)^2$ is analytic there, we simply apply Cauchy's integral formula in eq. (6.2.15).

$$I(t > 0) = i\frac{e^{i(2i)t}}{(2i + i)^2} = -i\frac{e^{-2t}}{9}. \tag{6.3.34}$$

If $t < 0$ we then need form a closed *clockwise* contour C by closing the integral along the real line in the lower half plane. Here, $e^{i\omega t}/(\omega - 2i)$ is analytic, and we can invoke eq. (6.2.19),

$$\begin{aligned} I(t < 0) = i \oint_C \frac{d\omega}{2\pi i} \frac{e^{i\omega t}}{(\omega + i)^2(\omega - 2i)} &= -i\frac{d}{d\omega}\left(\frac{e^{i\omega t}}{\omega - 2i}\right)_{\omega=-i} \\ &= -ie^t\frac{1 - 3t}{9} \end{aligned} \tag{6.3.35}$$

To summarize,

$$\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega t}}{(\omega + i)^2(\omega - 2i)} = -i\frac{e^{-2t}}{9}\Theta(t) - ie^t\frac{1 - 3t}{9}\Theta(-t), \tag{6.3.36}$$

where $\Theta(t)$ is the step function.

We can check this result as follows. Since $I(t = 0) = -i/9$ can be evaluated independently, this indicates we should expect the $I(t)$ to be continuous there: $I(t = 0^+) = I(t = -0^+) = -i/9$. Also notice, if we apply a t-derivative on $I(t)$ and interchange the integration and derivative operation, each $\mathrm{d}/\mathrm{d}t$ amounts to a $i\omega$. Therefore, we can check the following differential equations obeyed by $I(t)$:

$$\left(\frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}t}+i\right)^2\left(\frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}t}-2i\right)I(t)=\delta(t), \tag{6.3.37}$$

$$\left(\frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}t}+i\right)^2 I(t)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{e^{i\omega t}}{\omega-2i}=i\Theta(t)e^{-2t}, \tag{6.3.38}$$

$$\left(\frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}t}-2i\right)I(t)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{e^{i\omega t}}{(\omega+i)^2}=-i\Theta(-t)ite^t=\Theta(-t)te^t. \tag{6.3.39}$$

Problem 6.17. Evaluate

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}z}{z^3+i}. \tag{6.3.40}$$

Hint: What are the roots $z^3=-1$? Does it matter how you 'close-the-contour'? □

Problem 6.18. Integral Representation of Step Function Show that the integral representation of the step function $\Theta(t)$, defined in eq. (5.1.5), is

$$\Theta(t)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi i}\frac{e^{i\omega t}}{\omega-i0^+} \tag{6.3.41}$$

$$=\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{ie^{-i\omega t}}{\omega+i0^+}. \tag{6.3.42}$$

Differentiating this result and recalling eq. (5.1.13) then yields the integral representation of the Dirac $\delta-$function:

$$\Theta'(t)=\delta(t)=\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}e^{i\omega t}. \tag{6.3.43}$$

The $\omega-i0^+$ in eq. (6.3.41) means the purely imaginary root lies very slightly above 0; alternatively one would view it as an instruction to deform the contour by making an infinitesimally small counterclockwise semi-circle going slightly below the real axis around the origin. Whereas the $\omega+i0^+$ means the purely imaginary root lies very slightly below 0.

Next, use the integral representation in eq. (6.3.41) to justify the result

$$\Theta(0)=\frac{1}{2}. \tag{6.3.44}$$

Finally, let a and b be non-zero real numbers. Evaluate

$$I(a,b)\equiv\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi i}\frac{e^{i\omega a}}{\omega+ib}. \tag{6.3.45}$$

Hint: You may wish to consider $a>0$ and $a<0$ cases separately. □

Problem 6.19. (From Arfken et al. [18]) Sometimes this "closing-the-contour" trick need not involve closing the contour at infinity. Show by contour integration that

$$I \equiv \int_0^\infty \frac{(\ln x)^2}{1+x^2} \mathrm{d}x = \frac{\pi^3}{8}. \tag{6.3.46}$$

Hint: Put $x = z \equiv e^t$ and try to evaluate the integral now along the contour that runs along the real line from $t = -R$ to $t = R$ – for $R \gg 1$ – then along a vertical line from $t = R$ to $t = R + i\pi$, then along the horizontal line from $t = R + i\pi$ to $t = -R + i\pi$, then along the vertical line back to $t = -R$; then take the $R \to +\infty$ limit. □

Problem 6.20. Evaluate

$$I(a) \equiv \int_{-\infty}^\infty \frac{\sin(ax)}{x} \mathrm{d}x, \qquad a \in \mathbb{R}. \tag{6.3.47}$$

Hint(s): First convert the sine into exponentials and deform the contour along the real line into one that makes a infinitesimally small semi-circular detour around the origin $z = 0$. The semi-circle can be clockwise, passing above $z = 0$ or counterclockwise, going below $z = 0$. Make sure you justify why making such a small deformation does not affect the answer. □

Problem 6.21. Evaluate

$$I(t) \equiv \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{e^{-i\omega t}}{(\omega - ia)^2(\omega + ib)^2}, \qquad t \in \mathbb{R};\ a, b > 0. \tag{6.3.48}$$

□

Problem 6.22. Gamma Function and Complex Reciprocal Powers Justify the following result:

$$\int_0^\infty t^{z-1} e^{-\mu t} \mathrm{d}t = \frac{\Gamma(z)}{\mu^z}, \tag{6.3.49}$$

$$\mathrm{Re}(z) > -1, \qquad \mathrm{Re}(\mu) > 0. \tag{6.3.50}$$

Applications of eq. (6.3.49) may be found in (quantum) field theory calculations. □

Counting zeros of analytic functions Within a simply connected domain $\mathfrak{D}$ on the complex plane, such that C denotes the counterclockwise path along its boundary, let us show that the following integral

$$N = \frac{1}{2\pi i} \oint_C \frac{f'(z)}{f(z)} \mathrm{d}z \tag{6.3.51}$$

counts the number of zeros of f lying inside C – provided f is analytic there. Note that, if an analytic function vanishes at $z = z_0$, then in that neighborhood it can be Taylor expanded as $f(z) = c_n(z - z_0)^n + c_{n+1}(z - z_0)^{n+1} + \dots$. The $n \geq 1$ here is an integer; and we count $f(z)$ as having n zeros at $z = z_0$. The total number of zeros counts all the distinct $\{z_0\}$ but with each of their associated multiplicities included. For example, $f(z) = (z-1)(z-3)^2$ has three zeros on the entire complex plane; while $f(z) = z(z - \pi)$ has two.

Since $f(z)$ is analytic, $f'(z)/f(z)$ is analytic unless $f(z) = 0$. We may therefore deform our contour C and break it up into tiny ones $\{C_i\}$, with one encircling each and every zero. Suppose the ith zero is at $z = z_i$, and $f(z) = C_i(z - z_i)^{n_i}(1 + \mathcal{O}(z - z_i))$ there, for n_i a positive integer, then

$$\frac{1}{2\pi i}\oint_{C_i}\frac{f'(z)}{f(z)}\mathrm{d}z = \frac{1}{2\pi i}\oint_{C_i}\frac{C_i n_i(z - z_i)^{n_i - 1}(1 + \mathcal{O}(z - z_i))}{C_i(z - z_i)^{n_i}(1 + \mathcal{O}(z - z_i))}\mathrm{d}z. \tag{6.3.52}$$

Shrinking the radius of C_i to zero, we obtain

$$\frac{1}{2\pi i}\oint_{C_i}\frac{f'(z)}{f(z)}\mathrm{d}z = \frac{n_i}{2\pi i}\oint_{C_i}\frac{\mathrm{d}z}{z - z_i} = n_i. \tag{6.3.53}$$

The sum over the contributions from over all the $\{C_i\}$ is

$$\oint_C \frac{f'(z)}{f(z)}\mathrm{d}z = \sum_i \oint_{C_i}\frac{f'(z)}{f(z)}\mathrm{d}z = \sum_i n_i. \tag{6.3.54}$$

As an example, let $f(z) = \sum_{\ell=0}^{n} q_\ell z^\ell$ be a nth order polynomial, where $\{q_\ell\}$ are arbitrary non-zero complex coefficients. Consider an infinitely large circular contour centered at $z = 0$. Then,

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}\mathrm{d}z = \frac{1}{2\pi i}\lim_{|z|\to\infty}\oint_C \frac{\sum_{\ell=1}^{n} q_\ell z^{\ell-1}\ell}{\sum_{\ell=0}^{n} q_\ell z^\ell}\mathrm{d}z \tag{6.3.55}$$

$$= \frac{n}{2\pi i}\lim_{|z|\to\infty}\oint_C \frac{\mathrm{d}z}{z}\frac{\sum_{\ell=1}^{n} q_\ell z^{\ell-n}(\ell/n)}{\sum_{\ell=0}^{n} q_\ell z^{\ell-n}}. \tag{6.3.56}$$

As $|z| \to \infty$, all the positive powers of $1/z$ drop out, leaving only the q_n in the numerator and denominator. This leads us to conclude that the number of zeroes is

$$\frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)}\mathrm{d}z = n; \tag{6.3.57}$$

a result consistent with the fundamental theorem of algebra.

Integral Representations of Infinite Series There are certain types of infinite series that can be converted into a contour integral, which then allows the former to be evaluated by deforming the contour appropriately. The four main types of infinite sums we will address here are as follows.[51] For f analytic in the region near the entire real line,

$$\sum_{n=-\infty}^{+\infty} f(n) = \oint \frac{f(z)}{\tan(\pi z)}\frac{\mathrm{d}z}{2i} \tag{6.3.58}$$

$$\sum_{n=-\infty}^{+\infty} (-)^n n \cdot f(n) = \oint \frac{f(z)}{\sin(\pi z)}\frac{\mathrm{d}z}{2i} \tag{6.3.59}$$

$$\sum_{n=-\infty}^{+\infty} f\left(n + \frac{1}{2}\right) = -\oint f(z)\cdot\tan(\pi z)\frac{\mathrm{d}z}{2i} \tag{6.3.60}$$

[51] See Table 11.2 of Arfken et al [18].

$$\sum_{n=-\infty}^{+\infty} (-)^n f\left(n+\frac{1}{2}\right) = -\oint \frac{f(z)}{\cos(\pi z)} \frac{\mathrm{d}z}{2i}. \tag{6.3.61}$$

These counter-clockwise closed contours wrap around the real axis, where we have assumed f itself has no singularities.

Problem 6.23. Verify that (location of the singularities, associated residues) of $\pi \cot(\pi z)$, $\pi \csc(\pi z)$, $-\pi \tan(\pi z)$, and $-\pi \sec(\pi z)$ are respectively $(n, 1)$, $(n, (-)^n n)$, $(n+1/2, 1)$ and $(n, (-)^n)$. Here $n = 0, \pm 1, \pm 2, \pm 3, \dots$ is an arbitrary integer. Also investigate the behavior of $|\cot(\pi z)|$, $|\csc(\pi z)|$, $|\tan(\pi z)|$ and $|\sec(\pi z)|$ for large $|z|$. $\square$

If the infinite series does not run over all integers, it is sometimes possible to massage it into such a form. In any case, let us consider a simple example, for $a \in \mathbb{R}$ non-integer:

$$S(a) \equiv \sum_{n=-\infty}^{+\infty} \frac{1}{n^2 + a^2}. \tag{6.3.62}$$

We may exploit eq. (6.3.58) by identifying $F(z) = 1/(z^2+a^2)$ and choosing a contour that wraps around the real line but does not contain the singularities at either $\pm i|a|$. Now, by expanding the contour to an infinite radius circle but making the appropriate detours – coming up or down to skirt the poles at $\pm i|a|$ before returning to infinity – we may see that

$$S(a) = \oint \frac{(z-i|a|)^{-1}(z+i|a|)^{-1}}{\tan(\pi z)} \frac{\mathrm{d}z}{2i} \tag{6.3.63}$$

$$= -\sum_{z=\pm ia} \mathrm{Res} \frac{\pi}{(z-i|a|)(z+i|a|)\tan(\pi z)} \tag{6.3.64}$$

$$= -\frac{\pi}{2ia\tan(i\pi a)} + \frac{\pi}{2ia\tan(-i\pi a)} = \frac{\pi}{a\tanh(\pi a)}. \tag{6.3.65}$$

Problem 6.24. Mellin-Barnes Justify the following Mellin-Barnes representation:

$$(X+Y)^{-\lambda} = \frac{1}{\Gamma(\lambda)} \int_{-i\infty}^{+i\infty} \frac{\mathrm{d}z}{2\pi i} \Gamma(z+\lambda)\Gamma(-z) \frac{Y^z}{X^{z+\lambda}}. \tag{6.3.66}$$

Here, all the poles of $\Gamma(z+\lambda)$ lie to the left of the otherwise vertical integration contour; whereas all the poles of $\Gamma(-z)$ lie to its right.

Iterate eq. (6.3.66) to obtain

$$\left(\sum_{a=1}^{L} u_a\right)^{-\sigma} = \frac{1}{\Gamma[\sigma] u_L^\sigma} \prod_{a=1}^{L-1} \left(\int_{-i\infty}^{i\infty} \frac{\mathrm{d}s_a}{2\pi i} \Gamma[-s_a] \left(\frac{u_a}{u_L}\right)^{s_a}\right) \Gamma\left[\sum_{b=1}^{L-1} s_b + \sigma\right]. \tag{6.3.67}$$

The vertical contour for the s_i integral, where $i \in \{1, \dots, L-1\}$, lies to the right of the poles of $\Gamma(s_i + s_{i-1} + \dots + s_1 + \sigma)$; namely, s_i[left poles] $= -n - \sigma - s_1 - s_2 - \dots - s_{i-1}$, $n = 0, 1, 2, \dots$. While it lies to the left of the poles of $\Gamma[-s_i]$; i.e., s_i[right poles] $= n$, $n = 0, 1, 2, \dots$.

These formulas has applications in (quantum) field theory calculations.

Hints: Assume the infinite arc joining the positive and negative Im ends of the vertical contour do not contribute to the integral – this may be proved by studying the asymptotic behavior of the $\Gamma-$function. Then show that eq. (6.3.66) is equivalent to eq. (6.2.29). $\square$

Principal Values and Hilbert Transforms When integrating $1/x$ over the real line, an ambiguity occurs in the neighborhood of $x \approx 0$. Depending on how we skirt the singularity, different answers are obtained. The principal value, or Pr, is defined as follows. For $a < c < b$,

$$\mathrm{Pr}\int_a^b \frac{f(x)}{x-c}\mathrm{d}x \equiv \lim_{\epsilon\to 0}\left(\int_a^{c-\epsilon} + \int_{c+\epsilon}^b\right)\frac{f(x)}{x-c}\mathrm{d}x. \tag{6.3.68}$$

If the upper and lower limits goes to $+\infty$ and $-\infty$ respectively,

$$\mathrm{Pr}\int_{-\infty}^{+\infty} \frac{f(x)}{x-c}\mathrm{d}x \equiv \lim_{\substack{R\to+\infty \\ \epsilon\to 0}}\left(\int_{-R}^{c-\epsilon} + \int_{c+\epsilon}^R\right)\frac{f(x)}{x-c}\mathrm{d}x. \tag{6.3.69}$$

As a simple example, let us compute

$$I(a) \equiv \mathrm{Pr}\int_{\mathbb{R}} \frac{\mathrm{d}z}{z-a}, \tag{6.3.70}$$

where $a \in \mathbb{R}$. Let us consider the $I(a)$; plus a clockwise infinitesimal semi-circular contour protruding into the positive Im portion of the complex $z-$plane; plus an infinitely large clockwise semi-circular one joining $+\infty$ to $-\infty$ on the Re $z-$line. This forms a closed contour C, which yields the following result via taking the residue at $z = a$:

$$\oint_{\mathrm{CW}} \frac{\mathrm{d}z}{z-a} = -2\pi i = \left(\mathrm{Pr}\int + \int_{\cap} + \int_{\bigcup}\right)\frac{\mathrm{d}z}{z-a}. \tag{6.3.71}$$

The integral along infinitesimal semi-circle may be computed as

$$\int_{\cap} \frac{\mathrm{d}z}{z-a}\bigg|_{z-a=\epsilon e^{i\theta},\pi\leq\theta\leq 0} = \int_{\pi}^{0} \frac{\epsilon e^{i\theta} i\mathrm{d}\theta}{\epsilon e^{i\theta}} = -i\pi. \tag{6.3.72}$$

Similarly, the infinitely large semi-circle yields

$$\int_{\bigcup} \frac{\mathrm{d}z}{z}\bigg|_{\lim_{R\to\infty} z-a=Re^{i\theta},\theta:0\to-\pi} = \lim_{R\to\infty}\int_0^{-\pi} \frac{Re^{i\theta} i\mathrm{d}\theta}{Re^{i\theta}} = -i\pi. \tag{6.3.73}$$

Altogether,

$$\mathrm{Pr}\int \frac{\mathrm{d}z}{z-a} = -2\pi i - \left(\int_{\cap} + \int_{\bigcup}\right)\frac{\mathrm{d}z}{z-a} = 0; \tag{6.3.74}$$

a fact that could be deduced readily by observing that $1/(z-a)$ is odd under $(z-a) \to -(z-a)$.

We will now employ the Pr value integral to relate the real and imaginary parts of a complex function $f(x)$ evaluated on the real line ($x \in \mathbb{R}$), when f itself is analytic and bounded in the upper half plane. In particular, let us first consider

$$0 = \left(\mathrm{Pr}\int_{-\infty}^{+\infty} + \int_{\cap_1} + \int_{\cap_2}\right)\frac{f(z)}{z-x}\frac{\mathrm{d}z}{2\pi i}. \tag{6.3.75}$$

Like the example above, the $\int_{\cap_1}$ is an infinitesimal semi-circle right above $z = x$, where x itself lies on the real line; whereas $\int_{\cap_2}$ refers to the infinitely large semi-circle $\lim_{R\to\infty} z = Re^{i\theta}$ for $0 \leq \theta \leq \pi$. As long as f itself is bounded, the integral $\int_{\cap_2}$ over the infinite semi-circle is zero, and we are left with

$$\text{Pr}\int_{-\infty}^{+\infty} \frac{f(z)}{z-x}\frac{\mathrm{d}z}{2\pi i} = -\lim_{\epsilon\to 0^+}\int_{\pi}^{0} \frac{f(x+\epsilon e^{i\theta})}{\epsilon e^{i\theta}}\frac{\epsilon e^{i\theta} i \mathrm{d}\theta}{2\pi i} \tag{6.3.76}$$

$$= \frac{f(x)}{2}. \tag{6.3.77}$$

If $f = u + iv$, where u and v are real,

$$u + iv = \text{Pr}\int_{-\infty}^{+\infty} \frac{v(z) - iu(z)}{z-x}\frac{\mathrm{d}z}{\pi}. \tag{6.3.78}$$

Therefore,

$$\text{Re } f(x) = \text{Pr}\int_{-\infty}^{+\infty} \frac{\text{Im } f(z)}{z-x}\frac{\mathrm{d}z}{\pi} \tag{6.3.79}$$

$$\text{Im } f(x) = -\text{Pr}\int_{-\infty}^{+\infty} \frac{\text{Re } f(z)}{z-x}\frac{\mathrm{d}z}{\pi}. \tag{6.3.80}$$

These are known as Hilbert transform pairs.

Suppose $\widetilde{F}(\omega)$ is bounded and analytic on the upper half $\omega-$plane; and suppose further $\widetilde{F}(\omega)^* = \widetilde{F}(-\omega)$ for real ω – this happens, for instance, when $\widetilde{F}(\omega) = u(\omega) + iv(\omega)$ is the frequency transform of a real function of time – then we may step back a little and recognize

$$\widetilde{F}(\omega > 0) = -i\text{Pr}\left(\int_{-\infty}^{0} + \int_{0}^{+\infty}\right) \frac{\widetilde{F}(x)}{x-\omega}\frac{\mathrm{d}x}{\pi} \tag{6.3.81}$$

$$= -i\text{Pr}\left(\int_{0}^{\infty} \frac{\widetilde{F}(-x)}{-x-\omega}\frac{\mathrm{d}x}{\pi} + \int_{0}^{+\infty} \frac{\widetilde{F}(x)}{x-\omega}\frac{\mathrm{d}x}{\pi}\right) \tag{6.3.82}$$

$$= -i\text{Pr}\int_{0}^{\infty} \left(-\frac{u(x) - iv(x)}{x+\omega} + \frac{u(x) + iv(x)}{x-\omega}\right)\frac{\mathrm{d}x}{\pi} \tag{6.3.83}$$

$$u(\omega) + iv(\omega) = 2\text{Pr}\int_{0}^{\infty} \left(\frac{-i\omega\cdot u(x)}{x^2-\omega^2} + \frac{x\cdot v(x)}{x^2-\omega^2}\right)\frac{\mathrm{d}x}{\pi}. \tag{6.3.84}$$

That is, for real $F(t)$, its frequency transform obeys the following Hilbert transforms.

$$\text{Re } \widetilde{F}(\omega > 0) = \frac{2}{\pi}\text{Pr}\int_{0}^{\infty} \frac{x\cdot\text{Im } \widetilde{F}(x)}{x^2-\omega^2}\mathrm{d}x \tag{6.3.85}$$

$$\text{Im } \widetilde{F}(\omega > 0) = -\frac{2\omega}{\pi}\text{Pr}\int_{0}^{\infty} \frac{\text{Re } \widetilde{F}(x)}{x^2-\omega^2}\mathrm{d}x. \tag{6.3.86}$$

This allows the results to refer only to positive (and, hence, physical) frequencies ω.

Problem 6.25. Pr and Dirac's δ-function Notice, when we skirt the pole at $x = \omega$ on the real line by making an infinitesimal semi-circle above it, we pick up (negative) half the residue:

$$\lim_{\epsilon \to 0} \int_{\theta=\pi}^{\theta=0} \left. \frac{f(z)}{z-\omega} \right|_{z-\omega=\epsilon e^{i\theta}} \mathrm{d}z = -i\pi f(\omega). \tag{6.3.87}$$

Keeping this in mind, explain the following identities, for $x \in \mathbb{R}$:

$$\frac{1}{x - \omega \pm i0^+} = \mathrm{Pr}\frac{1}{x-\omega} \mp i\pi\delta\left(x-\omega\right), \tag{6.3.88}$$

where $\delta(\dots)$ is the Dirac delta function.

Hints: These identities are meant to be understood under an integral sign. Moreover, $x - \omega \pm i0^+$ means the pole(s) is at $x = \omega \mp i0^+$; i.e., the upper sign is displaced slightly downwards and the lower sign upwards on the complex $\omega-$plane. □

6.4 Branch Points, Branch Cuts

Branch points and Riemann sheets A branch point of a function $f(z)$ is a point z_0 on the complex plane such that going around z_0 in an infinitesimally small circle does not give you back the same function value. That is,

$$f\left(z_0 + \epsilon \cdot e^{i\theta}\right) \neq f\left(z_0 + \epsilon \cdot e^{i(\theta+2\pi)}\right), \qquad 0 < \epsilon \ll 1. \tag{6.4.1}$$

Example I One example is the power z^α, for α non-integer. Zero is a branch point because, for $0 < \epsilon \ll 1$, we may considering circling it $n \in \mathbb{Z}^+$ times.

$$(\epsilon e^{2\pi n i})^\alpha = \epsilon^\alpha e^{2\pi n \alpha i} \neq \epsilon^\alpha. \tag{6.4.2}$$

If $\alpha = 1/2$, then circling zero twice would bring us back to the same function value. If $\alpha = 1/m$, where m is a positive integer, we would need to circle zero m times to get back to the same function value. What this is teaching us is that, to define the function $f(z) = z^{1/m}$ properly, we need m "Riemann sheets" of the complex plane. To see this, we first define a cut line along the positive real line and proceed to explore the function f by sampling its values along a continuous line. If we start from a point slightly above the real axis, $z^{1/m}$ there is defined as $|z|^{1/m}$, where the positive root is assumed here. As we move around the complex plane, let us use polar coordinates to write $z = \rho e^{i\theta}$; once θ runs beyond 2π, i.e., once the contour circles around the origin more than one revolution, we exit the first complex plane and enter the second. For example, when z is slightly above the real axis on the second sheet, we define $z = |z|^{1/m} e^{i2\pi/m}$; and anywhere else on the second sheet we have $z = |z|^{1/m} e^{i(2\pi/m)+i\theta}$, where θ is still measured with respect to the real axis. We can continue this process, circling the origin, with each increasing counterclockwise revolution taking us from one sheet to the next. On the nth sheet our function reads $z = |z|^{1/m} e^{i(2\pi n/m)+i\theta}$. It is the mth sheet that needs to be joined with the very first sheet, because by the mth sheet we have covered all the m solutions of what we mean by taking the mth root of a complex number. (If we had explored the function using a clockwise path instead, we'd migrated from the first sheet to the mth sheet, then to the

$(m-1)$th sheet and so on.) Finally, if α were not rational – it is not the ratio of two integers – we would need an infinite number of Riemann sheets to fully describe z^α as a complex differentiable function of z.

The presence of the branch cut(s) is necessary because we need to join one Riemann sheet to the next, so as to construct an analytic function mapping the full domain back to the complex plane. However, as long as one Riemann sheet is joined to the next so that the function is analytic across this boundary, and as long as the full domain is mapped properly onto the complex plane, the location of the branch cut(s) is arbitrary. For example, for the $f(z) = z^\alpha$ case above, as opposed to the real line, we can define our branch cut to run along the radial line $\{\rho e^{i\theta_0} | \rho \geq 0\}$ for any $0 < \theta_0 \leq 2\pi$. All we are doing is re-defining where to join one sheet to another, with the nth sheet mapping one copy of the complex plane $\{\rho e^{i(\theta_0+\varphi)} | \rho \geq 0, 0 \leq \varphi < 2\pi\}$ to $\{|z|^\alpha e^{i\alpha(\theta_0+\varphi)} | \rho \geq 0, 0 \leq \varphi < 2\pi\}$. Of course, in this new definition, the $2\pi - \theta_0 \leq \varphi < 2\pi$ portion of the nth sheet would have belonged to the $(n+1)$th sheet in the old definition – but, taken as a whole, the collection of all relevant Riemann sheets still cover the same domain as before.

Example II ln is another example. You already know the answer but let us work out the complex derivative of $\ln z$. Because $e^{\ln z} = z$, we have

$$(e^{\ln z})' = e^{\ln z} \cdot (\ln z)' = z \cdot (\ln z)' = 1. \tag{6.4.3}$$

This implies,

$$\frac{\mathrm{d} \ln z}{\mathrm{d}z} = \frac{1}{z}, \qquad z \neq 0, \tag{6.4.4}$$

which in turn says $\ln z$ is analytic away from the origin. We may now consider making m infinitesimal circular trips around $z = 0$.

$$\ln(\epsilon e^{i2\pi m}) = \ln(\epsilon e^{i2\pi m}) = \ln \epsilon + i2\pi m \neq \ln \epsilon. \tag{6.4.5}$$

Just as for $f(z) = z^\alpha$ when α is irrational, it is in fact not possible to return to the same function value – the more revolutions you take, the further you move in the imaginary direction. $\ln(z)$ for $z = x + iy$ actually maps the mth Riemann sheet to a horizontal band on the complex plane, lying between $2\pi(m-1) \leq \operatorname{Im} \ln(z) \leq 2\pi m$.

Breakdown of Laurent series To understand the need for multiple Riemann sheets further, it is instructive to go back to our discussion of the Laurent series using an annulus around the isolated singular point, which lead up to eq. (6.2.34). For both $f(z) = z^\alpha$ and $f(z) = \ln(z)$, the branch point is at $z = 0$. If we had used a single complex plane, with say a branch cut along the positive real line, $f(z)$ would not even be continuous – let alone analytic – across the $z = x > 0$ line: $f(z = x + i0^+) = x^\alpha \neq f(z = x - i0^+) = x^\alpha e^{i2\pi\alpha}$, for instance. Therefore the derivation there would not go through, and a Laurent series for either z^α or $\ln z$ about $z = 0$ cannot be justified. But as far as integration is concerned, provided we keep track of how many times the contour wraps around the origin – and therefore how many Riemann sheets have been transversed – both z^α and $\ln z$ are analytic once all relevant Riemann sheets have been taken into account. For example, let us do $\oint_C \ln(z) \mathrm{d}z$, where C begins from the point $z_1 \equiv r_1 e^{i\theta_1}$ and loops around the origin n times and ends on the point $z_2 \equiv r_2 e^{i\theta_2 + i2\pi n}$ for $(n \geq 1)$-integer. Across these n sheets and away from $z = 0$, $\ln(z)$ is analytic. We may therefore invoke Cauchy's theorem in eq. (6.2.5) to deduce the result depends on the path only through its 'winding number' n. Because $(z\ln(z)-z)' = \ln z$,

$$\int_{z_1}^{z_2} \ln(z)\mathrm{d}z = r_2 e^{i\theta_2}\left(\ln r_2 + i(\theta_2 + 2\pi n) - 1\right) - r_1 e^{i\theta_1}\left(\ln r_1 + i\theta_1 - 1\right). \qquad (6.4.6)$$

Likewise, for the same integration contour C,

$$\int_{z_1}^{z_2} z^\alpha \mathrm{d}z = \frac{r_2^{\alpha+1}}{\alpha+1} e^{i(\alpha+1)(\theta_2+2\pi n)} - \frac{r_1^{\alpha+1}}{\alpha+1} e^{i(\alpha+1)\theta_1}. \qquad (6.4.7)$$

Branches On the other hand, the purpose of defining a branch cut, is that it allows us to define a single-valued function on a *single* complex plane – a branch of a multivalued function – as long as we agree never to cross over this cut when moving about on the complex plane. For example, a branch cut along the negative real line means $\sqrt{z} = \sqrt{r}e^{i\theta}$ with $-\pi < \theta < \pi$; you don't pass over the cut line along $z < 0$ when you move around on the complex plane.

Another common example is given by the following branch of $\sqrt{z^2-1}$:

$$\sqrt{z+1}\sqrt{z-1} = \sqrt{r_1 r_2} e^{i(\theta_1+\theta_2)/2}, \qquad (6.4.8)$$

where $z+1 \equiv r_1 e^{i\theta_1}$ and $z-1 \equiv r_2 e^{i\theta_2}$; and $\sqrt{r_1 r_2}$ is the positive square root of $r_1 r_2 > 0$. By circling the branch point you can see the function is well defined if we cut along $-1 < z < +1$, because $(\theta_1+\theta_2)/2$ goes from 0 to $(\theta_1+\theta_2)/2 = 2\pi$.[52] Otherwise, if the cut is defined as $z < -1$ (on the negative real line) together with $z > 1$ (on the positive real line), the branch points at $z = \pm 1$ cannot be circled and the function is still well defined and single-valued.

Yet another example is given by the Legendre function

$$Q_0(z) = \ln\left[\frac{z+1}{z-1}\right]. \qquad (6.4.9)$$

The branch points, where the argument of the ln goes to zero, is at $z = \pm 1$. $Q_\nu(z)$ is usually defined with a cut line along $-1 < z < +1$ on the real line. Let's circle the branch points counterclockwise, with

$$z+1 \equiv r_1 e^{i\theta_1} \qquad \text{and} \qquad z-1 \equiv r_2 e^{i\theta_2} \qquad (6.4.10)$$

as before. Then,

$$Q_0(z) = \ln\left[\frac{z+1}{z-1}\right] = \ln\frac{r_1}{r_2} + i\left(\theta_1 - \theta_2\right). \qquad (6.4.11)$$

After one closed loop, we go from $\theta_1 - \theta_2 = 0 - 0 = 0$ to $\theta_1 - \theta_2 = 2\pi - 2\pi = 0$; there is no jump. When x lies on the real line between -1 and 1, $Q_0(x)$ is then defined as

$$Q_0(x) = \frac{1}{2}Q_0(x+i0^+) + \frac{1}{2}Q_0(x-i0^+), \qquad (6.4.12)$$

[52] Arfken et al. goes through various points along this circling-the-($z = \pm 1$) process, but the main point is that there is no jump after a complete circle, unlike what you'd get circling the branch point of, say $z^{1/3}$. On the other hand, you may want to use the $z+1 \equiv r_1 e^{i\theta_1}$ and $z-1 \equiv r_2 e^{i\theta_2}$ parametrization here and understand how many Riemann sheets it would take define the whole $\sqrt{z^2-1}$.

where the $i0^+$ in the first term on the right means the real line is approached from the upper half plane and the second term means it is approached from the lower half plane. What does that give us? Approaching from above means $\theta_1 = 0$ and $\theta_2 = \pi$; so $\ln(z + i0^+ + 1)/(z + i0^+ - 1) = \ln|(z+1)/(z-1)| - i\pi$. Approaching from below means $\theta_1 = 2\pi$ and $\theta_2 = \pi$; therefore $\ln(z - i0^+ + 1)/(z - i0^+ - 1) = \ln|(z+1)/(z-1)| + i\pi$. Hence the average of the two yields

$$Q_0(x) = \ln\left[\frac{1+x}{1-x}\right], \qquad -1 < x < +1. \tag{6.4.13}$$

because the imaginary parts cancel while $|z+1| = x+1$ and $|z-1| = 1-x$ in this region.

Example I Let us exploit the following branch of natural log

$$\ln z = \ln r + i\theta, \qquad z = re^{i\theta}, \qquad 0 \leq \theta < 2\pi \tag{6.4.14}$$

to evaluate the integral encountered in eq. (6.3.46).

$$I \equiv \int_0^\infty \frac{(\ln x)^2}{1+x^2}\mathrm{d}x = \frac{\pi^3}{8}. \tag{6.4.15}$$

To begin we will actually consider

$$I' \equiv \lim_{\substack{R\to\infty \\ \epsilon\to 0}} \oint_{C_1+C_2+C_3+C_4} \frac{(\ln z)^2}{1+z^2}\mathrm{d}z, \tag{6.4.16}$$

where C_1 runs over $z \in (-\infty, -\epsilon]$ (for $0 \leq \epsilon \ll 1$), C_2 over the infinitesimal semi-circle $z = \epsilon e^{i\theta}$ (for $\theta \in [\pi, 0]$), C_3 over $z \in [\epsilon, +\infty)$ and C_4 over the (infinite) semi-circle $Re^{i\theta}$ (for $R \to +\infty$ and $\theta \in [0, \pi]$).

First, we show that the contribution from C_2 and C_4 are zero once the limits $R \to \infty$ and $\epsilon \to 0$ are taken.

$$\begin{aligned} \left|\lim_{\epsilon\to 0}\int_{C_2}\frac{(\ln z)^2}{1+z^2}\mathrm{d}z\right| &= \left|\lim_{\epsilon\to 0}\int_\pi^0 i\mathrm{d}\theta\epsilon e^{i\theta}\frac{(\ln\epsilon + i\theta)^2}{1+\epsilon^2 e^{2i\theta}}\right| \\ &\leq \lim_{\epsilon\to 0}\int_0^\pi \mathrm{d}\theta\epsilon|\ln\epsilon + i\theta|^2 = 0. \end{aligned} \tag{6.4.17}$$

and

$$\begin{aligned} \left|\lim_{R\to\infty}\int_{C_4}\frac{(\ln z)^2}{1+z^2}\mathrm{d}z\right| &= \left|\lim_{R\to\infty}\int_0^\pi i\mathrm{d}\theta Re^{i\theta}\frac{(\ln R + i\theta)^2}{1+R^2 e^{2i\theta}}\right| \\ &\leq \lim_{R\to\infty}\int_0^\pi \mathrm{d}\theta|\ln R + i\theta|^2/R = 0. \end{aligned} \tag{6.4.18}$$

Moreover, I' can be evaluated via the residue theorem; within the closed contour, the integrand blows up at $z = i$.

$$I' \equiv 2\pi i \lim_{\substack{R\to\infty \\ \epsilon\to 0}} \oint_{C_1+C_2+C_3+C_4} \frac{(\ln z)^2}{(z+i)(z-i)}\frac{\mathrm{d}z}{2\pi i}$$

$$= 2\pi i \frac{(\ln i)^2}{2i} = \pi(\ln(1) + i(\pi/2))^2 = -\frac{\pi^3}{4}. \tag{6.4.19}$$

This means the sum of the integral along C_1 and C_3 yields $-\pi^3/4$. If we use polar coordinates along both C_1 and C_2, namely $z = re^{i\theta}$,

$$\int_\infty^0 \mathrm{d}re^{i\pi}\frac{(\ln r + i\pi)^2}{1 + r^2 e^{i2\pi}} + \int_0^\infty \frac{(\ln r)^2}{1+r^2}\mathrm{d}r = -\frac{\pi^3}{4} \tag{6.4.20}$$

$$\int_0^\infty \mathrm{d}r\frac{2(\ln r)^2 + i2\pi\ln r - \pi^2}{1+r^2} = -\frac{\pi^3}{4} \tag{6.4.21}$$

We may equate the real and imaginary parts of both sides. The imaginary one, in particular, says

$$\int_0^\infty \mathrm{d}r\frac{\ln r}{1+r^2} = 0, \tag{6.4.22}$$

while the real part now hands us

$$\begin{aligned} 2I &= \pi^2 \int_0^\infty \frac{\mathrm{d}r}{1+r^2} - \frac{\pi^3}{4} \\ &= \pi^2 \left[\arctan(r)\right]_{r=0}^{r=\infty} - \frac{\pi^3}{4} = \frac{\pi^3(2-1)}{4} = \frac{\pi^3}{4} \end{aligned} \tag{6.4.23}$$

We have managed to solve for the integral I

Problem 6.26. If x is a real number, justify the identity

$$\ln(x + i0^+) = \ln|x| + i\pi\Theta(-x), \tag{6.4.24}$$

where Θ is the step function. Hence,

$$\partial_x \ln(x + i0^+) = \frac{1}{x} - i\pi\delta(x). \tag{6.4.25}$$

These identities occurs in classical and quantum field theories. □

Example II Consider the integral, where $0 < \mathrm{Re}\ z < 1$,

$$I(z) \equiv \int_0^\infty \frac{t^{z-1}}{t+1}\mathrm{d}t. \tag{6.4.26}$$

This integration from 0 to ∞ may be viewed as a contour running just above the positive real line. If we rotate the contour by $2\pi - 0^+$, notice that

$$\mathrm{d}t\frac{t^{z-1}}{t+1} \to e^{i2\pi}\mathrm{d}t\frac{e^{2\pi i(z-1)}t^{z-1}}{t \cdot e^{i2\pi} + 1}. \tag{6.4.27}$$

Let us now let C be the contour that runs from 0 to ∞ just slightly above the real line, then counter-clockwise around the infinite radius circle centered at the origin, then from ∞ back to

0 just slightly below the real line. Because the magnitude of the integral over the infinite radius circle is bounded by

$$\lim_{R\to\infty}\int_0^{2\pi}\left|Rie^{i\theta}\mathrm{d}\theta\frac{R^{z-1}e^{i(z-1)\theta}}{Re^{i\theta}+1}\right| \leq \lim_{R\to\infty}\int_0^{2\pi}\mathrm{d}\theta\left|R^{iz_{\mathrm{I}}-1+z_{\mathrm{R}}}e^{-z_{\mathrm{I}}\theta+i(z_{\mathrm{R}}-1)\theta}\right| \tag{6.4.28}$$

$$\to \lim_{R\to 0} R^{z_{\mathrm{R}}-1}=0, \tag{6.4.29}$$

since $-1<z_{\mathrm{R}}-1<0$. Therefore,

$$\int_C \frac{t^{z-1}}{t+1}\mathrm{d}t = (1-e^{2\pi z i})\int_0^\infty \frac{t^{z-1}}{t+1}\mathrm{d}t \tag{6.4.30}$$

$$= 2\pi i e^{i(z-1)\pi}. \tag{6.4.31}$$

Because $1-e^{2\pi z i}=e^{\pi z i}(-)2i\sin(\pi z)$, we must therefore have

$$\int_0^\infty \frac{t^{z-1}}{t+1}\mathrm{d}t = \frac{2\pi i e^{i(z-1)\pi}}{e^{\pi z i}(-)2i\sin(\pi z)} = \frac{\pi}{\sin(\pi z)}. \tag{6.4.32}$$

There is an application of the formula in eq. (6.4.32) in the derivation of the Gamma function identity in eq. (6.2.47). Following Lebedev [5] we employ the following trick, by first focusing on the range $0<\mathrm{Re}\ z<1$,

$$\Gamma(z)\Gamma(1-z)=\int_0^{+\infty}t^{z-1}e^{-t}\mathrm{d}t\cdot\int_0^{+\infty}t'^{-z}e^{-t'}\mathrm{d}t'. \tag{6.4.33}$$

Define $u\equiv t+t'$ and $v\equiv t/t'$; one would find

$$t=\frac{uv}{1+v}, \qquad \mathrm{d}t=\frac{v(1+v)\mathrm{d}u+u\mathrm{d}v}{(1+v)^2}; \tag{6.4.34}$$

$$t'=\frac{u}{1+v}, \qquad \mathrm{d}t'=\frac{(1+v)\mathrm{d}u-u\mathrm{d}v}{(1+v)^2}; \tag{6.4.35}$$

and

$$\Gamma(z)\Gamma(1-z)=\int_0^{+\infty}e^{-u}\mathrm{d}u\cdot\int_0^\infty\frac{v^{z-1}}{1+v}\mathrm{d}v=\int_0^\infty\frac{v^{z-1}}{1+v}\mathrm{d}v. \tag{6.4.36}$$

We recover eq. (6.2.47) once we invoke eq. (6.4.32) and the principle of analytic continuation.

Problem 6.27. Problem from Arfken et al. [18] For $-1<a<1$, show that

$$\int_0^\infty \mathrm{d}x\frac{x^a}{(x+1)^2}=\frac{\pi a}{\sin(\pi a)}. \tag{6.4.37}$$

Hint: Define a branch cut along the positive real line, and follow a similar strategy as the one employed to tackle the integral in eq. (6.4.32). □

Problem 6.28. By converting the following real integral into a complex closed loop one, explain why

$$I(\eta; a, b) \equiv \int_a^b \frac{\mathrm{d}z}{(\eta - z)\sqrt{(b-z)(z-a)}}, \tag{6.4.38}$$

for positive square root $\sqrt{\cdot}$ and positive real numbers $b > a > 0$ is

$$I(\eta; a, b) = \frac{\pi}{\sqrt{(\eta - a)(\eta - b)}} \tag{6.4.39}$$

when $\eta > b > a$; and

$$I(\eta; a, b) = -\frac{\pi}{\sqrt{(a - \eta)(b - \eta)}} \tag{6.4.40}$$

when $\eta < a < b$. Hint: First choose the branch cut to run along a straight line from a to b; and consider the closed loop integral running just above it from $a \to b$ and then running just below it from $b \to a$. □

Problem 6.29. Two Zeroes Define the natural log as $\ln z = \ln |z| + i\theta$, where $z = |z|e^{i\theta}$ and $-\pi < \theta < \pi$. Proof that, for $\alpha, \beta > 0$,

$$f(z) \equiv (z + \alpha)^2 - \beta \ln z \tag{6.4.41}$$

has precisely two zeroes on the complex plane, with the branch cut defined for the natural log above. Prove that these two zeroes are distinct except when the minimum of $f(z)$ lies exactly on the positive real line. Hint: You should be able to show that, for $z = |z|e^{i\theta}$, $(f'(z)/f(z))\mathrm{d}z \to 2i\mathrm{d}\theta$ as $|z| \to \infty$. Do not forget to show that the integral along the branch cut does not contribute to the counting of zeroes. □

6.5 *Mellin Transform

[53]In eq. (5.8.25) we introduced the Mellin transform of a function $f(r)$ defined on the positive real line $r \geq 0$:

$$\widetilde{f}(z) \equiv \int_0^\infty f(r') r'^{z-1} \mathrm{d}r', \tag{6.5.1}$$

$$z = s + 1 + i\nu \in \mathbb{C}. \tag{6.5.2}$$

We shall assume this integral converges on some open strip $a < (\mathrm{Re}(z) = s + 1) < b$ on the complex $z-$plane; whereas $\nu = \mathrm{Im}(z)$. In fact, $\widetilde{f}(z)$ is an analytic function of the complex variable z; and $\widetilde{f}(z) \to 0$ as $\mathrm{Im}(z) \to \pm\infty$.

Previously, we obtained its inverse by linear algebraic arguments, by first deriving the basis functions which are simultaneous eigenstates of the (Hermitian) generator of the dilatation operator.[54] Here, we will take a more pragmatic approach, by multiplying both sides of eq. (6.5.1)

[53]I benefited greatly from *The Handbook of Transforms* [8] while preparing this section.

[54]Quick recap: The real number s is defined as $D_s(\lambda)\,|r\rangle = \lambda^s\,|\lambda \cdot r\rangle$; while $D_s(e^\epsilon) = e^{-i\epsilon \cdot \mathcal{E}}$ with ν being the (real) eigenvalue of the Hermitian generator $\mathcal{E}_s$.

with $\exp(-i\nu \ln r)$, and integrating over ν.

$$\int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} \widetilde{f}(z) e^{-i\nu \ln r} = \int_0^\infty \mathrm{d}r' f(r') \int_{\mathbb{R}} \frac{\mathrm{d}\nu}{2\pi} r'^s e^{i\nu(\ln r' - \ln r)} \tag{6.5.3}$$

$$= r^{s+1} f(r). \tag{6.5.4}$$

We may transform the integration variable ν to z, so that $\mathrm{d}\nu = \mathrm{d}z/i$, $e^{-i\nu \ln r} = e^{-z \ln r} e^{(s+1)\ln r}$, and the path of integration now runs along the vertical line $\mathrm{Re}(z) + i\mathbb{R}$. We have arrived at a contour integral representation of the inverse Mellin transform in eq. (5.8.19):

$$f(r \geq 0) = \int_{\mathrm{Re}(z)-i\infty}^{\mathrm{Re}(z)+i\infty} \frac{\mathrm{d}z}{2\pi i} \frac{\widetilde{f}(z)}{r^z}. \tag{6.5.5}$$

If we understand the analytic structure of $\widetilde{f}(z)$ – for e.g., where its poles lie – this vertical contour may be distorted appropriately for easier evaluation of $f(r)$.

Range of Re(z) in Mellin: An Example from [8] The range of $\mathrm{Re}(z)$ where the Mellin transform converges is an *crucial* part of its definition. As the *Handbook of Transforms* [8] explains, it is possible to obtain the same Mellin transform $\widehat{f}(z)$ from different starting $f(r)$s; i.e., the inverse Mellin transform is seemingly not unique if one is not careful about keeping track of $\mathrm{Re}(z)$'s range of validity. A simple example is to compare, for $r_0 > 0$ and $x \in \mathbb{R}$,

$$\widetilde{f}_1(z) \equiv \int_0^\infty \Theta(r - r_0) r^x \cdot r^{z-1} \mathrm{d}r \tag{6.5.6}$$

$$= -\frac{r_0^{x+z}}{x+z}, \qquad \mathrm{Re}(z) < -x \tag{6.5.7}$$

versus

$$\widetilde{f}_2(z) \equiv \int_0^\infty (\Theta(r) - \Theta(r - r_0)) r^x \cdot r^{z-1} \mathrm{d}r \tag{6.5.8}$$

$$= -\frac{r_0^{x+z}}{x+z}, \qquad \mathrm{Re}(z) > -x. \tag{6.5.9}$$

The range of validity of $\mathrm{Re}(z)$ for $\widetilde{f}_{1,2}$ do not overlap at all; yet, they are the same expression. Let us attempt to re-construct the integrands of equations (6.5.6) and (6.5.8),

$$f_1(r) \equiv \Theta(r - r_0) r^x, \tag{6.5.10}$$

$$f_2(r) \equiv (\Theta(r) - \Theta(r - r_0)) r^x. \tag{6.5.11}$$

For $f_1(r)$, eq. (6.5.5) applied to eq. (6.5.6) now reads

$$f_1(r) = \int_{\substack{\mathrm{Re}(z) < -x \\ \mathrm{Im}(z) \in \mathbb{R}}} \frac{\mathrm{d}\ \mathrm{Im}(z)}{2\pi i} \frac{-r_0^{x+z} \cdot (x+z)^{-1}}{r^z} \tag{6.5.12}$$

$$= -r_0^x \int_{\substack{\mathrm{Re}(z) < -x \\ \mathrm{Im}(z) \in \mathbb{R}}} \frac{\mathrm{d}\ \mathrm{Im}(z)}{2\pi i} \frac{\exp\left(z \ln(r_0/r)\right)}{x+z}. \tag{6.5.13}$$

Whenever $r_0 > r$, the exponential $\exp(z\ln(r_0/r))$ in the second line would blow up as $\text{Re}(z) \to +\infty$. This prompts us to close the contour on the negative $\text{Re}(z)$ part of the complex $z-$plane. But since the original vertical line contour has $\text{Re}(z) < -x$, the integrand is analytic within resulting closed contour and the result is zero. On the other hand, whenever $r > r_0$, the exponential $\exp(z\ln(r_0/r))$ would diverge as $\text{Re}(z) \to -\infty$. Closing the contour on the $\text{Re}(z) > 0$ half of the complex $z-$plane then yields

$$f_1(r) = -\Theta(r - r_0)\text{Res}_{z=-x}\left(-\frac{r_0^x}{x+z}\exp\left(z\ln(r_0/r)\right)\right) \tag{6.5.14}$$

$$= \Theta(r - r_0)r_0^x(r_0/r)^{-x}. \tag{6.5.15}$$

Problem 6.30. Re-construct $f_2(r)$ by taking the inverse transform of eq. (6.5.8). $\square$

Gamma Function Referring to eq. (6.2.44), we see that $\Gamma(z)$ for $\text{Re}(z) > -1$ may be viewed as the Mellin transform of the exponential $f(t) \equiv \exp(-t)$; i.e.,

$$\Gamma(z) \equiv \widetilde{f}(z) = \int_0^\infty f(t)t^{z-1}\mathrm{d}t. \tag{6.5.16}$$

The inversion formula in eq. (6.5.5) says, for $\text{Re}(z) > 0$,

$$e^{-t} = \int_{\text{Re}(z)-i\infty}^{\text{Re}(z)+i\infty}\frac{\mathrm{d}z}{2\pi i}\frac{\Gamma(z)}{t^z}. \tag{6.5.17}$$

Let us verify this statement by direct evaluation of the right hand side. Specifically, let us close the vertical contour, joining its positive and negative Im ends via an infinitely large arc running on the $\text{Re}(z) < 0$ half of the complex $z-$plane. (We will, in due course, justify why this infinite arc contributes zero to the integral.) Recalling eq. (6.2.62),

$$\int_{\text{Re}(z)-i\infty}^{\text{Re}(z)+i\infty}\frac{\mathrm{d}z}{2\pi i}\frac{\Gamma(z)}{t^z} = \sum_{n=0}^{+\infty}\text{Res}_{z=-n}\left(\frac{\Gamma(z)}{t^z}\right) \tag{6.5.18}$$

$$= \frac{(-t)^n}{n!}, \tag{6.5.19}$$

which is the Taylor series expansion of e^{-t}.

Properties of Mellin Transforms Since eq. (6.5.1) defines an analytic function of z, we can take its derivative readily. Recalling $\partial_z r^z = r^z \cdot \ln r$, we deduce for $n = 0, 1, 2, 3, \ldots$,

$$\frac{\mathrm{d}^n\widetilde{f}(z)}{\mathrm{d}z^n} = \int_0^\infty(\ln r)^n f(r)r^{z-1}\mathrm{d}r. \tag{6.5.20}$$

Allowing us to draw the correspondence

$$(\mathrm{d}/\mathrm{d}z)^n \leftrightarrow (\ln r)^n. \tag{6.5.21}$$

Also, eq. (6.5.1) immediately implies

$$\widetilde{f}(z_1 + z_2) = \int_0^\infty r^{z_1}f(r)\cdot r^{z_2-1}\mathrm{d}r \tag{6.5.22}$$

$$= \int_0^\infty r^{z_2} f(r) \cdot r^{z_1 - 1} \mathrm{d}r. \tag{6.5.23}$$

Multiplication of the original function $f(r)$ by a power law (i.e., $f(r) \to r^{z'} f(r)$) amounts to an additive shift of the argument z in the Mellin transform ($\widetilde{f}(z) \to \widetilde{f}(z + z')$).

Next, we perform the Mellin transform of the nth derivative of $f(r)$. For all continuous functions $f(r = 0)$ that are regular at $r = 0$, integration-by-parts allow us to deduce – whenever $\mathrm{Re} z > 1$ –

$$\int_0^\infty \frac{\mathrm{d}^n f(r)}{\mathrm{d}r^n} \cdot r^{z-1} \mathrm{d}r = (-)^n \int_0^\infty f(r) \frac{\mathrm{d}^n}{\mathrm{d}r^n} \cdot r^{z-1} \mathrm{d}r \tag{6.5.24}$$

$$= (-)^n (z-1)(z-2)(z-3) \dots (z-n) \widetilde{f}(z). \tag{6.5.25}$$

The correspondence here is

$$(-)^n (z-1)(z-2)(z-3) \dots (z-n) \leftrightarrow (\mathrm{d}/\mathrm{d}r)^n. \tag{6.5.26}$$

Problem 6.31. Justify the correspondence

$$(-)^n z^n \leftrightarrow (r\mathrm{d}/\mathrm{d}r)^n, \tag{6.5.27}$$

where $n = 0, 1, 2, 3, \dots$. □

YZ: To be continued. Summation of series. Relation to Fourier and Laplace transform. Radial. Variational problem: what's the state that minimizes the variance/uncertainty? Zeta versus Theta functions. Moments. Probability/Statistics. Inverse transform. Mellin-Barnes. Gamma function asymptotics. Fourier and Hankel transforms. Asymptotic expansions.

7 Special and Approximation Techniques in Calculus

Integration is usually much harder than differentiation. Any function $f(x)$ you can build out of powers, logs, trigonometric functions, etc., can usually be readily differentiated.[55] But to integrate a function in closed form you have to know another function $g(x)$ whose derivative yields $f(x)$; that's the essential content of the fundamental theorem of calculus.

$$\int f(x)\mathrm{d}x \stackrel{?}{=} \int g'(x)\mathrm{d}x = g(x) + \text{constant} \tag{7.0.1}$$

Here, I will discuss integration techniques that I feel are not commonly found in standard treatments of calculus. Among them, some techniques will show how to extract approximate answers from integrals. This is, in fact, a good place to highlight the importance of approximation techniques in physics. For example, most of the predictions from quantum field theory – our fundamental framework to describe elementary particle interactions at the highest energies/smallest distances – is based on perturbation theory.

7.1 Gaussian integrals

As a start, let us consider the following "Gaussian" integral:

$$I_G(a) \equiv \int_{-\infty}^{+\infty} e^{-ax^2}\mathrm{d}x, \tag{7.1.1}$$

where $\mathrm{Re}(a) > 0$. (Why is this restriction necessary?) Let us suppose that $a > 0$ for now. Then, we may consider squaring the integral, i.e., the 2-dimensional (2D) case:

$$(I_G(a))^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-ax^2}e^{-ay^2}\mathrm{d}x\mathrm{d}y. \tag{7.1.2}$$

You might think "doubling" the problem is only going to make it harder, not easier. But let us now view (x, y) as Cartesian coordinates on the 2D plane and proceed to change to polar coordinates, $(x, y) = r(\cos\phi, \sin\phi)$; this yields $\mathrm{d}x\mathrm{d}y = \mathrm{d}\phi\mathrm{d}r \cdot r$.

$$(I_G(a))^2 = \int_{-\infty}^{+\infty} e^{-a(x^2+y^2)}\mathrm{d}x\mathrm{d}y = \int_0^{2\pi}\mathrm{d}\phi\int_0^{+\infty}\mathrm{d}r \cdot re^{-ar^2} \tag{7.1.3}$$

The integral over ϕ is straightforward; whereas the radial one now contains an additional r in the integrand – this is exactly what makes the integral do-able.

$$\begin{aligned}(I_G(a))^2 &= 2\pi\int_0^{+\infty}\mathrm{d}r\frac{1}{-2a}\partial_r e^{-ar^2} \\ &= \left[\frac{-\pi}{a}e^{-ar^2}\right]_{r=0}^{r=\infty} = \frac{\pi}{a}\end{aligned} \tag{7.1.4}$$

[55]The ease of differentiation ceases once you start dealing with "special functions"; see, for e.g., here for a discussion on how to differentiate the Bessel function $J_\nu(z)$ with respect to its order ν.

Because e^{-ax^2} is a positive number if a is positive, we know that $I_G(a>0)$ must be a positive number too. Since $(I_G(a))^2 = \pi/a$ the Gaussian integral itself is just the positive square root

$$\int_{-\infty}^{+\infty} e^{-ax^2}\mathrm{d}x = \sqrt{\frac{\pi}{a}}, \qquad \mathrm{Re}(a) > 0. \tag{7.1.5}$$

Because both sides of eq. (7.1.5) can be differentiated readily with respect to a (for $a \neq 0$), by analytic continuation, even though we started out assuming a is positive, we may now relax that assumption and only impose $\mathrm{Re}(a) > 0$. If you are uncomfortable with this analytic continuation argument, you can also tackle the integral directly. Suppose $a = \rho e^{i\delta}$, with $\rho > 0$ and $-\pi/2 < \delta < \pi/2$. Then we may rotate the contour for the x integration from $x \in (-\infty, +\infty)$ to the contour C defined by $z \equiv e^{-i\delta/2}\xi$, where $\xi \in (-\infty, +\infty)$. (The 2 arcs at infinity contribute nothing to the integral – can you prove it?)

$$\begin{aligned} I_G(a) &= \int_{\xi=-\infty}^{\xi=+\infty} e^{-\rho e^{i\delta}(e^{-i\delta/2}\xi)^2}\mathrm{d}(e^{-i\delta/2}\xi) \\ &= \frac{1}{e^{i\delta/2}}\int_{\xi=-\infty}^{\xi=+\infty} e^{-\rho\xi^2}\mathrm{d}\xi \end{aligned} \tag{7.1.6}$$

At this point, since $\rho > 0$ we may refer to our result for $I_G(a>0)$ and conclude

$$\int_{-\infty}^{+\infty} e^{-ax^2}\mathrm{d}x = \frac{1}{e^{i\delta/2}}\sqrt{\frac{\pi}{\rho}} = \sqrt{\frac{\pi}{\rho e^{i\delta}}} = \sqrt{\frac{\pi}{a}}, \qquad -\frac{\pi}{2} < (\delta \equiv \arg[a]) < \frac{\pi}{2}. \tag{7.1.7}$$

Problem 7.1. Compute, for $\mathrm{Re}(a) > 0$,

$$\int_{0}^{+\infty} e^{-ax^2}\mathrm{d}x, \qquad \text{for } \mathrm{Re}(a) > 0 \tag{7.1.8}$$

$$\int_{-\infty}^{+\infty} e^{-ax^2}x^n\mathrm{d}x, \qquad \text{for } n \text{ odd} \tag{7.1.9}$$

$$\int_{-\infty}^{+\infty} e^{-ax^2}x^n\mathrm{d}x, \qquad \text{for } n \text{ even} \tag{7.1.10}$$

$$\int_{0}^{+\infty} e^{-ax^2}x^\beta\mathrm{d}x, \qquad \text{for } \mathrm{Re}(\beta) > -1 \tag{7.1.11}$$

Hint: For the very last integral, consider the change of variables $x' \equiv \sqrt{a}x$, and refer to eq. (6.2.44). □

Problem 7.2. Explain why, for $\mathrm{Im}\ a > 0$ and arbitrary complex b,

$$\int_{\mathbb{R}} \exp\left[iak^2 + ibk\right]\mathrm{d}k = \exp\left[-\frac{ib^2}{4a}\right]\sqrt{\frac{i\pi}{a}}. \tag{7.1.12}$$

Remember the square root is a multi-valued function – explain what it means here. □

Problem 7.3. Gamma function at half-integers From equations (6.2.44) and (6.2.46), explain why for $n = 0, 1, 2, 3, \dots$,

$$\Gamma\left(n+\frac{1}{2}\right) = \sqrt{\pi}\frac{(2n-1)!!}{2^n}; \tag{7.1.13}$$

where the double factorial is $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)$ with $(-1)!! \equiv 1$. Hint: First work out $\Gamma(1/2)$. □

Problem 7.4. Solid Angle in $D \geq 2$ space dimensions There are many applications of the Gaussian integral in physics. Here, we give an application in geometry, and calculate the solid angle in D spatial dimensions. In D-space, the solid angle Ω_{D-1} subtended by a sphere of radius r is defined through the relation

$$\text{Surface area of sphere} \equiv \Omega_{D-1} \cdot r^{D-1}. \tag{7.1.14}$$

Since r is the only length scale in the problem, and since area in D-space has to scale as [Length^{D-1}], we see that Ω_{D-1} is independent of the radius r. Moreover, the volume of a spherical shell of radius r and thickness $\mathrm{d}r$ must be the area of the sphere times $\mathrm{d}r$. Now, argue that the D dimensional integral in spherical coordinates becomes

$$(I_G(a=1))^D = \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} e^{-\vec{x}^2} = \Omega_{D-1}\int_0^\infty \mathrm{d}r \cdot r^{D-1}e^{-r^2}. \tag{7.1.15}$$

Next, evaluate $(I_G(a=1))^D$ directly. Then use the results of the previous problem to compute the last equality of eq. (7.1.15). At this point you should arrive at

$$\Omega_{D-1} = \frac{2\pi^{D/2}}{\Gamma(D/2)}, \tag{7.1.16}$$

where Γ is the Gamma function. Now, for integer $n \geq 0$, $\Gamma(n+1) = n!$. Therefore, the solid angle of a sphere situated in even dimensions is

$$\Omega_{\text{odd }(D-1)} = \frac{2\pi^{D/2}}{((D/2)-1)!}. \tag{7.1.17}$$

Prove that, in odd dimensions,

$$\Omega_{\text{even }D-1=2k} = \frac{2^{k+1}\pi^k}{(2k-1)!!} = \frac{2^{\frac{D+1}{2}}\pi^{\frac{D-1}{2}}}{(D-2)!!}, \tag{7.1.18}$$

$$k = 1, 2, 3, \dots; \tag{7.1.19}$$

where $(2k-1)!! = 1 \cdot 3 \cdot 5 \dots (2k-3) \cdot (2k-1)$. Hint: Use the Gamma function identity in eq. (7.1.13). □

Problem 7.5. Doubling-the-integrals The technique of 'doubling-the-integrals' we exploited in eq. (7.1.2) to evaluate the Gaussian integral $I_G(a)$ in eq. (7.1.1) may be employed to other integrals. In this problem, we will see how the Beta function $B[\mu,\nu] = \Gamma(\mu)\Gamma(\nu)/\Gamma(\mu+\nu)$ in terms of the Gamma function $\Gamma(z)$ may be evaluated in such a manner through its integral representation in eq. (6.2.74).

- First show that

$$\int_0^1 \mathrm{d}z(1-z)^{\mu-1}z^{\nu-1} = \int_0^\infty \mathrm{d}x\frac{x^{\mu-1}}{(1+x)^{\mu+\nu}}. \tag{7.1.20}$$

  Hint: Try putting $z \equiv 1/(1+x)$.

- Second, starting with the product of two Γ-functions using their integral representations in eq. (6.2.44), show that

$$\begin{aligned}\Gamma(\mu)\Gamma(\nu) &= 2\int_0^{\frac{\pi}{2}} \mathrm{d}\theta \sin^{2\mu-1}[\theta]\cos^{2\nu-1}[\theta] \int_0^\infty \mathrm{d}r\ r^{\mu+\nu-1}e^{-r} \\ &= \Gamma[\mu+\nu]\int_0^\infty \mathrm{d}x\frac{x^{\mu-1}}{(1+x)^{\mu+\nu}}. \end{aligned}\tag{7.1.21}$$

  Hints: To obtain the first equality, convert the e^{-t} occurring in eq. (6.2.44) into a Gaussian; i.e., $t \equiv t'^2$. Then use polar coordinates to integrate over the first quadrant. To obtain the second equality, convert to sine and cosine into tangent $\tau \equiv \tan(\theta)$ via $\sin\theta = \tau/\sqrt{1+\tau^2}$ and $\cos\theta = 1/\sqrt{1+\tau^2}$.

  We may summarize our findings as follows:

$$\begin{aligned}2\int_0^{\frac{\pi}{2}} \mathrm{d}\theta \sin^{2\mu-1}[\theta]\cos^{2\nu-1}[\theta] = \int_0^\infty \mathrm{d}x\frac{x^{\mu-1}}{(1+x)^{\mu+\nu}} &= \int_0^1 \mathrm{d}z(1-z)^{\mu-1}z^{\nu-1} \\ &= B[\mu,\nu] = B[\nu,\mu]. \end{aligned}\tag{7.1.22}$$

- Finally, show that

$$\int_0^\infty \frac{\mathrm{d}\lambda\ \lambda^{\mu-1}}{(a+2b\lambda)^{\nu+\mu}} = \frac{B[\mu,\nu]}{2^\nu a^\nu b^\mu}. \tag{7.1.23}$$

□

7.2 Integral Representation of the Gamma function

The integral representation of the Gamma function in eq. (6.2.44) can used to convert a inverse power law into an exponential, via

$$\frac{1}{b^z} = \frac{1}{\Gamma(z)}\int_0^{+\infty} \mathrm{d}t t^{z-1}e^{-bt}, \tag{7.2.1}$$

for Re $b > 0$ and Re $z > -1$. (If Re $b < 0$ the exponential will blow up as $t \to \infty$.) This formula finds applications, for instance, in Feynman diagram calculations.

To show this we simply assume that $b > 0$ first. Then it is just a matter of re-scaling $t' \equiv bt$.

$$\int_0^{+\infty} \mathrm{d}t t^{z-1}e^{-bt} = \frac{1}{b^z}\int_0^{+\infty} \mathrm{d}(bt)(bt)^{z-1}e^{-(bt)} \tag{7.2.2}$$

$$= \frac{1}{b^z} \int_0^{+\infty} \mathrm{d}t' t'^{z-1} e^{-t'}. \tag{7.2.3}$$

The case for Re $b > 0$ can simply be obtained by analytic continuation, since both left and right hand sides exist and are analytic.

Example Let us consider the following $D-$dimensional integral, where we shall assume Re $z > -1$,

$$I(\vec{x}) \equiv \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{e^{i\vec{k}\cdot\vec{x}}}{(\vec{k}^2)^z}. \tag{7.2.4}$$

First we re-scale $\vec{q} \equiv |\vec{x}|\vec{k}$, denote $\hat{n} \equiv \vec{x}/|\vec{x}|$, and apply eq. (7.2.1).

$$\begin{aligned} I(\vec{x}) &= \frac{1}{|\vec{x}|^{D-2z}} \int_{\mathbb{R}^D} \frac{\mathrm{d}^D q}{(2\pi)^D} \frac{e^{i\vec{q}\cdot\hat{\hat{n}}}}{(\vec{q}^2)^z} \\ &= \frac{1}{\Gamma(z)|\vec{x}|^{D-2z}} \int_0^\infty \mathrm{d}t\ t^{z-1} \int_{\mathbb{R}^D} \frac{\mathrm{d}^D q}{(2\pi)^D} e^{i\vec{q}\cdot\hat{n} - t\vec{q}^2} \end{aligned} \tag{7.2.5}$$

Next, we recognize that the $\vec{q}-$integral may be factored into D one dimensional integrals, where each and every may be tackled using eq. (7.1.12). That yields:

$$\begin{aligned} I(\vec{x}) &= \frac{1}{\Gamma(z)|\vec{x}|^{D-2z}} \int_0^\infty \mathrm{d}t\ t^{z-1} \prod_{j=1}^{D} \left(\int_{\mathbb{R}} \frac{\mathrm{d}q^j}{2\pi} e^{iq^j \hat{n}^j - t(q^j)^2} \right) \\ &= \frac{1}{\Gamma(z)|\vec{x}|^{D-2z}} \int_0^\infty \mathrm{d}t\ t^{z-1} \prod_{j=1}^{D} \left(\frac{\sqrt{\pi}}{2\pi t^{1/2}} \exp\left[-\frac{(\hat{n}^j)^2}{4t} \right] \right) \\ &= \frac{1}{\Gamma(z)|\vec{x}|^{D-2z}} \int_0^\infty \frac{\mathrm{d}t}{4^{z-1}} \frac{(4t)^{z-1}}{(4\pi t)^{D/2}} \exp\left[-\frac{1}{4t} \right]. \end{aligned} \tag{7.2.6}$$

Finally, putting $1/4t \equiv \eta$ hands us the integral representation of the Gamma function in eq. (6.2.44).

$$I(\vec{x}) = \frac{1}{\Gamma(z)|\vec{x}|^{D-2z}} \int_0^\infty \frac{\mathrm{d}\eta}{4^z \pi^{D/2}} \eta^{\frac{D}{2}-z-1} e^{-\eta}. \tag{7.2.7}$$

We have obtained the result

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{e^{i\vec{k}\cdot\vec{x}}}{(\vec{k}^2)^z} = \frac{\Gamma[\frac{D}{2} - z]}{4^z \pi^{D/2} \Gamma[z] |\vec{x}|^{D-2z}}, \tag{7.2.8}$$

which allows us to analytically continue z for all complex z. $\square$

Problem 7.6. Massive case As a generalization, show that

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{e^{i\vec{k}\cdot\vec{r}}}{(\vec{k}^2 + m^2)^a} = \frac{2m^{D-2a}}{(2\sqrt{\pi})^D \Gamma(a)} \left(\frac{2}{mr} \right)^{\frac{D}{2}-a} K_{\frac{D}{2}-a}(mr), \tag{7.2.9}$$

where $r \equiv |\vec{x}|$ and $K_\nu(z)$ is the modified Bessel function. For $a = 1$ and $D = 3$, you should also find that

$$\int_{\mathbb{R}^3} \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot\vec{r}}}{\vec{k}^2 + m^2} = \frac{\exp(-m \cdot r)}{4\pi r}. \tag{7.2.10}$$

Hint: You may wish to utilize the following integral representation of the modified Bessel function $K_\nu(z)$:

$$K_\nu[z] = \frac{1}{2}\left(\frac{z}{2}\right)^\nu \int_0^\infty \frac{\mathrm{d}t}{t^{\nu+1}} \exp\left[-t - \frac{z^2}{4t}\right]. \tag{7.2.11}$$

Remark For large argument $|z| \gg 1$,

$$K_\nu(z) \sim \sqrt{\frac{\pi}{2z}} \exp(-z).$$

Therefore the above integral carries the behavior for $mr \gg 1$:

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{e^{i\vec{k}\cdot\mathbf{r}}}{(\vec{k}^2 + m^2)^a} \sim \frac{1}{2^{D-\frac{1}{2}} \pi^{\frac{D-1}{2}} \Gamma(a)} \left(\frac{2m}{r}\right)^{\frac{D}{2}-a} \frac{\exp(-mr)}{\sqrt{mr}}. \tag{7.2.12}$$

In fundamental physics, the Coulomb potential ($\propto 1/r^{D-2}$) corresponds to a force mediated by a massless $m = 0$ particle such as the photon. If the force carrier is massive $m > 0$, on the other hand, this result yields the corresponding 'Yukawa potential', whose range is significantly shorter due to the exponential suppression $\exp(-mr)$. □

Problem 7.7. Show that

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D z}{(\vec{z}^2 + \Delta)^\sigma} = \frac{\pi^{D/2} \Gamma\left(\sigma - \frac{D}{2}\right)}{\Gamma(\sigma) \Delta^{\sigma - \frac{D}{2}}}. \tag{7.2.13}$$

What restrictions on Δ and σ need to be imposed? □

Problem 7.8. By considering the $z \to 0$ limit of eq. (7.2.8), justify the following result.

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} e^{i\vec{k}\cdot\vec{x}} \ln \frac{\vec{k}^2}{m^2} = -\frac{\Gamma[\frac{D}{2}]}{(\sqrt{\pi}|\vec{x}|)^D}; \tag{7.2.14}$$

for some arbitrary constant $m^2 > 0$, where $[m^2] = [\vec{k}^2] = 1/[\vec{x}^2]$. Hint: $z^{-\epsilon} = 1 - \epsilon \ln z + \mathcal{O}(\epsilon^2)$. □

Feynman-Schwinger Parameters There is

7.3 Complexification

Sometimes complexifying the integral makes it easier. Here's a simple example from Matthews and Walker [15].

$$I = \int_0^\infty \mathrm{d}x e^{-ax} \cos(\lambda x), \qquad a > 0,\ \lambda \in \mathbb{R}. \tag{7.3.1}$$

If we regard $\cos(\lambda x)$ as the real part of $e^{i\lambda x}$,

$$\begin{aligned} I &= \mathrm{Re} \int_0^\infty \mathrm{d}x e^{-(a-i\lambda)x} \\ &= \mathrm{Re} \left[\frac{e^{-(a-i\lambda)x}}{-(a-i\lambda)} \right]_{x=0}^{x=\infty} \\ &= \mathrm{Re} \frac{1}{a-i\lambda} = \mathrm{Re} \frac{a+i\lambda}{a^2+\lambda^2} = \frac{a}{a^2+\lambda^2} \end{aligned} \tag{7.3.2}$$

Problem 7.9. What is

$$\int_0^\infty \mathrm{d}x e^{-ax} \sin(\lambda x), \qquad a > 0, \; \lambda \in \mathbb{R}? \tag{7.3.3}$$

7.4 Differentiation under the integral sign (Leibniz's theorem)

Differentiation under the integral sign, or Leibniz's theorem, is the result

$$\frac{\mathrm{d}}{\mathrm{d}z} \int_{a(z)}^{b(z)} \mathrm{d}s F(z,s) = b'(z) F(z, b(z)) - a'(z) F(z, a(z)) + \int_{a(z)}^{b(z)} \mathrm{d}s \frac{\partial F(z,s)}{\partial z}. \tag{7.4.1}$$

Problem 7.10. By using the limit definition of the derivative, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}z} H(z) = \lim_{\delta \to 0} \frac{H(z+\delta) - H(z)}{\delta}, \tag{7.4.2}$$

argue the validity of eq. (7.4.1). □

Why this result is useful for integration can be illustrated by some examples. The art involves creative insertion of some auxiliary parameter α in the integrand. Let's start with

$$\Gamma(n+1) = \int_0^\infty \mathrm{d}t t^n e^{-t}, \qquad n \text{ a positive integer.} \tag{7.4.3}$$

For $\mathrm{Re}(n) > -1$ this is in fact the definition of the Gamma function. We introduce the parameter as follows

$$I_n(\alpha) = \int_0^\infty \mathrm{d}t t^n e^{-\alpha t}, \qquad \alpha > 0, \tag{7.4.4}$$

and notice

$$\begin{aligned} I_n(\alpha) &= (-\partial_\alpha)^n \int_0^\infty \mathrm{d}t e^{-\alpha t} = (-\partial_\alpha)^n \frac{1}{\alpha} \\ &= (-)^n(-1)(-2)\dots(-n)\alpha^{-1-n} = n!\alpha^{-1-n} \end{aligned} \tag{7.4.5}$$

By setting $\alpha = 1$, we see that the Gamma function $\Gamma(z)$ evaluated at integer values of z returns the factorial.

$$\Gamma(n+1) = I_n(\alpha = 1) = n!. \tag{7.4.6}$$

Next, we consider a trickier example:

$$\int_{-\infty}^{\infty} \frac{\sin(x)}{x} \mathrm{d}x. \tag{7.4.7}$$

This can be evaluated via a contour integral. But here we do so by introducing a $\alpha \in \mathbb{R}$,

$$I(\alpha) \equiv \int_{-\infty}^{\infty} \frac{\sin(\alpha x)}{x} \mathrm{d}x. \tag{7.4.8}$$

Observe that the integral is odd with respect to α, $I(-\alpha) = -I(\alpha)$. Differentiating once,

$$I'(\alpha) = \int_{-\infty}^{\infty} \cos(\alpha x) \mathrm{d}x = \int_{-\infty}^{\infty} e^{i\alpha x} \mathrm{d}x = 2\pi\delta(\alpha). \tag{7.4.9}$$

($\cos(\alpha x)$ can be replaced with $e^{i\alpha x}$ because the $i\sin(\alpha x)$ portion integrates to zero.) Remember the derivative of the step function $\Theta(\alpha)$ is the Dirac δ-function $\delta(\alpha)$: $\Theta'(z) = \Theta'(-z) = \delta(z)$. Taking into account $I(-\alpha) = -I(\alpha)$, we can now deduce the answer to take the form

$$I(\alpha) = \pi\left(\Theta(\alpha) - \Theta(-\alpha)\right) = \pi \mathrm{sgn}(\alpha), \tag{7.4.10}$$

There is no integration constant here because it will spoil the property $I(-\alpha) = -I(\alpha)$. What remains is to choose $\alpha = 1$,

$$I(1) = \int_{-\infty}^{\infty} \frac{\sin(x)}{x} \mathrm{d}x = \pi. \tag{7.4.11}$$

Problem 7.11. An alternate means of obtaining this result is to first consider

$$I(\alpha > 0) \equiv \int_0^{\infty} e^{-\alpha \cdot x} \frac{\sin(x)}{x}. \tag{7.4.12}$$

Differentiate $I(\alpha)$ with respect to α and recall the result $\partial_\alpha \arctan(\alpha) = 1/(1+\alpha^2)$ to arrive at eq. (7.4.10). ☐

Problem 7.12. Show that

$$I(\alpha) = \int_0^{\pi} \ln\left[1 - 2\alpha\cos(x) + \alpha^2\right] \mathrm{d}x = 2\pi\Theta(|\alpha| - 1) \cdot \ln\alpha, \qquad |\alpha| \neq 1, \tag{7.4.13}$$

by differentiating once with respect to α, changing variables to $t \equiv \tan(x/2)$, and then using complex analysis. (*Do not* copy the solution from Wikipedia!) You may need to consider the cases $|\alpha| > 1$ and $|\alpha| < 1$ separately. Here, $\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$. ☐

Problem 7.13. Natural Log: Integral Representation Show that, for $x > 0$, the natural logarithm admits the integral representation

$$\ln x = \int_0^{\infty} \frac{\mathrm{d}s}{s}\left(e^{is} - e^{ixs}\right). \tag{7.4.14}$$

This formula finds application even in quantum field theory. Hint: Rotate the s contour on the complex plane to convert the exponents into purely real ones. Then treat the resulting integral as a function of x: Can you derive a first order differential equation it satisfies? What is the appropriate boundary condition? ☐

Problem 7.14. Cauchy-Frullani Integral For $\alpha > 0$, show that

$$\int_0^\infty \frac{f(\alpha x) - f(x)}{x} \mathrm{d}x = (f(\infty) - f(0)) \ln \alpha. \tag{7.4.15}$$

Then deduce that

$$\int_0^\infty \frac{f(ax) - f(bx)}{x} \mathrm{d}x = (f(\infty) - f(0)) \ln(a/b) \tag{7.4.16}$$

for $a, b > 0$. By introducing a suitable damping factor, can you recover eq. (7.4.14) from eq. (7.4.16)? □

7.5 Symmetry

You may sometimes need to do integrals in higher than one dimension. If it arises from a physical problem, it may exhibit symmetry properties you should definitely exploit. The case of rotational symmetry is a common and important one, and we shall focus on it here. A simple example is as follows. In 3-dimensional (3D) space, we define

$$I(\vec{k}) \equiv \int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega_{\widehat{n}}}{4\pi} e^{i\vec{k}\cdot\widehat{n}}. \tag{7.5.1}$$

The $\int_{\mathbb{S}^2} \mathrm{d}\Omega$ means we are integrating the unit radial vector $\widehat{n}$ with respect to the solid angles on the sphere; $\vec{k} \cdot \vec{x}$ is just the Euclidean dot product. For example, if we use spherical coordinates, the Cartesian components of the unit vector would be

$$\widehat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{7.5.2}$$

and $\mathrm{d}\Omega = \mathrm{d}(\cos\theta)\mathrm{d}\phi$. The key point here is that we have a rotationally invariant integral. In particular, the (θ, ϕ) here are measured with respect to some (x^1, x^2, x^3)-axes. If we rotated them to some other (orthonormal) (x'^1, x'^2, x'^3)-axes related via some rotation matrix $R^i{}_j$,

$$\widehat{n}^i(\theta, \phi) = R^i{}_j \widehat{n}'^j(\theta', \phi'), \tag{7.5.3}$$

where $\det R^i{}_j = 1$; in matrix notation $\widehat{n} = R\widehat{n}'$ and $R^{\mathrm{T}} R = \mathbb{I}$. Then $\mathrm{d}(\cos\theta)\mathrm{d}\phi = \mathrm{d}\Omega = \mathrm{d}\Omega' \det R^i{}_j = \mathrm{d}\Omega' = \mathrm{d}(\cos\theta')\mathrm{d}\phi'$, and

$$I(R\vec{k}) = \int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega_{\widehat{n}}}{4\pi} e^{i\vec{k}\cdot(R^{\mathrm{T}}\widehat{n})} = \int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega'_{\widehat{n}'}}{4\pi} e^{i\vec{k}\cdot\widehat{n}'} = I(\vec{k}). \tag{7.5.4}$$

In other words, because R was an arbitrary rotation matrix, $I(\vec{k}) = I(|\vec{k}|)$; the integral cannot possibly depend on the direction of $\vec{k}$, but only on the magnitude $|\vec{k}|$. That in turn means we may as well pretend $\vec{k}$ points along the x^3-axis, so that the dot product $\vec{k} \cdot \widehat{n}'$ only involved the $\cos\theta \equiv \widehat{n}' \cdot \widehat{e}_3$.

$$I(|\vec{k}|) = \int_0^{2\pi} \mathrm{d}\phi \int_{-1}^{+1} \frac{\mathrm{d}(\cos\theta)}{4\pi} e^{i|\vec{k}|\cos\theta} = \frac{e^{i|\vec{k}|} - e^{-i|\vec{k}|}}{2i|\vec{k}|}. \tag{7.5.5}$$

We arrive at

$$\int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega_{\widehat{n}}}{4\pi} e^{i\vec{k}\cdot\widehat{n}} = \frac{\sin|\vec{k}|}{|\vec{k}|}. \tag{7.5.6}$$

Problem 7.15. With $\widehat{n}$ denoting the unit radial vector in 3−space, evaluate

$$I(\vec{x}) = \int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega_{\widehat{n}}}{|\vec{x}-\vec{r}|}, \qquad \vec{r} \equiv r\widehat{n}. \tag{7.5.7}$$

Note that the answer for $|\vec{x}| > |\vec{r}| = r$ differs from that when $|\vec{x}| < |\vec{r}| = r$. Can you explain the physical significance? Hint: This can be viewed as an electrostatics problem. □

Problem 7.16. A problem that combines both rotational symmetry and the higher dimensional version of "differentiation under the integral sign" is the (tensorial) integral

$$\int_{\mathbb{S}^2} \frac{\mathrm{d}\Omega}{4\pi} \widehat{n}^{i_1}\widehat{n}^{i_2}\dots\widehat{n}^{i_N}, \tag{7.5.8}$$

where $\widehat{n}$ is the unit radial vector in 3−dimensional flat space; N is an integer greater than or equal to 1. The answer for odd N can be understood by asking, how does the integrand and the measure $\mathrm{d}\Omega_{\widehat{n}}$ transform under a parity flip of the coordinate system, namely under $\widehat{n} \to -\widehat{n}$? What's the answer for even N? Hint: consider differentiating eq. (7.5.6) with respect to $k^{i_1}, \dots, k^{i_N}$; how is that related to the Taylor expansion of $\sin(|\vec{k}|)/|\vec{k}|$? (There is some combinatorics to consider here.) Also consider carrying out the calculation explicitly for the first few cases; e.g.. for $N = 1, 2, 3, 4$. □

Problem 7.17. Can you generalize eq. (7.5.6) to D spatial dimensions? Namely, evaluate

$$\int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\Omega_{\widehat{n}} e^{i\vec{k}\cdot\hat{n}}. \tag{7.5.9}$$

The $\vec{k}$ is an arbitrary vector in D-space and $\widehat{n}$ is the unit radial vector in the same. Hint: You should find

$$\int_{\mathbb{S}^{D-1}} \mathrm{d}\Omega_{\widehat{n}} e^{i\vec{k}\cdot\hat{n}} = \left(\int_{\mathbb{S}^{D-2}} \mathrm{d}\Omega_{\widehat{n}}\right) \int_0^\pi (\sin\theta)^{D-2} e^{i|\vec{k}|\cos\theta} \mathrm{d}\theta. \tag{7.5.10}$$

Then refer to eq. 10.9.4 of the NIST page here. Using the results of eq. (7.5.9) or otherwise, explain why the tensor integral involving odd powers of the radial vector is zero.

$$\int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\Omega \widehat{n}^{i_1}\dots\widehat{n}^{i_{2\ell}}\widehat{n}^{i_{2\ell+1}} = 0, \qquad \ell = 0, 1, 2, \dots \tag{7.5.11}$$

Then verify that the integral over even powers of $\widehat{n}$ delivers the following result:

$$\int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\Omega \widehat{n}^{i_1}\dots\widehat{n}^{i_{2\ell}} = \frac{\pi^{\frac{D}{2}}}{2^{\ell-1}\Gamma[\frac{D-2}{2}+\ell+1]} \sum \left(\text{Full contractions of } k^{i_1}\dots k^{i_{2\ell}}\right). \tag{7.5.12}$$

Here, I have defined a contraction between a pair of k's by replacing them with the corresponding Kronecker delta. For e.g., contraction of k^ik^j yields δ^{ij}; full contraction of $k^{i_1}k^{i_2}k^{i_3}k^{i_4}$ would yield

$$\delta^{i_1i_2}\delta^{i_3i_4} + \delta^{i_1i_3}\delta^{i_2i_4} + \delta^{i_1i_4}\delta^{i_2i_3}. \tag{7.5.13}$$

Hints: You may first explain why

$$\frac{\partial^{2\ell}}{\partial k^{i_1}\dots\partial k^{i_{2\ell}}}k^{2\ell} = 2^\ell\cdot\ell!\sum\left(\text{Full contractions of } k^{i_1}\dots k^{i_{2\ell}}\right). \tag{7.5.14}$$

If you work out the first few cases, you should find:

$$\frac{\partial^2}{\partial k^{i_1}\partial k^{i_2}}k^2 = 2\delta^{i_1 i_2}, \tag{7.5.15}$$

$$\frac{\partial^4}{\partial k^{i_1}\dots\partial k^{i_4}}k^4 = 8\left(\delta^{i_1 i_2}\delta^{i_3 i_4}+\delta^{i_1 i_3}\delta^{i_2 i_4}+\delta^{i_1 i_4}\delta^{i_2 i_3}\right), \tag{7.5.16}$$

$$\begin{aligned}\frac{\partial^6}{\partial k^{i_1}\dots\partial k^{i_6}}k^6 = 48(&\delta^{i_1 i_2}\delta^{i_3 i_4}\delta^{i_5 i_6}+\delta^{i_1 i_3}\delta^{i_2 i_4}\delta^{i_5 i_6}+\delta^{i_1 i_4}\delta^{i_2 i_3}\delta^{i_5 i_6}\\
&\delta^{i_1 i_2}\delta^{i_3 i_5}\delta^{i_4 i_6}+\delta^{i_1 i_3}\delta^{i_2 i_5}\delta^{i_4 i_6}+\delta^{i_1 i_5}\delta^{i_2 i_3}\delta^{i_4 i_6}\\
&\delta^{i_1 i_2}\delta^{i_5 i_4}\delta^{i_3 i_6}+\delta^{i_1 i_5}\delta^{i_2 i_4}\delta^{i_3 i_6}+\delta^{i_1 i_4}\delta^{i_2 i_5}\delta^{i_3 i_6}\\
&\delta^{i_1 i_5}\delta^{i_3 i_4}\delta^{i_2 i_6}+\delta^{i_1 i_3}\delta^{i_5 i_4}\delta^{i_2 i_6}+\delta^{i_1 i_4}\delta^{i_5 i_3}\delta^{i_2 i_6}\\
&\delta^{i_5 i_2}\delta^{i_3 i_4}\delta^{i_1 i_6}+\delta^{i_5 i_3}\delta^{i_2 i_4}\delta^{i_1 i_6}+\delta^{i_5 i_4}\delta^{i_2 i_3}\delta^{i_1 i_6}).\end{aligned} \tag{7.5.17}$$

These results could then be used to extract eq. (7.5.12) from eq. (7.5.9). ☐

Tensor integrals Next, we consider the following integral involving two arbitrary vectors $\vec{a}$ and $\vec{k}$ in 3D space.[56]

$$I\left(\vec{a},\vec{k}\right) = \int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{n}}\frac{\vec{a}\cdot\widehat{n}}{1+\vec{k}\cdot\widehat{n}} \tag{7.5.18}$$

First, we write it as $\vec{a}$ dotted into a vector integral $\vec{J}$, namely

$$I\left(\vec{a},\vec{k}\right) = \vec{a}\cdot\vec{J}, \qquad \vec{J}\left(\vec{k}\right) \equiv \int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{n}}\frac{\widehat{n}}{1+\vec{k}\cdot\widehat{n}}. \tag{7.5.19}$$

Let us now consider replacing $\vec{k}$ with a rotated version of $\vec{k}$. This amounts to replacing $\vec{k}\to R\vec{k}$, where R is an orthogonal 3×3 matrix of unit determinant, with $R^{\mathrm{T}}R = RR^{\mathrm{T}} = \mathbb{I}$. We shall see that $\vec{J}$ transforms as a vector $\vec{J}\to R\vec{J}$ under this same rotation. This is because $\int \mathrm{d}\Omega_{\widehat{n}}\to\int\mathrm{d}\Omega_{\widehat{n}'}$, for $\widehat{n}'\equiv R^{\mathrm{T}}\widehat{n}$, and

$$\begin{aligned}\vec{J}\left(R\vec{k}\right) &= \int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{n}}\frac{R(R^{\mathrm{T}}\widehat{n})}{1+\vec{k}\cdot(R^{\mathrm{T}}\widehat{n})}\\
&= R\int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{n}'}\frac{\widehat{n}'}{1+\vec{k}\cdot\widehat{n}'} = R\vec{J}(\vec{k}).\end{aligned} \tag{7.5.20}$$

But the only vector that $\vec{J}$ depends on is $\vec{k}$. Therefore the result of $\vec{J}$ has to be some scalar function f times $\vec{k}$.

$$\vec{J} = f\cdot\vec{k}, \qquad \Rightarrow \qquad I\left(\vec{a},\vec{k}\right) = f\vec{a}\cdot\vec{k}. \tag{7.5.21}$$

[56]This example was taken from Matthews and Walker [15].

To calculate f we now dot both sides with $\vec{k}$.

$$f = \frac{\vec{J} \cdot \vec{k}}{\vec{k}^2} = \frac{1}{\vec{k}^2} \int_{\mathbb{S}^2} \mathrm{d}\Omega_{\widehat{n}} \frac{\vec{k} \cdot \widehat{n}}{1 + \vec{k} \cdot \widehat{n}} \tag{7.5.22}$$

At this point, the nature of the remaining scalar integral is very similar to the one we've encountered previously. Choosing $\vec{k}$ to point along the $\widehat{e}_3$ axis,

$$\begin{aligned} f &= \frac{2\pi}{\vec{k}^2} \int_{-1}^{+1} \mathrm{d}(\cos\theta) \frac{|\vec{k}| \cos\theta}{1 + |\vec{k}| \cos\theta} \\ &= \frac{2\pi}{\vec{k}^2} \int_{-1}^{+1} \mathrm{d}c \left(1 - \frac{1}{1 + |\vec{k}| c}\right) = \frac{4\pi}{\vec{k}^2} \left(1 - \frac{1}{2|\vec{k}|} \ln\left(\frac{1 + |\vec{k}|}{1 - |\vec{k}|}\right)\right). \end{aligned} \tag{7.5.23}$$

Therefore,

$$\int_{\mathbb{S}^2} \mathrm{d}\Omega_{\widehat{n}} \frac{\vec{a} \cdot \widehat{n}}{1 + \vec{k} \cdot \widehat{n}} = \frac{4\pi\left(\vec{k} \cdot \vec{a}\right)}{\vec{k}^2} \left(1 - \frac{1}{2|\vec{k}|} \ln\left(\frac{1 + |\vec{k}|}{1 - |\vec{k}|}\right)\right). \tag{7.5.24}$$

This technique of reducing tensor integrals into scalar ones find applications even in quantum field theory calculations.

Problem 7.18. Calculate

$$A^{ij}(\vec{a}) \equiv \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{k^i k^j}{\vec{k}^2 + (\vec{k} \cdot \vec{a})^4}, \tag{7.5.25}$$

where $\vec{a}$ is some (dimensionless) vector in 3D Euclidean space. Do so by first arguing that this integral transforms as a tensor in D-space under rotations. In other words, if ${R^i}_j$ is a rotation matrix, under the rotation

$$a^i \to {R^i}_j a^j, \tag{7.5.26}$$

we have

$$A^{ij}({R^k}_l a^l) = {R^i}_l {R^j}_k A^{kl}(\vec{a}). \tag{7.5.27}$$

Hint: The only rank-2 tensors available here are δ^{ij} and $a^i a^j$, so we must have

$$A^{ij} = f_1 \delta^{ij} + f_2 a^i a^j. \tag{7.5.28}$$

To find $\psi_{1,2}$ take the trace and also consider $A^{ij} a_i a_j$. $\square$

7.6 Integrals involving Bessel functions

7.7 Asymptotic expansion of integrals

[57] Many solutions to physical problems, say arising from some differential equations, can be expressed as integrals. Moreover the "special functions" of mathematical physics, whose properties

[57] The material in this section is partly based on Chapter 3 of Matthews and Walker's "*Mathematical Methods of Physics*" [15]; and the latter portions are heavily based on Chapter 6 of Bender and Orszag's "*Advanced mathematical methods for scientists and engineers*" [16].

are well studied – Bessel, Legendre, hypergeometric, etc. – all have integral representations. Often we wish to study these functions when their arguments are very large, and it is then useful to have techniques to extract an answer from these integrals in such a limit. This topic is known as the "asymptotic expansion of integrals".

7.7.1 Integration-by-parts (IBP)

In this section we will discuss how to use integration-by-parts (IBP) to approximate integrals.

Previously we evaluated

$$\frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-t^2} \mathrm{d}t = 1. \tag{7.7.1}$$

The erf function is defined as

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}t e^{-t^2}. \tag{7.7.2}$$

Its small argument limit can be obtained by Taylor expansion,

$$\begin{aligned} \mathrm{erf}(x \ll 1) &= \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}t \left(1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \dots\right) \\ &= \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \dots\right). \end{aligned} \tag{7.7.3}$$

But what about its large argument limit $\mathrm{erf}(x \gg 1)$? We may write

$$\begin{aligned} \mathrm{erf}(x) &= \frac{2}{\sqrt{\pi}} \left(\int_0^\infty \mathrm{d}t - \int_x^\infty \mathrm{d}t\right) e^{-t^2} \\ &= 1 - \frac{2}{\sqrt{\pi}} I(x), \qquad\qquad I(x) \equiv \int_x^\infty \mathrm{d}t e^{-t^2}. \end{aligned} \tag{7.7.4}$$

Integration-by-parts may be employed as follows.

$$\begin{aligned} I(x) &= \int_x^\infty \mathrm{d}t \frac{1}{-2t} \partial_t e^{-t^2} = \left[\frac{e^{-t^2}}{-2t}\right]_{t=x}^{t=\infty} - \int_x^\infty \mathrm{d}t \partial_t \left(\frac{1}{-2t}\right) e^{-t^2} \\ &= \frac{e^{-x^2}}{2x} - \int_x^\infty \mathrm{d}t \frac{e^{-t^2}}{2t^2} = \frac{e^{-x^2}}{2x} - \int_x^\infty \mathrm{d}t \frac{1}{2t^2(-2t)} \partial_t e^{-t^2} \\ &= \frac{e^{-x^2}}{2x} - \frac{e^{-x^2}}{4x^3} + \int_x^\infty \mathrm{d}t \frac{3}{4t^4} e^{-t^2} \end{aligned} \tag{7.7.5}$$

Problem 7.19. After n integration by parts,

$$\int_x^\infty \mathrm{d}t e^{-t^2} = e^{-x^2} \sum_{\ell=1}^{n} (-)^{\ell-1} \frac{1 \cdot 3 \cdot 5 \dots (2\ell - 3)}{2^\ell x^{2\ell-1}} - (-)^n \frac{1 \cdot 3 \cdot 5 \dots (2n-1)}{2^n} \int_x^\infty \mathrm{d}t \frac{e^{-t^2}}{t^{2n}}. \tag{7.7.6}$$

This result can be found in Matthew and Walker, but can you prove it more systematically by mathematical induction? For a fixed x, find the n such that the next term generated by integration-by-parts is larger than the previous term. This series does not converge – why? □

If we drop the remainder integral in eq. (7.7.6), the resulting series does not converge as $n \to \infty$. However, for large $x \gg 1$, it is not difficult to argue that the first few terms do offer an excellent approximation, since each subsequent term is suppressed relative to the previous by a $1/x$ factor.[58]

Problem 7.20. Using integration-by-parts, develop a large $x \gg 1$ expansion for

$$I(x) \equiv \int_x^\infty \mathrm{d}t \frac{\sin(t)}{t}. \tag{7.7.7}$$

Hint: Consider instead $\int_x^\infty \mathrm{d}t \frac{\exp(it)}{t}$. □

What is an asymptotic series? A Taylor expansion of say e^x

$$e^x = 1 + x + \frac{x}{2!} + \frac{x^3}{3!} + \dots \tag{7.7.8}$$

converges for all $|x|$. In fact, for a fixed $|x|$, we know summing up more terms of the series

$$\sum_{\ell=0}^{N} \frac{x^\ell}{\ell!}, \tag{7.7.9}$$

– the larger N we go – the closer to the actual value of e^x we would get.

An asymptotic series of the sort we have encountered above, and will be doing so below, is a series of the sort

$$S_N(x) = A_0 + \frac{A_1}{x} + \frac{A_2}{x^2} + \dots + \frac{A_N}{x^N}. \tag{7.7.10}$$

For a fixed $|x|$ the series oftentimes diverges as we sum up more and more terms ($N \to \infty$). However, for a fixed N, it can usually be argued that as $x \to +\infty$ the $S_N(x)$ becomes an increasingly better approximation to the object we derived it from in the first place.

As Matthews and Walker [15] further explains:

> "*. . . an asymptotic series may be added, multiplied, and integrated to obtain the asymptotic series for the corresponding sum, product and integrals of the corresponding functions. Also, the asymptotic series of a given function is unique, but . . . An asymptotic series does not specify a function uniquely.*"

Problem 7.21. Gaussian Distributions And Their Tails The Gaussian probability distribution (aka, normal distribution or Bell curve) is defined as

$$P(x, \bar{x}; \sigma) \equiv \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}, \qquad x, \bar{x} \in \mathbb{R}, \sigma > 0. \tag{7.7.11}$$

[58] In fact, as observed by Matthews and Walker [15], since this is an oscillating series, the optimal n to truncate the series is the one right before the smallest.

By probability distribution, we mean $P(x,\bar{x};\sigma)\mathrm{d}x$ is the probability of obtaining an outcome to lie between x and $x+\mathrm{d}x$. It must be normalized such that the total probability is unity,

$$\int_{-\infty}^{+\infty} P(x,\bar{x};\sigma)\mathrm{d}x = 1. \tag{7.7.12}$$

Normal distributions show up prominently in statistical analysis. An example of its application is the *Central Limit Theorem*, which tells us: the random errors associated with the measurements of a given observable is expected to be Gaussian distributed, provided these errors are due to a large number of independent factors. In what follows, we shall define

$$\langle f \rangle \equiv \int_{-\infty}^{+\infty} f(x)P(x,\bar{x};\sigma)\mathrm{d}x. \tag{7.7.13}$$

Mean (aka Average) Show that the average value of making a large number of measurements – i.e., the mean – is

$$\langle x \rangle = \bar{x}. \tag{7.7.14}$$

Variance and Standard Deviation Next, show that the variance is

$$\langle (x-\bar{x})^2 \rangle = \langle x^2 \rangle - \bar{x}^2 = \sigma^2. \tag{7.7.15}$$

The standard deviation is defined as the square root of the variance, and is therefore σ.

Higher Point Functions Show that $\langle (x-\bar{x})^n \rangle = 0$, for all odd $n = 1,3,5,\dots$; and for even $n = 2,4,6,\dots$,

$$\langle (x-\bar{x})^n \rangle = (n-1)!!\sigma^n; \tag{7.7.16}$$

where $n!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (n-2) \cdot n$. That is, you will discover – perhaps not surprisingly – that the answer depends only on σ. The generalization of this result to perturbative Quantum Field Theory reads: all even point functions are determined by the two point function via Wick's theorem.

Tails and Asymptotics Even though the tail end of a Gaussian distribution is exponentially suppressed, and hence constitutes a small fraction of the total population, these 'outliers' are often important because of their extreme characteristics. For example, the IQ of humans are typically modeled as normal distributed. If this is accurate, that means there are exponentially few people who are extremely smart; but because they are significantly more intelligent than the rest of us, they will tend to stand out from the crowd – make significantly more contributions – in tasks that are cognitively demanding.

Different Means, Same Variance Suppose group A has an IQ mean of $\bar{x}_<$ and group B has a higher IQ mean of $\bar{x}_>$; and suppose the IQ variances $\{\sigma\}$ of the two groups are the same. If the total population of A is N_A and that of B is N_B, show that for both $(x_\star - \bar{x}_<)/\sigma \gg 1$ and $(x_\star - \bar{x}_>)/\sigma \gg 1$,

$$\frac{\text{Number of people from group } A \text{ with IQ greater or equal to } x_\star}{\text{Number of people from group } B \text{ with IQ greater or equal to } x_\star} \tag{7.7.17}$$

$$\sim \frac{N_A}{N_B}\left(\frac{x_\star - \bar{x}_>}{x_\star - \bar{x}_<}\right)\exp\left(-\frac{(2x_\star - \bar{x}_> - \bar{x}_<)(\bar{x}_> - \bar{x}_<)}{2\sigma^2}\right)\left(1 - \frac{\sigma^2}{(x_\star - \bar{x}_<)^2} + \frac{\sigma^2}{(x_\star - \bar{x}_>)^2} + \dots\right),$$

where $\sim$ means 'asymptotic to'. For a 'cut-off' $x_\star$ much larger than the means $\bar{x}_<$ and $\bar{x}_>$, the exponential factor teaches us:

> If the populations of two groups are comparable ($N_A \sim N_B$) then the more elite the selection – namely, the higher the threshold $x_\star$ – the more the outcome will be exponentially dominated by the group with the higher mean.

Same Mean, Different Variances Next, suppose groups A and B have the same IQ average $\bar{x}$; but suppose the IQ standard deviation $\sigma_<$ of A is smaller than the IQ standard deviation $\sigma_>$ of B. If the total population of A is N_A and that of B is N_B, show that

$$\frac{\text{Number of people from group } A \text{ with IQ greater or equal to } x_\star}{\text{Number of people from group } B \text{ with IQ greater or equal to } x_\star} \tag{7.7.18}$$

$$\sim \frac{N_A}{N_B}\left(\frac{\sigma_<}{\sigma_>}\right)\exp\left(-\frac{(x_\star - \bar{x})^2(\sigma_>^2 - \sigma_<^2)}{2\sigma_>^2\sigma_<^2}\right)\left(1 - \frac{\sigma_<^2}{(x_\star - \bar{x})^2} + \frac{\sigma_>^2}{(x_\star - \bar{x})^2} + \dots\right). \tag{7.7.19}$$

For a 'cut-off' $x_\star$ much larger than the mean $\bar{x}$, the $\sigma_</\sigma_>$ multiplied by the exponential factor teaches us:

> If the populations of two groups are comparable ($N_A \sim N_B$) then the more elite the selection – namely, the higher the threshold $x_\star$ – the more the outcome will be exponentially dominated by the group with the higher variance.

We used IQ as a specific example, but these conclusions would of course hold for any observable whose possible outcomes are governed by the Gaussian distribution. □

7.7.2 Laplace's Method, Method of Stationary Phase, Steepest Descent

Exponential suppression The asymptotic methods we are about to encounter in this section rely on the fact that, the integrals we are computing really receive most of their contribution from a small region of the integration domain, because outside of this region the integrand itself becomes highly exponentially suppressed. A basic illustration of this is

$$I(\alpha) = \int_0^\alpha e^{-t} = 1 - e^{-\alpha}. \tag{7.7.20}$$

As $\alpha \to \infty$ we have $I(\infty) = 1$. Even though it takes an infinite range of integration to obtain 1, we see that most of the contribution ($\gg 99\%$) comes from $t = 0$ to $t \sim \mathcal{O}(10)$. For example, $e^{-5} \approx 6.7 \times 10^{-3}$ and $e^{-10} \approx 4.5 \times 10^{-5}$. You may also think about evaluating this integral numerically; what this shows is that it is not necessary to sample your integrand out to very large t to get an accurate answer.[59]

[59] In the Fourier transform section I pointed out how, if you merely need to resolve the coarser features of your wave profile, then provided the short wavelength modes do not have very large amplitudes, only the coefficients of the modes with longer wavelengths need to be known accurately. Here, we shall see some integrals only require us to know their integrands in a small region, if all we need is an approximate (but oftentimes highly accurate) answer. This is a good rule of thumb to keep in mind when tackling difficult, apparently complicated, problems in physics: focus on the most relevant contributions to the final answer, and often this will simplify the problem-solving process.

Laplace's Method We now turn to integrals of the form

$$I(\alpha) = \int_a^b f(t) e^{\alpha \phi(t)} \mathrm{d}t \tag{7.7.21}$$

where both f and ϕ are real. (There is no need to ever consider the complex f case since it can always be split into real and imaginary parts.) We will consider the $\alpha \to +\infty$ limit and try to extract the leading order behavior of the integral.

The main strategy goes roughly as follows. Find the location of the maximum of $\phi(t)$ – say it is at $t = c$. This can occur in between the limits of integration $a < c < b$ or at one of the end points $c = a$ or $c = b$. As long as $f(c) \neq 0$, we may expand both $f(t)$ and $\phi(t)$ around $t = c$. For simplicity we display the case where $a < c < b$:

$$I(\alpha) \sim e^{\alpha\phi(c)} \int_{c-\kappa}^{c+\kappa} \left(f(c) + (t-c) f'(c) + \dots\right) \exp\left(\alpha \left\{ \frac{\phi^{(p)}(c)}{p!} (t-c)^p + \dots \right\}\right) \mathrm{d}t, \tag{7.7.22}$$

where we have assumed the first non-zero derivative of ϕ is at the pth order, and κ is some small number ($\kappa < |b - a|$) such that the expansion can be justified, because the errors incurred from switching from $\int_a^b \to \int_{c-\kappa}^{c+\kappa}$ are exponentially suppressed. (Since $\phi(t = c)$ is maximum, $\phi'(c)$ is usually – but not always! – zero.) Then, term by term, these integrals, oftentimes after a change of variables, can be tackled using the Gamma function integral representation in eq. (6.2.44), namely

$$\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t, \qquad \mathrm{Re}(z) > 0, \tag{7.7.23}$$

by extending the former's limits to infinity, $\int_{c-\kappa}^{c+\kappa} \to \int_{-\infty}^{+\infty}$. This last step, like the expansion in eq. (7.7.22), is usually justified because the errors incurred are again exponentially small.

Examples The first example, where $\phi'(c) \neq 0$, is related to the integral representation of the parabolic cylinder function; for $\mathrm{Re}(\nu) > 0$,

$$I(\alpha) = \int_0^{100} t^{\nu-1} e^{-t^2/2} e^{-\alpha \cdot t} \mathrm{d}t. \tag{7.7.24}$$

Here, $\phi(t) = -t$ and its maximum is at the lower limit of integration. For large t the integrand is exponentially suppressed, and we expect the contribution to arise mainly for $t \in [0, \text{a few})$. In this region we may Taylor expand $e^{-t^2/2}$. Term-by-term, we may then extend the upper limit of integration to infinity, provided we can justify the errors incurred are small enough for $\alpha \gg 1$.

$$\begin{aligned} I(\alpha \to \infty) &\sim \int_0^\infty t^{\nu-1} \left(1 - \frac{t^2}{2} + \dots\right) e^{-\alpha \cdot t} \mathrm{d}t \\ &= \int_0^\infty \frac{(\alpha \cdot t)^{\nu-1}}{\alpha^{\nu-1}} \left(1 - \frac{(\alpha \cdot t)^2}{2\alpha^2} + \dots\right) e^{-(\alpha \cdot t)} \frac{\mathrm{d}(\alpha \cdot t)}{\alpha} \\ &= \frac{\Gamma(\nu)}{\alpha^\nu} \left(1 + \mathcal{O}\left(\alpha^{-2}\right)\right). \end{aligned} \tag{7.7.25}$$

The second example is

$$
\begin{aligned}
I(\alpha \to \infty) &= \int_0^{88} \frac{\exp(-\alpha \cosh(t))}{\sqrt{\sinh(t)}} \mathrm{d}t \\
&\sim \int_0^\infty \frac{\exp\left(-\alpha \left\{1 + \frac{t^2}{2} + \dots\right\}\right)}{\sqrt{t}\sqrt{1 + t^2/6 + \dots}} \mathrm{d}t \\
&\sim e^{-\alpha} \int_0^\infty \frac{(\alpha/2)^{1/4} \exp\left(-(\sqrt{\alpha/2}t)^2\right)}{\sqrt{\sqrt{\alpha/2}t}} \frac{\mathrm{d}(\sqrt{\alpha/2}t)}{\sqrt{\alpha/2}}.
\end{aligned} \tag{7.7.26}
$$

To obtain higher order corrections to this integral, we would have to be expand both the exp and the square root in the denominator. But the $t^2/2 + \dots$ comes multiplied with a α whereas the denominator is α-independent, so you'd need to make sure to keep enough terms to ensure you have captured all the contributions to the next- and next-to-next leading corrections, etc. We will be content with just the dominant behavior: we put $z \equiv t^2 \Rightarrow \mathrm{d}z = 2t\mathrm{d}t = 2\sqrt{z}\mathrm{d}t$.

$$
\begin{aligned}
\int_0^{88} \frac{\exp(-\alpha \cosh(t))}{\sqrt{\sinh(t)}} \mathrm{d}t &\sim \frac{e^{-\alpha}}{(\alpha/2)^{1/4}} \int_0^\infty z^{\left(1 - \frac{1}{4} - \frac{1}{2}\right) - 1} e^{-z} \frac{\mathrm{d}z}{2} \\
&= e^{-\alpha} \frac{\Gamma(1/4)}{2^{3/4} \alpha^{1/4}}.
\end{aligned} \tag{7.7.27}
$$

In both examples, the integrand really behaves very differently from the first few terms of its expanded version for $t \gg 1$, but the main point here is – it doesn't matter! The error incurred, for very large α, is exponentially suppressed anyway. If you care deeply about rigor, you may have to prove this assertion on a case-by-case basis; see Example 7 and 8 of Bender & Orszag's Chapter 6 [16] for careful discussions of two specific integrals.

Stirling's formula As an example of Laplace's method, let us use it to obtain a large $n \gg 1$ limit representation of the factorial $n! = \Gamma(n+1)$. In fact, what follows applies to any large positive n; i.e., not necessarily integer.

$$
\Gamma(n+1) = \int_0^\infty t^n e^{-t} \mathrm{d}t = \int_0^\infty e^{n \ln(t) - t} \mathrm{d}t. \tag{7.7.28}
$$

It appears here that $\phi(t) = \ln(t)$ and the maximum is at $t = \infty$. Actually, let us first re-scale $t \to n \cdot t$.

$$
\Gamma(n+1) = n \int_0^\infty e^{n(\ln(n \cdot t) - t)} \mathrm{d}t = n e^{n \ln n} \int_0^\infty e^{n(\ln(t) - t)} \mathrm{d}t. \tag{7.7.29}
$$

In this form, comparison with eq. (7.7.21) tells us $\phi(t) = \ln(t) - t$ and $f(t) = 1$. Moreover the maximum of $\phi(t)$ is at $t = 1$ because

$$
\phi'(t) = \frac{\mathrm{d}}{\mathrm{d}t} (\ln(t) - t) = \frac{1}{t} - 1 = 0. \tag{7.7.30}
$$

Now let us define $t = 1 + x$, so that x runs from -1 to $+\infty$. The factorial then reads

$$
\Gamma(n+1) = n^{n+1} \int_{-1}^\infty e^{n(\ln(1+x) - x - 1)} \mathrm{d}x. \tag{7.7.31}
$$

Now, the exponent begins as a Gaussian near $x \approx 0$ because

$$\ln(1+x) - x = -\frac{x^2}{2} + \frac{x^3}{3} + \mathcal{O}(x^4). \tag{7.7.32}$$

Inserting an explicit Taylor series of $\exp[n(\ln(1+x) - x + \frac{x^2}{2})]$ into our integral at hand while by extending the lower limit to $-\infty$ yields

$$\Gamma(n+1 \to \infty) \sim \frac{n^{n+1}}{e^n} \int_{-\infty}^{\infty} e^{-\frac{n}{2}x^2} \exp\left[n\left(\ln(1+x) - x + \frac{x^2}{2}\right)\right] \mathrm{d}x; \tag{7.7.33}$$

$$n! \sim \sqrt{2\pi}\frac{n^{n+\frac{1}{2}}}{e^n}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + \mathcal{O}(n^{-5})\right). \tag{7.7.34}$$

Problem 7.22. Fractional Error in Sterling's Approximation Can you estimate – say, using integration-by-parts – the fractional error incurred when extending the lower integration limit from -1 in (7.7.31) to $-\infty$ in eq. (7.7.33)? $\square$

Problem 7.23. What is the leading behavior of

$$I(\alpha) \equiv \int_0^{50.12345+e^{\sqrt{2}}+\pi^{\sqrt{e}}} e^{-\alpha \cdot t^{\pi}} \sqrt{1+\sqrt{t}}\mathrm{d}t \tag{7.7.35}$$

in the limit $\alpha \to +\infty$? And, how does the first correction scale with α? $\square$

Problem 7.24. What is the leading behavior of

$$I(\alpha) = \int_{-\pi/2}^{\pi/2} \frac{e^{-\alpha\cos(t)^2}}{(\cos(t))^p}\mathrm{d}t, \tag{7.7.36}$$

for $0 \leq p < 1$, in the limit $\alpha \to +\infty$? Note that there are two maximums of $\phi(t)$ here. $\square$

Method of Stationary Phase We now consider the case where the exponent is purely imaginary,

$$I(\alpha) = \int_a^b f(t) e^{i\alpha\phi(t)}\mathrm{d}t. \tag{7.7.37}$$

Here, both f and ϕ are real. As we did previously, we will consider the $\alpha \to +\infty$ limit and try to extract the leading order behavior of the integral.

What will be very useful, to this end, is the following lemma.

> The **Riemann-Lebesgue lemma** states that $I(\alpha \to \infty)$ in eq. (7.7.37) goes to zero provided: (I) $\int_a^b |f(t)|\mathrm{d}t < \infty$; (II) $\phi(t)$ is continuously differentiable; and (III) $\phi(t)$ is not constant over a finite range within $t \in [a,b]$.

We will not prove this result, but it is heuristically very plausible: as long as $\phi(t)$ is not constant, the $\exp[i\alpha\phi(t)]$ fluctuates wildly as $\alpha \to +\infty$ on the $t \in [a,b]$ interval. For large enough α, $f(t)$ will be roughly constant over 'each period' of $\exp[i\alpha\phi(t)]$, which in turn means $f(t)\exp[i\alpha\phi(t)]$ will integrate to zero over this same 'period'.

Case I: $\phi(t)$ has no turning points The first implication of the Riemann-Lebesgue lemma is that, if $\phi'(t)$ is not zero anywhere within $t \in [a,b]$; and as long as $f(t)/\phi'(t)$ is smooth enough within $t \in [a,b]$ and exists on the end points; then we can use integration-by-parts to show that the integral in eq. (7.7.37) has to scale as $1/\alpha$ as $\alpha \to \infty$.

$$\begin{aligned} I(\alpha) &= \int_a^b \frac{f(t)}{i\alpha\phi'(t)} \frac{\mathrm{d}}{\mathrm{d}t} e^{i\alpha\phi(t)} \mathrm{d}t \\ &= \frac{1}{i\alpha} \left\{ \left[\frac{f(t)}{\phi'(t)} e^{i\alpha\phi(t)} \right]_a^b - \int_a^b e^{i\alpha\phi(t)} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{f(t)}{\phi'(t)} \right) \mathrm{d}t \right\}. \end{aligned} \tag{7.7.38}$$

The integral on the second line within the curly brackets is one where Riemann-Lebesgue applies. Therefore it goes to zero relative to the (boundary) term preceding it, as $\alpha \to \infty$. Therefore what remains is

$$\int_a^b f(t) e^{i\alpha\phi(t)} \mathrm{d}t \sim \frac{1}{i\alpha} \left[\frac{f(t)}{\phi'(t)} e^{i\alpha\phi(t)} \right]_a^b, \qquad \alpha \to +\infty, \qquad \phi'(a \leq t \leq b) \neq 0. \tag{7.7.39}$$

Case II: $\phi(c)$ has at least one turning point If there is at least one point where the phase is stationary, $\phi'(a \leq c \leq b) = 0$, then provided $f(c) \neq 0$, we shall see that the dominant behavior of the integral in eq. (7.7.37) scales as $1/\alpha^{1/p}$, where p is the lowest order derivative of ϕ that is non-zero at $t = c$. Because $1/p < 1$, the $1/\alpha$ behavior we found above is sub-dominant to $1/\alpha^{1/p}$ – hence the need to analyze the two cases separately.

Let us, for simplicity, assume the stationary point is at a, the lower limit. We shall discover the leading behavior to be

$$\int_a^b f(t) e^{i\alpha\phi(t)} \mathrm{d}t \sim f(a) \exp\left(i\alpha\phi(a) \pm i\frac{\pi}{2p} \right) \frac{\Gamma(1/p)}{p} \left(\frac{p!}{\alpha|\phi^{(p)}(a)|} \right)^{1/p}, \tag{7.7.40}$$

where $\phi^{(p)}(a)$ is first non-vanishing derivative of $\phi(t)$ at the stationary point $t = a$; while the $+$ sign is to be chosen if $\phi^{(p)}(a) > 0$ and $-$ if $\phi^{(p)}(a) < 0$.

To understand eq. (7.7.40), we decompose the integral into

$$I(\alpha) = \int_a^{a+\kappa} f(t) e^{i\alpha\phi(t)} \mathrm{d}t + \int_{a+\kappa}^b f(t) e^{i\alpha\phi(t)} \mathrm{d}t. \tag{7.7.41}$$

The second integral scales as $1/\alpha$, as already discussed, since we assume there are no stationary points there. The first integral, which we shall denote as $S(\alpha)$, may be expanded in the following way provided κ is chosen appropriately:

$$S(\alpha) = \int_a^{a+\kappa} (f(a) + \dots) e^{i\alpha\phi(a)} \exp\left(\frac{i\alpha}{p!} (t-a)^p \phi^{(p)}(a) + \dots \right) \mathrm{d}t. \tag{7.7.42}$$

To convert the oscillating exp into a real, dampened one, let us rotate our contour. Around $t = a$, we may change variables to $t - a \equiv \rho e^{i\theta} \Rightarrow (t-a)^p = \rho^p e^{ip\theta} = i\rho^p$ (i.e., $\theta = \pi/(2p)$) if $\phi^{(p)}(a) > 0$; and $(t-a)^p = \rho^p e^{ip\theta} = -i\rho^p$ (i.e., $\theta = -\pi/(2p)$) if $\phi^{(p)}(a) < 0$. Since our stationary point is at the lower limit, this is for $\rho > 0$.[60]

$$
\begin{aligned}
& S(\alpha \to \infty) \\
& \sim f(a) e^{i\alpha\phi(a)} e^{\pm i\pi/(2p)} \int_0^{+\infty} \exp\left(-\frac{\alpha}{p!}|\phi^{(p)}(a)|\rho^p\right) \frac{\mathrm{d}(\rho^p)}{p \cdot \rho^{p-1}} \qquad (7.7.43) \\
& \sim \frac{1}{p} f(a) e^{i\alpha\phi(a)} e^{\pm i\pi/(2p)} \left(\frac{\alpha}{p!}|\phi^{(p)}(a)|\right)^{-1/p} \int_0^{+\infty} \left(\frac{\alpha}{p!}|\phi^{(p)}(a)|s\right)^{\frac{1}{p}-1} \exp\left(-\frac{\alpha}{p!}|\phi^{(p)}(a)|s\right) \mathrm{d}\left(\frac{\alpha}{p!}|\phi^{(p)}(a)|s\right).
\end{aligned}
$$

This establishes the result in eq. (7.7.40).

Problem 7.25. Bessel function Starting from the following integral representation of the Bessel function

$$
J_n(x) = \frac{1}{\pi}\int_0^{\pi} \cos(n\theta - x\sin\theta)\,\mathrm{d}\theta \qquad (7.7.44)
$$

where $n = 0, 1, 2, 3, \dots$, show that the leading behavior as $x \to +\infty$ is

$$
J_n(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right). \qquad (7.7.45)
$$

Hint: Express the cosine as the real part of an exponential. Note the stationary point is two-sided, but it is fairly straightforward to deform the contour appropriately.

Method of Steepest Descent We now allow our exponent $\phi = u + iv$ (in the $e^{\alpha\phi}$) to be complex.

$$
I(\alpha) = \int_C f(t) e^{\alpha u(t)} e^{i\alpha v(t)} \mathrm{d}t, \qquad (7.7.46)
$$

The f, u and v are real; C is some contour on the complex t plane; and as before we will study the $\alpha \to \infty$ limit. We will assume $u + iv$ forms an analytic function of t.

The method of steepest descent is the strategy to deform the contour C to some C' such that it lies on a constant-phase path – i.e., where the imaginary part of the exponent does not change along it.

$$
I(\alpha) = e^{i\alpha v} \int_{C'} f(t) e^{\alpha u(t)} \mathrm{d}t \qquad (7.7.47)
$$

The reason for doing so is that the constant phase contour also coincides with the steepest descent one of the real part of the exponent $u(t)$ – unless the contour passes through a saddle point, where more than one steepest descent paths can intersect. The dominant region of contribution to a given integral will be centered around the maximum value of $u(t)$ *along* a steepest descent path joining its upper and lower limits. We may then employ Laplace's method to obtain an asymptotic series.

[60] If p is even, and if the stationary point is not one of the end points, observe that we can choose $\theta = \pm(\pi/(2p) + \pi) \Rightarrow e^{ip\theta} = \pm i$ for the $\rho < 0$ portion of the contour – i.e., run a straight line rotated by θ through the stationary point – and the final result would simply be twice of eq. (7.7.40).

To understand this further we recall that the gradient is perpendicular to the lines of constant potential, i.e., the gradient points along the curves of most rapid change. Assuming $u+iv$ is an analytic function, and denoting $t = x+iy$ (for x and y real), the Cauchy-Riemann equations they obey

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v \tag{7.7.48}$$

means the dot product of their gradients is zero:

$$\vec{\nabla} u \cdot \vec{\nabla} v = \partial_x u \partial_x v + \partial_y u \partial_y v = \partial_y v \partial_x v - \partial_x v \partial_y v = 0. \tag{7.7.49}$$

A constant phase line – namely, the contour line where v is constant – is necessarily perpendicular to $\vec{\nabla} v$. But since $\vec{\nabla} u \cdot \vec{\nabla} v = 0$ in the relevant region of the 2D complex ($t = x+iy$)-plane where $u(t)+iv(t)$ is assumed to be analytic, a constant phase line must therefore be (anti)parallel to $\vec{\nabla} u$, the direction of most rapid change of the real amplitude $e^{\alpha \cdot u}$. To sum:

> A constant$-v(t)$ path on the complex t plane corresponds to a path of most-rapid-change of $u(t)$.

We will examine the following simple example:

$$I(\alpha) = \int_0^1 \ln(t) e^{i\alpha t} \mathrm{d}t. \tag{7.7.50}$$

We identify $f(t) = \ln(t)$ and $\phi(t) = it = ix - y$ for $t = x+iy$, where x and y are real and imaginary portions of t. Constant phase would therefore mean constant x. Thus, let's deform the contour $\int_0^1$ so it becomes the sum of the straight lines C_1, C_2 and C_3. C_1 runs from $t=0$ along the positive imaginary axis to infinity. C_2 runs horizontally from $i\infty$ to $i\infty+1$. Then C_3 runs from $i\infty+1$ back down to 1. There is no contribution from C_2 because the integrand there is $\ln(i\infty)e^{-\alpha\infty}$, which is zero for positive α.

$$I(\alpha) = i\int_0^\infty \ln(it) e^{-\alpha \cdot t} \mathrm{d}t - ie^{i\alpha} \int_0^\infty \ln(1+it) e^{-\alpha \cdot t} \mathrm{d}t. \tag{7.7.51}$$

Notice the exponents in both integrands have now zero (and therefore constant) phases. Moreover, for $\alpha \gg 1$, the dominant region contribution to the two integrals is $t \in [0, \mathcal{O}(\text{few})]$ since $e^{-\alpha \cdot t}$ is maximum at $t=0$ and falls off rapidly for large t.

$$\begin{aligned} I(\alpha) &= i\int_0^\infty \ln(i(\alpha \cdot t)/\alpha) e^{-(\alpha \cdot t)} \frac{\mathrm{d}(\alpha \cdot t)}{\alpha} - ie^{i\alpha} \int_0^\infty \ln(1+i(\alpha \cdot t)/\alpha) e^{-(\alpha \cdot t)} \frac{\mathrm{d}(\alpha \cdot t)}{\alpha} \\ &= i\int_0^\infty (\ln(z) - \ln(\alpha) + i\pi/2) e^{-z} \frac{\mathrm{d}z}{\alpha} - ie^{i\alpha} \int_0^\infty \left(i\frac{z}{\alpha} + \mathcal{O}(\alpha^{-2}) \right) e^{-z} \frac{\mathrm{d}z}{\alpha}. \end{aligned} \tag{7.7.52}$$

The only integral that remains unfamiliar is the first one

$$\begin{aligned} \int_0^\infty e^{-z} \ln(z) = \frac{\partial}{\partial \mu}\bigg|_{\mu=1} \int_0^\infty e^{-z} e^{(\mu-1)\ln(z)} &= \frac{\partial}{\partial \mu}\bigg|_{\mu=1} \int_0^\infty e^{-z} z^{\mu-1} \\ &= \Gamma'(1) = -\gamma_{\mathrm{E}} \end{aligned} \tag{7.7.53}$$

The $\gamma_{\mathrm{E}} = 0.577216\ldots$ is known as the Euler-Mascheroni constant. At this point,

$$\int_0^1 \ln(t) e^{i\alpha \cdot t} \mathrm{d}t \sim \frac{i}{\alpha} \left(-\gamma_{\mathrm{E}} - \ln(\alpha) + i\frac{\pi}{2} - \frac{ie^{i\alpha}}{\alpha} \left\{ 1 + \mathcal{O}(\alpha^{-1}) \right\} \right), \qquad \alpha \to +\infty. \tag{7.7.54}$$

Problem 7.26. Perform an asymptotic expansion of

$$I(k) \equiv \int_{-1}^{+1} e^{ikx^2} \mathrm{d}x \tag{7.7.55}$$

using the steepest descent method. Hint: Find the point $t = t_0$ on the real line where the phase is stationary. Then deform the integration contour such that it passes through t_0 and has a stationary phase everywhere. Can you also tackle $I(k)$ using integration-by-parts? $\square$

7.7.3 Wronskians and Completeness of Bessel Functions

In this section we will encounter the Wronskian. It will be used in conjunction with the asymptotic results for $J_\nu(z)$,

$$J_{\pm\nu}(z) \sim \sqrt{\frac{2}{\pi z}} \cos\left(z \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right)\left(1 + \mathcal{O}(z^{-1})\right); \tag{7.7.56}$$

to derive generalizations of the completeness relations of Bessel $J_\nu(kr)$.

Wronskian The Wronskian Wr of two functions f and g with respect to some variable z, is defined as

$$\mathrm{Wr}_z\,(f, g) = f \cdot \partial_z g - g \cdot \partial_z f. \tag{7.7.57}$$

This object is important in the study of the solution to ordinary differential equations. Let us suppose the pair $f_{1,2}(z)$ obeys

$$p_2(z) f''(z) + p_1(z) f'(z) + p_0(z) f(z) = 0 \tag{7.7.58}$$

We then record the following facts.

- If $\mathrm{Wr}_z(f_1, f_2) \neq 0$, then $f_{1,2}(z)$ are linearly independent. Equivalently, if these $f_{1,2}$ are linearly dependent, their Wronskian $\mathrm{Wr}_z(f_1(z), f_2(z))$ is necessarily zero. If $\mathrm{Wr}_z(f_1(z), f_2(z))$ is zero, however, it does *not* necessarily imply they are linearly dependent – and example is the pair $f_1(x) = x|x|$ and $f_2(x) = x^2$.

- Furthermore, it is not difficult to verify, if eq. (7.7.58) is taken into account, the Wronskian itself obeys the 1st order ODE

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathrm{Wr}_z(f_1, f_2) = -\frac{p_1(z)}{p_2(z)}\mathrm{Wr}_z(f_1, f_2), \tag{7.7.59}$$

  which immediately implies the Wronskian can be determined, up to an overall multiplicative constant, without the need to know explicitly the pair of homogeneous solutions $f_{1,2}$,

$$\mathrm{Wr}_z(f_1, f_2) = W_0 \exp\left(-\int^z \frac{p_1(z'')}{p_2(z'')}\mathrm{d}z''\right), \qquad W_0 = \text{constant}. \tag{7.7.60}$$

- If we "rotate" from one pair of linearly independent solutions (f_1, f_2) to another (g_1, g_2) via a constant invertible matrix ${M_\mathrm{I}}^\mathrm{J}$,

$$f_\mathrm{I}(z) = {M_\mathrm{I}}^\mathrm{J} g_\mathrm{J}(z), \qquad \mathrm{I}, \mathrm{J} \in \{1, 2\}, \ \det {M_\mathrm{I}}^\mathrm{J} \neq 0; \tag{7.7.61}$$

then

$$\mathrm{Wr}_z(f_1, f_2) = \left(\det {M_\mathrm{I}}^\mathrm{J}\right) \mathrm{Wr}_z(g_1, g_2). \tag{7.7.62}$$

Let us now compute the Wronskian for the pair $J_{\pm\nu}(z)$. For Bessel equation in eq. (5.4.117) with m replaced with μ, we may identify

$$p_2 = 1 \qquad \text{and} \qquad p_2 = z^{-1}. \tag{7.7.63}$$

The solution in eq. (7.7.60) indicates,

$$\mathrm{Wr}_z \left(J_\nu(z), J_{-\nu}(z)\right) \propto \frac{1}{z}. \tag{7.7.64}$$

According to eq. (5.4.118), $J_{\pm\mu}(z)$ is proportional to z^ν multiplied by an even power series. Hence, the derivative of $J_\nu(z)$ goes as $z^{\nu-1}$ multiplied by an even power series. When we multiply $J_\alpha(z) \cdot J'_\beta(z)$, therefore, we expect it to yield an expression that goes as $z^{\alpha+\beta-1}$ multiplied by an even power series. But eq. (7.7.64) tells us that the Wronskian scales precisely as z^{-1}. Hence, the contributions to the higher powers of z arising from the even power series portion of eq. (5.4.118) must cancel out.

$$\begin{aligned} \mathrm{Wr}_z \left(J_\nu(z), J_{-\nu}(z)\right) &= \frac{(z/2)^\nu}{\nu!} \cdot \frac{\mathrm{d}}{\mathrm{d}z} \frac{(z/2)^{-\nu}}{(-\nu)!} - \frac{(z/2)^{-\nu}}{(-\nu)!} \cdot \frac{\mathrm{d}}{\mathrm{d}z} \frac{(z/2)^\nu}{\nu!} && (7.7.65) \\ &= -\frac{2\nu}{z \cdot (-\nu)!\nu!}. && (7.7.66) \end{aligned}$$

In the discussion leading up to eq. (6.2.47) below, we derive the result $\nu! = \nu \cdot (\nu - 1)!$ and $(\nu - 1)!(-\nu)! = \pi / \sin(\pi z)$. This means

$$\mathrm{Wr}_z \left(J_\nu(z), J_{-\nu}(z)\right) = -\frac{2\sin(\pi \cdot \nu)}{z\pi}. \tag{7.7.67}$$

The pair $J_{\pm\nu}(z)$ are linearly independent as long as ν is not an integer. This leads us to the following problem that investigates the relationship between $J_\nu(k \cdot r)$ and $J_{-\nu}(q \cdot r)$.

Problem 7.27. We will now discover

$$\int_0^\infty \mathrm{d}r \cdot r J_{-\nu}(kr) J_\nu(qr) = \cos(\pi\nu) \frac{\delta(k - q)}{\sqrt{k \cdot q}} + \mathrm{Pr}\left(\frac{2\sin[\pi\nu]}{\pi(q^2 - k^2)} \left(\frac{q}{k}\right)^\nu\right), \quad k, q > 0. \tag{7.7.68}$$

First, verify eq. (7.7.59) and the results

$$\begin{aligned} \mathrm{Wr}_r \left(\sqrt{r} J_\mu(kr), \sqrt{r} J_\nu(qr)\right) &= (qr) J_\mu(kr) J'_\nu(qr) - (kr) J_\mu(kr) J'_\nu(qr), && (7.7.69) \\ \frac{\mathrm{d}}{\mathrm{d}r} \mathrm{Wr}_r \left(\sqrt{r} J_\mu(kr), \sqrt{r} J_\nu(qr)\right) &= \left((k^2 - q^2) r - \frac{\mu^2 - \nu^2}{r}\right) J_\mu(kr) J_\nu(qr). && (7.7.70) \end{aligned}$$

Can you use the second line to deduce eq. (7.7.68)? **YZ: Incomplete!** □

7.8 Fourier Transforms

We have seen how the Fourier transform pairs arise within the linear algebra of states represented in some position basis corresponding to some D dimensional infinite flat space. Denoting the state/function as f, and using Cartesian coordinates, the pairs read

$$f(\vec{x}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \widetilde{f}(\vec{k}) e^{i\vec{k}\cdot\vec{x}} \tag{7.8.1}$$

$$\widetilde{f}(\vec{k}) = \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} f(\vec{x}) e^{-i\vec{k}\cdot\vec{x}} \tag{7.8.2}$$

[61]By inserting eq. (7.8.2) into eq. (7.8.1) we may obtain the integral representation of the δ-function

$$\delta^{(D)}(\vec{x} - \vec{x}') = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{7.8.3}$$

In physical applications, almost any function residing in infinite space can be Fourier transformed. The meaning of the Fourier expansion in eq. (7.8.1) is that of resolving a given profile $f(\vec{x})$ – which can be a wave function of an elementary particle, or a component of an electromagnetic signal – into its basis wave vectors. Remember the magnitude of the wave vector is the reciprocal of the wave length, $|\vec{k}| \sim 1/\lambda$.[62] Heuristically, this indicates the coarser features in the profile – those you'd notice at first glance – come from the modes with longer wavelengths, small $|\vec{k}|$ values. The finer features requires us to know accurately the Fourier coefficients of the waves with very large $|\vec{k}|$, i.e., short wavelengths.

In many physical problems we only need to understand the coarser features, the Fourier modes up to some inverse wavelength $|\vec{k}| \sim \Lambda_{\mathrm{UV}}$. (This in turn means Λ_{UV} lets us *define* what we mean by coarse ($\equiv |\vec{k}| < \Lambda_{\mathrm{UV}}$) and fine ($\equiv |\vec{k}| > \Lambda_{\mathrm{UV}}$) features.) In fact, it is often *not possible* to experimentally probe the Fourier modes of very small wavelengths, or equivalently, phenomenon at very short distances, because it would expend too much resources to do so. For instance, it much easier to study the overall appearance of the desk you are sitting at – its physical size, color of its surface, etc. – than the atoms that make it up. This is also the essence of why it is very difficult to probe quantum aspects of gravity: humanity does not currently have the resources to construct a powerful enough accelerator to understand elementary particle interactions at the energy scales where quantum gravity plays a significant role.

Problem 7.28. A simple example illustrating how Fourier transforms help us understand the coarse ($\equiv$ long wavelength) versus fine ($\equiv$ short wavelength) features of some profile is to consider a Gaussian of width σ, but with some small oscillations added on top of it.

$$f(x) = \exp\left(-\frac{1}{2}\left(\frac{x - x_0}{\sigma}\right)^2\right)(1 + \epsilon\sin(\omega x)), \qquad |\epsilon| \ll 1. \tag{7.8.4}$$

[61]Always check the Fourier conventions of the literature you are reading. Here, we have a $1/(2\pi)$ for every (inverse space) $k-$dimension; and no factors of (2π)s in the real space ones. This is not a universal convention.

[62]In older books, Fourier space is often dubbed 'reciprocal space'.

Assume that the wavelength of the oscillations is much shorter than the width of the Gaussian, $1/\omega \ll \sigma$. Find the Fourier transform $\widetilde{f}(k)$ of $f(x)$ and comment on how, discarding the short wavelength coefficients of the Fourier expansion of $f(x)$ still reproduces its gross features, namely the overall shape of the Gaussian itself. Notice, however, if ϵ is not small, then the oscillations – and hence the higher $|\vec{k}|$ modes – cannot be ignored. □

Problem 7.29. Find the inverse Fourier transform of the "top hat" in 3 dimensions:

$$\widetilde{f}(\vec{k}) \equiv \Theta\left(\Lambda - |\vec{k}|\right) \tag{7.8.5}$$

$$f(\vec{x}) = ? \tag{7.8.6}$$

Bonus problem: Can you do it for arbitrary D dimensions? Hint: You may need to know how to write down spherical coordinates in D dimensions. Then examine eq. 10.9.4 of the NIST page here. □

Problem 7.30. Show that the Fourier transform of a Gaussian in $D-$dimensions

$$f(\vec{x}) = \exp\left(-x^i M_{ij} x^j\right), \tag{7.8.7}$$

where M_{ij} is a real symmetric $D \times D$ matrix with positive eigenvalues is

$$\pi^{D/2}\left(\det \widehat{M}\right)^{-\frac{1}{2}}. \tag{7.8.8}$$

Your result should justify the statement:

The Fourier transform of a Gaussian is another Gaussian.

Hint: Diagonalize M_{ij}. □

Problem 7.31. Real functions If $f(\vec{x})$ is real, show that $\widetilde{f}(\vec{k})^* = \widetilde{f}(-\vec{k})$. Similarly, if $f(\vec{x})$ is a real periodic function in D-space, show that the Fourier series coefficients in eq. (5.3.31) and (5.3.32) obey $\widetilde{f}(n^1, \dots, n^D)^* = \widetilde{f}(-n^1, \dots, -n^D)$. □

Problem 7.32. Parity In this problem, we shall see how the parity (anti-)symmetry on a given function $f(x)$ is inherited by its Fourier transform counterpart. That is, proceed to show that

$$f(-x) = f(x) \qquad \text{iff} \qquad \widetilde{f}(-k) = \widetilde{f}(k); \tag{7.8.9}$$

which in turn implies

$$\widetilde{f}(k) = \widetilde{f}(|k|), \tag{7.8.10}$$

$$f(x) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} \cos(k \cdot x)\, \widetilde{f}(|k|). \tag{7.8.11}$$

Likewise, show that

$$f(-x) = -f(x) \qquad \text{iff} \qquad \widetilde{f}(-k) = -\widetilde{f}(k); \tag{7.8.12}$$

which in turn implies

$$\widetilde{f}(k) = \text{sgn}(k) \cdot \widetilde{f}(|k|), \tag{7.8.13}$$

$$f(x) = i \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} \sin(k \cdot x)\, \widetilde{f}(|k|). \tag{7.8.14}$$

Can you write down the analogous Fourier/inverse Fourier series for even and odd parity periodic functions on $\mathbb{R}$? □

Cosine and Sine Transforms Equations (7.8.11) and (7.8.14) can be written as

$$f(x) = \int_0^\infty \frac{\mathrm{d}k}{\pi} \cos(kx)\widetilde{f}(k), \qquad (\text{Even } f(x)), \tag{7.8.15}$$

$$f(x) = i \int_0^\infty \frac{\mathrm{d}k}{\pi} \sin(kx)\widetilde{f}(k), \qquad (\text{Odd } f(x)); \tag{7.8.16}$$

and the inverse transforms as

$$\widetilde{f}(k) = \int_0^\infty \mathrm{d}x \cos(kx) f(x), \qquad (\text{Even } f(x)), \tag{7.8.17}$$

$$\widetilde{f}(k) = -i \int_0^\infty \mathrm{d}x \sin(kx) f(x), \qquad (\text{Odd } f(x)). \tag{7.8.18}$$

The corresponding completeness relations are

$$\int_0^\infty \frac{\mathrm{d}k}{\pi} \cos(kx)\cos(kx') = \delta(k - k'), \tag{7.8.19}$$

$$\int_0^\infty \frac{\mathrm{d}k}{\pi} \sin(kx)\sin(kx') = \delta(k - k'). \tag{7.8.20}$$

Sine and cosine are orthogonal:

$$\int_0^\infty \frac{\mathrm{d}k}{\pi} \cos(kx)\sin(kx') = 0. \tag{7.8.21}$$

Notice these integrals only involve the positive real line; i.e., all relevant variables x, x' and k can be taken as positive. This is due to the (anti)symmetry of the $f(x)$ under parity $x \leftarrow -x$.

Moreover, if we were given a function $f(x)$ defined only on the positive real line, we may extend it to the entire real line by defining it as an even or odd function – namely, $f(x < 0) \equiv f(-x)$ for even or $f(x < 0) \equiv -f(-x)$ for odd. Then, these cosine and sine transforms will continue to apply. Since the choice of even or odd 'continuation' is arbitrary, such considerations therefore indicate, either the pair of equations (7.8.15) and (7.8.17) or the pair (7.8.16) and (7.8.18) are equally valid pair of transforms for any function defined on the half line $\mathbb{R}^+$.

Problem 7.33. Fourier Transforms of Derivatives One reason why Fourier transforms are important for solving differential equations, is that, in Fourier space, derivatives become $\vec{k}-$vectors and differential equations transform into algebraic ones. Explain the correspondence

$$\partial_{x^i} \leftrightarrow i k^i \tag{7.8.22}$$

by Fourier decomposing $\partial_i f(\vec{x})$; i.e., write down the right hand side of eq. (7.8.1) after replacing its left hand side with $\partial_{x^i} f(\vec{x})$. □

Problem 7.34. Relation to Hankel Transform Explain why the triplet equations (7.8.15), (7.8.17), and (7.8.19); as well as equations (7.8.16), (7.8.18), and (7.8.20) are special cases of equations (5.4.130), (5.4.131), (5.4.136), and (5.4.137). Hint: What is Bessel $J_\nu(z)$ when $\nu = \pm 1/2$? $\square$

Problem 7.35. For a complex $f(\vec{x})$, show that

$$\int_{\mathbb{R}^D} \mathrm{d}^D x |f(\vec{x})|^2 = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} |\widetilde{f}(\vec{k})|^2, \tag{7.8.23}$$

$$\int_{\mathbb{R}^D} \mathrm{d}^D x M^{ij} \partial_i f(\vec{x})^* \partial_j f(\vec{x}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} M^{ij} k_i k_j |\widetilde{f}(\vec{k})|^2, \tag{7.8.24}$$

where you should assume the matrix M^{ij} does not depend on position $\vec{x}$.

Convolution Theorem Define the convolution of two functions F and G as

$$f(\vec{x}) \equiv \int_{\mathbb{R}^D} \mathrm{d}^D y F(\vec{x} - \vec{y}) G(\vec{y}). \tag{7.8.25}$$

Prove that the Fourier transform of the convolution between F and G is the product of their individual Fourier transforms:

$$\widetilde{f}(\vec{k}) = \widetilde{F}(\vec{k}) \widetilde{G}(\vec{k}). \tag{7.8.26}$$

You may need to employ the integral representation of the δ-function; or invoke linear algebraic arguments. $\square$

Problem 7.36. Yukawa Potential In 3D space, what is the Fourier transform of

$$V(r) \equiv \frac{\exp(-m \cdot r)}{4\pi r}, \tag{7.8.27}$$

where $r \equiv |\vec{x}|$? Hint: See Problem (7.6). $\square$

7.8.1 Damped Driven Simple Harmonic Oscillator

Many physical problems – from RLC circuits to perturbative Quantum Field Theory (pQFT) – reduces to some variant of the driven damped harmonic oscillator.[63] We will study it in the form of the 2nd order ordinary differential equation (ODE)

$$m \cdot \ddot{x}(t) + f \cdot \dot{x}(t) + k \cdot x(t) = F(t), \qquad f, k > 0, \tag{7.8.28}$$

[63] In pQFT the different Fourier modes of (possibly multiple) fields are the harmonic oscillators. If the equations are nonlinear, that means modes of different momenta drive/excite each other. Similar remarks apply for different fields that appear together in their differential equations. If you study fields residing in an expanding universe like ours, you'll find that the expansion of the universe provides friction and hence each Fourier mode behaves as a damped oscillator. The quantum aspects include the perspective that the Fourier modes themselves are both waves propagating in spacetime as well as particles that can be localized, say by the silicon wafers of the detectors at the Large Hadron Collider (LHC) in Geneva. These particles – the Fourier modes – can also be created from and absorbed by the vacuum.

where each dot represents a time derivative; for e.g., $\ddot{x} \equiv \mathrm{d}^2x/\mathrm{d}t^2$. You can interpret this equation as Newton's second law (in 1D) for a particle with trajectory $x(t)$ of mass m. The f term corresponds to some frictional force that is proportional to the velocity of the particle itself; the $k > 0$ refers to the spring constant, if the particle is in some locally-parabolic potential; and $F(t)$ is some other time-dependent external force. For convenience we will divide both sides by m and re-scale the constants and $F(t)$ so that our ODE now becomes

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \Omega^2 x(t) = F(t), \qquad \Omega, \gamma > 0. \tag{7.8.29}$$

We will perform a Fourier analysis of this problem by transforming both the trajectory and the external force,

$$x(t) = \int_{-\infty}^{+\infty} \widetilde{x}(\omega) e^{-i\omega t} \frac{\mathrm{d}\omega}{2\pi}, \qquad F(t) = \int_{-\infty}^{+\infty} \widetilde{F}(\omega) e^{-i\omega t} \frac{\mathrm{d}\omega}{2\pi}. \tag{7.8.30}$$

I will first find the particular solution $x_p(t)$ for the trajectory due to the presence of the external force $F(t)$, through the Green's function $G(t-t')$ of the differential operator $(\mathrm{d}/\mathrm{d}t)^2 + 2\gamma(\mathrm{d}/\mathrm{d}t) + \Omega^2$. I will then show the fundamental importance of the Green's function by showing how you can obtain the homogeneous solution to the damped simple harmonic oscillator equation, once you have specified the position $x(t')$ and velocity $\dot{x}(t')$ at some initial time t'. (This is, of course, to be expected, since we have a 2nd order ODE.)

First, we begin by taking the Fourier transform of the ODE itself.

Problem 7.37. Damped SHO in Frequency Space Show that, in frequency space, eq. (7.8.29) is

$$\left(-\omega^2 - 2i\omega\gamma + \Omega^2\right) \widetilde{x}(\omega) = \widetilde{F}(\omega). \tag{7.8.31}$$

In effect, each time derivative $\mathrm{d}/\mathrm{d}t$ is replaced with $-i\omega$. We see that the differential equation in eq. (7.8.29) is converted into an algebraic one in eq. (7.8.31). This is one reason why the Fourier transform is useful for solving differential equations. □

Inhomogeneous (particular) solution For $F \neq 0$, we may infer from eq. (7.8.31) that the particular solution – the part of $\widetilde{x}(\omega)$ that is due to $\widetilde{F}(\omega)$ – is

$$\widetilde{x}_p(\omega) = \frac{\widetilde{F}(\omega)}{-\omega^2 - 2i\omega\gamma + \Omega^2}, \tag{7.8.32}$$

which in turn implies

$$\begin{aligned} x_p(t) &= \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega t} \frac{\widetilde{F}(\omega)}{-\omega^2 - 2i\omega\gamma + \Omega^2} \\ &= \int_{-\infty}^{+\infty} \mathrm{d}t' G(t-t') F(t'); \end{aligned} \tag{7.8.33}$$

where

$$G(t-t') = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} \frac{e^{-i\omega(t-t')}}{-\omega^2 - 2i\omega\gamma + \Omega^2}. \tag{7.8.34}$$

To get to eq. (7.8.33) we have inserted the inverse Fourier transform

$$\widetilde{F}(\omega) = \int_{-\infty}^{+\infty} \mathrm{d}t' e^{-i\omega t'} F(t'). \tag{7.8.35}$$

Problem 7.38. Damped SHO: Retarded Green's Function Show that the Green's function in eq. (7.8.34) obeys the damped harmonic oscillator equation eq. (7.8.29), but driven by a impulsive force ("point-source-at-time t'")

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}}{\mathrm{d}t} + \Omega^2\right) G(t-t') = \left(\frac{\mathrm{d}^2}{\mathrm{d}t'^2} - 2\gamma\frac{\mathrm{d}}{\mathrm{d}t'} + \Omega^2\right) G(t-t') = \delta(t-t'), \tag{7.8.36}$$

so that eq. (7.8.33) can be interpreted as the $x_p(t)$ sourced/driven by the superposition of impulsive forces over all times, weighted by $F(t')$. Explain why the differential equation with respect to t' has a different sign in front of the 2γ term. By "closing the contour" appropriately, verify that eq. (7.8.34) yields

$$G(t-t') = \Theta(t-t')\mathcal{G}(t-t'), \tag{7.8.37}$$

$$\mathcal{G}(t-t') \equiv e^{-\gamma(t-t')}\frac{\sin\left(\sqrt{\Omega^2-\gamma^2}(t-t')\right)}{\sqrt{\Omega^2-\gamma^2}}. \tag{7.8.38}$$

Explain what happens when $\Omega^2 < \gamma^2$. $\square$

Distributional Calculus We may check this Green's function result by direct differentiation, provided we invoke some $\delta-$function identities. Firstly, if $\mathrm{d}/\mathrm{d}t$ is denoted by an over-dot, the first and second derivatives are

$$\dot{G}(t-t') = \delta(t-t')\mathcal{G}(t-t') + \Theta(t-t')\dot{\mathcal{G}}(t-t'); \tag{7.8.39}$$

and

$$\ddot{G} = \delta'(t-t')\mathcal{G} + 2\delta(t-t')\dot{\mathcal{G}} + \Theta(t-t')\ddot{\mathcal{G}}. \tag{7.8.40}$$

The $\delta(t-t')$ enforces $t = t'$, for otherwise otherwise it is zero, anything multiplying it must also be evaluated at $t = t'$. Furthermore, to deal with the $\delta'(t-t')$ term, we first recognize that – by the same reasoning in the previous sentence – $z\delta(z) = 0$. Differentiating this expression once,

$$z\delta'(z) = -\delta(z). \tag{7.8.41}$$

Multiplying this relation by more powers of z informs us that, for $n \geq 2$,

$$z^n\delta'(z) = 0. \tag{7.8.42}$$

As long as $f(z)$ may be Taylor expanded about $z = 0$, this teaches us that

$$f(z)\delta'(z) = \big(f(0) + z\cdot f'(0) + (z^2/2)\cdot f''(0) + \dots\big)\,\delta'(z) \tag{7.8.43}$$
$$= f(0)\delta'(z) - f'(0)\delta(z). \tag{7.8.44}$$

Altogether, we now gather

$$\begin{aligned}&\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}+2\gamma\frac{\mathrm{d}}{\mathrm{d}t}+\Omega^2\right)G(t-t')\\&=\delta'(t-t')\mathcal{G}(0)+\delta(t-t')\left(2\gamma\mathcal{G}(0)+\dot{\mathcal{G}}(0)\right)+\Theta(t-t')\left(\ddot{\mathcal{G}}+2\gamma\dot{\mathcal{G}}+\Omega^2\mathcal{G}\right)\end{aligned} \tag{7.8.45}$$

For the Green's function itself to obey $\ddot{G}+2\gamma\dot{G}+\Omega^2G=\delta(t-t')$, that means the $\delta'(t-t')$ and $\Theta(t-t')$ terms must both be zero; i.e.,

$$\mathcal{G}(0)=0; \tag{7.8.46}$$

and $\mathcal{G}(t-t')$ must be a homogeneous solution of the damped SHO equation,

$$\ddot{\mathcal{G}}+2\gamma\dot{\mathcal{G}}+\Omega^2\mathcal{G}=0. \tag{7.8.47}$$

What remains is the $\delta(t-t')$ term, which we need its coefficient to be unity. Because eq. (7.8.46) eliminates the $2\gamma\mathcal{G}(0)$, we are left with

$$\frac{\mathrm{d}\mathcal{G}(t=t')}{\mathrm{d}t}=-\frac{\mathrm{d}\mathcal{G}(t'=t)}{\mathrm{d}t'}=1. \tag{7.8.48}$$

Problem 7.39. Checking eq. (7.8.38) Verify equations (7.8.46), (7.8.47) and (7.8.48) directly using eq. (7.8.38). □

Problem 7.40. Damped SHO: Dirichlet Green's Function Verify that

$$G_s(s,s')=\frac{e^{-\gamma(s-s')}}{\sqrt{\Omega^2-\gamma^2}}\frac{\sin\left(\sqrt{\Omega^2-\gamma^2}(s_<-t_\mathrm{i})\right)\cdot\sin\left(\sqrt{\Omega^2-\gamma^2}(s_>-t_\mathrm{f})\right)}{\sin\left(\sqrt{\Omega^2-\gamma^2}(t_\mathrm{f}-t_\mathrm{i})\right)}, \tag{7.8.49}$$

with $s_<=\min(s,s')$ and $s_>=\max(s,s')$, obeys the Dirichlet boundary conditions at the initial and final times, i.e., t_i and t_f:

$$G_s(s=t_\mathrm{f},t_\mathrm{i},s')=0=G_s(s,s'=t_\mathrm{f},t_\mathrm{i}); \tag{7.8.50}$$

as well as the equations

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}s^2}+2\gamma\frac{\mathrm{d}}{\mathrm{d}s}+\Omega^2\right)G_s(s,s')=\delta(s-s')=\left(\frac{\mathrm{d}^2}{\mathrm{d}s'^2}-2\gamma\frac{\mathrm{d}}{\mathrm{d}s'}+\Omega^2\right)G_s(s,s'). \tag{7.8.51}$$

What choice of contour in eq. (7.8.34) would produce the Dirichlet Green's function in eq. (7.8.49)? □

Causality Notice the Green's function obeys causality. Any force $F(t')$ from the future of t, i.e., $t'>t$, does not contribute to the trajectory in eq. (7.8.33) due to the step function $\Theta(t-t')$ in eq. (7.8.37). That is,

$$x_p(t)=\int_{-\infty}^{t}\mathrm{d}t'\mathcal{G}(t-t')F(t'). \tag{7.8.52}$$

Initial value formulation and homogeneous solutions With the Green's function $G(t-t')$ at hand and the particular solution sourced by $F(t)$ understood – let us now move on to use $G(t-t')$ to obtain the homogeneous solution of the damped simple harmonic oscillator. Let $x_h(t)$ be the homogeneous solution satisfying

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}+2\gamma\frac{\mathrm{d}}{\mathrm{d}t}+\Omega^2\right)x_h(t)=0. \tag{7.8.53}$$

We then start by examining the following integral

$$\begin{aligned} I(t,t') \equiv \int_{t'}^{\infty}\mathrm{d}t''\Big\{ & x_h(t'')\left(\frac{\mathrm{d}^2}{\mathrm{d}t''^2}-2\gamma\frac{\mathrm{d}}{\mathrm{d}t''}+\Omega^2\right)G(t-t'') \\ & - G(t-t'')\left(\frac{\mathrm{d}^2}{\mathrm{d}t''^2}+2\gamma\frac{\mathrm{d}}{\mathrm{d}t''}+\Omega^2\right)x_h(t'')\Big\}. \end{aligned} \tag{7.8.54}$$

Using the equations (7.8.36) and (7.8.53) obeyed by $G(t-t')$ and $x_h(t)$, we may immediately infer that

$$I(t,t')=\int_{t'}^{\infty}\mathrm{d}t'' x_h(t'')\delta(t-t'')=\Theta(t-t')x_h(t). \tag{7.8.55}$$

(The step function arises because, if t lies outside of $[t',\infty)$, and is therefore less than t', the integral will not pick up the δ-function contribution and the result would be zero.) On the other hand, we may in eq. (7.8.54) cancel the Ω^2 terms, and then integrate-by-parts one of the derivatives from the $\ddot{G}$, $\dot{G}$, and $\ddot{x}_h$ terms.

$$\begin{aligned} I(t,t') = & \left[x_h(t'')\left(\frac{\mathrm{d}}{\mathrm{d}t''}-2\gamma\right)G(t-t'')-G(t-t'')\frac{\mathrm{d}x_h(t'')}{\mathrm{d}t''}\right]_{t''=t'}^{t''=\infty} \\ & + \int_{t'}^{\infty}\mathrm{d}t''\Big(-\frac{\mathrm{d}x_h(t'')}{\mathrm{d}t''}\frac{\mathrm{d}G(t-t'')}{\mathrm{d}t''}+2\gamma\frac{\mathrm{d}x_h(t'')}{\mathrm{d}t''}G(t-t'') \\ & + \frac{\mathrm{d}G(t-t'')}{\mathrm{d}t''}\frac{\mathrm{d}x_h(t'')}{\mathrm{d}t''}-2\gamma G(t-t'')\frac{\mathrm{d}x_h(t'')}{\mathrm{d}t''}\Big). \end{aligned} \tag{7.8.56}$$

Observe that the integral on the second and third lines is zero because the integrands cancel. Moreover, because of the $\Theta(t-t')$ (namely, causality), we may assert $\lim_{t'\to\infty}G(t-t')=G(t'>t)=0$. Recalling eq. (7.8.55), we have arrived at

$$\Theta(t-t')x_h(t)=G(t-t')\frac{\mathrm{d}x_h(t')}{\mathrm{d}t'}+\left(2\gamma G(t-t')+\frac{\mathrm{d}G(t-t')}{\mathrm{d}t}\right)x_h(t'). \tag{7.8.57}$$

Because we have not made any assumptions about our trajectory – except it satisfies the homogeneous equation in eq. (7.8.53) – we have shown that, for an arbitrary initial position $x_h(t')$ and velocity $\dot{x}_h(t')$, the Green's function $G(t-t')$ can in fact also be used to obtain the homogeneous solution for $t>t'$, where $\Theta(t-t')=1$. In particular, since $x_h(t')$ and $\dot{x}_h(t')$ are freely specifiable, they must be completely independent of each other. Furthermore, the right hand side of eq. (7.8.57) must span the 2-dimensional space of solutions to eq. (7.8.53). Therefore, the coefficients of $x_h(t')$ and $\dot{x}_h(t')$ must in fact be the two linearly independent homogeneous solutions to $x_h(t)$,

$$x_h^{(1)}(t) = G(t > t') = e^{-\gamma(t-t')} \frac{\sin\left(\sqrt{\Omega^2 - \gamma^2}(t - t')\right)}{\sqrt{\Omega^2 - \gamma^2}}, \tag{7.8.58}$$

$$\begin{aligned} x_h^{(2)}(t) &= 2\gamma G(t > t') + \partial_t G(t > t') \\ &= e^{-\gamma(t-t')} \left(\frac{\gamma \cdot \sin\left(\sqrt{\Omega^2 - \gamma^2}(t - t')\right)}{\sqrt{\Omega^2 - \gamma^2}} + \cos\left(\sqrt{\Omega^2 - \gamma^2}(t - t')\right) \right). \end{aligned} \tag{7.8.59}$$

[64]That $x_h^{(1,2)}$ must be independent for any $\gamma > 0$ and Ω^2 is worth reiterating, because this is a potential issue for the damped harmonic oscillator equation when $\gamma = \Omega$. We can check directly that, in this limit, $x_h^{(1,2)}$ remain linearly independent. On the other hand, if we had solved the homogeneous equation by taking the real (or imaginary part) of an exponential; namely, try

$$x_h(t) = \text{Re}\ e^{i\omega t}, \tag{7.8.61}$$

we would find, upon inserting eq. (7.8.61) into eq. (7.8.53), that

$$\omega = \omega_\pm \equiv i\gamma \pm \sqrt{\Omega^2 - \gamma^2}. \tag{7.8.62}$$

This means, when $\Omega = \gamma$, we obtain repeated roots and the otherwise linearly independent solutions

$$x_h^{(\pm)}(t) = \text{Re}\ e^{-\gamma t \pm i\sqrt{\Omega^2 - \gamma^2} t} \tag{7.8.63}$$

become linearly dependent there – both $x_h^{(\pm)}(t) = e^{-\gamma t}$.

Problem 7.41. Compute the Wronskian

$$\text{Wr}[x_h^{(1)}, x_h^{(2)}] = x_h^{(1)} \dot{x}_h^{(2)} - \dot{x}_h^{(1)} x_h^{(2)} \tag{7.8.64}$$

of the homogeneous solutions in equations (7.8.58) and (7.8.59), so as to confirm their linear independence. □

Problem 7.42. Explain why the real or imaginary part of a complex solution to a homogeneous real linear differential equation is also a solution. Now, start from eq. (7.8.61) and verify that eq. (7.8.63) are indeed solutions to eq. (7.8.53) for $\Omega \neq \gamma$. Comment on why the presence of t' in equations (7.8.58) and (7.8.59) amount to arbitrary constants multiplying the homogeneous solutions in eq. (7.8.63). □

---

[64]Note that

$$\frac{\mathrm{d}G(t - t')}{\mathrm{d}t} = \Theta(t - t') \frac{\mathrm{d}}{\mathrm{d}t} \left(e^{-\gamma(t-t')} \frac{\sin\left(\sqrt{\Omega^2 - \gamma^2}(t - t')\right)}{\sqrt{\Omega^2 - \gamma^2}} \right) = \Theta(t - t') \frac{\mathrm{d}\mathcal{G}(t - t')}{\mathrm{d}t}. \tag{7.8.60}$$

Although differentiating $\Theta(t-t')$ gives $\delta(t-t')$, its coefficient is proportional to $\sin(\sqrt{\Omega^2 - \gamma^2}(t - t'))/\sqrt{\Omega^2 - \gamma^2}$, which is zero when $t = t'$, even if $\Omega = \gamma$.

Problem 7.43. Suppose for some initial time t_0, $x_h(t_0) = 0$ and $\dot{x}_h(t_0) = V_0$. There is an external force given by

$$F(t) = \text{Im}\left(e^{-(t/\tau)^2} e^{i\mu t}\right), \qquad \text{for } -2\pi n/\mu \leq t \leq 2\pi n/\mu, \qquad \mu > 0, \qquad . \tag{7.8.65}$$

and $F(t) = 0$ otherwise. (n is an integer greater than 1.) Solve for the motion $x(t > t_0)$ of the damped simple harmonic oscillator, in terms of t_0, V_0, τ, μ and n. $\square$

Problem 7.44. Boundary Value Problem We now turn to using the Green's function in eq. (7.8.37) to solve the boundary value problem, as opposed to the initial value one of eq. (7.8.57). Specifically, we now wish to solve eq. (7.8.53) but subject to the boundary conditions, where $\tau > \tau'$:

$$x_h(\tau') = x_\text{i} \qquad \text{and} \qquad x_h(\tau) = x_\text{f}. \tag{7.8.66}$$

Verify that the solution $x_h(t)$ reads

$$x_h\left(\tau' \leq t \leq \tau\right) = \mathcal{G}(t - \tau') \frac{x_\text{f} + \partial_{\tau'}\mathcal{G}(\tau - \tau')x_\text{i}}{\mathcal{G}(\tau - \tau')} - \partial_{\tau'}\mathcal{G}(t - \tau')x_\text{i}; \tag{7.8.67}$$

where we have employed the homogeneous-solution portion of the retarded Green's function in eq. (7.8.38). Hints: You may find equations (7.8.46), (7.8.47) and (7.8.48) useful. $\square$

Problem 7.45. Feynman, Retarded and Advanced Green's Functions In this problem, we shall explore different choices of integration contours $\{C\}$ occurring within the integral representation of the simple harmonic oscillator (SHO) Green's function. This has applications not just to the SHO itself; but to the study of classical and quantum field theory. For $\Omega > 0$,

$$G_\text{SHO}(T) = -\int_\mathbb{R} \frac{\text{d}\omega}{2\pi} \frac{e^{-i\omega T}}{\omega^2 - \Omega^2}, \tag{7.8.68}$$

$$\ddot{G}_\text{SHO} + \Omega^2 G_\text{SHO} = \delta(T). \tag{7.8.69}$$

The retarded contour G^+_SHO is defined to be the one running along the real line, but skirting *above* the poles at $\omega = \pm\Omega$. Show that this yields

$$G^+_\text{SHO}(T) = -\int_\mathbb{R} \frac{\text{d}\omega}{2\pi} \frac{e^{-i\omega T}}{(\omega + i0^+)^2 - \Omega^2} \tag{7.8.70}$$

$$= \Theta(T)\frac{\sin(\Omega T)}{\Omega}. \tag{7.8.71}$$

The advanced contour G^+_SHO is defined to be the one running along the real line, but skirting *below* the poles at $\omega = \pm\Omega$. Show that this yields

$$G^-_\text{SHO}(T) = -\int_\mathbb{R} \frac{\text{d}\omega}{2\pi} \frac{e^{-i\omega T}}{(\omega - i0^+)^2 - \Omega^2} \tag{7.8.72}$$

$$= \Theta(-T)\frac{\sin(-\Omega T)}{\Omega}. \tag{7.8.73}$$

The Feynman contour G^F_{SHO} is defined to be the one running along the real line, but skirting *below* the pole at $\omega = -\Omega$ but *above* the pole at $\omega = +\Omega$. Show that this yields

$$G^F_{\text{SHO}}(T) = -\int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \frac{e^{-i\omega T}}{\omega^2 - \Omega^2 + i0^+} \tag{7.8.74}$$

$$= \frac{i}{2}\frac{\cos(\Omega T)}{\Omega} + \frac{1}{2}G^+_{\text{SHO}}(T) + \frac{1}{2}G^-_{\text{SHO}}(T). \tag{7.8.75}$$

Notice the retarded and advanced Green's functions can in fact be extracted from the Feynman one. □

7.8.2 Causality and Analytic Properties in Frequency Space; Laplace Transform

Many physical quantities depend on other quantities through a non-local convolution. An example is what we have just witnessed – the trajectory of a dampened harmonic oscillator $x(t)$ driven by an external source $F(t)$:

$$x(t) = \int_{-\infty}^{+\infty} G^+(t - t')F(t'), \tag{7.8.76}$$

where G is the appropriate retarded Green's function. Here, we will simply regard eq. (7.8.76) as the generic relation between some signal x engendered by some source F. The time t may thus be regarded as some 'observation time' and t' is the 'emission time' of the signal. If we expect cause to precede effect, then this causality requirement demands that G^+ be strictly zero when the detection time t is less than the source-emission time t'; namely,

$$G^+(\tau) = \Theta(\tau)\mathcal{G}(\tau), \tag{7.8.77}$$

where $\Theta(\tau > 0) = 1$ and $\Theta(\tau < 0) = 0$.

Let us now recognize that *causality* ($t > t'$ is necessary for a non-zero signal) implies a certain *analyticity* of the frequency transform of G^+. The key reason is, the frequency transform of a retarded function G^+ of elapsed time $\tau \equiv t - t'$ always involves an integral over only the positive real axis:

$$\widetilde{G}^+(\omega) = \int_{\mathbb{R}} \Theta(\tau)\mathcal{G}(\tau)e^{i\omega\tau}\mathrm{d}\tau = \int_0^{\infty} \mathcal{G}(\tau)e^{i\omega\tau}\mathrm{d}\tau. \tag{7.8.78}$$

For physical applications ω is real since it corresponds to the frequency of some system. But we may attempt to analytically continue $\widetilde{G}(\omega)$ off the real line using this integral representation. In particular, if we decompose the frequency into real ω_{R} and imaginary parts ω_{I}; namely, $\omega \equiv \omega_{\text{R}} + i\omega_{\text{I}}$,

$$\widetilde{G}^+(\omega) = \int_0^{\infty} \mathcal{G}(\tau)e^{i\omega_{\text{R}}\tau}e^{-\omega_{\text{I}}\tau}\mathrm{d}\tau. \tag{7.8.79}$$

We see that when $\omega_{\text{I}} < 0$, the integral representation of $\widetilde{G}^+(\omega)$ will likely not converge because $e^{-\omega_{\text{I}}t} = e^{|\omega_{\text{I}}|t} \to \infty$ as $t \to \infty$. On the other hand, if

$$\int_0^{\infty} |\mathcal{G}(\tau)|\mathrm{d}\tau < \infty, \tag{7.8.80}$$

then we are guaranteed that $\widetilde{G}^+(\omega)$ is analytic for $\text{Im}\ \omega \equiv \omega_{\rm I} > 0$ because its derivative converges:

$$|\partial_\omega \widetilde{G}^+(\omega)| = \left| i \int_0^\infty \tau \cdot \mathcal{G}(\tau) e^{i\omega\tau} \mathrm{d}\tau \right| \tag{7.8.81}$$

$$\leq \int_0^\infty |\mathcal{G}(\tau)| \cdot \tau e^{-\omega_{\rm I}\tau} \mathrm{d}\tau. \tag{7.8.82}$$

> Any retarded object $G^+(\tau)$ of the form in eq. (7.8.77) that satisfies eq. (7.8.80) has a frequency transform $\widetilde{G}^+(\omega)$ that is analytic in the positive imaginary portion of the complex $\omega-$plane – i.e., causality in real time implies analyticity in the positive imaginary frequency space.

If we are given the frequency transform $\widetilde{G}(\omega)$ of some Green's function $G(\tau)$, we may reconstruct it via the Fourier decomposition

$$G(\tau) = \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega\tau} \widetilde{G}(\omega). \tag{7.8.83}$$

As we have argued above, if $G(\tau)$ is retarded – i.e., it vanishes for $\tau < 0$ – then its frequency space counterpart ought to be analytic for $\text{Im}\ \omega \equiv \omega_{\rm I} > 0$ and, moreover, should be damped to 0 as $\omega_{\rm I} \to +\infty$. The latter means, whenever $\tau < 0$, we should then be able to trivially 'close-the-contour' for the above $\omega-$integral by joining $\pm\infty$ on the real line with an infinitely large semi-circle in the positive Im half of the $\omega-$plane; namely, $e^{\omega_{\rm I} t} \to 0$ for $\omega_{\rm I} \to +\infty$. Since $\widetilde{G}(\omega)$ is analytic for $\omega_{\rm I} > 0$, by Cauchy's theorem we must then have

$$G(\tau < 0) = \oint_{\substack{\text{semi-circle} \\ \omega_{\rm I} > 0}} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega_{\rm R}\tau} e^{-\omega_{\rm I}|\tau|} \widetilde{G}(\omega) = 0. \tag{7.8.84}$$

We may check our understanding against the damped SHO Green's function in equations (7.8.34) and (7.8.37), with $\tau \equiv t - t'$. We see that $\mathcal{G}(\tau) = \exp(-\gamma\tau)\sin(\sqrt{\Omega^2 - \gamma^2}\tau)/\sqrt{\Omega^2 - \gamma^2}$ obeys eq. (7.8.80) for $\Omega, \gamma > 0$. Hence, $\widetilde{G}(\omega_{\rm I} > 0)$ must be analytic – indeed,

$$\widetilde{G}(\omega) = \frac{1}{-\omega^2 - 2i\omega\gamma + \Omega^2} = -\frac{1}{(\omega - \omega_+)(\omega - \omega_-)}, \tag{7.8.85}$$

where

$$\omega_\pm = -i\gamma \pm \sqrt{\Omega^2 - \gamma^2}. \tag{7.8.86}$$

Both poles $\omega_\pm$ lie in the negative imaginary ω plane even if $\Omega < \gamma$.

YZ: Is a complex function $\widetilde{G}^+(\omega)$ completely determined by poles and branch cuts? Examples. Quasinormal modes?

7.8.3 Multi-dimensional Gaussian Probability Distributions

Moments of Multi-Dimensional Gaussians Many physical and statistical problems involve the '$n-$point' function

$$C^{i_1 i_2 \dots i_n}[O] \equiv \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} x^{i_1} x^{i_2} \dots x^{i_n} P[\vec{x}, O], \tag{7.8.87}$$

$$P[\vec{x},O] \equiv \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x}\right] \tag{7.8.88}$$

or its generalization. Here, O is a real symmetric $D \times D$ matrix with strictly positive eigenvalues. (If O were real but not symmetric, its anti-symmetric part would drop out in the $\vec{x}^{\mathrm{T}} O \vec{x}$.) Let us now proceed to show that

$$C[O] \equiv \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} P[\vec{x},O] = 1; \tag{7.8.89}$$

the odd $n = 2\ell + 1$ case is zero,

$$C^{i_1 \dots i_{2\ell+1}}[O] = 0; \tag{7.8.90}$$

and the even $n = 2\ell$ case involves components of O^{-1}, the matrix inverse of O; namely,

$$C^{i_1 \dots i_{2\ell}}[O] = \frac{1}{2^\ell \ell!} (O^{-1})^{\{i_1 i_2} (O^{-1})^{i_3 i_4} \dots (O^{-1})^{i_{2\ell-1} i_{2\ell}\}}. \tag{7.8.91}$$

If we define a *contraction* between the indices i and j to be $(O^{-1})^{ij} = (O^{-1})^{ji}$; then we may phrase the even $n = 2\ell$ result as the sum over all full contractions between the n indices – the 2^ℓ will drop out because O^{-1} is necessarily symmetric and $\ell!$ will also vanish because of symmetry under the interchange of any two O^{-1}s –

$$C^{i_1 \dots i_{2\ell}}[O] = \sum \left(\text{Full contractions of } \{i_1, \dots, i_{2\ell}\}\right). \tag{7.8.92}$$

For example, the $n = 2$ case is simply

$$C^{ij}[O] = \frac{1}{2^1} (O^{-1})^{\{ij\}} = (O^{-1})^{ij}; \tag{7.8.93}$$

the $n = 4$ case yields

$$C^{abij}[O] = \frac{1}{2^2 2!} (O^{-1})^{\{ab} (O^{-1})^{ij\}} \tag{7.8.94}$$

$$= (O^{-1})^{ab} (O^{-1})^{ij} + (O^{-1})^{ai} (O^{-1})^{bj} + (O^{-1})^{aj} (O^{-1})^{bi}; \tag{7.8.95}$$

and so on.

Let us begin with $C[O]$. Since O is real and symmetric, it must be diagonalizable through an orthogonal transformation U; namely,

$$\vec{x}^{\mathrm{T}} O \vec{x} = (U\vec{x})^{\mathrm{T}} \Lambda (U\vec{x}), \tag{7.8.96}$$

where Λ is the diagonal matrix containing the eigenvalues $\{\lambda_i\}$ of O. That is, $O = U^{\mathrm{T}} \Lambda U$, where $\Lambda = \mathrm{diag}[\lambda_1, \dots, \lambda_D]$. Since $|\det U| = 1$, a change-of-variables $\vec{x}' = U\vec{x}$ now informs us

$$C[O] = \sqrt{\frac{\det O}{(2\pi)^D}} \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \exp\left[-\frac{1}{2}\vec{x}'^{\mathrm{T}} \Lambda \vec{x}'\right] \tag{7.8.97}$$

$$= \sqrt{\frac{\det O}{(2\pi)^D}} \prod_{i=1}^{D} \int_{\mathbb{R}} \mathrm{d}x'^i \exp\left[-\frac{1}{2}\lambda_i (x'^i)^2\right]. \tag{7.8.98}$$

We see why the eigenvalues of O need to be positive – for, otherwise, the corresponding Gaussian integral would not converge. Employing eq. (7.1.5) now hands us

$$C[O] = \sqrt{\frac{\det O}{(2\pi)^D}} \prod_{i=1}^{D} \sqrt{\frac{2\pi}{\lambda_i}}. \tag{7.8.99}$$

Further recognizing that the determinant of a matrix is simply the product of its eigenvalues, we conclude $C[O] = 1$.

The odd $n = 2\ell + 1$ case follows from the fact that the integrand is odd under parity. Upon a change-of-variables $\vec{x} \equiv -\vec{y}$,

$$\begin{aligned} C^{i_1 \dots i_{2\ell+1}}[O] &\propto \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} x^{i_1} x^{i_2} \dots x^{i_{2\ell+1}} \exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x}\right] \\ &= \int_{\mathbb{R}^D} \mathrm{d}^D\vec{y} (-)^{2\ell+1} y^{i_1} y^{i_2} \dots y^{i_{2\ell+1}} \exp\left[-\frac{1}{2}\vec{y}^{\mathrm{T}} O \vec{y}\right]; \end{aligned} \tag{7.8.100}$$

which in turn implies

$$C^{i_1 \dots i_{2\ell+1}}[O] = -C^{i_1 \dots i_{2\ell+1}}[O] = 0. \tag{7.8.101}$$

The even $n = 2\ell$ case involves a trick that appears even in quantum field theory, where it is usually attributed to Julian Schwinger. We first recognize that the $n-$point function itself may be written as a $n-$derivative with respect to an auxiliary 'source' $\vec{J}$:

$$C^{i_1 \dots i_n}[O] = \frac{\partial}{\partial J^{i_1}} \dots \frac{\partial}{\partial J^{i_n}} \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x} + \vec{J} \cdot \vec{x}\right] \Bigg|_{\vec{J}=\vec{0}}. \tag{7.8.102}$$

Equivalently, we may regard the Fourier transform

$$\widetilde{P}[\vec{k}, O] \equiv \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x} + i\vec{k} \cdot \vec{x}\right] \tag{7.8.103}$$

[65]as a generating function of the $n-$point functions. By Taylor expanding with respect to $\vec{k}$,

$$\widetilde{P}[\vec{k}, O] = \sum_{n=0}^{\infty} \frac{i^n}{n!} k^{i_1} \dots k^{i_n} \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x}\right] x^{i_1} \dots x^{i_n} \tag{7.8.104}$$

$$= \sum_{n=0}^{\infty} \frac{i^n}{n!} k^{i_1} \dots k^{i_n} \frac{1}{i^n} \frac{\partial^n \widetilde{P}}{\partial k^{i_1} \dots k^{i_n}} \Bigg|_{\vec{k}=\vec{0}} \tag{7.8.105}$$

[65]For the purposes of this problem – I prefer to do $(1/i)\partial_{k^j}$ instead of $-(1/i)\partial_{k^j}$ in order to draw the parallel to Schwinger's source trick in QFT – I have employed an opposite sign convention; namely $e^{i\vec{k}\cdot\vec{x}}$ instead of $e^{-i\vec{k}\cdot\vec{x}}$.

$$= \sum_{n=0}^{\infty} \frac{i^n}{n!} k^{i_1} \dots k^{i_n} C^{i_1 \dots i_n}, \tag{7.8.106}$$

with the repeated indices $\{i_1, \dots, i_n\}$ summed over; we see that the coefficient of the nth order term is, in fact, the n point function itself.

Problem 7.46. By re-expressing

$$\vec{x} \equiv \vec{y} + O^{-1} i\vec{k}, \tag{7.8.107}$$

demonstrate that this leads to the 'completing-the-square'

$$-\frac{1}{2}\vec{x}^{\mathrm{T}} O \vec{x} + i\vec{k} \cdot \vec{x} = -\frac{1}{2}\vec{y}^{\mathrm{T}} O \vec{y} - \frac{1}{2}\vec{k}^{\mathrm{T}} O^{-1} \vec{k} \tag{7.8.108}$$

and $\mathrm{d}^D \vec{x} = \mathrm{d}^D \vec{y}$. Hint: Remember that $(O^{-1})^{\mathrm{T}} = O^{-1}$. □

Equations (7.8.107) and (7.8.108) tell us, our generating function is now transformed into

$$\widetilde{P}[\vec{k}, O] = \exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}} O^{-1} \vec{k}\right] \int_{\mathbb{R}^D} \mathrm{d}^D \vec{y} \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\vec{y}^{\mathrm{T}} O \vec{y}\right] \tag{7.8.109}$$

$$= \exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}} O^{-1} \vec{k}\right] C[0] = \exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}} O^{-1} \vec{k}\right]. \tag{7.8.110}$$

The Taylor expansion of $\widetilde{P}$ now reads

$$\widetilde{P}[\vec{k}, O] = \sum_{\ell=0}^{\infty} \frac{(i^2)^\ell}{2^\ell \ell!} k^{a_1} k^{b_1} \dots k^{a_\ell} k^{b_\ell} \cdot (O^{-1})^{a_1 b_1} \dots (O^{-1})^{a_\ell b_\ell}. \tag{7.8.111}$$

That only even powers of $\vec{k}$ appear immediately informs us that all odd point functions vanish; a fact we deduced by parity arguments earlier. Taking $n = 2\ell$ (even) derivatives with respect to $\vec{k}$, followed by setting $\vec{k} = \vec{0}$, yields

$$C^{i_1 \dots i_n} = \frac{1}{i^n} \frac{\partial^n \widetilde{P}[\vec{k}, O]}{\partial k^{i_1} \dots k^{i_n}} \bigg|_{\vec{k}=\vec{0}} \tag{7.8.112}$$

$$= \frac{1}{2^\ell \ell!} \frac{\partial k^{\{a_1}}{\partial k^{i_1}} \frac{\partial k^{b_1}}{\partial k^{i_2}} \dots \frac{\partial k^{a_\ell}}{\partial k^{i_{2\ell-1}}} \frac{\partial k^{b_\ell\}}}{\partial k^{i_{2\ell}}} (O^{-1})^{a_1 b_1} \dots (O^{-1})^{a_\ell b_\ell} \tag{7.8.113}$$

$$= \frac{1}{2^\ell \ell!} \delta^{\{a_1}{}_{i_1} \delta^{b_1}{}_{i_2} \dots \delta^{a_\ell}{}_{i_{2\ell-1}} \delta^{b_\ell\}}{}_{i_{2\ell}} (O^{-1})^{a_1 b_1} \dots (O^{-1})^{a_\ell b_\ell} \tag{7.8.114}$$

$$= \frac{1}{2^\ell \ell!} (O^{-1})^{\{i_1 i_2} \dots (O^{-1})^{i_{2\ell-1} i_{2\ell}\}}. \tag{7.8.115}$$

For each fixed permutation, out of the $(2\ell)!$ ones, there is a 'double counting' for each $(O^{-1})^{mn} = (O^{-1})^{nm}$; hence, since there are ℓ O^{-1}s the 2^ℓ will cancel out. Likewise, there is a $\ell!$ 'over-counting' from the symmetry under the mutual interchange of the O^{-1}s; for e.g., $O^{ij} O^{ab} = O^{ab} O^{ij}$. As an illustration of these general arguments, look at the $n = 4$ case:

$$C^{abij}[O] = \frac{1}{2^2 2!} \Big((O^{-1})^{ab} (O^{-1})^{ij} + (O^{-1})^{ab} (O^{-1})^{ji} + (O^{-1})^{ai} (O^{-1})^{bj}$$

$$
\begin{aligned}
&\qquad + (O^{-1})^{ai}(O^{-1})^{jb} + (O^{-1})^{aj}(O^{-1})^{bi} + (O^{-1})^{aj}(O^{-1})^{ib} \\
&+ (O^{-1})^{ba}(O^{-1})^{ij} + (O^{-1})^{ba}(O^{-1})^{ji} + (O^{-1})^{bi}(O^{-1})^{aj} \\
&\qquad + (O^{-1})^{bi}(O^{-1})^{ja} + (O^{-1})^{bj}(O^{-1})^{ai} + (O^{-1})^{bj}(O^{-1})^{ia} \\
&+ (O^{-1})^{ia}(O^{-1})^{bj} + (O^{-1})^{ia}(O^{-1})^{jb} + (O^{-1})^{ib}(O^{-1})^{aj} \\
&\qquad + (O^{-1})^{ib}(O^{-1})^{ja} + (O^{-1})^{ij}(O^{-1})^{ab} + (O^{-1})^{ij}(O^{-1})^{ba} \\
&+ (O^{-1})^{ja}(O^{-1})^{bi} + (O^{-1})^{ja}(O^{-1})^{ib} + (O^{-1})^{jb}(O^{-1})^{ai} \\
&\qquad + (O^{-1})^{jb}(O^{-1})^{ia} + (O^{-1})^{ji}(O^{-1})^{ab} + (O^{-1})^{ji}(O^{-1})^{ba}\Big). \qquad (7.8.116)
\end{aligned}
$$

To sum, these considerations have led us to the formula in eq. (7.8.91).

Problem 7.47. Non-Zero One Point Function Let us now consider the following probability density. For constant $\vec{x}_0$ and real orthogonal $D \times D$ matrix with positive eigenvalue,

$$
P[\vec{x}, O, \vec{x}_0] \equiv \sqrt{\frac{\det O}{(2\pi)^D}} \exp\left[-\frac{1}{2}\left(\vec{x} - \vec{x}_0\right)^{\mathrm{T}} O \left(\vec{x} - \vec{x}_0\right)\right]. \qquad (7.8.117)
$$

- Show that P here is normalized properly; i.e.,

$$
\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} P[\vec{x}, O, \vec{x}_0] = 1. \qquad (7.8.118)
$$

- Now, let us define

$$
C^{i_1 i_2 \dots i_n}[O, \vec{x}_0] \equiv \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}(x - x_0)^{i_1}(x - x_0)^{i_2} \dots (x - x_0)^{i_n} P[\vec{x}, O, \vec{x}_0]. \qquad (7.8.119)
$$

  Show that the odd $n = 2\ell + 1$ case is zero,

$$
C^{i_1 \dots i_{2\ell+1}}[O, \vec{x}_0] = 0. \qquad (7.8.120)
$$

  Explain why this means $\vec{x}_0$ is the one point function – namely,

$$
\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} P[\vec{x}, O, \vec{x}_0] x^i = x_0^i. \qquad (7.8.121)
$$

  Next, show that the even $n = 2\ell$ case involves components of O^{-1}, the matrix inverse of O; namely,

$$
\begin{aligned}
C^{i_1 \dots i_{2\ell}}[O, \vec{x}_0] &= \frac{1}{2^\ell \ell!}(O^{-1})^{\{i_1 i_2}(O^{-1})^{i_3 i_4} \dots (O^{-1})^{i_{2\ell-1} i_{2\ell}\}}, \qquad (7.8.122) \\
&= \sum \left(\text{Full contractions of } \{i_1, \dots, i_{2\ell}\}\right). \qquad (7.8.123)
\end{aligned}
$$

  Hint: Identify the right shift-of-variables.

- What we have just discussed is the most general multi-dimensional Gaussian probability distribution. To see this, show that the most general quadratic-in-$\vec{x}$ exponent may always be massaged into the form in eq. (7.8.117). Specifically, show that P there is equivalent to

$$\exp\left[-\frac{1}{2}\vec{A}^{\mathrm{T}}O^{-1}\vec{A}\right]\sqrt{\frac{\det O}{(2\pi)^{D}}}\exp\left[-\frac{1}{2}\vec{x}^{\mathrm{T}}O\vec{x}+\vec{A}\cdot\vec{x}\right], \tag{7.8.124}$$

for some real vector $\vec{A}$, if and only if

$$\vec{A}=O\vec{x}_{0}. \tag{7.8.125}$$

- Next, explain why

$$\widetilde{P}[\vec{k},O,\vec{x}_{0}]\equiv\int_{\mathbb{R}^{D}}\mathrm{d}^{D}\vec{x}\sqrt{\frac{\det O}{(2\pi)^{D}}}\exp\left[-\frac{1}{2}\left(\vec{x}-\vec{x}_{0}\right)^{\mathrm{T}}O\left(\vec{x}-\vec{x}_{0}\right)+i\vec{k}\cdot\left(\vec{x}-\vec{x}_{0}\right)\right] \tag{7.8.126}$$

$$=\exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}}O^{-1}\vec{k}\right] \tag{7.8.127}$$

is the generating function for the $n-$point functions $\{C^{i_{1}\dots i_{n}}[O,\vec{x}_{0}]\}$.

- Also explain why

$$\widetilde{P}_{\chi}[\vec{k},O,\vec{x}_{0}]\equiv\int_{\mathbb{R}^{D}}\mathrm{d}^{D}\vec{x}\sqrt{\frac{\det O}{(2\pi)^{D}}}\exp\left[-\frac{1}{2}\left(\vec{x}-\vec{x}_{0}\right)^{\mathrm{T}}O\left(\vec{x}-\vec{x}_{0}\right)+i\vec{k}\cdot\vec{x}\right]$$

$$=\exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}}O^{-1}\vec{k}+i\vec{k}\cdot\vec{x}_{0}\right] \tag{7.8.128}$$

is the generating function for the alternate $n-$point function

$$\chi^{i_{1}i_{2}\dots i_{n}}[O,\vec{x}_{0}]\equiv\int_{\mathbb{R}^{D}}\mathrm{d}^{D}\vec{x}x^{i_{1}}x^{i_{2}}\dots x^{i_{n}}P[\vec{x},O,\vec{x}_{0}]. \tag{7.8.129}$$

- Define the shifted quantity

$$i\vec{k}_{s}\equiv i\vec{k}+O\vec{x}_{0}. \tag{7.8.130}$$

Show that eq. (7.8.128) may be packaged as

$$\widetilde{P}_{\chi}[\vec{k},O,\vec{x}_{0}]=\exp\left[-\frac{1}{2}\vec{k}_{s}^{\mathrm{T}}O^{-1}\vec{k}_{s}-\frac{1}{2}\vec{x}_{0}^{\mathrm{T}}O\vec{x}_{0}\right]. \tag{7.8.131}$$

Finally, explain why

$$\left(\frac{1}{i}\frac{\partial}{\partial k^{i_{1}}}-x_{0}^{i_{1}}\right)\left(\frac{1}{i}\frac{\partial}{\partial k^{i_{2}}}-x_{0}^{i_{2}}\right)\dots\left(\frac{1}{i}\frac{\partial}{\partial k^{i_{n}}}-x_{0}^{i_{n}}\right)\exp\left[-\frac{1}{2}\vec{k}_{s}^{\mathrm{T}}O^{-1}\vec{k}_{s}-\frac{1}{2}\vec{x}_{0}^{\mathrm{T}}O\vec{x}_{0}\right]\Bigg|_{\vec{k}=\vec{0}}$$

$$=\frac{1}{i^{n}}\frac{\partial}{\partial k^{i_{1}}}\frac{\partial}{\partial k^{i_{2}}}\dots\frac{\partial}{\partial k^{i_{n}}}\exp\left[-\frac{1}{2}\vec{k}^{\mathrm{T}}O^{-1}\vec{k}\right]\Bigg|_{\vec{k}=\vec{0}}; \tag{7.8.132}$$

and how this result is related to eq. (7.8.122). Note that the derivatives on both sides are with respect to the same $\vec{k}$.

□

Problem 7.48. Hermitian O Case Let O be a Hermitian matrix with strictly positive eigenvalues. Consider the following $(n,\bar{n})-$point function, for non-negative integers n and $\bar{n}$:

$$C^{i_1\dots i_n;j_1\dots j_{\bar{n}}}[O] \equiv \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}_{\mathrm{R}} \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}_{\mathrm{I}} x^{i_1}\dots x^{i_n}\overline{x^{j_1}\dots x^{j_{\bar{n}}}} \exp\left[-\vec{x}^\dagger O\vec{x}\right] \frac{\det O}{\pi^D}, \tag{7.8.133}$$

where the overbar denotes complex conjugation; and $\vec{x}$ is a complex $D-$component column.

Show that $C^{i_1\dots i_n;j_1\dots j_{\bar{n}}}[O]$ is non-zero only when $n=\bar{n}$. And when $n=\bar{n}$, defining each contraction between ij to be $(O^{-1})^{ij}$, show that

$$C^{i_1\dots i_n;j_1\dots j_n}[O] = \sum\left(\text{Full contractions between the } (i,j) \text{ indices}\right); \tag{7.8.134}$$

where the contraction only involves i_a and j_b indices; i.e., none between pairs of is nor between pairs of js. For example, the $n=1$ case is

$$C^{i;j}[O] = (O^{-1})^{ij}; \tag{7.8.135}$$

the $n=2$ case is

$$C^{ab;ij}[O] = (O^{-1})^{ai}(O^{-1})^{bj} + (O^{-1})^{aj}(O^{-1})^{bi}; \tag{7.8.136}$$

the $n=3$ case is

$$\begin{aligned} C^{abc;ij\ell}[O] = &(O^{-1})^{ai}(O^{-1})^{bj}(O^{-1})^{c\ell} + (O^{-1})^{ai}(O^{-1})^{b\ell}(O^{-1})^{cj} \\ &+ (O^{-1})^{aj}(O^{-1})^{bi}(O^{-1})^{c\ell} + (O^{-1})^{aj}(O^{-1})^{b\ell}(O^{-1})^{ci} \\ &+ (O^{-1})^{a\ell}(O^{-1})^{bi}(O^{-1})^{cj} + (O^{-1})^{a\ell}(O^{-1})^{bj}(O^{-1})^{ci}; \end{aligned} \tag{7.8.137}$$

and so on. □

7.8.4 Statistics: Convolution and the Central Limit Theorem

Moments of a Probability Distribution If $P(x)$ denotes the probability distribution of some real variable $x \in [a,b]$, that means $P(x)\mathrm{d}x$ gives the probability that the outcome of a particular measurement be found between x and $x+\mathrm{d}x$. The total probability must be 1:

$$\int_{\mathbb{R}} P(x)\mathrm{d}x = 1. \tag{7.8.138}$$

Without loss of generality, we will always integrate over the real line. Any probability distribution $P(x)$ that is only defined within some interval, say $x \in [a,b]$, can always be extended to the entire real line by defining $P(x<a) = 0 = P(x>b)$.

A set of commonly used statistic are the moments, or $(n \geq 1)-$point functions (for $n = 1,2,3,\dots$):

$$\langle x^n \rangle \equiv \int_{\mathbb{R}} x^n P(x)\mathrm{d}x. \tag{7.8.139}$$

For example, the *mean* or *average* x_0 is the ($n = 1$) one point function,

$$x_0 \equiv \langle x \rangle \equiv \int_{\mathbb{R}} xP(x)\mathrm{d}x \qquad \text{(mean)}. \tag{7.8.140}$$

The variance σ^2 is the average of the square of the deviation from the mean,

$$\sigma^2 \equiv \langle (x - x_0)^2 \rangle = \int_{\mathbb{R}} (x - x_0)^2 P(x)\mathrm{d}x \tag{7.8.141}$$

$$= \langle x^2 \rangle - \langle x \rangle^2 \qquad \text{(variance)}; \tag{7.8.142}$$

which we see is, in turn, equal to the difference between the $n = 2$ point function and the square of the $n = 1$ function. Moreover, the *standard deviation* σ is the (positive) square root of the variance; it measures the 'spread' or 'width' of a given distribution – the Gaussian being a prime example.

The Fourier transform $\widetilde{P}$ of the probability distribution is the *generating function* of these $n-$point moments. To see this, we simply perform a Taylor expansion of the exponential,

$$\widetilde{P}(k) = \int_{\mathbb{R}} e^{ikx} P(x)\mathrm{d}x = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \int_{\mathbb{R}} x^n P(x)\mathrm{d}x \tag{7.8.143}$$

$$= 1 + \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle. \tag{7.8.144}$$

(That the first term is unity, is simply the statement that total probability is necessarily one.) Equivalently, we may note that

$$\langle x^n \rangle = \frac{1}{i^n} \frac{\partial \widetilde{P}(k=0)}{\partial k^n}. \tag{7.8.145}$$

In other words, if we are given all the moments $\{\langle x^{n \geq 1} \rangle\}$ of a probability distribution, we may – at least in principle – reconstruct its Fourier transform because the former constitutes its Taylor series coefficients. And since the Fourier transform is invertible, namely

$$P(x) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} \widetilde{P}(k) e^{-ikx}, \tag{7.8.146}$$

that implies the set $\{\langle x^{n \geq 1} \rangle\}$, whenever they exist,[66] fully characterize the $P(x)$ itself.

Problem 7.49. Probability Conservation Explain why probability conservation in eq. (7.8.138) implies $\widetilde{P}(k = 0) = 1$. □

Cumulants Instead of using the $n-$moments, an alternate manner to characterize a probability distribution $P(x)$ is through its *cumulants* $\{\kappa_n\}$, defined by expressing its Fourier transform as an exponential,

$$\widetilde{P}(k) \equiv \exp\left(\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n \right); \tag{7.8.147}$$

[66] An example of a probability distribution with infinite (even $n \geq 2$)$-$moments is the Cauchy-Lorentz(-Breit-Wigner) $P(x) = (a/\pi)(x^2 + a^2)^{-1}$, for some fixed positive a.

or, equivalently,

$$\ln \widetilde{P}(k) \equiv \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n. \tag{7.8.148}$$

The sum on the right hand side of eq. (7.8.148) begins at $n = 1$ so as to ensure total probability is unity; i.e., $\widetilde{P}(k = 0) = 1$ implies $\kappa_0 = 0$. Hence, we may define these cumulants as the $n-$th derivative of the natural logarithm of $\widetilde{P}(k)$ with respect to (ik), evaluated at $k = 0$.

$$\kappa_{n \geq 1} = \frac{1}{i^n} \frac{\partial^n \ln \widetilde{P}(k = 0)}{\partial k^n} \tag{7.8.149}$$

That $\ln \widetilde{P}(k)$ is a power series can, in fact, be shown by verifying eq. (7.8.149) exists for all n. For a given n, the derivatives on the right hand side of eq.(7.8.148) lead to sum of products of the form $(\prod_a (i^{-n_a} \partial_k^{n_a} \widetilde{P})^{m_a})/\widetilde{P}^l \Big|_{k=0} = (\prod_a (\langle x^{n_a} \rangle)^{m_a})$, for sets of integers $\{n_1, \dots, m_1, \dots, l\}$.

We may solve the cumulants in terms of the moments, and vice versa. This shows that, whenever both sets exist, they fully describe $P(x)$ itself. Firstly, we may Taylor expand the natural log of eq. (7.8.144) and compare the coefficient of its $(ik)^n/n!$ term with that of eq. (7.8.148). Recalling that $\ln(1 + z) = -\sum_{n=1} (-z)^n/n$, we may compute the first few terms explicitly as

$$\kappa_1 = \langle x \rangle \equiv x_0, \tag{7.8.150}$$
$$\kappa_2 = \langle (x - x_0)^2 \rangle \equiv \sigma^2, \tag{7.8.151}$$
$$\kappa_3 = \langle (x - x_0)^3 \rangle, \tag{7.8.152}$$
$$\kappa_4 = \langle (x - x_0)^4 \rangle - 3\langle (x - x_0)^2 \rangle^2; \tag{7.8.153}$$

etc. Similarly, we may Taylor expand eq. (7.8.147) and compare the coefficients of its $(ik)^n/n!$ term with that of eq. (7.8.144). This procedure yields

$$\langle x^2 \rangle = \kappa_1^2 + \kappa_2, \tag{7.8.154}$$
$$\langle x^3 \rangle = \kappa_1^3 + 3\kappa_1\kappa_2 + \kappa_3, \tag{7.8.155}$$
$$\langle x^4 \rangle = \kappa_1^4 + 6\kappa_2\kappa_1^2 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + \kappa_4; \tag{7.8.156}$$

etc. Of course, we could also have began from eq. (7.8.150) to solve for $\langle x \rangle$ in terms of κ_1; followed by eq. (7.8.151) to solve for $\langle x^2 \rangle$ in terms of κ_1 and κ_2; and so on.

Problem 7.50. Verify the results in equations (7.8.150)-(7.8.153) and equations (7.8.154)-(7.8.156). Hint: When computing the N-th order term $(ik)^N/N!$ from the natural log of eq. (7.8.144) or from eq. (7.8.147), the summations involved may be truncated at $n = N$. □

Cumulants for Gaussian As an example, let us write down the Fourier transform of the 1D Gaussian distribution

$$P(x) = \frac{e^{-\frac{1}{2}\left(\frac{x - x_0}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}}. \tag{7.8.157}$$

According to eq. (7.8.128),

$$\widetilde{P}(k) = \exp\left[ikx_0 + \frac{(ik)^2}{2!}O^{-1}\right]. \tag{7.8.158}$$

In the 1D case, $O = 1/\sigma^2$, and we may therefore read off

$$\kappa_1 = x_0, \qquad \kappa_2 = \sigma^2, \qquad \text{and} \qquad \kappa_{n\geq 3} = 0. \tag{7.8.159}$$

On the other hand, if the only non-zero cumulants of some distribution $P(x)$ are the first two $\kappa_{1,2}$, i.e., if its Fourier transform takes the form in eq. (7.8.158), then by simply taking the inverse transform we shall see that $P(x)$ must be Gaussian. In fact, it is $\kappa_2 \neq 0$ that is crucial here, since $\kappa_1 = x_0 = 0$ merely implies the mean is zero. Hence:

> A probability distribution $P(x)$ is Gaussian if and only if its second cumulant is strictly non-zero ($\kappa_2 \neq 0$) while all higher ones are zero ($\kappa_{n\geq 3} = 0$).

Translation and Re-Scaling If we now consider the $n-$point moments, not of x itself but of its translated cousin $x - a$ for constant real a,

$$\langle (x-a)^n \rangle = \int_{\mathbb{R}} (x-a)^n P(x) \mathrm{d}x; \tag{7.8.160}$$

the associated generating function would now read

$$\int_{\mathbb{R}} e^{ik(x-a)} P(x) \mathrm{d}x = e^{-ika}\widetilde{P}(k) = \exp\left(ik(\kappa_1 - a) + \sum_{n=2}^{\infty} \frac{(ik)^n}{n!}\kappa_n\right), \tag{7.8.161}$$

where eq. (7.8.147) was used in the right hand side. This informs us, under such a shift $x \to x-a$, the corresponding cumulants are all invariant $\kappa_n \to \kappa_n$ except the $n=1$ case, where $\kappa_1 \to \kappa_1 - a$.

Similarly, we may now consider computing the $n-$point moments of not x itself but of $\lambda \cdot x$, for some real λ:

$$\langle (\lambda \cdot x)^n \rangle = \lambda^n \int_{\mathbb{R}} x^n P(x) \mathrm{d}x. \tag{7.8.162}$$

The associated generating function would then read

$$\int_{\mathbb{R}} e^{ik(\lambda \cdot x)} P(x) \mathrm{d}x = \widetilde{P}\left(\lambda \cdot k\right) = \exp\left(\sum_{n=1}^{\infty} \frac{(ik)^n}{n!}\lambda^n \kappa_n\right). \tag{7.8.163}$$

Under a re-scalinng $x \to \lambda \cdot x$, therefore, the cumulants become re-scaled as $\kappa_n \to \lambda^n \kappa_n$.

Problem 7.51. Normalized Cumulants: Gaussian Case Show that

$$\exp\left(\frac{(ik)^2}{2!}\right) = 1 + \sum_{n=1}^{\infty} \frac{(ik)^{2n}}{(2n)!}(2n-1)!!, \tag{7.8.164}$$

where $0!! \equiv 1$ and for n a positive integer, $(n \geq 1)!! = 1 \cdot 3 \cdot 5 \cdots (n-2) \cdot n$. Use this result and the $\langle ((x-x_0)/\sigma)^n \rangle-$generating function

$$\int_{\mathbb{R}} \mathrm{d}x e^{ik\left(\frac{x-x_0}{\sigma}\right)} P(x), \tag{7.8.165}$$

where $P(x)$ is the Gaussian in eq. (7.8.157), to explain why eq. (7.7.16) holds. $\square$

Problem 7.52. Normalized Probability Distributions Let us now consider computing the $n-$moments of $(x-x_0)/\sigma$, where x_0 is the mean and σ the standard deviation. This amounts to 'centering' the distribution at the mean; and using the standard deviation σ as the benchmark scale of the statistical problem at hand.

Define the $n-$moments

$$\left\langle \left(\frac{x-x_0}{\sigma}\right)^n \right\rangle = \int_{\mathbb{R}} \left(\frac{x-x_0}{\sigma}\right)^n P(x)\mathrm{d}x; \tag{7.8.166}$$

as well as the cumulants

$$\exp\left(\sum_{n=1}^{\infty} \frac{(ik)^n}{n!}\kappa'_n\right) \equiv \int_{\mathbb{R}} \exp\left(ik\frac{x-x_0}{\sigma}\right) P(x)\mathrm{d}x. \tag{7.8.167}$$

Show that

$$\kappa'_1 = \left\langle \frac{x-x_0}{\sigma} \right\rangle = 0, \tag{7.8.168}$$

$$\kappa'_2 = \left\langle \left(\frac{x-x_0}{\sigma}\right)^2 \right\rangle = 1, \tag{7.8.169}$$

$$\kappa'_3 = \left\langle \left(\frac{x-x_0}{\sigma}\right)^3 \right\rangle, \tag{7.8.170}$$

$$\kappa'_4 = \left\langle \left(\frac{x-x_0}{\sigma}\right)^4 \right\rangle - 3\left\langle \left(\frac{x-x_0}{\sigma}\right)^2 \right\rangle^2; \tag{7.8.171}$$

and

$$\left\langle \left(\frac{x-x_0}{\sigma}\right)^4 \right\rangle = 3\kappa_2'^2 + \kappa'_4. \tag{7.8.172}$$

Argue that, more generally, that

$$\kappa'_{n\geq 2} = \kappa_n/\sigma^n. \tag{7.8.173}$$

Hint: Going from x to $(x-x_0)/\sigma$ involves a re-scaling followed by a translation. □

We have already seen earlier that knowing all the cumulants $\{\kappa_n\}$ is equivalent to knowing all the moments $\{\langle x^n \rangle\}$; which in turn is equivalent to – through the Fourier transform – knowing the probability distribution $P(x)$ itself. Let us now observe that knowing the 'normalized cumulants' $\{\kappa'_n\}$ defined in Problem (7.52) plus the mean x_0 and variance σ^2 is also equivalent to knowing the moments, since these $\{\kappa'_{n\geq 2}\}$ are merely re-scaled versions of the $\{\kappa_n\}$.

Problem 7.53. Almost Gaussian Distribution Consider the following probability distribution

$$P(x) \equiv \frac{\mathcal{N}}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2 \left(1+\epsilon\left(\frac{x-x_0}{\sigma_0}\right)^2\right)\right), \tag{7.8.174}$$

$$x, x_0 \in \mathbb{R}, \sigma_0 > 0. \tag{7.8.175}$$

We dub it 'almost Gaussian' because we shall assume $0 < \epsilon \ll 1$; though, note that the tail ends of the distribution, where $|x - x_0|/\sigma_0 \gg 1$, differ dramatically from those of the Gaussian. Here, $\mathcal{N}$ is the constant that renders $\int P(x)\mathrm{d}x = 1$. Show that, up to first order in ϵ,

$$\mathcal{N} = 1 + \frac{3}{2}\epsilon + \mathcal{O}(\epsilon^2). \tag{7.8.176}$$

Explain why all the following odd-n moments are exactly zero,

$$\langle (x - x_0)^n \rangle = 0, \qquad n = 1, 3, 5, 7, \dots. \tag{7.8.177}$$

Compute $\langle (x - x_0)^2 \rangle$, $\langle (x - x_0)^4 \rangle$, κ_2, and κ_4 up to first order in ϵ. Can you explain why $\kappa_n = 0$ for odd n? And, for n even, why does $\kappa_{n \geq 4} \to 0$ as $\epsilon \to 0$? Hint: Recall the higher cumulants for the Gaussian distribution. $\square$

Problem 7.54. Non-Perturbative Follow-Up to Problem (7.53) We now consider the case where ϵ in eq. (7.8.174) is positive but arbitrary. First show that

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left(\frac{x - x_0}{\sigma_0}\right)^2 \left(1 + \epsilon\left(\frac{x - x_0}{\sigma_0}\right)^2\right)\right) \mathrm{d}x = \frac{\sigma_0 \cdot e^{\frac{1}{16\epsilon_0}}}{2\sqrt{\epsilon_0}} K_{-\frac{1}{4}}\left(\frac{1}{16\epsilon_0}\right); \tag{7.8.178}$$

where $K_\nu(z)$ is the modified Bessel function.

Use this result to show that

$$\left\langle \left(\frac{x - x_0}{\sigma_0}\right)^2 \right\rangle = (4\epsilon)^{-1} \left(\frac{K_{\frac{3}{4}}\left(\frac{1}{16\epsilon}\right)}{K_{\frac{1}{4}}\left(\frac{1}{16\epsilon}\right)} - 1\right), \tag{7.8.179}$$

$$\left\langle \left(\frac{x - x_0}{\sigma_0}\right)^4 \right\rangle = (8\epsilon^2)^{-1} \left(1 + 4\epsilon - \frac{K_{\frac{3}{4}}\left(\frac{1}{16\epsilon}\right)}{K_{\frac{1}{4}}\left(\frac{1}{16\epsilon}\right)}\right). \tag{7.8.180}$$

(Hint: Consider re-scaling $\sigma_0 \to \sigma_0/\sqrt{\lambda}$ and $\epsilon \to \epsilon/\lambda^2$ for positive λ.) Next, compute

$$\left\langle \left(\frac{x - x_0}{\sigma_0}\right)^4 \right\rangle - 3\left\langle \left(\frac{x - x_0}{\sigma_0}\right)^2 \right\rangle^2, \tag{7.8.181}$$

which is closely related to κ_4'. Then, proceed to take the small and large ϵ limits of these results. You should find agreement with your results in Problem (7.53) when $0 < \epsilon \ll 1$; but discover that the large ϵ results yield inverse powers – for e.g., they begin at order $1/\epsilon$ – and therefore cannot be obtained by summing up a power series in ϵ. This $1/\epsilon$ behavior in the large ϵ limit is often dubbed "non-perturbative in ϵ", especially in the QFT literature. $\square$

Convolution Theorem and Multiple Distributions As is often the case, the outcome of a particular measurement receives contributions from more than one factor. For example, the height of a person depends on his/her genes, nutrition, environment, life history, etc. Suppose the random variable x arises from N independent contributions; i.e., $x = x_1 + x_2 + \dots + x_N$, where the probability of finding the I-th value to lie between x_I and $x_\mathrm{I} + \mathrm{d}x_\mathrm{I}$ is $P_\mathrm{I}(x_\mathrm{I})\mathrm{d}x_\mathrm{I}$. Since these are independent, the probability of obtaining the set $\{x_1, \dots, x_N\}$ is $\prod_{\mathrm{I}=1}^{N}(P_\mathrm{I}(x_\mathrm{I})\mathrm{d}x_\mathrm{I})$. However, what we measure is usually not the individual contributions but the net x itself. Therefore, the question is instead:

What is the probability of obtaining a given x, which is the sum of the individual x_{I}s?

This involves taking the above probability, but integrating over all $\{x_{\mathrm{I}}\}$ consistent with the constraint $x = \sum_{\mathrm{I}} x_{\mathrm{I}}$:

$$P(x)\mathrm{d}x \equiv \prod_{\mathrm{I}=1}^{N} \left(\int \mathrm{d}x_{\mathrm{I}} P_{\mathrm{I}}(x_{\mathrm{I}}) \right) \delta \left(x - \sum_{\mathrm{J}=1}^{N} x_{\mathrm{J}} \right) \mathrm{d}x. \tag{7.8.182}$$

This is the multiple distribution generalization of a convolution integral. For instance, when there are only two P_{I}s, we recover the $D = 1$ version of eq. (7.8.25):

$$P(x) \equiv \int \mathrm{d}x_1 \int \mathrm{d}x_2 P_1(x_1) P_2(x_2) \delta\left(x - x_1 - x_2\right) \tag{7.8.183}$$

$$= \int \mathrm{d}x' P_1(x') P_2(x - x'). \tag{7.8.184}$$

Problem 7.55. Check that total probability in eq. (7.8.182) is unity; namely, verify that $\int P(x)\mathrm{d}x = 1$. □

If we now employ the integral representation of the Dirac $\delta-$function,

$$\delta\left(x - \sum_{\mathrm{J}=1}^{N} x_{\mathrm{J}}\right) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} e^{-ik(x - \sum_{\mathrm{J}} x_{\mathrm{J}})}, \tag{7.8.185}$$

we recognize that the net probability density $P(x)$ is in fact the inverse Fourier decomposition of the product of the Fourier transforms of the individual Ps:

$$P(x) = \int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi} e^{-ikx} \prod_{\mathrm{I}=1}^{N} \widetilde{P}_{\mathrm{I}}(k). \tag{7.8.186}$$

By definition, $P(x) = \int \mathrm{d}k (2\pi)^{-1} e^{-ikx} \widetilde{P}(k)$. Reading $\widetilde{P}(k)$ off the coefficient of the exponential, we conclude that

$$\widetilde{P}(k) = \prod_{\mathrm{I}=1}^{N} \widetilde{P}_{\mathrm{I}}(k). \tag{7.8.187}$$

This is, in fact, the convolution theorem – the Fourier transform of a given convolution of two or more probability distributions $\{P_{\mathrm{I}}\}$ is the product of the Fourier transform of each and every individual distribution $\widetilde{P}_{\mathrm{I}}$. Moreover, one key reason why we introduced the concept of cumulants is because, we now learn that cumulants of the net probability distribution P is the sum of the cumulants of the N independent individual Ps that contribute to a given random variable x. Specifically, if $\kappa_n^{(\mathrm{I})}$ is the n-th cumulant of P_{I} keeping in mind eq. (7.8.147), we may read off the coefficient of $(ik)^n/m!$ in the exponent, via

$$\exp\left(\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n\right) = \prod_{\mathrm{I}=1}^{N} \exp\left(\frac{ik}{1!}\kappa_1^{(\mathrm{I})} + \frac{(ik)^2}{2!}\kappa_2^{(\mathrm{I})} + \frac{(ik)^3}{3!}\kappa_3^{(\mathrm{I})} + \dots\right) \tag{7.8.188}$$

$$= \exp\left(\sum_{\mathrm{I}=1}^{N}\left\{\frac{ik}{1!}\kappa_1^{(\mathrm{I})} + \frac{(ik)^2}{2!}\kappa_2^{(\mathrm{I})} + \frac{(ik)^3}{3!}\kappa_3^{(\mathrm{I})} + \dots\right\}\right) \tag{7.8.189}$$

$$= \exp\left(\sum_{n=1}^{\infty}\frac{(ik)^n}{n!}\kappa_n\right); \tag{7.8.190}$$

namely,

$$\kappa_n = \sum_{\mathrm{I}=1}^{N}\kappa_n^{(\mathrm{I})}. \tag{7.8.191}$$

We may, in turn, use these cumulants to obtain the $n-$moments of $P(x)$ itself.

Central Limit Theorem We shall now witness that, if P receives contributions from N independent $\{P_{\mathrm{I}}\}$; then in the limit as $N \to \infty$, the *normalized* cumulants of P higher than two will tend to zero. And since the Fourier transform of a Gaussian is another Gaussian, that teaches us:

The normalized version of $P(x)$ tends to a Gaussian as $N \to \infty$.

Starting from $P(x)$ itself, first recall from eq. (7.8.150) that its mean is simply the first cumulant:

$$\langle x\rangle \equiv x_0 = \kappa_1 = \sum_{\mathrm{I}=1}^{N}\kappa_1^{(\mathrm{I})} = \sum_{\mathrm{I}=1}^{N}x_0^{(\mathrm{I})}, \tag{7.8.192}$$

where $x_0^{(\mathrm{I})}$ is the mean of the I-th distribution P_{I}. The second cumulant, according to eq. (7.8.151) is the net variance σ^2:

$$\sigma^2 \equiv \langle(x-x_0)^2\rangle = \kappa_2 = \sum_{\mathrm{I}=1}^{N}\kappa_2^{(\mathrm{I})} \equiv N\overline{\kappa}_2; \tag{7.8.193}$$

where we have defined $\overline{\kappa}_2$ to be the average second cumulant of the $\{P_{\mathrm{I}}\}$. Since the second cumulant is involves the average of the *square* of the deviation from the mean of each P_{I}, it is reasonable to expect its average over the N independent distributions to be non-zero: $\overline{\kappa}_2 > 0$.

Next, we shall compute the cumulants of the normalized distribution. Using $\sigma = \sqrt{N\cdot\overline{\kappa}_2}$,

$$\int_{\mathbb{R}}\mathrm{d}x \exp\left(ik\frac{x-x_0}{\sqrt{N\overline{\kappa}_2}}\right)P(x) \equiv \exp\left(\sum_{n=1}^{\infty}\frac{(ik)^n}{n!}\kappa'_n\right). \tag{7.8.194}$$

Recall from eq. (7.8.168) and (7.8.169) that $\kappa'_1 = 0$ and $\kappa'_2 = 1$; whereas eq. (7.8.173) says

$$\kappa'_{n\geq 2} = \frac{\kappa_n}{(N\cdot\overline{\kappa}_2)^{n/2}}. \tag{7.8.195}$$

That is, as $N \to \infty$ the higher (normalized) cumulants κ'_n decays to zero as long as the average variance $\overline{\kappa}_2$ does not go to zero faster than $1/N$. More explicitly we have

$$\lim_{N\to\infty}\int_{\mathbb{R}}\mathrm{d}x \exp\left(ik\frac{x-x_0}{\sqrt{N\overline{\kappa}_2}}\right)P(x) = \exp\left(-\frac{k^2}{2}\right) \tag{7.8.196}$$

and – keeping in mind Problem (7.51) – hence,

$$\lim_{N\to\infty}P(x) = \exp\left(-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2\right). \tag{7.8.197}$$

7.8.5 Fourier Series

Consider a periodic function $f(x)$ with period L, meaning

$$f(x+L) = f(x). \tag{7.8.198}$$

Then its Fourier series representation is given by

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{i\frac{2\pi n}{L}x}, \tag{7.8.199}$$
$$C_n = \frac{1}{L}\int_{\text{one period}} \mathrm{d}x' f(x') e^{-i\frac{2\pi n}{L}x'}.$$

(I have derived this in our linear algebra discussion.) The Fourier series can be viewed as the discrete analog of the Fourier transform. In fact, one way to go from the Fourier series to the Fourier transform, is to take the infinite box limit $L \to \infty$. Just as the meaning of the Fourier transform is the decomposition of some wave profile into its continuous infinity of wave modes, the Fourier series can be viewed as the discrete analog of that. One example is that of waves propagating on a guitar or violin string – the string (of length L) is tied down at the end points, so the amplitude of the wave ψ has to vanish there

$$\psi(x=0) = \psi(x=L) = 0. \tag{7.8.200}$$

Even though the Fourier series is supposed to represent the profile ψ of a periodic function, there is nothing to stop us from imagining duplicating our guitar/violin string infinite number of times. Then, the decomposition in (7.8.199) applies, and is simply the superposition of possible vibrational modes allowed on the string itself.

Problem 7.56. (From Riley et al.) Find the Fourier series representation of the Dirac comb, i.e., find the $\{C_n\}$ in

$$\sum_{n=-\infty}^{\infty} \delta(x+nL) = \sum_{n=-\infty}^{\infty} C_n e^{i\frac{2\pi n}{L}x}, \qquad x \in \mathbb{R}. \tag{7.8.201}$$

Then prove the *Poisson summation formula*; where for an arbitrary function $f(x)$ and its Fourier transform $\widetilde{f}$,

$$\sum_{n=-\infty}^{\infty} f(x+nL) = \frac{1}{L}\sum_{n=-\infty}^{\infty} \widetilde{f}\left(\frac{2\pi n}{L}\right) e^{i\frac{2\pi n}{L}x}. \tag{7.8.202}$$

Hint: Note that

$$f(x+nL) = \int_{-\infty}^{+\infty} \mathrm{d}x' f(x')\delta(x-x'+nL). \tag{7.8.203}$$

$\square$

Problem 7.57. Gibbs Phenomenon The Fourier series of a discontinuous function suffers from what is known as the Gibbs phenomenon – near the discontinuity, the Fourier series does not fit the actual function very well. As a simple example, consider the periodic function $f(x)$ where within a period $x \in [0, L)$,

$$f(x) = -1, \qquad -L/2 \leq x \leq 0 \tag{7.8.204}$$
$$= 1, \qquad 0 \leq x \leq L/2. \tag{7.8.205}$$

Find its Fourier series representation

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{i\frac{2\pi n}{L}x}. \tag{7.8.206}$$

Since this is an odd function, you should find that the series becomes a sum over sines – cosine is an even function – which in turn means you can rewrite the summation as one only over positive integers n. Truncate this sum at $N = 20$ and $N = 50$, namely

$$f_N(x) \equiv \sum_{n=-N}^{N} C_n e^{i\frac{2\pi n}{L}x}, \tag{7.8.207}$$

and find a computer program to plot $f_N(x)$ as well as $f(x)$ in eq. (7.8.204). You should see the $f_N(x)$ over/undershooting the $f(x)$ near the latter's discontinuities, even for very large $N \gg 1$.[67] $\square$

YZ: Is there a more general way to understand this Gibbs phenomenon?

7.9 JWKB solution to $-\epsilon^2\psi''(x) + U(x)\psi(x) = 0$, for $0 < \epsilon \ll 1$

Many physicists encounter for the first time the following Jeffreys-Wentzel-Kramers-Brillouin (JWKB; aka WKB) method and its higher dimensional generalization, when solving the Schrödinger equation – and are told that the approximation amounts to the semi-classical limit where Planck's constant tends to zero, $\hbar \to 0$. Here, I want to highlight its general nature: it is not just applicable to quantum mechanical problems but oftentimes finds relevance when the wavelength of the solution at hand can be regarded as 'small' compared to the other length scales in the physical setup. The statement that electromagnetic waves in curved spacetimes or non-trivial media propagate predominantly on the null cone in the (effective) geometry, is in fact an example of such a 'short wavelength' approximation.

We will focus on the 1D case. Many physical problems reduce to the following 2nd order linear ordinary differential equation (ODE):

$$-\epsilon^2\psi''(x) + U(x)\psi(x) = 0, \tag{7.9.1}$$

where ϵ is a "small" (usually fictitious) parameter. This second order ODE is very general because both the Schrödinger and the (frequency space) Klein-Gordon equation with some potential reduces to this form. (Also recall that the first derivative terms in all second order

[67] See §5.7 of James Nearing's Math Methods book for a pedagogical discussion of how to estimate both the location and magnitude of the (first) maximum overshoot.

ODEs may be removed via a redefinition of ψ.) The main goal of this section is to obtain its approximate solutions.

We will use the ansatz

$$\psi(x) = \sum_{\ell=0}^{\infty} \epsilon^\ell \alpha_\ell(x) e^{iS(x)/\epsilon}.$$

Plugging this into our ODE, we obtain

$$0 = \sum_{\ell=0}^{\infty} \epsilon^\ell \left(\alpha_\ell(x)\left(S'(x)^2 + U(x)\right) - i\left(\alpha_{\ell-1}(x)S''(x) + 2S'(x)\alpha'_{\ell-1}(x)\right) - \alpha''_{\ell-2}(x)\right) \tag{7.9.2}$$

with the understanding that $\alpha_{-2}(x) = \alpha_{-1}(x) = 0$. We need to set the coefficients of ϵ^ℓ to zero. The first two terms ($\ell = 0, 1$) give us solutions to $S(x)$ and $\alpha_0(x)$.

$$0 = \alpha_0\left(S'(x)^2 + U(x)\right) \qquad \Rightarrow \qquad S_\pm(x) = \sigma_0 \pm i\int^x \mathrm{d}x'\sqrt{U(x')};\ \sigma_0 = \text{const.}$$

$$0 = -i\epsilon\left(2\alpha'_0(x)S'(x) + \alpha_0(x)S''(x)\right), \qquad \Rightarrow \qquad \alpha_0(x) = \frac{C_0}{U(x)^{1/4}}$$

(While the solutions $S_\pm(x)$ contains two possible signs, the $\pm$ in S' and S'' factors out of the second equation and thus α_0 does not have two possible signs.)

Problem 7.58. Recursion relation for higher order terms By considering the $\ell \geq 2$ terms in eq. (7.9.2), show that there is a recursion relation between $\alpha_\ell(x)$ and $\alpha_{\ell+1}(x)$. Can you use them to deduce the following two linearly independent JWKB solutions?

$$0 = -\epsilon^2\psi''_\pm(x) + U(x)\psi_\pm(x) \tag{7.9.3}$$

$$\psi_\pm(x) = \frac{1}{U(x)^{1/4}}\exp\left[\mp\frac{1}{\epsilon}\int^x \mathrm{d}x'\sqrt{U(x')}\right]\sum_{\ell=0}^{\infty}\epsilon^\ell Q_{(\ell|\pm)}(x), \tag{7.9.4}$$

$$Q_{(\ell|\pm)}(x) = \pm\frac{1}{2}\int^x \frac{\mathrm{d}x'}{U(x')^{1/4}}\frac{\mathrm{d}^2}{\mathrm{d}x'^2}\left(\frac{Q_{(\ell-1|\pm)}(x')}{U(x')^{1/4}}\right), \qquad Q_{(0|\pm)}(x) \equiv 1 \tag{7.9.5}$$

To lowest order

$$\psi_\pm(x) = \frac{1}{U^{1/4}(x)}\exp\left[\mp\frac{1}{\epsilon}\int^x \mathrm{d}x'\sqrt{U[x']}\right](1 + \mathcal{O}[\epsilon]). \tag{7.9.6}$$

Note: in these solutions, the $\sqrt{\cdot}$ and $\sqrt[4]{\cdot}$ are positive roots. □

JWKB Counts Derivatives In terms of the $Q_{(n)}$s we see that the JWKB method is really an approximation that works whenever each dimensionless derivative $\mathrm{d}/\mathrm{d}x$ acting on some power of $U(x)$ yields a smaller quantity, i.e., roughly speaking $\mathrm{d}\ln U(x)/\mathrm{d}x \sim \epsilon \ll 1$; this small derivative approximation is related to the short wavelength approximation. Also notice from the exponential $\exp[iS/\epsilon] \sim \exp[\pm(i/\epsilon)\int\sqrt{-U}]$ that the $1/\epsilon$ indicates an integral (namely, an inverse derivative). To sum:

The ficticious parameter $\epsilon \ll 1$ in the JWKB solution of $-\epsilon^2\psi'' + U\psi = 0$ counts the number of derivatives; whereas $1/\epsilon$ is an integral. The JWKB approximation works well whenever each additional dimensionless derivative acting on some power of U yields a smaller and smaller quantity.

Breakdown and connection formulas There is an important aspect of JWKB that I plan to discuss in detail in a future version of these lecture notes. From the $1/\sqrt[4]{U(x)}$ prefactor of the solution in eq. (7.9.4), we see the approximation breaks down at $x = x_0$ whenever $U(x_0) = 0$. The JWKB solutions on either side of $x = x_0$ then need to be joined by matching onto a valid solution in the region $x \sim x_0$. One common approach is to replace U with its first non-vanishing derivative, $U(x) \to ((x - x_0)^n/n!)U^{(n)}(x_0)$; if $n = 1$, the corresponding solutions to the 2nd order ODE are Airy functions – see, for e.g., Sakurai's *Modern Quantum Mechanics* for a discussion. Another approach, which can be found in Matthews and Walker [15], is to complexify the JWKB solutions, perform analytic continuation, and match them on the complex plane.

8 Calculus of Variation

8.1 Fundamentals

[68]In single-variable calculus, the turning points $\{x_i | i = 1, 2, 3, \dots\}$ of a function $f(x)$ of the real variable x are characterized by the equations

$$f'(x_i) = 0. \tag{8.1.1}$$

That is, the local extremum – minimum, maximum, or inflection point – occurs whenever the first order variation $\mathrm{d}f = f'(x)\mathrm{d}x$ is zero. In this section, we will study the *calculus of variation*, where we will determine the equations obeyed by the function $q(\lambda)$ of some real variable λ such that the following *action* integral involving it and its first derivative

$$S[q] \equiv \int_{\tau'}^{\tau} L\left(\lambda, q(\lambda), q'(\lambda)\right) \mathrm{d}\lambda \tag{8.1.2}$$

is extremized, provided the boundary conditions

$$q(\tau') = q_1 \qquad \text{and} \qquad q(\tau) = q_2 \tag{8.1.3}$$

are specified. To elaborate what it means for an action such as eq. (8.1.2) to be extremized means, suppose we found such a trajectory $\bar{q}(\lambda)$. Then *all* the 'nearby' trajectories $q(\lambda) = \bar{q}(\lambda) + \delta q(\lambda)$, for 'small' δq, would either yield larger (local minimum), smaller (local maximum), or the same (local inflection) value for S.

Here, we will assume the *Lagrangian* $L(\lambda, a, b)$ is a given differentiable function of the variables λ, a, and b. This sort of 'variational principle' problems occur throughout physics; from the Lagrangian-Hamiltonian formulation of classical mechanics and field theory, the principle of least time in ray optics, geodesics in curved (space)time, etc.

Euler-Lagrange Equations in 1D The answer is, to extremize S in eq. (8.1.2), the analogy to eq. (8.1.1) is the *Euler-Lagrange* equation

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial q'(\lambda)} - \frac{\partial L}{\partial q(\lambda)} = 0, \tag{8.1.4}$$

– whose expanded form reads

$$\frac{\partial^2 L}{\partial \lambda \partial q'} + q'' \frac{\partial^2 L}{\partial q'^2} + q' \frac{\partial^2 L}{\partial q \partial q'} - \frac{\partial L}{\partial q} = 0 \tag{8.1.5}$$

– subject to the boundary conditions in eq. (8.1.3). In equations (8.1.4) and (8.1.5) and the rest of this Chapter, whenever the derivative $\partial/\partial\lambda$ is acting on the Lagrangian, it is carried out with q and q' held fixed. Likewise, whenever $\partial/\partial q'$ is acting on L it is carried out with λ and q held fixed; and $\partial L/\partial q$ is performed with λ and q' held fixed.

Derivation of Euler-Lagrange eq. (8.1.4) If S is extremized by $q(\lambda)$, that means upon a slight perturbation $\delta q(\lambda)$, where we replace in eq. (8.1.2)

$$q(\lambda) \to q(\lambda) + \delta q(\lambda) \qquad \text{and} \tag{8.1.6}$$

[68]In writing this Chapter, I have consulted Arfken et al. [18], Byron and Fuller [14] and Morse and Feshbach [13].

$$q'(\lambda) \to q'(\lambda) + \frac{\mathrm{d}}{\mathrm{d}\lambda}\delta q(\lambda), \tag{8.1.7}$$

the first-order-variation of S – the portion of S linear in δq – must vanish. This is analogous to the turning point condition in eq. (8.1.1), where the first order variation of $f(x)$ is zero upon replacing $x \to x_i + \mathrm{d}x$. Note that all trajectories need to obey the boundary conditions in eq. (8.1.3); this includes both $q(\lambda)$ and the perturbed $q(\lambda) + \delta q(\lambda)$. Therefore, we must have

$$\delta q(\tau') = 0 = \delta q(\tau). \tag{8.1.8}$$

We are now ready to carry out the perturbation as a (formal) power series in q:

$$\begin{aligned} S[q + \delta q] &= \int_{\tau'}^{\tau} L\left(\lambda, q(\lambda) + \delta q(\lambda), q'(\lambda) + \delta q'(\lambda)\right) \mathrm{d}\lambda & (8.1.9)\\ &= S[q] + \delta_1 S[q] + \mathcal{O}\left(\delta q^2\right), & (8.1.10) \end{aligned}$$

where

$$\begin{aligned} \delta_0 S[q] &\equiv \int_{\tau'}^{\tau} L\left(\lambda, q(\lambda), q'(\lambda)\right) \mathrm{d}\lambda, & (8.1.11)\\ \delta_1 S[q] &\equiv \int_{\tau'}^{\tau} \left(\frac{\partial L}{\partial q(\lambda)}\delta q(\lambda) + \frac{\partial L}{\partial q'(\lambda)}\frac{\mathrm{d}}{\mathrm{d}\lambda}\delta q(\lambda)\right) \mathrm{d}\lambda. & (8.1.12) \end{aligned}$$

From this calculation, we see that the $\partial L/\partial q$ arises from varying q but holding the $q'(\lambda)$ argument of L fixed; whereas $\partial L/\partial q'$ comes about due to the variation of $q'(\lambda)$ but holding q argument of L fixed. Next, we integrate-by-parts the derivative in $\mathrm{d}\delta q/\mathrm{d}\lambda$,

$$\delta_1 S[q] = \int_{\tau'}^{\tau} \delta q(\lambda)\left(\frac{\partial L}{\partial q(\lambda)} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial q'(\lambda)}\right)\mathrm{d}\lambda + \left[\frac{\partial \mathcal{L}}{\partial q'(\lambda)}\delta q(\lambda)\right]_{\lambda=\tau'}^{\lambda=\tau}. \tag{8.1.13}$$

But the boundary terms vanish because eq. (8.1.8) reminds us the trajectory perturbation has to be trivial there. At this juncture, we see that the first-order-variation is zero iff

$$\delta_1 S[q] = \int_{\tau'}^{\tau} \delta q(\lambda)\left(\frac{\partial L}{\partial q(\lambda)} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial q'(\lambda)}\right)\mathrm{d}\lambda = 0 \tag{8.1.14}$$

for arbitrary small perturbations $\delta q(\tau' \leq \lambda \leq \tau)$. Now, if the integral

$$\int_{\tau'}^{\tau} \delta q(\lambda) F(\lambda) \mathrm{d}\lambda \tag{8.1.15}$$

vanishes for *arbitrary* small perturbations δq and if F *were not* exactly zero, then we may simply choose δq to be a smoothed-out 'top hat' within some small $\lambda-$region where F is either strictly positive or negative – assuming, of course, F itself is continuous. But then the integral would necessarily produce a corresponding positive or negative number, contradicting the fact that it has to be zero. This contradiction implies $F = 0$. In turn, it teaches us the factor multiplying δq in $\delta_1 S$ must therefore be zero, leading us to the Euler-Lagrange equations (8.1.4). $\square$

'Constant-of-Motion' Let us also note that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(q'\frac{\partial L}{\partial q'}-L\right)=q''\frac{\partial L}{\partial q'}+q'^2\frac{\partial^2 L}{\partial q'\partial q}+q''q'\frac{\partial^2 L}{\partial q'^2}+q'\frac{\partial^2 L}{\partial q'\partial\lambda}$$
$$-q'\frac{\partial L}{\partial q}-q''\frac{\partial L}{\partial q'}-\frac{\partial L}{\partial\lambda} \tag{8.1.16}$$
$$=q'\left(q'\frac{\partial^2 L}{\partial q'\partial q}+q''\frac{\partial^2 L}{\partial q'^2}+\frac{\partial^2 L}{\partial q'\partial\lambda}-\frac{\partial L}{\partial q}\right)-\frac{\partial L}{\partial\lambda} \tag{8.1.17}$$

Comparison with eq. (8.1.5) then teaches us, as long as $q'\neq 0$, the following is equivalent form of the Euler-Lagrange equation (8.1.4):

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(q'\frac{\partial L}{\partial q'}-L\right)+\frac{\partial L}{\partial\lambda}=0. \tag{8.1.18}$$

In particular, notice if L depends on λ only through q and q', then if $q(\lambda)$ is a solution to eq. (8.1.4) and $q'\neq 0$,

$$q'\frac{\partial L}{\partial q'}-L=E=\text{constant}. \tag{8.1.19}$$

Second Order to First Order ODE The existence of a 'constant-of-motion' in eq. (8.1.19) allows certain classes of second order ordinary differential equations (ODEs) to be converted into first order ones. This, in turn, allows the solutions to be determined – albeit sometimes only implicitly – in terms of integrals involving the expressions arising from the original ODE. For example, consider:

$$q''(\lambda)+U'(q(\lambda))=0. \tag{8.1.20}$$

One can readily verify that the $\lambda-$independent Lagrangian that gives rise to this equation is

$$L(q,q')=\frac{1}{2}q'(\lambda)^2-U(q(\lambda)). \tag{8.1.21}$$

If $q(\lambda)$ is a solution then according to eq. (8.1.19) the corresponding constant E is

$$q'\frac{\partial L}{\partial q'}-L=\frac{1}{2}q'^2+U(q)=E \tag{8.1.22}$$
$$F(q)\equiv\int\frac{\mathrm{d}q}{\sqrt{2}\sqrt{E-U(q)}}=\pm(\lambda-\lambda_0). \tag{8.1.23}$$

If the above integral can be performed, we would then have a function $F(q)$ on the left hand side; and if its inverse F^{-1} can be determined, we may then apply it on both sides to obtain $q(\lambda)=F^{-1}(\pm(\lambda-\lambda_0))$. Moreover, the two constants E and λ_0 are then determined by imposing an appropriate pair of initial and/or boundary conditions.

As an example, suppose $U(q)=(\omega^2/2)x^2$; i.e., we wish to solve

$$q''+\omega^2 q=0, \tag{8.1.24}$$

The solution is a linear combination of $\sin(\omega\lambda)$ and $\cos(\omega\lambda)$. But let us use the conserved quantity E and integrate

$$F(q) = \int \frac{\mathrm{d}q}{\sqrt{2}\sqrt{E-(\omega^2/2)q^2}} = \frac{1}{\omega}\arcsin\left(\frac{\omega q}{\sqrt{2E}}\right) = \pm(\lambda-\lambda_0) \tag{8.1.25}$$

$$q(\lambda) = \pm\frac{\sqrt{2E}}{\omega}\sin\left(\omega(\lambda-\lambda_0)\right). \tag{8.1.26}$$

Since E was arbitrary, the overall coefficient may be replaced with an arbitrary constant A:

$$q(\lambda) = A\sin\left(\omega(\lambda-\lambda_0)\right). \tag{8.1.27}$$

Problem 8.1. Higher Derivatives Suppose we demand the action be extremized, but now the Lagrangian depends on λ, as well as q, $q', \dots, q^{(n)}$; i.e., from the zeroth through the $(n>1)$th derivative $q^{(n)} \equiv \mathrm{d}^n q/\mathrm{d}\lambda^n$. Show that the Euler-Lagrange equations now become

$$(-)^{n+1}\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n}\frac{\partial L}{\partial q^{(n)}} + (-)^n\frac{\mathrm{d}^{n-1}}{\mathrm{d}\lambda^{n-1}}\frac{\partial L}{\partial q^{(n-1)}} + \dots - \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\frac{\partial L}{\partial q^{(2)}} + \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial q^{(1)}} = \frac{\partial L}{\partial q}. \tag{8.1.28}$$

What are the appropriate boundary conditions? (Hint: You should find, $2n$ of them are needed.) Explain why the highest $\lambda-$derivative that occurs in eq. (8.1.28) is $q^{(2n)}$. In fact, you should be able to argue that eq. (8.1.28) is linear in $q^{(2n)}$. □

Euler-Lagrange Equations in Arbitary Dimensions It is not difficult to generalize the preceding discussion to arbitrary dimensions $D \geq 1$. Suppose $\vec{q}(\tau' \leq \lambda \leq \tau)$ joins $\vec{y}'$ to $\vec{y}$; and suppose we have a Lagrangian built out of $\vec{q}$ and its first derivative $\vec{q}'(\lambda)$. Then if $\vec{q}$ extremizes the action

$$S = \int_{\tau'}^{\tau} L\left(\lambda, \vec{q}, \dot{\vec{q}}\right)\mathrm{d}\lambda, \tag{8.1.29}$$

$$\vec{q}(\lambda=\tau') = \vec{y}', \qquad \vec{q}(\lambda=\tau) = \vec{y}; \tag{8.1.30}$$

with the end points $\vec{y}'$ and $\vec{y}$ held fixed, the trajectory itself obeys the $D-$dimensional Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{q}^i} = \frac{\partial L}{\partial q^i}, \tag{8.1.31}$$

$$\ddot{q}^j\frac{\partial^2 L}{\partial \dot{q}^j\partial\dot{q}^i} + \dot{q}^j\frac{\partial^2 L}{\partial q^j\partial\dot{q}^i} + \frac{\partial L}{\partial\lambda\partial\dot{q}^i} = \frac{\partial L}{\partial q^i}; \tag{8.1.32}$$

where $\dot{q}^i \equiv \mathrm{d}q^i/\mathrm{d}\lambda$; and $\partial L/\partial\dot{\vec{q}}$ is carried out with $\vec{q}$ held fixed while $\dot{\vec{q}}$ is held fixed in $\partial L/\partial\vec{q}$.

Overall Constants Don't Matter Notice: both the 1D (8.1.4) and multi-dimensional (8.1.31) Euler-Lagrange equations are linear in the Lagrangian L. Hence, for arbitrary constant $C \neq 0$, both the re-scaled Lagrangian $C \cdot L$ and original L would yield the same equations. As far as the Euler-Lagrange equations as concerned, we may therefore drop all overall multiplicative constants in L.

Proof As with the 1D case, we consider perturbations of the path

$$\vec{q}(\lambda) \to \vec{q} + \delta\vec{q}, \tag{8.1.33}$$

$$\dot{\vec{q}}(\lambda) \to \dot{\vec{q}} + \frac{\mathrm{d}}{\mathrm{d}\lambda}\delta\vec{q}, \tag{8.1.34}$$

but since the end points are held fixed, we shall demand the perturbations vanish at the end points

$$\delta\vec{q}(\tau') = \vec{0} = \delta\vec{q}(\tau). \tag{8.1.35}$$

We now examine the ensuing perturbations of the action. Denoting $\delta_{n\geq 1}S$ to be the part of $S[\vec{q} + \delta\vec{q}]$ containing precisely n powers of $\delta\vec{q}$:

$$S[\vec{q}] \to S[\vec{q} + \delta\vec{q}] = S[\vec{q}] + \delta_1 S[\vec{q}] + \delta_2 S[\vec{q}] + \dots \tag{8.1.36}$$

$$\delta_1 S = \int_{\tau'}^{\tau} \left(\frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^i}\frac{\mathrm{d}\delta q^i}{\mathrm{d}\lambda} \right) \mathrm{d}\lambda \tag{8.1.37}$$

$$= \int_{\tau'}^{\tau} \delta q^i \left(\frac{\partial L}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{q}^i} + \frac{\mathrm{d}}{\mathrm{d}\lambda}\left\{ \frac{\partial L}{\partial \dot{q}^i}\delta q^i(\lambda) \right\} \right) \mathrm{d}\lambda \tag{8.1.38}$$

$$= \int_{\tau'}^{\tau} \delta q^i \left(\frac{\partial L}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{q}^i} \right) \mathrm{d}\lambda + \left[\frac{\partial L}{\partial \dot{q}^i}\delta q^i(\lambda) \right]_{\lambda=\tau'}^{\lambda=\tau}. \tag{8.1.39}$$

There is an implicit sum over $i = 1, 2, \dots, D$; i.e., there are D independent variations from each component of $\vec{q}$. By eq. (8.1.35) the last term of the last line is zero. Since each component is independent from the rest, we may recall the above 1D argument – that $\int_{\tau'}^{\tau} \delta q \cdot F \mathrm{d}\lambda = 0$ for *arbitrary* δq implies $F = 0$ – to infer eq. (8.1.31) has been recovered, if we demand $\delta_1 S = 0$, the action is stationary under first order in perturbations.

Min, Max or Inflection Point? Both the 1D (8.1.4) and multi-dimensional (8.1.31) Euler-Lagrange equations are necessary and sufficient conditions for an *extremum*: $\vec{q}$ extremizes its action S if and only if its corresponding eq. (8.1.31) holds. However, the solution itself does not inform us if it provides a (local) minimum, maximum or inflection point. To answer this, we will compute below the second order corrections to the action $\delta_2 S$ induced by perturbations about the solution.

Second Order ODEs Notice, from the expanded form in eq. (8.1.31), the Euler-Lagrange equations are second order ones because the L itself depends only on $\vec{q}$ and its first derivative. In fact, since the only dependence on the second derivative $\ddot{q}$ in eq. (8.1.31) appears in its first term, we see it must be linear in $\ddot{q}$. From the theory of ODEs, we know that a second order differential equation requires two distinct conditions to be specified in order for a unique solution to be obtained. This is consistent with the pair of boundary conditions $\vec{q}(\tau')$ and $\vec{q}(\tau)$ needed to render the boundary terms zero – recall equations (8.1.35) and (8.1.39) – so as to ensure the variation of S is zero at first order.

Perturbation Theory Physical applications often require the use of perturbation theory, because exact closed-form solutions are hard to come by. As we will discover below, much of dynamics is encoded via the action principle we are exploring here. To this end, let us record the following useful observation, a corollary of sorts of equations (8.1.31) employed in (8.1.39):

> If an action is perturbed away from a solution $\vec{q}(\lambda)$ of its corresponding Euler-Lagrange equations, namely $S[\vec{q}] \to S[\vec{q} + \delta\vec{q}]$ but $\delta\vec{q}$ is now arbitrary and no longer needs to vanish at the end points; then – up to additive boundary terms – its first order perturbation vanishes.

Problem 8.2. When L is λ-independent. An important point to remember:

> If the Lagrangian $L(\vec{q}(\lambda), \dot{\vec{q}}(\lambda))$ does not depend on λ explicitly, the solutions to its Euler-Lagrange equations always admit at least one constant E.

Show that it is given by

$$E = \dot{q}^i \cdot \frac{\partial L}{\partial \dot{q}^i} - L. \tag{8.1.40}$$

This is of course the multi-dimensional generalization of eq. (8.1.19). Below, we shall see that the right hand side of eq. (8.1.40) is the *Hamiltonian*. In many cases, it is simply the conserved energy. □

Problem 8.3. Cyclic (or, Ignorable) Coordinate When the Lagrangian does not depend explicitly on a coordinate q^k (for some fixed $1 \leq k \leq D$) we say it is *cyclic* or *ignorable*. Explain why, if L is independent of q^k, then

$$p_k \equiv \frac{\partial L}{\partial \dot{q}^k} \tag{8.1.41}$$

is a constant-of-motion. □

Is the Lagrangian unique? Before moving on to tackle some concrete examples, we inquire: Is the Lagrangian $L(\lambda, \vec{q}, \dot{\vec{q}})$ unique? Such a question arises in physics because, in many situations – including the study of the fundamental interactions of Nature – the starting point of any theoretical investigation begins with the specification of the Lagrangian, oftentimes guided by symmetry requirements. The (non-)uniqueness of the Lagrangian is therefore tied to the (non-)uniqueness of our theoretical starting point.

More specifically, suppose we have two Lagrangians, $L_1(\lambda, \vec{q}, \dot{\vec{q}})$ and $L_2(\lambda, \vec{q}, \dot{\vec{q}})$, and they yield the same Euler-Lagrange equations; namely,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_1}{\partial \dot{q}^i} - \frac{\partial L_1}{\partial q^i} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_2}{\partial \dot{q}^i} - \frac{\partial L_2}{\partial q^i}. \tag{8.1.42}$$

How are L_1 and L_2 related?[69] We will now show that, for the 1D case, two Lagrangians that depend on only λ, $q(\lambda)$ and $q'(\lambda)$ (i.e., no higher derivatives) that give the same Euler-Lagrange equations, can only differ up to an additive total time derivative:

$$L_1(\lambda, q, q') = L_2(\lambda, q, q') + \frac{\mathrm{d}F(\lambda, q)}{\mathrm{d}\lambda}. \tag{8.1.43}$$

(The F cannot depend on q' or higher derivatives.) The answer for the $D-$dimensional case is similar, but we leave it to the reader to prove in Problem (9.100). This result also teaches us,

[69] The reader may rightly object that eq. (8.1.42) is too strict of a requirement. For e.g., since both sides are actually equal to zero, they can each be multiplied by a different non-zero factor and the resulting equations-of-motion are still the same; i.e., we may only require $(\mathrm{d}/\mathrm{d}\lambda)(\partial L_1/\partial \dot{q}^i) - \partial L_1/\partial q^i = A\{(\mathrm{d}/\mathrm{d}\lambda)(\partial L_2/\partial \dot{q}^i) - \partial L_2/\partial q^i\}$, for $A(\lambda, \vec{q}, \dot{\vec{q}}, \ddot{\vec{q}}) \neq 0$. We leave this more general case for the reader's analysis.

the actions constructed from L_1 and L_2 are extremized by the same solutions; and they differ only by boundary terms, which are assumed to be fixed:

$$\int_{\tau'}^{\tau} L_1(\lambda, q, q')\mathrm{d}\lambda = \int_{\tau'}^{\tau} L_2(\lambda, q, q')\mathrm{d}\lambda + F(\tau, q(\tau)) - F(\tau', q(\tau')) \,. \tag{8.1.44}$$

The key to this proof of the (near-)uniqueness of the Lagrangian is the *Poincaré Lemma*, which we shall state for the 2 variable case here:

> Within a simply-connected region of 'space' parametrized by $\{x, y\}$, a two component object of two variables $V_i(x, y) = (V_x, V_y)$ (i.e., the i refers to either the x or y component) is a pure gradient – namely, $V_i(x, y) = \partial_i \varphi(x, y)$ – if and only if $\partial_x V_y - \partial_y V_x = 0$.

We write $L_1(\lambda, q, q') = L_2(\lambda, q, q') + \Delta L(\lambda, q, q')$, where ΔL is *defined* to be the difference between L_1 and L_2. Then, noting that the Euler-Lagrange operator in eq. (8.1.4) is linear, we see when it acts on ΔL, the result must be identically zero. Using its expanded version in eq. (8.1.5),

$$q'' \frac{\partial^2 \Delta L}{\partial q'^2} + q' \frac{\partial^2 \Delta L}{\partial q \partial q'} + \frac{\partial^2 \Delta L}{\partial \lambda \partial q'} - \frac{\partial \Delta L}{\partial q} = 0. \tag{8.1.45}$$

Note that this cannot depend on some specific form of q, q', and q''. It is identically zero due to the form of ΔL. Since ΔL depends on q, q' but not on q'' or higher derivatives, the only occurrence of the second derivative is thus in the first term. ΔL itself must in fact be at most linear in q'; namely,

$$\frac{\partial^2 \Delta L}{\partial q'^2} = 0. \tag{8.1.46}$$

Now we apply the Euler-Lagrange operator to

$$\Delta L = V_1(\lambda, q) + q' V_2(\lambda, q). \tag{8.1.47}$$

Plugging this back into eq. (8.1.45) hands us

$$q' \frac{\partial V_2}{\partial q} + \frac{\partial V_2}{\partial \lambda} - q' \frac{\partial V_2}{\partial q} - \frac{\partial V_1}{\partial q} = 0 \tag{8.1.48}$$

$$\frac{\partial V_2}{\partial \lambda} - \frac{\partial V_1}{\partial q} = 0. \tag{8.1.49}$$

If we identify $(x^1, x^2) \equiv (\lambda, q)$, we may re-write the second line in the form occurring in the Poincaré Lemma; i.e.,

$$\partial_1 V_2 - \partial_2 V_1 = 0. \tag{8.1.50}$$

At this point, the Poincaré Lemma itself tells us (V_1, V_2) is a gradient:

$$(V_1, V_2) = (\partial_\lambda F(\lambda, q), \partial_q F(\lambda, q)) \,, \tag{8.1.51}$$

for some scalar F that depends only on λ and q (and not on higher derivatives of q). According to eq. (8.1.47), ΔL itself is now

$$\Delta L = \partial_\lambda F + q' \partial_q F = \frac{\mathrm{d}F(\lambda, q)}{\mathrm{d}\lambda}. \tag{8.1.52}$$

In other words, our two starting Lagrangians L_1 and L_2 differs at most by a total time derivative.

Problem 8.4. Verify explicitly that the Euler-Lagrange equations applied to $\mathrm{d}F(\lambda, q(\lambda))/\mathrm{d}\lambda$ is indeed zero:

$$\left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial}{\partial q'} - \frac{\partial}{\partial q}\right)\frac{\mathrm{d}F(\lambda, q(\lambda))}{\mathrm{d}\lambda} = 0. \tag{8.1.53}$$

This provides an explicit check that both L_1 and L_2 in eq. (8.1.43) do indeed produce the same differential equations. □

Problem 8.5. One of the key assumptions we have made is that the Lagrangian itself depends on (the independent) variable λ; as well as (the dependent) variable $q(\lambda)$ and its first derivative $q'(\lambda)$ – but not on its higher derivatives $q''(\lambda)$, $q'''(\lambda)$, etc. Suppose we allow L to depend also on the second derivative; namely, $L(\lambda, q, q', q'')$; but still require that its Euler-Lagrange equations in eq. (8.1.28) yield second order differential equations. Argue that L can only be linear in q'':

$$L(\lambda, q, q', q'') = L_1(\lambda, q, q') + q'' \cdot L_2(\lambda, q, q'). \tag{8.1.54}$$

How many boundary conditions would you need to ensure its action $\int_{\tau'}^{\tau} L\mathrm{d}\lambda$ is extremized? Comment on its consistency with the second order nature of the differential equation. Hints: Vary the part of the action that depends on $q'' \cdot L_2$. You should find that $\delta q(\tau') = 0 = \delta q(\tau)$ in eq. (8.1.8) is not sufficient to render the boundary terms zero. **YZ: Chih-Chien tells me that $\partial^3 L/\partial q^3 = 0$ too. Very constrained!**

This problem teaches us, while it is (formally) possible to obtain second order equations from higher order Lagrangians, the associated variational principle requires too many boundary conditions – i.e., it yields an over-determined set of ODEs – and is therefore generically inconsistent. □

Example: Shortest Distance in 2D The infinitesimal distance on the (x, y) plane is $\mathrm{d}\ell = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \sqrt{1 + (\mathrm{d}y/\mathrm{d}x)^2}\mathrm{d}x$. Assuming $y(x)$ is not multi-valued – i.e., assuming the path is not too curvy so that there is more than one y value for a given x – we may then write the total length spanned by $y(x_1 \leq x \leq x_2)$ as

$$\ell = \int_{x_0}^{x_1} \sqrt{1 + y'(x)^2}\mathrm{d}x. \tag{8.1.55}$$

Eq. (8.1.4) says

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial\sqrt{1 + y'(x)^2}}{\partial y'(x)} = \frac{\partial\sqrt{1 + y'(x)^2}}{\partial y}, \tag{8.1.56}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{y'(x)}{\sqrt{1 + y'(x)^2}} = 0; \tag{8.1.57}$$

which tells us $y'(x)/\sqrt{1 + y'(x)^2}$ is some $x-$independent constant C. Hence, the slope $y'(x)$ can be solved in terms of some other constant related to C, which is exactly the condition of a straight line. If the end points are (x_0, y_0) and (x_1, y_1), the unique solution is

$$y(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0). \tag{8.1.58}$$

An alternate manner to derive the same result is to appeal to the constancy of eq. (8.1.19).

$$y' \frac{\partial \sqrt{1+y'^2}}{\partial y'} - \sqrt{1+y'^2} = -C \tag{8.1.59}$$

$$\frac{1}{\sqrt{1+y'^2}} = C \tag{8.1.60}$$

Again, we see that $y'(x)$ must be some constant related to C. Note that one may always make a line longer by adding more wiggles in between, for instance; hence, there cannot be a longest line. This straight line solution must therefore be a minimum length one.

Problem 8.6. Straight line in infinite flat $D-$space It is possible to remove this restrictive assumption that $y(x)$ cannot be multi-valued by introducing an auxiliary parameter λ such that $\vec{z}(\lambda)$ parametrizes some path in infinite (flat) space. In fact, what follows works for any dimension $D \geq 2$. Show that the shortest path between the fixed points $\vec{y}_0$ and $\vec{y}_1$ is

$$\vec{q}(0 \leq \lambda \leq 1) = \vec{y}_0 + f(\lambda)\left(\vec{y}_1 - \vec{y}_0\right), \tag{8.1.61}$$

– where f is an arbitrary monotonically increasing function subject to $f(0) = 0$ and $f(1) = 1$ – by extremizing the length integral

$$\ell = \int_{\vec{y}_1}^{\vec{y}_2} \sqrt{\mathrm{d}\vec{q} \cdot \mathrm{d}\vec{q}} = \int_0^1 \sqrt{\dot{\vec{q}}^2} \mathrm{d}\lambda \tag{8.1.62}$$

using the Euler-Lagrange equations (8.1.31). Can you explain why this extremum is a minimum? Hint: $\dot{\vec{q}}^2$ is strictly non-negative. □

Example: Ray Optics and Fermat's Principle of Least Time In a medium, we may postulate that the effective speed of light now depends on its location,

$$\frac{|\mathrm{d}\vec{x}|}{\mathrm{d}t} = \frac{c}{n(\vec{x})}, \tag{8.1.63}$$

where c is the speed of light in vacuum and $n > 1$ is the refractive index; namely, the speed of light in the medium c/n is *slower* than that in vacuum.

Fermat's principle states that light rays follow paths that take the *least* time. In other words, we need to minimize

$$c \int \mathrm{d}t \equiv c \cdot \Delta t = \int_{\vec{z}'}^{\vec{z}} n(\vec{x}) |\mathrm{d}\vec{x}| \tag{8.1.64}$$

$$= \int_{\tau'}^{\tau} n\left(\vec{x}(\lambda)\right) \sqrt{\dot{\vec{x}}(\lambda)^2} \mathrm{d}\lambda. \tag{8.1.65}$$

The Euler-Lagrange equations are

$$\frac{1}{\sqrt{\dot{\vec{x}}^2}} \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{n(\vec{x})}{\sqrt{\dot{\vec{x}}^2}} \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\right) = \frac{\partial n(\vec{x})}{\partial x^i}. \tag{8.1.66}$$

If we exploit the infinitesimal length as our parameter $\mathrm{d}\ell = \sqrt{\dot{\vec{x}}}\mathrm{d}\lambda$,

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\left(n\left(\vec{x}\right)\frac{\mathrm{d}x^i}{\mathrm{d}\ell}\right) = \frac{\partial n\left(\vec{x}\right)}{\partial x^i}. \tag{8.1.67}$$

This ray equation has an analog even in curved (static) spacetimes – see Problem (11.15) and its eq. (11.3.118).

Problem 8.7. Action Principle for General Linear 2nd Order ODE The general linear second order ordinary differential equation defined within some interval $z \in [a, b]$, phrased in the 'Hermitian form' (cf. eq. (5.11.9)), with some fixed 'source' J is

$$\frac{1}{w(z)}\frac{\mathrm{d}}{\mathrm{d}z}\left(w(z)p_2(z)\frac{\mathrm{d}\psi(z)}{\mathrm{d}z}\right) + p_0(z)\psi(z) = J(z). \tag{8.1.68}$$

Show that this can be recovered by extremizing

$$S[\psi] \equiv \int_a^b w(z)\left(\frac{p_2(z)\psi'(z)^2 - p_0(z)\psi(z)^2}{2} + \psi(z)J(z)\right)\mathrm{d}z. \tag{8.1.69}$$

As far as the differential equation is concerned, we shall assume that $w(z), p_2(z) > 0$ within $z \in [a, b]$ except perhaps at isolated points. □

Is a minimum? An example We now turn to an example of a variational problem where more than one possible solution exists, and therefore more analysis is necessary to figure out which one is the true minimum. Consider two circular loops of radius r_0 lying flat on the (x, y) plane in 3D (flat) space, where both are centered at $(0, 0)$. The $z-$coordinate of one of them is $+z_0$ and the other is $-z_0$. That is, we have the two rings described by

$$\vec{x}_\pm = r_0\left(\cos\phi, \sin\phi, \pm z_0/r_0\right), \qquad 0 \leq \phi < 2\pi. \tag{8.1.70}$$

Now, we ask:

> If an infinitesimally thin 2D membrane joins the two rings – Arfken et al [18] calls it a soap bubble – what is the shape that yields the minimum area? And, what is this minimum area?

By stretching this 2D membrane, we see that there is no upper limit to how large its area can be. Hence, no solution can be the global maximum. Also, since there cannot be a zero area solution, there *has to be* a global minimum.

Next, by axial symmetry, we may view the area spanned by this membrane as the surface of revolution gotten by rotating a curve on the $(0, y, z)$ plane joining the two points $(0, r_0, \pm z_0)$. Thus, the area is given by

$$\begin{aligned} A &= 2\pi \int y\sqrt{\mathrm{d}y^2 + \mathrm{d}z^2} \\ &= 2\pi \int_{-z_0}^{z_0} y(z)\sqrt{1 + y'(z)^2}\mathrm{d}z, \qquad y(\pm z_0) = r_0. \end{aligned} \tag{8.1.71}$$

The (re-scaled) Lagrange here is $L(y, y') = y(z)\sqrt{1+y'(z)^2}$, and its associated Euler-Lagrange equation is

$$\frac{\mathrm{d}}{\mathrm{d}z}\frac{\partial}{\partial y'}\left(y\sqrt{1+y'^2}\right) = \frac{\partial}{\partial y}\left(y\sqrt{1+y'^2}\right), \tag{8.1.72}$$

$$\frac{1}{\sqrt{1+y'^2}}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{y \cdot y'}{\sqrt{1+y'^2}}\right) = \frac{1}{\sqrt{1+y'^2}}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{2}\frac{1}{\sqrt{1+y'^2}}\frac{\mathrm{d}y^2}{\mathrm{d}z}\right) = 1, \tag{8.1.73}$$

$$\frac{\mathrm{d}^2 y^2}{\mathrm{d}\ell^2} = 2; \tag{8.1.74}$$

where we have defined the infinitesimal arc length $\mathrm{d}\ell$ via the relation

$$\mathrm{d}\ell = \sqrt{1+y'^2}\mathrm{d}z. \tag{8.1.75}$$

The general solution to $\mathrm{d}^2 y^2/\mathrm{d}\ell^2 = 2$ is

$$y(\ell) = \sqrt{C_0 + C_1\ell + \ell^2} \tag{8.1.76}$$

for constants C_0 and C_1. If the total arc length is ℓ_0, we parametrize $-\ell_0/2 \le \ell \le \ell_0/2$ (so that $\int_{-\ell_0/2}^{+\ell_0/2}\mathrm{d}\ell = \ell_0$), then $y(\pm\ell_0/2) = r_0$ and $\sqrt{C_0 \pm C_1\ell_0 + \ell_0^2} = r_0$. The solution to our 2 ring setup in terms of ℓ is therefore

$$y(-\ell_0/2 \le \ell \le \ell_0/2) = \sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2}. \tag{8.1.77}$$

The area may, in turn, be obtained as a function of the total arc length

$$A = 2\pi\int_{-\ell_0/2}^{\ell_0/2} y\mathrm{d}\ell \tag{8.1.78}$$

$$= \pi\left[\ell\sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2} + (r_0^2 - (\ell_0/2)^2)\ln\left[\ell + \sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2}\right]\right]_{\ell=-\ell_0/2}^{\ell=+\ell_0/2} \tag{8.1.79}$$

$$= \pi\left(r_0 \cdot \ell_0 + (r_0^2 - (\ell_0/2)^2)\ln\left[\frac{r_0 + \ell_0/2}{r_0 - \ell_0/2}\right]\right). \tag{8.1.80}$$

Next, we turn to solving for z. From eq. (8.1.75),

$$\frac{\mathrm{d}\ell}{\mathrm{d}z} = \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\mathrm{d}\ell}{\mathrm{d}z}\right)^2}, \tag{8.1.81}$$

$$\frac{\mathrm{d}z}{\mathrm{d}\ell} = \sqrt{\frac{r_0^2 - (\ell_0/2)^2}{r_0^2 - (\ell_0/2)^2 + \ell^2}}, \tag{8.1.82}$$

$$z(\ell) = \frac{1}{2}\sqrt{r_0^2 - (\ell_0/2)^2}\ln\left(\frac{\sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2} + \ell}{\sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2} - \ell}\right) + C_2. \tag{8.1.83}$$

For z to be real, we need

$$r_0 \ge \ell_0/2. \tag{8.1.84}$$

The constant C_2 is determined via the boundary condition $z(\pm\ell_0/2) = \pm z_0$.

$$\pm z_0 = \pm\frac{1}{2}\sqrt{r_0^2 - (\ell_0/2)^2}\ln\left(\frac{r_0 + \ell_0/2}{r_0 - \ell_0/2}\right) + C_2 \tag{8.1.85}$$

Thus, $C_2 = 0$ and we arrive at the following relation between the total arc length and the two parameters (r_0, z_0) of the setup.

$$z(-\ell_0/2 \leq \ell \leq +\ell_0/2) = \frac{1}{2}\sqrt{r_0^2 - (\ell_0/2)^2}\ln\left(\frac{\sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2} + \ell}{\sqrt{r_0^2 - (\ell_0/2)^2 + \ell^2} - \ell}\right), \tag{8.1.86}$$

$$z_0 = \sqrt{r_0^2 - (\ell_0/2)^2}\text{arctanh}\left(\frac{\ell_0}{2r_0}\right); \tag{8.1.87}$$

where we have recognized, whenever $0 \leq z \leq 1$, $(1/2)\ln[(1+z)/(1-z)] = \text{arctanh}(z)$.

If we set $\xi_0 \equiv z_0/r_0$ and $\lambda_0 \equiv \ell_0/(2r_0)$, eq. (8.1.87) is transformed into

$$\xi_0 = \sqrt{1 - \lambda_0^2}\cdot\text{arctanh}(\lambda_0). \tag{8.1.88}$$

For a solution to exist we see that

$$0 \leq \lambda_0 \leq 1. \tag{8.1.89}$$

The maximum of eq. (8.1.88) may be obtained numerically; it occurs at $\lambda_0 \approx 0.833556$; and the maximum itself is

$$\max\left(\sqrt{1 - \lambda_0^2}\cdot\text{arctanh}(\lambda_0)\right) \approx 0.662743 \geq \xi_0 = \frac{z_0}{r_0}. \tag{8.1.90}$$

Furthermore, whenever z_0/r_0 satisfies this bound, there are *two* solutions for λ_0 at a given ξ_0. To compare them, we first re-write eq. (8.1.80) as

$$A(\lambda_0) = 2\pi r_0^2\left(\lambda_0 + \left(1 - \lambda_0^2\right)\text{arctanh}(\lambda_0)\right). \tag{8.1.91}$$

Using $\partial_x \text{arctanh} x = 1/(1-x^2)$, this area function $A(\lambda_0)$ increases from 0 and maxes out when

$$\lambda_0\cdot\text{arctanh}(\lambda_0) = 1, \tag{8.1.92}$$

which is (as can be checked readily) the same condition for maximum $z(\lambda_0)$ – namely, $\lambda_0 \approx 0.833556$. This in turn implies

$$\max A = A(\lambda_0 = 0.833556\dots) \approx (1.19968\dots)(2\pi r_0^2). \tag{8.1.93}$$

After maxing out, $A(\lambda_0)$ then falls to 1 as $\lambda_0 \to 1$.

At this juncture, let us record the existence of the *discontinous* Goldschmidt solution: two flat circular discs of combined area $2(\pi r_0^2)$, one attached to each of the radius r_0 rings at $z = \pm z_0$. (Why is this a solution? Such solutions solve the homogeneous Laplace equation on the (x, y) plane – see Problem (9.75) and the discussion leading up to eq. (9.7.29).) Now, the constant $A = 2\pi r_0^2$ line intersects the $A(\lambda_0)$ graph at $\lambda_0 \approx 0.584376$ and at $\lambda_0 = 1$. The area of the 2D membrane $A(0.584376\dots \le \lambda_0 \le 1)$ is larger than the $2\pi r_0^2$ whereas $A(0 \le \lambda_0 \le 0.584376\dots)$ is smaller than $2\pi r_0^2$. Hence, the minimum area for $0 \le \lambda_0 \le 0.584376\dots$ is given by $A(\lambda_0)$; whereas for $0.584376\dots \le \lambda_0 \le 1$ it is given by the Goldschmidt solution.

Finally, if we re-write eq. (8.1.77) as

$$\ell = \operatorname{sgn}(\ell) \cdot \left| y^2 + (\ell_0/2)^2 - r_0^2 \right|^{\frac{1}{2}}, \tag{8.1.94}$$

this converts eq. (8.1.86) into

$$\tanh\left(\frac{z}{\sqrt{r_0^2 - (\ell_0/2)^2}}\right) = \frac{\operatorname{sgn}(\ell) \cdot |y^2 + (\ell_0/2)^2 - r_0^2|^{\frac{1}{2}}}{y}, \tag{8.1.95}$$

$$y(z) = \sqrt{r_0^2 - (\ell_0/2)^2} \cosh\left(\frac{z}{\sqrt{r_0^2 - (\ell_0/2)^2}}\right). \tag{8.1.96}$$

To summarize: Eq. (8.1.96) provides the curve $y(z)$ on the $(0, y, z)$ plane whose surface of revolution around the $z-$axis produces a minimal area surface. For a fixed separation $2z_0$ between the two rings at the two ends of the membrane, there are actually two solutions; it is the one with shorter total arc length ℓ_0 that has the minimal area, at least for $\ell_0 > 0$ up to $\ell_0/(2r_0) = \lambda_0 = 0.564376\dots$, where $A(\lambda_0) = 2\pi r_0^2$ and $z(\lambda_0 = 0.564376\dots)/r_0 = 0.527697\dots$. Beyond this cross-over value of λ_0, the minimal area surface is given by the 2 disjoint circular discs fixed on the rings, with area given by $2\pi r_0^2$.

Problem 8.8. Catenoid from 'Constant-of-Motion' Use the 'constant-of-motion' given in eq. (8.1.19) to directly integrate the $\mathrm{d}y/\mathrm{d}z$ equations arising from the Lagrangian $L = y\sqrt{1 + y'(z)^2}$. Upon invoking parity $(z \leftrightarrow -z)$ arguments, you should find, for constant χ, the solution

$$y(z) = \chi \cosh(z/\chi). \tag{8.1.97}$$

This route yields the shape of the minimal surface more quickly, though you still have to push the analysis further to understand why $\chi = \sqrt{r_0^2 - (\ell_0/2)^2}$; why there are two solutions for a given z_0; etc. $\square$

Second Variation and Normal Modes Is a given solution to the Euler-Lagrange equations (8.1.31) a (locally) minimum, maximum or inflection 'point'? To study this question we now compute the second order perturbations to the action in eq. (8.1.36).

$$\delta_2 S = \int_{\tau'}^{\tau} \left(\frac{1}{2} \frac{\partial^2 L}{\partial q^i \partial q^j} \delta q^i \delta q^j + \frac{1}{2} \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \frac{\mathrm{d}\delta q^i}{\mathrm{d}\lambda} \frac{\mathrm{d}\delta q^j}{\mathrm{d}\lambda} + \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} \delta q^i \frac{\mathrm{d}\delta q^j}{\mathrm{d}\lambda} \right) \mathrm{d}\lambda \tag{8.1.98}$$

$$= \int_{\tau'}^{\tau} \delta q^i \cdot \delta_2 L_{ij} \cdot \delta q^j \mathrm{d}\lambda. \tag{8.1.99}$$

In the second line, we have integrated-by-parts the derivative acting on the $\mathrm{d}\delta q^i/\mathrm{d}\lambda$ term in the middle term of the first line, using the conditions in eq. (8.1.35) to drop the boundary terms. The $\delta_2 L_{ij}$ is therefore an operator acting on δq^j:

$$\delta_2 L_{ij} \delta q^j \equiv \frac{1}{2} \frac{\partial^2 L}{\partial q^i \partial q^j} \delta q^j - \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \frac{\mathrm{d}}{\mathrm{d}\lambda} \delta q^j \right) + \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} \frac{\mathrm{d}}{\mathrm{d}\lambda} \delta q^j. \tag{8.1.100}$$

We may generalize the second order perturbation of the action to a matrix element of $\delta_2 L_{ij}$. Let the relevant vector space be that of all deviation trajectories $\{\delta\vec{q}(\tau' \leq \lambda \leq \tau)\}$ obeying the Dirichlet boundary conditions in eq. (8.1.35).

$$\left\langle \delta a^i \left| \delta_2 L_{ij} \right| \delta b^j \right\rangle \equiv \int_{\tau'}^{\tau} \delta a^i \cdot \delta_2 L_{ij} \cdot \delta b^j \mathrm{d}\lambda \tag{8.1.101}$$

The second order perturbed action itself now reads

$$\delta_2 S = \left\langle \delta q^i \left| \delta_2 L_{ij} \right| \delta q^j \right\rangle. \tag{8.1.102}$$

Let us examine its adjoint. Since everything here is real by assumption,

$$\left\langle \delta a^i \left| \delta_2 L_{ij} \right| \delta b^j \right\rangle = \overline{\left\langle \delta a^i \left| \delta_2 L_{ij} \right| \delta b^j \right\rangle} = \left\langle \delta b^j \left| (\delta_2 L_{ij})^\dagger \right| \delta a^i \right\rangle \tag{8.1.103}$$

$$= \int_{\tau'}^{\tau} \delta a^i \left\{ \frac{1}{2} \frac{\partial^2 L}{\partial q^i \partial q^j} \delta b^j - \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \frac{\mathrm{d}}{\mathrm{d}\lambda} \delta b^j \right) + \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} \frac{\mathrm{d}}{\mathrm{d}\lambda} \delta b^j \right\} \mathrm{d}\lambda \tag{8.1.104}$$

$$= \int_{\tau'}^{\tau} \delta b^j \left(\delta_2 L_{ij} \right)^\dagger \delta a^i \mathrm{d}\lambda \equiv \int_{\tau'}^{\tau} \delta b^j \delta_2 {L^\dagger}_{ji} \delta a^i \mathrm{d}\lambda; \tag{8.1.105}$$

where

$$\delta_2 {L^\dagger}_{ji} \delta a^i \equiv \frac{1}{2} \frac{\partial^2 L}{\partial q^i \partial q^j} \delta a^i - \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \frac{\mathrm{d}}{\mathrm{d}\lambda} \delta a^i \right) - \frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} \delta a^i \right). \tag{8.1.106}$$

Whenever $\partial^2 L/(\partial q^i \partial \dot{q}^j) = 0$ If the $\partial^2 L/(\partial q^i \partial \dot{q}^j)$ term were in fact absent, we see that $\delta_2 L_{ij}$ is in fact Hermitian. We may then proceed to expand δq^i in its complete set of eigenvectors $\{\widehat{e}^i_\xi\}$, which obeys

$$\delta_2 L_{ij}(\lambda) \widehat{e}^j_\xi(\lambda) = \xi \cdot \widehat{e}^i_\xi(\lambda), \tag{8.1.107}$$

$$\sum_\xi \widehat{e}^i_\xi(\lambda')^* \widehat{e}^j_\xi(\lambda) = \sum_\xi \widehat{e}^i_\xi(\lambda') \widehat{e}^j_\xi(\lambda)^* = \delta^{ij} \delta(\lambda' - \lambda), \tag{8.1.108}$$

$$\int_{\tau'}^{\tau} \delta_{ij} \widehat{e}^i_{\xi'}(\lambda)^* \widehat{e}^j_\xi(\lambda) \mathrm{d}\lambda = \delta^{\xi'}{}_{\xi}. \tag{8.1.109}$$

The expansion itself proceeds as

$$\delta q^i(\lambda) = \sum_\xi \widehat{e}^i_\xi(\lambda) \int_{\tau'}^{\tau} \mathrm{d}\lambda' \left(\widehat{e}^j_\xi(\lambda')^* \delta q^j(\lambda') \right) \tag{8.1.110}$$

$$\equiv \sum_\xi \widehat{e}^i_\xi(\lambda) \left\langle \xi \right| \delta q \rangle \tag{8.1.111}$$

$$\delta_2 L_{ij}(\lambda) \delta q^j(\lambda) = \sum_\xi \xi \cdot \widehat{e}^i_\xi(\lambda) \left\langle \xi \right| \delta q \rangle; \tag{8.1.112}$$

leading us to deduce,

$$\delta_2 S = \int_{\tau'}^{\tau} \delta q^i \delta_2 L_{ij} \delta q^j \mathrm{d}\lambda \tag{8.1.113}$$

$$= \sum_{\xi',\xi} \xi \int_{\tau'}^{\tau} \widehat{e}^i_{\xi'}(\lambda)^* \widehat{e}^i_{\xi}(\lambda) \mathrm{d}\lambda \cdot \overline{\langle \xi' | \delta q \rangle} \langle \xi | \delta q \rangle \tag{8.1.114}$$

$$= \sum_{\xi} \xi \cdot | \langle \xi | \delta q \rangle |^2. \tag{8.1.115}$$

Therefore, if all eigenvalues $\{\xi\}$ of $\delta_2 L_{ij}$ are positive, we have a (local) minimum. And, if all eigenvalues of $\delta_2 L_{ij}$ are negative, we have a (local) maximum.

As a simple example, we consider the case of the 2D straight line, obtained from the Lagrangian $L = \sqrt{1 + y'(x)^2}$. Because it has dependence only on y' and not on y, we may readily compute

$$\delta_2 L \cdot \delta y = -\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{\partial^2}{\partial y'^2} \left(\sqrt{1 + y'^2} \right) \frac{\mathrm{d}}{\mathrm{d}x} \delta y(x) \right) \tag{8.1.116}$$

$$= -\frac{1}{2(1 + y'^2)^{3/2}} \frac{\mathrm{d}^2}{\mathrm{d}x^2} \delta y(x). \tag{8.1.117}$$

For a straight line, $y'(x)$ and hence the factor multiplying $(\mathrm{d}/\mathrm{d}x)^2$ is constant. If the end points are (x_0, y_0) and (x_1, y_1), we require

$$\delta y(x_0) = 0 = \delta y(x_1). \tag{8.1.118}$$

Then, the (un-normalized) eigenfunctions are

$$\delta y_n(x) = \sin \left(\frac{n\pi}{x_1 - x_0} (x - x_0) \right). \tag{8.1.119}$$

for $n = 1, 2, 3, \dots$; and the corresponding eigenvalues can be read off the equation

$$\delta_2 L \cdot \delta y_n = \frac{1}{2(1 + y'^2)^{3/2}} \left(\frac{n\pi}{x_1 - x_0} \right)^2 \delta y_n. \tag{8.1.120}$$

As expected, all of them are positive because the straight line is the minimum length line. On the other hand, in this example, one may simply expand the action about the solution up second order in the deviation δy without integration-by-parts. (Remember – see eq. '*Observation*' right after eq. (8.1.39) – the first order perturbation off the solution always vanishes, as long as δy is zero at the end points.) One would find

$$\delta_2 S = \frac{1}{2(1 + y'^2)^{3/2}} \int_{x_0}^{x_1} (\delta y'(x))^2 \mathrm{d}x, \tag{8.1.121}$$

a manifestly positive quantity; without the need to compute eigenvalues of differential operators.

Problem 8.9. Simple Harmonic Oscillator To study a more nuanced example, we turn to the simple harmonic oscillator (SHO), whose Lagrangian is given by

$$L \left(\vec{q}, \dot{\vec{q}} \right) \equiv \frac{1}{2} \dot{\vec{q}}(\lambda)^2 - \frac{\Omega^2}{2} \vec{q}(\lambda)^2. \tag{8.1.122}$$

Show that extremizing $S \equiv \int_{\tau'}^{\tau} L \mathrm{d}\lambda$ leads to the SHO equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} + \Omega^2\right) \vec{q}(\lambda) = 0. \tag{8.1.123}$$

Explain why $\delta_2 L_{ij}$ in eq. (8.1.100) is Hermitian in this case, and demonstrate it is given by

$$\delta_2 L_{ij} \cdot \delta q^j = -\frac{\delta^{ij}}{2} \left(\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} + \Omega^2\right) \delta q^j(\lambda). \tag{8.1.124}$$

For fixed initial and final times τ' and τ, explain why the eigenvalues are

$$\lambda_n = \frac{\Omega^2}{2} \left(\left(\frac{n\pi}{\Omega(\tau - \tau')}\right)^2 - 1 \right), \tag{8.1.125}$$

where $n = 1, 2, 3, \ldots$. Remember, for the SHO, π/Ω is the half period. Hence, whenever the elapsed time $\tau - \tau'$ is *shorter* than the half-period, then $n\pi/(\Omega(\tau - \tau')) > 1$ and all eigenvalues are positive – the SHO solution in eq. (7.8.67) (with $x_h \to \vec{q}$, $t \to \lambda$, $x_{\mathrm{i}} \to \vec{x}'$ and $x_{\mathrm{f}} \to \vec{x}$) in fact minimizes its action. However, once $\tau - \tau'$ is greater than π/Ω, there will necessarily be $\lfloor \pi/(\Omega(\tau - \tau')) \rfloor \geq 1$ negative eigenvalues: the solution becomes an inflection point in the sense that $\delta_2 S$ can be either negative or positive depending on 'how much' of $\delta\vec{q}$ lies on the negative eigenvalue subspace versus the positive eigenvalue subspace. □

Caustics, Conjugate Points, Eigenvalue Sign Flips Earlier, in the discussion enveloping eq. (5.11.226), we recognized the formation of caustics in the simple harmonic oscillator's (SHO's) motion: starting from $(\tau', \vec{x}_0)$, there are an infinity of solutions that will converge onto $(\tau, (-)^n \vec{x}_0)$ for $\tau = \tau' + n\pi/\omega$; i.e., whenever a half-integer multiple of the half-period π/ω has elapsed. Here in eq. (8.1.125), we see that the differential operator associated with the second order variation of its action about the solution yields an eigenvalue that flips sign whenever the SHO itself passes through a caustic at half-periods. For the first half-period, $0 < \tau - \tau' < \pi/\omega$, every eigenvalue $\lambda_n > 0$; but for the second half-period $\pi/\omega < \tau - \tau' < 2\pi/\omega$, the first eigenvalue flips sign $\lambda_1 < 0$ while the rest are still positive $\lambda_{n \geq 2} > 0$. At the k-th half-period $k\pi/\omega < \tau - \tau' < (k+1)\pi/\omega$, the first k eigenvalues have flipped over to become negative; while the rest remain positive. That an eigenvalue associated with the second order variation of the action flips sign whenever the trajectory passes through a caustic is a generic phenomenon, and is known as the Morse Index theorem.

Problem 8.10. Damped Simple Harmonic Oscillator (SHO): Version 1 Extremize the following action, for $a(t) \neq 0$.

$$S \equiv \int_{\tau'}^{\tau} \frac{a(t)^3}{2} \left(\dot{\vec{q}}(t)^2 - \Omega^2 \vec{q}(t)^2\right) \mathrm{d}t \tag{8.1.126}$$

and show that it leads to 'time-dependent' damped SHO, as long as $a(t) \neq 0$:

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + 3\frac{\dot{a}(t)}{a(t)}\frac{\mathrm{d}}{\mathrm{d}t} + \Omega^2\right) \vec{q}(t) = 0. \tag{8.1.127}$$

An analogous scenario arises in cosmological applications. What does $a(t)$ need to be in order to recover time-independent friction? That is, solve $3(\dot{a}/a) = 2\gamma$, for constant $\gamma > 0$. Hint: $a(t) = a_0 \exp((2/3)\gamma \cdot t)$. □

Problem 8.11. General Variation of Action Show that, under a simultaneous variation of both the path

$$\vec{q} \to \vec{q} + \delta\vec{q} \tag{8.1.128}$$

as well as the initial and final times

$$\tau \to \tau + \mathrm{d}\tau, \tag{8.1.129}$$
$$\tau' \to \tau' + \mathrm{d}\tau'; \tag{8.1.130}$$

the general action $S \equiv \int_{\tau'}^{\tau} L(s, \vec{q}, \dot{\vec{q}})\mathrm{d}s$ transforms as

$$S \to S + \delta_1 S \tag{8.1.131}$$

$$\delta_1 S = L(\tau, \vec{q}(\tau), \dot{\vec{q}}(\tau))\mathrm{d}\tau - L(\tau', \vec{q}(\tau'), \dot{\vec{q}}(\tau'))\mathrm{d}\tau' + \left(\frac{\partial L}{\partial \dot{q}^i}\delta q^i\right)_{s=\tau} - \left(\frac{\partial L}{\partial \dot{q}^i}\delta q^i\right)_{s=\tau'} + \int_{\tau'}^{\tau} \delta q^i(s)\left(\frac{\partial L}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}s}\frac{\partial L}{\partial \dot{q}^i}\right)\mathrm{d}s. \tag{8.1.132}$$

If $\vec{q}$ actually satisfies the Euler-Lagrange equations, then this first order variation of S may be viewed as

$$\delta_1 S = \partial_\tau S \mathrm{d}\tau + \partial_{\tau'} S \mathrm{d}\tau' + \frac{\partial S}{\partial q^i(\tau)}\delta q^i(\tau) + \frac{\partial S}{\partial q^i(\tau')}\delta q^i(\tau'). \tag{8.1.133}$$

The derivatives with respect to $q^i(\tau)$ and $q^i(\tau')$ are to be viewed as derivatives with respect to, respectively, the final and initial positions. □

8.2 Dealing With Constraints

Multi-Variable Calculus Review Suppose we wish to find the extremum of the function $f(\vec{x})$, where $\vec{x}$ here is simply a shorthand for $D \geq 1$ variables such that $\partial_i f(\vec{x})$ denotes the derivative with respect to ith variable x^i. Suppose we also wish to simultaneously satisfy $N \leq D - 1$ constraints of the following form

$$g_{\mathrm{I}}(\vec{x}) = 0, \qquad \mathrm{I} \in \{1, 2, \dots, N\}. \tag{8.2.1}$$

(For example, $g(\vec{x}) = \vec{x}^2 - R^2$ would describe a sphere of radius R.) Then the desired extremum $\{\vec{x}_\ell | \ell = 1, 2, \dots\}$ of $f(\vec{x})$ subject to the constraints in eq. (8.2.1) is provided by the following. First introduce the function

$$F(\vec{x}) = f(\vec{x}) - \lambda^{\mathrm{I}} g_{\mathrm{I}}(\vec{x}), \tag{8.2.2}$$

where the *Lagrange multipliers* $\{\lambda^{\mathrm{I}} | \mathrm{I} = 1, 2, \dots, N\}$ are constants; and an implicit sum over I is implied. (Notice, when $\vec{x}$ satisfies eq. (8.2.1), $F(\vec{x}) = f(\vec{x})$.) Then, the extremum must simultaneously satisfy the N equations encoded by (8.2.1) and render the D first derivatives of F zero:

$$g_{\mathrm{I}}(\vec{x}_\ell) = 0 = \partial_i F(\vec{x}_\ell). \tag{8.2.3}$$

The $N + D$ equations in eq. (8.2.3) will yield solutions for not only the turning points $\{\vec{x}_\ell | \ell = 1, 2, \dots\}$ – there could be more than one – but also the N Lagrange multipliers $\{\lambda^{\mathrm{I}}\}$.

Note that $N < D$, otherwise eq. (8.2.5) would form an over-determined system: for e.g., in 3D space, one constraint defines a 2D surface; two constraints would defined a 1D line (intersection of two 2D surfaces); and three constraints would yield *at most* a single point; beyond that, there will generically be no solution unless some of the constraints are actually degenerate.

Variational Principle with 'Local' Constraints A similar strategy applies to the calculus of variation in the presence of constraints. Suppose the equations implied by some Lagrangian $L(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda))$ – together with the usual boundary conditions

$$\vec{q}(\tau) = \vec{y} \qquad \text{and} \qquad \vec{q}(\tau') = \vec{y}' \tag{8.2.4}$$

– are additionally subject to the N 'local-in-λ' constraint(s)

$$G_{\mathrm{I}}\left(\lambda, \vec{q}, \dot{\vec{q}}\right) = 0, \qquad \mathrm{I} = 1, 2, 3, \dots, N. \tag{8.2.5}$$

Then, like its multi-variable counterpart, we introduce N $\lambda-$*dependent* Lagrange multipliers $\{\Lambda^{\mathrm{I}}(\lambda) | \mathrm{I} = 1, 2, 3, \dots, N\}$ and proceed to extremize the action

$$S = \int_{\tau'}^{\tau} \left(L\left(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda)\right) - \Lambda^{\mathrm{I}}(\lambda) G_{\mathrm{I}}\left(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda)\right)\right) \mathrm{d}\lambda. \tag{8.2.6}$$

That is, we simply need to solve

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_{\mathrm{M}}}{\partial \dot{q}^i} = \frac{\partial L_{\mathrm{M}}}{\partial q^i}, \tag{8.2.7}$$

$$L_{\mathrm{M}} \equiv L - \Lambda^{\mathrm{I}} G_{\mathrm{I}}; \tag{8.2.8}$$

for both $\{\vec{q}(\lambda)\}$ – which has to obey eq. (8.1.35) as well as the constraint(s) (8.2.5) – and the N Lagrange multipliers $\{\Lambda_{\mathrm{I}}(\lambda)\}$. There are altogether $N + D$ equations for the N λs and D components of $\vec{q}$.

Variational Principle with Integral Constraints Instead of the 'local' constraints eq. (8.2.5), suppose we now study the equations implied by some Lagrangian $L(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda))$, subject to the same boundary conditions $\vec{q}(\tau) = \vec{y}$ and $\vec{q}(\tau') = \vec{y}'$ in eq. (8.2.4) – but now apply the following N integral constraints

$$\int_{\tau'}^{\tau} G_{\mathrm{I}}\left(\lambda, \vec{q}, \dot{\vec{q}}\right) \mathrm{d}\lambda = \text{constant}, \qquad \mathrm{I} = 1, 2, 3, \dots, N. \tag{8.2.9}$$

We introduce N $\lambda-$*independent* Lagrange multipliers $\{\Lambda^{\mathrm{I}} | \mathrm{I} = 1, 2, 3, \dots, N\}$ and proceed to extremize the action

$$S = \int_{\tau'}^{\tau} \left(L\left(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda)\right) - \Lambda^{\mathrm{I}} \cdot G_{\mathrm{I}}\left(\lambda, \vec{q}(\lambda), \dot{\vec{q}}(\lambda)\right)\right) \mathrm{d}\lambda. \tag{8.2.10}$$

That is, we simply need to solve

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_{\mathrm{M}}}{\partial \dot{q}^i} = \frac{\partial L_{\mathrm{M}}}{\partial q^i}, \tag{8.2.11}$$

$$L_{\mathrm{M}} \equiv L - \Lambda^{\mathrm{I}} G_{\mathrm{I}}; \tag{8.2.12}$$

for both $\{\vec{q}(\lambda)\}$ – which has to obey eq. (8.1.35) as well as the constraint(s) (8.2.9) – and the N Lagrange multipliers $\{\Lambda_{\mathrm{I}}\}$. There are altogether $N + D$ equations for the N λs and D components of $\vec{q}$.

Example: Maximum Area for Fixed Perimeter If you are given an infinitesimally thin string that forms a closed loop of fixed length ℓ, what is the 2D shape that yields the largest area? Observe, if there were no fixed-length constraint, the largest area is infinite; while for a fixed length ℓ, the smallest area is zero, since the string loop can readily be squashed to a line.

If we use Cartesian coordinates $\vec{r} = (x(\theta), y(\theta), 0)$ on the $(1,2)-$plane, the area swept out by the trajectory is given by

$$A = \frac{1}{2} \int_0^{2\pi} \mathrm{d}\theta \left(\vec{r}(\theta) \times \frac{\mathrm{d}\vec{r}(\theta)}{\mathrm{d}\theta} \right) \cdot \widehat{e}_3 \tag{8.2.13}$$

$$= \frac{1}{2} \int_0^{2\pi} \mathrm{d}\theta \left(x \cdot y' - x' \cdot y \right). \tag{8.2.14}$$

(By the right hand rule,

$$\frac{1}{2} \left(\vec{r}(\theta) \times \frac{\mathrm{d}\vec{r}(\theta)}{\mathrm{d}\theta} \right) \cdot \widehat{e}_3 \mathrm{d}\theta = \frac{1}{2} |\vec{r}| \cdot |\mathrm{d}\vec{r}| \sin\varphi \tag{8.2.15}$$

is a positive infinitesimal area; in fact, it is the area of the triangle, whose base is $|\vec{r}|$ and perpendicular height is $|\mathrm{d}\vec{r}| \sin\varphi$, where φ is the angle between $\mathrm{d}\vec{r}(\theta)/\mathrm{d}\theta$ and $\vec{r}$.)

The length swept out by this same trajectory $\vec{r}(\theta)$ is, in turn,

$$\ell = \int_0^{2\pi} \sqrt{x'^2 + y'^2} \mathrm{d}\theta. \tag{8.2.16}$$

Hence, we form the Lagrangian

$$L = \frac{1}{2}(x \cdot y' - x' \cdot y) - \Lambda \cdot \sqrt{x'^2 + y'^2}. \tag{8.2.17}$$

The Euler-Lagrange equations are

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial L}{\partial x'} = \frac{\partial L}{\partial x} \tag{8.2.18}$$

$$\Lambda \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{x'}{\sqrt{x'^2 + y'^2}} = -y'; \tag{8.2.19}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial L}{\partial y'} = \frac{\partial L}{\partial y} \tag{8.2.20}$$

$$\Lambda \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{y'}{\sqrt{x'^2 + y'^2}} = x'. \tag{8.2.21}$$

Integrating them leads us to the first order equations

$$\Lambda \frac{x'}{\sqrt{x'^2 + y'^2}} = y_0 - y \qquad \text{and} \qquad \Lambda \frac{y'}{\sqrt{x'^2 + y'^2}} = x - x_0; \tag{8.2.22}$$

for constants (x_0, y_0). Squaring both sides and adding the two equations immediately hands us the equation of a circle centered at (x_0, y_0) with radius Λ.

$$(x - x_0)^2 + (y - y_0)^2 = \Lambda^2 \tag{8.2.23}$$

Now, since the perimeter is fixed at $\ell = 2\pi\Lambda$, that means we may re-express

$$\Lambda = \ell/(2\pi). \tag{8.2.24}$$

The parametric solution is therefore

$$(x, y) = (x_0, y_0) + \frac{\ell}{2\pi}(\cos\theta, \sin\theta). \tag{8.2.25}$$

Problem 8.12. Isoperimetric Problem: Maximum Area? Did we really obtain a maximum area? Perturb the solution in eq. (8.2.25), but using spherical coordinates, namely

$$(x, y) = (x_0, y_0) + \left(\frac{\ell}{2\pi} + \delta r(\theta)\right)(\cos\theta, \sin\theta); \tag{8.2.26}$$

and proceed to insert it into

$$S = \int_0^{2\pi} \left(\frac{1}{2}(x \cdot y' - x' \cdot y) - \frac{\ell}{2\pi}\left(\sqrt{x'^2 + y'^2} - \frac{\ell}{2\pi}\right)\right) \mathrm{d}\theta; \tag{8.2.27}$$

where we have added to the Lagrangian in eq. (8.2.17) the constant $\Lambda \cdot \ell/(2\pi)$ so that the result for S is the actual area. Show that, up to second order in the deviation vector $\delta\vec{r} \equiv (\delta x, \delta y, 0)$,

$$S = \frac{\ell^2}{4\pi} - \frac{1}{2}\int_0^{2\pi} \left\{\delta r'(\theta)^2 - \delta r(\theta)^2\right\} \mathrm{d}\theta + \mathcal{O}\left(\delta\vec{r}^3\right) \tag{8.2.28}$$

Explain why the second order in δr term on the right hand side is non-negative. Hint: You should find the eigenvalues corresponding to $\{\xi\}$ in eq. (8.1.115) to be $\xi_n = -(1/2)(n^2 - 1)$, for $n = 1, 2, 3, \ldots$, and hence strictly non-positive. □

Problem 8.13. Fixed Area, Minimum Perimeter For a fixed 2D area A_0 whose boundary is a simply connected curve, prove that the boundary with the shortest perimeter is that of a circle. *Bonus*: Can you argue why this is a minimum perimeter? □

Problem 8.14. Reciprocity Prove the following reciprocity relation. Extremizing the action

$$A \equiv \int_{\tau'}^{\tau} L_A\left(\lambda, \vec{q}, \dot{\vec{q}}\right) \mathrm{d}\lambda \tag{8.2.29}$$

subject to the requirement that the integral

$$B \equiv \int_{\tau'}^{\tau} L_B\left(\lambda, \vec{q}, \dot{\vec{q}}\right) \mathrm{d}\lambda \tag{8.2.30}$$

be held constant is equivalent – up to a redefinition of the relevant Lagrange multiplier – to extremizing the latter integral in eq. (8.2.30) while holding the former one in eq. (8.2.29) fixed. $\square$

Example: Maximum Entropy with Fixed Variance If $P(x)\mathrm{d}x$ is the probability of some outcome to lie within x and $x + \mathrm{d}x$, then the total probability being one means

$$\int_{\mathbb{R}} P(x)\mathrm{d}x = 1. \tag{8.2.31}$$

(We are assuming x runs over the entire real line.) In statistical physics, the *entropy* of P itself is defined as $\int_{\mathbb{R}}(-P \ln P)\mathrm{d}x$. Let us ask:

> What is the P that maximizes entropy but yields a fixed variance (aka 2−point function)?

This tells us we should examine the action

$$S \equiv \int_{\mathbb{R}} \mathrm{d}x \left(-P(x)\ln P(x) - \Lambda^1 P(x) - \Lambda^2 \cdot P(x) \cdot x^2\right). \tag{8.2.32}$$

The constraints are eq. (8.2.31) and

$$\int_{\mathbb{R}} \mathrm{d}x \left(P(x) \cdot x^2\right) = \text{constant}; \tag{8.2.33}$$

while the Euler-Lagrange equations are

$$0 = \frac{\partial}{\partial P}\left(-P(x)\ln P(x) - \Lambda^1 \cdot P(x) - \Lambda^2 \cdot P(x) \cdot x^2\right) \tag{8.2.34}$$

$$\ln P(x) = -1 - \Lambda^1 - \Lambda^2 \cdot x^2 \tag{8.2.35}$$

$$P(x) = e^{-\Lambda^2 \cdot x^2 - \Lambda^1 - 1}. \tag{8.2.36}$$

[70]We can already see $P(x)$ must be a Gaussian, with Λ^2 (which must be positive – why?) related to its variance and Λ^1 to the overall normalization to ensure total probability is unity.

$$1 = \int_{\mathbb{R}} P(x)\mathrm{d}x = \int_{\mathbb{R}} e^{-\Lambda^2 \cdot x^2 - \Lambda^1 - 1}\mathrm{d}x = \sqrt{\frac{\pi}{\Lambda^2}}e^{-\Lambda^1 - 1} \tag{8.2.37}$$

Hence, $P(x) = \sqrt{\Lambda^2/\pi}\exp(-\Lambda^2 x^2)$ and if we define the variance as

$$\sigma^2 \equiv \int_{\mathbb{R}} \mathrm{d}x \left(P(x) \cdot x^2\right); \tag{8.2.38}$$

[70]Notice, if this were a problem without constraints, i.e., if $\Lambda^1 = 0 = \Lambda^2$, the solution $\ln P = -1$ would not make much sense. In general, extremizing an action $\int L(\lambda, q)\mathrm{d}\lambda$ that does not involve derivatives of q merely leads to the conclusion that L itself does not depend on q. Can you see why?

then a direct integration hands us

$$\sigma^2 = \sqrt{\frac{\Lambda^2}{\pi}} \int_{\mathbb{R}} e^{-\Lambda^2 x^2} \cdot x^2 \mathrm{d}x = \frac{1}{2\Lambda^2}. \tag{8.2.39}$$

Gathering our result, and assuming $\sigma > 0$,

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right). \tag{8.2.40}$$

Problem 8.15. Maximum? Show that the Gaussian is indeed the maximum entropy probability distribution subject to the constraints (8.2.31) and (8.2.33). Hint: Consider perturbing $P(x) \to P(x) + \delta P(x)$. You should find, for $P(x) = (\sqrt{2\pi}\sigma)^{-1} \exp(-(1/2)(x/\sigma)^2)$,

$$S[P + \delta P] = 1 - \sigma\sqrt{\frac{\pi}{2}} \int_{\mathbb{R}} \left(e^{\frac{1}{2}(x/\sigma)^2} \delta P(x)^2\right) \mathrm{d}x + \mathcal{O}\left(\delta P^3\right). \tag{8.2.41}$$

Notice the second term on the right hand side is negative. □

Problem 8.16. Fixed Mean and Variance Solve for the maximum entropy probability distribution with both its one point (aka mean) and two point functions fixed. That is, re-do the above analysis with the additional constraint

$$\int_{\mathbb{R}} \mathrm{d}x \left(P(x) \cdot x\right) = \text{constant}. \tag{8.2.42}$$

Hint: You should find

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x - x_0}{\sigma}\right)^2\right), \tag{8.2.43}$$

where x_0 is the one point function. □

8.3 Lagrangian Dynamics in Classical Physics

As already alluded to, one of the key reasons why the variational or (stationary) action principle – and, hence, the calculus of variation – is so important, is because *all* of fundamental physics (electroweak, strong and gravitational interactions) may be succinctly described by one. Furthermore, Richard Feynman's path integral extends the applicability of the action from the classical to the quantum domain. Of course, much of non-relativistic physics may also be done using it too.

Non-Relativistic Particle Mechanics Let us begin with the most common case occurring within classical mechanics. In a flat space with Cartesian coordinates $\{\vec{x}\}$, choose the Lagrangian to be the difference between kinetic T and potential energy V:

$$L\left(\vec{x}(t), \dot{\vec{x}}(t)\right) \equiv T - V \equiv \frac{m}{2}\dot{\vec{x}}^2 - V(\vec{x}). \tag{8.3.1}$$

A short calculation indicates

$$\frac{\partial L}{\partial \vec{x}} = -\vec{\nabla}_{\vec{x}} V, \tag{8.3.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\vec{x}}} \equiv \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = m\ddot{\vec{x}}. \tag{8.3.3}$$

Euler-Lagrange in eq. (8.1.31) then tells us, mass times acceleration is the negative gradient of the potential energy; namely, Newton's second law for conservative forces is captured by the statement $L = T - V$:

$$m\ddot{x}(t) = -\vec{\nabla} V(\vec{x}) \equiv \text{Force}. \tag{8.3.4}$$

Noether: Continuous Symmetries and Conserved Quantities The Lagrangian formulation of dynamics allowed Emmy Noether to elucidate the connection between continuous symmetries and their corresponding conserved quantities. This important relationship continues to hold even in quantum mechanics and (quantum) field theory. As a start, let us examine Noether's theorem for the system encoded within the following Lagrangian of N particles $\{\vec{x}_{\mathrm{I}} | \mathrm{I} = 1, 2, \dots, N\}$ expressed in Cartesian coordinates:

$$L = \sum_{\mathrm{I}=1}^{N} \frac{M_{\mathrm{I}}}{2}\dot{\vec{x}}_{\mathrm{I}}^2 - \frac{1}{2}\sum_{\mathrm{J}\neq\mathrm{I}} V\left(|\vec{x}_{\mathrm{I}} - \vec{x}_{\mathrm{J}}|\right). \tag{8.3.5}$$

Time Translation Symmetry Apart from the time-dependence of the trajectories themselves $\{\vec{x}_{\mathrm{I}}\}$, notice the Lagrangian itself does not depend on time explicitly. That is, upon replacing $t \to t + t_0$ for arbitrary t_0, the L retains the *same form* – we dub this 'time-translation symmetry'. Let us examine its consequence for infinitesimal time displacements. Upon $t \to t + \mathrm{d}t$,

$$\vec{x}_{\mathrm{I}}(t) \to \vec{x}_{\mathrm{I}}(t) + \delta\vec{x}_{\mathrm{I}}, \tag{8.3.6}$$

$$\delta\vec{x}_{\mathrm{I}}(t) = \dot{\vec{x}}_{\mathrm{I}}(t)\mathrm{d}t. \tag{8.3.7}$$

Since the L itself is $t-$independent, it would therefore transform as[71]

$$L\left(\{\vec{x}_{\mathrm{I}}, \dot{\vec{x}}_{\mathrm{I}}\}\right) \to L\left(\{\vec{x}_{\mathrm{I}}, \dot{\vec{x}}_{\mathrm{I}}\}\right) + \frac{\partial L}{\partial x^i_{\mathrm{J}}}\delta x^i_{\mathrm{J}} + \frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}}\frac{\mathrm{d}}{\mathrm{d}t}\delta x^i_{\mathrm{J}} + \mathcal{O}\left(\delta x^2\right) \tag{8.3.8}$$

$$= L\left(\{\dot{\vec{x}}_{\mathrm{I}}, \dot{\vec{x}}_{\mathrm{I}}\}\right) + \left(\frac{\partial L}{\partial x^i_{\mathrm{J}}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}}\right)\delta x^i_{\mathrm{J}} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}}\delta x^i_{\mathrm{J}}\right) + \mathcal{O}\left(\delta x^2\right); \tag{8.3.9}$$

where there is an implicit sum over both the particle label J and its spatial component index i. At this point, we have actually not yet employed eq. (8.3.7); but if now proceed to do so, together with the Euler-Lagrange equations – which will eliminate the middle group of terms in eq. (8.3.9) – we will obtain:

$$L\left(\{\dot{\vec{x}}_{\mathrm{I}}, \dot{\vec{x}}_{\mathrm{I}}\}\right) \to L\left(\{\dot{\vec{x}}_{\mathrm{I}}, \dot{\vec{x}}_{\mathrm{I}}\}\right) + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}}\dot{x}^i_{\mathrm{J}}\right)\mathrm{d}t + \mathcal{O}\left(\delta x^2\right). \tag{8.3.10}$$

[71] There is actually no need to re-do the calculation below; simply read it off eq. (8.1.38), where we were perturbing the L by perturbing the trajectories, with the identifications $\vec{q} \leftrightarrow \vec{x}_{\mathrm{I}}$ and $\delta\vec{q} = \delta\vec{x}_{\mathrm{I}}$.

On the other hand, L itself may simply be treated as a function of time $L(t) \equiv L(\{\vec{x}_{\rm I}(t), \dot{\vec{x}}_{\rm I}(t)\})$, due to its *implicit* dependence through the trajectories $\{\vec{x}_{\rm I}(t)\}$. Under $t \to t + \mathrm{d}t$, it must transform as

$$L(t) \to L(t+\mathrm{d}t) = L(t) + \mathrm{d}t \cdot \frac{\mathrm{d}L(t)}{\mathrm{d}t} + \mathcal{O}\left(\mathrm{d}t^2\right). \tag{8.3.11}$$

Equating the right hand sides of (8.3.10) and (8.3.11), and extracting the coefficient of the $\mathrm{d}t$ terms, we arrive at the following total time derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^i_{\rm J}}\dot{x}^i_{\rm J} - L\right) = 0. \tag{8.3.12}$$

Below, you will verify that the quantity inside the parenthesis, $(\partial L/\partial \dot{x}^i_{\rm J})\dot{x}^i_{\rm J} - L$, is the total energy of the system.

> Time-translation symmetry – i.e., the absence of *explicit* time-dependence of L – implies total energy is conserved.

It is worth reiterating, this result is not true for any old trajectory, but only the 'physical' ones; namely, those that actually satisfy the Euler-Lagrange equations.

Problem 8.17. Total Energy Verify that the conserved quantity inside the parenthesis of eq. (8.3.12) constructed from the L of eq. (8.3.5) is in fact the total energy E:

$$\frac{\partial L}{\partial \dot{x}^i_{\rm J}}\dot{x}^i_{\rm J} - L = \sum_{\rm I=1}^{N}\left(\frac{M_{\rm I}}{2}\dot{\vec{x}}^2_{\rm I} + \frac{1}{2}\sum_{\rm J \neq I} V\left(|\vec{x}_{\rm I} - \vec{x}_{\rm J}|\right)\right) \equiv E. \tag{8.3.13}$$

Total kinetic plus potential energy is a constant because L does not depend explicitly on t.

Explicit Time Dependence Suppose L does depend on time explicitly. Show that eq. (8.3.12) would now read instead

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}^i_{\rm J}}\dot{x}^i_{\rm J} - L\right) = -\left(\frac{\partial L}{\partial t}\right)_{\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\}}; \tag{8.3.14}$$

and, hence, is no longer a "total time derivative is zero" statement. □

Space-Translation Symmetry Next, we observe that the L in eq. (8.3.5) remains the same under a simultaneous replacement

$$\vec{x}_{\rm I} \to \vec{x}_{\rm I} + \vec{a}; \tag{8.3.15}$$

for all $\mathrm{I} = 1, \dots, N$ and for constant but arbitrary displacement $\vec{a}$. The interpretation is that the underlying space that these particles inhabits is *homogeneous* – it does not matter whether the particles are interacting in some region $\mathcal{D}$ or in some other $\mathcal{D}'$ displaced by $\vec{a}$ relative to $\mathcal{D}$; the particles' mutual interactions remain identical.

A similar Taylor expansion calculation would yield eq. (8.3.9), but with $\delta x^i_{\rm I} = a^i$ for all the particles. Invoking the Euler-Lagrange equations,

$$L \to L + \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{\rm J}\frac{\partial L}{\partial \dot{x}^i_{\rm J}}a^i\right) + \mathcal{O}(a^2). \tag{8.3.16}$$

But, recall L actually remains the same under eq. (8.3.15). Therefore all the order $\vec{a}$ and higher terms would necessarily have to vanish. And since $\vec{a}$ was arbitrary, we arrive at:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{\mathrm{J}=1}^{N} \frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}}\right) = 0. \tag{8.3.17}$$

Below, you will verify that the quantity inside the parenthesis, $\sum_{\mathrm{J}} \partial L/\partial \dot{x}^i_{\mathrm{J}}$, is the total momentum of the system.

> Space-translation symmetry – i.e., the invariance of L under a constant spatial displacement of all N bodies – implies total momentum is conserved.

Again, this result is not true for any old trajectory, but only the 'physical' ones – those that actually satisfy the Euler-Lagrange equations.

Problem 8.18. Total Momentum Verify that the conserved quantity inside the parenthesis of eq. (8.3.17) constructed from the L of eq. (8.3.5) is in fact the total momentum:

$$\sum_{\mathrm{J}=1}^{N} \frac{\partial L}{\partial \dot{x}^i_{\mathrm{J}}} = \sum_{\mathrm{J}=1}^{N} M_{\mathrm{J}} \dot{\vec{x}}_{\mathrm{J}} \equiv \vec{P}. \tag{8.3.18}$$

'Breaking' Space-Translation Invariance Suppose we add an extra potential U to L in eq. (8.3.5) that is no longer space-translation invariant.

$$L = \sum_{\mathrm{I}=1}^{N}\left(\frac{M_{\mathrm{I}}}{2}\dot{\vec{x}}^2_{\mathrm{I}} - \frac{1}{2}\sum_{\mathrm{J}\neq\mathrm{I}} V\left(|\vec{x}_{\mathrm{I}} - \vec{x}_{\mathrm{J}}|\right)\right) - U\left(\{\vec{x}_{\mathrm{I}}\}\right) \tag{8.3.19}$$

Show that eq. (8.3.17) would instead become

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{\mathrm{J}=1}^{N} M_{\mathrm{J}}\dot{\vec{x}}_{\mathrm{J}}\right) = -\sum_{\mathrm{J}=1}^{N} \vec{\nabla}_{\vec{x}_{\mathrm{J}}} U. \tag{8.3.20}$$

These are actually the equations obeyed by the center-of-mass of the whole system – can you elaborate? If we wish to model a constant external force $\vec{F}_{\mathrm{ext}}$, show that

$$U = -\vec{F}_{\mathrm{ext}} \cdot \vec{X}_{\mathrm{CM}}, \tag{8.3.21}$$

where $\vec{X}_{\mathrm{CM}} \equiv (\sum_{\mathrm{J}} M_{\mathrm{J}}\vec{x}_{\mathrm{J}})/(\sum_{\mathrm{K}} M_{\mathrm{K}})$ is the center-of-mass coordinate vector. □

Problem 8.19. Rotational Symmetry and Angular Momentum Rotation is defined as the linear transformation on position Cartesian vectors such that their lengths are preserved; i.e., it is an orthogonal matrix. That is, for all $\mathrm{I} = 1, \dots, N$, if we replace $\vec{x}_{\mathrm{I}}$ with its rotated version $\vec{y}_{\mathrm{I}}$ via

$$x^i_{\mathrm{I}} \to \widehat{R}^i{}_j x^j_{\mathrm{I}} \equiv y^i_{\mathrm{I}}; \tag{8.3.22}$$

then for all orthogonal transformations $\widehat{R}^{\mathrm{T}}\widehat{R}=\mathbb{I}$,

$$\vec{x}_{\mathrm{I}}\cdot\vec{x}_{\mathrm{J}}=\vec{x}_{\mathrm{I}}\cdot\widehat{R}^{\mathrm{T}}\widehat{R}\cdot\vec{x}_{\mathrm{J}}=\vec{y}_{\mathrm{I}}\cdot\vec{y}_{\mathrm{J}}. \tag{8.3.23}$$

This immediately informs us, the Lagrangian L in eq. (8.3.5) is invariant under such a rotation operation applied simultaneously to all $\{\vec{x}_{\mathrm{I}}\}$, as long as $\widehat{R}$ is $t-$independent.

$$\dot{\vec{x}}_{\mathrm{I}}\cdot\dot{\vec{x}}_{\mathrm{I}}=\dot{\vec{x}}_{\mathrm{I}}\cdot\widehat{R}^{\mathrm{T}}\widehat{R}\cdot\dot{\vec{x}}_{\mathrm{I}}=\dot{\vec{y}}_{\mathrm{I}}\cdot\dot{\vec{y}}_{\mathrm{I}} \tag{8.3.24}$$

$$|\vec{x}_{\mathrm{I}}-\vec{x}_{\mathrm{J}}|=|\vec{y}_{\mathrm{I}}-\vec{y}_{\mathrm{J}}| \tag{8.3.25}$$

The rotation matrix $\widehat{R}^i{}_j$ has been discussed in Chapter (5.5). We borrow the following result. For infinitesimal rotations,

$$\widehat{R}^i{}_j=\delta^i{}_j+\omega^i{}_j+\mathcal{O}(\omega^2); \tag{8.3.26}$$

where $\omega^i{}_j$ is anti-symmetric

$$\omega^i{}_j=\omega_{ij}=-\omega_{ji}=-\omega^j{}_i \tag{8.3.27}$$

and is otherwise 'small' but arbitrary. Show that the corresponding conserved quantity is the total angular momentum

$$J^{ij}=\sum_{\mathrm{J}=1}^{N}x_{\mathrm{J}}^{[i}\left(M_{\mathrm{J}}\dot{x}_{\mathrm{J}}^{j]}\right)=\sum_{\mathrm{J}=1}^{N}M_{\mathrm{J}}\left(x_{\mathrm{J}}^i\dot{x}_{\mathrm{J}}^j-x_{\mathrm{J}}^j\dot{x}_{\mathrm{J}}^i\right). \tag{8.3.28}$$

Hint: You should be able to explain why $\delta x_{\mathrm{J}}^i=\omega^i{}_j x_{\mathrm{J}}^j$. □

Galilean Transformation and Symmetry Galilean transformations describe the transformation of one inertial frame $(t,\vec{x})$ to another $(t',\vec{x}')$ within the classical dynamics encoded within eq. (8.3.5):

$$\vec{x}=\widehat{R}\cdot\vec{x}'+\vec{a}+\vec{V}\cdot t' \qquad \text{and} \qquad t=t'; \tag{8.3.29}$$

where $\widehat{R}$ is a constant orthogonal matrix obeying $\widehat{R}^{\mathrm{T}}\widehat{R}=\mathbb{I}$; $\vec{a}$ is a constant displacement vector; and $\vec{V}$ is a constant velocity – i.e., $\vec{V}\cdot t+\vec{a}$ is the constant velocity displacement vector connecting one inertial frame to another. This contains the spatial rotations and translations discussed earlier.

Specifically, the equations of motion

$$M_{\mathrm{I}}\ddot{\vec{x}}_{\mathrm{I}}=-\sum_{\mathrm{J}\neq\mathrm{I}}\vec{\nabla}_{\vec{x}_{\mathrm{I}}}V\left(|\vec{x}_{\mathrm{I}}-\vec{x}_{\mathrm{J}}|\right) \tag{8.3.30}$$

are left invariant in form upon such a Galilean transformation in eq. (8.3.29):

$$M_{\mathrm{I}}\ddot{\vec{x}}'_{\mathrm{I}}=-\sum_{\mathrm{J}\neq\mathrm{I}}\vec{\nabla}_{\vec{x}'_{\mathrm{I}}}V\left(|\vec{x}'_{\mathrm{I}}-\vec{x}'_{\mathrm{J}}|\right) \tag{8.3.31}$$

because

$$\ddot{\vec{x}}_{\mathrm{I}}=\widehat{R}\cdot\ddot{\vec{x}}'_{\mathrm{I}}, \tag{8.3.32}$$

$$\vec{x}_{\mathrm{I}}-\vec{x}_{\mathrm{J}}=\widehat{R}\cdot\left(\vec{x}'_{\mathrm{I}}-\vec{x}'_{\mathrm{J}}\right), \tag{8.3.33}$$

$$|\vec{x}_{\mathrm{I}}-\vec{x}_{\mathrm{J}}|=|\vec{x}'_{\mathrm{I}}-\vec{x}'_{\mathrm{J}}|. \tag{8.3.34}$$

Problem 8.20. Show that the Galilean transformations in eq. (8.3.29) implies

$$\vec{\nabla}_{\vec{x}} = \widehat{R} \cdot \vec{\nabla}_{\vec{x}'}. \tag{8.3.35}$$

This completes the argument that the equations in eq. (8.3.30) are left invariant in form – can you explain why? □

Problem 8.21. Noether Charge of Galilean Boost Let us specialize eq. (8.3.29) to the sub-class of Galilean boosts:

$$\vec{x} = \vec{x}' + \vec{V} \cdot t' \qquad \text{and} \qquad t = t'. \tag{8.3.36}$$

The associated Lagrangian in eq. (8.3.5) is not invariant under eq. (8.3.36). Verify, it is shifted by a constant plus a time derivative:

$$L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\}\right) = L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\} \to \{\vec{x}'_{\rm I}, \dot{\vec{x}}'_{\rm I}\}\right) + \frac{1}{2}\sum_{\rm I} M_{\rm I}\vec{V}^2 + \frac{\mathrm{d}}{\mathrm{d}t}\sum_{\rm I} M_{\rm I}\vec{x}'_{\rm I} \cdot \vec{V}. \tag{8.3.37}$$

Then demonstrate that the conserved quantity associated with the Galilean 'boost' due to $\vec{V}$ – i.e, its Noether charge – is related to

$$\vec{P} \cdot t - \sum_{\rm I} M_{\rm I}\vec{x}_{\rm I}, \tag{8.3.38}$$

where $\vec{P} = \sum_{\rm I} M_{\rm I}\dot{\vec{x}}_{\rm I}$ is the total linear momentum (constant whenever $\{\vec{x}_{\rm I}\}$ solve eq. (8.3.30)) and the second term can be dubbed the total dipole moment. Can you see why this is simply the statement that, in the absence of external forces, the center-of-mass $\sum_{\rm I} M_{\rm I}\vec{x}'_{\rm I}/\sum_{\rm J} M_{\rm J}$ of such a closed system must undergo constant velocity motion? □

Noether's Charges: Ambiguity Strictly speaking, the conserved quantities in Noether's theorem is ambiguous up to overall multiplicative and additive constants. In particular, they were read off as the quantities occurring inside a total time derivative, $(\mathrm{d}/\mathrm{d}t)$(conserved quantity) $= 0$; but if some X is $t-$independent so is $\alpha X + \beta$ for constants α and β.

Problem 8.22. Galilean Relativity More generally, under the Galilean transformations in eq. (8.3.29), suppose a Lagrangian of positions and velocities transform as

$$L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\}\right) = L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\} \to \{\vec{x}'_{\rm I}, \dot{\vec{x}}'_{\rm I}\}\right) + \frac{\mathrm{d}}{\mathrm{d}t}F\left(\{\vec{x}'_{\rm I}\}\right) + \text{constant}, \tag{8.3.39}$$

where F is a function of the positions $\{\vec{x}'_{\rm I}\}$ but not velocities $\{\dot{\vec{x}}'_{\rm I}\}$. Show that such a Lagrangian yields Euler-Lagrange equations that transform covariantly as

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{x}^i_{\rm J}} - \frac{\partial}{\partial x^i_{\rm J}}\right) L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\}\right) = 0 = \widehat{R}^{ij}\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{x}'^j_{\rm J}} - \frac{\partial}{\partial x'^j_{\rm J}}\right) L\left(\{\vec{x}_{\rm I}, \dot{\vec{x}}_{\rm I}\} \to \{\vec{x}'_{\rm I}, \dot{\vec{x}}'_{\rm I}\}\right). \tag{8.3.40}$$

This in fact leads to the statement of Galilean relativity:

> The laws of physics encoded within a Lagrangian that obeys eq. (8.3.37) takes the same form in all inertial frames related by Galilean transformations. □

2 Body Problem With Central Potential The 2 body problem plays a key role in real world applications: from the electron orbiting a proton (i.e., the H atom); to Earth orbiting the Sun; to scattering experiments in atomic, nuclear and particle physics. Here, we will study its non-relativistic version, assuming the potential energy between the two bodies – with masses M_1 and M_2 located at $\vec{x}_1$ and $\vec{x}_2$ respectively – only depends on the distance between them $|\vec{x}_1 - \vec{x}_2|$. The 2 body Lagrangian is

$$L_{2\mathrm{B}} = \frac{M_1}{2}\dot{\vec{x}}_1^2 + \frac{M_2}{2}\dot{\vec{x}}_2^2 - V\left(|\vec{x}_1 - \vec{x}_2|\right). \tag{8.3.41}$$

If we do a change-of-coordinates to the center-of-mass coordinate

$$\vec{X}_{\mathrm{CM}} \equiv \frac{M_1\vec{x}_1 + M_2\vec{x}_2}{M_1 + M_2}, \tag{8.3.42}$$

$$\vec{\Delta} \equiv \vec{x}_1 - \vec{x}_2; \tag{8.3.43}$$

the Lagrangian will transform into

$$L_{2\mathrm{B}} = \frac{M_1 + M_2}{2}\dot{\vec{X}}_{\mathrm{CM}}^2 + \frac{\mu}{2}\dot{\vec{\Delta}}^2 - V\left(\Delta\right); \tag{8.3.44}$$

where the *reduced mass* μ is given by the relation

$$\mu \equiv \frac{M_1 M_2}{M_1 + M_2}. \tag{8.3.45}$$

Eq. (8.3.44) teaches us, the non-relativistic 2 body problem with a central potential U can always be reduced to a free particle with mass $M_1 + M_2$; plus a 1 body problem subject to the same potential V. The resulting Euler-Lagrange equations are: acceleration-free motion of the center-of-mass,

$$\ddot{\vec{X}}_{\mathrm{CM}} = 0; \tag{8.3.46}$$

and the reduced mass' motion driven by the central potential,

$$\mu\ddot{\vec{\Delta}} = -V'(\Delta)\widehat{\Delta}. \tag{8.3.47}$$

On the other hand, since the problem is spherically symmetric, we may exploit spherical coordinates:

$$\vec{\Delta} = r\left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)\right). \tag{8.3.48}$$

Problem 8.23. Equations-of-Motion in Spherical Coordinates If we denote

$$U \equiv V/\mu \tag{8.3.49}$$

– remember Lagrangians are only defined up to overall constants – show that the Lagrangian for $\vec{\Delta}$–motion can be written as

$$L_{\Delta} \equiv \frac{1}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin(\theta)^2\dot{\phi}^2\right) - U(r). \tag{8.3.50}$$

Then show that the radial motion is governed by

$$\ddot{r} - r\left(\dot{\theta}^2 + \sin^2(\theta)\dot{\phi}^2\right) + U'(r) = 0; \tag{8.3.51}$$

and the angular motion obeys the following equations, as long as $r \neq 0$ and $\sin\theta \neq 0$.

$$\ddot{\theta} + 2\frac{\dot{r}}{r}\dot{\theta} - \sin(\theta)\cos(\theta)\dot{\phi}^2 = 0 \tag{8.3.52}$$

$$\ddot{\phi} + 2\left(\frac{\dot{r}}{r} + \dot{\theta}\cot(\theta)\right)\dot{\phi} = 0. \tag{8.3.53}$$

Notice equations (8.3.52) and (8.3.53) arises from the (reduced) Lagrangian

$$L_\Omega \equiv \frac{1}{2}r^2\left(\dot{\theta}^2 + \sin(\theta)^2\dot{\phi}^2\right). \tag{8.3.54}$$

Explain why this L_Ω can be derived from eq. (8.3.48) by pretending r is $t-$independent while taking time derivatives; namely, when computing $L_\Omega = (1/2)\partial_t\vec{\Delta}\cdot\partial_t\vec{\Delta}$. Further explain why this, in turn, tells us L_Ω is spherically symmetric. □

The spherical symmetry of eq. (8.3.52) and (8.3.53) are important because we may now argue that, since the choice of the 3−axis is arbitrary, we may always orient the axes such that all motion takes place on the equatorial $\theta = \pi/2$ plane. First, by spherical symmetry, we may orient our axes so that the initial velocity $\dot{\Delta}(t_0)$ at some time t_0 lies on the equatorial plane.[72] Therefore, we only need to show that, given an initial velocity that lies on the $(1,2)$ plane, it will always remain so. To this end, notice from eq. (8.3.52) that whenever $\theta = \pi/2$ and $\dot{\theta} = 0$ (i.e., velocity tangent to the equatorial plane) at a given time t_0, the acceleration of the altitude angle θ is zero: $\ddot{\theta} = 0$. But that means over the next time step $t_0 \to t_0 + \mathrm{d}t$, the velocity will remain confined to the equatorial plane. But since this is true for any t_0, that means motion is confined to the equatorial plane for all time. We now merely have to solve the equatorial plane $\theta = \pi/2$ versions of equations (8.3.51), (8.3.52) and (8.3.53):

$$\ddot{r} - r\dot{\phi}^2 + U'(r) = 0, \tag{8.3.55}$$

$$\ddot{\phi} + 2\frac{\dot{r}}{r}\dot{\phi} = 0. \tag{8.3.56}$$

Problem 8.24. Equatorial Plane Lagrangian Since we have argued that all motion can be assumed to take place on the equatorial plane, that means equations (8.3.55) and (8.3.56) should be derivable from the Lagrangian eq. (8.3.50) by restricting it to $\theta = \pi/2$ from the outset:

$$L_\Delta \equiv \frac{1}{2}\left(\dot{r}^2 + r^2\dot{\phi}^2\right) - U(r). \tag{8.3.57}$$

The reasoning is that, since motion is known to takes place strictly on the equatorial plane anyway, when demanding the extremum of its associated action there is no need to vary θ away

[72] For instance, $\vec{\Delta}(t_0)$ can be defined to be parallel to the positive 1−axis $\widehat{e}_1$; and the $(2,3)$ plane can then be rotated until the component of $\partial_t\vec{\Delta}(t_0)$ perpendicular to $\widehat{e}_1$ is parallel to the positive 2−axis $\widehat{e}_2$.

from $\pi/2$ at all. Show explicitly that it is indeed the case; i.e., obtain equations (8.3.55) and (8.3.56) from eq. (8.3.57).

Explain why the energy E and angular momentum ℓ, defined by

$$E \equiv \frac{1}{2}\left(\dot{r}^2 + r^2\dot{\phi}^2\right) + U(r) \qquad \text{and} \qquad \ell \equiv r^2\dot{\phi}, \tag{8.3.58}$$

are constants-of-motion. (Hint: Recall equations (8.1.40) and (8.1.41).) Explain why the constancy of ℓ in eq. (8.3.58) amounts to the (generalized) Kepler's second law of motion:

> The area per unit time swept out by the displacement vector of the point mass is the constant $\ell/2$.

If A is the total area swept out in one (2π) orbit, T is the orbital time period, $f = 1/T$ is the frequency, and $\Omega = 2\pi f$ is the angular frequency, explain why

$$A = \frac{\ell}{2}T = \frac{\ell}{2f} = \frac{\pi\ell}{\Omega}. \tag{8.3.59}$$

Finally, evaluate J^{12} in eq. (8.3.28) for the 2−body problem at hand; and verify

$$J^{12} = r^2\dot{\phi}. \tag{8.3.60}$$

This confirms that ℓ is indeed the conserved angular momentum. □

In eq. (8.3.58), we may insert $\dot{\phi} = \ell/r^2$ into the energy conservation equation to state

$$E = \frac{1}{2}\dot{r}^2 + U_{\text{eff}}(r), \tag{8.3.61}$$

$$U_{\text{eff}}(r) \equiv \frac{\ell^2}{2r^2} + U(r). \tag{8.3.62}$$

This may be interpreted as a 1D setup: the total (constant) energy E is equals to the 'kinetic' $\dot{r}^2/2$ plus 'potential' U_{eff}. Hence, for a fixed E, the motion of $\vec{\Delta}$ takes place between the maximum and minimum r is given by the intersection of the horizontal E line with the effective potential $U_{\text{eff}}(r)$. Additionally, we may perform the change-of-derivatives $\mathrm{d}r/\mathrm{d}t = (\mathrm{d}r/\mathrm{d}\phi)(\mathrm{d}\phi/\mathrm{d}t)$ in eq. (8.3.58) and deduce

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\frac{\ell}{r^2}\right)^2 = 2\left(E - U_{\text{eff}}(r)\right); \tag{8.3.63}$$

where we have again replaced $\dot{\phi} \to \ell/r^2$. Re-arranging,

$$\frac{\mathrm{d}\phi}{\mathrm{d}r} = \frac{\ell}{r^2\sqrt{2\left(E - U_{\text{eff}}(r)\right)}}, \tag{8.3.64}$$

$$\phi - \phi_0 = \int \frac{\ell \cdot \mathrm{d}r}{r^2\sqrt{2\left(E - U_{\text{eff}}(r)\right)}}. \tag{8.3.65}$$

We see that this equation allows us – at least in principle – to obtain $r(\phi)$, the shape of the trajectory as a function of the azimuthal angle ϕ.

Problem 8.25. Kepler Problem: Planetary Motion Consider two bodies of masses M_1 and M_2 interacting through a gravitational (Newtonian) potential energy given by

$$V = -\frac{G_{\rm N} M_1 M_2}{|\vec{x}_1 - \vec{x}_2|}. \tag{8.3.66}$$

Show that the effective potential is

$$U_{\rm eff}(r) = \frac{\ell^2}{2r^2} - \frac{G_{\rm N}(M_1 + M_2)}{r}, \tag{8.3.67}$$

with a minimum located at $r_\star \equiv \ell^2/(G_{\rm N}(M_1 + M_2))$ and

$$U_{\rm eff}(r_\star) = -\frac{1}{2}\left(\frac{G_{\rm N}(M_1 + M_2)}{\ell}\right)^2 \equiv U_\star. \tag{8.3.68}$$

Make a sketch of $U_{\rm eff}(r)$. (Hint: It should go to positive infinity as $r \to 0$; dip below zero where there is only one minimum; and then gradually rise back up to zero as $r \to \infty$.) Explain why unbound orbits exist whenever $E \geq 0$; and why bound ones exist for

$$U_\star \leq E < 0. \tag{8.3.69}$$

Explain why $E = 0$ is a 'marginal' case; and why $E = U_\star$ refers to a circular orbit.

Show that the integral in eq. (8.3.64) yields

$$\phi - \phi_0 = \arccos\left(\frac{1}{\sqrt{2}\sqrt{E - U_\star}}\left(\frac{\ell}{r} - \sqrt{-2U_\star}\right)\right) \tag{8.3.70}$$

and

$$r(\phi) = \frac{\ell/\sqrt{-2U_\star}}{1 + \sqrt{1 - E/U_\star} \cdot \cos(\phi - \phi_0)}. \tag{8.3.71}$$

Eccentricity Remember from eq. (8.3.68) that the minimum effective potential is negative, $U_\star < 0$. Hence, the eccentricity

$$e \equiv \sqrt{1 - E/U_\star} \tag{8.3.72}$$

is greater than unity for unbound orbits ($e \geq 1$), with $e = 1$ being the marginal case; and less than unity for bound orbits ($0 \leq e < 1$). For circular motion $E = U_\star$ and the eccentricity is zero: $e = 0$.

Semi-Major Axis The minimum and maximum radii, according to eq. (8.3.71), occur at $\phi - \phi_0 = 0$ and $\phi - \phi_0 = \pi$ respectively. Explain why the sum of $r_{\rm min} \equiv r(\phi - \phi_0 = 0)$ and $r_{\rm max} \equiv r(\phi - \phi_0 = \pi)$ is twice of the semi-major axis a, and hence:

$$a = \frac{\ell^2}{G_{\rm N}(M_1 + M_2)(1 - e^2)}. \tag{8.3.73}$$

Use these facts to also demonstrate

$$E = -(1-e^2)\left(\frac{G_{\rm N}(M_1+M_2)}{\sqrt{2}\ell}\right)^2 = -\frac{G_{\rm N}(M_1+M_2)}{2a}, \tag{8.3.74}$$

$$\ell = \sqrt{1-e^2}\frac{G_{\rm N}^{2/3}(M_1+M_2)^{2/3}}{\Omega^{1/3}}. \tag{8.3.75}$$

Kepler's Third Law Employ the relation in eq. (8.3.59) between the area A swept out by the orbit and its period $T = 2\pi/\Omega$ – a consequence of the (generalized) Kepler's second law – to deduce

$$\frac{\ell}{\Omega} = a^2\sqrt{1-e^2} \tag{8.3.76}$$

and

$$\Omega^2 = \frac{G_{\rm N}(M_1+M_2)}{a^3}. \tag{8.3.77}$$

Finally, explain why eq. (8.3.71) may be re-expressed as

$$r(\phi) = \sqrt[3]{\frac{G_{\rm N}(M_1+M_2)}{\Omega^2}}\frac{1-e^2}{1+e\cdot\cos(\phi-\phi_0)}. \tag{8.3.78}$$

□

Problem 8.26. Kepler Problem: Laplace-Runge-Lenz Vector Assuming that Newtonian gravity holds – namely,

$$\ddot{\vec{\Delta}} = -\frac{G_{\rm N}(M_1+M_2)}{\Delta^2}\widehat{\Delta}, \tag{8.3.79}$$

$$\widehat{\Delta} \equiv \frac{\vec{\Delta}}{\Delta} \tag{8.3.80}$$

– verify directly that the following Laplace-Runge-Lenz vector $\vec{A}$ is a constant-of-motion.

$$\vec{A} \equiv -\dot{\vec{\Delta}} \times \left(\dot{\vec{\Delta}} \times \vec{\Delta}\right) - G_{\rm N}(M_1+M_2)\widehat{\Delta}. \tag{8.3.81}$$

That this vector $\vec{A}$ is constant, is the result of the $1/r$ character of the Newtonian potential. Hence, it is also constant for the corresponding classical electrodynamics problem with its $1/r$ Coulomb potential. Upon quantization, the Laplace-Runge-Lenz vector may be exploited to obtain the energy levels of the (spin-less) hydrogen atom from purely algebraic means. □

Constrained Particle Motion Suppose our non-relativistic particle experiences a potential energy of $V(\vec{x})$ but is further constrained by the N constraints $\{G_{\rm I}(\vec{x}) = 0\}$, where $\vec{x}$ is the position of the particle in 3D Cartesian coordinates. This means our modified Lagrangian, with N Lagrange multipliers $\{\Lambda^{\rm I}\}$, is now

$$L = \frac{m}{2}\dot{\vec{x}}(t)^2 - V(\vec{x}(t)) - \Lambda^{\rm I}(t)\cdot G_{\rm I}(\vec{x}(t)). \tag{8.3.82}$$

By viewing the Lagrange multipliers as extra 'coordinates', the constraint equations themselves may be viewed as part of the full set of Euler-Lagrange equations. In particular, since L does not depend on $\dot{\Lambda}^{\mathrm{I}}$, we not only have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i} = 0, \tag{8.3.83}$$

but also

$$G_{\mathrm{I}}(\vec{x}) = -\frac{\partial L}{\partial \Lambda^{\mathrm{I}}} = 0. \tag{8.3.84}$$

For simplicity, let us focus on the case where there is only one constraint: $\Lambda^{\mathrm{I}} \cdot G_{\mathrm{I}} \to \Lambda \cdot G$. Applying the Euler-Lagrange equations now hand us

$$m\ddot{\vec{x}} = -\vec{\nabla}_{\vec{x}} V(\vec{x}) - \Lambda \cdot \vec{\nabla}_{\vec{x}} G(\vec{x}). \tag{8.3.85}$$

Remember that $\vec{\nabla} G(\vec{x})$ points towards the direction of most rapid change in G; whereas G is constant in the direction perpendicular to $\vec{\nabla} G(\vec{x})$ – i.e., $\vec{\nabla} G$ is normal to the constraint surface. In detail, if $\widehat{\xi}$ is an arbitrary a unit vector parallel to this surface,

$$\widehat{\xi} \cdot \vec{\nabla} G = 0. \tag{8.3.86}$$

We thus see the dynamics *along* the constraint surface does not involve G:

$$m\ddot{\vec{x}} \cdot \widehat{\xi} = -\vec{\nabla} V \cdot \widehat{\xi} \tag{8.3.87}$$

If we further normalize $\vec{\nabla} G$ to unit length; namely,

$$\widehat{n} \equiv \vec{\nabla} G(\vec{x}) / |\vec{\nabla} G(\vec{x})|; \tag{8.3.88}$$

then

$$m\ddot{\vec{x}} \cdot \widehat{n} = -\vec{\nabla}_{\vec{x}} V(\vec{x}) \cdot \widehat{n} - \Lambda |\vec{\nabla}_{\vec{x}} G(\vec{x})|. \tag{8.3.89}$$

In other words, the normal force

$$\vec{F}_{\mathrm{normal}} = -\Lambda |\vec{\nabla}_{\vec{x}} G(\vec{x})| \, \widehat{n} \tag{8.3.90}$$

is what keeps the particle stuck on the constant G surface.

Energy Conservation for Constrained Motion According to eq. (8.1.40), whenever V and G are time-independent, the total energy

$$E = \dot{x}^i \frac{\partial L}{\partial \dot{x}^i} - L = \frac{m}{2}\dot{\vec{x}}^2 + V(\vec{x}) + \Lambda^{\mathrm{I}} \cdot G_{\mathrm{I}}(\vec{x}) \tag{8.3.91}$$

is a constant-of-motion. If we further impose the constraints $\{G_{\mathrm{I}} = 0\}$, the total energy no longer depends on the Lagrange multiplier terms:

$$E = \left(\frac{m}{2}\dot{\vec{x}}^2 + V(\vec{x})\right)_{\{G_{\mathrm{I}}(\vec{x})=0\}}. \tag{8.3.92}$$

Intrinsic Coordinates Instead of implementing constraints through Lagrange multipliers in eq. (8.3.82), we may employ the $(D-N)$ coordinates $\vec{\xi}$ intrinsic to the constraint surface; i.e., $\vec{x}(\vec{\xi})$ parametrize the $(D-N)$–dimensional surface. Since $G_{\text{I}}(\vec{x}) = 0$ when the constraints are satisfied, the Lagrangian in eq. (8.3.82) translates to

$$L = \frac{1}{2} g_{\text{AB}}(\vec{\xi}) \dot{\xi}^{\text{A}} \dot{\xi}^{\text{B}} - V(\vec{\xi}), \tag{8.3.93}$$

where $V(\vec{\xi}) \equiv V(\vec{x}(\vec{\xi}))$ and by the chain rule,

$$g_{\text{AB}} \equiv \frac{\partial \vec{x}}{\partial \xi^{\text{A}}} \cdot \frac{\partial \vec{x}}{\partial \xi^{\text{B}}}. \tag{8.3.94}$$

(The indices A and B run from 1 through N.) The Euler-Lagrange equations read

$$\frac{D^2 \xi^{\text{A}}}{\text{d}t^2} = -g^{\text{AB}} \partial_{\xi^{\text{B}}} V(\vec{\xi}), \tag{8.3.95}$$

$$\frac{D^2 \xi^{\text{A}}}{\text{d}t^2} \equiv \ddot{\xi}^{\text{A}} + \Gamma^{\text{A}}{}_{\text{BF}} \dot{\xi}^{\text{B}} \dot{\xi}^{\text{F}}, \tag{8.3.96}$$

$$\Gamma^{\text{A}}{}_{\text{BF}} \equiv \frac{1}{2} g^{\text{AC}} \left(\partial_{\text{B}} g_{\text{FC}} + \partial_{\text{F}} g_{\text{BC}} - \partial_{\text{C}} g_{\text{BF}} \right), \tag{8.3.97}$$

where g^{AB} is defined as the inverse matrix of g_{AB}; namely, $g^{\text{AC}} g_{\text{CB}} = \delta^{\text{A}}{}_{\text{B}}$.

Problem 8.27. Starting from eq. (8.3.93), show its the Euler-Lagrange equations yield eq. (8.3.95). *Bonus*: Can you show that eq. (8.3.85) is equivalent to eq. (8.3.95) by transforming $\vec{x} \to \vec{x}(\vec{\xi})$? □

Constraint before Variation? What we have just discovered is, there is actually no need to introduce Lagrange multipliers if we can find generalized coordinates $\{\xi^{\text{A}}\}$ intrinsic to the constraint surface. Remember, the variational principle of non-relativistic classical mechanics says that motion extremizes the integral of the difference between kinetic and potential energy, with the boundary conditions fixed. But if we already know beforehand that the motion *has* to take place on some constraint surface, there is no need to consider the variation of the trajectory away from the surface.

Likewise, for the 2 body problem with a central potential $V(r)$, we were able to argue that all the motion had to take place on a fixed 2D plane, which we then chose to be $\theta = \pi/2$ in the spherical coordinate system. This allowed us to obtain the equatorial plane Lagrangian in eq. (8.3.57), since there is no need to consider variation of the trajectory away from $\theta = \pi/2$.

Problem 8.28. In general, however, if we do not know beforehand how the motion may be constrained, then it is illegal to constrain the coordinates *before* deriving the corresponding Euler-Lagrange equations.

As an example, in the 2 body problem with a central potential, we may apply the conservation of angular momentum $\dot{\phi} = \ell/r^2$ to eq. (8.3.55):

$$\ddot{r} - \frac{\ell^2}{r^3} + U'(r) = 0. \tag{8.3.98}$$

Now insert $\dot{\phi} = \ell/r^2$ into the equatorial plane Lagrangian in eq. (8.3.57); then derive the corresponding Euler-Lagrange equations. You should find

$$\ddot{r} + \frac{\ell^2}{r^3} + U'(r) = 0; \tag{8.3.99}$$

where the ℓ^2 term now has the wrong sign.

In words: we need to consider all trajectories (r, ϕ) and pick the ones that extremize $T - V$ on the equatorial plane; whereas $\dot{\phi} = \ell/r^2$ means we have already constrained the trajectories to only a subset of all 2D motion. The analogous problem in ordinary calculus is: do not evaluate a function at a particular value *before* differentiation; for e.g., if $f(z) = \sin(z + a)$ then $f'(a) = \cos(2a) \neq \partial_a \sin(2a)$. $\square$

Example: Circular Motion Immersed in Constant Gravity In this problem we will examine motion of a point particle confined on a vertical circle immersed in a constant gravitational field. The coordinates of the circle are

$$(x(\theta), y(\theta)) = (0, R) + r(\cos\theta, \sin\theta), \tag{8.3.100}$$

where the positive $y-$direction ($\theta = \pi/2$) points upwards; and the gravitational force per unit mass is

$$\vec{F}_g = -g(0, 1). \tag{8.3.101}$$

If we exploit 2D polar coordinates, the Lagrangian per unit mass is therefore

$$L = \frac{1}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - g(r\sin(\theta) + R) - \lambda \cdot (r - R). \tag{8.3.102}$$

We have chosen a gravitational potential so that it is zero at the bottom of the circle. Moreover, the total conserved energy is, evaluated on the circular trajectory $r = R$ is

$$E = \frac{R^2}{2}\dot{\theta}^2 + gR(\sin(\theta) + 1). \tag{8.3.103}$$

At the bottom of the circle, $\theta = (3/2)\pi$, and all the energy is kinetic. Hence, if we denote v as the particle's velocity at the bottom of the circle, $E = v^2/2$ and

$$\frac{v^2}{2} = \frac{R^2}{2}\dot{\theta}^2 + gR(\sin(\theta) + 1). \tag{8.3.104}$$

The Euler-Lagrange equations are

$$R\dot{\theta}^2 - g\sin(\theta) = \lambda(t) \tag{8.3.105}$$

$$\ddot{\theta} + (g/R)\cos(\theta) = 0. \tag{8.3.106}$$

For course, for small oscillations around the bottom, $\varphi \equiv \theta - (3\pi/2)$, the second line would yield the SHO equation with angular frequency given by $\sqrt{g/R}$:

$$\ddot{\varphi} + (g/R)\varphi = 0. \tag{8.3.107}$$

We notice eq. (8.3.106) alone is sufficient for understanding motion on the circle itself. However, eq. (8.3.105) in fact provides additional information. Recall that the normal force is what holds the particle in place. If equations (8.3.82) and (8.3.102) are compared; we may identify

$$G = r - R; \tag{8.3.108}$$

and therefore eq. (8.3.90) now reads

$$\vec{F}_{\text{normal}} = -\Lambda\widehat{r}. \tag{8.3.109}$$

If we solve for $\dot{\theta}^2/2$ in eq. (8.3.104),

$$R\dot{\theta}^2 = \frac{2}{R}\left(\frac{v^2}{2} - gR(\sin(\theta)+1)\right), \tag{8.3.110}$$

and insert this into eq. (8.3.105),

$$\begin{aligned} \lambda(t) &= \frac{v^2}{R} - 2g(\sin(\theta)+1) - g\sin(\theta) \\ &= \frac{v^2}{R} - g(3\sin(\theta)+2). \end{aligned} \tag{8.3.111}$$

Keeping in mind eq. (8.3.109), we see that at the bottom of the circle, the particle experiences normal force of

$$\theta = \frac{3\pi}{2}: \qquad \vec{F}_{\text{normal}} = -\widehat{r}\left(\frac{v^2}{R} - g\right). \tag{8.3.112}$$

Now, if the particle were held in place by a rigid rod, say, the normal force $\lambda(t)$ can be either positive or negative. However, if the particle were moving along the inner surface of a circular hoop, then $\lambda > 0$ on physical grounds. Eq. (8.3.111) then indicates – if we define $\theta \equiv \phi + \pi/2$ so that ϕ refers to the top of the circle – the particle cannot access the region

$$\cos(\theta) \geq \frac{1}{3}\left(\frac{v^2}{gR} - 2\right). \tag{8.3.113}$$

There are non-trivial solutions whenever the rightmost factor is less than unity, which in turn is equivalent to

$$v \leq \sqrt{5gR}. \tag{8.3.114}$$

This example illustrates why it is sometimes useful to introduce a Lagrange multiplier even if we know how to describe the constraint surface with intrinsic coordinates (here, $r = R$ and $\xi = \theta$). It yields the normal force holding the particle in place, i.e., eq. (8.3.90), which is lost in the intrinsic formulation. For the problem at hand, knowing the normal force allows us to deduce when the constrained motion itself breaks down.

Problem 8.29. Ellipsoidal Motion Immersed in Constant Gravity Consider a unit mass particle moving along the inner surface of the following ellipsoidal hoop immersed in a constant gravitational force $\vec{F}_g$:

$$(x(\theta), y(\theta)) = (0, \rho_0) + \left(\sqrt{\rho_0^2 - R^2}\sin\theta, \rho_0\cos\theta\right), \qquad \rho_0 > R > 0 \tag{8.3.115}$$

$$\vec{F}_g = -g(0,1). \tag{8.3.116}$$

Starting from the ellipsoidal coordinates

$$(x(\rho,\theta), y(\rho,\theta)) = \left(\sqrt{\rho^2 - R^2}\sin\theta, \rho_0\cos\theta\right), \tag{8.3.117}$$

first argue that the appropriate Lagrangian is

$$L = \frac{1}{2}\left(\dot{\theta}^2\left(\rho^2 - R^2\cos^2(\theta)\right) + \dot{\rho}^2\left(\frac{\rho^2\sin^2(\theta)}{\rho^2 - R^2} + \cos^2(\theta)\right)\right)$$
$$- g(\rho\cdot\cos(\theta) + \rho_0) - \lambda(\rho - \rho_0). \tag{8.3.118}$$

Show that, in order to reach the top of the ellipsoid at $\theta = 0$, if the particle begins at the bottom $(\theta = \pi)$ with speed v, it must be faster than $v \geq \sqrt{g\cdot R^2/\rho_0}\sqrt{5(\rho_0/R)^2 - 1}$. □

Problem 8.30. Brachistochrone curve Restricting our attention to the vertical (x, y) plane, suppose a point mass starts from $(0,0)$ and takes a path to (x_0, y_0) under the influence of the downward gravitational force per unit mass of $-g\widehat{y}$, where $\widehat{y}$ is the unit vector pointing (upwards) parallel to the positive $y-$axis. We wish to derive the path that takes the *least* time.

Let us solve y in terms of x, so that the path is now $(x, y) = (x, y(x))$. If we set zero potential energy at $y = 0$, since the mass started from rest, by the conservation of energy

$$0 = \frac{1}{2}(\dot{x}(t)^2 + \dot{y}(t)^2) + gy(t) \tag{8.3.119}$$

$$= \frac{1}{2}\dot{x}^2(1 + (\mathrm{d}y/\mathrm{d}x)^2) + gy. \tag{8.3.120}$$

- By recognizing that $\mathrm{d}t = \mathrm{d}x/\dot{x}(t)$ show that the total time taken Δt is given by the integral

$$\Delta t = \int_0^{x_0} \mathrm{d}x\sqrt{\frac{1 + y'(x)^2}{-2gy(x)}}. \tag{8.3.121}$$

- To minimize Δt, use the constant of motion in eq. (8.1.19) to show that

$$y'(x) = -\sqrt{-1 - \frac{\zeta_0^2}{y}}, \tag{8.3.122}$$

for some constant ζ_0. Since $y < 0$ by assumption, we see that $0 \leq -y/\zeta_0^2 \leq 1$ for the square root to be real.

- Integrate this $y'(x)$ equation to arrive at the following relationship between y and x:

$$\zeta_0^2 \left(\arcsin \sqrt{-\frac{y}{\zeta_0^2}} - \sqrt{-\frac{y}{\zeta_0^2}\left(1 + \frac{y}{\zeta_0^2}\right)} \right) = x - \eta_0, \tag{8.3.123}$$

  where η_0 is an arbitrary constant.

- Explain why $(x(\psi = 0), y(\psi = 0)) = (0, 0)$ implies $\eta_0 = 0$. Then deduce that, for a fixed endpoint (x_0, y_0), the remaining constant is given by the solution to

$$\arcsin \sqrt{-\frac{y_0}{\zeta_0^2}} - \sqrt{-\frac{y_0}{\zeta_0^2}\left(1 + \frac{y_0}{\zeta_0^2}\right)} = \frac{x_0}{\zeta_0^2}. \tag{8.3.124}$$

- Now, put $\psi = \arcsin \sqrt{-y/\zeta_0^2}$ and deduce the following parametric solution in terms of ψ:

$$x(\psi) = \frac{\zeta_0^2}{2} \left(2\psi - \sin(2\psi)\right), \tag{8.3.125}$$
$$y(\psi) = -\zeta_0^2 \sin^2(\psi). \tag{8.3.126}$$

  Be sure to explain why the initial ψ is 0; whereas the final ψ is $\arcsin \sqrt{-y_0/\zeta_0^2}$.

It is apparently possible to solve the same problem but with kinetic friction included; see the Wolfram MathWorld page here. □

8.4 Hamiltonian Dynamics in Classical Physics

We now turn to a different but equivalent formulation of classical mechanics: Hamiltonian dynamics. It is also a key starting point for understanding quantum mechanics, because – as we shall see below – the Hamiltonian H of a physical system is itself the core object governing its time evolution.

Given a Lagrangian $L(t, \vec{q}, \dot{\vec{q}})$, the corresponding Hamiltonian is defined through the *Legendre transform*

$$H\left(t, \vec{q}, \vec{p}\right) \equiv p_i \dot{q}^i(\vec{q}, \vec{p}) - L(t, \vec{q}, \vec{p}), \tag{8.4.1}$$

where the momentum conjugate to $\vec{q}$ is defined as

$$p_i(t) \equiv \left(\frac{\partial L(t, \vec{q}, \dot{\vec{q}})}{\partial \dot{q}^i} \right)_{t, \vec{q}}. \tag{8.4.2}$$

In terms of p_i, the Euler-Lagrange equations (8.1.31) now reads

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \left(\frac{\partial L(t, \vec{q}, \dot{\vec{q}})}{\partial q^i} \right)_{t, \dot{\vec{q}}}. \tag{8.4.3}$$

The definition for p_i in eq. (8.4.2) relates the three groups of objects: $\{\vec{q}, \vec{p}, \dot{\vec{q}}\}$. This usually, though not always, means we may also solve $\dot{\vec{q}}$ in terms of $\vec{q}$ and $\vec{p}$. (Some jargon: When we may solve $\dot{\vec{q}}$ in terms of $\vec{q}$ and $\vec{p}$, then the system is dubbed *non-degenerate*; and, if $\dot{\vec{q}}$ cannot be solved in terms of $\vec{q}$ and $\vec{p}$ the system is dubbed *degenerate*.)

Let us vary the Hamiltonian $H(t, \vec{q}, \vec{p})$ to first order, by perturbing the time

$$t \to t + \mathrm{d}t; \tag{8.4.4}$$

as well as the position and its conjugate momentum:

$$q^i \to q^i + \delta q^i \qquad \text{and} \qquad p_i \to p_i + \delta p_i. \tag{8.4.5}$$

We obtain

$$\delta H = \left(\frac{\partial H}{\partial t}\right)_{\vec{q},\vec{p}} \mathrm{d}t + \left(\frac{\partial H}{\partial q^i}\right)_{t,\vec{p}} \delta q^i + \left(\frac{\partial H}{\partial p_i}\right)_{t,\vec{q}} \delta p_i. \tag{8.4.6}$$

On the other hand, by varying the Legendre transform,

$$\delta H = \delta p_i \dot{q}^i + p_i \frac{\mathrm{d}}{\mathrm{d}t}\delta q^i - \left(\frac{\partial L}{\partial t}\right)_{\vec{q},\dot{\vec{q}}} \mathrm{d}t - \left(\frac{\partial L}{\partial q^i}\right)_{t,\dot{\vec{q}}} \delta q^i - \left(\frac{\partial L}{\partial \dot{q}^i}\right)_{t,\vec{q}} \frac{\mathrm{d}}{\mathrm{d}t}\delta q^i. \tag{8.4.7}$$

If we apply the definition eq. (8.4.2), we see the $(\mathrm{d}/\mathrm{d}t)\delta q^i$ terms drop out and

$$\delta H = -\left(\frac{\partial L}{\partial t}\right)_{\vec{q},\dot{\vec{q}}} \mathrm{d}t - \left(\frac{\partial L}{\partial q^i}\right)_{t,\dot{\vec{q}}} \delta q^i + \dot{q}^i \delta p_i. \tag{8.4.8}$$

Comparing the coefficients of $\mathrm{d}t$, δq^i and δp_i in equations (8.4.6) and (8.4.8), we arrive at the following relations between the first derivatives of the Hamiltonian and Lagrangian:

$$\left(\frac{\partial H(t, \vec{q}, \vec{p})}{\partial t}\right)_{\vec{q},\vec{p}} = -\left(\frac{\partial L(t, \vec{q}, \dot{\vec{q}})}{\partial t}\right)_{\vec{q},\dot{\vec{q}}}, \tag{8.4.9}$$

$$\left(\frac{\partial H(t, \vec{q}, \vec{p})}{\partial q^i}\right)_{t,\vec{p}} = -\left(\frac{\partial L(t, \vec{q}, \dot{\vec{q}})}{\partial q^i}\right)_{t,\dot{\vec{q}}}, \tag{8.4.10}$$

$$\left(\frac{\partial H(t, \vec{q}, \vec{p})}{\partial p_i}\right)_{t,\vec{q}} = \dot{q}^i(t). \tag{8.4.11}$$

It is worth reiterating, equations (8.4.9), (8.4.10) and (8.4.11) follow simply from the *definitions* in (8.4.1) and (8.4.2). But if we now employ the dynamics as encoded in the Euler-Lagrange equations in eq. (8.4.3), we arrive at *Hamilton's equations*:

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = -\left(\frac{\partial H(t, \vec{q}, \vec{p})}{\partial q^i}\right)_{t,\vec{p}} \qquad \text{and} \qquad \frac{\mathrm{d}q^i(t)}{\mathrm{d}t} = \left(\frac{\partial H(t, \vec{q}, \vec{p})}{\partial p_i}\right)_{t,\vec{q}}. \tag{8.4.12}$$

Example For non-relativistic classical mechanics with Lagrangian

$$L = \frac{1}{2} m \dot{\vec{x}}^2 - V(\vec{x}); \tag{8.4.13}$$

by identifying $\vec{x} \equiv \vec{q}$, the conjugate momentum is

$$p_i = \frac{\partial L}{\partial \dot{x}^i} = m\dot{x}^i. \tag{8.4.14}$$

The Hamiltonian is therefore

$$H(\vec{x}, \vec{p}) = p_i \dot{x}^i - L = \frac{\vec{p}^2}{m} - \left(\frac{\vec{p}^2}{2m} - V(\vec{x})\right) \tag{8.4.15}$$

$$= \frac{\vec{p}^2}{2m} + V(\vec{x}). \tag{8.4.16}$$

Hamilton's equations are

$$\dot{p}_i = -\partial_{x^i} V(\vec{x}) \tag{8.4.17}$$

and

$$\dot{x}^i = p_i/m. \tag{8.4.18}$$

Configuration versus Phase Space We witness the key distinction between the Lagrangian versus the Hamiltonian formulations. The former yields second order equations-of-motion (provided L depends only on $\vec{q}$ and $\dot{\vec{q}}$) for a single set of $t-$dependent variables $D-$component object $\vec{q}(\lambda)$; whereas the latter yields first order ones for the pair of $D-$component *phase space* coordinates $(\vec{q}(\lambda), \vec{p}(\lambda))$. This means, to solve the dynamics encoded within these ODEs, we need to provide $2D$ independent boundary and/or initial conditions for $\vec{q}(t)$ within the Lagrangian formalism in eq. (8.1.31). While, in the Hamiltonian formulation of dynamics, only one set of initial or final conditions are needed each of the $\vec{q}(t)$ and $\vec{p}(t)$ in eq. (8.4.12); yielding a total of $2D$ set of conditions.

Problem 8.31. Lagrangian from Hamiltonian We have managed to derive Hamiltonian dynamics from its Lagrangian counterpart. Now, derive Lagrangian dynamics from its Hamiltonian counterpart; and, hence, prove their equivalence. That is, start from some $H(t, \vec{q}(t), \vec{p}(t))$ and assume Hamilton's equations in (8.4.12) hold. Then apply the definition in eq. (8.4.1), but in the form

$$L\left(t, \vec{q}(t), \dot{\vec{q}}(t)\right) = p_i(t)\dot{q}^i(t) - H\left(t, \vec{q}(t), \dot{\vec{q}}(t)\right), \tag{8.4.19}$$

where all the $\vec{p}$ has been replaced with $\vec{q}$ and $\dot{\vec{q}}$ using the $\dot{\vec{q}}$ equation in eq. (8.4.12) – assuming it is non-degenerate; namely, $\vec{p}$ may be uniquely solved in terms of $\vec{q}$ and $\dot{\vec{q}}$.

As an example: starting from

$$H(\vec{q}, \vec{p}) = \frac{\vec{p}^2}{2m} + V(\vec{q}), \tag{8.4.20}$$

use Hamilton's equations to recover eq. (8.3.4) and the definition eq. (8.4.19) to reconstruct the Lagrangian $L(\vec{q}, \dot{\vec{q}})$ in eq. (8.3.1).

Hints: Vary L the way we varied H above – first, as a function of $(t, \vec{q}, \dot{\vec{q}})$; and second, by varying its definition in eq. (8.4.19) with H written in terms of $\vec{q}$ and $\vec{p}$ so that Hamilton's equations in eq. (8.4.12) may be employed. Upon doing the latter, you should find

$$\delta L = \frac{\mathrm{d}}{\mathrm{d}t}\left\{\left(\frac{\partial L(t,\vec{q},\dot{\vec{q}})}{\partial \dot{q}^i}\right)_{t,\vec{q}} \delta q^i\right\} - \left(\frac{\partial H(t,\vec{q},\vec{p})}{\partial t}\right)_{\vec{q},\vec{p}} \mathrm{d}t. \tag{8.4.21}$$

By comparing the δL from the two routes just described, you should then be able to recover equations (8.1.31) and (8.4.9). ☐

Action for Hamilton's Equations We now turn to the action whose extremum would yield Hamilton's equations in (8.4.12):

$$S[q,p] \equiv \int_{t'}^{t} \Big(p(s)\dot{q}(s) - H\left(s, q(s), p(s)\right)\Big)\mathrm{d}s. \tag{8.4.22}$$

That is, by treating (q, p) as independent variables, we now demand that the action be stationary under both variations

$$q(t) \to q(t) + \delta q(t) \qquad \text{and} \qquad p(t) \to p(t) + \delta p(t); \tag{8.4.23}$$

with the position $q(t)$ subject to the boundary conditions in eq. (8.1.3), so that $\delta q(t) = 0 = \delta q(t')$. (No boundary conditions are necessary for the momentum p.) The variation with respect to p yields

$$\delta_{0,1} S[q,p] = \int_{t'}^{t} \delta p \left(\dot{q} - \left(\frac{\partial H}{\partial p}\right)_{t,q}\right)\mathrm{d}t''. \tag{8.4.24}$$

Whereas the variation with respect to $q(t)$ hands us

$$\delta_{1,0} S[q,p] = -\int_{t'}^{t} \delta q \left(\dot{p} + \left(\frac{\partial H}{\partial q}\right)_{t,p}\right)\mathrm{d}t''. \tag{8.4.25}$$

The setting to zero the coefficients of δp and δq in equations (8.4.24) and (8.4.25) indeed yields eq. (8.4.12).

Problem 8.32. Hamilton's Equations: Action Principle in $D-$dimensions Verify that the $D-$dimensional version of eq. (8.4.22) is

$$S[\vec{q},\vec{p}] \equiv \int_{t'}^{t} \left(p_i(s)\frac{\mathrm{d}q^i(s)}{\mathrm{d}s} - H\left(s, \vec{q}(s), \vec{p}(s)\right)\right)\mathrm{d}s. \tag{8.4.26}$$

That is, vary eq. (8.4.26) with respect to $\vec{q}$ and $\vec{p}$ to recover eq. (8.4.12). Notice, the integrands of equations (8.4.22) and (8.4.26) are nothing but the Legendre transform that yields the Lagrangian from the Hamiltonian, except it is written in terms of $(\vec{q}, \vec{p})$. In fact, it is these forms of the action that appear within the path integral formulation of quantum mechanics and quantum field theory.

Next, we turn to the types of phase space coordinate transformations – known otherwise as *canonical transformations* – that would leave this S invariant up to an additive constant. ☐

Canonical Transformations and Symmetry of Hamilton's Equations If we write Hamilton's equations in (8.4.12) in the matrix form

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} q^i(t) \\ p_a(t) \end{bmatrix} = \begin{bmatrix} 0 & (\mathbb{I}_{D\times D})^i{}_j \\ -(\mathbb{I}_{D\times D})_a{}^b & 0 \end{bmatrix}\begin{bmatrix} \partial_{q^b} \\ \partial_{p_j} \end{bmatrix} H(t,\vec{q},\vec{p}), \tag{8.4.27}$$

where 0 and $\mathbb{I}_{D\times D}$ are, respectively, the zero and identity matrix in arbitrary $D-$dimensions. By defining

$$\vec{r} \equiv (\vec{q},\vec{p})^{\mathrm{T}} \qquad \text{and} \qquad \widehat{J} \equiv \begin{bmatrix} 0 & \mathbb{I}_{D\times D} \\ -\mathbb{I}_{D\times D} & 0 \end{bmatrix}; \tag{8.4.28}$$

Hamilton's equations (8.4.27) can be expressed as

$$\dot{r}^{\mathrm{A}} = \widehat{J}^{\mathrm{AB}}\partial_{r^{\mathrm{B}}} H. \tag{8.4.29}$$

The indices A and B run over the D spatial ones of q^i and the D spatial ones of p_i; so for e.g.,

$$\widehat{J}^{\mathrm{AB}}\partial_{r^{\mathrm{B}}} = \widehat{J}^{\mathrm{A}i_q}\partial_{q^{i_q}} + \widehat{J}^{\mathrm{A}i_p}\partial_{p_{i_p}}. \tag{8.4.30}$$

If we now consider performing a transformation of the phase space coordinates $\vec{r} \equiv (q^i, p_a)$ into say $\vec{R} \equiv (Q^i(\vec{q},\vec{p}), P_a(\vec{q},\vec{p}))$; while treating the Hamiltonian H as a scalar under these transformations; then the form of Hamilton's equations (8.4.27) is preserved; namely,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} Q^i(t) \\ P_a(t) \end{bmatrix} &= \begin{bmatrix} 0 & (\mathbb{I}_{D\times D})^i{}_j \\ -(\mathbb{I}_{D\times D})_a{}^b & 0 \end{bmatrix}\begin{bmatrix} \partial_{Q^b} \\ \partial_{P_j} \end{bmatrix} H\left(t,\vec{Q},\vec{P}\right) \\ \dot{R}^{\mathrm{A}} &= \widehat{J}^{\mathrm{AB}}\partial_{R^{\mathrm{B}}} H\left(t,\vec{R}\right); \end{aligned} \tag{8.4.31}$$

if the Jacobian matrix $\partial\vec{R}/\partial\vec{r}$ itself – defined via the relation

$$\begin{bmatrix} \partial_{q^b} \\ \partial_{p_j} \end{bmatrix} = \partial_{\vec{r}} = \frac{\partial R^{\mathrm{B}}}{\partial\vec{r}}\partial_{R^{\mathrm{B}}} = \begin{bmatrix} \frac{\partial\vec{Q}}{\partial\vec{q}} & \frac{\partial\vec{P}}{\partial\vec{q}} \\ \frac{\partial\vec{Q}}{\partial\vec{p}} & \frac{\partial\vec{P}}{\partial\vec{p}} \end{bmatrix}\begin{bmatrix} \partial_{Q^b} \\ \partial_{P_j} \end{bmatrix} \tag{8.4.32}$$

– is time-independent, invertible, and obeys

$$\left(\frac{\partial\vec{R}}{\partial\vec{r}}\right)\cdot\widehat{J}\cdot\left(\frac{\partial\vec{R}}{\partial\vec{r}}\right)^{\mathrm{T}} = \widehat{J}; \tag{8.4.33}$$

or, equivalently,

$$\frac{\partial R^{\mathrm{I}}}{\partial r^{\mathrm{A}}}\cdot\widehat{J}^{\mathrm{AB}}\cdot\frac{\partial R^{\mathrm{J}}}{\partial r^{\mathrm{B}}} = \widehat{J}^{\mathrm{IJ}}. \tag{8.4.34}$$

Here, the A-th row B-th column of the matrix $\partial\vec{R}/\partial\vec{r}$ is $(\partial\vec{R}/\partial\vec{r})^{\mathrm{AB}} \equiv \partial R^{\mathrm{A}}/\partial r^{\mathrm{B}}$.

Some jargon: a Jacobian $\partial\vec{R}/\partial\vec{r}$ is *symplectic* if eq. (8.4.33) holds. Also, a given phase-space coordinate transformation is *canonical* if its Jacobian is symplectic; namely, canonical transformations yield sympletic Jacobians that leave Hamilton's equations form invariant. Finally, the set of all $2D\times 2D$ real matrices $\{\widehat{M}\}$ satisfying $\widehat{M}\cdot\widehat{J}\cdot\widehat{M}^{\mathrm{T}} = \widehat{J}$ forms the group $\mathrm{Sp}_{2D,\mathbb{R}}$.

To understand the preceding statements more explicitly, we begin from eq. (8.4.27). The chain rule tells us, as long as the transformations $(\vec{q}(t), \vec{p}(t)) = (\vec{q}(\vec{Q}, \vec{P}), \vec{p}(\vec{Q}, \vec{P}))$ do not depend explicitly on the time t; then, for instance,

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{q}\left(\vec{Q}, \vec{P}\right) = \frac{\mathrm{d}Q^i(t)}{\mathrm{d}t}\frac{\partial \vec{q}}{\partial Q^i} + \frac{\mathrm{d}P_a(t)}{\mathrm{d}t}\frac{\partial \vec{q}}{\partial P_a}. \tag{8.4.35}$$

Hence,

$$\frac{\partial r^{\mathrm{A}}}{\partial R^{\mathrm{I}}}\frac{\mathrm{d}R^{\mathrm{I}}}{\mathrm{d}t} = \widehat{J}^{\mathrm{AB}}\frac{\partial R^{\mathrm{J}}}{\partial r^{\mathrm{B}}}\partial_{R^{\mathrm{J}}} H. \tag{8.4.36}$$

Upon recognizing that $(\partial R^{\mathrm{I}}/\partial r^{\mathrm{A}})(\partial r^{\mathrm{A}}/\partial R^{\mathrm{J}}) = \delta^{\mathrm{I}}{}_{\mathrm{J}}$; we may therefore multiply both sides of eq. (8.4.36) from the left by $\partial \vec{R}/\partial \vec{r}$, and arrive at eq. (8.4.31) by imposing the symmetry condition in eq. (8.4.33).

Poisson Brackets and Canonical Transformations If we work out the $D = 1$ case of eq. (8.4.33), the Jacobian $\partial \vec{R}/\partial \vec{r}$ is sympletic iff

$$\begin{bmatrix} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial Q}{\partial q} & \frac{\partial P}{\partial q} \\ \frac{\partial Q}{\partial p} & \frac{\partial P}{\partial p} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{8.4.37}$$

$$\begin{bmatrix} 0 & \frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q} \\ -\left(\frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q}\right) & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{8.4.38}$$

We see that the transformation $\vec{r} \to \vec{r}(\vec{R})$ is canonical iff

$$\frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q} = 1. \tag{8.4.39}$$

Problem 8.33. Sympletic Transformations in D-dimensions For $D > 1$ dimensions, show that eq. (8.4.33) reads instead

$$\begin{bmatrix} \frac{\partial Q^i}{\partial q^a}\frac{\partial Q^j}{\partial p_a} - \frac{\partial Q^i}{\partial p_a}\frac{\partial Q^j}{\partial q^a} & \frac{\partial Q^i}{\partial q^a}\frac{\partial P_j}{\partial p_a} - \frac{\partial Q^i}{\partial p_a}\frac{\partial P_j}{\partial q^a} \\ -\left(\frac{\partial Q^i}{\partial q^a}\frac{\partial P_j}{\partial p_a} - \frac{\partial Q^i}{\partial p_a}\frac{\partial P_j}{\partial q^a}\right) & \frac{\partial P_i}{\partial q^a}\frac{\partial P_j}{\partial p_a} - \frac{\partial P_i}{\partial p_a}\frac{\partial P_j}{\partial q^a} \end{bmatrix} = \begin{bmatrix} 0 & \delta^i{}_j \\ -\delta^i{}_j & 0 \end{bmatrix}. \tag{8.4.40}$$

These considerations motivate us to define, for spatial dimension $D \geq 1$, the Poisson bracket of two arbitrary scalar functions $f(t, \vec{q}, \vec{p})$ and $g(t, \vec{q}, \vec{p})$ as

$$\{f, g\}_{\vec{q},\vec{p}} \equiv \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}, \tag{8.4.41}$$

so that eq. (8.4.39) now takes the form $\{Q, P\}_{q,p} = 1$; and eq. (8.4.40)

$$\begin{bmatrix} \{Q^i, Q^j\}_{\vec{q},\vec{p}} & \{Q^i, P_j\}_{\vec{q},\vec{p}} \\ -\{Q^i, P_j\}_{\vec{q},\vec{p}} & \{P_i, P_j\}_{\vec{q},\vec{p}} \end{bmatrix} = \begin{bmatrix} 0 & \mathbb{I}_{D\times D} \\ -\mathbb{I}_{D\times D} & 0 \end{bmatrix}. \tag{8.4.42}$$

which in turn is equivalent to

$$\{Q^i, Q^j\}_{\vec{q},\vec{p}} = 0 = \{P_i, P_j\}_{\vec{q},\vec{p}} \qquad \text{and} \qquad \{Q^i, P_j\}_{\vec{q},\vec{p}} = \delta^i{}_j. \tag{8.4.43}$$

Next, verify the identities

$$\left\{Q^i, Q^j\right\}_{\vec{Q},\vec{P}} = 0 = \left\{P_i, P_j\right\}_{\vec{Q},\vec{P}} \qquad \text{and} \qquad \left\{Q^i, P_j\right\}_{\vec{Q},\vec{P}} = \delta^i{}_j. \tag{8.4.44}$$

We thus conclude:

The transformation $\vec{r} \to \vec{r}(\vec{R})$ is canonical iff the 3 distinct sets of Poisson brackets between the generalized coordinates $\vec{Q}$ and its conjugate momentum $\vec{P}$ are preserved:

$$\left\{r^{\mathrm{A}}, r^{\mathrm{B}}\right\} = \widehat{J}^{\mathrm{AB}} = \left\{R^{\mathrm{A}}, R^{\mathrm{B}}\right\}. \tag{8.4.45}$$

The canonical Poisson bracket in eq. (8.4.43) is closely related to its linear algebra commutator cousins $[X^i, X^j] = 0 = [P_i, P_j]$ and $[X^i, P_j] = i\delta^i{}_j$, where X^i is the position operator in Cartesian coordinates and P_j is the momentum. □

From equations (8.4.27) and (8.4.28), we see that the (i,j) entries of the upper-right $D \times D$ identity matrix sub-block of $\widehat{J}$ can be identified with (q^i, p_j) in that $\dot{q}^i = \widehat{J}^{ij}\partial_{p_j}H = \delta^{ij}\partial_{p_j}H$; whereas the (i,j) entries of the lower-left $D \times D$ identity matrix sub-block of $\widehat{J}$ can be identified with (p_i, q^j) in that $\dot{p}_i = \widehat{J}^{ij}\partial_{q^j}H = -\delta^{ij}\partial_{q^j}H$. We may therefore recognize that the Poisson bracket itself may be re-expressed as

$$\{f, g\}_{\vec{q},\vec{p}} = \frac{\partial f}{\partial r^{\mathrm{A}}}\widehat{J}^{\mathrm{AB}}\frac{\partial g}{\partial r^{\mathrm{B}}}. \tag{8.4.46}$$

In this form, we may readily recognize that the Poisson bracket inherits the anti-symmetric character of $\widehat{J}$; i.e., $\widehat{J}^{ij} = -\widehat{J}^{ji}$ implies

$$\{f, g\} = -\{g, f\}. \tag{8.4.47}$$

Moreover, we may now recognize that the four different sub-blocks of $\widehat{J}$ are in fact Poisson brackets of $\vec{q}$s and $\vec{p}$s.

$$\text{Upper Left Block} \qquad \left\{q^a, q^b\right\}_{\vec{q},\vec{p}} = \widehat{J}^{q^a q^b} = 0, \tag{8.4.48}$$

$$\text{Upper Right Block} \qquad \left\{q^a, p_b\right\}_{\vec{q},\vec{p}} = \widehat{J}^{q^a p_b} = \delta^a{}_b, \tag{8.4.49}$$

$$\text{Lower Left Block} \qquad \left\{p_a, q^b\right\}_{\vec{q},\vec{p}} = \widehat{J}^{p_a q^b} = -\delta_a{}^b, \tag{8.4.50}$$

$$\text{Lower Right Block} \qquad \left\{p_a, p_b\right\}_{\vec{q},\vec{p}} = \widehat{J}^{p_a p_b} = 0. \tag{8.4.51}$$

Suppressing the $\{\vec{q}, \vec{p}\}$ subscript for now, we may surmise:

$$\widehat{J} = \begin{bmatrix} \{\vec{q}, \vec{q}\} & \{\vec{q}, \vec{p}\} \\ \{\vec{p}, \vec{q}\} & \{\vec{p}, \vec{p}\} \end{bmatrix} = \begin{bmatrix} 0 & \delta^i{}_j \\ -\delta^i{}_j & 0 \end{bmatrix}. \tag{8.4.52}$$

Furthermore, for $\vec{R} \equiv (\vec{Q}, \vec{P})$, if we perform the transformation $\vec{r} \to \vec{r}(\vec{R})$, the Poisson bracket itself transforms as

$$\frac{\partial f}{\partial r^{\mathrm{A}}}\widehat{J}^{\mathrm{AB}}\frac{\partial g}{\partial r^{\mathrm{B}}} = \{f, g\}_{\vec{q},\vec{p}} = \frac{\partial f}{\partial R^{\mathrm{I}}}\frac{\partial R^{\mathrm{I}}}{\partial r^{\mathrm{A}}}\widehat{J}^{\mathrm{AB}}\frac{\partial R^{\mathrm{J}}}{\partial r^{\mathrm{B}}}\frac{\partial g}{\partial R^{\mathrm{J}}} \tag{8.4.53}$$

$$= \frac{\partial f}{\partial R^{\mathrm{I}}} \left(\left(\frac{\partial \vec{R}}{\partial \vec{r}} \right) \cdot \widehat{J} \cdot \left(\frac{\partial \vec{R}}{\partial \vec{r}} \right)^{\mathrm{T}} \right)^{\mathrm{IJ}} \frac{\partial g}{\partial R^{\mathrm{J}}}. \tag{8.4.54}$$

Recall that, the *form* of Hamilton's equations is left invariant iff eq. (8.4.33) holds. This, in turn, leads to the following observation:

> A given phase-space coordinate transformation $\vec{r} \equiv (\vec{q}, \vec{p}) \to \vec{r}(\vec{Q}, \vec{P}) \equiv \vec{r}(\vec{R})$ is canonical – it leaves the form of Hamilton's equations invariant – iff it leaves Poisson brackets invariant: $\{f, g\}_{\vec{q},\vec{p}} = \{f, g\}_{\vec{Q},\vec{P}}$.

Of course, we should not be surprised by this statement, since we have already seen above that, for the transformation to be canonical, (f, g) do not even need to be arbitrary functions of the phase space coordinates; but the *new* phase space coordinates $(\vec{Q}, \vec{P})$ themselves.

Example: Lagrangian to Hamiltonian The Lagrangian $L(t, \vec{q}, \dot{\vec{q}})$ is a scalar under arbitrary coordinate transformations $\vec{q} \to \vec{q}(\vec{Q})$.

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0 = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{Q}^i} - \frac{\partial L}{\partial Q^i} \tag{8.4.55}$$

By construction, Hamilton's equations derived from either $L(t, \vec{q}, \dot{\vec{q}})$ or $L(t, \vec{Q}, \dot{\vec{Q}})$ must take the same form; i.e., either

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i}; \tag{8.4.56}$$

or

$$\dot{Q}^i = \frac{\partial H}{\partial P_i} \qquad \text{and} \qquad \dot{P}_i = -\frac{\partial H}{\partial Q^i}. \tag{8.4.57}$$

That is, the transformation bringing $(\vec{q}, \vec{p} = \partial L/\partial \dot{\vec{q}})$ to $(\vec{Q}, \vec{P} = \partial L/\partial \dot{\vec{Q}})$ must be canonical. To check this explicitly, we first begin with the relation

$$\dot{q}^a = \frac{\mathrm{d}}{\mathrm{d}t} q^a \left(\vec{Q}(t) \right) = \frac{\partial q^a}{\partial Q^j} \frac{\mathrm{d}Q^j}{\mathrm{d}t}. \tag{8.4.58}$$

Therefore

$$\left(\frac{\partial L}{\partial \dot{Q}^i} \right)_{t,\vec{Q}} = \left(\frac{\partial L}{\partial \dot{q}^a} \right)_{t,\vec{q}} \frac{\partial}{\partial \dot{Q}^i} \left\{ \frac{\partial q^a}{\partial Q^j} \dot{Q}^j \right\}_{\vec{Q}} = \left(\frac{\partial L}{\partial \dot{q}^a} \right)_{t,\vec{q}} \frac{\partial q^a}{\partial Q^i}; \tag{8.4.59}$$

and we deduce the relation between the 'old' and 'new' momentum to be

$$P_i = \frac{\partial q^a}{\partial Q^i} p_a. \tag{8.4.60}$$

Since $\vec{q} \to \vec{q}(\vec{Q})$ depends on $\vec{Q}$ but not on $\vec{p}$, the Jacobian $\partial \vec{q}/\partial \vec{Q}$ does not depend on $\vec{p}$ either. This allows us to compute the following first derivatives.

$$\frac{\partial P_i}{\partial p_a} = \frac{\partial q^a}{\partial Q^i} \qquad \text{and} \qquad \frac{\partial P_i}{\partial q^a} = \frac{\partial}{\partial q^a} \left(\frac{\partial q^l}{\partial Q^i} \right) p_l = \frac{\partial Q^c}{\partial q^a} \frac{\partial q^l}{\partial Q^c \partial Q^i} p_l. \tag{8.4.61}$$

Exploiting these results, we may then compute

$$\left\{Q^i, Q^j\right\} = \frac{\partial Q^i}{\partial q^a}\frac{\partial Q^j}{\partial p_a} - \frac{\partial Q^i}{\partial p_a}\frac{\partial Q^j}{\partial q^a} = 0 \tag{8.4.62}$$

$$\left\{P_i, P_j\right\} = \frac{\partial}{\partial q^a}\left(\frac{\partial q^l}{\partial Q^i}\right) p_l \cdot \frac{\partial q^a}{\partial Q^i} - (i \leftrightarrow j) \tag{8.4.63}$$

$$= \frac{\partial^2 q^l}{\partial Q^j \partial Q^i} p_l - (i \leftrightarrow j) = 0. \tag{8.4.64}$$

Finally,

$$\left\{Q^i, P_j\right\} = \frac{\partial Q^i}{\partial q^a}\frac{\partial q^a}{\partial Q^j} - \frac{\partial Q^i}{\partial p_a}\frac{\partial P_j}{\partial q^a} \tag{8.4.65}$$

$$= \frac{\partial Q^i}{\partial Q^j} = \delta^i{}_j. \tag{8.4.66}$$

We have thus checked that $(\partial R^{\mathrm{I}}/\partial r^{\mathrm{A}})\widehat{J}^{\mathrm{AB}}(\partial R^{\mathrm{J}}/\partial r^{\mathrm{B}}) = \widehat{J}^{\mathrm{IJ}}$; i.e., the $(\vec{q},\vec{p})$ to $(\vec{Q},\vec{P})$ transformation is canonical.

Problem 8.34. Simple Harmonic Oscillator (SHO) The SHO is described by the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2. \tag{8.4.67}$$

Verify the following $x \to x(\phi, \rho)$ and $p \to p(\phi, \rho)$ are canonical transformations, with ϕ being 'position' and ρ 'momentum':

$$x = \sqrt{\frac{2\rho}{m\omega}}\cos\phi \qquad \text{and} \qquad p = \sqrt{2\rho \cdot m\omega}\sin\phi. \tag{8.4.68}$$

' One way to show this is to simply verify:

$$\{f(x,p), g(x,p)\}_{\phi,\rho} = \partial_x f \partial_p g - \partial_p f \partial_x g. \tag{8.4.69}$$

In terms of the new coordinates, further verify that

$$H = \omega\rho, \tag{8.4.70}$$

$$\dot{\phi} = \omega \qquad \text{and} \qquad \dot{\rho} = 0. \tag{8.4.71}$$

These calculations tell us, the solutions to the SHO with fixed energy E sweeps out circles on the (x,p) phase space with the angular velocity given by ω and radius proportional to $\sqrt{\rho} = \sqrt{E/\omega}$. The pair (ϕ, ρ) is an example of angle-action variables, where ρ the conjugate momentum to ϕ is constant and $\dot{\phi}$ itself is $\phi-$independent. □

Problem 8.35. Verify the following properties of the Poisson bracket.

Linearity It is linear. For all constants α and β,

$$\{\alpha \cdot f + \beta g, h\} = \alpha\{f, h\} + \beta\{g, h\}. \tag{8.4.72}$$

Product Rule It obeys the 'product rule':

$$\{f \cdot g, h\} = \{f, h\}g + f\{g, h\}. \tag{8.4.73}$$

Jacobi Identity Finally, it satisfies the Jacobi identity

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0. \tag{8.4.74}$$

Warning: This is a tedious calculation.

The properties of the Poisson bracket in equations (8.4.47), (8.4.72), (8.4.73), and (8.4.74) have parallels to those of the commutator. □

Since canonical transformations leave Hamilton's equations form-invariant, and since Hamilton's equations follow from extremizing the action S in eq. (8.4.26), canonical transformations must therefore leave the action S invariant up to an additive constant. Let us now show that is indeed the case at least for $D = 1$ by proving that the object

$$\Delta_H(q, p) \equiv p\mathrm{d}q - P\mathrm{d}Q \tag{8.4.75}$$

is a pure gradient

$$\Delta_H = \partial_q\Sigma(q, p)\mathrm{d}q + \partial_p\Sigma(q, p)\mathrm{d}p \tag{8.4.76}$$

iff the Poisson bracket in eq. (8.4.43) holds. For, we may transform $Q \to Q(q, p)$ and see that

$$\begin{aligned} \Delta_H(q, p) &= (p - P\partial_q Q)\mathrm{d}q - P\partial_p Q\mathrm{d}p && (8.4.77) \\ &\equiv V_q\mathrm{d}q + V_p\mathrm{d}p. && (8.4.78) \end{aligned}$$

By the 2D version of the Poincaré lemma invoked above for proving the (non-)uniqueness of the Lagrangian, we may say that (V_q, V_p) is a pure gradient iff its 'curl' is zero:

$$\begin{aligned} \partial_p V_q - \partial_q V_p &= (1 - \partial_p P\partial_q Q - P\partial_p\partial_q Q) + (\partial_q P\partial_p Q + P\partial_q\partial_p Q) && (8.4.79) \\ \Leftrightarrow \qquad 1 &= \partial_q Q\partial_p P - \partial_p Q\partial_q P. && (8.4.80) \end{aligned}$$

You will prove the $D-$dimensional version of this statement in Problem (9.101) below. Here, we simply record:

$$p_i\mathrm{d}q^i - P_i\mathrm{d}Q^i = \partial_{q^i}\Sigma(\vec{q}, \vec{p})\mathrm{d}q^i + \partial_{p_i}\Sigma(\vec{q}, \vec{p})\mathrm{d}p_i \tag{8.4.81}$$

iff the canonical Poisson brackets in eq. (8.4.43) is valid. Furthermore, we then have, under such a canonical transformation,

$$\begin{aligned} S &= \int_{t'}^{t} \left(P_i(s)\dot{Q}^i(s) - H\left(s, \vec{q}(\vec{Q}, \vec{P}), \vec{p}(\vec{Q}, \vec{P})\right)\right)\mathrm{d}s + \int_{t'}^{t} \frac{\mathrm{d}\Sigma}{\mathrm{d}t}\mathrm{d}t && (8.4.82) \\ &= \int_{t'}^{t} \left(P_i(s)\dot{Q}^i(s) - H\left(s, \vec{Q}, \vec{P}\right)\right)\mathrm{d}s + \left[\Sigma(\vec{q}(s), \vec{p}(s))\right]_{s=t'}^{s=t}; && (8.4.83) \\ H\left(s, \vec{Q}, \vec{P}\right) &\equiv H\left(s, \vec{q}(\vec{Q}, \vec{P}), \vec{p}(\vec{Q}, \vec{P})\right). && (8.4.84) \end{aligned}$$

It should be possible to arrange boundary conditions such that the Σ remains a constant and the integral itself is extremized with respect to both $\vec{Q}$ and $\vec{P}$.

Problem 8.36. Infinitesimal Time Evolution Generates Canonical Transformations Consider the infinitesimal time development of $(\vec{q}, \vec{p})$; i.e.,

$$(\vec{q}(t), \vec{p}(t)) \to \left(\vec{Q}(t), \vec{P}(t)\right) \equiv (\vec{q}(t+\mathrm{d}t), \vec{p}(t+\mathrm{d}t)) \tag{8.4.85}$$

$$= (\vec{q}(t), \vec{p}(t)) + (\dot{\vec{q}}(t), \dot{\vec{p}}(t))\mathrm{d}t + \mathcal{O}\left(\mathrm{d}t^2\right). \tag{8.4.86}$$

Assuming these $(\vec{q}(t), \vec{p}(t))$ obey Hamilton's equations, prove that $(\vec{Q}(\vec{q}, \vec{p}), \vec{P}(\vec{q}, \vec{p}))$ defines an infinitesimal canonical transformation. □

Problem 8.37. Space- and Momentum-Translations Verify the following 'kinematical' Poisson bracket results:

$$\{p_i, f(t, \vec{q}, \vec{p})\} = -\partial_{q_i} f(t, \vec{q}, \vec{p}), \tag{8.4.87}$$

$$\left\{q^i, f(t, \vec{q}, \vec{p})\right\} = \partial_{p_i} f(t, \vec{q}, \vec{p}). \tag{8.4.88}$$

The first result may be interpreted as a small displacement in momentum space; and the second as one in position space. That is, under the phase phase displacement $\vec{q} \to \vec{q}+\mathrm{d}\vec{q}$ and $\vec{p} \to \vec{p}+\mathrm{d}\vec{p}$, we may assert

$$f(\vec{q}, \vec{p}) \to f(\vec{q}, \vec{p}) - \{p_i, f(\vec{q}, \vec{p})\} \,\mathrm{d}q^i + \left\{q^i, f(\vec{q}, \vec{p})\right\} \mathrm{d}p_i + \mathcal{O}\left((\mathrm{d}q)^2, (\mathrm{d}p)^2, \mathrm{d}q\mathrm{d}p\right). \tag{8.4.89}$$

Since Poisson brackets implement differentiation, for finite but constant displacements $(\vec{q}, \vec{p}) \to (\vec{q}+\vec{\xi}, \vec{p}+\vec{\Pi})$, the ensuing Taylor expansion may be written as a sum over nested Poisson brackets:

$$\begin{aligned} & f(t, \vec{q}+\vec{\xi}, \vec{p}+\vec{\Pi}) \\ & \quad = f(t, \vec{q}, \vec{p}) + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \left(\xi^i \partial_{q^i} + \Pi^i \partial_{p_i}\right)^\ell f(t, \vec{q}, \vec{p}) \qquad (8.4.90) \\ & \quad = f(t, \vec{q}, \vec{p}) + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \left\{-\xi^{i_\ell} p_{i_\ell} + \Pi^{i_\ell} q^{i_\ell}, \left\{\dots \left\{-\xi^{i_1} p_{i_1} + \Pi^{i_1} q^{i_1}, f(t, \vec{q}, \vec{p})\right\} \dots\right\}\right\}. \qquad (8.4.91) \end{aligned}$$

In the second line, the linearity of Poisson brackets in eq. (8.4.72) was employed.

Time-Evolution Next, use Hamilton's equations (8.4.12) to show that, for an arbitrary function $f(t, \vec{q}, \vec{p})$,

$$\frac{\mathrm{d}}{\mathrm{d}t} f\left(t, \vec{q}(t), \vec{p}(t)\right) = \left(\frac{\partial f}{\partial t}\right)_{\vec{q}, \vec{p}} + \{f, H\}. \tag{8.4.92}$$

In particular, if f does not depend explicitly on time, that means under finite time evolution $t \to t + \Delta t$, the $f\left(\vec{q}(t), \vec{p}(t)\right)$ will evolve into

$$\begin{aligned} & f\left(\vec{q}(t+\Delta t), \vec{p}(t+\Delta t)\right) \\ & \quad = f\left(\vec{q}(t), \vec{p}(t)\right) + \sum_{\ell=1}^{+\infty} \frac{\Delta t^\ell}{\ell!} \underbrace{\Big\{ \dots \Big\{ \Big\{ f\left(\vec{q}(t), \vec{p}(t)\right), H \Big\}, H \Big\} \dots, H \Big\}}_{\ell \text{ Poisson brackets}}. \qquad (8.4.93) \end{aligned}$$

Whereas eq. (8.4.87) and (8.4.88) do not depend on H and hence are 'kinematical', this result is dynamical: it tells us how f would evolve whenever the $(\vec{q}, \vec{p})$ satisfy Hamilton's equations. There is even a quantum mechanical version of eq. (8.4.92), where the $(\vec{q}, \vec{p})$ are promoted to operators while the Poisson bracket $\{\cdot, \cdot\}$ is replaced with the commutator $(1/i)[\cdot, \cdot]$. □

Conserved quantities Eq. (8.4.92) is an important result because it allows us to identify quantities that are conserved under time evolution governed by Hamilton's equations (8.4.12). For a start, let us observe that since the Poisson bracket is anti-symmetric, the Poisson bracket of H with itself must necessarily be zero: $\{H, H\} = 0$. Therefore, if H itself does not depend explicitly on time, it must be a conserved quantity – i.e., energy itself:

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\vec{q}, \vec{p}) = \frac{\partial H}{\partial t} = 0. \tag{8.4.94}$$

Furthermore, if q^k is an ignorable (or, cyclic) coordinate of some Lagrangian L, then the Hamiltonian $H = \dot{q}^i (\partial L/\partial \dot{q}^i) - L$ is necessarily independent of q^k too. Thus, its conjugate momentum must be conserved by Hamilton's equation:

$$\dot{p}_k = -\frac{\partial H}{\partial q^k} = 0. \tag{8.4.95}$$

According to eq. (8.4.92) we may also express this as

$$\dot{p}_k = \{p_k, H\} = 0. \tag{8.4.96}$$

Observe that, if some dynamical variables A and B are constants of motion – and if H itself is time-independent – then

$$\{A, H\} = 0 = \{B, H\}. \tag{8.4.97}$$

Furthermore, by the Jacobi identity in eq. (8.4.74),

$$\{A, \{B, H\}\} + \{B, \{H, A\}\} + \{H, \{A, B\}\} = \{H, \{A, B\}\} = 0. \tag{8.4.98}$$

That is, if A and B are conserved; so is $\{A, B\}$.

Previously, we saw that linear momentum was conserved due to space-translation symmetry. For the 2-body problem, for instance, a space-translation symmetric Lagrangian would have a potential that depends only on the relative displacement $\vec{\Delta} \equiv \vec{x}_1 - \vec{x}_2$ and not on their sum $\vec{x}_1 + \vec{x}_2$. In fact, referring to eq. (8.3.44), we may deduce its Hamiltonian to be

$$H_{2\mathrm{B}} = \frac{\vec{P}_{\mathrm{CM}}^2}{2(M_1 + M_2)} + \frac{\vec{P}_\mu^2}{2\mu} + V\left(|\vec{\Delta}|\right). \tag{8.4.99}$$

That tells us $\vec{X}_{\mathrm{CM}}$ is a cyclic coordinate and $\vec{P}_{\mathrm{CM}}$ is a constant-of-motion.

Problem 8.38. Angular Momentum Conservation If $\vec{L} \equiv \vec{\Delta} \times \vec{P}_\mu$, verify that

$$\left\{\vec{L}, H_{2\mathrm{B}}\right\} = 0; \tag{8.4.100}$$

and therefore angular momentum is conserved. You may assume that $\{\Delta^i, P_\mu^j\} = \delta^{ij}$. □

Problem 8.39. Lie Algebra of SO_3 Next, verify the following analog of the Lie algebra of SO_3. If L^i is the ith Cartesian component of the angular momentum,

$$\left\{L^a, L^b\right\} = i\epsilon^{abc} L^c, \tag{8.4.101}$$

$$\left\{\vec{L}^2, L^a\right\} = 0, \qquad \vec{L}^2 \equiv L^b L^b; \tag{8.4.102}$$

where $\epsilon^{123} \equiv 1$. Again, you may assume the canonical relation $\{\Delta^i, P_\mu^j\} = \delta^{ij}$. □

Symmetry We now define a symmetry to be one generated by a canonical transformation such that the Hamiltonian remains invariant. For example, in the $H_{2\mathrm{B}}$ above, displacing $\vec{X}_{\mathrm{CM}} \to \vec{X}_{\mathrm{CM}} + \vec{a}$ for any constant $\vec{a}$ leaves the Hamiltonian invariant; so does the simultaneous replacements $P^i_\mu \to \widehat{R}^{ij} P^j_\mu$ and $\Delta^i \to \widehat{R}^{ij}\Delta^j$ for some time independent rotation matrix $\widehat{R}$.

For infinitesimal transformation

$$\vec{q} \to \vec{q} + \delta\vec{q}, \tag{8.4.103}$$
$$\vec{p} \to \vec{p} + \delta\vec{p}; \tag{8.4.104}$$

the Hamiltonian goes as $H \to H + \delta_1 H + \dots$, where the first order perturbation may be computed via Taylor expansion as

$$\delta_1 H = \partial_{q^i} H \cdot \delta q^i + \partial_{p_i} H \cdot \delta p_i. \tag{8.4.105}$$

In Problem (9.101) below, the 'small' displacements $\delta\vec{q}$ and $\delta\vec{p}$ are canonical transformation, up to first order in $\delta\vec{q}$ and $\delta\vec{p}$, iff they take the following general 'pure gradients' form

$$\delta q^i = \partial_{p_i}\left(A(\vec{q},\vec{p}) + C_p(\vec{p})\right), \tag{8.4.106}$$
$$\delta p_i = -\partial_{q^i}\left(A(\vec{q},\vec{p}) - C_q(\vec{q})\right), \tag{8.4.107}$$

where C_q and C_p only depends on $\vec{q}$ and $\vec{p}$ respectively. Therefore,

$$\delta_1 H = \partial_{q^i} H \partial_{p_i}\left(A(\vec{q},\vec{p}) + C_p(\vec{p})\right) - \partial_{p_i} H \partial_{q^i}\left(A(\vec{q},\vec{p}) - C_q(\vec{q})\right) \tag{8.4.108}$$
$$= \{H, A\} + \partial_{q^i} H \partial_{p_i} C_p(\vec{p}) + \partial_{p_i} H \partial_{q^i} C_q(\vec{q}). \tag{8.4.109}$$

If the $\vec{r}$ satisfies Hamilton's equations,

$$\delta H = -\frac{\mathrm{d}A}{\mathrm{d}t} - \frac{\mathrm{d}p_i}{\mathrm{d}t}\partial_{p_i} C_p(\vec{p}) + \frac{\mathrm{d}q^i}{\mathrm{d}t}\partial_{q^i} C_q(\vec{q}) \tag{8.4.110}$$
$$= -\frac{\mathrm{d}}{\mathrm{d}t}\left(A(\vec{q},\vec{p}) + C_p(\vec{p}) - C_q(\vec{q})\right). \tag{8.4.111}$$

To sum: If the canonical transformation $\vec{r} \to \vec{r} + \delta\vec{r}$ (cf. equations (8.4.106) and (8.4.107)) generates an infinitesimal symmetry – it leaves H invariant up to first order – then $A(\vec{q},\vec{p}) + C_p(\vec{p}) - C_q(\vec{q})$ must be a constant-of-motion.[73]

Ostrogradski: Why is Newton's 2nd Law 2nd order in time? Just as we considered the higher derivatives version of the Lagrangian formulation in Problem (8.1), we may seek an analogous one for Hamiltonian dynamics. This will lead us to an insight regarding why Newton's second law is a differential equation involving only up to two – but no higher – time derivatives.

The starting point for us will be the Lagrangian that now depends on $\vec{q}$, $\dot{\vec{q}}$, $\dots$, $\vec{q}^{(n)} \equiv \mathrm{d}^n\vec{q}/\mathrm{d}t^n$.

$$S = \int_{t'}^{t} L\left(s, \vec{q}(s), \dot{\vec{q}}(s), \dots, \vec{q}^{(n)}(s)\right)\mathrm{d}s \tag{8.4.112}$$

[73] Actually, invariance of H under canonical transformation is a tad too strict for conserved quantities to be found. Rather, H only needs to go as $H \to H + \mathrm{d}\Psi/\mathrm{d}t$ for some scalar Ψ. Then, the corresponding conserved quantity under the infinitesimal canonical transformations would be $A + C_p - C_q - \Psi$.

To require that the dynamics – solution to $\vec{q}$ – extremize S will require $2n$ boundary conditions. The Euler-Lagrange equations now read (cf. eq. (8.1.28)):

$$\frac{\partial L}{\partial \vec{q}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \vec{q}^{(1)}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 \frac{\partial L}{\partial \vec{q}^{(2)}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-1} \frac{\partial L}{\partial \vec{q}^{(n-1)}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n} \frac{\partial L}{\partial \vec{q}^{(n)}} = 0. \tag{8.4.113}$$

Ostrogradski's Hamiltonian dynamics for such a higher derivative theory is then formulated as follows.[74] First define the A-th $\vec{Q}$ variable as

$$\vec{Q}_{(\mathrm{A})} \equiv \frac{\mathrm{d}^{\mathrm{A}-1}\vec{q}(t)}{\mathrm{d}t^{\mathrm{A}-1}} \equiv \vec{q}^{(\mathrm{A}-1)}, \tag{8.4.114}$$

$$\mathrm{A} \in \{1, 2, \dots, n\}; \tag{8.4.115}$$

so that there are n $\vec{Q}$s in total: $\vec{Q}_{(1)} = \vec{q}$, $\vec{Q}_{(2)} = \dot{\vec{q}}$, $\dots$, $\vec{Q}_{(n)} = \vec{q}^{(n-1)}$. As for their corresponding conjugate momentum variables,

$$\vec{P}^{(\mathrm{A})} \equiv \frac{\partial L}{\partial \vec{q}^{(\mathrm{A})}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right) \frac{\partial L}{\partial \vec{q}^{(\mathrm{A}+1)}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 \frac{\partial L}{\partial \vec{q}^{(\mathrm{A}+2)}} + \dots + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-\mathrm{A}} \frac{\partial L}{\partial \vec{q}^{(n)}}; \tag{8.4.116}$$

with A again running from 1 through n. For instance, the 1st and 2nd momenta are

$$\vec{P}^{(1)} \equiv \frac{\partial L}{\partial \dot{\vec{q}}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right) \frac{\partial L}{\partial \ddot{\vec{q}}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 \frac{\partial L}{\partial \vec{q}^{(3)}} + \dots + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-1} \frac{\partial L}{\partial \vec{q}^{(n)}}, \tag{8.4.117}$$

$$\vec{P}^{(2)} \equiv \frac{\partial L}{\partial \ddot{\vec{q}}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right) \frac{\partial L}{\partial \vec{q}^{(3)}} + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 \frac{\partial L}{\partial \vec{q}^{(4)}} + \dots + \left(-\frac{\mathrm{d}}{\mathrm{d}t}\right)^{n-2} \frac{\partial L}{\partial \vec{q}^{(n)}}; \tag{8.4.118}$$

whereas the $(n-1)$th and nth momenta are

$$\vec{P}^{(n-1)} = \frac{\partial L}{\partial \vec{q}^{(n-1)}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \vec{q}^{(n)}} \tag{8.4.119}$$

$$\text{and} \qquad \vec{P}^{(n)} = \frac{\partial}{\partial \vec{q}^{(n)}} L\left(t, \vec{Q}_{(1)}, \vec{Q}_{(2)}, \dots, \vec{Q}_{(n)}, \vec{q}^{(n)}\right). \tag{8.4.120}$$

We have used eq. (8.4.114) in eq. (8.4.120) to replace the first $n-1$ derivatives of $\vec{q}$ with their corresponding $\vec{Q}$s. This then brings us to the key 'no degeneracy' assumption of Ostrogradski: that the highest time derivative $\vec{q}^{(n)}$ may be solved in terms of $\vec{Q}_{(1)} = \vec{q}$, $\vec{Q}_{(2)} = \dot{\vec{q}}$, $\dots$, $\vec{Q}_{(n)} = \vec{q}^{(n-1)}$ and $\vec{P}^{(n)}$ through eq. (8.4.120). The Hamiltonian for such a higher derivative system is then defined as

$$\begin{aligned} &H\left(t, \vec{Q}_{(1)}, \dots, \vec{Q}_{(n)}, \vec{P}^{(1)}, \dots, \vec{P}^{(n)}\right) \\ &\qquad \equiv \vec{P}^{(\mathrm{A})} \cdot \dot{\vec{Q}}_{(\mathrm{A})} - L\left(t, \vec{Q}_{(1)}, \dots, \vec{Q}_{(n)}, \vec{q}^{(n)}\left(\vec{Q}_{(1)}, \dots, \vec{Q}_{(n)}, \vec{P}^{(n)}\right)\right); \end{aligned} \tag{8.4.121}$$

where the $\cdot$ is the usual Euclidean dot product and the A in $\vec{P}^{(\mathrm{A})} \cdot \dot{\vec{Q}}_{(\mathrm{A})}$ runs from 1 to n so that

$$\vec{P}^{(\mathrm{A})} \cdot \dot{\vec{Q}}_{(\mathrm{A})} = \vec{P}^{(1)} \cdot \vec{Q}_{(2)} + \dots + \vec{P}^{(n-1)} \cdot \vec{Q}_{(n)} + \vec{P}^{(n)} \cdot \vec{q}^{(n)}\left(\vec{Q}_{(1)}, \dots, \vec{Q}_{(n)}, \vec{P}^{(n)}\right), \tag{8.4.122}$$

[74] This section follows Woodard [20], which I recommend the reader to read for further details and examples; see also [21].

because $\mathrm{d}\vec{Q}_{(\mathrm{A})}/\mathrm{d}t = \mathrm{d}\vec{q}^{(\mathrm{A}-1)}/\mathrm{d}t = \vec{Q}_{(\mathrm{A}+1)}$. In terms of these $\vec{Q}$s and $\vec{P}$s, Hamilton's equations are

$$\frac{\mathrm{d}\vec{Q}_{(\mathrm{A})}}{\mathrm{d}t} = \frac{\partial H}{\partial \vec{P}^{(\mathrm{A})}} \qquad \text{and} \qquad \frac{\mathrm{d}\vec{P}^{(\mathrm{A})}}{\mathrm{d}t} = -\frac{\partial H}{\partial \vec{Q}_{(\mathrm{A})}}. \tag{8.4.123}$$

Problem 8.40. Verify that the Ostrogradsky-Hamilton equations (8.4.123) implies the Euler-Lagrange ones in eq. (8.4.113). □

Ostrogradsky's Instability One of the main observations regarding Ostrogradsky's Hamiltonian H in eq. (8.4.121) is that the Lagrangian on the right hand side is independent of the first $n-1$ momenta, $\vec{P}^{(1)}$ through $\vec{P}^{(n-1)}$, and therefore the Hamiltonian itself is in fact *linear* in them due to the first group of terms $\vec{P}^{(\mathrm{A})} \cdot \dot{\vec{Q}}_{(\mathrm{A})}$ – see eq. (8.4.122). Now, whenever H itself is time-independent, as is usually the case when the system is completely isolated, then H is a constant-of-motion when evaluated on the solutions to the $\vec{Q}$s and $\vec{P}$s. This 'on-shell' H may therefore be identified as 'energy'. However, because H is linear in the first $n-1$ momenta, this 'energy' no longer has a lower limit. For a given physical system, as long as its $\vec{P}^{(1)}$ through $\vec{P}^{(n-1)}$ may be decreased as much as one desires, H itself can then be lowered arbitrarily. If there is only one object in the system, this may not pose an issue. But if there are two or more objects in the system, each governed by such higher momenta Hamiltonians, this would mean one object O may *increase* its 'energy' arbitrarily by absorbing it from others. That this may continue indefinitely is because, since there is no lower bound to their 'energies', the other objects may simply lower theirs and impart it to O. This describes a runaway situation, where infinite amount of positive energy may extracted by absorbing it from others. Since we do not observe such unstable behavior in Nature, we expect physical systems to admit differential equations that are at most second order in time.

Problem 8.41. Verify explicitly, that whenever L and therefore H does not depend explicitly on time, $\mathrm{d}H/\mathrm{d}t = 0$ when evaluated on the solutions to eq. (8.4.123). □

Field Redefinitions; Use post-Coulombic L?

8.5 *Dissipative Systems: A Simple Example

9 Differential Geometry of Curved Spaces

9.1 Preliminaries: Metric, Tangent Vectors and Curvature

Being fluent in the mathematics of differential geometry is mandatory if you wish to understand Einstein's General Relativity, humanity's current theory of gravity. But it also gives you a coherent framework to understand the multi-variable calculus you have previously learned. Importantly, it allows you to do calculus in *arbitrary* coordinate systems and dimensions other than the 3 spatial ones you are familiar with. In this section I will provide a practical introduction to differential geometry, and will show you how to recover from it what you have encountered in 2D/3D vector calculus. My goal here is that you will understand the subject well enough to perform concrete calculations, without worrying too much about the more abstract notions like, for e.g., what a manifold is.

I will assume you have an intuitive sense of what space means – after all, we live in it! Spacetime is simply space with an extra time dimension appended to it, although the notion of 'distance' in spacetime is a bit more subtle than that in space alone. To specify the (local) geometry of a space or spacetime means we need to understand how to express distances in terms of the coordinates we are using. For example, in Cartesian coordinates (x, y, z) and by invoking Pythagoras' theorem, the square of the distance $(\mathrm{d}\ell)^2$ between (x, y, z) and $(x+\mathrm{d}x, y+\mathrm{d}y, z+\mathrm{d}z)$ in flat (aka Euclidean) space is

$$(\mathrm{d}\ell)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2. \tag{9.1.1}$$

[75]As already alluded to, a significant amount of machinery in differential geometry involves understanding how to employ arbitrary coordinate systems – and switching between different ones. For instance, we may convert the Cartesian coordinates flat space of eq. (9.1.1) into spherical coordinates,

$$(x, y, z) \equiv r\left(\sin\theta\cdot\cos\phi, \sin\theta\cdot\sin\phi, \cos\theta\right), \tag{9.1.2}$$

and find

$$(\mathrm{d}\ell)^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2). \tag{9.1.3}$$

The geometries in eq. (9.1.1) and eq. (9.1.3) are exactly the same. All we have done is to express them in different coordinate systems.

Conventions This is a good place to (re-)introduce the Einstein summation and the index convention. First, instead of (x, y, z), we can instead use $x^i \equiv (x^1, x^2, x^3)$; here, the superscript does not mean we are raising x to the first, second and third powers. A derivative

[75]In 4-dimensional flat space*time*, with time t in addition to the three spatial coordinates $\{x, y, z\}$, the infinitesimal distance is given by a modified form of Pythagoras' theorem: $\mathrm{d}s^2 \equiv (\mathrm{d}t)^2 - (\mathrm{d}x)^2 - (\mathrm{d}y)^2 - (\mathrm{d}z)^2$. (The opposite sign convention, i.e., $\mathrm{d}s^2 \equiv -(\mathrm{d}t)^2 + (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2$, is also equally valid.) Why the "time" part of the distance differs in sign from the "space" part of the metric would lead us to a discussion of the underlying Lorentz symmetry. Because I wish to postpone the latter for the moment, I will develop differential geometry for curved spaces, not curved spacetimes. Despite this restriction, rest assured most of the subsequent formulas do carry over to curved spacetimes by simply replacing Latin/English alphabets with Greek ones – see the "Conventions" paragraph below.

with respect to the ith coordinate is $\partial_i \equiv \partial/\partial x^i$. The advantage of such a notation is its compactness: we can say we are using coordinates $\{x^i\}$, where $i \in \{1,2,3\}$.[76] Not only that, we can employ Einstein's summation convention, which says all repeated indices are automatically summed over their relevant range. For example, eq. (9.1.1) now reads:

$$(\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2 = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \equiv \sum_{1\leq i,j\leq 3} \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{9.1.4}$$

(We say the indices of the $\{\mathrm{d}x^i\}$ are being contracted with those of δ_{ij}.) The symbol δ_{ij} is known as the Kronecker delta, defined as

$$\delta_{ij} = 1, \qquad i = j, \tag{9.1.5}$$
$$= 0, \qquad i \neq j. \tag{9.1.6}$$

Of course, δ_{ij} is simply the ij component of the identity matrix. Already, we can see δ_{ij} can be readily defined in an arbitrary D dimensional space, by allowing i,j to run from 1 through D. With these conventions, we can re-express the change of variables from eq. (9.1.1) and eq. (9.1.3) as follows. First write

$$\xi^i \equiv (r, \theta, \phi); \tag{9.1.7}$$

which are subject to the restrictions

$$r \geq 0, \qquad 0 \leq \theta \leq \pi, \qquad \text{and} \qquad 0 \leq \phi < 2\pi. \tag{9.1.8}$$

Next, remember the chain rule

$$\mathrm{d}x^i = \frac{\partial x^i}{\partial \xi^a}\mathrm{d}\xi^a. \tag{9.1.9}$$

Then (9.1.1) becomes

$$\delta_{ab}\mathrm{d}x^a\mathrm{d}x^b = \delta_{ab}\frac{\partial x^a}{\partial \xi^i}\frac{\partial x^b}{\partial \xi^j}\mathrm{d}\xi^i\mathrm{d}\xi^j = \frac{\partial \vec{x}}{\partial \xi^i}\cdot\frac{\partial \vec{x}}{\partial \xi^j}\mathrm{d}\xi^i\mathrm{d}\xi^j, \tag{9.1.10}$$

where in the second equality we have, for convenience, expressed the contraction with the Kronecker delta as an ordinary (vector calculus) dot product. At this point, let us notice, if we call the coefficients of the quadratic form g_{ij}; for example, $\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \equiv g_{ij}\mathrm{d}x^i\mathrm{d}x^j$, we have

$$g_{i'j'}(\vec{\xi}) = \frac{\partial \vec{x}}{\partial \xi^i}\cdot\frac{\partial \vec{x}}{\partial \xi^j}. \tag{9.1.11}$$

The primes on the indices are there to remind us this equation is not $g_{ij}(\vec{x}) = \delta_{ij}$, the components written in the Cartesian coordinates, but rather the ones written in spherical coordinates $\vec{\xi} =$

[76] It is common to use the English alphabets to denote space coordinates and Greek letters to denote spacetime ones. We will adopt this convention, but note that it is not a universal one; so be sure to check the notation of whatever you are reading.

(r, θ, ϕ). In fact, eq. (9.1.11) really holds for transforming flat space in Cartesian coordinates to *any* curvilinear coordinates $\{\vec{\xi}\}$. Moreover for $g_{ij}(\vec{x}) = \delta_{ij}$, what we are finding in eq. (9.1.10) is

$$g_{i'j'}(\vec{\xi}) = g_{ab}(\vec{x}) \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j}. \tag{9.1.12}$$

Let's proceed to work out the above dot products. Firstly,

$$\frac{\partial \vec{x}}{\partial r} = (\sin\theta \cdot \cos\phi, \sin\theta \cdot \sin\phi, \cos\theta) \equiv \widehat{r}, \tag{9.1.13}$$
$$\frac{\partial \vec{x}}{\partial \theta} = r\,(\cos\theta \cdot \cos\phi, \cos\theta \cdot \sin\phi, -\sin\theta) \equiv r\widehat{\theta}, \tag{9.1.14}$$
$$\frac{\partial \vec{x}}{\partial \phi} = r\,(-\sin\theta \cdot \sin\phi, \sin\theta \cdot \cos\phi, 0) \equiv r\sin(\theta)\widehat{\phi}. \tag{9.1.15}$$

The $\widehat{r}$ is the unit radial vector; the $\widehat{\theta}$ is the unit vector tangent to the longitude lines; and $\widehat{\phi}$ is that tangent to the latitude lines.

Next, a direct calculation should return the results

$$g_{r\theta} = g_{\theta r} = \frac{\partial \vec{x}}{\partial r} \cdot \frac{\partial \vec{x}}{\partial \theta} = 0, \qquad g_{r\phi} = g_{\phi r} = \frac{\partial \vec{x}}{\partial r} \cdot \frac{\partial \vec{x}}{\partial \phi} = 0, \qquad g_{\theta\phi} = g_{\phi\theta} = \frac{\partial \vec{x}}{\partial \theta} \cdot \frac{\partial \vec{x}}{\partial \phi} = 0; \tag{9.1.16}$$

and

$$g_{rr} = \frac{\partial \vec{x}}{\partial r} \cdot \frac{\partial \vec{x}}{\partial r} \equiv \left(\frac{\partial \vec{x}}{\partial r}\right)^2 = 1, \tag{9.1.17}$$
$$g_{\theta\theta} = \left(\frac{\partial \vec{x}}{\partial \theta}\right)^2 = r^2, \tag{9.1.18}$$
$$g_{\phi\phi} = \left(\frac{\partial \vec{x}}{\partial \phi}\right)^2 = r^2 \sin^2(\theta). \tag{9.1.19}$$

Altogether, these yield eq. (9.1.3).

If the $g_{ab}(\vec{x})$ in eq. (9.1.12) were not simply δ_{ab}, the coordinate transformation computation would of course not amount to merely taking dot products. Instead, we may phrase it as a matrix multiplication. Regarding $\partial x^i/\partial \xi^a$ as the ia component of the matrix $\partial x/\partial \xi$, eq. (9.1.12) is then the ij component of

$$\widehat{g}(\vec{\xi}) = \left(\frac{\partial x}{\partial \xi}\right)^{\mathrm{T}} \widehat{g}(\vec{x}) \frac{\partial x}{\partial \xi}. \tag{9.1.20}$$

Problem 9.1. Verify that the Jacobian matrix $\partial x^i/\partial(r, \theta, \phi)^a$ encountered above can be cast as the following matrix product:

$$\frac{\partial x^i}{\partial (r, \theta, \phi)^a} = \begin{bmatrix} \widehat{r} & \widehat{\theta} & \widehat{\phi} \end{bmatrix} \mathrm{diag}\,(1, r, r\sin\theta)\,, \tag{9.1.21}$$
$$\widehat{r} \equiv (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^{\mathrm{T}}, \tag{9.1.22}$$

$$\widehat{\theta} \equiv \frac{\partial \widehat{r}}{\partial \theta} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)^{\mathrm{T}}, \tag{9.1.23}$$

$$\sin\theta \cdot \widehat{\phi} \equiv \frac{\partial \widehat{r}}{\partial \phi} = \sin\theta \cdot (-\sin\phi, \cos\phi, 0)^{\mathrm{T}}. \tag{9.1.24}$$

The $\widehat{r}$, $\widehat{\theta}$, and $\widehat{\phi}$ are the unit vectors, written as 3−component columns, pointing along the respective r, θ and ϕ coordinate lines at a given point in space. Use this result to carry out the matrix multiplication in eq. (9.1.20), so as to verify that eq. (9.1.3) follows from eq. (9.1.1). □

General spatial metric In a generic curved space, the square of the infinitesimal distance between the neighboring points $\vec{x}$ and $\vec{x}+\mathrm{d}\vec{x}$, which we will continue to denote as $(\mathrm{d}\ell)^2$, is no longer given by eq. (9.1.1) – because we cannot expect Pythagoras' theorem to apply. But by scaling arguments it should still be quadratic in the infinitesimal distances $\{\mathrm{d}x^i\}$. The most general of such expression is

$$(\mathrm{d}\ell)^2 = g_{ij}(\vec{x})\mathrm{d}x^i\mathrm{d}x^j. \tag{9.1.25}$$

Since it measures distances, g_{ij} needs to be real, as long as the coordinates $\{x^i\}$ themselves are real. It is also symmetric, since any antisymmetric portion would drop out of the summation in eq. (9.1.25) anyway. (Why?) Finally, because we are discussing curved spaces for now, g_{ij} needs to have strictly positive eigenvalues.

Additionally, given g_{ij}, we can proceed to define the inverse metric g^{ij} in any real coordinate system, as the matrix inverse of g_{ij}:

$$g^{ij}g_{jl} \equiv \delta^i_{\ l} \qquad \Leftrightarrow \qquad g^{ij} \equiv (g^{-1})_{ij}. \tag{9.1.26}$$

This, in turn, implies

$$g^{ij} = g^{ji} \tag{9.1.27}$$

itself must be real and symmetric.

Everything else in a differential geometric calculation follows from the curved metric in eq. (9.1.25), once it is specified for a given setup:[77] the ensuing Christoffel symbols, Riemann/Ricci tensors, covariant derivatives/curl/divergence; what defines straight lines; parallel transportation; etc.

Distances If you are given a path $\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)$ between the points $\vec{x}(\lambda_1) = \vec{x}_1$ and $\vec{x}(\lambda_2) = \vec{x}_2$, then $\mathrm{d}\vec{x} = (\mathrm{d}\vec{x}/\mathrm{d}\lambda)\mathrm{d}\lambda$, and the distance swept out by this path is given by the integral

$$\ell = \int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} \sqrt{g_{ij}\left(\vec{x}(\lambda)\right)\mathrm{d}x^i\mathrm{d}x^j} = \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \sqrt{g_{ij}\left(\vec{x}(\lambda)\right)\frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j(\lambda)}{\mathrm{d}\lambda}}. \tag{9.1.28}$$

The $\mathrm{d}x^i/\mathrm{d}\lambda$ is an example of a *tangent vector*; it describes the 'velocity' at $\vec{x}(\lambda)$.

[77]As with most physics texts on differential geometry, we will ignore torsion in this Chapter, but will discuss it briefly in the next, §(11).

Problem 9.2. 1D Coordinate Invariance Show that the definition in eq. (9.1.28) yields an infinitesimal distance that is invariant under an arbitrary change of the parameter λ, as long as the transformation is orientation preserving. That is, suppose we replace $\lambda \to \lambda(\lambda')$ and thus $\mathrm{d}\lambda = (\mathrm{d}\lambda/\mathrm{d}\lambda')\mathrm{d}\lambda'$ – then as long as $\mathrm{d}\lambda/\mathrm{d}\lambda' > 0$, we have

$$\mathrm{d}\ell = \mathrm{d}\lambda\sqrt{g_{ij}\left(\vec{x}\right)\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}} = \mathrm{d}\lambda'\sqrt{g_{ij}\left(\vec{x}\right)\frac{\mathrm{d}x^i}{\mathrm{d}\lambda'}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda'}}; \tag{9.1.29}$$

and hence

$$\ell = \int_{\lambda'_1}^{\lambda'_2} \mathrm{d}\lambda'\sqrt{g_{ij}\left(\vec{x}(\lambda')\right)\frac{\mathrm{d}x^i(\lambda')}{\mathrm{d}\lambda'}\frac{\mathrm{d}x^j(\lambda')}{\mathrm{d}\lambda'}}, \tag{9.1.30}$$

where $\lambda(\lambda'_{1,2}) = \lambda_{1,2}$. The parameter λ is really a coordinate of the 1D path swept out by $\vec{x}(\lambda)$; parameterization invariance here simply amounts to the statement that *any* 1D coordinate may be used to describe distances/paths.

Affine Parameterization Why can we always choose λ such that

$$\sqrt{g_{ij}\left(\vec{x}(\lambda)\right)\frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j(\lambda)}{\mathrm{d}\lambda}} = a \qquad (\equiv \text{constant}), \tag{9.1.31}$$

i.e., the square root factor can be made constant along the entire path linking $\vec{x}_1$ to $\vec{x}_2$? Hint: A key step may be to explain why we may always solve λ in terms of λ' (or, vice versa) through the relation

$$\mathrm{d}\lambda\sqrt{g_{ij}\left(\vec{x}\right)\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}} = a \cdot \mathrm{d}\lambda'. \tag{9.1.32}$$

(Up to a re-scaling and a 1D translation, this amounts to using the path length itself as the parameter λ' – can you see it?) Now, suppose eq. (9.1.31) holds, explain why the square of the distance integral in eq. (9.1.28) may then be expressed as

$$\ell^2 = (\lambda_2 - \lambda_1)\int_{\lambda_1}^{\lambda_2} g_{ij}(\vec{z}(\lambda))\frac{\mathrm{d}z^i}{\mathrm{d}\lambda}\frac{\mathrm{d}z^j}{\mathrm{d}\lambda}\mathrm{d}\lambda. \tag{9.1.33}$$

Hint: Use the constancy of the square root factor to solve $\lambda_2 - \lambda_1$ in terms of ℓ. □

Tangent Vectors as Directional Derivatives In Euclidean space, we may define vectors by drawing a directed straight line between one point to another. In curved space, the notion of a 'straight line' is not straightforward, and as such we no longer try to implement such a definition of a vector. Instead, the notion of tangent vectors, and their higher rank tensor generalizations, now play central roles in curved space(time) geometry and physics. Imagine, for instance, a thin layer of water flowing over an undulating 2D surface – an example of a tangent vector on a curved space is provided by the velocity of an infinitesimal volume within the flow.

More generally, let $\vec{x}(\lambda)$ denote the trajectory swept out by an infinitesimal volume of fluid as a function of (fictitious) time λ, transversing through a $(D \geq 2)-$dimensional space. (The $\vec{x}$ need not be Cartesian coordinates.) If $f(\vec{x})$ is a function defined on this $D-$space, we may compute the time derivative of f along this trajectory via the chain rule:

$$\frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda} = \frac{\mathrm{d}f(\vec{x}(\lambda))}{\mathrm{d}\lambda} = \frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda}\partial_{x^i} f \equiv v^i(\vec{x})\partial_i f. \tag{9.1.34}$$

We may then define the tangent vector $v^i(\lambda) \equiv \mathrm{d}x^i/\mathrm{d}\lambda$. Conversely, given a vector field $v^i(\vec{x})$, i.e., a $(D \geq 2)-$component object defined at every point in space with dimensions $[v^i] = [x^i]/[\lambda]$, we may find a trajectory $\vec{x}(\lambda)$ such that $\mathrm{d}\vec{x}/\mathrm{d}\lambda = \vec{v}(\vec{x}(\lambda))$. (This amounts to integrating an ODE, and in this context is why $\vec{x}(\lambda)$ is called the *integral curve of* v^i.) In other words, tangent vectors do fit the mental picture that the name suggests, as 'little arrows' based at each point in space, describing the local 'velocity' of some (perhaps fictitious) flow.

You may readily check that tangent vectors at a given point p in space do indeed form a vector space. However, we have written the components v^i but did not explain what their basis vectors were. Geometrically speaking, v tells us in what direction and how quickly to move away from the point p. This can be formalized by recognizing that the number of independent directions that one can move away from p corresponds to the number of independent partial derivatives on some arbitrary (scalar) function defined on the curved space; namely $\partial_i f(\vec{x})$ for $i = 1, 2, \dots, D$, where $\{x^i\}$ are the coordinates used. Furthermore, the set of $\{\partial_i\}$ do span a vector space, based at p. We would thus say that any tangent vector v is a superposition of partial derivatives:

$$v \equiv v^i(\vec{x}) \frac{\partial}{\partial x^i} \equiv v^i(x^1, x^2, \dots, x^D) \frac{\partial}{\partial x^i} \equiv v^i \partial_i. \tag{9.1.35}$$

> To sum: the $\{\partial_i\}$ are the basis kets based at a given point in the curved space, allowing us to enumerate all the independent directions along which we may compute the 'time derivative' $\mathrm{d}/\mathrm{d}\lambda = v^i \partial_i$ of any f at the same point. The rightmost term of eq. (9.1.34) also indicates, this notion of 'time derivatives' is independent of the choice of coordinates $\{x^i\}$.

Tangent Vectors: Flat Space Example In flat space we may always employ Cartesian coordinates $\{\vec{x}\}$, but for many applications may choose instead to switch to some other curvilinear ones $\{\vec{\xi}\}$. Suppose now we wish to describe some trajectory $\vec{\xi}(\lambda)$ – in classical mechanics λ would be the time t – and compute, say, its first and second derivatives with respect to some λ. If we begin with $\vec{x}$ but view them as a function of $\vec{\xi}(\lambda)$,

$$\dot{x}^i \equiv \frac{\mathrm{d}x^i(\vec{\xi}(\lambda))}{\mathrm{d}\lambda} = \frac{\partial x^i}{\partial \xi^a} \frac{\mathrm{d}\xi^a}{\mathrm{d}\lambda}; \tag{9.1.36}$$

and

$$\ddot{x}^i \equiv \frac{\mathrm{d}^2 x^i(\vec{\xi}(\lambda))}{\mathrm{d}\lambda^2} = \frac{\partial x^i}{\partial \xi^a} \frac{\mathrm{d}^2 \xi^a}{\mathrm{d}\lambda^2} + \frac{\partial^2 x^i}{\partial \xi^a \partial \xi^b} \frac{\mathrm{d}\xi^a}{\mathrm{d}\lambda} \frac{\mathrm{d}\xi^b}{\mathrm{d}\lambda}. \tag{9.1.37}$$

Now $\{\partial x^i/\partial \xi^a | i = 1, \dots, D\}$, for fixed a, are simply the Cartesian components of the tangent vector parallel to the ξ^a-axis:

$$\partial_{\xi^\ell} = \frac{\partial x^i}{\partial \xi^\ell} \partial_{x^i}. \tag{9.1.38}$$

Next, define the unit-length versions of these tangent vectors $\{\partial_{\xi^\ell}\}$ as

$$\widehat{e}_\ell \equiv \widehat{e}^a_\ell \partial_{x^a} \equiv \left| \frac{\partial \vec{x}}{\partial \xi^\ell} \right|^{-1} \frac{\partial x^i}{\partial \xi^\ell} \partial_{x^i}, \tag{9.1.39}$$

$$\left|\partial\vec{x}/\partial\xi^\ell\right| \equiv \sqrt{\delta_{ab}(\partial x^a/\partial\xi^\ell)(\partial x^b/\partial\xi^\ell)}. \tag{9.1.40}$$

For simplicity, let us further assume the curvilinear coordinates $\vec{\xi}$ are orthogonal, such that

$$g_{ab}(\vec{\xi}) = \frac{\partial\vec{x}}{\partial\xi^a}\cdot\frac{\partial\vec{x}}{\partial\xi^b} = \delta_{ab}\left(\frac{\partial\vec{x}}{\partial\xi^a}\right)^2, \tag{9.1.41}$$

$$\delta_{ab}\widehat{e}^a_m\widehat{e}^b_n = \delta_{mn}. \tag{9.1.42}$$

If we expand the velocity V and acceleration A, namely

$$V \equiv \dot{x}^i\partial_{x^i} \qquad \text{and} \qquad A \equiv \ddot{x}^i\partial_{x^i}, \tag{9.1.43}$$

in terms of these orthonormal basis $\{\widehat{e}_\ell\}$,

$$V \equiv V^{\widehat{\ell}}(\vec{\xi})\widehat{e}_\ell \qquad \text{and} \qquad A \equiv A^{\widehat{\ell}}(\vec{\xi})\widehat{e}_\ell; \tag{9.1.44}$$

then since we are in flat space, we may extract the ℓth component of V in the $\vec{\xi}$ orthonormal basis $\{\widehat{e}_\ell\}$ by dot product,

$$V^{\widehat{\ell}}(\vec{\xi}) = \vec{V}\cdot\widehat{e}_\ell = \dot{x}^m\delta_{mn}\widehat{e}^n_\ell \tag{9.1.45}$$

$$= \frac{1}{|\partial\vec{x}/\partial\xi^\ell|}\frac{\mathrm{d}\xi^a}{\mathrm{d}\lambda}\left(\frac{\partial\vec{x}}{\partial\xi^a}\cdot\frac{\partial\vec{x}}{\partial\xi^\ell}\right) = \left|\frac{\partial\vec{x}}{\partial\xi^\ell}\right|\frac{\mathrm{d}\xi^\ell}{\mathrm{d}\lambda}, \tag{9.1.46}$$

where there is no sum over ℓ and we have employed eq. (9.1.41). We see from eq. (9.1.39) that

$$\left|\frac{\partial\vec{x}}{\partial\xi^\ell}\right|\widehat{e}_\ell = \sqrt{g_{\ell\ell}}\cdot\widehat{e}_\ell = \frac{\partial}{\partial\xi^\ell}; \tag{9.1.47}$$

which means the velocity itself is

$$V(\vec{\xi}) = V^{\widehat{\ell}}\widehat{e}_\ell = \sum_{\ell=1}^{D}\dot{\xi}^\ell\left|\frac{\partial\vec{x}}{\partial\xi^\ell}\right|\widehat{e}_\ell \tag{9.1.48}$$

$$= \sum_{\ell=1}^{D}\dot{\xi}^\ell\sqrt{g_{\ell\ell}}\cdot\widehat{e}_\ell = \dot{\xi}^\ell\partial_{\xi^\ell}. \tag{9.1.49}$$

Similarly, the ℓth component of the acceleration in the orthonormal basis $\{\widehat{e}_\ell\}$ is given by a similar dot product calculation,

$$A^{\widehat{\ell}}(\vec{\xi}) = \ddot{x}^m\delta_{mn}\widehat{e}^n_\ell = \frac{1}{|\partial\vec{x}/\partial\xi^\ell|}\left(\frac{\partial\vec{x}}{\partial\xi^a}\frac{\mathrm{d}^2\xi^a}{\mathrm{d}\lambda^2} + \frac{\partial^2\vec{x}}{\partial\xi^b\partial\xi^c}\frac{\mathrm{d}\xi^b}{\mathrm{d}\lambda}\frac{\mathrm{d}\xi^c}{\mathrm{d}\lambda}\right)\cdot\frac{\partial\vec{x}}{\partial\xi^\ell} \tag{9.1.50}$$

$$= \frac{g_{a\ell}}{|\partial\vec{x}/\partial\xi^\ell|}\left(\frac{\mathrm{d}^2\xi^a}{\mathrm{d}\lambda^2} + g^{af}\frac{\partial^2\vec{x}}{\partial\xi^b\partial\xi^c}\cdot\frac{\partial\vec{x}}{\partial\xi^f}\frac{\mathrm{d}\xi^b}{\mathrm{d}\lambda}\frac{\mathrm{d}\xi^c}{\mathrm{d}\lambda}\right) \tag{9.1.51}$$

$$= \left|\frac{\partial\vec{x}}{\partial\xi^\ell}\right|\left(\frac{\mathrm{d}^2\xi^\ell}{\mathrm{d}\lambda^2} + g^{\ell f}\frac{\partial^2\vec{x}}{\partial\xi^b\partial\xi^c}\cdot\frac{\partial\vec{x}}{\partial\xi^f}\frac{\mathrm{d}\xi^b}{\mathrm{d}\lambda}\frac{\mathrm{d}\xi^c}{\mathrm{d}\lambda}\right). \tag{9.1.52}$$

Defining the *symmetrization* operation

$$T_{\{ab\}} \equiv T_{ab} + T_{ba} = T_{\{ba\}} \tag{9.1.53}$$

for an arbitrary object with indices ab; let us then consider

$$\frac{1}{2}\partial_{\xi^{\{b}} g_{c\}f} - \frac{1}{2}\partial_{\xi^f} g_{bc} = \frac{1}{2}\partial_{\xi^{\{b}} \left(\frac{\partial \vec{x}}{\partial \xi^{c\}}} \cdot \frac{\partial \vec{x}}{\partial \xi^f} \right) - \frac{1}{2}\partial_{\xi^f} \left(\frac{\partial \vec{x}}{\partial \xi^b} \cdot \frac{\partial \vec{x}}{\partial \xi^c} \right) \tag{9.1.54}$$

$$= \frac{\partial \vec{x}}{\partial \xi^b \partial \xi^c} \cdot \frac{\partial \vec{x}}{\partial \xi^f} + \frac{1}{2}\frac{\partial \vec{x}}{\partial \xi^{\{b}} \cdot \frac{\partial \vec{x}}{\partial \xi^{c\}} \partial \xi^f} - \frac{1}{2}\frac{\partial \vec{x}}{\partial \xi^{\{b}} \cdot \frac{\partial \vec{x}}{\partial \xi^{c\}} \partial \xi^f} \tag{9.1.55}$$

$$= \frac{\partial^2 \vec{x}}{\partial \xi^b \partial \xi^c} \cdot \frac{\partial \vec{x}}{\partial \xi^f}. \tag{9.1.56}$$

Hence, we may express

$$g^{\ell f}(\vec{\xi}) \frac{\partial^2 \vec{x}}{\partial \xi^c \partial \xi^b} \cdot \frac{\partial \vec{x}}{\partial \xi^f} = \frac{1}{2} g^{\ell f} \left(\partial_{\xi^{\{b}} g_{c\}f}(\vec{\xi}) - \frac{1}{2}\partial_{\xi^f} g_{bc}(\vec{\xi}) \right) \equiv \Gamma^{\ell}{}_{bc}[\vec{\xi}]. \tag{9.1.57}$$

These $\{\Gamma^{\ell}{}_{bc}[\vec{\xi}]\}$ are known as *Christoffel symbols*, where the second equality holds for arbitrary metrics and coordinate systems; whereas the leftmost expression is valid for arbitrary curvilinear coordinates but in flat space. In Problem (9.40) below, you will provide a different proof of the leftmost expression of eq. (9.1.57).

At this point, the acceleration reads

$$\ddot{x}^i \partial_{x^i} = A^{\hat{\ell}} \widehat{e}_\ell = \frac{D^2 \xi^\ell}{\mathrm{d}\lambda^2} \partial_{\xi^\ell} = \sum_{\ell=1}^{D} \frac{D^2 \xi^\ell}{\mathrm{d}\lambda^2} \left| \frac{\partial \vec{x}}{\partial \xi^\ell} \right| \widehat{e}_\ell, \tag{9.1.58}$$

where – by inserting eq. (9.1.57) into eq. (9.1.52) – the acceleration components are

$$\frac{D^2 \xi^a}{\mathrm{d}\lambda^2} \equiv \ddot{\xi}^a + \Gamma^a{}_{bc} \dot{\xi}^b \dot{\xi}^c. \tag{9.1.59}$$

These $\{D^2 \xi^a / \mathrm{d}\lambda^2\}$ will turn out to be *covariant*; i.e., transforming as a vector in arbitrary curvilinear coordinates.

Integral Curve from Tangent Vector To obtain the integral curve – sometimes dubbed *lines of flow* [13] – from its tangent vector, we note that $v^i(\vec{x}) = \mathrm{d}x^i/\mathrm{d}\lambda \Leftrightarrow \mathrm{d}\lambda = \mathrm{d}x^i/v^i$ (no sum over i). This in turn implies

$$\frac{\mathrm{d}x^1}{v^1(\vec{x})} = \frac{\mathrm{d}x^2}{v^2(\vec{x})} = \dots = \frac{\mathrm{d}x^D}{v^D(\vec{x})}. \tag{9.1.60}$$

In certain circumstances, these may be integrated to determine the integral curves. Taking a 3D example from Morse and Feshbach's [13] Chapter 1, if

$$v^x = -ay, \qquad v^y = ax, \qquad v^z = b(x^2 + y^2); \tag{9.1.61}$$

then we have

$$-\frac{\mathrm{d}x}{ay} = \frac{\mathrm{d}y}{ax} = \frac{\mathrm{d}z}{b(x^2 + y^2)}. \tag{9.1.62}$$

Equating the first and second terms, $-x\mathrm{d}x = y\mathrm{d}y$, which in turn yields

$$\mathrm{d}\left(x^2 + y^2\right) = 0 \tag{9.1.63}$$
$$x^2 + y^2 = \rho^2 (\equiv \text{constant}); \tag{9.1.64}$$

whereas equating the second and third factors hand us

$$\frac{\mathrm{d}y}{a(\mathrm{sgn}(x))\sqrt{\rho^2 - y^2}} = \frac{\mathrm{d}z}{b\rho^2}. \tag{9.1.65}$$

Its integral yields, for constant z_0, the 2D surface $\arctan(y/x) = a \cdot (z - z_0)/(b\rho^2)$, whose intersection with the 2D cylinder $x^2 + y^2 = \rho^2$ then defines a flow line for a fixed pair (ρ, z_0).

Problem 9.3. Problem 1.3 of Morse & Feshbach [13] In 3D flat space parametrized by Cartesian coordinates (x, y, z), show that the integral curves to

$$v^i(x, y, z) = \left(2xz, 2yz, a^2 + z^2 - x^2 - y^2\right), \tag{9.1.66}$$

for constant a, are given by the intersection between the following 2D surfaces

$$\frac{y}{x} = \tan\varphi \qquad \text{and} \qquad \frac{x^2 + y^2 + z^2 + a^2}{\sqrt{x^2 + y^2}} = r_0, \tag{9.1.67}$$

where each flow line is labeled by the constants $\{\varphi, r_0\}$. Hints: One approach is to first re-write v^i in the spherical coordinates (r, θ, ϕ) of eq. (9.1.2). Show that

$$v = (a^2 + r^2)\cos(\theta)\partial_r + \frac{r^2 - a^2}{r}\sin(\theta)\partial_\theta. \tag{9.1.68}$$

Explain why this implies

$$\frac{r^2 + a^2}{r \cdot r_0} = \sin\theta \qquad \text{and} \qquad \phi = \varphi, \tag{9.1.69}$$

where r_0 and φ refer to some fixed radius and azimuth angle respectively. □

Parallel transport and (intrinsic) curvature Roughly speaking, a curved space is one where the usual rules of Euclidean (flat) space no longer apply. For example, Pythagoras' theorem does not hold; and the sum of the three angles of an extended triangle is not π.

The quantitative criteria to distinguish a curved space from a flat one, is to parallel transport a tangent vector $v^i(\vec{x})$ around a closed loop on a coordinate grid. If, upon bringing it back to the same location $\vec{x}$, the tangent vector is the same one we started with – *for all possible coordinate loops* – then the space is flat. Otherwise the space is curved. In particular, if you parallel transport a vector around an infinitesimal closed loop formed by two pairs of coordinate lines, starting from any one of its corners, and if the resulting vector is compared with original one, you would find that the difference is proportional to the Riemann curvature tensor $R^i{}_{jkl}$. More specifically, suppose v^i is parallel transported along a parallelogram, from $\vec{x}$ to $\vec{x} + \mathrm{d}\vec{y}$; then to

$\vec{x}+\mathrm{d}\vec{y}+\mathrm{d}\vec{z}$; then to $\vec{x}+\mathrm{d}\vec{z}$; then back to $\vec{x}$. Then, denoting the end result as v'^i, we would find that

$$v'^i - v^i \propto R^i{}_{jkl} v^j \mathrm{d}y^k \mathrm{d}z^l. \tag{9.1.70}$$

Therefore, whether or not a geometry is locally curved is determined by this tensor. Of course, we have not defined what parallel transport actually is; to do so requires knowing the covariant derivative – but let us first turn to a simple example where our intuition still holds.

Example: 2-sphere A common textbook example of a curved space is the *surface* of a sphere of some fixed radius, sitting in 3D flat space, parametrized by the usual spherical coordinates $(0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$.[78] Start at the north pole with the tangent vector $v = \partial_\theta$ pointing towards the equator with azimuthal direction $\phi = \phi_0$. Let us parallel transport v along itself, i.e., with $\phi = \phi_0$ fixed, until we reach the equator itself. At this point, the vector is perpendicular to the equator, pointing towards the South pole. Next, we parallel transport v along the equator from $\phi = \phi_0$ to some other longitude $\phi = \phi_0'$; here, v is still perpendicular to the equator, and still pointing towards the South pole. Finally, we parallel transport it back to the North pole, along the $\phi = \phi_0'$ line. Back at the North pole, v now points along the $\phi = \phi_0'$ longitude line and no longer along the original $\phi = \phi_0$ line. Therefore, v does not return to itself after parallel transport around a closed loop: the $2-$sphere is a curved surface. This same example also provides us a triangle whose sum of its internal angles is $2(\pi/2) + |\phi_0 - \phi_0'| > \pi$.[79] Finally, notice in this 2-sphere example, the question of what a straight line means – let alone using it to define a vector, as one might do in flat space – does not produce a clear answer.

Comparing tangent vectors at different locations That tangent vectors do not, in general, remain the same under parallel transportation in a curved space tells us comparing tangent vectors based at different locations is not a straightforward procedure, compared to the situation in flat space. For, if $\vec{v}(\vec{x})$ is to be compared to $\vec{w}(\vec{x}')$ by parallel transporting the former to $\vec{x}'$; different results may be obtained by simply choosing different paths to get from $\vec{x}$ to $\vec{x}'$.

Intrinsic vs Extrinsic Curvature A 2D cylinder (embedded in 3D flat space) formed by rolling up a flat rectangular piece of paper has a surface that is *intrinsically* flat – the Riemann tensor is zero everywhere because the intrinsic geometry of the surface is the same flat metric before the paper was rolled up. However, the paper as viewed by an ambient 3D observer does have an *extrinsic* curvature due to its cylindrical shape. To characterize extrinsic curvature mathematically – at least in the case where we have a $D-1$ dimensional surface situated in a D dimensional space – one would erect a vector perpendicular to the surface in question and parallel transport it along this same surface: the latter is flat if the vector remains parallel; otherwise it is curved. In curved spacetimes, when this vector refers to the flow of time and is perpendicular to some spatial surface, the extrinsic curvature also describes its time evolution.

[78] Any curved space can in fact always be viewed as a curved surface residing in a higher dimensional flat space.

[79] The $2-$sphere has positive curvature; whereas a saddle has negative curvature, and would support a triangle whose angles add up to less than π. In a very similar spirit, the Cosmic Microwave Background (CMB) sky contains hot and cold spots, whose angular size provide evidence that we reside in a spatially flat universe. See the Wilkinson Microwave Anisotropy Probe (WMAP) pages here and here.

9.2 Locally Flat Coordinates, General Tensors, Volumes

Locally flat (aka Riemann normal) coordinates and symmetries It is a mathematical fact that, given some fixed point y_0^i on the curved space, one can find coordinates y^i such that locally the metric does become flat:

$$\lim_{\vec{y}\to\vec{y}_0} g_{ij}(\vec{y}) = \delta_{ij} - \frac{1}{3}R_{ikjl}(\vec{y}_0)\ (y-y_0)^k(y-y_0)^l + \dots, \tag{9.2.1}$$

with a similar result for curved spacetimes. In this "locally flat" coordinate system, the first corrections to the flat Euclidean metric is quadratic in the displacement vector $\vec{y}-\vec{y}_0$, and $R_{ikjl}(\vec{y}_0)$ is the Riemann tensor – which is the chief measure of curvature – evaluated at $\vec{y}_0$. In a curved space*time*, that geometry can always be viewed as locally flat is why the mathematics you are encountering here is the appropriate framework for reconciling gravity as a force, Einstein's equivalence principle, and the Lorentz symmetry of Special Relativity.

Note that under spatial rotations $\{\widehat{R}^i{}_j\}$, which obeys $\widehat{R}^a{}_i\widehat{R}^b{}_j\delta_{ab} = \delta_{ij}$; and translations $\{a^i\}$; if we define in Euclidean space the following change-of-Cartesian coordinates (from $\vec{x}$ to $\vec{x}'$)

$$x^i \equiv \widehat{R}^i{}_j x'^j + a^i; \tag{9.2.2}$$

the flat metric would retain the same form

$$\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j = \delta_{ab}\widehat{R}^a{}_i\widehat{R}^b{}_j\mathrm{d}x'^i\mathrm{d}x'^j = \delta_{ij}\mathrm{d}x'^i\mathrm{d}x'^j. \tag{9.2.3}$$

A similar calculation would tell us flat Euclidean space is invariant under parity flips, i.e., $x'^k \equiv -x^k$ for some fixed k. To sum:

> At a given point in a curved space, it is always possible to find a coordinate system – i.e., a geometric viewpoint/'frame' – such that the space is flat up to distances of $\mathcal{O}(1/|\max R_{ijlk}(\vec{y}_0)|^{1/2})$, and hence 'locally' invariant under rotations, translations, and reflections.

This is why it took a while before humanity came to recognize we live on the curved surface of the (approximately spherical) Earth: locally, the Earth's surface looks flat!

'Dot Product' of Tangent Vectors This local flatness allows us to interpret the contraction of tangent vectors with the metric. For two distinct vectors v^i and w^i, if we choose to evaluate $g_{ij}v^iw^j$ in a locally flat region, it becomes an ordinary dot product in Euclidean space:

$$g_{ij}v^iw^j \to \delta_{ij}v^iw^j = \vec{v}\cdot\vec{w}. \tag{9.2.4}$$

For instance, $g_{ij}v^iv^j$ is the square of the length of $\vec{v}$; whereas

$$\frac{g_{ij}v^iw^j}{\sqrt{(g_{ab}v^av^b)(g_{mn}w^mw^n)}} \equiv \cos\theta \tag{9.2.5}$$

is the cosine of the angle between $\vec{v}$ and $\vec{w}$.

Coordinate-transforming the metric Note that, in the context of eq. (9.1.25), $\vec{x}$ is not a vector in Euclidean space, but rather another way of denoting x^a without introducing too many dummy indices $\{a, b, \dots, i, j, \dots\}$. Also, x^i in eq. (9.1.25) are not necessary Cartesian coordinates, but can be completely arbitrary. The metric $g_{ij}(\vec{x})$ can viewed as a 3×3 (or $D \times D$, in D dimensions) matrix of functions of $\vec{x}$, telling us how the notion of distance vary as one moves about in the space. Just as we were able to translate from Cartesian coordinates to spherical ones in Euclidean 3-space, in this generic curved space, we can change from $\vec{x}$ to $\vec{\xi}$, i.e., one arbitrary coordinate system to another, so that

$$g_{ij}\left(\vec{x}\right)\mathrm{d}x^i \mathrm{d}x^j = g_{ij}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^i(\vec{\xi})}{\partial \xi^a}\frac{\partial x^j(\vec{\xi})}{\partial \xi^b}\mathrm{d}\xi^a \mathrm{d}\xi^b \equiv g_{ab}(\vec{\xi})\mathrm{d}\xi^a \mathrm{d}\xi^b. \tag{9.2.6}$$

We can attribute all the coordinate transformation to how it affects the components of the metric:

$$g_{ab}(\vec{\xi}) = g_{ij}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^i(\vec{\xi})}{\partial \xi^a}\frac{\partial x^j(\vec{\xi})}{\partial \xi^b}. \tag{9.2.7}$$

The left hand side are the metric components in $\vec{\xi}$ coordinates. The right hand side consists of the Jacobians $\partial x/\partial \xi$ contracted with the metric components in $\vec{x}$ coordinates – but now with the $\vec{x}$ replaced with $\vec{x}(\vec{\xi})$, their corresponding expressions in terms of $\vec{\xi}$. Recall too, we have already noted in eq. (9.1.20) that eq. (9.2.7) may be calculated via matrix multiplication.

Inverse metric Previously, we defined g^{ij} to be the matrix inverse of the metric tensor g_{ij}. Starting with Problem (9.6) below, we shall see this in turn implies g^{ij} are the components of the tensor

$$g^{ij}(\vec{x})\partial_i \otimes \partial_j, \tag{9.2.8}$$

whose transformation properties are 'opposite' to g_{ij} with lower indices. Here, we have used $\otimes$ to indicate we are taking the tensor product of the partial derivatives ∂_i and ∂_j. In $g_{ij}\left(\vec{x}\right)\mathrm{d}x^i \mathrm{d}x^j$ we really should also have $\mathrm{d}x^i \otimes \mathrm{d}x^j$, but I prefer to stick with the more intuitive idea that the metric (with lower indices) is the sum of squares of distances. Just as we know how $\mathrm{d}x^i$ transforms under $\vec{x} \to \vec{x}(\vec{\xi})$, we also can work out how the partial derivatives transform.

$$g^{ij}(\vec{x})\frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} = g^{ab}\left(\vec{x}(\vec{\xi})\right)\frac{\partial \xi^i}{\partial x^a}\frac{\partial \xi^j}{\partial x^b}\frac{\partial}{\partial \xi^i} \otimes \frac{\partial}{\partial \xi^j} \tag{9.2.9}$$

In terms of its components, we can read off their transformation rules:

$$g^{ij}(\vec{\xi}) = g^{ab}\left(\vec{x}(\vec{\xi})\right)\frac{\partial \xi^i}{\partial x^a}\frac{\partial \xi^j}{\partial x^b}. \tag{9.2.10}$$

The left hand side is the inverse metric written in the $\vec{\xi}$ coordinate system, whereas the right hand side involves the inverse metric written in the $\vec{x}$ coordinate system – contracted with two Jacobian's $\partial \xi/\partial x$ – except all the $\vec{x}$ are replaced with the expressions $\vec{x}(\vec{\xi})$ in terms of $\vec{\xi}$.

(Inverse) Jacobians A technical point: here and below, the Jacobian $\partial x^a(\vec{\xi})/\partial \xi^j$ can be calculated in terms of $\vec{\xi}$ by direct differentiation if we have defined $\vec{x}$ in terms of $\vec{\xi}$, namely $\vec{x}(\vec{\xi})$. But the Jacobian $(\partial \xi^i/\partial x^a)$ in terms of $\vec{\xi}$ requires a matrix inversion. For, by the chain rule,

$$\frac{\partial x^i}{\partial \xi^l}\frac{\partial \xi^l}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i{}_j, \qquad \text{and} \qquad \frac{\partial \xi^i}{\partial x^l}\frac{\partial x^l}{\partial \xi^j} = \frac{\partial \xi^i}{\partial \xi^j} = \delta^i{}_j. \tag{9.2.11}$$

In other words, given $\vec{x} \to \vec{x}(\vec{\xi})$, we can compute $\mathcal{J}^a_{\ i} \equiv \partial x^a/\partial \xi^i$ in terms of $\vec{\xi}$, with a being the row number and i as the column number. Then find the inverse, i.e., $(\mathcal{J}^{-1})^a_{\ i}$ and identify it with $\partial \xi^a/\partial x^i$ in terms of $\vec{\xi}$.

Problem 9.4. Let x^i be Cartesian coordinates and

$$\xi^i \equiv (r, \theta, \phi) \tag{9.2.12}$$

be the usual spherical coordinates; see eq. (9.1.2). Calculate $\partial \xi^i/\partial x^a$ in terms of $\vec{\xi}$ and thereby, from the flat inverse metric δ^{ij} in Cartesian coordinates, find the inverse metric $g^{ij}(\vec{\xi})$ in the spherical coordinate system. You should find:

$$g^{ij}(\vec{\xi}) = \delta^{ab} \frac{\partial \xi^i}{\partial x^a} \frac{\partial \xi^j}{\partial x^b} = \text{diag}\left[1, \frac{1}{r^2}, \frac{1}{r^2 \sin(\theta)^2}\right]. \tag{9.2.13}$$

Hint: Compute $\partial x^i/\partial(r,\theta,\phi)^a$. How do you get $\partial(r,\theta,\phi)^a/\partial x^i$ from it? You may also find the factorized form of the Jacobian matrix in eq. (9.1.21) to be useful here. $\square$

Kets and Bras Just as the $\{\partial_i\}$ are the ket's, the basis tangent vectors at a given point in space, we may now identify the infinitesimal distances $\{\mathrm{d}x^i\}$ as the basis dual vectors (the bra's) through the definition

$$\left\langle \mathrm{d}x^i \middle| \partial_j \right\rangle \equiv \delta^i_{\ j}, \qquad \forall i, j. \tag{9.2.14}$$

Why this is a useful perspective is due to the following. Let us consider an infinitesimal variation of our arbitrary function at $\vec{x}$:

$$\mathrm{d}f = \partial_i f(\vec{x}) \mathrm{d}x^i. \tag{9.2.15}$$

Then, given a vector field v, we can employ eq. (9.2.14) to construct the derivative of the latter along the former, at some point $\vec{x}$, by

$$\langle \mathrm{d}f | \, v \rangle = v^j \partial_i f(\vec{x}) \left\langle \mathrm{d}x^i \middle| \partial_j \right\rangle = v^i \partial_i f(\vec{x}). \tag{9.2.16}$$

This describes a flux of sorts: if v^i were flowing parallel to the constant f surface, $\langle \mathrm{d}f | \, v \rangle$ would be zero, for instance.

What about the inner products $\langle \mathrm{d}x^i | \, \mathrm{d}x^j \rangle$ and $\langle \partial_i | \, \partial_j \rangle$? They are

$$\left\langle \mathrm{d}x^i \middle| \mathrm{d}x^j \right\rangle = g^{ij} \qquad \text{and} \qquad \langle \partial_i | \, \partial_j \rangle = g_{ij}. \tag{9.2.17}$$

This is because, for real metrics g_{ij},

$$g_{ij} \left| \mathrm{d}x^j \right\rangle \equiv |\partial_i\rangle \qquad \Leftrightarrow \qquad g_{ij} \left\langle \mathrm{d}x^j \right| \equiv \langle \partial_i | \, ; \tag{9.2.18}$$

or, equivalently,

$$\left| \mathrm{d}x^j \right\rangle \equiv g^{ij} \, |\partial_i\rangle \qquad \Leftrightarrow \qquad \left\langle \mathrm{d}x^j \right| \equiv g^{ij} \, \langle \partial_i | \, . \tag{9.2.19}$$

These also allow us to see

$$v^i \left|\partial_i\right\rangle = v^i g_{ij} \left|\mathrm{d}x^j\right\rangle = v_j \left|\mathrm{d}x^j\right\rangle, \tag{9.2.20}$$

where the lower index object v_j is obtained from its upper index cousin via contraction with the metric,

$$v_j \equiv v^i g_{ij}. \tag{9.2.21}$$

Likewise,

$$v_i \left|\mathrm{d}x^i\right\rangle = v_i g^{ij} \left|\partial_j\right\rangle = v^j \left|\partial_j\right\rangle, \tag{9.2.22}$$

where the upper index object v^j is obtained from its lower index cousin via contraction with the inverse metric,

To sum:

> At a given point in a curved space, one may define two different vector spaces – one spanned by the basis tangent vectors $\{|\partial_i\rangle\}$ (whose length^2 is given by the metric g_{ij}) and another by its dual 'bras' $\{|\mathrm{d}x^i\rangle\}$ (whose length^2 is given by the inverse metric g^{ij}). These two vector spaces are connected through the metric g_{ij} and its inverse.

Incidentally, we may express the 'dot product' between $\vec{v}$ and $\vec{w}$ using the bra-ket notation:

$$\langle \vec{v} | \vec{w} \rangle = v^i w^j \langle \partial_i | \partial_j \rangle = g_{ij} v^i w^j. \tag{9.2.23}$$

Line integral The metric, which measures distance, is a superposition of the square of infinitesimal displacements $\{\mathrm{d}x^i\}$. On the other hand, we may readily integrate the infinitesimal displacement in the ith direction, namely $\int_a^b \mathrm{d}x^i$; or the variation of a general function $f(\vec{x})$ along an arbitrary path $\vec{x}(\lambda)$ linking $\vec{x}(\lambda = \lambda_1) = \vec{z}'$ and $\vec{x}(\lambda = \lambda_2) = \vec{z}$,

$$\int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} \mathrm{d}f = \int_{\lambda_1}^{\lambda_2} \partial_i f(\vec{x}) \frac{\mathrm{d}x^i}{\mathrm{d}\lambda} \mathrm{d}\lambda \tag{9.2.24}$$

$$= \int_{\lambda_1}^{\lambda_2} \frac{\mathrm{d}f(\vec{x})}{\mathrm{d}\lambda} \mathrm{d}\lambda = f(\vec{z}) - f(\vec{z}'). \tag{9.2.25}$$

In this case, we see that the result is actually path independent; and only depends on the end points $\vec{z}'$ and $\vec{z}$.

More generally, we may define the $1-$*form* as a general superposition of the infinitesimal displacements

$$A \equiv A_i(\vec{x}) \mathrm{d}x^i. \tag{9.2.26}$$

If we then define a path $\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)$ in space, with the end points

$$\vec{x}(\lambda_1) = \vec{z}' \qquad \text{and} \qquad \vec{x}(\lambda_2) = \vec{z}; \tag{9.2.27}$$

we may then proceed to define the line integral

$$\int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} A \equiv \int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} A_i \mathrm{d}x^i = \int_{\lambda_1}^{\lambda_2} A_i \left(\vec{x}(\lambda)\right) \frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda} \mathrm{d}\lambda. \tag{9.2.28}$$

The line integral that occurs in 3D vector calculus, is then merely the flat Cartesian version of the above, where $A_i \mathrm{d}x^i = \vec{A} \cdot \mathrm{d}\vec{x} = \vec{A} \cdot \vec{v}\mathrm{d}\lambda$, where $\vec{v}(\lambda) \equiv \mathrm{d}\vec{x}(\lambda)/\mathrm{d}\lambda$.

Note, in general, the line integral in eq. (9.2.28) gives a result that depends not only on the end points $\vec{z}'$ and $\vec{z}$ but on the path $\vec{x}(\lambda)$ as well. Choosing a different path would yield a different answer, even if the end points are fixed. To see this, we promote the path $\vec{x}(\lambda)$ into a one-parameter family of paths; i.e., $\vec{x}(\lambda) \to \vec{Y}(\lambda, \sigma)$, where σ is a continuous (real) parameter – running over, say, the interval $[\sigma_1, \sigma_2]$ – which distinguishes one path from another, but still joining the same end point $\vec{z}'$ and $\vec{z}$:

$$\vec{Y}(\lambda_1, \sigma_1 \leq \sigma \leq \sigma_2) = \vec{z}' \qquad \text{and} \qquad \vec{Y}(\lambda_2, \sigma_1 \leq \sigma \leq \sigma_2) = \vec{z}. \tag{9.2.29}$$

That the path is fixed at the end points means

$$\partial_\sigma \vec{Y}(\lambda_1, \sigma) = \vec{0} = \partial_\sigma \vec{Y}(\lambda_2, \sigma). \tag{9.2.30}$$

The line integral in eq. (9.2.28) now becomes

$$\int_{\lambda_1}^{\lambda_2} A_j(\vec{Y}) \frac{\partial Y^j}{\partial \lambda} \mathrm{d}\lambda. \tag{9.2.31}$$

If this line integral did not depend on the specific path chosen, the answer must be independent of σ. The derivative of the line integral with respect to σ must be zero:

$$0 = \int_{\lambda_1}^{\lambda_2} \left(\partial_\sigma Y^i \partial_i A_j \partial_\lambda Y^j + A_j \partial_\sigma \partial_\lambda Y^j\right) \mathrm{d}\lambda \tag{9.2.32}$$

$$= \int_{\lambda_1}^{\lambda_2} \partial_\sigma Y^i \partial_\lambda Y^j \left(\partial_i A_j - \partial_j A_i\right) \mathrm{d}\lambda + \left[A_j \partial_\sigma Y^j\right]_{\lambda=\lambda_1}^{\lambda=\lambda_2} \tag{9.2.33}$$

$$= \int_{\lambda_1}^{\lambda_2} \partial_\sigma Y^i \partial_\lambda Y^j F_{ij} \mathrm{d}\lambda, \tag{9.2.34}$$

where eq. (9.2.30) was used to set to zero the boundary terms and

$$F_{ij} \equiv \partial_i A_j - \partial_j A_i \equiv \partial_{[i} A_{j]}. \tag{9.2.35}$$

Below, we shall see that the anti-symmetric F_{ij} is in fact the generalized version of the 'curl' in vector calculus, which in differential geometry is more commonly dubbed 'exterior derivative'. Since we have not made any assumptions about $\vec{Y}$ except that it joins $\vec{z}'$ to $\vec{z}$ and that it is a continuous family of paths paramterized by σ, we may therefore conclude

> The line integral (from, say, $\vec{z}'$ to $\vec{z}$) of the one form $A_i \mathrm{d}x^i$ is path-independent within a simply connected domain if and only if $F_{ij} = \partial_{[i} A_{j]} = 0$ within the same domain.

If the line integral were indeed path-independent, we may, for a fixed based point $\vec{z}'$, define the scalar f at $\vec{z}$ via the integral

$$f(\vec{z}) \equiv \int_{\lambda_1}^{\lambda_2} A_j(\vec{x}(\lambda)) \frac{\mathrm{d}x^j}{\mathrm{d}\lambda} \mathrm{d}\lambda \tag{9.2.36}$$

for *any* path $\vec{x}(\lambda)$ that joins $\vec{z}'$ to $\vec{z}$. (This would not be a sensible definition if different paths yielded different answers for $f(\vec{z})$.) Let us now suppose we have $\vec{x}(\lambda, \vec{z}', \vec{z})$ from above linking $\vec{z}'$ to $\vec{z}$ as λ runs from λ_1 to λ_2, but now for *arbitrary* $\vec{z}'$ and $\vec{z}$. That is, we should view the path as a function not only of λ but also of $\vec{z}'$ and $\vec{z}$; namely,

$$\vec{x}(\lambda_1, \vec{z}', \vec{z}) = \vec{z}' \qquad \text{and} \qquad \vec{x}(\lambda_2, \vec{z}', \vec{z}) = \vec{z}; \tag{9.2.37}$$

which in turn implies

$$\partial_{z^i} x^j(\lambda_1, \vec{z}', \vec{z}) = 0 \qquad \text{and} \qquad \partial_{z^i} x^j(\lambda_2, \vec{z}', \vec{z}) = \delta^j_i. \tag{9.2.38}$$

Let us then compute the gradient of f at $\vec{z}$. First, eq. (9.2.36) now reads

$$f(\vec{z}) = \int_{\lambda_1}^{\lambda_2} A_j\left(\vec{x}(\lambda, \vec{z}', \vec{z})\right) \frac{\partial x^j}{\partial \lambda} \mathrm{d}\lambda, \tag{9.2.39}$$

and the gradient, in turn, is

$$\begin{aligned}
\partial_{z^i} f(\vec{z}) &= \int_{\lambda_1}^{\lambda_2} \left(\frac{\partial x^a}{\partial z^i} \partial_{x^a} A_b(\vec{x}) \frac{\partial x^b}{\partial \lambda} + A_b(\vec{x}) \frac{\partial^2 x^b}{\partial z^i \partial \lambda} \right) \mathrm{d}\lambda && (9.2.40) \\
&= \left[A_b(\vec{x}) \frac{\partial x^b}{\partial z^i} \right]_{\lambda=\lambda_1}^{\lambda=\lambda_2} + \int_{\lambda_1}^{\lambda_2} \left(\frac{\partial x^a}{\partial z^i} \partial_{x^a} A_b(\vec{x}) \frac{\partial x^b}{\partial \lambda} - \frac{\partial x^a}{\partial \lambda} \partial_{x^a} A_b(\vec{x}) \frac{\partial x^b}{\partial z^i} \right) \mathrm{d}\lambda && (9.2.41) \\
&= A_i(\vec{z}) + \int_{\lambda_1}^{\lambda_2} \frac{\partial x^a}{\partial z^i} \frac{\partial x^b}{\partial \lambda} F_{ab}(\vec{x}) \mathrm{d}\lambda. && (9.2.42)
\end{aligned}$$

Since $F_{ij} = 0$ – the definition of f was path independent – we have arrived at

$$\partial_i f = A_i. \tag{9.2.43}$$

Integrating both sides of this equation would recover eq. (9.2.24) up to an additive constant. This, in turn, informs us

> The one-form A_i is a pure gradient (namely, $A_i = \partial_i f$) within a simply connected domain if and only if its 'curl' (or, exterior derivative) is zero: $F_{ij} = \partial_{[i} A_{j]} = 0$.

Finally, notice this entire discussion on line integrals made no reference to the metric g_{ij}.

Problem 9.5. Straight line integral If $F_{ij} = \partial_{[i} A_{j]} = 0$, we have argued that *any* path linking $\vec{z}'$ to $\vec{z}$ would yield a scalar whose gradient is A_i itself. Verify via a direct calculation (or otherwise) that the straight line path

$$\vec{x}(0 \leq \lambda \leq 1) = \vec{z}' + \lambda(\vec{z} - \vec{z}') \tag{9.2.44}$$

leads to

$$f(\vec{z}) = \int_0^1 \mathrm{d}\lambda A_j \left(\vec{z}' + \lambda(\vec{z} - \vec{z}')\right) (z^j - z'^j); \tag{9.2.45}$$

and that

$$\partial_{z^i} f(\vec{z}) = A_i(\vec{z}). \tag{9.2.46}$$

Hint: By carrying out the derivative on the line integral, you should be able to first show that

$$\partial_i f(\vec{z}) = \int_0^1 \mathrm{d}\lambda \left(A_i \left(\vec{x}(\lambda)\right) + \lambda \partial_{x^{[i}} A_{j]} \left(\vec{x}(\lambda)\right) (z^j - z'^j) + \lambda \partial_{x^j} A_i \left(\vec{x}(\lambda)\right) (z^j - z'^j) \right). \tag{9.2.47}$$

Then, proceed to employ $F_{ij} = 0$ and consider how $\mathrm{d}/\mathrm{d}\lambda$ acts on $A_i(\vec{x}(\lambda))$. □

General tensor A *scalar* φ is an object with no indices that transforms as

$$\varphi(\vec{\xi}) = \varphi\left(\vec{x}(\vec{\xi})\right). \tag{9.2.48}$$

That is, take $\varphi(\vec{x})$ and simply replace $\vec{x} \to \vec{x}(\vec{\xi})$ to obtain $\varphi(\vec{\xi})$. An example of a scalar field is the temperature $T(\vec{x})$ of an uneven, hence curved, 2D surface. Perhaps somewhat less obvious, the coordinates we endow to a given curved space(time) are also scalars – the intersections of their 'equipotential' surfaces are in fact the grid lines that allow us to parametrize the space(time) itself. For instance, in 3D flat space parametrized by spherical coordinates (r, θ, ϕ), the equipotential surfaces of the radial coordinate are simply the surface of a sphere with radius r. Their intersection with constant θ surfaces form latitude lines; and with constant ϕ surfaces form longitude ones.

A *vector* $v^i(\vec{x})\partial_i$ transforms as, by the chain rule,

$$v^i(\vec{x})\frac{\partial}{\partial x^i} = v^i(\vec{x}(\vec{\xi}))\frac{\partial \xi^j}{\partial x^i}\frac{\partial}{\partial \xi^j} \equiv v^j(\vec{\xi})\frac{\partial}{\partial \xi^j}. \tag{9.2.49}$$

If we attribute all the transformations to the components, the components in the $\vec{x}$-coordinate system $v^i(\vec{x})$ is related to those in the $\vec{y}$-coordinate system $v^i(\vec{\xi})$ through the relation

$$v^i(\vec{\xi}) = v^a(\vec{x}(\vec{\xi}))\frac{\partial \xi^i}{\partial x^a}. \tag{9.2.50}$$

Similarly, a 1-form $A_i \mathrm{d}x^i$ transforms, by the chain rule,

$$A_i(\vec{x})\mathrm{d}x^i = A_i(\vec{x}(\vec{\xi}))\frac{\partial x^i}{\partial \xi^j}\mathrm{d}\xi^j \equiv A_j(\vec{\xi})\mathrm{d}\xi^j. \tag{9.2.51}$$

If we again attribute all the coordinate transformations to the components; the ones in the $\vec{x}$-system $A_i(\vec{x})$ is related to the ones in the $\vec{\xi}$-system $A_i(\vec{\xi})$ through

$$A_j(\vec{\xi}) = A_i(\vec{x}(\vec{\xi}))\frac{\partial x^i}{\partial \xi^j}. \tag{9.2.52}$$

By taking tensor products of $\{|\partial_i\rangle\}$ and $\{\langle \mathrm{d}x^i|\}$, we may define a *rank* $\binom{N}{M}$ *tensor* T as an object with N "upper indices" and M "lower indices" that transforms as

$$T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M}(\vec{\xi}) = T^{a_1 a_2 \dots a_N}{}_{b_1 b_2 \dots b_M}\left(\vec{x}(\vec{\xi})\right) \frac{\partial \xi^{i_1}}{\partial x^{a_1}} \dots \frac{\partial \xi^{i_N}}{\partial x^{a_N}} \frac{\partial x^{b_1}}{\partial \xi^{j_1}} \dots \frac{\partial x^{b_M}}{\partial \xi^{j_M}}. \tag{9.2.53}$$

The left hand side are the tensor components in $\vec{\xi}$ coordinates and the right hand side are the Jacobians $\partial x/\partial \xi$ and $\partial \xi/\partial x$ contracted with the tensor components in $\vec{x}$ coordinates – but now with the $\vec{x}$ replaced with $\vec{x}(\vec{\xi})$, their corresponding expressions in terms of $\vec{\xi}$. This multi-indexed object should be viewed as the components of

$$T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M}(\vec{x}) \left| \frac{\partial}{\partial x^{i_1}} \right\rangle \otimes \dots \otimes \left| \frac{\partial}{\partial x^{i_N}} \right\rangle \otimes \left\langle \mathrm{d}x^{j_1} \right| \otimes \dots \otimes \left\langle \mathrm{d}x^{j_M} \right|. \tag{9.2.54}$$

[80]Above, we only considered T with all upper indices followed by all lower indices. Suppose we had $T^i{}_j{}^k$; it is the components of

$$T^i{}_j{}^k(\vec{x}) \left|\partial_i\right\rangle \otimes \left\langle \mathrm{d}x^j \right| \otimes \left|\partial_k\right\rangle. \tag{9.2.55}$$

Raising and lowering tensor indices The indices on a tensor are moved – from upper to lower, or vice versa – using the metric tensor. For example,

$$T^{m_1 \dots m_a}{}_i{}^{n_1 \dots n_b} = g_{ij} T^{m_1 \dots m_a j n_1 \dots n_b}, \tag{9.2.56}$$

$$T_{m_1 \dots m_a}{}^i{}_{n_1 \dots n_b} = g^{ij} T_{m_1 \dots m_a j n_1 \dots n_b}. \tag{9.2.57}$$

The key observation is the upper and lower indices transform 'oppositely' from each other because of eq. (9.2.11). Compare

$$V^i(\vec{\xi}) = \frac{\partial \xi^i}{\partial x^a} V^a\left(\vec{x}(\vec{\xi})\right) \equiv \mathcal{J}^i{}_a V^a\left(\vec{x}(\vec{\xi})\right) \tag{9.2.58}$$

versus

$$W_i(\vec{\xi}) = \frac{\partial x^a}{\partial \xi^i} W_a\left(\vec{x}(\vec{\xi})\right) \equiv W_a\left(\vec{x}(\vec{\xi})\right) \left(\mathcal{J}^{-1}\right)^a{}_i. \tag{9.2.59}$$

---

[80]Strictly speaking, when discussing the metric and its inverse above, we should also have respectively expressed them as $g_{ij}\left\langle \mathrm{d}x^i\right| \otimes \left\langle \mathrm{d}x^j\right|$ and $g^{ij}\left|\partial_i\right\rangle \otimes \left|\partial_j\right\rangle$, with the appropriate bras and kets enveloping the $\{\mathrm{d}x^i\}$ and $\{\partial_i\}$. We did not do so because we wanted to highlight the geometric interpretation of $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ as the square of the distance between $\vec{x}$ and $\vec{x}+\mathrm{d}\vec{x}$, where the notion of $\mathrm{d}x^i$ as (a component of) an infinitesimal 'vector' – as opposed to being a 1-form – is, in our opinion, more useful for building the reader's geometric intuition.

It may help the physicist reader to think of a scalar field in eq. (9.2.48) as an observable, such as the temperature $T(\vec{x})$ of the 2D undulating surface mentioned above. If you were provided such an expression for $T(\vec{x})$, together with an accompanying definition for the coordinate system $\vec{x}$; then, to convert this same temperature field to a different coordinate system (say, $\vec{\xi}$) one would, in fact, do $T(\vec{\xi}) \equiv T(\vec{x}(\vec{\xi}))$, because you'd want $\vec{\xi}$ to refer to the same point in space as $\vec{x} = \vec{x}(\vec{\xi})$. For a general tensor in eq. (9.2.54), the tensor components $T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M}$ may then be regarding as scalars describing some weighted superposition of the tensor product of basis vectors and 1-forms. Its transformation rules in eq. (9.2.53) are really a shorthand for the lazy physicist who does not want to carry the basis vectors/1-forms around in his/her calculations.

Hence, $v_i = g_{ij} v^j$ automatically converts the vector v^i into a tensor that transforms properly as 1−form; and similarly, $v^i = g^{ij} v_j$ automatically produces a vector from a 1−form v_i. In fact, recalling the "Kets and Bras" discussion above, we have for instance:

$$V^j \left|\partial_j\right\rangle = V^j (g_{ij} \left|\mathrm{d}x^i\right\rangle) = V_i \left|\mathrm{d}x^i\right\rangle = V_i (g^{ij} \left|\partial_j\right\rangle). \tag{9.2.60}$$

Because upper indices transform oppositely from lower indices, when we contract a upper and lower index, it now transforms as a scalar. For example,

$$\begin{aligned} A^i{}_l(\vec{\xi}) B^{lj}(\vec{\xi}) &= \frac{\partial \xi^i}{\partial x^m} A^m{}_a \left(\vec{x}(\vec{\xi})\right) \frac{\partial x^a}{\partial \xi^l} \frac{\partial \xi^l}{\partial x^c} B^{cn} \left(\vec{x}(\vec{\xi})\right) \frac{\partial \xi^j}{\partial x^n} \\ &= \frac{\partial \xi^i}{\partial x^m} A^m{}_a \left(\vec{x}(\vec{\xi})\right) \delta^a{}_c \, B^{cn} \left(\vec{x}(\vec{\xi})\right) \frac{\partial \xi^j}{\partial x^n} \\ &= \frac{\partial \xi^i}{\partial x^m} \frac{\partial \xi^j}{\partial x^n} A^m{}_c \left(\vec{x}(\vec{\xi})\right) B^{cn} \left(\vec{x}(\vec{\xi})\right). \end{aligned} \tag{9.2.61}$$

Moreover, we have the following equivalent scalars

$$v^i w_i = v^i (g_{ij} w^j) = (v_i g^{ij}) w_j = v_i w^i. \tag{9.2.62}$$

Altogether, these illustrate why we use the metric g_{ij} and its inverse g^{ij} to move indices: since they are always available in a given (curved) geometry, they provide a universal means to convert one tensor to another through movement of its indices. In fact, we may go further.

Problem 9.6. Invariant $\binom{1}{1}$ Tensor Suppose, in the $\vec{x}$ coordinate system, we have the tensor $T^i{}_j(\vec{x}) \equiv \delta^i{}_j$, which may be represented as a $D \times D$ identity matrix $\mathbb{I}_{D\times D}$. Show that, in any other coordinate system $\vec{\xi}$, this tensor remains the *same*:

$$T^i{}_j(\vec{\xi}) = \delta^i{}_j = T^i{}_j(\vec{x}). \tag{9.2.63}$$

The placement of indices is important here. Explain why $T_{ij}(\vec{x}) \equiv \delta_{ij}$ and $T^{ij}(\vec{x}) \equiv \delta^{ij}$ *do not* remain the same under coordinate transformations $\vec{x} \to \vec{x}(\vec{\xi})$. Hint: Provide an example. □

The invariance of $\delta^i{}_j$ is why we may simultaneously and consistently define g^{ij} as the matrix inverse of g_{ij} and assert that g^{ij} are the components of $g^{ij}\partial_i \otimes \partial_j$, with both indices of the g^{ij} transforming oppositely from those of g_{ij}. For, we may simply contract any pair of indices; say,

$$g^{ia}(\vec{x}) g_{ib}(\vec{x}) = \delta^a{}_b = g^{ia}(\vec{\xi}) g_{ib}(\vec{\xi}). \tag{9.2.64}$$

Quotient Rule A closely related result is known as the quotient theorem, which we shall phrase as follows. If

$$C(\vec{x}) \equiv A^{i_1 \dots i_N}(\vec{x}) B_{i_1 \dots i_N}(\vec{x}) \tag{9.2.65}$$

transforms as a scalar for *any* tensor $B_{i_1 \dots i_N}$, then $A^{i_1 \dots i_N}$ must be a tensor. For, since C is a scalar, we may switch from the $\vec{x}$ coordinate system to the $\vec{\xi}$ system by simplying inserting replacing $\vec{x} \to \vec{x}(\vec{\xi})$, namely

$$C'(\vec{\xi}) = C(\vec{x}(\vec{\xi})) = A^{i_1 \dots i_N}(\vec{x}(\vec{\xi})) B_{i_1 \dots i_N}(\vec{x}(\vec{\xi})). \tag{9.2.66}$$

But since $B_{i_1 \dots i_N}$ is a tensor, we may immediately assert

$$C(\vec{x}(\vec{\xi})) = A^{i_1 \dots i_N}(\vec{x}(\vec{\xi})) \frac{\partial \xi^{a_1}}{\partial x^{i_1}} \dots \frac{\partial \xi^{a_N}}{\partial x^{i_N}} B_{a'_1 \dots a'_N}(\vec{\xi}). \tag{9.2.67}$$

Since $B_{i_1 \dots i_N}$ is arbitrary this equation necessarily involves all components of $A^{i_1 \dots i_N}$, which then allows us to recognize

$$A^{a'_1 \dots a'_N}(\vec{\xi}) \equiv A^{i_1 \dots i_N}(\vec{x}(\vec{\xi})) \frac{\partial \xi^{a_1}}{\partial x^{i_1}} \dots \frac{\partial \xi^{a_N}}{\partial x^{i_N}} \tag{9.2.68}$$

– i.e., $A^{i_1 \dots i_N}$ can be viewed as a tensor – so that in the 'new' coordinate system

$$C(\vec{x}(\vec{\xi})) = A^{a'_1 \dots a'_N}(\vec{\xi}) B_{a'_1 \dots a'_N}(\vec{\xi}). \tag{9.2.69}$$

Problem 9.7. Prove that, if

$$C^{i_1 \dots i_S}{}_{k_1 \dots k_M} \equiv A^{i_1 \dots i_S j_1 \dots j_N} B_{k_1 \dots k_M j_1 \dots j_N} \tag{9.2.70}$$

transforms as a $\binom{S}{M}$ tensor for any $B_{k_1 \dots k_M j_1 \dots j_N}$ that transforms as a tensor; then $A^{i_1 \dots i_S j_1 \dots j_N}$ must be a tensor. Here, N, M, and S are arbitrary positive integers. $\square$

Problem 9.8. Cartesian Tensors in Flat Space In $D-$dimensional flat space with geometry $g_{ij} = \delta_{ij}$ parametrized by Cartesian coordinates $\vec{x}$, consider the Euclidean coordinate transformation (rotation plus spatial translation)

$$x^i = \widehat{R}^{ij} x'^j + a^i; \tag{9.2.71}$$

where $\widehat{R}$ is an orthogonal matrix obeying $\widehat{R}^{\mathrm{T}} \widehat{R} = \mathbb{I}_{D \times D}$; and $\vec{a}$ is constant. For an arbitrary tensor $T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M}(\vec{x})$, derive the relationship

$$T^{a'_1 a'_2 \dots a'_N}{}_{b'_1 b'_2 \dots b'_M}(\vec{x}') = T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M}(\vec{x} = \widehat{R}\vec{x}' + \vec{a}) \widehat{R}^{i_1 a_1} \dots \widehat{R}^{i_N a_N} \widehat{R}_{j_1 b_1} \dots \widehat{R}_{j_M b_M}; \tag{9.2.72}$$

where $\widehat{R}^{ab} = \widehat{R}_{ab}$, and $T^{a'_1 a'_2 \dots a'_N}{}_{b'_1 b'_2 \dots b'_M}$ are the components of the same tensor but in the $\vec{x}'$ coordinate system. Notice both upper and lower indices transform in the same manner. $\square$

Problem 9.9. Argue that, if a tensor $T^{i_1 i_2 \dots i_N}$ is zero in some coordinate system, it must be zero in any other coordinate system. $\square$

General covariance Tensors are ubiquitous in physics: the electric and magnetic fields can be packaged into one Faraday tensor $F_{\mu\nu}$; the energy-momentum-shear-stress tensor of matter $T_{\mu\nu}$ is what sources the curved geometry of spacetime in Einstein's theory of General Relativity; etc. The coordinate transformation rules in eq. (9.2.53) that defines a tensor is actually the statement that, the mathematical description of the physical world (the tensors themselves in eq. (9.2.54)) should not depend on the coordinate system employed. Any expression or equation with physical meaning – i.e., it yields quantities that can in principle be measured – must be put in a form that is generally covariant: either a scalar or tensor under coordinate transformations.[81]

[81] You may also demand your equations/quantities to be tensors/scalars under group transformations.

An example is, it makes no sense to assert that your new-found law of physics depends on g^{11}, the 11 component of the inverse metric – for, in what coordinate system is this law expressed in? What happens when we use a different coordinate system to describe the outcome of some experiment designed to test this law?

Above, we have already encountered the scalar $g_{ij}v^i w^j$, where $\vec{v}$ and $\vec{w}$ are arbitrary (tangent) vectors. Since it is a scalar, we may evaluate it in any coordinate system we wish. This is why it was legitimate to do so in a locally flat coordinate system so as to facilitate its interpretation as a curved space 'dot product'.

Below, we will show that the infinitesimal volume in curved space is given by $\mathrm{d}^D\vec{x}\sqrt{g(\vec{x})}$, where $g(\vec{x})$ is the determinant of the metric in the $\vec{x}-$coordinate basis. For this to make sense geometrically, you will show in Problem (9.10) below that it is in fact generally covariant – i.e., it takes the same form in any coordinate system:

$$\mathrm{d}^D\vec{x}\sqrt{g(\vec{x})} = \mathrm{d}^D\vec{\xi}\sqrt{g(\vec{\xi})}; \tag{9.2.73}$$

where $g(\vec{\xi})$ is the determinant of the metric but in the $\vec{\xi}-$coordinate basis.

Another aspect of general covariance is that, although tensor equations should hold in any coordinate system – if you suspect that two tensors quantities are actually equal, say

$$S^{i_1 i_2 \dots} = T^{i_1 i_2 \dots}, \tag{9.2.74}$$

it suffices to find one coordinate system to prove this equality. It is not necessary to prove this by using abstract indices/coordinates because, as long as the coordinate transformations are invertible, then once we have verified the equality in one system, the proof in any other follows immediately once the required transformations are specified. One common application of this observation is to apply the fact mentioned around eq. (9.2.1), that at any given point in a curved space(time), one can always choose coordinates where the metric there is flat. You will often find this "locally flat" coordinate system simplifies calculations – and perhaps even aids in gaining some intuition about the relevant physics, since the expressions usually reduce to their more familiar counterparts in flat space. To illustrate this using a simple example, we now answer the question: what is the curved analog of the infinitesimal volume, which we would usually write as $\mathrm{d}^D x$ in Cartesian coordinates?

Determinant of metric and the infinitesimal volume The determinant of the metric transforms as

$$\det g_{ij}(\vec{\xi}) = \det\left[g_{ab}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^a}{\partial \xi^i}\frac{\partial x^b}{\partial \xi^j}\right]. \tag{9.2.75}$$

Using the properties $\det A \cdot B = \det A \det B$ and $\det A^{\mathrm{T}} = \det A$, for any two square matrices A and B,

$$\det g_{ij}(\vec{\xi}) = \left(\det\frac{\partial x^a(\vec{\xi})}{\partial \xi^b}\right)^2 \det g_{ij}\left(\vec{x}(\vec{\xi})\right). \tag{9.2.76}$$

The square root of the determinant of the metric is often denoted as $\sqrt{|g|}$. It transforms as

$$\sqrt{\left|g(\vec{\xi})\right|} = \sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|}\left|\det\frac{\partial x^a(\vec{\xi})}{\partial \xi^b}\right|. \tag{9.2.77}$$

We have previously noted that, given any point $\vec{x}_0$ in the curved space, we can always choose local coordinates $\{\vec{x}\}$ such that the metric there is flat. This means at $\vec{x}_0$ the infinitesimal volume of space is $\mathrm{d}^D\vec{x}$ and $\det g_{ij}(\vec{x}_0) = 1$. Recall from multi-variable calculus that, whenever we transform $\vec{x} \to \vec{x}(\vec{\xi})$, the integration measure would correspondingly transform as

$$\mathrm{d}^D\vec{x} = \mathrm{d}^D\vec{\xi}\left|\det\frac{\partial x^i}{\partial \xi^a}\right|, \tag{9.2.78}$$

where $\partial x^i/\partial\xi^a$ is the Jacobian matrix with row number i and column number a. Comparing this multi-variable calculus result to eq. (9.2.77) specialized to our metric in terms of $\{\vec{x}\}$ but evaluated at $\vec{x}_0$, we see the determinant of the Jacobian *is* in fact the square root of the determinant of the metric in some other coordinates $\vec{\xi}$,

$$\sqrt{\left|g(\vec{\xi})\right|} = \left(\sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|}\left|\det\frac{\partial x^i(\vec{\xi})}{\partial\xi^a}\right|\right)_{\vec{x}=\vec{x}_0} = \left|\det\frac{\partial x^i(\vec{\xi})}{\partial\xi^a}\right|_{\vec{x}=\vec{x}_0}. \tag{9.2.79}$$

In flat space and by employing Cartesian coordinates $\{\vec{x}\}$, the infinitesimal volume (at some location $\vec{x} = \vec{x}_0$) is $\mathrm{d}^D\vec{x}$. What is its curved analog? What we have just shown is that, by going from $\vec{\xi}$ to a locally flat coordinate system $\{\vec{x}\}$,

$$\mathrm{d}^D\vec{x} = \mathrm{d}^D\vec{\xi}\left|\det\frac{\partial x^i(\vec{\xi})}{\partial\xi^a}\right|_{\vec{x}=\vec{x}_0} = \mathrm{d}^D\vec{\xi}\sqrt{|g(\vec{\xi})|}. \tag{9.2.80}$$

However, since $\vec{x}_0$ was an arbitrary point in our curved space, we have argued that, in a general coordinate system $\vec{\xi}$, the infinitesimal volume is given by

$$\mathrm{d}^D\vec{\xi}\sqrt{\left|g(\vec{\xi})\right|} \equiv \mathrm{d}\xi^1 \dots \mathrm{d}\xi^D\sqrt{\left|g(\vec{\xi})\right|}. \tag{9.2.81}$$

Problem 9.10. General Covariance of Volume Form Upon an orientation preserving change of coordinates $\vec{y} \to \vec{y}(\vec{\xi})$, where $\det\partial y/\partial\xi > 0$, show that

$$\mathrm{d}^D\vec{y}\sqrt{|g(\vec{y})|} = \mathrm{d}^D\vec{\xi}\sqrt{\left|g(\vec{\xi})\right|}. \tag{9.2.82}$$

Therefore calling $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}$ an infinitesimal volume is a generally covariant statement.

It is worth reiterating: $g(\vec{y})$ is the determinant of the metric written in the $\vec{y}$ coordinate system; whereas $g(\vec{\xi})$ is that of the metric written in the $\vec{\xi}$ coordinate system. The latter is *not* the same as the determinant of the metric written in the $\vec{y}$-coordinates, with $\vec{y}$ replaced with $\vec{y}(\vec{\xi})$; namely, be careful that the determinant is not a scalar. □

Since $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}$ is generally covariant – the same prescription may be employed to compute it in all coordinate systems – and since all curved spaces are locally flat in the Riemann normal coordinate system $\{y^i\}$, we may employ the latter to interpret $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|} = \mathrm{d}^D\vec{y}$ as the infinitesimal volume.

Volume integrals If $\varphi(\vec{x})$ is some scalar quantity, finding its volume integral within some domain $\mathfrak{D}$ in a generally covariant way can be now carried out using the infinitesimal volume we have uncovered; it reads

$$I \equiv \int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g(\vec{x})|} \varphi(\vec{x}). \tag{9.2.83}$$

In other words, I is the same result no matter what coordinates we used to compute the integral on the right hand side.

Example: Volume of sphere The sphere of radius R in flat 3D space can be described by $r \leq R$, where in spherical coordinates $\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)$. Therefore $\det g_{ij} = r^4(\sin\theta)^2$ and the sphere's volume reads

$$\mathrm{Vol}(r \leq R) = \int_{r \leq R} \mathrm{d}^3\vec{\xi} \sqrt{g(\vec{\xi})}, \qquad \xi^i \equiv (r, \theta, \phi) \tag{9.2.84}$$

$$= \int_0^R \mathrm{d}r r^2 \int_0^\pi \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi = \frac{4}{3}\pi R^3. \tag{9.2.85}$$

Problem 9.11. 2D Ellipses In flat 2D space, consider the following coordinate transformation from Cartesian coordinates (x^1, x^2) to ellipsoidal ones (ρ, φ):

$$\vec{x} = \left(\sqrt{\rho^2 - a^2} \sin\varphi, \rho\cos\varphi\right), \tag{9.2.86}$$

$$\rho \geq a > 0, \qquad \varphi \in (-\pi, +\pi]. \tag{9.2.87}$$

For a fixed ρ, explain why $\vec{x}(-\pi < \varphi \leq \pi)$ sweeps out an ellipse of semi-major axis ρ, semi-minor axis $\sqrt{\rho^2 - a^2}$, and eccentricity

$$e = a/\rho. \tag{9.2.88}$$

Verify that $\vec{x}_\pm \equiv (0, \pm a)$ are the foci, in that

$$|\vec{x} - \vec{x}_\pm| = \rho \mp a\cos\varphi \tag{9.2.89}$$

and therefore

$$|\vec{x} - \vec{x}_+| + |\vec{x} - \vec{x}_-| = 2\rho. \tag{9.2.90}$$

Moreover, if we parametrize

$$\vec{x} - \vec{x}_\pm \equiv r_\pm(\sin\phi_\pm, \cos\phi_\pm), \tag{9.2.91}$$

$$\phi_\pm \in (-\pi, +\pi]; \tag{9.2.92}$$

demonstrate that

$$r_\pm = \rho \frac{1 - e^2}{1 \pm e\cos\phi_\pm}. \tag{9.2.93}$$

Finally, show that the metric becomes

$$\mathrm{d}\vec{x}^2 = \frac{\rho^2 - a^2\cos^2\varphi}{\rho^2 - a^2}\left(\mathrm{d}\rho^2 + (\rho^2 - a^2)\mathrm{d}\varphi^2\right), \tag{9.2.94}$$

and the area within the ellipse of some fixed ρ is

$$A = \pi\rho^2\sqrt{1 - e^2}. \tag{9.2.95}$$

Hint: Start by showing that $\mathrm{d}^2\vec{x} = \mathrm{d}\rho\mathrm{d}\varphi(\rho^2 - a^2\cos^2\varphi)/\sqrt{\rho^2 - a^2}$. □

Problem 9.12. Spherical coordinates in D Space Dimensions In D dimensional flat space, let $\vec{x}$ be Cartesian coordinates; $r \equiv |\vec{x}|$ the radius; and $\widehat{n}_D$ be the unit radial vector. We may further denote the D-th unit vector as $\widehat{e}_D$; and $\widehat{n}_{D-1}$ as the unit radial vector, parametrized by the angles $\{0 \leq \theta^1 < 2\pi, 0 \leq \theta^2 \leq \pi, \dots, 0 \leq \theta^{D-2} \leq \pi\}$, in the plane perpendicular to $\widehat{e}_D$. Any vector $\vec{x}$ in this space can thus be expressed as

$$\vec{x}(r, \vec{\theta}) = r\widehat{n}_D(\vec{\theta}), \tag{9.2.96}$$

$$\widehat{n}_D(\vec{\theta}) = \cos(\theta^{D-1})\widehat{e}_D + \sin(\theta^{D-1})\widehat{n}_{D-1}, \qquad 0 \leq \theta^{D-1} \leq \pi. \tag{9.2.97}$$

(Can you see why this is nothing but the Gram-Schmidt process?) Just like in the 3D case, $r\cos(\theta^{D-1})$ is the projection of $\vec{x}$ along the $\widehat{e}_D$ direction; while $r\sin(\theta^{D-1})$ is that along the radial direction in the plane perpendicular to $\widehat{e}_D$.

- If $\mathrm{d}\Omega^2_{N-1}$ is the square of the infinitesimal solid angle in N spatial dimensions, where

$$\mathrm{d}\Omega^2_{N-1} \equiv \sum_{\mathrm{I,J}=1}^{N-1} \Omega^{(N-1)}_{\mathrm{IJ}}\mathrm{d}\theta^{\mathrm{I}}\mathrm{d}\theta^{\mathrm{J}}, \tag{9.2.98}$$

$$\Omega^{(N-1)}_{\mathrm{IJ}} \equiv \sum_{i,j=1}^{N} \delta_{ij}\frac{\partial\widehat{n}^i_N}{\partial\theta^{\mathrm{I}}}\frac{\partial\widehat{n}^j_N}{\partial\theta^{\mathrm{J}}}, \tag{9.2.99}$$

  first show that the Cartesian metric δ_{ij} in D-space transforms to

$$(\mathrm{d}\ell)^2 = \mathrm{d}r^2 + r^2\mathrm{d}\Omega^2_{D-1}. \tag{9.2.100}$$

  Since the sphere may be defined as the surface of constant radius r, namely $\mathrm{d}r = 0$, we see that $r^2\mathrm{d}\Omega^2_{D-1}$ must be the metric on the $(D-1)-$sphere of (fixed) radius r.

- Then show that square of the infinitesimal solid angle in $D-$space, i.e., the metric of the unit $(D-1)-$sphere, is related to that in $(D-1)-$space as

$$\mathrm{d}\Omega^2_{D-1} = (\mathrm{d}\theta^{D-1})^2 + (\sin\theta^{D-1})^2\mathrm{d}\Omega^2_{D-2}. \tag{9.2.101}$$

- Proceed to argue that the full D-metric in spherical coordinates is

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\left((\mathrm{d}\theta^{D-1})^2 + \sum_{\mathrm{I}=2}^{D-1} s^2_{D-1}\dots s^2_{D-\mathrm{I}+1}(\mathrm{d}\theta^{D-\mathrm{I}})^2\right), \tag{9.2.102}$$

$$\theta^1 \in [0, 2\pi), \qquad \theta^2, \dots, \theta^{D-1} \in [0, \pi]. \tag{9.2.103}$$

  Here, $s_{\mathrm{I}} \equiv \sin\theta^{\mathrm{I}}$.

- Write down the Cartesian coordinates $\vec{x}$ as a function of their spherical counterparts, for $D = 3, 4, 5$.
- Show that the determinant of the angular metric $\Omega^{(D-1)}_{\mathrm{IJ}}$ obeys a recursion relation

$$\det \Omega^{(D-1)}_{\mathrm{IJ}} = \left(\sin\theta^{D-1}\right)^{2(D-2)} \cdot \det \Omega^{(D-2)}_{\mathrm{IJ}}. \tag{9.2.104}$$

- Explain why this implies there is a recursion relation between the infinitesimal solid angle in D space and that in $(D-1)$ space. Moreover, show that the integration volume measure $\mathrm{d}^D\vec{x}$ in Cartesian coordinates then becomes, in spherical coordinates,

$$\mathrm{d}^D\vec{x} = \mathrm{d}r \cdot r^{D-1} \cdot \mathrm{d}\theta^1 \dots \mathrm{d}\theta^{D-1} \left(\sin\theta^{D-1}\right)^{D-2} \sqrt{\det \Omega^{(D-2)}_{\mathrm{IJ}}}. \tag{9.2.105}$$

- Use these results to prove the solid angle Ω_{D-1} subtended by a unit sphere in D spatial dimensions is

$$\Omega_{D-1} = \frac{2\pi^{D/2}}{\Gamma(D/2)}, \tag{9.2.106}$$

where $\Gamma(z)$ is the Gamma function, the generalization of the factorial away from non-negative integers. Hint: You may find the integral representation of the Beta function useful; see here. □

9.3 Symmetries (aka Isometries) and Infinitesimal Displacements

In some Cartesian coordinates $\{x^i\}$ the flat space metric is $\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j$. Suppose we chose a different set of axes for new Cartesian coordinates $\{x'^i\}$, the metric will still take the same form, namely $\delta_{ij}\mathrm{d}x'^i\mathrm{d}x'^j$. Likewise, on a 2-sphere the metric is $\mathrm{d}\theta^2 + (\sin\theta)^2\mathrm{d}\phi^2$ with a given choice of axes for the 3D space the sphere is embedded in; upon any rotation to a new axis, so the new angles are now (θ', ϕ'), the 2-sphere metric is still of the same form $\mathrm{d}\theta'^2 + (\sin\theta')^2\mathrm{d}\phi'^2$. All we have to do, in both cases, is swap the symbols $\vec{x} \to \vec{x}'$ and $(\theta, \phi) \to (\theta', \phi')$. The reason why we can simply swap symbols to express the same geometry in different coordinate systems, is because of the symmetries present: for flat space and the 2-sphere, the geometries are respectively indistinguishable under translation/rotation and rotation about its center.

Motivated by this observation that geometries enjoying symmetries (aka isometries) retain their *form* under an active coordinate transformation – one that corresponds to an actual displacement from one location to another[82] – we now consider a infinitesimal coordinate transformation as follows. Starting from $\vec{x}$, we define a new set of coordinates $\vec{x}'$ through an infinitesimal vector $\vec{\xi}(\vec{x})$,

$$\vec{x}' \equiv \vec{x} - \vec{\xi}(\vec{x}). \tag{9.3.1}$$

(The $-$ sign is for technical convenience.) We shall interpret this definition as an active coordinate transformation – given some location $\vec{x}$, we now move to a point $\vec{x}'$ that is displaced infinitesimally far away, with the displacement itself described by $-\vec{\xi}(\vec{x})$. On the other hand, since

[82] As opposed to a passive coordinate transformation, which is one where a different set of coordinates are used to describe the same location in the geometry.

$\vec{\xi}$ is assumed to be "small," we may replace in the above equation, $\vec{\xi}(\vec{x})$ with $\vec{\xi}(\vec{x}') \equiv \vec{\xi}(\vec{x} \to \vec{x}')$. This is because the error incurred would be of $\mathcal{O}(\xi^2)$.

$$\vec{x} = \vec{x}' + \vec{\xi}(\vec{x}') + \mathcal{O}(\xi^2) \qquad \Rightarrow \qquad \frac{\partial x^i}{\partial x'^a} = \delta^i_a + \partial_{a'}\xi^i(\vec{x}') + \mathcal{O}(\xi\partial\xi) \tag{9.3.2}$$

Also note that inverse Jacobian is

$$\frac{\partial x'^i}{\partial x^a} = \delta^i_a - \partial_{a'}\xi^i(\vec{x}') + \mathcal{O}(\xi\partial\xi). \tag{9.3.3}$$

How does this coordinate transformation change our metric?

$$\begin{aligned} g_{ij}(\vec{x})\,\mathrm{d}x^i\mathrm{d}x^j &= g_{ij}\left(\vec{x}' + \vec{\xi}(\vec{x}') + \dots\right)\left(\delta^i_a + \partial_{a'}\xi^i + \dots\right)\left(\delta^j_b + \partial_{b'}\xi^j + \dots\right)\mathrm{d}x'^a\mathrm{d}x'^b \\ &= \left(g_{ij}(\vec{x}') + \xi^c\partial_{c'}g_{ij}(\vec{x}') + \dots\right)\left(\delta^i_a + \partial_{a'}\xi^i + \dots\right)\left(\delta^j_b + \partial_{b'}\xi^j + \dots\right)\mathrm{d}x'^a\mathrm{d}x'^b \\ &\equiv \left(g_{ij}(\vec{x}') + \left(\pounds_{\vec{\xi}}g\right)_{ij}(\vec{x}') + \mathcal{O}(\xi^2)\right)\mathrm{d}x'^i\mathrm{d}x'^j, \end{aligned} \tag{9.3.4}$$

where the first-order-in-$\vec{\xi}$ terms are

$$\left(\pounds_{\vec{\xi}}g\right)_{ij}(\vec{x}') \equiv \xi^c(\vec{x}')\frac{\partial g_{ij}(\vec{x}')}{\partial x'^c} + g_{ia}(\vec{x}')\frac{\partial \xi^a(\vec{x}')}{\partial x'^j} + g_{ja}(\vec{x}')\frac{\partial \xi^a(\vec{x}')}{\partial x'^i}. \tag{9.3.5}$$

This $(\pounds_{\vec{\xi}}g)_{ij}$ is known as the Lie derivative of the metric along $\vec{\xi}$.

A point of clarification might be helpful. In eq. (9.3.4), we are not merely asking "What is $\mathrm{d}\ell^2$ at $\vec{x}'$?" The answer to that question would be $(\mathrm{d}\ell)^2_{\vec{x}'} = g_{ij}(\vec{x} - \vec{\xi}(\vec{x}))\mathrm{d}x^i\mathrm{d}x^j$, with no need to transform the $\mathrm{d}x^i$. Rather, here, we are performing a coordinate transformation from $\vec{x}$ to $\vec{x}'$, *induced* by an infinitesimal displacement via $\vec{x}' = \vec{x} - \vec{\xi}(\vec{x}) \Leftrightarrow \vec{x} = \vec{x}' + \vec{\xi}(\vec{x}') + \dots$ – where $\vec{x}$ and $\vec{x}'$ are infinitesimally separated. An elementary example would be to rotate the 2$-$sphere about the $z-$axis, so $\theta = \theta'$ but $\phi = \phi' + \epsilon$ for infinitesimal ϵ. Then, $\xi^i\partial_i = \epsilon\partial_\phi$.

At this point, we see that if the geometry enjoys a symmetry along the entire curve whose tangent vector is $\vec{\xi}$, then it must retain its form $g_{ij}(\vec{x})\mathrm{d}x^i\mathrm{d}x^j = g_{ij}(\vec{x}')\mathrm{d}x'^i\mathrm{d}x'^j$ and therefore,[83] obey the equations

$$\left(\pounds_{\vec{\xi}}g\right)_{ij} = 0, \qquad \text{(isometry along } \vec{\xi}\text{)}. \tag{9.3.6}$$

Conversely, if $(\pounds_{\vec{\xi}}g)_{ij} = 0$ everywhere in space, then starting from some point $\vec{x}$, we can make incremental displacements along the curve whose tangent vector is $\vec{\xi}$, and therefore find that the metric retain its form along its entirety. Now, a vector $\vec{\xi}$ that satisfies $(\pounds_{\vec{\xi}}g)_{ij} = 0$ is called a Killing vector and eq. (9.3.6) is known as Killing's equation. We may then summarize:

> A geometry enjoys an isometry along $\vec{\xi}$ if and only if $\vec{\xi}$ is a Killing vector satisfying eq. (9.3.6) in the given region of space.

[83] We reiterate, by the same *form*, we mean $g_{ij}(\vec{x})$ and $g_{ij}(\vec{x}')$ are the same functions if we treat $\vec{x}$ and $\vec{x}'$ as dummy variables. For example, $g_{33}(r,\theta) = (r\sin\theta)^2$ and $g_{3'3'}(r',\theta') = (r'\sin\theta')^2$ in the 2-sphere metric.

The Lie derivative of a metric along some vector $\vec{\xi}$, as given in eq. (9.3.5), is linear in $\vec{\xi}$. That implies, if $\vec{\xi}$ and $\vec{\chi}$ are two vector fields and a and b are *constants*, the linearity property

$$(\pounds_{a\cdot\vec{\xi}+b\cdot\vec{\chi}}g)_{ij} = a(\pounds_{\vec{\xi}}g)_{ij} + b(\pounds_{\vec{\chi}}g)_{ij}. \tag{9.3.7}$$

In turn, we may assert: A *constant* linear superposition of Killing vectors returns another Killing vector. It turns out, for a fixed dimension D, there can be at most $D(D+1)/2$ linearly independent Killing vectors – we shall explore the consequences of this in §(9.8).

Remark In the above 'general covariance' discussion, I emphasized the importance of expressing geometric or physical laws in the same form in all coordinate systems. You may therefore ask, can equations (9.3.5) and (9.3.6) be re-phrased as tensor equations? For, otherwise, how do we know the notion of symmetry in curved space(time) is itself a coordinate independent concept? See Problem (9.35) for an answer.

Problem 9.13. Can you justify the following statement?

> "If the metric g_{ij} is independent of one of the coordinates, say x^k, then ∂_k is a Killing vector of the geometry."

From this, explain why $\{\partial_k = \delta^j_k \partial_j | k = 1, 2, \dots, D\}$ are Killing vectors of flat spacetime $g_{ij} = \delta_{ij}$ written in Cartesian coordinates. □

Problem 9.14. Angular Momentum in Flat Space Verify the following vector fields written in Cartesian coordinates,

$$J^{ij} \equiv x^i \partial_j - x^j \partial_i = x^{[i}\delta^{j]k}\partial_k, \tag{9.3.8}$$

are Killing vectors for a fixed pair of distinct (i, j), with $i \neq j$. (In §(4) you learned how these are proportional to the generators of rotations.)

The i and j run over 1 through D, and since J^{ij} is anti-symmetric under $i \leftrightarrow j$ interchange, there are $(D^2 - D)/2$ of these Killing vector fields. They correspond to 'orbital' angular momentum in quantum mechanics. □

Problem 9.15. Angular momentum 'generators' and 2-sphere isometries The generators of rotation in 3D space are proportional to the following vectors:

$$J^{(1)} = -\sin(\phi)\partial_\theta - \cos(\phi)\cot(\theta)\partial_\phi, \tag{9.3.9}$$

$$J^{(2)} = \cos(\phi)\partial_\theta - \sin(\phi)\cot(\theta)\partial_\phi, \tag{9.3.10}$$

$$J^{(3)} = \partial_\phi. \tag{9.3.11}$$

In 3D, convert eq. (9.3.8) into spherical coordinates (r, θ, ϕ) and verify that J^{ij} in fact are equivalent to equations (9.3.9), (9.3.10), and (9.3.11). (See §(5.6.2) for a discussion, or Problem (9.16) below. Briefly: $J^{(1)}$ generates rotations on the $(2,3)$ plane; $J^{(2)}$ on the $(1,3)$ plane; and $J^{(3)}$ on the $(1,2)$ plane.) Next, verify directly that they satisfy the Killing equation (9.3.6) on the metric of the unit 2−sphere centered at $\vec{x} = \vec{0}$ in 3D flat space: $\mathrm{d}\ell^2 = \mathrm{d}\theta^2 + (\sin\theta)^2\mathrm{d}\phi^2$. □

Problem 9.16. Generator of Rotation Around $\widehat{n}$-Axis in 3D Let x^i be Cartesian coordinates and (r, θ, ϕ) be spherical ones in flat 3D space. If $\widehat{R}(\varphi \cdot \widehat{n})$ denotes the rotation matrix – see eq. (5.6.16) – implementing rotation by angle φ around the unit

$$\widehat{n}(\alpha, \beta) \equiv (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha) \tag{9.3.12}$$

axis; explain (for e.g., using geometric considerations) why the tangent vector to the flow generated by $\widehat{R}$ at a given location $\vec{x}$ is given by

$$J^i\partial_i \equiv \left.\frac{\mathrm{d}}{\mathrm{d}\varphi}\right|_{\vec{x}} = \left.\frac{\mathrm{d}x^i}{\mathrm{d}\varphi}\right|_{\varphi=0} \partial_{x^i}, \tag{9.3.13}$$

$$\frac{\mathrm{d}x^i}{\mathrm{d}\varphi} = \left(\frac{\mathrm{d}\widehat{R}(\varphi=0)}{\mathrm{d}\varphi}\right)^i{}_j x^j. \tag{9.3.14}$$

where in spherical coordinate basis

$$J^{a'} \cdot \partial_{(r,\theta,\phi)^a} = \frac{\partial(r,\theta,\phi)^a}{\partial x^b}\left(\frac{\mathrm{d}\widehat{R}(\varphi=0)}{\mathrm{d}\varphi}\right)^b{}_j x^j(r,\theta,\phi)\partial_{(r,\theta,\phi)^a}. \tag{9.3.15}$$

Check that $\vec{J}$ in Cartesian $\vec{x}-$basis is also given by the following vector calculus cross product:

$$\vec{J} = \frac{\mathrm{d}\vec{x}}{\mathrm{d}\varphi} = r\widehat{n} \times \widehat{r}, \tag{9.3.16}$$

$$\widehat{r}(\theta, \phi) \equiv (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{9.3.17}$$

Can you explain *why* this is true?

Use the rotation matrix in eq. (5.6.16) to deduce

$$J^{a'}(r,\theta,\phi) = r\left(0, \sin(\alpha)\sin(\beta-\phi), \cos(\alpha) - \cos(\beta-\phi)\cot(\theta)\sin(\alpha)\right)^{\mathrm{T}}. \tag{9.3.18}$$

You should check that this in fact recovers equations (9.3.9), (9.3.10), and (9.3.11). Since this tangent vector 'generates' rotation around the $\widehat{n}-$axis, which does not involve displacements in the radial direction, we may view

$$J^{\mathrm{A}}(\theta,\phi)\partial_{\mathrm{A}} = r\sin(\alpha)\sin(\beta-\phi)\partial_\theta + r\left(\cos(\alpha) - \cos(\beta-\phi)\cot(\theta)\sin(\alpha)\right)\partial_\phi, \tag{9.3.19}$$

as a tangent vector residing on the $2-$sphere. Since rotation is a symmetry operation, verify that this J^{A} satisfies Killing's equation (9.3.6) on $\mathbb{S}^2$. $\square$

Problem 9.17. Commutation Relations If A and B are vector fields, so that $Af(\vec{x}) \equiv A^i\partial_i f(\vec{x})$ for arbitrary scalar functions $\{f\}$, show that

$$[A, B]\, f \equiv A^i\partial_i\left(B^j\partial_j f\right) - B^i\partial_i\left(A^j\partial_j f\right) \tag{9.3.20}$$

$$= \left(A^i\partial_i B^j - B^i\partial_i A^j\right)\partial_j f. \tag{9.3.21}$$

That is, the commutator of two vector fields A and B is a vector field whose jth component is $A^i\partial_i B^j - B^i\partial_i A^j$.

$$[A,B]^j = A^i\partial_i B^j - B^i\partial_i A^j \tag{9.3.22}$$

Referring to the Killing vectors in equations (9.3.9) through (9.3.11), if we now define $L^{(a)} \equiv -iJ^{(a)}$, where $a \in \{1,2,3\}$, show that these $\{L^{(a)}\}$ obey the angular momentum algebra

$$\left[L^{(i)}, L^{(j)}\right] = i\epsilon^{ijk}L^{(k)}. \tag{9.3.23}$$

Since these $(J^{(1)}, J^{(2)}, J^{(3)})$ produce infinitesimal rotations, this problem is a direct verification that the geometry of the 2−sphere is invariant under rotations. □

Lie Derivative on Tensors Instead of taking the Lie derivative of the metric, we may also consider the coordinate transformation of arbitrary tensors induced by the infinitesimal displacement in equations (9.3.1) and (9.3.2). For instance, we have

$$A_i(\vec{x})\mathrm{d}x^i = A_i\left(\vec{x}' + \vec{\xi}(\vec{x}') + \dots\right)\left(\delta^i_a + \partial_{a'}\xi^i(\vec{x}')\right)\mathrm{d}x'^a \tag{9.3.24}$$

$$= A_i(\vec{x}')\mathrm{d}x'^i + \xi^{a'}(\vec{x}')\partial_{a'}A_i(\vec{x}')\mathrm{d}x'^i + \partial_{i'}\xi^{a'}(\vec{x}')A_a(\vec{x}')\mathrm{d}x'^i + \mathcal{O}(\vec{\xi}^2) \tag{9.3.25}$$

$$\equiv \left(A_i(\vec{x}\to\vec{x}') + (\pounds_{\vec{\xi}}A)_i(\vec{x}') + \mathcal{O}(\vec{\xi}^2)\right)\mathrm{d}x'^i \tag{9.3.26}$$

$$(\pounds_{\vec{\xi}}A)_i(\vec{x}') \equiv \xi^{a'}(\vec{x}')\partial_{a'}A_i(\vec{x}') + \partial_{i'}\xi^{a'}(\vec{x}')A_a(\vec{x}'); \tag{9.3.27}$$

and

$$v^i(\vec{x})\partial_{x^i} = v^i\left(\vec{x}' + \vec{\xi}(\vec{x}') + \dots\right)\left(\delta^a_i - \partial_{i'}\xi^{a'}(\vec{x}')\right)\partial_{x'^a} \tag{9.3.28}$$

$$= v^i(\vec{x}')\partial_{x'^i} + \xi^a(\vec{x}')\partial_{a'}v^i(\vec{x}')\partial_{x'^i} - \partial_{a'}\xi^i(\vec{x}')v^a(\vec{x}')\partial_{x'^i} + \mathcal{O}(\vec{\xi}^2) \tag{9.3.29}$$

$$\equiv \left(v^i(\vec{x}\to\vec{x}') + (\pounds_{\vec{\xi}}v)^i(\vec{x}') + \mathcal{O}(\vec{\xi}^2)\right)\partial_{x'^i} \tag{9.3.30}$$

$$(\pounds_{\vec{\xi}}v)^{i'}(\vec{x}') \equiv \xi^{a'}(\vec{x}')\partial_{a'}v^i(\vec{x}') - v^a(\vec{x}')\partial_{a'}\xi^{i'}(\vec{x}') \tag{9.3.31}$$

$$= \left[\vec{\xi}, \vec{v}\right]^{i'}; \tag{9.3.32}$$

where the last equality follows from the commutator definition in eq. (9.3.22). This is why $[A,B] = \pounds_A B$ is sometimes called the *Lie bracket*.

Problem 9.18. Lie Derivative of a General Tensor If $\vec{x}$ and $\vec{x}'$ are infinitesimally nearby coordinate systems related via eq. (9.3.1), show that $T^{i_1\dots i_N}{}_{j_1\dots j_M}(\vec{x})$ (the components of a given tensor in the $\vec{x}$ coordinate basis) and $T^{i'_1\dots i'_N}{}_{j'_1\dots j'_M}(\vec{x}')$ (the components of the same tensor but in the $\vec{x}'$ coordinate basis) are in turn related via

$$T^{i'_1\dots i'_N}{}_{j'_1\dots j'_M}(\vec{x}') = T^{i_1\dots i_N}{}_{j_1\dots j_M}(\vec{x}\to\vec{x}') + \left(\pounds_{\vec{\xi}}T\right)^{i_1\dots i_N}{}_{j_1\dots j_M}(\vec{x}\to\vec{x}') + \mathcal{O}(\vec{\xi}^2); \tag{9.3.33}$$

where the Lie derivative of the tensor reads

$$\left(\pounds_{\vec{\xi}}T\right)^{i_1\dots i_N}{}_{j_1\dots j_M}(\vec{x}) = \xi^l\partial_l T^{i_1\dots i_N}{}_{j_1\dots j_M}$$

$$- T^{l i_2 \dots i_N}{}_{j_1 \dots j_M} \partial_l \xi^{i_1} - \dots - T^{i_1 \dots i_{N-1} l}{}_{j_1 \dots j_M} \partial_l \xi^{i_N}$$
$$+ T^{i_1 \dots i_N}{}_{l j_2 \dots j_M} \partial_{j_1} \xi^l + \dots + T^{i_1 \dots i_N}{}_{j_1 \dots j_{M-1} l} \partial_{j_M} \xi^l. \tag{9.3.34}$$

The $\vec{x} \to \vec{x}'$ on the right hand side of eq. (9.3.33) means, the tensor $T^{i_1 \dots i_N}{}_{j_1 \dots j_M}$ and its Lie derivative are to be computed in the $\vec{x}-$coodinate basis – but $\vec{x}$ is to be replaced, component-by-component, with $\vec{x}'$ afterwards. □

What are coordinates? The commutator, or Lie bracket, of two vector fields we first encountered in Problem (9.17) and its relation to the Lie derivative in equations (9.3.31) and (9.3.32) allow us to now address a seemingly basic question: What are coordinates? Coordinates are a set of independent (real) parameters that allow us to unambiguously specify a given point in space. This, in particular, means we must be able to vary any of the coordinates without varying others; so as to move from one location to another without restriction. On the 2D flat space where $\vec{x}(\text{Cartesian}) = r(\cos\phi, \sin\phi)$, for instance, it must be possible to vary the azimuth angle ϕ without varying the radius r; and vice versa. On the other hand, we have just seen that the coordinate transformation of a vector $\vec{v}$ field induced by an infinitesimal displacement along some other vector $\vec{w}$, is given by their Lie bracket $[\vec{v}, \vec{w}]$. Hence, if we want to associate $v^i \partial_i \equiv \mathrm{d}/\mathrm{d}x^1$ with the derivative along some coordinate x^1 and $w^i \partial_i \equiv \mathrm{d}/\mathrm{d}x^2$ along some other coordinate x^2; their Lie bracket must vanish – i.e., there must be no change to $\vec{v}$ if we displace along $\vec{w}$ and vice versa, if their integral curves are coordinate lines. This generalizes to the following theorem:[84]

> A set of $1 < N \leq D$ vector fields $\{\mathrm{d}/\mathrm{d}\xi^i\}$ form a coordinate basis in the $D-$dimensional space they inhabit, if and only if they commute.

To elaborate: if these N vector fields commute, we may integrate them to find a $N-$dimensional coordinate grid within the $D-$dimensional space. Conversely, we are already accustomed to the fact that the partial derivatives with respect to the coordinates of space do, of course, commute amongst themselves. On the other hand, if these set of N vector fields do not necessarily commute, but are closed under their Lie bracket, namely

$$\left[\frac{\mathrm{d}}{\mathrm{d}\xi^i}, \frac{\mathrm{d}}{\mathrm{d}\xi^j}\right] = f_{ij}{}^k \frac{\mathrm{d}}{\mathrm{d}\xi^k}, \tag{9.3.35}$$

then it is still possible to find a set of N coordinates $\{y^i\}$ that parametrize the $N-$dimensional subspace whose tangent vectors are these $\{\mathrm{d}/\mathrm{d}\xi^i\}$; except the coordinates $\vec{y}$ are not necessarily the parameters $\vec{\xi}$.

Problem 9.19. Example of non-commuting vector fields[85] In 2D flat space, starting from Cartesian coordinates x^i, we may convert to cylindrical coordinates

$$(x^1, x^2) = r(\cos\phi, \sin\phi). \tag{9.3.36}$$

[84] See, for instance, Schutz [23] for a pedagogical discussion.
[85] Schutz [23] Exercise 2.1

The pair of vector fields $(\partial_r, \partial_\phi)$ do form a coordinate basis – it is possible to hold r fixed while going along the integral curve of ∂_ϕ and vice versa. However, show via a direct calculation that the following commutator involving the unit vector fields $\widehat{r}$ and $\widehat{\phi}$ is not zero:

$$\left[\widehat{r}, \widehat{\phi}\right] f(r, \phi) = -r^{-2}\partial_\phi f \neq 0; \tag{9.3.37}$$

where

$$\widehat{r} \equiv \cos(\phi)\partial_{x^1} + \sin(\phi)\partial_{x^2}, \tag{9.3.38}$$

$$\widehat{\phi} \equiv -\sin(\phi)\partial_{x^1} + \cos(\phi)\partial_{x^2}. \tag{9.3.39}$$

Therefore $\widehat{r}$ and $\widehat{\phi}$ do not form a coordinate basis. □

9.4 Orthonormal Frames

So far, we have been writing tensors in the coordinate basis – the basis vectors of our tensors are formed out of tensor products of $\{\mathrm{d}x^i\}$ and $\{\partial_i\}$. To interpret components of tensors, however, we need them written in an orthonormal basis. This amounts to using a uniform set of measuring sticks on all axes, i.e., a local set of (non-coordinate) Cartesian axes where one "tick mark" on each axis translates to the same length. Moreover, writing vectors $V^i\partial_i = V^{\widehat{i}}\widehat{e}_i$ in an orthonormal basis $\{\widehat{e}_i\}$ in flat space(time) reduces to the vector calculus practice of using unit length mutually perpendicular basis vectors.

As an example, suppose we wish to describe some fluid's velocity $v^x\partial_x + v^y\partial_y$ on a 2 dimensional flat space. In Cartesian coordinates $v^x(x, y)$ and $v^y(x, y)$ describe the velocity at some point $\vec{\xi} = (x, y)$ flowing in the x- and y-directions respectively. Suppose we used polar coordinates, however,

$$\xi^i = r(\cos\phi, \sin\phi). \tag{9.4.1}$$

The metric would read

$$(\mathrm{d}\ell)^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2. \tag{9.4.2}$$

The velocity now reads $v^r(\vec{\xi})\partial_r + v^\phi(\vec{\xi})\partial_\phi$, where $v^r(\vec{\xi})$ has an interpretation of "rate of flow in the radial direction". However, notice the dimensions of the v^ϕ is not even the same as that of v^r; if v^r were of [Length/Time], then $[v^\phi]$ is of 1/Time. At this point we recall – just as $\mathrm{d}r$ (which is dual to ∂_r) can be interpreted as an infinitesimal length in the radial direction, the arc length $r\mathrm{d}\phi$ (which is dual to $(1/r)\partial_\phi$) is the corresponding one in the perpendicular azimuthal direction. Let us introduce the following notation for the vector fields

$$\varepsilon_{\widehat{r}} \equiv \varepsilon_{\widehat{r}}{}^i\partial_i \equiv \partial_r \qquad \Leftrightarrow \qquad \varepsilon_{\widehat{r}}{}^i = \delta^i_r; \tag{9.4.3}$$

and

$$\varepsilon_{\widehat{\phi}} \equiv \varepsilon_{\widehat{\phi}}{}^i\partial_i \equiv r^{-1}\partial_\phi \qquad \Leftrightarrow \qquad \varepsilon_{\widehat{\phi}}{}^i = r^{-1}\delta^i_\phi. \tag{9.4.4}$$

A direct calculation now reveals their orthonormal character:

$$\langle \varepsilon_{\widehat{r}} | \, \varepsilon_{\widehat{r}} \rangle = \langle \partial_r | \, \partial_r \rangle = g_{rr} = 1 \tag{9.4.5}$$

$$\left\langle \varepsilon_{\widehat{\phi}} \middle| \, \varepsilon_{\widehat{\phi}} \right\rangle = r^{-2} \langle \partial_\phi | \, \partial_\phi \rangle = \frac{g_{\phi\phi}}{r^2} = 1 \tag{9.4.6}$$

$$\left\langle \varepsilon_{\widehat{r}} | \, \varepsilon_{\widehat{\phi}} \right\rangle = \langle \partial_r | \, r^{-1} \partial_\phi \rangle = r^{-1} g_{r\phi} = 0. \tag{9.4.7}$$

The $\varepsilon_{\widehat{r}}$ and $\varepsilon_{\widehat{\phi}}$ here may therefore be identified with, respectively, the unit vectors $\widehat{r}$ and $\widehat{\phi}$ in multi-variable calculus.

Using the considerations thus far as a guide, we would now express the velocity at $\vec{\xi}$ as

$$\begin{aligned} v &= v^{\widehat{r}} \varepsilon_{\widehat{r}} + v^{\widehat{\phi}} \varepsilon_{\widehat{\phi}} & (9.4.8) \\ &= v^r \frac{\partial}{\partial r} + (r \cdot v^\phi) \left(\frac{1}{r} \frac{\partial}{\partial \phi} \right) & (9.4.9) \\ &= \frac{\mathrm{d}r}{\mathrm{d}\lambda} \frac{\partial}{\partial r} + \left(r \cdot \frac{\mathrm{d}\phi}{\mathrm{d}\lambda} \right) \left(\frac{1}{r} \frac{\partial}{\partial \phi} \right), & (9.4.10) \end{aligned}$$

so that now $v^{\widehat{\phi}} \equiv r \cdot v^\phi = (r \mathrm{d}\phi)/\mathrm{d}\lambda$ may be interpreted as the velocity in the azimuthal direction.

More formally, given a (real, symmetric) metric g_{ij} we may always find a orthogonal transformation $O^a{}_i$ that diagonalizes it; and by absorbing into this transformation the eigenvalues of the metric, the orthonormal frame fields emerge:

$$\begin{aligned} g_{ij} \mathrm{d}x^i \mathrm{d}x^j &= \sum_{a,b} \left(O^a{}_i \cdot \lambda_a \delta_{ab} \cdot O^b{}_j \right) \mathrm{d}x^i \mathrm{d}x^j \\ &= \sum_{a,b} \left(\sqrt{\lambda_a} O^a{}_i \cdot \delta_{ab} \cdot \sqrt{\lambda_b} O^b{}_j \right) \mathrm{d}x^i \mathrm{d}x^j \\ &= \left(\delta_{ab} \varepsilon^{\widehat{a}}{}_i \varepsilon^{\widehat{b}}{}_j \right) \mathrm{d}x^i \mathrm{d}x^j = \delta_{ab} \left(\varepsilon^{\widehat{a}}{}_i \mathrm{d}x^i \right) \left(\varepsilon^{\widehat{b}}{}_j \mathrm{d}x^j \right), & (9.4.11) \\ \varepsilon^{\widehat{a}}{}_i &\equiv \sqrt{\lambda_a} O^a{}_i, \qquad \text{(no sum over } a\text{)}. & (9.4.12) \end{aligned}$$

In the first equality, we have exploited the fact that any real symmetric matrix g_{ij} can be diagonalized by an appropriate orthogonal matrix $O^a{}_i$, with real eigenvalues $\{\lambda_a\}$; in fact, from matrix algebra, $O^a{}_i$ is the ith component of the ath eigenvector of the matrix g_{ij}. For the second equality, we have exploited the assumption that we are working in Riemannian spaces, where all eigenvalues of the metric are positive,[86] to take the positive square roots of the eigenvalues; in the third we have defined the orthonormal frame vector fields as $\varepsilon^{\widehat{a}}{}_i = \sqrt{\lambda_a} O^a{}_i$, with no sum over a. Finally, from eq. (9.4.11) and by defining the infinitesimal lengths

$$\varepsilon^{\widehat{a}} \equiv \varepsilon^{\widehat{a}}{}_i \mathrm{d}x^i, \tag{9.4.13}$$

we arrive at the following curved space parallel to Pythagoras' theorem in flat space:

$$(\mathrm{d}\ell)^2 = g_{ij} \mathrm{d}x^i \mathrm{d}x^j = \left(\varepsilon^{\widehat{1}} \right)^2 + \left(\varepsilon^{\widehat{2}} \right)^2 + \dots + \left(\varepsilon^{\widehat{D}} \right)^2. \tag{9.4.14}$$

[86] As opposed to semi-Riemannian/Lorentzian spaces, where the eigenvalue associated with the 'time' direction has a different sign from the rest.

The metric components are now

$$g_{ij} = \delta_{ab}\varepsilon^{\widehat{a}}{}_{i}\varepsilon^{\widehat{b}}{}_{j} = \left(\widehat{\varepsilon}^{\mathrm{T}}\widehat{\varepsilon}\right)_{ij}, \tag{9.4.15}$$

where in the second equality $\varepsilon^{\widehat{a}}{}_{i}$ is now regarded as the $a-$th row and $i-$th column of the matrix $\widehat{\varepsilon}$. Using $\det \widehat{A} \cdot \widehat{B} = \det \widehat{A} \det \widehat{B}$ and $\det \widehat{A}^{\mathrm{T}} = \det \widehat{A}$, the metric determinant now reads

$$\det g_{ij} = (\det \widehat{\varepsilon})^2 . \tag{9.4.16}$$

We say the metric on the right hand side of eq. (9.4.11) is written in an orthonormal frame, because in this basis $\{\varepsilon^{\widehat{a}}{}_{i}\mathrm{d}x^i | a = 1, 2, \dots, D\}$, the metric components are identical to the flat Cartesian ones. We have put a $\widehat{\cdot}$ over the a-index, to distinguish from the i-index, because the latter transforms as a tensor

$$\varepsilon^{\widehat{a}}{}_{i}(\vec{\xi}) = \varepsilon^{\widehat{a}}{}_{j}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^j(\vec{\xi})}{\partial \xi^i}. \tag{9.4.17}$$

This also implies the i-index can be moved using the metric:

$$\varepsilon^{\widehat{a}i}(\vec{x}) = g^{ij}(\vec{x})\varepsilon^{\widehat{a}}{}_{j}(\vec{x}), \qquad \varepsilon^{\widehat{a}}{}_{i}(\vec{x}) = g_{ij}(\vec{x})\varepsilon^{\widehat{a}j}(\vec{x}). \tag{9.4.18}$$

We may readily check that eq. (9.4.17) is the correct transformation rule because it is equivalent to eq. (9.2.7).

$$g_{a'b'}(\vec{\xi}) = \delta_{mn}\varepsilon^{\widehat{m}}{}_{a'}(\vec{\xi})\varepsilon^{\widehat{n}}{}_{b'}(\vec{\xi}) \tag{9.4.19}$$

$$= \delta_{mn}\varepsilon^{\widehat{m}}{}_{i}(\vec{x})\varepsilon^{\widehat{n}}{}_{j}(\vec{x})\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b} = g_{ij}\left(\vec{x}\right)\frac{\partial x^i}{\partial \xi^a}\frac{\partial x^j}{\partial \xi^b} \tag{9.4.20}$$

The $\widehat{a}$ index does not transform under coordinate transformations. But it can be rotated by an orthogonal matrix $R^{\widehat{a}}{}_{\widehat{b}}(\vec{x})$, which itself can depend on the space coordinates, while keeping the metric in eq. (9.4.11) the same object. By orthogonal matrix, we mean any $\widehat{R}$ that obeys

$$\widehat{R}^{\widehat{a}}{}_{\widehat{c}}\delta_{ab}\widehat{R}^{\widehat{b}}{}_{\widehat{f}} = \delta_{cf} \tag{9.4.21}$$

$$\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I}. \tag{9.4.22}$$

Upon the replacement

$$\varepsilon^{\widehat{a}}{}_{i}(\vec{x}) \to \widehat{R}^{\widehat{a}}{}_{\widehat{b}}(\vec{x})\varepsilon^{\widehat{b}}{}_{i}(\vec{x}), \tag{9.4.23}$$

we discover the invariance

$$g_{ij}\mathrm{d}x^i\mathrm{d}x^j \to \left(\delta_{ab}\widehat{R}^{\widehat{a}}{}_{\widehat{c}}\widehat{R}^{\widehat{b}}{}_{\widehat{f}}\right)\varepsilon^{\widehat{c}}{}_{i}\varepsilon^{\widehat{f}}{}_{j}\mathrm{d}x^i\mathrm{d}x^j \tag{9.4.24}$$

$$= \delta_{cf}\varepsilon^{\widehat{c}}{}_{i}\varepsilon^{\widehat{f}}{}_{j}\mathrm{d}x^i\mathrm{d}x^j = g_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{9.4.25}$$

Notice, this invariance holds even if the rotation $\widehat{R}^{\widehat{a}}{}_{\widehat{b}}(\vec{x})$ depended on space. The interpretation of eq. (9.4.23) is that the choice of local Cartesian-like (non-coordinate) axes are not unique; just as the Cartesian coordinate system in flat space can be redefined through a rotation $\widehat{R}$ obeying $\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I}$, these local axes can also be rotated freely at each and every $\vec{x}$. It is a consequence of this O_D symmetry that upper and lower orthonormal frame indices actually transform the same way. We begin by demanding that rank-1 tensors in an orthonormal frame transform as

$$V^{\widehat{a}'} = \widehat{R}^{\widehat{a}}{}_{\widehat{c}} V^{\widehat{c}}, \qquad V_{\widehat{a}'} = (\widehat{R}^{-1})^{\widehat{f}}{}_{\widehat{a}} V_{\widehat{f}} \tag{9.4.26}$$

so that

$$V^{\widehat{a}'} V_{\widehat{a}'} = V^{\widehat{a}} V_{\widehat{a}}. \tag{9.4.27}$$

But $\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I}$ means $\widehat{R}^{-1} = \widehat{R}^{\mathrm{T}}$ and thus the ath row and cth column of the inverse, namely $(\widehat{R}^{-1})^{\widehat{a}}{}_{\widehat{c}}$, is equal to the cth row and ath column of $\widehat{R}$ itself: $(\widehat{R}^{-1})^{\widehat{a}}{}_{\widehat{c}} = \widehat{R}^{\widehat{c}}{}_{\widehat{a}}$.

$$V_{\widehat{a}'} = \sum_f \widehat{R}^{\widehat{a}}{}_{\widehat{f}} V_{\widehat{f}}. \tag{9.4.28}$$

In other words, $V_{\widehat{a}}$ transforms just like $V^{\widehat{a}}$.

We have thus shown that the orthonormal frame index is moved by the Kronecker delta; $V^{\widehat{a}'} = V_{\widehat{a}'}$ for any vector written in an orthonormal frame, and in particular,

$$\varepsilon^{\widehat{a}}{}_i(\vec{x}) = \delta^{ab}\varepsilon_{\widehat{b}i}(\vec{x}) = \varepsilon_{\widehat{a}i}(\vec{x}). \tag{9.4.29}$$

In addition to the orthonormal basis 1−forms in eq. (9.4.13), we have thus also discovered the orthonormal basis vector fields

$$\varepsilon_{\widehat{a}} \equiv \varepsilon_{\widehat{a}}{}^i \partial_i. \tag{9.4.30}$$

Let us check that these new basis frame fields $\{\varepsilon_{\widehat{a}}\}$ and $\{\varepsilon^{\widehat{a}}\}$ are indeed of unit length.

$$\varepsilon^{\widehat{f}}{}_j \varepsilon^{\widehat{b}j} = \varepsilon^{\widehat{f}}{}_j \varepsilon^{\widehat{b}}{}_k g^{jk} = \delta^{fb}, \tag{9.4.31}$$

$$\varepsilon_{\widehat{f}}{}^j \varepsilon_{\widehat{b}j} = \varepsilon_{\widehat{f}}{}^j \varepsilon_{\widehat{b}}{}^k g_{jk} = \delta_{fb}. \tag{9.4.32}$$

To understand this we begin with the diagonalization of the metric, $\delta_{cf}\varepsilon^{\widehat{c}}{}_i \varepsilon^{\widehat{f}}{}_j = g_{ij}$. Contracting both sides with the orthonormal frame vector $\varepsilon^{\widehat{b}j}$,

$$\delta_{cf}\varepsilon^{\widehat{c}}{}_i \varepsilon^{\widehat{f}}{}_j \varepsilon^{\widehat{b}j} = \varepsilon^{\widehat{b}}{}_i, \tag{9.4.33}$$

$$(\varepsilon^{\widehat{b}j}\varepsilon_{\widehat{f}j})\varepsilon^{\widehat{f}}{}_i = \varepsilon^{\widehat{b}}{}_i. \tag{9.4.34}$$

If we let M denote the matrix $M^b{}_f \equiv (\varepsilon^{\widehat{b}j}\varepsilon_{\widehat{f}j})$, then we have $i = 1, 2, \dots, D$ matrix equations $M \cdot \varepsilon_i = \varepsilon_i$. As long as the determinant of g_{ab} is non-zero, then $\{\varepsilon_i\}$ are linearly independent vectors spanning $\mathbb{R}^D$ (see eq. (9.4.16)). Since every ε_i is an eigenvector of M with eigenvalue one, that means $M = \mathbb{I}$, and we have proved eq. (9.4.31).

To summarize,

$$g_{ij} = \delta_{ab}\varepsilon^{\widehat{a}}{}_i \varepsilon^{\widehat{b}}{}_j, \qquad g^{ij} = \delta^{ab}\varepsilon_{\widehat{a}}{}^i \varepsilon_{\widehat{b}}{}^j,$$
$$\delta_{ab} = g_{ij}\varepsilon_{\widehat{a}}{}^i \varepsilon_{\widehat{b}}{}^j, \qquad \delta^{ab} = g^{ij}\varepsilon^{\widehat{a}}{}_i \varepsilon^{\widehat{b}}{}_j. \tag{9.4.35}$$

Problem 9.20. Orthonormal Frame from Inverse Metric Starting from the real and symmetric inverse metric g^{ij}, explain why

$$g^{ij} = \sum_{1 \leq a,b \leq D} \frac{O^a{}_i}{\sqrt{\lambda_a}} \delta_{ab} \frac{O^b{}_j}{\sqrt{\lambda_b}}, \tag{9.4.36}$$

where the orthogonal transformation $O^a{}_i$ is the same as that in eq. (9.4.11). Use $\varepsilon_{\widehat{a}}{}^i = g^{ij} \varepsilon^{\widehat{a}}{}_j$ to show that

$$\varepsilon_{\widehat{a}}{}^i = O^a{}_i / \sqrt{\lambda_a}. \tag{9.4.37}$$

Why is this consistent with

$$g^{ij} \partial_i \partial_j = \delta^{ab} \varepsilon_{\widehat{a}}{}^i \varepsilon_{\widehat{b}}{}^j \partial_i \partial_j = (\varepsilon_{\widehat{1}})^2 + \cdots + \left(\varepsilon_{\widehat{D}}\right)^2 \tag{9.4.38}$$

in eq. (9.4.35)? □

Problem 9.21. Verify the orthonormal conditions

$$\left\langle \varepsilon^{\widehat{a}} \middle| \varepsilon^{\widehat{b}} \right\rangle = \delta^{ab} \qquad \text{and} \qquad \left\langle \varepsilon_{\widehat{a}} \middle| \varepsilon_{\widehat{b}} \right\rangle = \delta_{ab}, \tag{9.4.39}$$

where $\varepsilon^{\widehat{a}} \equiv \varepsilon^{\widehat{a}}{}_i \mathrm{d}x^i$ and $\varepsilon_{\widehat{a}} \equiv \varepsilon_{\widehat{a}}{}^i \partial_i$. (Note: $\varepsilon_{\widehat{a}} \neq \delta_{ab} \varepsilon^{\widehat{b}}$.) What is $\left\langle \varepsilon^{\widehat{a}} \middle| \varepsilon_{\widehat{b}} \right\rangle$? □

Problem 9.22. Rotation Matrix If $\{\varepsilon^{\widehat{a}}{}_i\}$ and $\{\varepsilon^{\widehat{a}'}{}_i\}$ are two distinct sets of orthonormal frame fields, let us define the $D \times D$ matrix

$$\widehat{R}^{ab} \equiv \varepsilon^{\widehat{a}}{}_i \varepsilon^{\widehat{b}' i}. \tag{9.4.40}$$

Prove that $\widehat{R}^{ab}$ is not only an orthogonal matrix, it relates one set of the orthonormal frame fields to another via the relation

$$\varepsilon^{\widehat{a}}{}_i = \widehat{R}^{ab} \varepsilon^{\widehat{b}'}{}_i. \tag{9.4.41}$$

The next problem provides a simple example. □

Problem 9.23. Orthonormal Frame from Non-Orthogonal Coordinates In this problem, we shall work out the orthonormal frame fields in a non-orthogonal coordinate system. Starting from 2D flat space $\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2$, consider the transformation

$$x = x'^1 \qquad \text{and} \qquad y = x'^2 + \epsilon \cdot x'^1. \tag{9.4.42}$$

Show that

$$\mathrm{d}\ell^2 = (1+\epsilon^2)(\mathrm{d}x'^1)^2 + 2\epsilon \cdot \mathrm{d}x'^1 \mathrm{d}x'^2 + (\mathrm{d}x'^2)^2, \tag{9.4.43}$$

$$g_{i'j'}(\vec{x}') \doteq \begin{bmatrix} 1+\epsilon^2 & \epsilon \\ \epsilon & 1 \end{bmatrix}. \tag{9.4.44}$$

Treat the metric as a matrix – do not be too worried about the position (up versus down) of the indices in this problem – and demonstrate that orthonormal eigensystem is given by

$$g_{i'j'}v_{\mathrm{I}}^j = \lambda_{\mathrm{I}} v_{\mathrm{I}}^i \equiv \frac{2+\epsilon^2-\epsilon\sqrt{4+\epsilon^2}}{2} v_{\mathrm{I}}^i, \tag{9.4.45}$$

$$v_{\mathrm{I}}^j = \frac{\sqrt{2}}{\sqrt[4]{4+\epsilon^2}\sqrt{\sqrt{4+\epsilon^2}-\epsilon}} \left[\frac{\epsilon-\sqrt{4+\epsilon^2}}{2}, 1\right]^{\mathrm{T}}; \tag{9.4.46}$$

and

$$g_{i'j'}v_{\mathrm{II}}^j = \lambda_{\mathrm{II}} v_{\mathrm{II}}^i \equiv \frac{2+\epsilon^2+\epsilon\sqrt{4+\epsilon^2}}{2} v_{\mathrm{II}}^i, \tag{9.4.47}$$

$$v_{\mathrm{II}}^j = \frac{\sqrt{2}}{\sqrt[4]{4+\epsilon^2}\sqrt{\sqrt{4+\epsilon^2}+\epsilon}} \left[\frac{\epsilon+\sqrt{4+\epsilon^2}}{2}, 1\right]^{\mathrm{T}}. \tag{9.4.48}$$

In other words, recalling eq. (9.4.11),

$$g_{i'j'} = \left(\begin{bmatrix} \sqrt{\lambda_{\mathrm{I}}} v_{\mathrm{I}}^1 & \sqrt{\lambda_{\mathrm{I}}} v_{\mathrm{I}}^2 \\ \sqrt{\lambda_{\mathrm{II}}} v_{\mathrm{II}}^1 & \sqrt{\lambda_{\mathrm{II}}} v_{\mathrm{II}}^2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \sqrt{\lambda_{\mathrm{I}}} v_{\mathrm{I}}^1 & \sqrt{\lambda_{\mathrm{I}}} v_{\mathrm{I}}^2 \\ \sqrt{\lambda_{\mathrm{II}}} v_{\mathrm{II}}^1 & \sqrt{\lambda_{\mathrm{II}}} v_{\mathrm{II}}^2 \end{bmatrix} \right)_{ij} \tag{9.4.49}$$

and therefore

$$\varepsilon^{\mathrm{I}}{}_i = \sqrt{\lambda_{\mathrm{I}}} v_{\mathrm{I}}^i \tag{9.4.50}$$

$$\varepsilon^{\mathrm{II}}{}_i = \sqrt{\lambda_{\mathrm{II}}} v_{\mathrm{II}}^i. \tag{9.4.51}$$

Of course, the solution to the orthonormal frame fields is not unique; but this problem steps through the diagonalization process to illustrate the generic algorithm. In this case, an easier method is to write down the solution – by inspection – using the original (x, y) system:

$$\varepsilon^{\mathrm{I}'}{}_i \mathrm{d}x'^i = \mathrm{d}x = \mathrm{d}x'^1 \tag{9.4.52}$$

$$\varepsilon^{\mathrm{II}'}{}_i \mathrm{d}x'^i = \mathrm{d}y = \epsilon \cdot \mathrm{d}x'^1 + \mathrm{d}x'^2. \tag{9.4.53}$$

Can you show that $\{\varepsilon^{\mathrm{I}'}{}_i, \varepsilon^{\mathrm{II}'}{}_i\}$ are a rotated versions of $\{\varepsilon^{\mathrm{I}}{}_i, \varepsilon^{\mathrm{II}}{}_i\}$; i.e., find $\widehat{R}^{\mathrm{A}}{}_{\mathrm{B}}$, for A and B running over I and II, where $\varepsilon^{\mathrm{A}}{}_i = \widehat{R}^{\mathrm{A}}{}_{\mathrm{B}} \varepsilon^{\mathrm{B}'}{}_i$? Hint: $\varepsilon^{\mathrm{A}}{}_i \varepsilon^{\mathrm{B}}{}_j g^{ij} = \delta^{\mathrm{AB}}$ may be useful here. □

Tensor Components in Orthonormal Basis Now, any tensor with written in a coordinate basis can be converted to one in an orthonormal basis by contracting with the orthonormal frame fields $\varepsilon^{\widehat{a}}{}_i$ in eq. (9.4.11). For example, the velocity field in an orthonormal frame is

$$v^{\widehat{a}} = \varepsilon^{\widehat{a}}{}_i v^i. \tag{9.4.54}$$

For the two dimension example above,

$$(\mathrm{d}r)^2 + (r\mathrm{d}\phi)^2 = \delta_{rr}(\mathrm{d}r)^2 + \delta_{\phi\phi}(r\mathrm{d}\phi)^2, \tag{9.4.55}$$

allowing us to read off the only non-zero components of the orthonormal frame fields are

$$\varepsilon^{\widehat{r}}{}_{r} = 1, \qquad \varepsilon^{\widehat{\phi}}{}_{\phi} = r; \tag{9.4.56}$$

which in turn implies

$$v^{\widehat{r}} = \varepsilon^{\widehat{r}}{}_{r} v^{r} = v^{r}, \qquad v^{\widehat{\phi}} = \varepsilon^{\widehat{\phi}}{}_{\phi} v^{\phi} = r\ v^{\phi}. \tag{9.4.57}$$

More generally, what we are doing here is really switching from writing the same tensor in coordinates basis $\{\mathrm{d}x^i\}$ and $\{\partial_i\}$ to an orthonormal basis $\{\varepsilon^{\widehat{a}}{}_{i}\mathrm{d}x^i\}$ and $\{\varepsilon_{\widehat{a}}{}^{i}\partial_i\}$. For example,

$$T_{ijk}{}^{l} \left\langle \mathrm{d}x^i \right| \otimes \left\langle \mathrm{d}x^j \right| \otimes \left\langle \mathrm{d}x^k \right| \otimes \left| \partial_l \right\rangle = T_{\widehat{i}\widehat{j}\widehat{k}}{}^{\widehat{l}} \left\langle \varepsilon^{\widehat{i}} \right| \otimes \left\langle \varepsilon^{\widehat{j}} \right| \otimes \left\langle \varepsilon^{\widehat{k}} \right| \otimes \left| \varepsilon_{\widehat{l}} \right\rangle \tag{9.4.58}$$

$$\varepsilon^{\widehat{i}} \equiv \varepsilon^{\widehat{i}}{}_{a} \mathrm{d}x^a \qquad\qquad \varepsilon_{\widehat{i}} \equiv \varepsilon_{\widehat{i}}{}^{a}\partial_a. \tag{9.4.59}$$

To sum: the formula that converts a general tensor in a coordinate basis to the same in an orthonormal one is

$$T^{\widehat{a}_1 \dots \widehat{a}_M}{}_{\widehat{b}_1 \dots \widehat{b}_N} = T^{i_1 \dots i_M}{}_{j_1 \dots j_N} \varepsilon^{\widehat{a}_1}{}_{i_1} \dots \varepsilon^{\widehat{a}_M}{}_{i_M} \varepsilon_{\widehat{b}_1}{}^{j_1} \dots \varepsilon_{\widehat{b}_N}{}^{j_N}. \tag{9.4.60}$$

Problem 9.24. Explain why the 'inverse' transformation of eq. (9.4.60) is

$$T^{\widehat{a}_1 \dots \widehat{a}_M}{}_{\widehat{b}_1 \dots \widehat{b}_N} \varepsilon_{\widehat{a}_1}{}^{i_1} \dots \varepsilon_{\widehat{a}_M}{}^{i_M} \varepsilon^{\widehat{b}_1}{}_{j_1} \dots \varepsilon^{\widehat{b}_N}{}_{j_N} = T^{i_1 \dots i_M}{}_{j_1 \dots j_N}. \tag{9.4.61}$$

Hint: Insert eq. (9.4.60) into the left hand side of (9.4.61), followed by using the appropriate relation in eq. (9.4.35). □

Even though the physical dimension of the whole tensor $[T]$ must necessarily be consistent, because the $\{\mathrm{d}x^i\}$ and $\{\partial_i\}$ do not have the same dimensions – compare, for e.g., $\mathrm{d}r$ versus $\mathrm{d}\theta$ in spherical coordinates – the components of tensors in a coordinate basis do not all have the same dimensions, making their interpretation difficult. By using orthonormal frame fields as defined in eq. (9.4.59), we see that

$$\sum_a \left(\varepsilon^{\widehat{a}}\right)^2 = \delta_{ab} \varepsilon^{\widehat{a}}{}_{i} \varepsilon^{\widehat{b}}{}_{j} \mathrm{d}x^i \mathrm{d}x^j = g_{ij} \mathrm{d}x^i \mathrm{d}x^j \tag{9.4.62}$$

$$\left[\varepsilon^{\widehat{a}}\right] = \text{Length}; \tag{9.4.63}$$

and

$$\sum_a \left(\varepsilon_{\widehat{a}}\right)^2 = \delta^{ab} \varepsilon_{\widehat{a}}{}^{i} \varepsilon_{\widehat{b}}{}^{j} \partial_i \partial_j = g^{ij} \partial_i \partial_j \tag{9.4.64}$$

$$\left[\varepsilon_{\widehat{a}}\right] = 1/\text{Length}; \tag{9.4.65}$$

which in turn implies, for instance, the consistency of the physical dimensions of the orthonormal components $T_{\widehat{i}\widehat{j}\widehat{k}}{}^{\widehat{l}}$ in eq. (9.4.58):

$$[T_{\widehat{i}\widehat{j}\widehat{k}}{}^{\widehat{l}}][\varepsilon^{\widehat{i}}]^3[\varepsilon_{\widehat{l}}] = [T], \tag{9.4.66}$$

$$\left[T_{\widehat{i}\widehat{j}\widehat{k}}{}^{\widehat{l}}\right] = \frac{[T]}{\text{Length}^2}. \tag{9.4.67}$$

This consistency of physical dimensions of tensor components written in an orthonormal basis is a key reason why it is in such a basis – and not the coordinate ones – that allows for their physical or geometric interpretation.

(Curved) Dot Product Revisited We have already noted that the generalization of the dot product between two (tangent) vectors $\vec{U}$ and $\vec{V}$ at some location $\vec{x}$ is $\vec{U}(\vec{x}) \cdot \vec{V}(\vec{x}) \equiv g_{ij}(\vec{x})U^i(\vec{x})V^j(\vec{x})$. Previously, we have justified this interpretation by using a locally flat coordinate system. Alternatively, we may also exploit the orthonormal frame:

$$\vec{U}(\vec{x}) \cdot \vec{V}(\vec{x}) = \delta_{ij} U^{\widehat{i}}(\vec{x}) V^{\widehat{j}}(\vec{x}). \tag{9.4.68}$$

Problem 9.25. Find the orthonormal frame fields $\{\varepsilon^{\widehat{a}}{}_i\}$ in 3-dimensional Cartesian, Spherical and Cylindrical coordinate systems. Hint: Just like the 2D case above, by packaging the metric $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ appropriately, you can read off the frame fields without further work. □

Problem 9.26. Physical Dimension of Tensor Components in Orthonormal Basis Show that every component of a general $\binom{M}{N}$ tensor $T^{i_1 \dots i_M}{}_{j_1 \dots j_N}$, when written in an orthonormal frame, has the same physical dimension

$$\left[T^{\widehat{i}_1 \dots \widehat{i}_M}{}_{\widehat{j}_1 \dots \widehat{j}_N}\right] = [T] \cdot (\text{Length})^{M-N}. \tag{9.4.69}$$

Equivalently, each tensor has one and only one physical scale associated with it. □

9.5 Covariant derivatives, Parallel Transport, Geodesics

Covariant Derivative How do we take derivatives of tensors in such a way that we get back a tensor in return? To start, let us see that the partial derivative of a tensor is not a tensor. Consider

$$\begin{aligned}\frac{\partial T_j(\vec{\xi})}{\partial \xi^i} &= \frac{\partial x^a}{\partial \xi^i}\frac{\partial}{\partial x^a}\left(T_b\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^b}{\partial \xi^j}\right) \\ &= \frac{\partial x^a}{\partial \xi^i}\frac{\partial x^b}{\partial \xi^j}\frac{\partial T_b\left(\vec{x}(\vec{\xi})\right)}{\partial x^a} + \frac{\partial^2 x^b}{\partial \xi^j \partial \xi^i} T_b\left(\vec{x}(\vec{\xi})\right). \end{aligned} \tag{9.5.1}$$

The second derivative $\partial^2 x^b/\partial\xi^i\partial\xi^j$ term is what spoils the coordinate transformation rule we desire. To fix this, we introduce the concept of the covariant derivative ∇, which is built out of the partial derivative and the Christoffel symbols $\Gamma^i{}_{jk}$, which in turn is built out of the metric tensor,

$$\Gamma^i{}_{jk} = \frac{1}{2} g^{il}\left(\partial_j g_{kl} + \partial_k g_{jl} - \partial_l g_{jk}\right). \tag{9.5.2}$$

Notice the Christoffel symbol is symmetric in its lower indices: $\Gamma^i{}_{jk} = \Gamma^i{}_{kj}$. For flat space expressed in Cartesian coordinates, we also see immediately that $g_{ij} = \delta_{ij}$ implies $\Gamma^i{}_{jk} = 0$ since every term in eq. (9.5.2) involves the first derivatives $\partial_i g_{jk}$.

[87]For a scalar φ the covariant derivative is just the partial derivative

$$\nabla_i \varphi = \partial_i \varphi. \tag{9.5.3}$$

For a $\binom{0}{1}$ or $\binom{1}{0}$ tensor, its covariant derivative reads

$$\nabla_i T_j = \partial_i T_j - \Gamma^l{}_{ij} T_l, \tag{9.5.4}$$
$$\nabla_i T^j = \partial_i T^j + \Gamma^j{}_{il} T^l. \tag{9.5.5}$$

Under $\vec{x} \to \vec{x}(\vec{\xi})$, we have,

$$\nabla_{\xi^i} \varphi(\vec{\xi}) = \frac{\partial x^a}{\partial \xi^i} \nabla_{x^a} \varphi\left(\vec{x}(\vec{\xi})\right), \tag{9.5.6}$$
$$\nabla_{\xi^i} T_j(\xi) = \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \nabla_{x^a} T_b\left(\vec{x}(\vec{\xi})\right). \tag{9.5.7}$$

For a general $\binom{N}{M}$ tensor, we have

$$\begin{aligned} \nabla_k T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M} &= \partial_k T^{i_1 i_2 \dots i_N}{}_{j_1 j_2 \dots j_M} \\ &\quad + \Gamma^{i_1}{}_{kl} T^{l i_2 \dots i_N}{}_{j_1 j_2 \dots j_M} + \Gamma^{i_2}{}_{kl} T^{i_1 l \dots i_N}{}_{j_1 j_2 \dots j_M} + \dots + \Gamma^{i_N}{}_{kl} T^{i_1 \dots i_{N-1} l}{}_{j_1 j_2 \dots j_M} \\ &\quad - \Gamma^l{}_{k j_1} T^{i_1 \dots i_N}{}_{l j_2 \dots j_M} - \Gamma^l{}_{k j_2} T^{i_1 \dots i_N}{}_{j_1 l \dots j_M} - \dots - \Gamma^l{}_{k j_M} T^{i_1 \dots i_N}{}_{j_1 \dots j_{M-1} l}. \end{aligned} \tag{9.5.8}$$

[88]By using eq. (9.5.1) we may infer how the Christoffel symbols themselves must transform – they are not tensors. Firstly,

$$\begin{aligned} \nabla_{\xi^i} T_j(\vec{\xi}) &= \partial_{\xi^i} T_j(\vec{\xi}) - \Gamma^l{}_{ij}(\vec{\xi}) T_l(\vec{\xi}) \\ &= \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \partial_{x^a} T_b\left(\vec{x}(\vec{\xi})\right) + \left(\frac{\partial^2 x^b}{\partial \xi^j \partial \xi^i} - \Gamma^l{}_{ij}(\vec{\xi}) \frac{\partial x^b(\vec{\xi})}{\partial \xi^l}\right) T_b\left(\vec{x}(\vec{\xi})\right) \end{aligned} \tag{9.5.9}$$

On the other hand,

$$\begin{aligned} \nabla_{\xi^i} T_j(\vec{\xi}) &= \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \nabla_{x^a} T_b\left(\vec{x}(\vec{\xi})\right) \\ &= \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \left\{\partial_{x^a} T_b\left(\vec{x}(\vec{\xi})\right) - \Gamma^l{}_{ab}\left(\vec{x}(\vec{\xi})\right) T_l\left(\vec{x}(\vec{\xi})\right)\right\} \end{aligned} \tag{9.5.10}$$

Comparing equations (9.5.9) and (9.5.10),

$$\left(\Gamma^l{}_{ij}(\vec{\xi}) \frac{\partial x^k(\vec{\xi})}{\partial \xi^l} - \frac{\partial^2 x^k}{\partial \xi^j \partial \xi^i}\right) T_k\left(\vec{x}(\vec{\xi})\right) = \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \Gamma^k{}_{ab}\left(\vec{x}(\vec{\xi})\right) T_k\left(\vec{x}(\vec{\xi})\right). \tag{9.5.11}$$

---

[87]For now, we will take the practical approach, and focus on *how* to construct the covariant derivative given a metric g_{ij}. In §(11.3) below, we will return to understand *why* the covariant derivative takes the form it does.

[88]The semi-colon and comma are sometimes used to denote, respectively, the covariant and partial derivatives. For example, $\nabla_l \nabla_i T^{jk} \equiv T^{jk}{}_{;il}$ and $T^{ij}{}_{,k} \equiv \partial_k T^{ij}$.

Since $T_k(\vec{x})$ is arbitrary for now, this leads us to relate the Christoffel symbol written in $\vec{\xi}$ coordinates $\Gamma^l{}_{ij}(\vec{\xi})$ and that written in $\vec{x}$ coordinates $\Gamma^l{}_{ij}(\vec{x})$.

$$\Gamma^l{}_{ij}(\vec{\xi})\frac{\partial x^k(\vec{\xi})}{\partial \xi^l} = \frac{\partial^2 x^k}{\partial \xi^j \partial \xi^i} + \frac{\partial x^a}{\partial \xi^i}\frac{\partial x^b}{\partial \xi^j}\Gamma^k{}_{ab}\left(\vec{x}(\vec{\xi})\right) \tag{9.5.12}$$

As long as the coordinate transformation $\partial x^k/\partial \xi^j$ is invertible, we may contract both sides with $\partial \xi^s/\partial x^k$ to obtain

$$\Gamma^l{}_{ij}(\vec{\xi}) = \Gamma^k_{mn}\left(\vec{x}(\vec{\xi})\right)\frac{\partial \xi^l}{\partial x^k(\vec{\xi})}\frac{\partial x^m(\vec{\xi})}{\partial \xi^i}\frac{\partial x^n(\vec{\xi})}{\partial \xi^j} + \frac{\partial \xi^l}{\partial x^k(\vec{\xi})}\frac{\partial^2 x^k(\vec{\xi})}{\partial \xi^j \partial \xi^i}. \tag{9.5.13}$$

On the right hand side, all $\vec{x}$ have been replaced with $\vec{x}(\vec{\xi})$.[89]

To sum: the Christoffel symbol *does not* transform like a tensor. The second derivative terms on the right hand side of eq. (9.5.13) which spoil the tensor transformation rules, in fact cancel out the second derivative terms from the partial derivative acting on the Jacobian in eq. (9.5.1). For the covariant derivative on a multi-rank tensor in eq. (9.5.8), there needs to be a Christoffel symbol contracted with each index because – upon coordinate transformation – it helps cancel the partial derivative acting on the Jacobian contracted with the same index.

Product Rule The covariant derivative, like its partial derivative counterpart, obeys the product rule. Suppressing the indices, if T_1 and T_2 are both tensors, we have

$$\nabla\left(T_1 T_2\right) = (\nabla T_1)T_2 + T_1(\nabla T_2). \tag{9.5.14}$$

Unlike partial derivatives, repeated covariant derivatives do not commute; hence, make sure you keep track of the order of operations. For instance,

$$\nabla_a \nabla_b T^{ij} \neq \nabla_b \nabla_a T^{ij}. \tag{9.5.15}$$

Problem 9.27. Commutator of Covariant Derivatives on Scalar Show that double covariant derivatives on a scalar field *do* commute: $\nabla_i\nabla_j\varphi = \nabla_j\nabla_i\varphi$. If we define

$$[\nabla_i, \nabla_j] \equiv \nabla_i\nabla_j - \nabla_j\nabla_i, \tag{9.5.16}$$

this may be expressed as $[\nabla_i, \nabla_j]\varphi = 0$. □

Metric compatibility As you will see below, the metric is parallel transported in all directions,

$$\nabla_i g_{jk} = \nabla_i g^{jk} = 0. \tag{9.5.17}$$

Combined with the product rule in eq. (9.5.14), this means when raising and lowering of indices of a covariant derivative of a tensor, the metric may be passed in and out of the ∇. For example,

$$\nabla_j(g_{ia}T^{kal}) = \nabla_j g_{ia}\cdot T^{kal} + g_{ia}\nabla_j T^{kal} = g_{ia}\nabla_j T^{kal}$$

[89]We note in passing that in gauge theory – which encompasses humanity's current description of the non-gravitational forces (electromagnetic-weak $(\mathrm{SU}_2)_{\text{left-handed fermions}} \times (U_1)_{\text{hypercharge}}$ and strong nuclear $(\mathrm{SU}_3)_{\text{color}}$) – the fundamental fields there $\{A^b{}_\mu\}$ transforms (in a group theory sense) in a very similar fashion as the Christoffel symbols do (under a coordinate transformation) in eq. (9.5.13).

$$= \nabla_j T^{k\ l}_{\ i} . \tag{9.5.18}$$

Christoffel symbols as 'rotation coefficients' I have introduced the Christoffel symbol here by showing how it allows us to define a derivative operator on a tensor that returns a tensor. I should mention here that, alternatively, it is also possible to view $\Gamma^i{}_{jk}$ as 'rotation matrices,' describing the failure of parallel transporting the basis bras $\{\langle \mathrm{d}x^i|\}$ and kets $\{|\partial_i\rangle\}$ as they are moved from one point in space to a neighboring point infinitesimally far away. Specifically,

$$\nabla_i \left\langle \mathrm{d}x^j \right| = -\Gamma^j{}_{ik} \left\langle \mathrm{d}x^k \right| \qquad \text{and} \qquad \nabla_i \left|\partial_j\right\rangle = \Gamma^l{}_{ij} \left|\partial_l\right\rangle . \tag{9.5.19}$$

By projecting $\left\langle \mathrm{d}x^k \right|$ on the right equality and recalling eq. (9.2.14),

$$\Gamma^k{}_{ij} = \left\langle \mathrm{d}x^k \left|\nabla_i\right| \partial_j \right\rangle . \tag{9.5.20}$$

Within this perspective, the tensor components are scalars. The product rule then yields, for instance,

$$\begin{aligned} \nabla_i \left(V_a \left\langle \mathrm{d}x^a \right|\right) &= \left(\nabla_i V_a\right) \left\langle \mathrm{d}x^a \right| + V_a \nabla_i \left\langle \mathrm{d}x^a \right| \\ &= \left(\partial_i V_j - V_a \Gamma^a{}_{ij}\right) \left\langle \mathrm{d}x^j \right| . \end{aligned} \tag{9.5.21}$$

$$\begin{aligned} \nabla_i \left(V^a \left|\partial_a\right\rangle\right) &= \left(\nabla_i V^a\right) \left|\partial_a\right\rangle + V^a \nabla_i \left|\partial_a\right\rangle \\ &= \left(\partial_i V^a + \Gamma^a{}_{ij} V^j\right) \left|\partial_a\right\rangle . \end{aligned} \tag{9.5.22}$$

Parallel transport Now that we have introduced the covariant derivative, we may finally define what (invariance under) parallel transport actually is.

Let v^i be a (tangent) vector field and $T^{j_1 \dots j_N}$ be some tensor. (Here, the placement of indices on the T is not important, but we will assume for convenience, all of them are upper indices.) We say that the tensor T is invariant under parallel transport along the vector $\vec{v}$ when

$$v^i \nabla_i T^{j_1 \dots j_N} = 0. \tag{9.5.23}$$

Problem 9.28. Christoffel symbols, Parallel Transport on 2−sphere Employ eq. (9.5.2) to calculate the Christoffel symbols of the metric on the 2-sphere with unit radius,

$$(\mathrm{d}\ell)^2 = \mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2. \tag{9.5.24}$$

That is, the non-zero components of the metric are $g_{\theta\theta} = g_{11} = 1 = g^{11} = g^{\theta\theta}$ and $g_{\phi\phi} = g_{22} = (\sin\theta)^2 = 1/g^{22} = 1/g^{\theta\theta}$. For example,

$$\Gamma^\theta{}_{\phi\phi} = \frac{1}{2} g^{\theta i} \left(\partial_\phi g_{\phi i} + \partial_\phi g_{\phi i} - \partial_i g_{\phi\phi}\right) . \tag{9.5.25}$$

Due to the diagonal nature of the 2−sphere metric, $g^{\theta i} = \delta^i_\theta$,

$$\Gamma^\theta{}_{\phi\phi} = \frac{1}{2} \left(2\partial_\phi g_{\phi\theta} - \partial_\theta (\sin\theta)^2\right) = -\sin\theta\cos\theta. \tag{9.5.26}$$

You should find the other non-trivial components to be

$$\Gamma^\phi{}_{\phi\theta} = \Gamma^\phi{}_{\theta\phi} = \cot\theta. \tag{9.5.27}$$

In the coordinate system (θ, ϕ), define the vector $v^i = (v^\theta, v^\phi) = (1, 0)$, i.e., $v = \partial_\theta$. This is the vector tangent to the sphere, at a given location $(0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$ on the sphere, such that it points away from the North and towards the South pole, along a constant longitude line. Show that it is parallel transported along itself, as quantified by the statement

$$v^i \nabla_i v^j = \nabla_\theta v^j = 0. \tag{9.5.28}$$

Next, show that $\vec{v}$ is parallel transported along the equator; i.e., explain why it is equivalent to

$$(\partial_\phi)^i \nabla_i v^j = 0, \tag{9.5.29}$$

and verify it via a direct calculation. □

Riemann and Ricci tensors I will not use them very much in the rest of our discussion in this Chapter, but I should still highlight that the Riemann and Ricci tensors are fundamental to describing curvature. The Riemann tensor is built out of the Christoffel symbols via

$$R^i{}_{jkl} = \partial_k \Gamma^i{}_{lj} - \partial_l \Gamma^i{}_{kj} + \Gamma^i{}_{sk}\Gamma^s{}_{lj} - \Gamma^i{}_{sl}\Gamma^s{}_{kj}. \tag{9.5.30}$$

The failure of parallel transport of some vector V^i around an infinitesimally small coordinate loop, is characterized by

$$[\nabla_k, \nabla_l]V^i \equiv (\nabla_k \nabla_l - \nabla_l \nabla_k)V^i = R^i{}_{jkl}V^j, \tag{9.5.31}$$

$$[\nabla_k, \nabla_l]V_j \equiv (\nabla_k \nabla_l - \nabla_l \nabla_k)V_j = -R^i{}_{jkl}V_i. \tag{9.5.32}$$

The generalization to higher rank tensors is

$$\begin{aligned}
&[\nabla_i, \nabla_j]T^{k_1 \dots k_N}{}_{l_1 \dots l_M} \\
&= R^{k_1}{}_{aij} T^{a k_2 \dots k_N}{}_{l_1 \dots l_M} + R^{k_2}{}_{aij} T^{k_1 a k_3 \dots k_N}{}_{l_1 \dots l_M} + \dots + R^{k_N}{}_{aij} T^{k_1 \dots k_{N-1} a}{}_{l_1 \dots l_M} \\
&\qquad - R^a{}_{l_1 ij} T^{k_1 \dots k_N}{}_{a l_2 \dots l_M} - R^a{}_{l_2 ij} T^{k_1 \dots k_N}{}_{l_1 a l_3 \dots l_M} - \dots - R^a{}_{l_M ij} T^{k_1 \dots k_N}{}_{l_1 \dots l_{M-1} a}.
\end{aligned} \tag{9.5.33}$$

This illustrates the point alluded to earlier – covariant derivatives commute iff space is flat; i.e., iff the Riemann tensor is zero.

The Riemann tensor obeys the following symmetries.

$$R_{ijab} = R_{abij}, \qquad R_{ijab} = -R_{jiab}, \qquad R_{abij} = -R_{abji}, \qquad R_{i[jmn]} = 0. \tag{9.5.34}$$

The Riemann tensor also obeys the Bianchi identities[90]

$$R^i{}_{[jkl]} = \nabla_{[i} R^{jk}{}_{lm]} = 0. \tag{9.5.35}$$

In D dimensions, the Riemann tensor has $D^2(D^2-1)/12$ algebraically independent components. In particular, in $D = 1$ dimension, space is always flat because $R_{1111} = -R_{1111} = 0$.

[90] The symbol $[\dots]$ means the indices within it are fully anti-symmetrized; in particular, $T_{[ijk]} = T_{ijk} - T_{ikj} - T_{jik} + T_{jki} - T_{kji} + T_{kij}$. We will have more to say about this operation later on.

The Ricci tensor is defined as the non-trivial contraction of a pair of the Riemann tensor's indices.

$$R_{jl} \equiv R^i{}_{jil}. \tag{9.5.36}$$

It is symmetric

$$R_{ij} = R_{ji}. \tag{9.5.37}$$

Finally the Ricci scalar results from a contraction of the Ricci tensor's indices.

$$\mathcal{R} \equiv g^{jl} R_{jl}. \tag{9.5.38}$$

Contracting eq. (9.5.35) appropriately yields the Bianchi identities involving the Ricci tensor and scalar

$$\nabla^i \left(R_{ij} - \frac{g_{ij}}{2} \mathcal{R} \right) = 0. \tag{9.5.39}$$

This is a good place to pause and state, the Christoffel symbols in eq. (9.5.2), covariant derivatives, and the Riemann/Ricci tensors, etc., are in general very tedious to compute. If you ever have to do so on a regular basis, say for research, I highly recommend familiarizing yourself with one of the various software packages available that could do them for you.

Geodesics Recall the distance integral in eq. (9.1.28). If you wish to determine the shortest path (aka geodesic) between some given pair of points $\vec{x}_1$ and $\vec{x}_2$, you will need to minimize eq. (9.1.28). This is a 'calculus of variation' problem. The argument runs as follows. Suppose you found the path $\vec{z}(\lambda)$ that yields the shortest ℓ. Then, if you consider a slight variation $\delta\vec{z}$ of the path, namely consider

$$\vec{x}(\lambda) = \vec{z}(\lambda) + \delta\vec{z}(\lambda), \tag{9.5.40}$$

we must find the contribution to ℓ at first order in $\delta\vec{z}$ to be zero. This is analogous to the vanishing of the first derivatives of a function at its minimum.[91] In other words, in the integrand of eq. (9.1.28) we must replace

$$g_{ij}\left(\vec{x}(\lambda)\right) \to g_{ij}\left(\vec{z}(\lambda) + \delta\vec{z}(\lambda)\right) = g_{ij}\left(\vec{z}(\lambda)\right) + \delta z^k(\lambda) \frac{\partial g_{ij}\left(\vec{z}(\lambda)\right)}{\partial z^k} + \mathcal{O}(\delta z^2) \tag{9.5.41}$$

$$\frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda} \to \frac{\mathrm{d}z^i(\lambda)}{\mathrm{d}\lambda} + \frac{\mathrm{d}\delta z^i(\lambda)}{\mathrm{d}\lambda}. \tag{9.5.42}$$

Since $\delta\vec{z}$ was arbitrary, at first order, its coefficient within the integrand must vanish. If we further specialize to affine parameters λ – i.e., such that

$$\sqrt{g_{ij}(\mathrm{d}z^i/\mathrm{d}\lambda)(\mathrm{d}z^j/\mathrm{d}\lambda)} = \text{constant along the entire path } \vec{z}(\lambda) \tag{9.5.43}$$

[91]There is some smoothness condition being assumed here. For instance, the tip of the pyramid (or a cone) is the maximum height achieved, but the derivative slightly away from the tip is negative in all directions.

– then one would arrive at the following second order non-linear ODE. Minimizing the distance ℓ between $\vec{x}_1$ and $\vec{x}_2$ leads to the shortest path $\vec{z}(\lambda)$ ($\equiv$ geodesic) obeying:

$$0 = \frac{\mathrm{d}^2 z^i}{\mathrm{d}\lambda^2} + \Gamma^i{}_{jk}\left(g_{ab}(\vec{z})\right)\frac{\mathrm{d}z^j}{\mathrm{d}\lambda}\frac{\mathrm{d}z^k}{\mathrm{d}\lambda}, \tag{9.5.44}$$

with the boundary conditions

$$\vec{z}(\lambda_1) = \vec{x}_1, \qquad \vec{z}(\lambda_2) = \vec{x}_2. \tag{9.5.45}$$

You will verify this discussion in Problem (9.29) below.

The converse is also true, in that – if the geodesic equation in eq. (9.5.44) holds, then g_{ij} $(\mathrm{d}z^i/\mathrm{d}\lambda)(\mathrm{d}z^j/\mathrm{d}\lambda)$ is a constant along the entire geodesic; and, hence, λ is affine. Denoting $\ddot{z}^i \equiv \mathrm{d}^2 z^i/\mathrm{d}\lambda^2$ and $\dot{z}^i \equiv \mathrm{d}z^i/\mathrm{d}\lambda$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(g_{ij}\dot{z}^i\dot{z}^j\right) &= 2\ddot{z}^i\dot{z}^j g_{ij} + \dot{z}^k\partial_k g_{ij}\dot{z}^i\dot{z}^j \\ &= 2\ddot{z}^i\dot{z}^j g_{ij} + \dot{z}^k\dot{z}^i\dot{z}^j\left(\partial_k g_{ij} + \partial_i g_{kj} - \partial_j g_{ik}\right)\end{aligned} \tag{9.5.46}$$

Note that the last two terms inside the parenthesis of the second equality cancel. The reason for inserting them is because the expression contained within the parenthesis is related to the Christoffel symbol; keeping in mind eq. (9.5.2),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(g_{ij}\dot{z}^i\dot{z}^j\right) &= 2\dot{z}^i\left\{\ddot{z}^j g_{ij} + \dot{z}^k\dot{z}^j g_{il}\frac{g^{lm}}{2}\left(\partial_k g_{jm} + \partial_j g_{km} - \partial_m g_{jk}\right)\right\} \\ &= 2g_{il}\dot{z}^i\left\{\ddot{z}^l + \dot{z}^k\dot{z}^j\Gamma^l{}_{kj}\right\} = 0.\end{aligned} \tag{9.5.47}$$

The last equality follows because the expression in the $\{\dots\}$ of eq. (9.5.47) is the right hand side of eq. (9.5.44). This constancy of g_{ij} $(\mathrm{d}z^i/\mathrm{d}\lambda)(\mathrm{d}z^j/\mathrm{d}\lambda)$ is useful for solving the geodesic equation itself.

In §(11) below, we will also see why eq. (9.5.44) is equivalent to the statement that some unit length vector field $v^i(\vec{x}) = \mathrm{d}x^i/\mathrm{d}\lambda$, obeying $v^i v^j g_{ij} = 1$, is parallel transported along itself:

$$v^i\nabla_i v^j = 0. \tag{9.5.48}$$

Problem 9.29. Geodesics: Action Principle and Noether's Theorem Show that the affine parameter form of the geodesic (9.5.44) follows from demanding the distance-squared integral of eq. (9.1.33) be extremized:

$$\ell^2 = (\lambda_2 - \lambda_1)\int_{\lambda_1}^{\lambda_2}\left(g_{ij}\left(\vec{z}(\lambda)\right)\frac{\mathrm{d}z^i}{\mathrm{d}\lambda}\frac{\mathrm{d}z^j}{\mathrm{d}\lambda}\right)\mathrm{d}\lambda. \tag{9.5.49}$$

[92]That is, show that eq. (9.5.44) follows from applying the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_g}{\partial\dot{z}^i} - \frac{\partial L_g}{\partial z^i} = 0 \tag{9.5.50}$$

[92]Some jargon: In the General Relativity literature, $\ell^2/2$ (half of eq. (9.5.49)) is known as Synge's world function.

to the Lagrangian

$$L_g \equiv \frac{1}{2} g_{ij} \dot{z}^i \dot{z}^j, \qquad\qquad \dot{z}^i \equiv \frac{\mathrm{d}z^i}{\mathrm{d}\lambda}. \tag{9.5.51}$$

In fact, you should find that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\partial L_g}{\partial \dot{z}^i} - \frac{\partial L_g}{\partial z^i} = g_{ib} \frac{D^2 z^b}{\mathrm{d}\lambda^2}, \tag{9.5.52}$$

$$\frac{D^2 z^j}{\mathrm{d}\lambda^2} \equiv \ddot{z}^j + \Gamma^j{}_{ab} \dot{z}^a \dot{z}^b. \tag{9.5.53}$$

Equivalently, you may also directly perturb the Lagrangian in eq. (9.5.51) by replacing $\vec{z} \to \vec{z} + \delta \vec{z}$, to find that its first-order-in-δz piece is

$$\delta_1 L_g = \frac{\mathrm{d}}{\mathrm{d}\lambda} \left(\delta z^i g_{ij} \dot{z}^j \right) - \delta z^i g_{ij} \frac{D^2 z^j}{\mathrm{d}\lambda^2}. \tag{9.5.54}$$

If the geodesic equation (9.5.44) is satisfied by $\vec{z}(\lambda)$, argue that the integral in eq. (9.5.49) yields the square of the geodesic distance between $\vec{x}_1 \equiv \vec{z}(\lambda_1)$ and $\vec{x}_2 \equiv \vec{z}(\lambda_2)$. (Hints: Remember eq. (9.1.28) and the constancy of L.) Also show that eq. (9.5.49) takes the same form under the re-scaling

$$\lambda = a \cdot \lambda' + b \tag{9.5.55}$$

for constants a and b; namely,

$$\ell^2 = (\lambda'_2 - \lambda'_1) \int_{\lambda'_1}^{\lambda'_2} \left(g_{ij} \left(\vec{z}(\lambda') \right) \frac{\mathrm{d}z^i}{\mathrm{d}\lambda'} \frac{\mathrm{d}z^j}{\mathrm{d}\lambda'} \right) \mathrm{d}\lambda'. \tag{9.5.56}$$

Conserved quantities from symmetries Finally, suppose ∂_k is a Killing vector. Explain why

$$\frac{\partial L_g}{\partial \dot{z}^k} \text{ is constant along the geodesic } \vec{z}(\lambda). \tag{9.5.57}$$

This is an example of Noether's theorem. For example, in flat Euclidean space, since the metric in Cartesian coordinates is a constant δ_{ij}, all the $\{\partial_i | i = 1, 2, \dots, D\}$ are Killing vectors. Therefore, from $L = (1/2)\delta_{ij} \dot{z}^i \dot{z}^j$, and we have

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\mathrm{d}z^i}{\mathrm{d}\lambda} = 0 \qquad \Rightarrow \qquad \frac{\mathrm{d}z^i}{\mathrm{d}\lambda} = \text{constant}. \tag{9.5.58}$$

This is, in fact, the statement that the center of mass of an isolated system obeying Newtonian mechanics moves with a constant velocity – total momentum is conserved. By re-writing the Euclidean metric in spherical coordinates, provide the proper definition of angular momentum (about the $D-$axis) and proceed to prove that it is conserved.

More generally, suppose ξ^i is a Killing vector. Can you show that

$$\xi^k \left(\vec{z} \right) \frac{\partial L_g}{\partial \dot{z}^k} \tag{9.5.59}$$

is constant along a geodesic? Hint: Recall the coordinate transformation we performed earlier, $\vec{x} \to \vec{x} + \vec{\xi}$. How is that related to the first order variation of L_g? Answer:

$$\delta_1 L_g = \frac{1}{2} \dot{z}^a \dot{z}^b (\pounds_\xi g)_{ab}. \tag{9.5.60}$$

When evaluated on a given path $\vec{z}(\lambda)$, also explain why $\vec{x} \to \vec{x} + \vec{\xi}$ amounts to $\dot{z}^i \to \dot{z}^i + \partial_a \xi^i \dot{z}^a$ and the identification $\delta z^i \equiv \xi^i$. Then compare eq. (9.5.54) to (9.5.60) $\square$

How to Solve For Geodesics The algorithm for solving geodesic paths in a curved space can thus be summarized as follows.

- Set the Lagrangian L_g in eq. (9.5.51) to a positive constant. From the freedom to re-scale λ in eq. (9.5.55), we see that L_g may be set to any positive constant because

$$L_g[\lambda = a \cdot \lambda' + b] = L_g[\lambda']/a^2. \tag{9.5.61}$$

  One scheme is to set $L_g = 1/2$ and $\lambda_1 = 0$; then eq. (9.5.49) tells us $\ell = \lambda_2$. Another scheme is to set $L_g = L_0^2/2$, $\lambda_1 = 0$, and $\lambda_2 = 1$. Then the Lagrangian evaluated on the geodesic solution is the geodesic length itself; i.e., eq. (9.5.49) becomes $\ell = L_0$.

- Make sure to *first* exploit all possible conserved quantities arising from the symmetries present in the geometry. These will provide you with a set of first order ordinary differential equations instead of the second order ones in eq. (9.5.44).

- Only after exploiting the constancy of L_g as well as the constant Noether charges of the geometry's isometries do you then turn to solving the remaining geodesic equations (9.5.44) themselves.

Problem 9.30. Non-Affinely Parametrized Geodesics It is of course not necessary to solve geodesics in terms of their affine parameters. We may directly minimize the length integral in eq. (9.1.28). This implies the Euler-Lagrange equations are to be applied to the Lagrangian

$$L'_g \equiv \sqrt{g_{ij} \dot{z}^i \dot{z}^j} \equiv \sqrt{2L_g}, \tag{9.5.62}$$

where the λ in $\dot{z} \equiv \mathrm{d}z^i/\mathrm{d}\lambda$ no longer needs to be affinely parametrized. This form of the geodesic equation is useful if, for instance, you wish to solve the solution in terms of one of the coordinates (say, z^k, for some k).

Show that the explicit form of $(\mathrm{d}/\mathrm{d}\lambda)(\partial L'_g/\partial \dot{z}^i) = \partial L'_g/\partial z^i$, the non-affinely parametrized geoedesic equation, reads as follows:

$$\ddot{z}^i + \Gamma^i{}_{ab} \dot{z}^a \dot{z}^b = \dot{z}^i \frac{\mathrm{d}}{\mathrm{d}\lambda} \ln \sqrt{g_{ab} \dot{z}^a \dot{z}^b}. \tag{9.5.63}$$

Hint: You may find it useful to first work out the relevant Euler-Lagrange equations in terms of L_g; then employ eq. (9.5.52) and the definition for L_g at the very end. $\square$

Geodesics in Flat Space Let us start with the example of geodesics in flat space, using Cartesian coordinates $\{\vec{x}\}$. The affinely parametrized geodesic Lagrangian is

$$L_g = \frac{1}{2}\delta_{ij}\dot{z}^i\dot{z}^j, \tag{9.5.64}$$

where each overdot denotes a derivative with respect to the affine parameter λ. The geodesic equation is

$$\ddot{z} = 0. \tag{9.5.65}$$

This is just the constant acceleration problem in classical mechanics. The solution, joining $\vec{x}_1$ to $\vec{x}_2$ is the straight line

$$\vec{z}(0 \leq \lambda \leq 1) = \vec{x}_1 + \lambda(\vec{x}_2 - \vec{x}_1). \tag{9.5.66}$$

The square of the geodesic length, i.e., eq. (9.5.49), is therefore

$$\ell(\vec{x}_1 \leftrightarrow \vec{x}_2)^2 = (\vec{x}_1 - \vec{x}_2)^2. \tag{9.5.67}$$

Translation Symmetries From the symmetries point-of-view, the metric δ_{ij} is independent of all the Cartesian coordinates. Hence, we may identify the conserved momentum $\{p_i\}$ as

$$\frac{\partial L_g}{\partial \dot{z}^i} = \dot{z}^i = p_i. \tag{9.5.68}$$

This immediately leads to $\vec{z}(\lambda) = \vec{p}\cdot\lambda + \vec{c}$ for constant $\{c^i\}$. The Lagrangian L_g is then

$$L_g = \frac{p^2}{2}, \qquad p^2 \equiv \delta^{ij}p_ip_j. \tag{9.5.69}$$

If we set $L_g = 1/2$ so that $p = 1$; by choosing $\lambda_1 = 0$, we have $\vec{z}(0) = \vec{c} = \vec{x}_1$ and $\vec{z}(\lambda_2) = \lambda_2 \cdot \vec{p} + \vec{x}_1 = \vec{x}_2$. Since $\vec{p}$ is now of unit length, we see that

$$\lambda_2 \cdot \vec{p} = \vec{x}_2 - \vec{x}_1 \tag{9.5.70}$$

implies

$$\lambda_2 = |\vec{x}_2 - \vec{x}_1|. \tag{9.5.71}$$

On the other hand, if we choose $L_g = L_0^2/2$, $\lambda_1 = 0$ and $\lambda_2 = 1$; then $\vec{z}(0) = \vec{c} = \vec{x}_1$ whereas $\vec{z}(1) = \vec{p} + \vec{x}_1 = \vec{x}_2$. Hence, $\vec{p} = \vec{x}_2 - \vec{x}_1$ and

$$L_0 = p = |\vec{x}_2 - \vec{x}_1|. \tag{9.5.72}$$

Geodesics on a 2−sphere The length of a geodesic on the 2−sphere, joining (θ_1, ϕ_1) and (θ_2, ϕ_2), is

$$\ell = \int_{\theta_1,\phi_1}^{\theta_2,\phi_2} \sqrt{\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2} \tag{9.5.73}$$

$$= \int_{\theta_1}^{\theta_2} \sqrt{1+\sin^2(\theta)\phi'(\theta)^2}\mathrm{d}\theta \equiv \int_{\theta_1}^{\theta_2} L_g \mathrm{d}\theta. \tag{9.5.74}$$

Instead of using an affinely parametrized geodesic, we may directly minimize this definition of the length. Since the Lagrangian $L_g(\theta, \phi'(\theta))$ is independent of ϕ, we must have

$$\frac{\partial L_g}{\partial \phi'} = \frac{\sin^2(\theta)\cdot\phi'(\theta)}{\sqrt{1+\sin^2(\theta)\phi'(\theta)^2}} = \chi \qquad (\equiv \text{const.}). \tag{9.5.75}$$

This may be solved for

$$\phi'(\theta) = \frac{\chi}{\sin(\theta)\sqrt{\sin^2(\theta)-\chi^2}}, \tag{9.5.76}$$

For the square root to remain real, we see that $|\chi| \leq \sin^2(\theta)$. Integrating both sides with respect to θ,

$$\phi(\theta) = \arccos\left(\frac{\chi}{\sqrt{1-\chi^2}}\cot(\theta)\right) + \phi_0, \tag{9.5.77}$$

where ϕ_0 is an integration constant. Since $-1 \leq \chi \leq +1$, we may identify $\chi = -\cos\theta_0$ for some arbitrary angle θ_0, and deduce

$$-\cot(\theta_0)\cot(\theta) = \cos(\phi-\phi_0), \tag{9.5.78}$$
$$\sin(\theta)\sin(\theta_0)\cos(\phi-\phi_0) + \cos(\theta_0)\cos(\theta) = 0. \tag{9.5.79}$$

Applying $\cos(\phi-\phi_0) = \cos(\phi)\cos(\phi_0) + \sin(\phi)\sin(\phi_0)$ to eq. (9.5.79), we obtain the equation of a plane intersecting the 2−sphere that also passes through the origin:

$$\widehat{r}(\theta,\phi)\cdot\widehat{B} = 0, \tag{9.5.80}$$
$$\widehat{r} = \left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)\right)^{\mathrm{T}}, \tag{9.5.81}$$
$$\widehat{B} = \left(\sin(\theta_0)\cos(\phi_0), \sin(\theta_0)\sin(\phi_0), \cos(\theta_0)\right)^{\mathrm{T}}. \tag{9.5.82}$$

If the geodesic joins (θ_1, ϕ_1) and (θ_2, ϕ_2), therefore, we may simply construct the normal vector by the cross product

$$\widehat{B} = \pm\frac{\widehat{r}(\theta_1,\phi_1)\times\widehat{r}(\theta_2,\phi_2)}{|\widehat{r}(\theta_1,\phi_1)\times\widehat{r}(\theta_2,\phi_2)|}. \tag{9.5.83}$$

Problem 9.31. Geodesics on 2-Sphere: Affine Case The affinely parametrized form of the geodesic equation on the $(D \geq 2)$−sphere is tackled in §(9.8) below. It turns out to be useful to view the unit D−sphere as the $\vec{x}\cdot\vec{x} = 1$ surface in $D+1$ flat space, where the $\vec{x}$ are Cartesian coordinates. Then, as derived in eq. (9.8.22), the geodesic $X^i(\lambda)$ itself obeys

$$\ddot{X}^i(\lambda) + X^i(\lambda) = 0. \tag{9.5.84}$$

In this problem, we shall discover an explicit representation of the geodesic joining some location $\widehat{r}_1$ on the 2−sphere to some other location $\widehat{r}_2$, where $\widehat{r}_{1,2}$ are unit vectors in 3D flat space. The strategy is to use the rotation matrix in eq. (5.6.20) to rotate $\widehat{r}_1$ to $\widehat{r}_2$, by first defining

$$\widehat{n} \equiv \frac{\widehat{r}_1 \times \widehat{r}_2}{\sqrt{1-(\widehat{r}_1 \cdot \widehat{r}_2)^2}}. \tag{9.5.85}$$

Show that the rotation matrix in eq. (5.6.20) applied to $\widehat{r}_1$ is

$$X^i(\theta) \equiv \widehat{R}^{ij}_{\widehat{n}}(\theta)\widehat{r}^j_1 \tag{9.5.86}$$

$$= \cos\theta \cdot \widehat{r}^i_1 - \sin\theta \frac{\widehat{r}^i_1(\widehat{r}_1 \cdot \widehat{r}_2) - \widehat{r}^i_2}{\sqrt{1-(\widehat{r}_1 \cdot \widehat{r}_2)^2}}. \tag{9.5.87}$$

Verify the boundary conditions

$$X^i(\theta = 0) = \widehat{r}^i_1 \qquad \text{and} \qquad X^i(\theta = \arccos \widehat{r}_1 \cdot \widehat{r}_2) = \widehat{r}^i_2; \tag{9.5.88}$$

as well as the geodesic equation $\ddot{X} + X = 0$. Explain why these results tell us the world function σ on the 2−sphere, half the square of the geodesic distance between $\widehat{r}_1$ and $\widehat{r}_2$, is given by

$$\sigma(\widehat{r}_1, \widehat{r}_2) \equiv \frac{1}{2}\ell(\widehat{r}_1 \leftrightarrow \widehat{r}_2)^2 = \frac{1}{2}\left(\arccos \widehat{r}_1 \cdot \widehat{r}_2\right)^2. \tag{9.5.89}$$

Of course, this geodesic distance $\ell(\widehat{r}_1 \leftrightarrow \widehat{r}_2) = \arccos \widehat{r}_1 \cdot \widehat{r}_2$ is the minimum if $\arccos \widehat{r}_1 \cdot \widehat{r}_2$ is smaller than π; namely, if the smaller circle is taken. If the greater circle is taken instead, the geodesic distance would be $2\pi - \arccos \widehat{r}_1 \cdot \widehat{r}_2$. If we encircle the 2−sphere on the plane containing the origin and $\widehat{r}_{1,2}$ multiple times while going between $\widehat{r}_1$ and $\widehat{r}_2$, we would need to add to these two solutions the appropriate multiple of 2π. □

Christoffel symbols from Lagrangian Instead of computing the Christoffel symbols using the formula in eq. (9.5.2), we may instead use the variational principle encoded eq. (9.5.49) to obtain its components. That is, starting from the Lagrangian in eq. (9.5.51), one may compute the geodesic equation (9.5.44) and read off ${\Gamma^i}_{ab}$ as the coefficient of $\dot{z}^a \dot{z}^b$ for $a = b$; and *half* of the coefficient of $\dot{z}^a \dot{z}^b$ for $a \neq b$.
Example I As a first example, let us extract the Christoffel symbols of the 2D flat metric in polar coordinates

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\phi^2. \tag{9.5.90}$$

The Lagrangian in eq. (9.5.51) is

$$L_g = \frac{1}{2}\dot{r}^2 + \frac{1}{2}r^2\dot{\phi}^2. \tag{9.5.91}$$

The Euler-Lagrange equations are

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_g}{\partial \dot{r}} = \frac{\partial L_g}{\partial r} \tag{9.5.92}$$

$$\ddot{r} = r\dot{\phi}^2 \tag{9.5.93}$$

$$\ddot{r} - r\dot{\phi}^2 = \ddot{r} + \Gamma^r{}_{\phi\phi}\dot{\phi}^2 = 0; \tag{9.5.94}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_g}{\partial \dot{\phi}} = \frac{\partial L_g}{\partial \phi} \tag{9.5.95}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(r^2\dot{\phi}) = 0 \tag{9.5.96}$$

$$\ddot{\phi} + \frac{2}{r}\dot{r}\dot{\phi} = \ddot{\phi} + \Gamma^\phi{}_{r\phi}\dot{r}\dot{\phi} + \Gamma^\phi{}_{\phi r}\dot{\phi}\dot{r} = 0. \tag{9.5.97}$$

We see that

$$\Gamma^r{}_{\phi\phi} = -r; \tag{9.5.98}$$

whereas, due to its symmetric character,

$$\Gamma^\phi{}_{r\phi} = \Gamma^\phi{}_{\phi r} = 1/r. \tag{9.5.99}$$

This latter point is worth reiterating: for $a \neq b$, the coefficient of $\dot{z}^a\dot{z}^b$ in the geodesic equation $\ddot{z}^i + (\dots)\dot{z}^a\dot{z}^b + \dots = 0$ is *twice* of $\Gamma^i{}_{ab}$, because – with no sum over a and b –

$$\Gamma^i{}_{ab}\dot{z}^a\dot{z}^b + \Gamma^i{}_{ba}\dot{z}^b\dot{z}^a = 2\Gamma^i{}_{ab}\dot{z}^a\dot{z}^b. \tag{9.5.100}$$

The rest of the Christoffel symbols of the 2D polar coordinates flat metric are zero because they do not appear in the geodesic equation; for e.g., $\Gamma^r{}_{rr} = 0$.
Example II Next, let us consider the following $D-$dimensional metric:

$$\mathrm{d}\ell^2 \equiv a(\vec{x})^2\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}, \tag{9.5.101}$$

where $a(\vec{x})$ is an arbitrary function. The Lagrangian in eq. (9.5.51) is now

$$L = \frac{1}{2}a^2\delta_{ij}\dot{z}^i\dot{z}^j, \qquad \dot{z}^i \equiv \frac{\mathrm{d}z^i}{\mathrm{d}\lambda}. \tag{9.5.102}$$

Applying the Euler-Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{z}^i} - \frac{\partial L}{\partial z^i} = 0 \tag{9.5.103}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(a^2\dot{z}^i\right) - a\partial_i a\dot{\vec{z}}^2 = 0 \tag{9.5.104}$$

$$2a\dot{z}^j\partial_j a\ \dot{z}^i + a^2\ddot{z}^i - a\partial_i a\dot{\vec{z}}^2 = 0 \tag{9.5.105}$$

$$\ddot{z}^i + \left(\frac{\partial_j a}{a}\delta^i_l + \frac{\partial_l a}{a}\delta^i_j - \frac{\partial_i a}{a}\delta_{lj}\right)\dot{z}^l\dot{z}^j = \ddot{z}^i + \Gamma^i{}_{lj}\dot{z}^l\dot{z}^j = 0. \tag{9.5.106}$$

Using $\{\dots\}$ to indicate symmetrization of the indices, we have derived

$$\begin{aligned}\Gamma^i{}_{lj} &= \frac{1}{a}\left(\partial_{\{j}a\delta^i_{l\}} - \partial_i a\delta_{lj}\right)\\ &= \left(\delta^k_{\{j}\delta^i_{l\}} - \delta^{ki}\delta_{lj}\right)\partial_k \ln a.\end{aligned} \tag{9.5.107}$$

Problem 9.32. Geodesics: Hamiltonian Formulation An alternate but equivalent manner to solve the geodesics in a given geometry, is through the Hamiltonian formulation. Define the conjugate momentum p_i to the coordinate z^i as

$$p_i \equiv \frac{\partial L_g}{\partial \dot{z}^i} = g_{ij}\dot{z}^j, \tag{9.5.108}$$

where L is the Lagrangian in eq. (9.5.51); and further define the Hamiltonian H through the Legendre transform

$$H(\vec{z}, \vec{p}) \equiv p_i \dot{z}^i(\vec{z}, \vec{p}) - L(\vec{z}, \vec{p}). \tag{9.5.109}$$

This relation between H and L assumes all the $\{\dot{z}^i \equiv \mathrm{d}z^i/\mathrm{d}\lambda\}$ has been re-expressed in terms of $\vec{z}$ and $\vec{p}$. Now demonstrate that the Hamiltonian H is equal to the Lagrangian L; in particular, you should find that

$$H = \frac{1}{2} g^{ij} p_i p_j. \tag{9.5.110}$$

Can you prove via a direct calculation that H, and therefore L, is a constant of motion? (In fact, Hamiltonian dynamics tells us, as long as L does not explicitly depend on the affine parameter λ, the right hand side of eq. (9.5.109) is necessarily a constant of motion.)

Geodesic Equations Show that Hamilton's equations

$$\frac{\mathrm{d}z^i}{\mathrm{d}\lambda} = \frac{\partial H}{\partial p_i} = g^{ij} p_j, \tag{9.5.111}$$

$$\frac{\mathrm{d}p_i}{\mathrm{d}\lambda} = -\frac{\partial H}{\partial z^i} = -\frac{1}{2}(\partial_i g^{ab}) p_a p_b \tag{9.5.112}$$

are equivalent to the geodesic equation (9.5.44). Hint: You may need to use the 'integration-by-parts' identity $(\partial_i g^{ab}) g_{bc} = -g^{ab}\partial_i g_{bc}$. Why is this true? □

Problem 9.33. RNC and Christoffel Symbols It is always possible to find a coordinate system with coordinates $\vec{y}$ such that, as $\vec{y} \to \vec{y}_0$, the Christoffel symbols vanish

$$\Gamma^k{}_{ij}(\vec{y}_0) = 0. \tag{9.5.113}$$

Can you demonstrate why this is true from the 'local flatness principle' encoded in eq. (9.2.1)? Use this observation to explain why the product rule *has* to hold for covariant derivatives.

Hint: it is important that, locally, the first deviation from flat space is quadratic in the displacement vector $(y - y_0)^i$. □

Remark That there is always an orthonormal frame where the metric is flat – recall eq. (9.4.11) – as well as the existence of a locally flat coordinate system, is why the measure of curvature, in particular the Riemann tensor in eq. (9.5.30), depends on first and second derivatives of the metric. Specifically, when eq. (9.5.113) holds but space is curved, we would have from eq. (9.5.30),

$$R^i{}_{jmn}(\vec{y}_0) = \partial_m \Gamma^i{}_{nj}(\vec{y}_0) - \partial_n \Gamma^i{}_{mj}(\vec{y}_0). \tag{9.5.114}$$

Problem 9.34. Christoffel $\Gamma_{ijk} \equiv g_{il}\Gamma^l{}_{jk}$ contains as much information as $\partial_i g_{ab}$ Why do the Christoffel symbols take on the form in eq. (9.5.2)? It comes from assuming that the Christoffel symbol obeys the symmetry

$$\Gamma^i{}_{jk} = \Gamma^i{}_{kj} \tag{9.5.115}$$

– this is the torsion-free condition – and demanding that the covariant derivative of a metric is a zero tensor,

$$\nabla_i g_{jk} = 0. \tag{9.5.116}$$

This can be expanded as

$$\nabla_i g_{jk} = 0 = \partial_i g_{jk} - \Gamma^l{}_{ij} g_{lk} - \Gamma^l{}_{ik} g_{jl}. \tag{9.5.117}$$

Expand also $\nabla_j g_{ki}$ and $\nabla_k g_{ij}$, and show that

$$2\Gamma^l{}_{ij} g_{lk} = \partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}. \tag{9.5.118}$$

Divide both sides by 2 and contract both sides with g^{km} to obtain $\Gamma^m{}_{ij}$ in eq. (9.5.2).

Remark Incidentally, while eq. (9.5.2) tells us the Christoffel symbol can be written in terms of the first derivatives of the metric; eq. (9.5.117) indicates the first derivative of the metric can also always be expressed in terms of the Christoffel symbols. In other words, $\partial_i g_{ab}$ contains as much information as $\Gamma^i{}_{jk}$, provided of course that g_{ij} itself is known. $\square$

Problem 9.35. Covariant form of $(\pounds_{\vec{\xi}} g)_{ij}$ Show that the Lie derivative of the metric $(\pounds_{\vec{\xi}} g)_{ij}$ in eq. (9.3.5) can be re-written in a more covariant looking expression

$$\left(\pounds_{\vec{\xi}} g\right)_{ij} (\vec{x}') = \nabla_i \xi_j + \nabla_j \xi_i \equiv \nabla_{\{i}\xi_{j\}}. \tag{9.5.119}$$

Some nomenclature: $(\pounds_{\vec{\xi}} g)_{ij} = 0$ is known as Killing's equation, and a vector that satisfies Killing's equation is called a Killing vector. Showing that $(\pounds_{\vec{\xi}} g)_{ij}$ is a tensor indicates such a characterization of symmetry is a generally covariant statement.

Hint: Convert all partial derivatives into covariant ones by adding/subtracting Christoffel symbols appropriately; for instance $\partial_a \xi^i = \nabla_a \xi^i - \Gamma^i{}_{ab}\xi^b$. $\square$

We may now rephrase the discussion leading up to eq. (9.3.5) as follows. Under an infinitesimal coordinate transformation $\vec{x} = \vec{x}' + \vec{\xi}(\vec{x}')$, where $\vec{\xi}$ is considered 'small', the metric transforms as

$$g_{ij}(\vec{x})\, \mathrm{d}x^i \mathrm{d}x^j = \left(g_{ij}(\vec{x} \to \vec{x}') + \nabla_{\{i}\xi_{j\}}(\vec{x} \to \vec{x}') + \mathcal{O}(\vec{\xi}^2)\right) \mathrm{d}x'^i \mathrm{d}x'^j. \tag{9.5.120}$$

The metric is said to enjoy a symmetry along $\vec{\xi}$ iff $\nabla_{\{i}\xi_{j\}} = 0$ along its integral curve.

Problem 9.36. Lie vs. Covariant Derivative Explain why the partial derivatives on the right hand side of eq. (11.8.51) may be replaced with covariant ones, namely

$$\left(\pounds_{\vec{\xi}} T\right)^{i_1 \dots i_N}{}_{j_1 \dots j_M} = \xi^l \nabla_l T^{i_1 \dots i_N}{}_{j_1 \dots j_M}$$

$$- T^{l i_2 \dots i_N}{}_{j_1 \dots j_M} \nabla_l \xi^{i_1} - \dots - T^{i_1 \dots i_{N-1} l}{}_{j_1 \dots j_M} \nabla_l \xi^{i_N}$$
$$+ T^{i_1 \dots i_N}{}_{l j_2 \dots j_M} \nabla_{j_1} \xi^l + \dots + T^{i_1 \dots i_N}{}_{j_1 \dots j_{M-1} l} \nabla_{j_M} \xi^l. \tag{9.5.121}$$

(Hint: First explain why $\partial_a \xi^b = \nabla_a \xi^b - \Gamma^b{}_{al} \xi^l$.) That the Lie derivative of a tensor can be expressed in terms of covariant derivatives indicates the former is a tensor. □

Problem 9.37. Prove that the tensor $T_{i_1}{}^{i_2 \dots i_N}$ is zero if and only if the corresponding tensor $T_{i_1 i_2 \dots i_N}$ is zero. Then, using the product rule, explain why $\nabla_i g_{jk} = 0$ implies $\nabla_i g^{jk} = 0$. Hint: start with $\nabla_i (g_{aj} g_{bk} g^{jk})$. □

Problem 9.38. 3D Flat Space Christoffel Symbols in Spherical Coordinates Calculate the Christoffel symbols of the 3-dimensional Euclidean metric in Cartesian coordinates δ_{ij}. Then calculate the Christoffel symbols for the same space, but in spherical coordinates:

$$(\mathrm{d}\ell)^2 = \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2). \tag{9.5.122}$$

To start you off, the non-zero components of the metric are

$$g_{rr} = 1, \qquad g_{\theta\theta} = r^2, \qquad g_{\phi\phi} = r^2 (\sin\theta)^2; \tag{9.5.123}$$
$$g^{rr} = 1, \qquad g^{\theta\theta} = r^{-2}, \qquad g^{\phi\phi} = \frac{1}{r^2 (\sin\theta)^2}. \tag{9.5.124}$$

You should find

$$\begin{aligned} &\Gamma^r{}_{\theta\theta} = -r, \qquad \Gamma^r{}_{\phi\phi} = -r(\sin\theta)^2, \\ &\Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \frac{1}{r}, \qquad \Gamma^\theta{}_{\phi\phi} = -\cos\theta \cdot \sin\theta, \\ &\Gamma^\phi{}_{r\phi} = \Gamma^\phi{}_{\phi r} = \frac{1}{r}, \qquad \Gamma^\phi{}_{\theta\phi} = \Gamma^\phi{}_{\phi\theta} = \cot\theta. \end{aligned} \tag{9.5.125}$$

How are these Christoffel symbols of 3D Euclidean space in spherical coordinates related to those of the 2-sphere in equations (9.5.26) and (9.5.27)? (This should serve as an independent check of your computations.) Hint: Relate the 2D and 3D versions of g_{ij} for $i, j \in \{\theta, \phi\}$; then followed by the $\Gamma^\theta{}_{ij}$ and $\Gamma^\phi{}_{ij}$. □

Problem 9.39. Christoffel Symbols: Cartesian to Curvilinear Derive the flat space Christoffel symbols in spherical coordinates from their Cartesian counterparts using eq. (9.5.13). That is, if $\vec{x}$ are Cartesian and $\vec{\xi}$ are curvilinear coordinates,

$$\Gamma^l{}_{ij}(\vec{\xi}) = \frac{\partial \xi^l}{\partial x^k(\vec{\xi})} \frac{\partial^2 x^k(\vec{\xi})}{\partial \xi^j \partial \xi^i}. \tag{9.5.126}$$

This lets you cross-check your results in Problem (9.38); you should also feel free to use software to help. The answer is given in eq. (9.5.125).

To provide an example for this latter method, let us calculate the Christoffel symbols of 2D flat space written in cylindrical coordinates $\xi^i \equiv (r, \phi)$,

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\phi, \qquad r \geq 0, \; \phi \in [0, 2\pi). \tag{9.5.127}$$

This means the non-zero components of the metric are

$$g_{rr} = 1, \qquad g_{\phi\phi} = r^2, \qquad g^{rr} = 1, \qquad g^{\phi\phi} = r^{-2}. \tag{9.5.128}$$

Keeping the diagonal nature of the metric in mind, let us start with

$$\begin{aligned} \Gamma^r{}_{ij} &= \frac{1}{2} g^{rk} \left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij} \right) = \frac{1}{2} g^{rr} \left(\partial_i g_{jr} + \partial_j g_{ir} - \partial_r g_{ij} \right) \\ &= \frac{1}{2} \left(\delta^r_j \partial_i g_{rr} + \delta^r_i \partial_j g_{rr} - \delta^\phi_i \delta^\phi_j \partial_r r^2 \right) = -\delta^\phi_i \delta^\phi_j r. \end{aligned} \tag{9.5.129}$$

In the third equality we have used the fact that the only g_{ij} that depends on r (and therefore yield a non-zero r-derivative) is $g_{\phi\phi}$. Now for the

$$\begin{aligned} \Gamma^\phi{}_{ij} &= \frac{1}{2} g^{\phi\phi} \left(\partial_i g_{j\phi} + \partial_j g_{i\phi} - \partial_\phi g_{ij} \right) \\ &= \frac{1}{2r^2} \left(\delta^\phi_j \partial_i g_{\phi\phi} + \delta^\phi_i \partial_j g_{\phi\phi} \right) = \frac{1}{2r^2} \left(\delta^\phi_j \delta^r_i \partial_r r^2 + \delta^\phi_i \delta^r_j \partial_r r^2 \right) \\ &= \frac{1}{r} \left(\delta^\phi_j \delta^r_i + \delta^\phi_i \delta^r_j \right). \end{aligned} \tag{9.5.130}$$

If we had started from Cartesian coordinates x^i,

$$x^i = r(\cos\phi, \sin\phi), \tag{9.5.131}$$

we know the Christoffel symbols in Cartesian coordinates are all zero, since the metric components are constant. If we wish to use eq. (9.5.13) to calculate the Christoffel symbols in (r, ϕ), the first term on the right hand side is zero and what we need are the $\partial x/\partial\xi$ and $\partial^2 x/\partial\xi\partial\xi$ matrices. The first derivative matrices are

$$\frac{\partial x^i}{\partial \xi^j} = \begin{bmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{bmatrix}^i{}_j \tag{9.5.132}$$

$$\frac{\partial \xi^i}{\partial x^j} = \left(\left(\frac{\partial x}{\partial \xi} \right)^{-1} \right)^i{}_j = \begin{bmatrix} \cos\phi & \sin\phi \\ -r^{-1}\sin\phi & r^{-1}\cos\phi \end{bmatrix}^i{}_j, \tag{9.5.133}$$

whereas the second derivative matrices are

$$\frac{\partial^2 x^1}{\partial \xi^i \xi^j} = \begin{bmatrix} 0 & -\sin\phi \\ -\sin\phi & -r\cos\phi \end{bmatrix} \tag{9.5.134}$$

$$\frac{\partial^2 x^2}{\partial \xi^i \xi^j} = \begin{bmatrix} 0 & \cos\phi \\ \cos\phi & -r\sin\phi \end{bmatrix}. \tag{9.5.135}$$

Therefore, from eq. (9.5.13),

$$\begin{aligned} \Gamma^r{}_{ij}(r, \phi) &= \frac{\partial r}{\partial x^k} \frac{\partial x^k}{\partial \xi^i \partial \xi^j} \\ &= \cos\phi \cdot \begin{bmatrix} 0 & -\sin\phi \\ -\sin\phi & -r\cos\phi \end{bmatrix} + \sin\phi \cdot \begin{bmatrix} 0 & \cos\phi \\ \cos\phi & -r\sin\phi \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -r \end{bmatrix}. \end{aligned} \tag{9.5.136}$$

Similarly,

$$\Gamma^{\phi}{}_{ij}(r,\phi) = \frac{\partial \phi}{\partial x^k} \frac{\partial x^k}{\partial \xi^i \partial \xi^j} \tag{9.5.137}$$

$$= -r^{-1} \sin\phi \begin{bmatrix} 0 & -\sin\phi \\ -\sin\phi & -r\cos\phi \end{bmatrix} + r^{-1}\cos\phi \begin{bmatrix} 0 & \cos\phi \\ \cos\phi & -r\sin\phi \end{bmatrix} = \begin{bmatrix} 0 & r^{-1} \\ r^{-1} & 0 \end{bmatrix}.$$

Of course, the generalization of this Cartesian to Spherical method works in non-flat geometries too, as long as we already know the Γ's on the right hand side of (9.5.13). □

Problem 9.40. Flat Space Curvilinear Christoffel Symbols Prove eq. (9.1.57) from eq. (9.5.126). Note that eq. (9.1.57) allows you to compute the curvilinear Christoffel symbols in flat space using the inverse metric $g^{ab}(\vec{\xi})$ and the first and second derivatives of $\vec{x}(\vec{\xi})$ without computing the inverse Jacobian $\partial \xi^l / \partial x^i$. Which formula is more convenient depends on whether you know the inverse metric or the inverse Jacobian. □

Higher Time Derivatives Since tangent vectors are (fictitious) velocities, $v^i = \mathrm{d}x^i/\mathrm{d}\lambda$, could we construct from them the corresponding accelerations, jerks, etc.? If we were in flat space using Cartesian coordinates, the acceleration a^i would be $v^j \partial_j v^i = (\mathrm{d}x^j/\mathrm{d}\lambda)\partial_j v^i$; and higher 'time' derivatives would follow from acting $v^j \partial_j$ repeatedly on v^i. In curved spaces or in flat ones but using curvilinear coordinates, we would instead have to apply the covariant derivative along $\vec{v}$, namely $v^j \nabla_j$ instead of $v^j \partial_j$, to ensure the end result still transforms as a vector. Hence, the covariant acceleration and jerk are, respectively,

$$\frac{D^2 x^i}{\mathrm{d}\lambda^2} = v^j \nabla_j v^i, \tag{9.5.138}$$

$$\frac{D^3 x^i}{\mathrm{d}\lambda^3} = v^k \nabla_k (v^j \nabla_j v^i). \tag{9.5.139}$$

When $v^i = \mathrm{d}x^i/\mathrm{d}\lambda$, we therefore have

$$\left(\frac{D}{\mathrm{d}\lambda}\right)^i{}_j = \delta^i{}_j \frac{\mathrm{d}}{\mathrm{d}\lambda} + v^l \Gamma^i{}_{lj}. \tag{9.5.140}$$

Problem 9.41. 2D Non-Relativistic Classical Mechanics: Polar Coordinates Show that the components of acceleration

$$A = \ddot{x}^i(t)\partial_{x^i} = \left(\ddot{y}^j + \Gamma^j{}_{mn}\dot{y}^m \dot{y}^n\right)\partial_{y^j} \tag{9.5.141}$$

in 2D flat space in polar coordinates $\vec{x}(\vec{y} \equiv (r,\phi)) = r(\cos\phi, \sin\phi)$ written in an orthonormal frame are

$$A^{\widehat{r}} = \ddot{r}(t) - \dot{\phi}(t)^2 r(t), \tag{9.5.142}$$

$$A^{\widehat{\phi}} = r(t)\ddot{\phi}(t) + 2\dot{r}(t)\dot{\phi}(t). \tag{9.5.143}$$

This recovers uniform circular motion when r and $\dot{\phi} \equiv \omega$ are constant: $A^{\widehat{a}}\widehat{e}_a = -(\omega^2 r)\widehat{e}_r$. □

Problem 9.42. Force-Free Classical Mechanics On The 2-Sphere The force-free version of Newton's second law in an arbitrary curved space can be viewed as geodesic motion:

$$D^2 z^i/\mathrm{d}t^2 = 0, \tag{9.5.144}$$

where z^i is the trajectory of the point mass and the affine parameter is now time t. On the unit 2−sphere, show that this force-free Newton's second law becomes

$$\ddot{\theta}(t) - \sin\theta(t)\cos\theta(t)\cdot\dot{\phi}(t)^2 = 0, \tag{9.5.145}$$

$$\ddot{\phi}(t) + 2\cot\theta(t)\cdot\dot{\theta}(t)\dot{\phi}(t) = 0. \tag{9.5.146}$$

Previously, we have verified that ∂_θ is parallel transported onto itself along the equator. Suppose at $t=0$ we release two masses m_1 and m_2 from the equator with velocities tangent, respectively, to the fixed azimuthal angles ϕ_1 and ϕ_2 lines. If their initial velocities (at $t=0$) are both $-v_0\partial_\theta$, with the − sign indicating they are pointing towards the North Pole, show that their geodesic (i.e., force-free) trajectories are

$$(\theta_{1,2}, \phi_{1,2}) = \left(\frac{\pi}{2} - v_0 t, \phi_{1,2}\right). \tag{9.5.147}$$

Then, show that the geodesic distance between the two trajectories at a given time – right up to their meeting at the North Pole – is

$$\begin{aligned} \ell(m_1 \leftrightarrow m_2) &= \arccos\left(\cos(v_0 t)^2\cos(\Delta\phi) + \sin(v_0 t)^2\right) && (9.5.148)\\ &= \cos(v_0 t)|\Delta\phi| + \mathcal{O}(\Delta\phi^3), && (9.5.149)\end{aligned}$$

where $\Delta\phi \equiv \phi_2 - \phi_1$; and in the second equality we have assumed $\Delta\phi$ is small enough that we may retain only the first order term. Finally, argue that, when the geodesic trajectories are close enough – so that their local region of space is nearly flat – the displacement vector joining m_1 to m_2 is

$$\vec{\Delta}_{1\to 2} = \cos(v_0 t)\Delta\phi\cdot\varepsilon_{\widehat{\phi}}, \tag{9.5.150}$$

where $\varepsilon_{\widehat{\phi}} = (\cos(v_0 t))^{-1}\partial_\phi$ is the unit vector tangent to the constant-θ (latitude) lines.

This example illustrates: on the curved space of the 2−sphere, not only do initially parallel geodesic lines eventually meet, the relative displacement between nearby geodesics does not undergo constant velocity motion, unlike their flat space counterparts. Hence, there are no global inertial frames as defined by Newton's first law of motion. □

Methods of Computing Christoffel Symbols At this juncture, we may summarize the following methods of calculating Christoffel symbols.

- Do it by brute force, using eq. (9.5.2).
- Use the Lagrangian method: apply the Euler-Lagrangian equations to the Lagrangian $L_g = (1/2)g_{ij}\dot{z}^i\dot{z}^j$, and read off the Christoffel symbols from the $\Gamma^i{}_{ab}\dot{z}^a\dot{z}^b$ terms of the resulting ODEs.

- *If working in flat space(time)*, the Christoffel symbols in a curvilinear coordinate system $\vec{\xi}$ can be obtained through its relation to the Cartesian ones $\vec{x}$ through eq. (9.1.57) or (9.5.126). $\square$

Variation of the metric and divergence of tensors Suppose we perturb the metric slightly

$$g_{ij} \to g_{ij} + h_{ij}, \tag{9.5.151}$$

where the components of h_{ij} are to be viewed as "small", and its indices are moved with the metric; for e.g.,

$$h^i{}_j = g^{ia} h_{aj}. \tag{9.5.152}$$

The inverse metric will become

$$g^{ij} \to g^{ij} - h^{ij} + h^{ik} h_k{}^j + \mathcal{O}\left(h^3\right), \tag{9.5.153}$$

then the square root of the determinant of the metric will change as

$$\sqrt{|g|} \to \sqrt{|g|} \left(1 + \frac{1}{2} g^{ab} h_{ab} + \mathcal{O}(h^2)\right). \tag{9.5.154}$$

Problem 9.43. Use the matrix identity in eq. (5.6.152), where for any square matrix X,

$$\det e^X = e^{\mathrm{Tr}[X]}, \tag{9.5.155}$$

to prove eq. (9.5.154). (The Tr X means the trace of the matrix X – sum over its diagonal terms.) Hint: Start with $\det(g_{ij} + h_{ij}) = \det(g_{ij}) \cdot \det(\delta^a{}_b + h^a{}_b)$. Then massage $\delta^i{}_j + h^i{}_j = \exp(\ln(\delta^i_j + h^i{}_j))$. $\square$

Problem 9.44. Use eq. (9.5.154) and the definition of the Christoffel symbol to show that

$$\partial_i \ln \sqrt{|g|} = \frac{1}{2} g^{ab} \partial_i g_{ab} = \Gamma^s{}_{is}. \tag{9.5.156}$$

This formula is of use in understanding the generalization of 'divergence' in multi-variable calculus to that in differential geometry of curved space(time)s. $\square$

Problem 9.45. Divergence of tensors. By exploiting eq. (9.5.156), verify the following formulas for the divergence of a vector V^i, a fully antisymmetric rank-($N \leq D$) tensor $F^{i_1 i_2 \dots i_N}$ and a symmetric tensor $S^{ij} = S^{ji}$,

$$\nabla_i V^i = \frac{\partial_i \left(\sqrt{|g|} V^i\right)}{\sqrt{|g|}}, \tag{9.5.157}$$

$$\nabla_j F^{j i_2 \dots i_N} = \frac{\partial_j \left(\sqrt{|g|} F^{j i_2 \dots i_N}\right)}{\sqrt{|g|}}, \tag{9.5.158}$$

$$\nabla_i S^{ij} = \frac{\partial_i\left(\sqrt{|g|}S^{ij}\right)}{\sqrt{|g|}} + \Gamma^j{}_{ab}S^{ab}. \tag{9.5.159}$$

Note that, fully antisymmetric means, swapping any pair of indices (say, $i_a \leftrightarrow i_b$) costs a minus sign,

$$F^{i_1\dots i_{a-1}i_a i_{a+1}\dots i_{b-1}i_b i_{b+1}\dots i_N} = -F^{i_1\dots i_{a-1}i_b i_{a+1}\dots i_{b-1}i_a i_{b+1}\dots i_N}. \tag{9.5.160}$$

Comment on how these expressions, equations (9.5.157)-(9.5.159), transform under a coordinate transformation, i.e., $\vec{x} \to \vec{x}(\vec{\xi})$. $\square$

Gradient of a scalar It is worth highlighting that the gradient of a scalar, with upper indices, depends on the metric; whereas the covariant derivative on the same scalar, with lower indices, does not.

$$\nabla^i\varphi = g^{ij}\nabla_j\varphi = g^{ij}\partial_j\varphi. \tag{9.5.161}$$

This means, even in flat space, $\nabla^i\varphi$ is not always equal to $\nabla_i\varphi$. (They are equal in Cartesian coordinates.) For instance, in spherical coordinates (r, θ, ϕ), where

$$g^{ij} = \mathrm{diag}(1, r^{-2}, r^{-2}(\sin\theta)^{-2}); \tag{9.5.162}$$

the gradient of a scalar is

$$\nabla^i\varphi = \left(\partial_r\varphi, r^{-2}\partial_\theta\varphi, r^{-2}(\sin\theta)^{-2}\partial_\phi\varphi\right). \tag{9.5.163}$$

while the same object with lower indices is simply

$$\nabla_i\varphi = (\partial_r\varphi, \partial_\theta\varphi, \partial_\phi\varphi). \tag{9.5.164}$$

Problem 9.46. For an arbitrary scalar $f(\vec{x})$, its variation is $\mathrm{d}f = \partial_i f \mathrm{d}x^i$ and its gradient is $\vec{\nabla}f \equiv \nabla^i f \partial_i$. Explain why their 'squares' are equal:

$$\langle \mathrm{d}f | \, \mathrm{d}f \rangle = \left\langle \vec{\nabla}f \middle| \vec{\nabla}f \right\rangle = \left\langle \mathrm{d}f \middle| \vec{\nabla}f \right\rangle. \tag{9.5.165}$$

$\square$

Laplacian of a scalar The Laplacian of a scalar ψ can be thought of as the divergence of its gradient. In 3D vector calculus you would write is as $\vec{\nabla}^2$ but in curved spaces we may also write it as $\square$ or $\nabla_i\nabla^i$:

$$\square\psi \equiv \vec{\nabla}^2\psi = \nabla_i\nabla^i\psi = g^{ij}\nabla_i\nabla_j\psi. \tag{9.5.166}$$

A short calculation would confirm that this formula does indeed return the familiar flat space $(g_{ij} = \delta_{ij})$ Laplacian in Cartesian coordinates,

$$\vec{\nabla}^2\psi = \delta^{ij}\partial_i\partial_j\psi, \tag{9.5.167}$$

because $\Gamma^i{}_{mn} = 0$. Of course, even in flat space, using curvilinear coordinates would yield more complicated expressions because the Christoffel symbols would be non-zero. However, as the next problem shows, there is actually no need for dealing with $\Gamma^i{}_{mn}$ explicitly when computing the Laplacian.

Problem 9.47. Show that the Laplacian of a scalar can be written more explicitly in terms of the determinant of the metric and the inverse metric as

$$\vec{\nabla}^2\psi \equiv \nabla_i \nabla^i \psi = \frac{1}{\sqrt{|g|}}\partial_i \left(\sqrt{|g|} g^{ij}\partial_j \psi\right). \tag{9.5.168}$$

Hint: Use eq. (9.5.157). □

For instance, the unit circle (embedded in 2D flat space) can be described by

$$\mathrm{d}\ell^2 = \mathrm{d}\phi^2, \qquad \phi \in [0, 2\pi). \tag{9.5.169}$$

The square root of the determinant of the metric is 1 and the inverse metric is $g^{\phi\phi} = 1$ and

$$\vec{\nabla}^2_{\mathbb{S}^1}\psi = \partial_\phi^2 \psi. \tag{9.5.170}$$

For the 2−sphere, the square root of the determinant of its metric in (θ, ϕ) coordinates is $\sin\theta$ and its inverse metric is $g^{ij} = (1, 1/\sin^2\theta)$. Therefore,

$$\vec{\nabla}_{\mathbb{S}^2}\psi = \frac{1}{\sin\theta}\left(\partial_\theta\left(\sin\theta \cdot \partial_\theta \psi\right) + \partial_\phi\left((\sin\theta)^{1-2}\partial_\phi\psi\right)\right) \tag{9.5.171}$$

$$= \partial_\chi\left((1-\chi^2)\partial_\chi\psi\right) + \frac{\vec{\nabla}_{\mathbb{S}^1}\psi}{1-\chi^2}, \qquad \chi \equiv \cos\theta. \tag{9.5.172}$$

Problem 9.48. Laplacian on $\mathbb{S}^D$ If $\vec{\nabla}^2_{\mathbb{S}^D}$ denotes the Laplacian on the $(D \geq 2)$−sphere, show that it is related to the Laplacian on the $D-1$ sphere via the relation

$$\vec{\nabla}^2_{\mathbb{S}^D}\psi = \frac{\partial_{c_D}\left((1-c_D^2)^{D/2}\partial_{c_D}\psi\right)}{(1-c_D^2)^{\frac{D-2}{2}}} + \frac{\vec{\nabla}^2_{\mathbb{S}^{D-1}}\psi}{1-c_D^2} \tag{9.5.173}$$

where $c_D \equiv \cos\theta^D$ is the cosine of the D−th angle. (Hint: Recall equations (9.2.101) and (9.2.104).) This relation allows us to diagonalize the Laplacian on the D−sphere for $D > 2$. Assume that

$$\vec{\nabla}^2_{\mathbb{S}^D} Y^{(D)}(\vec{\theta}) = -\ell_D(\ell_D + D - 1) Y^{(D)}(\vec{\theta}), \tag{9.5.174}$$

where $Y^{(D)}$ here denotes the unit norm eigenfunction of the D−sphere Laplacian – i.e., $Y^{(D)}$ is the D−dimensional spherical harmonic – and $\ell_{D\geq 2} \geq 0$ is an integer. Proceed to verify the following recursion relation between the spherical harmonic on the D−sphere and that on the $(D-1)$−sphere:

$$Y^{(D)}(\theta^1, \dots, \theta^D) = \sqrt{\frac{(2\ell_D + D - 1)\cdot(\ell_D - \ell_{D-1})!}{2\Gamma(\ell_D + \ell_{D-1} + D - 1)}} \times \frac{P^{\ell_{D-1}+\frac{D}{2}-1}_{\ell_D + \frac{D}{2}-1}\left(\cos\theta^D\right)}{(\sin\theta^D)^{\frac{D}{2}-1}} Y^{(D-1)}(\theta^1, \dots, \theta^{D-1}), \tag{9.5.175}$$

where $P^{\mu}_{\nu}(\cos\theta^D)$ is the Legendre function. The unit norm constraint, of course, only fixes the prefactor up to a multiplicative phase factor. Hint: You may find DLMF §14.8, §14.9(ii) and 14.17.6 useful. **YZ: A note of caution. I have assumed the latter DLMF formula holds for half-integer degree and order as well; i.e., for odd D. This needs to be checked.**

The $D-$dimensional spherical harmonic can therefore be labeled by D integers $\{\ell_1, \dots, \ell_D\}$:

$$Y_{\ell_1,\ell_2,\dots,\ell_D}(\vec{\theta}) = \prod_{\text{i}=2}^{D} \left(\sqrt{\frac{(2\ell_i + i - 1)\cdot(\ell_i - \ell_{i-1})!}{2\Gamma(\ell_i + \ell_{i-1} + i - 1)}} \frac{P^{\ell_{i-1}+\frac{i}{2}-1}_{\ell_i+\frac{i}{2}-1}(\cos\theta^i)}{(\sin\theta^i)^{\frac{i}{2}-1}} \right) \cdot \frac{e^{i\ell_1\theta^1}}{\sqrt{2\pi}}, \tag{9.5.176}$$

$$\theta^1 \in [0, 2\pi), \qquad \theta^2, \dots, \theta^D \in [0, \pi], \tag{9.5.177}$$

$$\ell_1 \in \{0, \pm1, \pm2, \pm3, \dots, \pm\ell_2\}, \qquad \ell_2, \ell_3, \dots, \ell_D \in \{0, 1, 2, 3, \dots\}; \tag{9.5.178}$$

where we have made the replacement $Y^{(D)} \to Y_{\ell_1,\dots,\ell_D}$. Comment on how the restriction of the range of ℓ_D depends on ℓ_{D-1}. □

Problem 9.49. Parity of Spherical Harmonics

Remarks on Scalar $\vec{\nabla}^2$ on Tensor In practical computations, one may encounter the scalar $\vec{\nabla}^2(\cdot) \equiv \nabla_i\nabla^i(\cdot) = |g|^{-1/2}\partial_a(|g|^{1/2}g^{ab}\partial_b\cdot)$ acting a tensor. For instance, electromagnetic calculations often lead us to

$$\vec{\nabla}^2 A_i \equiv \frac{1}{\sqrt{|g|}}\partial_a\left(\sqrt{|g|}g^{ab}\partial_b A_i\right), \tag{9.5.179}$$

where A_i is the flat space vector potential in the Cartesian coordinate system. If we recall the 'Covariant Derivative' discussion at the beginning of this section, we would recognize that this is not a tensor under coordinate transformations. To obtain a tensor expression, we would have to add to eq. (9.5.179) terms that involve Christoffel symbols, because the latter would transform in such a way that would cancel the non-tensorial portion of $\vec{\nabla}^2 A_i$ expressed in a different coordinate system – i.e., the derivatives acting on the Jacobian contracted with the vector potential. But all these *do not* imply we cannot compute the Laplacian portion of $\vec{\nabla}^2 A_i$ in a different coordinate system $\vec{x} = \vec{x}(\vec{y})$. Specifically, if we choose to remain in the same coordinate basis $\{\mathrm{d}x^i\}$ then the vector potential itself reads

$$A_i(\vec{x})\mathrm{d}x^i = A_i(\vec{x}(\vec{y}))\mathrm{d}x^i \tag{9.5.180}$$

– namely, the $A_i(\vec{x}) = A_i(\vec{x}(\vec{y}))$ are now treated as scalars – but if we now compute the scalar Laplacian acting on it with respect to $\vec{y}$ instead, then we must have the relation

$$\begin{aligned}\vec{\nabla}^2 A_i &= \frac{1}{\sqrt{|g(\vec{x})|}}\partial_{x^a}\left(\sqrt{|g(\vec{x})|}g^{ab}(\vec{x})\partial_{x^b}A_i(\vec{x})\right) \\ &= \frac{1}{\sqrt{|g'(\vec{y})|}}\partial_{y^m}\left(\sqrt{|g'(\vec{y})|}g'^{mn}(\vec{y})\partial_{y^b}A_i(\vec{x}(\vec{y}))\right);\end{aligned} \tag{9.5.181}$$

where $|g'(\vec{y})|$ denotes the determinant of the metric $g'_{mn}(\vec{y}) = (\partial x^a/\partial y^m)(\partial x^b/\partial y^n)g_{ab}(\vec{x}(\vec{y}))$ expressed in the new $\vec{y}$-coordinate system and $g'^{mn}(\vec{y})$ is its inverse. We reiterate: that the second equality of eq. (9.5.181) has to follow from its first, is because we are now effectively treating A_i as scalars under coordinate transformations.

Problem 9.50. Example: Scalar Laplacian in Flat Space To further understand the transformation from $\vec{x} \to \vec{y}$ in the first and second equalities of eq. (9.5.181), let us specialize to flat space Cartesian coordinates: $\mathrm{d}\ell^2 = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j$. Consider transforming $\vec{x}$ to some other coordinate system $\vec{x} = \vec{x}(\vec{y})$. Calculus tells us,

$$\begin{aligned}\vec{\nabla}^2 A_i &= \delta^{ab}\partial_{x^a}\partial_{x^b}A_i(\vec{x}) = \delta^{ab}\partial_{x^a}\left(\frac{\partial y^n}{\partial x^b}\partial_{y^n}A_i\right) \\ &= \delta^{ab}\frac{\partial y^m}{\partial x^a}\frac{\partial y^n}{\partial x^b}\partial_{y^m}\partial_{y^n}A_i(\vec{x}(\vec{y})) + \delta^{ab}\frac{\partial y^n}{\partial x^a \partial x^b}\partial_{y^n}A_i(\vec{x}(\vec{y})). \end{aligned} \tag{9.5.182}$$

Explain why the final equality is, within this context, equivalent to the second equality of eq. (9.5.181). Hint: Refer to equations (9.5.13) and (9.5.156). □

9.6 Levi-Civita (Pseudo-)Tensor, Hodge Dual, Exterior Derivative

Levi-Civita (Pseudo-)Tensor We have just seen how to write the divergence in any curved or flat space. We will now see that the curl from vector calculus also has a differential geometric formulation as an antisymmetric tensor, which will allow us to generalize the former to not only curved spaces but also arbitrary dimensions greater than 2. But first, we introduce the Levi-Civita tensor, and with it, the Hodge dual.

In D spatial dimensions we first define a Levi-Civita *symbol*

$$\epsilon_{i_1 i_2 \dots i_{D-1} i_D}. \tag{9.6.1}$$

It is defined by the following properties.

- It is completely antisymmetric in its indices. This means swapping any of the indices $i_a \leftrightarrow i_b$ (for $a \neq b$) will return

$$\epsilon_{i_1 i_2 \dots i_{a-1} i_a i_{a+1} \dots i_{b-1} i_b i_{b+1} \dots i_{D-1} i_D} = -\epsilon_{i_1 i_2 \dots i_{a-1} i_b i_{a+1} \dots i_{b-1} i_a i_{b+1} \dots i_{D-1} i_D}. \tag{9.6.2}$$

- For a given ordering of the D distinct coordinates $\{x^i | i = 1, 2, 3, \dots, D\}$, $\epsilon_{123\dots D} \equiv 1$. Below, we will have more to say about this choice.

These are sufficient to define every component of the Levi-Civita symbol. From the first definition, if any of the D indices are the same, say $i_a = i_b$, then the Levi-Civita symbol returns zero. (Why?) From the second definition, when all the indices are distinct, $\epsilon_{i_1 i_2 \dots i_{D-1} i_D}$ is a $+1$ if it takes even number of swaps to go from $\{1, \dots, D\}$ to $\{i_1, \dots, i_D\}$; and is a -1 if it takes an odd number of swaps to do the same. We may summarize:

$$\epsilon_{i_1 \dots i_D} = \text{sign of permutation from } \{1, \dots, D\} \text{ to } \{i_1, \dots, i_D\}. \tag{9.6.3}$$

For example, in the 2 and 3 dimensional cases and in Cartesian coordinates (x^1, x^2, x^3),

$$1 = \epsilon_{12} = -\epsilon_{21}, \tag{9.6.4}$$

$$1 = \epsilon_{123} = -\epsilon_{213} = -\epsilon_{321} = -\epsilon_{132} = \epsilon_{231} = \epsilon_{312}. \tag{9.6.5}$$

The Levi-Civita *tensor* $\widetilde{\epsilon}_{i_1 \dots i_D}$ is defined as

$$\widetilde{\epsilon}_{i_1 i_2 \dots i_D} \equiv \sqrt{|g|}\epsilon_{i_1 i_2 \dots i_D}. \tag{9.6.6}$$

Let us understand why it is a (pseudo-)tensor. Because the Levi-Civita *symbol* is just a multi-index array of ± 1 and 0, it does not change under coordinate transformations. Equation (9.2.77) then implies

$$\sqrt{|g(\vec{\xi})|}\epsilon_{a_1 a_2 \dots a_D} = \sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|}\left|\det \frac{\partial x^i(\vec{\xi})}{\partial \xi^j}\right|\epsilon_{a_1 a_2 \dots a_D}. \tag{9.6.7}$$

On the right hand side, $\left|g\left(\vec{x}(\vec{\xi})\right)\right|$ is the absolute value of the determinant of g_{ij} written in the coordinates $\vec{x}$ but with $\vec{x}$ replaced with $\vec{x}(\vec{\xi})$.

If $\widetilde{\epsilon}_{i_1 i_2 \dots i_D}$ were a tensor, on the other hand, it must obey eq. (9.2.53),

$$\begin{aligned} \sqrt{|g(\vec{\xi})|}\epsilon_{a_1 a_2 \dots a_D} &\overset{?}{=} \sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|}\epsilon_{i_1 \dots i_D}\frac{\partial x^{i_1}}{\partial \xi^{a_1}} \dots \frac{\partial x^{i_D}}{\partial \xi^{a_D}}, \\ &= \sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|}\left(\det \frac{\partial x^i}{\partial \xi^j}\right)\epsilon_{a_1 \dots a_D}, \end{aligned} \tag{9.6.8}$$

where in the second line we have recalled the co-factor expansion determinant of any $D \times D$ matrix M in eq. (3.2.10),

$$\epsilon_{a_1 \dots a_D} \det M = \epsilon_{i_1 \dots i_D} M^{i_1}{}_{a_1} \dots M^{i_D}{}_{a_D}. \tag{9.6.9}$$

Comparing equations (9.6.7) and (9.6.8) tells us the Levi-Civita $\widetilde{\epsilon}_{a_1 \dots a_D}$ transforms as a tensor for *orientation-preserving* coordinate transformations, namely for all coordinate transformations obeying

$$\det \frac{\partial x^i}{\partial \xi^j} = \epsilon_{i_1 i_2 \dots i_D}\frac{\partial x^{i_1}}{\partial \xi^1}\frac{\partial x^{i_2}}{\partial \xi^2} \dots \frac{\partial x^{i_D}}{\partial \xi^D} > 0. \tag{9.6.10}$$

Parity flips This restriction on the sign of the determinant of the Jacobian means the Levi-Civita tensor is invariant under "parity", and is why I call it a pseudo-tensor. Parity flips are transformations that reverse the orientation of some coordinate axis, say $\xi^i \equiv -x^i$ (for some fixed i) and $\xi^j = x^j$ for $j \neq i$. For the Levi-Civita tensor,

$$\sqrt{g(\vec{x})}\epsilon_{i_1 \dots i_D} = \sqrt{g(\vec{\xi})}\left|\det \text{diag}[1, \dots, 1, \underbrace{-1}_{i\text{th component}}, 1, \dots, 1]\right|\epsilon_{i_1 \dots i_D} = \sqrt{g(\vec{\xi})}\epsilon_{i_1 \dots i_D}; \tag{9.6.11}$$

whereas, under the usual rules of coordinate transformations (eq. (9.2.53)) we would have expected a 'true' tensor $T_{i_1 \dots i_D}$ to behave, for instance, as

$$T_{(1)(2)\dots(i-1)(i)(i+1)\dots(D)}(\vec{x})\frac{\partial x^i}{\partial \xi^i} = -T_{(1)(2)\dots(i-1)(i)(i+1)\dots(D)}(\vec{\xi}). \tag{9.6.12}$$

Orientation of coordinate system What is orientation? It is the choice of how one orders the coordinates in use, say $(x^1, x^2, \dots, x^D)$, together with the convention that $\epsilon_{12\dots D} \equiv 1$.

In 2D flat spacetime, for example, we may choose the 'right-handed' (x^1, x^2) as Cartesian coordinates, $\epsilon_{12} \equiv 1$, and obtain the infinitesimal volume $\mathrm{d}^2\vec{x} = \mathrm{d}x^1\mathrm{d}x^2$. We can switch to cylindrical coordinates

$$\vec{x}(\vec{\xi}) = r(\cos\phi, \sin\phi). \tag{9.6.13}$$

so that

$$\frac{\partial x^i}{\partial r} = (\cos\phi, \sin\phi), \qquad \frac{\partial x^i}{\partial \phi} = r(-\sin\phi, \cos\phi), \qquad r \geq 0,\ \phi \in [0, 2\pi). \tag{9.6.14}$$

If we ordered $(\xi^1, \xi^2) = (r, \phi)$, we would have

$$\epsilon_{i_1 i_2}\frac{\partial x^{i_1}}{\partial r}\frac{\partial x^{i_2}}{\partial \phi} = \det\begin{bmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{bmatrix} = r(\cos\phi)^2 + r(\sin\phi)^2 = r. \tag{9.6.15}$$

If we instead ordered $(\xi^1, \xi^2) = (\phi, r)$, we would have

$$\epsilon_{i_1 i_2}\frac{\partial x^{i_1}}{\partial \phi}\frac{\partial x^{i_2}}{\partial r} = \det\begin{bmatrix} -r\sin\phi & \cos\phi \\ r\cos\phi & \sin\phi \end{bmatrix} = -r(\sin\phi)^2 - r(\cos\phi)^2 = -r. \tag{9.6.16}$$

We can see that going from (x^1, x^2) to $(\xi^1, \xi^2) \equiv (r, \phi)$ is orientation preserving; and we should also choose $\epsilon_{r\phi} \equiv 1$.[93]

Problem 9.51. By going from Cartesian coordinates (x^1, x^2, x^3) to spherical ones,

$$\vec{x}(\vec{\xi}) = r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{9.6.17}$$

determine what is the orientation preserving ordering of the coordinates of $\vec{\xi}$, and is $\epsilon_{r\theta\phi}$ equal $+1$ or -1? $\square$

Infinitesimal volume re-visited The infinitesimal volume we encountered earlier can really be written as

$$\mathrm{d}(\text{vol.}) = \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\epsilon_{12\dots D} = \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}, \tag{9.6.18}$$

so that under a coordinate transformation $\vec{x} \to \vec{x}(\vec{\xi})$, the necessarily positive infinitesimal volume written in $\vec{x}$ transforms into another positive infinitesimal volume, but written in $\vec{\xi}$:

$$\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\epsilon_{12\dots D} = \mathrm{d}^D\vec{\xi}\sqrt{\left|g(\vec{\xi})\right|}\epsilon_{12\dots D}. \tag{9.6.19}$$

Below, we will see that $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}$ in modern integration theory is viewed as a differential $D-$form.

[93] We have gone from a 'right-handed' coordinate system (x^1, x^2) to a 'right-handed' (r, ϕ); we could also have gone from a 'left-handed' one (x^2, x^1) to a 'left-handed' (ϕ, r) and this would still be orientation-preserving.

Problem 9.52. We may consider the infinitesimal volume in 3D flat space in Cartesian coordinates

$$\mathrm{d}(\text{vol.}) = \mathrm{d}x^1 \mathrm{d}x^2 \mathrm{d}x^3. \tag{9.6.20}$$

Now, let us switch to spherical coordinates $\vec{\xi}$, with the ordering in the previous problem. Show that it is given by

$$\mathrm{d}x^1 \mathrm{d}x^2 \mathrm{d}x^3 = \mathrm{d}^3\vec{\xi}\sqrt{|g(\vec{\xi})|}, \tag{9.6.21}$$

$$\sqrt{|g(\vec{\xi})|} = \epsilon_{i_1 i_2 i_3} \frac{\partial x^{i_1}}{\partial \xi^1} \frac{\partial x^{i_2}}{\partial \xi^2} \frac{\partial x^{i_3}}{\partial \xi^3}. \tag{9.6.22}$$

Can you compare $\sqrt{|g(\vec{\xi})|}$ with the volume of the parallelepiped formed by $\partial_{\xi^1} x^i$, $\partial_{\xi^2} x^i$ and $\partial_{\xi^3} x^i$?[94] $\square$

Cross-Product in Flat 3D, Right-hand rule Notice the notion of orientation in 3D is closely tied to the "right-hand rule" in vector calculus. Let $\vec{X}$ and $\vec{Y}$ be vectors in Euclidean 3-space. In Cartesian coordinates, where $g_{ij} = \delta_{ij}$, you may check that their cross product is

$$\left(\vec{X} \times \vec{Y}\right)^k = \epsilon^{ijk} X^i Y^j. \tag{9.6.23}$$

For example, if $\vec{X}$ is parallel to the positive x^1 axis and $\vec{Y}$ parallel to the positive x^2-axis, so that $\vec{X} = |\vec{X}|(1,0,0)$ and $\vec{Y} = |\vec{Y}|(0,1,0)$, the cross product reads

$$\left(\vec{X} \times \vec{Y}\right)^k \to |\vec{X}||\vec{Y}|\epsilon^{12k} = |\vec{X}||\vec{Y}|\delta_3^k, \tag{9.6.24}$$

i.e., it is parallel to the positive x^3 axis. (Remember k cannot be either 1 or 2 because ϵ^{ijk} is fully antisymmetric.) If we had chosen $\epsilon_{123} = \epsilon^{123} \equiv -1$, then the cross product would become the "left-hand rule". Below, I will continue to point out, where appropriate, how this issue of orientation arises in differential geometry.

Problem 9.53. Show that the Levi-Civita tensor with all upper indices is given by

$$\widetilde{\epsilon}^{i_1 i_2 \dots i_D} = \frac{\text{sgn } \det(g_{ab})}{\sqrt{|g|}} \epsilon_{i_1 i_2 \dots i_D}. \tag{9.6.25}$$

In curved spaces, the sign of the $\det g_{ab} = 1$; whereas in curved spacetimes it depends on the signature used for the flat metric.[95] Hint: Raise the indices by contracting with inverse metrics, then recall the cofactor expansion definition of the determinant. $\square$

[94] Because of the existence of locally flat coordinates $\{y^i\}$, the interpretation of $\sqrt{|g(\xi)|}$ as the volume of parallelepiped formed by $\{\partial_{\xi^1} y^i, \dots, \partial_{\xi^D} y^i\}$ actually holds very generally.

[95] See eq. (9.4.16) to understand why the sign of the determinant of the metric is always determined by the sign of the determinant of its flat counterpart.

Problem 9.54. Show that the covariant derivative of the Levi-Civita tensor is zero.

$$\nabla_j \widetilde{\epsilon}_{i_1 i_2 \dots i_D} = 0. \tag{9.6.26}$$

(Hint: Start by expanding the covariant derivative in terms of Christoffel symbols; then go through some combinatoric reasoning or invoke the equivalence principle.) From this, explain why the following equalities are true; for any vector V,

$$\nabla_j \left(\widetilde{\epsilon}^{i_1 i_2 \dots i_{D-2} jk} V_k \right) = \widetilde{\epsilon}^{i_1 i_2 \dots i_{D-2} jk} \nabla_j V_k = \widetilde{\epsilon}^{i_1 i_2 \dots i_{D-2} jk} \partial_j V_k. \tag{9.6.27}$$

Why is $\nabla_i V_j - \nabla_j V_i = \partial_i V_j - \partial_j V_i$ for any V_i? Hint: expand the covariant derivatives in terms of the partial derivatives and the Christoffel symbols. □

Combinatorics This is an appropriate place to state how to actually construct a fully antisymmetric tensor from a given tensor $T_{i_1 \dots i_N}$. Denoting $\Pi(i_1 \dots i_N)$ to be a permutation of the indices $\{i_1 \dots i_N\}$, the antisymmetrization procedure is given by

$$\begin{aligned} T_{[i_1 \dots i_N]} &= \sum_{\text{permutations } \Pi \text{ of } \{i_1, i_2, \dots, i_N\}}^{N!} \sigma_\Pi \cdot T_{\Pi(i_1 \dots i_N)} \\ &= \sum_{\text{even permutations } \Pi \text{ of } \{i_1, i_2, \dots, i_N\}} T_{\Pi(i_1 \dots i_N)} - \sum_{\text{odd permutations } \Pi \text{ of } \{i_1, i_2, \dots, i_N\}} T_{\Pi(i_1 \dots i_N)}. \end{aligned} \tag{9.6.28}$$

In words: for a rank$-N$ tensor, $T_{[i_1 \dots i_N]}$ consists of a sum of $N!$ terms. The first is $T_{i_1 \dots i_N}$. Each and every other term consists of T with its indices permuted over all the $N! - 1$ distinct remaining possibilities, multiplied by $\sigma_\Pi = +1$ if it took even number of index swaps to get to the given permutation, and $\sigma_\Pi = -1$ if it took an odd number of swaps. (The σ_Π is often called the sign of the permutation Π.) For example,

$$T_{[ij]} = T_{ij} - T_{ji}, \qquad T_{[ijk]} = T_{ijk} - T_{ikj} - T_{jik} + T_{jki} + T_{kij} - T_{kji}. \tag{9.6.29}$$

Can you see why eq. (9.6.28) yields a fully antisymmetric object? Consider any pair of distinct indices, say i_a and i_b, for $1 \leq (a \neq b) \leq N$. Since the sum on its right hand side contains every permutation (multiplied by the sign) – we may group the terms in the sum of eq. (9.6.28) into pairs, say $\sigma_{\Pi_\ell} T_{j_1 \dots i_a \dots i_b \dots j_N} - \sigma_{\Pi_\ell} T_{j_1 \dots i_b \dots i_a \dots j_N}$. That is, for a given term $\sigma_{\Pi_\ell} T_{j_1 \dots i_a \dots i_b \dots j_N}$ there must be a counterpart with $i_a \leftrightarrow i_b$ swapped, multipled by a minus sign, because – if the first term involved even (odd) number of swaps to get to, then the second must have involved an odd (even) number. If we now considered swapping $i_a \leftrightarrow i_b$ in every term in the sum on the right hand side of eq. (9.6.28),

$$T_{[i_1 \dots i_a \dots i_b \dots i_N]} = \sigma_{\Pi_\ell} T_{j_1 \dots i_a \dots i_b \dots j_N} - \sigma_{\Pi_\ell} T_{j_1 \dots i_b \dots i_a \dots j_N} + \dots, \tag{9.6.30}$$

$$T_{[i_1 \dots i_b \dots i_a \dots i_N]} = - \left(\sigma_{\Pi_\ell} T_{j_1 \dots i_a \dots i_b \dots j_N} - \sigma_{\Pi_\ell} T_{j_1 \dots i_b \dots i_a \dots j_N} + \dots \right). \tag{9.6.31}$$

Problem 9.55. Given $T_{i_1 i_2 \dots i_N}$, how do we construct a fully symmetric object from it, i.e., such that swapping any two indices returns the same object? □

Problem 9.56. Explain why $T_{i_1 \dots i_N} = 0$ if it is fully anti-symmetric and $N > D$. Hint: Try constructing, say, a rank-3 fully anti-symmetric tensor in $D = 2$ dimensions; write down its components. □

Problem 9.57. If the Levi-Civita symbol is subject to the convention $\epsilon_{12\ldots D} \equiv 1$, explain why it is equivalent to the following expansion in Kronecker δs.

$$\epsilon_{i_1 i_2 \ldots i_D} = \delta^1_{[i_1}\delta^2_{i_2}\ldots\delta^{D-1}_{i_{D-1}}\delta^D_{i_D]} \tag{9.6.32}$$

This formula is intimately related to eq. (9.6.3) – can you see why? □

Problem 9.58. Argue that

$$\begin{aligned} T_{[i_1\ldots i_N]} = T_{[i_1\ldots i_{N-1}]i_N} - T_{[i_N i_2\ldots i_{N-1}]i_1} - T_{[i_1 i_N i_3\ldots i_{N-1}]i_2} \\ - T_{[i_1 i_2 i_N i_4\ldots i_{N-1}]i_3} - \cdots - T_{[i_1\ldots i_{N-2} i_N]i_{N-1}}. \end{aligned} \tag{9.6.33}$$

In fact, you should be able to argue that, on the right hand side, the index lying outside of the $[\ldots]$ can be located anywhere else, i.e., not necessarily on the rightmost slot.

In words: to construct the fully anti-symmetric combination of N indices, anti-symmetrize the first $N-1$ indices. Then swap the first and Nth index of the this first group; then swap the second and N index; etc. Note: This identity holds even if there are other indices not being anti-symmetrized. □

Product of Levi-Civita tensors The product of two Levi-Civita tensors will be important for the discussions to come. We have

$$\widetilde{\epsilon}^{i_1\ldots i_N k_1\ldots k_{D-N}}\widetilde{\epsilon}_{j_1\ldots j_N k_1\ldots k_{D-N}} = \operatorname{sgn}\det(g_{ab})\cdot A_N \delta^{i_1}_{[j_1}\ldots\delta^{i_N}_{j_N]}, \qquad 1 \leq N \leq D, \tag{9.6.34}$$

$$\widetilde{\epsilon}^{k_1\ldots k_D}\widetilde{\epsilon}_{k_1\ldots k_D} = \operatorname{sgn}\det(g_{ab})\cdot A_0, \qquad A_{N\geq 0} \equiv (D-N)!. \tag{9.6.35}$$

(Remember $0! = 1! = 1$; also, $\delta^{i_1}_{[j_1}\ldots\delta^{i_N}_{j_N]} = \delta^{[i_1}_{j_1}\ldots\delta^{i_N]}_{j_N}$.) For instance, in $D = 2$ dimensions,

$$\widetilde{\epsilon}^{ij}\widetilde{\epsilon}_{ab} = \operatorname{sgn}\det(g_{mn})\cdot\delta^i{}_{[a}\delta^j{}_{b]}, \tag{9.6.36}$$

$$\widetilde{\epsilon}^{is}\widetilde{\epsilon}_{as} = \operatorname{sgn}\det(g_{mn})\cdot\delta^i{}_a, \tag{9.6.37}$$

$$\widetilde{\epsilon}^{sl}\widetilde{\epsilon}_{sl} = \operatorname{sgn}\det(g_{mn})\cdot 2; \tag{9.6.38}$$

and in $D = 3$ dimensions,

$$\widetilde{\epsilon}^{ijk}\widetilde{\epsilon}_{abc} = \operatorname{sgn}\det(g_{mn})\cdot\delta^i{}_{[a}\delta^j{}_b\delta^k{}_{c]}, \tag{9.6.39}$$

$$\widetilde{\epsilon}^{ijs}\widetilde{\epsilon}_{abs} = \operatorname{sgn}\det(g_{mn})\cdot\delta^i{}_{[a}\delta^j{}_{b]}, \tag{9.6.40}$$

$$\widetilde{\epsilon}^{isl}\widetilde{\epsilon}_{asl} = \operatorname{sgn}\det(g_{mn})\cdot 2\delta^i{}_a, \tag{9.6.41}$$

$$\widetilde{\epsilon}^{slm}\widetilde{\epsilon}_{slm} = \operatorname{sgn}\det(g_{mn})\cdot 6; \tag{9.6.42}$$

etc.

Proof Let us first understand why there are a bunch of Kronecker deltas on the right hand side of eq. (9.6.33), starting from the $N = D$ case – where no indices are contracted.

$$\operatorname{sgn}\det(g_{ab})\widetilde{\epsilon}^{i_1\ldots i_D}\widetilde{\epsilon}_{j_1\ldots j_D} = \epsilon_{i_1\ldots i_D}\epsilon_{j_1\ldots j_D} = \delta^{i_1}_{[j_1}\ldots\delta^{i_D}_{j_D]} \tag{9.6.43}$$

(This means $A_D = 1$.) The first equality follows from eq. (9.6.25). The second may seem a bit surprising, because the indices $\{i_1, \dots, i_D\}$ are attached to a completely different $\widetilde{\epsilon}$ tensor from the $\{j_1, \dots, j_D\}$. However, if we manipulate

$$\delta^{i_1}_{[j_1} \dots \delta^{i_D}_{j_D]} = \delta^{i_1}_{[1} \dots \delta^{i_D}_{D]} \sigma_j = \delta^{1}_{[1} \dots \delta^{D}_{D]} \sigma_i \sigma_j = \sigma_i \sigma_j = \epsilon_{i_1 \dots i_D} \epsilon_{j_1 \dots j_D}, \tag{9.6.44}$$

where $\sigma_i = 1$ if it took even number of swaps to re-arrange $\{i_1, \dots, i_D\}$ to $\{1, \dots, D\}$ and $\sigma_i = -1$ if it took odd number of swaps; similarly, $\sigma_j = 1$ if it took even number of swaps to re-arrange $\{j_1, \dots, j_D\}$ to $\{1, \dots, D\}$ and $\sigma_j = -1$ if it took odd number of swaps. But σ_i is precisely the Levi-Civita *symbol* $\epsilon_{i_1 \dots i_D}$ and likewise $\sigma_j = \epsilon_{j_1 \dots j_D}$. The ($\geq 1$)-contractions between the $\widetilde{\epsilon}$s can, in principle, be obtained by contracting the right hand side of (9.6.43). Because one contraction of the $(N+1)$ Kronecker deltas have to return N Kronecker deltas, by induction, we now see why the right hand side of eq. (9.6.34) takes the form it does for any N.

What remains is to figure out the actual value of A_N. We will do so recursively, by finding a relationship between A_N and A_{N-1}. We will then calculate A_1 and use it to generate all the higher A_Ns. Starting from eq. (9.6.34), and employing eq. (9.6.33),

$$\begin{aligned}
&\widetilde{\epsilon}^{i_1 \dots i_{N-1} \sigma k_1 \dots k_{D-N}} \widetilde{\epsilon}_{j_1 \dots j_{N-1} \sigma k_1 \dots k_{D-N}} = A_N \delta^{i_1}_{[j_1} \dots \delta^{i_{N-1}}_{j_{N-1}} \delta^{\sigma}_{\sigma]} \qquad (9.6.45)\\
&= A_N \left(\delta^{i_1}_{[j_1} \dots \delta^{i_{N-1}}_{j_{N-1}]} \delta^{\sigma}_{\sigma} - \delta^{i_1}_{[\sigma} \delta^{i_2}_{j_2} \dots \delta^{i_{N-1}}_{j_{N-1}]} \delta^{\sigma}_{j_1} - \delta^{i_1}_{[j_1} \delta^{i_2}_{\sigma} \delta^{i_3}_{j_3} \dots \delta^{i_{N-1}}_{j_{N-1}]} \delta^{\sigma}_{j_2} - \dots - \delta^{i_1}_{[j_1} \dots \delta^{i_{N-2}}_{j_{N-2}} \delta^{i_{N-1}}_{\sigma]} \delta^{\sigma}_{j_{N-1}} \right) \\
&= A_N \cdot (D - (N-1)) \delta^{i_1}_{[j_1} \dots \delta^{i_{N-1}}_{j_{N-1}]} \equiv A_{N-1} \delta^{i_1}_{[j_1} \dots \delta^{i_{N-1}}_{j_{N-1}]}.
\end{aligned}$$

(The last equality is a definition, because A_{N-1} *is* the coefficient of $\delta^{i_1}_{[j_1} \dots \delta^{i_{N-1}}_{j_{N-1}]}$.) We have the relationship

$$A_N = \frac{A_{N-1}}{D - (N-1)}. \tag{9.6.46}$$

If we contract every index, we have to sum over all the $D!$ (non-zero components of the Levi-Civita symbol)^2,

$$\widetilde{\epsilon}^{i_1 \dots i_D} \widetilde{\epsilon}_{i_1 \dots i_D} = \operatorname{sgn}\det(g_{ab}) \cdot \sum_{i_1, \dots, i_D} (\epsilon_{i_1 \dots i_D})^2 = \operatorname{sgn}\det(g_{ab}) \cdot D! \tag{9.6.47}$$

That means $A_0 = D!$. If we contracted every index but one,

$$\widetilde{\epsilon}^{i k_1 \dots k_D} \widetilde{\epsilon}_{j k_1 \dots k_D} = \operatorname{sgn}\det(g_{ab}) A_1 \delta^i_j. \tag{9.6.48}$$

Contracting the i and j indices, and invoking eq. (9.6.47),

$$\operatorname{sgn}\det(g_{ab}) \cdot D! = \operatorname{sgn}\det(g_{ab}) A_1 \cdot D \qquad \Rightarrow \qquad A_1 = (D-1)!. \tag{9.6.49}$$

That means we may use A_1 (or, actually, A_0) to generate all other $A_{N \geq 0}$s,

$$\begin{aligned}
A_N &= \frac{A_{N-1}}{(D-(N-1))} = \frac{1}{D-(N-1)} \frac{A_{N-2}}{D-(N-2)} = \dots \\
&= \frac{A_1}{(D-1)(D-2)(D-3)\dots(D-(N-1))} = \frac{(D-1)!}{(D-1)(D-2)(D-3)\dots(D-(N-1))}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(D-1)(D-2)(D-3)\dots(D-(N-1))(D-N)(D-(N+1))\dots 3\cdot 2\cdot 1}{(D-1)(D-2)(D-3)\dots(D-(N-1))} \\
&= (D-N)!. \qquad (9.6.50)
\end{aligned}
$$

Note that $0! = 1$, so $A_D = 1$ as we have found earlier. **YZ: This proof can be simplified.** $\square$

Example A simple application involves the cross product of vectors $\vec{a}$ and $\vec{b}$ in flat space. In a Cartesian basis, vector calculus says

$$\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta, \qquad (9.6.51)$$

where θ is the angle between the two vectors. We now derive the result

$$|\vec{a}\times\vec{b}| = |\vec{a}||\vec{b}|\sin\theta \qquad (9.6.52)$$

by considering

$$
\begin{aligned}
|\vec{a}\times\vec{b}|^2 &= (\vec{a}\times\vec{b})\cdot(\vec{a}\times\vec{b}) = \delta_{ij}(\vec{a}\times\vec{b})^i(\vec{a}\times\vec{b})^j \qquad (9.6.53) \\
&= \epsilon^{ijk}a^jb^k\epsilon^{imn}a^mb^n. \qquad (9.6.54)
\end{aligned}
$$

Using equations (9.6.40) and (9.6.51),

$$
\begin{aligned}
(\vec{a}\times\vec{b})^2 &= (\delta^j_m\delta^k_n - \delta^j_n\delta^k_m)a^jb^ka^mb^n = \vec{a}^2\vec{b}^2 - (\vec{a}\cdot\vec{b})^2 \qquad (9.6.55) \\
&= |\vec{a}|^2|\vec{b}|^2\left(1-\cos^2\theta\right). \qquad (9.6.56)
\end{aligned}
$$

The derivation is complete once we recognize $1-\cos^2\theta = \sin^2\theta$.

Problem 9.59. Matrix determinants revisited Explain why the cofactor expansion definition of a square matrix in eq. (3.2.1) can also be expressed as

$$\det A = \epsilon^{i_1 i_2 \dots i_{D-1} i_D} A^1{}_{i_1} A^2{}_{i_2} \dots A^{D-1}{}_{i_{D-1}} A^D{}_{i_D} \qquad (9.6.57)$$

provided we define $\epsilon^{i_1 i_2 \dots i_{D-1} i_D}$ in the same way we defined its lower index counterpart, including $\epsilon^{123\dots D} \equiv 1$. That is, why can we cofactor expand about either the rows or the columns of a matrix, to obtain its determinant? What does that tell us about the relation $\det A^{\mathrm{T}} = \det A$? Can you also prove, using our result for the product of two Levi-Civita symbols, that $\det(A\cdot B) = (\det A)(\det B)$? $\square$

Problem 9.60. 3D Identities Explain why in 3D,

$$\vec{A}\cdot(\vec{B}\times\vec{C}) = \vec{B}\cdot(\vec{C}\times\vec{A}) = \vec{C}\cdot(\vec{A}\times\vec{B}). \qquad (9.6.58)$$

Next, using eq. (9.6.34), derive the 3D vector cross product identities

$$
\begin{aligned}
\vec{A}\times(\vec{B}\times\vec{C}) &= (\vec{A}\cdot\vec{C})\vec{B} - (\vec{A}\cdot\vec{B})\vec{C}, \qquad (9.6.59) \\
(\vec{A}\times\vec{B})\cdot(\vec{C}\times\vec{E}) &= (\vec{A}\cdot\vec{C})(\vec{B}\cdot\vec{E}) - (\vec{A}\cdot\vec{E})(\vec{B}\cdot\vec{C}). \qquad (9.6.60)
\end{aligned}
$$

$\square$

Hodge dual We are now ready to define the Hodge dual. Given a fully antisymmetric rank-N tensor $T_{i_1 \dots i_N}$, its Hodge dual – which I shall denote as $\widetilde{T}^{j_1 \dots j_{D-N}}$ – is a fully antisymmetric rank-$(D-N)$ tensor whose components are

$$\widetilde{T}^{j_1 \dots j_{D-N}} \equiv \frac{1}{N!} \widetilde{\epsilon}^{j_1 \dots j_{D-N} i_1 \dots i_N} T_{i_1 \dots i_N}. \tag{9.6.61}$$

Invertible Note that the Hodge dual is an invertible operation, as long as we are dealing with fully antisymmetric tensors, in that given $\widetilde{T}^{j_1 \dots j_{D-N}}$ we can recover $T_{i_1 \dots i_N}$ and vice versa.[96] All you have to do is contract both sides with the Levi-Civita tensor, namely

$$T_{i_1 \dots i_N} = \mathrm{sgn}(\det g_{ab}) \frac{(-)^{N(D-N)}}{(D-N)!} \widetilde{\epsilon}_{i_1 \dots i_N j_1 \dots j_{D-N}} \widetilde{T}^{j_1 \dots j_{D-N}}. \tag{9.6.62}$$

In other words $\widetilde{T}^{j_1 \dots j_{D-N}}$ and $T_{i_1 \dots i_N}$ contain the same amount of information.

Problem 9.61. Using eq. (9.6.34), verify the proportionality constant $(-)^{N(D-N)} \mathrm{sgn} g$ in the inverse Hodge dual of eq. (9.6.62), and thereby prove that the Hodge dual is indeed invertible for fully antisymmetric tensors. □

Curl Because they contain the same amount of information, the curl of a vector field A_i can now either be defined as the antisymmetric rank-2 tensor

$$F_{ij} \equiv \partial_{[i} A_{j]} \tag{9.6.63}$$

or its rank-$(D-2)$ Hodge dual

$$\widetilde{F}^{i_1 i_2 \dots i_{D-2}} \equiv \frac{1}{2} \widetilde{\epsilon}^{i_1 i_2 \dots i_{D-2} jk} \partial_{[j} A_{k]}. \tag{9.6.64}$$

$(D=3)$-dimensional space is a special case where both the original vector field A^i and the Hodge dual $\widetilde{F}^i$ are rank-1 tensors. This is usually how electromagnetism is taught: that in 3D the magnetic field is a vector arising from the curl of the vector potential A_i:

$$B^i = \frac{1}{2} \widetilde{\epsilon}^{ijk} \partial_{[j} A_{k]} = \widetilde{\epsilon}^{ijk} \partial_j A_k. \tag{9.6.65}$$

In particular, when we specialize to 3D flat space with Cartesian coordinates:

$$\left(\vec{\nabla} \times \vec{A}\right)^1 = \epsilon^{123} \partial_2 A_3 + \epsilon^{132} \partial_3 A_2 = \partial_2 A_3 - \partial_3 A_2, \tag{9.6.66}$$

$$\left(\vec{\nabla} \times \vec{A}\right)^2 = \epsilon^{213} \partial_1 A_3 + \epsilon^{231} \partial_3 A_1 = \partial_3 A_1 - \partial_1 A_3, \tag{9.6.67}$$

$$\left(\vec{\nabla} \times \vec{A}\right)^3 = \epsilon^{312} \partial_1 A_2 + \epsilon^{321} \partial_2 A_1 = \partial_1 A_2 - \partial_2 A_1. \tag{9.6.68}$$

---

[96] The fully antisymmetric property is crucial here: any symmetric portion of a tensor contracted with the Levi-Civita tensor would be lost. For example, an arbitrary rank-2 tensor can always be decomposed as $T_{ij} = (1/2)T_{\{ij\}} + (1/2)T_{[ij]}$; then, $\widetilde{\epsilon}^{i_1 \dots i_{D-2} jk} T_{jk} = \widetilde{\epsilon}^{i_1 \dots i_{D-2} jk}((1/2)T_{\{jk\}} + (1/2)T_{[jk]}) = (1/2)\widetilde{\epsilon}^{i_1 \dots i_{D-2} jk} T_{[jk]}$. The symmetric part is lost because $\widetilde{\epsilon}^{i_1 \dots i_{D-2} jk} T_{\{jk\}} = -\widetilde{\epsilon}^{i_1 \dots i_{D-2} kj} T_{\{kj\}}$.

This recovers the usual definition of the curl in 3D vector calculus. But you may have noticed from equations (9.6.63) and (9.6.64), in any other dimension, that the magnetic field is really not a (rank-1) vector but should be viewed either as a rank-2 curl or a rank$-(D-2)$ Hodge dual of this curl. $\square$

Exterior Derivative (Generalized Curl), Divergence We can extend the definition of a curl of a vector field to that of a rank$-N \leq D-1$ fully antisymmetric $B_{i_1 \dots i_N}$ as

$$\nabla_{[\sigma} B_{i_1 \dots i_N]} = \partial_{[\sigma} B_{i_1 \dots i_N]}. \tag{9.6.69}$$

(Can you explain why the ∇ can be replaced with ∂? Notice too, this definition of the curl does not involve the metric.) Even though I prefer to call eq. (9.6.69) the 'generalized curl of B', in differential form nomenclature, eq. (9.6.69) is instead dubbed the *exterior derivative*. More specifically, for an arbitrary fully anti-symmetric rank$-N$ tensor B,

$$(\mathrm{d}B)_{i_0 i_1 \dots i_N} \equiv \frac{1}{N!} \partial_{[i_0} B_{i_1 \dots i_N]}. \tag{9.6.70}$$

With the Levi-Civita tensor, we can convert the curl of an antisymmetric tensor into the divergence of its dual,

$$\nabla_\ell \widetilde{B}^{j_1 \dots j_{D-N-1} \ell} = \frac{1}{N!} \widetilde{\epsilon}^{j_1 \dots j_{D-N-1} \ell i_1 \dots i_N} \nabla_\ell B_{i_1 \dots i_N} \tag{9.6.71}$$

$$= \frac{1}{(N+1)!} \cdot \widetilde{\epsilon}^{j_1 \dots j_{D-N-1} \ell i_1 \dots i_N} \frac{\partial_{[\ell} B_{i_1 \dots i_N]}}{N!} \tag{9.6.72}$$

$$= \widetilde{\mathrm{d}B}^{j_1 \dots j_{D-N-1}}. \tag{9.6.73}$$

In the first equality, we have used the fact that the Levi-Civita tensor is covariantly constant; recall eq. (9.6.26). Since $\partial_{[\sigma} B_{i_1 \dots i_N]}$ and its Hodge dual contains the same information, we may proceed to identify the two objects,

$$\nabla_\ell \widetilde{B}^{j_1 \dots j_{D-N-1} \ell} \qquad \leftrightarrow \qquad \partial_{[\ell} B_{i_1 \dots i_N]}. \tag{9.6.74}$$

For example, in 3D, the magnetic field can be viewed as not the curl of A_i but rather as the following divergence of its dual:

$$\nabla_j \widetilde{A}^{ij} = \widetilde{\epsilon}^{ijk} \nabla_j A_k = B^i. \tag{9.6.75}$$

The divergence of the dual of A_i *is* the (negative) curl of A_i.

Let us take the anti-symmetric derivative of $F_{ij} \equiv \partial_{[i} A_{j]}$.

$$\begin{aligned} \partial_{[i} F_{jk]} &= \partial_{[i} \partial_{[j} A_{k]]} = 2 \partial_{[i} \partial_j A_{k]} \\ &= \partial_{[i} \partial_j A_{k]} - \partial_{[j} \partial_i A_{k]} = 0. \end{aligned} \tag{9.6.76}$$

That is, the curl of F_{ij} is zero because it involves the difference between the same pair of partial derivatives, for e.g., $\partial_i \partial_j$ and $\partial_j \partial_i$. Likewise, if we take the fully anti-symmetric derivative of the 1$-$form $v_i \equiv \partial_i \varphi$,

$$\begin{aligned} \partial_{[i} v_{j]} &= \partial_{[i} \partial_{j]} \varphi \\ &= (\partial_i \partial_j - \partial_j \partial_i) \varphi = 0. \end{aligned} \tag{9.6.77}$$

Problem 9.62. Lie derivative commutes with d Proof that, when acting upon an $N-$form – i.e., a fully anti-symmetric $\omega_{i_1\dots i_N}$ – the Lie derivative along some vector ξ^i commutes with the exterior derivative:

$$\mathrm{d}\pounds_{\vec{\xi}}\omega = \pounds_{\vec{\xi}}\mathrm{d}\omega. \tag{9.6.78}$$

Hint: Work out both sides – a $N+1$ form – in index notation. □

Problem 9.63. In 3D vector calculus, we learn that the divergence of a curl is zero

$$\vec{\nabla}\cdot\left(\vec{\nabla}\times\vec{A}\right) = 0; \tag{9.6.79}$$

and the curl of a gradient is zero

$$\vec{\nabla}\times\vec{\nabla}\varphi = 0. \tag{9.6.80}$$

In 3D (curved) space with arbitrary coordinates, verify that equations (9.6.76) and (9.6.77) are simply the Hodge dual versions of equations (9.6.79) and (9.6.80). □

Problem 9.64. Generalized 3D Vector Calculus Identities Prove the following $D=3$ identities in arbitrary curved spaces:

$$\left(\vec{\nabla}\times\left(\vec{\nabla}\times\vec{A}\right)\right)^i = \nabla^i\left(\vec{\nabla}\cdot\vec{A}\right) - \vec{\nabla}^2 A^i + R^i{}_j A^j, \tag{9.6.81}$$

$$\vec{\nabla}\times\left(\varphi\cdot\vec{A}\right) = \vec{\nabla}\varphi\times\vec{A} + \varphi\vec{\nabla}\times\vec{A}, \tag{9.6.82}$$

$$\vec{\nabla}\left(\vec{A}\cdot\vec{B}\right) = \vec{\nabla}_{\vec{A}}\vec{B} + \vec{\nabla}_{\vec{B}}\vec{A} + \vec{A}\times\left(\vec{\nabla}\times\vec{B}\right) + \vec{B}\times\left(\vec{\nabla}\times\vec{A}\right), \tag{9.6.83}$$

$$\vec{\nabla}\cdot\left(\vec{A}\times\vec{B}\right) = \vec{B}\cdot\left(\vec{\nabla}\times\vec{A}\right) - \vec{A}\cdot\left(\vec{\nabla}\times\vec{B}\right), \tag{9.6.84}$$

$$\vec{\nabla}\times\left(\vec{A}\times\vec{B}\right) = \vec{A}\left(\vec{\nabla}\cdot\vec{B}\right) - \vec{B}\left(\vec{\nabla}\cdot\vec{A}\right) + \vec{\nabla}_{\vec{B}}\vec{A} - \vec{\nabla}_{\vec{A}}\vec{B}, \tag{9.6.85}$$

where $R^i{}_j$ is the Ricci tensor, φ is an arbitrary scalar field, and the covariant derivative along some vector $\vec{v}$ is denoted as $\vec{\nabla}_{\vec{v}} \equiv v^a\nabla_a$. □

Problem 9.65. Show, by contracting both sides of eq. (9.6.65) with an appropriate $\widetilde{\epsilon}$-tensor, that

$$\widetilde{\epsilon}_{ijk}B^k = \partial_{[i}A_{j]}. \tag{9.6.86}$$

Assume $\operatorname{sgn}\det(g_{ab}) = 1$. □

Problem 9.66. Explain why, any fully anti-symmetric tensor rank$-D$ tensor in D-dimensional space (say, F) must be proportional to the Levi-Civita tensor:

$$F_{i_1\dots i_D} = \varphi(\vec{x})\widetilde{\epsilon}_{i_1\dots i_D}. \tag{9.6.87}$$

Hint: How many independent components of $F_{i_1\dots i_D}$ are there?

Is the Hodge dual of a rank-D fully antisymmetric tensor $F_{i_1\dots i_D}$ invertible?

In electromagnetism, if the magnetic field is always defined as the Hodge dual of $\partial_{[i}A_{j]}$, what rank tensor is it in 2 spatial dimensions? Also explain why $\partial_{[i}A_{j]}$ must be zero in 1 spatial dimension – i.e., only electric fields can exist in 1 space dimensions. □

Problem 9.67. All 2D Metrics Are (Locally) Conformally Flat[97] A metric g_{ij} is said to be conformally flat if it is equal to the flat metric $\bar{g}_{ij}$ multiplied by a scalar function which we shall denote as Ω^2 (not to be confused with the solid angle):

$$g_{ij}(\vec{x}) = \Omega(\vec{x})^2 \bar{g}_{ij}. \tag{9.6.88}$$

Here in 2D, $\bar{g}_{ij} = \delta_{ij} = \text{diag}[1, 1]$ if we are working with a curved space; whereas (in the following Chapter) $\bar{g}_{ij} = \eta_{ij} = \text{diag}[1, -1]$ if we are dealing with a curved spacetime instead.

In this problem, we will prove that:

> In a 2D curved space(time), it is always possible to find a set of local coordinates such that the metric takes the conformally flat form in eq. (9.6.88).

Recall that coordinates, in this case $\vec{x}$, may be regarded as scalar fields in the curved space(time). Now, if their gradients with respect to a different set of coordinates $\vec{x}'$ – i.e., the one forms $\mathrm{d}x^1 = (\partial x^1/\partial x'^m)\mathrm{d}x'^m$ and $\mathrm{d}x^2 = (\partial x^2/\partial x'^m)\mathrm{d}x'^m$ – are required to be Hodge duals of each other, namely,

$$\frac{\partial x^1}{\partial x'^m} = \widetilde{\epsilon}_{m'}{}^{n'} \frac{\partial x^2}{\partial x'^n}, \tag{9.6.89}$$

show that

$$\frac{\partial x^1}{\partial x'^m}\frac{\partial x^2}{\partial x'^n} g^{m'n'}(\vec{x}') = 0, \tag{9.6.90}$$

and

$$\frac{\partial x^1}{\partial x'^m}\frac{\partial x^1}{\partial x'^n} g^{m'n'} = (\text{sgn}\det g)\,\frac{\partial x^2}{\partial x'^m}\frac{\partial x^2}{\partial x'^n} g^{m'n'}. \tag{9.6.91}$$

Suppose we begin with the metric $g_{i'j'}(\vec{x}')\mathrm{d}x'^i\mathrm{d}x'^j$. Explain why the above results demonstrate that $g^{ij}(\vec{x})\partial_{x^i} \otimes \partial_{x^j}$ is conformally flat.

Laplacian Verify that the scalar Laplacian in such a conformally flat coordinate system in 2D is proportional to the flat Cartesian one:

$$\vec{\nabla}^2\varphi = \Omega^{-2}\left(\partial_{x^1}^2 + \partial_{x^2}^2\right)\varphi. \tag{9.6.92}$$

Homogeneous solutions Furthermore, show that these $\vec{x}(\vec{x}')$ coordinates satisfying eq. (9.6.89) and its Hodge dual must be homogeneous solutions to the Laplace equation with respect to $\vec{x}'$:

$$\vec{\nabla}^2_{\vec{x}'} x^1(\vec{x}') = 0 = \vec{\nabla}^2_{\vec{x}'} x^2(\vec{x}'). \tag{9.6.93}$$

If x^1 and x^2 are good coordinates, then $\mathrm{d}x^1$ and $\mathrm{d}x^2$ obeying eq. (9.6.89) and its Hodge dual must be the gradient of some scalar function(s), which in turn means they must be curl-free. Check that it is indeed the case, as long as eq. (9.6.93) (Laplace's equation) is satisfied:

$$\nabla_{[a'}(\mathrm{d}x^1)_{b']} = 0 = \nabla_{[a'}(\mathrm{d}x^2)_{b']}. \tag{9.6.94}$$

[97]This problem is based on appendix 11C of [26].

Remarks This investigation reveals, given some definition $x^2 = x^2(\vec{x}')$ obeying $\vec{\nabla}_{\vec{x}'} x^2 = 0$, eq. (9.6.89) defines up to an additive constant

$$x^1(\vec{x}') = \int \widetilde{\mathrm{d}x^2} = \int \widetilde{\epsilon}_{m'}{}^{n'} \frac{\partial x^2}{\partial x'^n} \mathrm{d}x'^m. \tag{9.6.95}$$

(Suppose $x^1(\vec{x}')$ were given instead – can you write down the analogous integral representation of x^2?) Hence, this problem provides a constructive proof for the existence of 2D conformally flat (and orthogonal) coordinate systems. □

Problem 9.68. The 2-Sphere is Conformally Flat[98] This problem follows up on Problem (9.67) by showing that the 2-sphere metric

$$\mathrm{d}\ell^2 = \mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2 \tag{9.6.96}$$

may be put into the conformally flat forms

$$\begin{aligned}
\mathrm{d}\ell^2 &= \left(\frac{2}{1+\chi^2}\right)^2 \left((\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2\right) && (9.6.97)\\
&= \left(\frac{2}{1+z\bar{z}}\right)^2 \mathrm{d}z\mathrm{d}\bar{z} && (9.6.98)\\
&= \left(\frac{2}{1+\chi^2}\right)^2 \left((\mathrm{d}\chi)^2 + \chi^2(\mathrm{d}\phi)^2\right), && (9.6.99)\\
\chi &\equiv \sqrt{(x^1)^2 + (x^2)^2}; && (9.6.100)
\end{aligned}$$

where

$$\begin{aligned}
(x^1, x^2) &\equiv (\tan(\theta/2)\cos(\phi), \tan(\theta/2)\sin(\phi)), && (9.6.101)\\
z &\equiv x^1 + ix^2. && (9.6.102)
\end{aligned}$$

Begin by showing that the $\vec{x}$ satisfy the (scalar) Laplace's equation in the (θ, ϕ) system:

$$\vec{\nabla}^2_{\theta,\phi} x^1 = 0 = \vec{\nabla}^2_{\theta,\phi} x^2. \tag{9.6.103}$$

Next, show that eq. (9.6.89) is in fact satisfied; i.e., that $\mathrm{d}x^1$ is the Hodge dual of $\mathrm{d}x^2$. Perform the explicit coordinate transformations to verify $\mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2$ does become equations (9.6.97) through (9.6.99). The transformation from eq. (9.6.97) to eq. (9.6.99) is of course the usual Cartesian (x^1, x^2) to 2D polar (χ, ϕ) coordinates, with χ being the 'radial' coordinate. □

Problem 9.69. Parallel Propagators on 2-Sphere In this problem we are going to solve for a pair of basis unit length vector fields $\{\varepsilon_{\widehat{1}}{}^i, \varepsilon_{\widehat{2}}{}^i\}$ on the unit 2-sphere $\mathrm{d}\ell^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$ that are also parallel transported along the geodesic joining two arbitrary locations $\widehat{r}$ and $\widehat{r}'$. From a 3D flat space perspective, these locations may be viewed as unit vectors

$$\widehat{r} \equiv (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{9.6.104}$$

[98] In fact, the round sphere in any dimension 2 or greater is conformally flat.

$$\widehat{r}' \equiv (\sin\theta'\cos\phi', \sin\theta'\sin\phi', \cos\theta') \,. \tag{9.6.105}$$

A key ingredient is the world function $\sigma = (1/2)(\arccos \widehat{r} \cdot \widehat{r}')^2$ derived in eq. (9.5.89); it is half the square of the geodesic distance between $\widehat{r}$ and $\widehat{r}'$ lying on the unit $2-$sphere. First show that it obeys

$$\nabla_a \sigma \nabla^a \sigma = 2\sigma \tag{9.6.106}$$

and – denoting $k^a \equiv \nabla^a \sigma$ – the non-affinely parametrized geodesic equation

$$k^a \nabla_a k^b = k^b. \tag{9.6.107}$$

Explain why, therefore,

$$\varepsilon_{\widehat{1}}{}^i \equiv \frac{\nabla^i \sigma}{\sqrt{2\sigma}} = \nabla^i \sqrt{\sigma} \tag{9.6.108}$$

is not only unit length, it is tangent to the geodesic path joining $\widehat{r}$ and $\widehat{r}'$; namely, verify $\varepsilon_{\widehat{1}}{}^a \nabla_a \varepsilon_{\widehat{1}}{}^i = 0$. Next, define the second vector field to be its Hodge dual:

$$\varepsilon_{\widehat{2}}{}^i \equiv \frac{\widetilde{\epsilon}^{ij} \nabla_j \sigma}{\sqrt{2\sigma}} = \widetilde{\epsilon}^{ij} \nabla_j \sqrt{\sigma}. \tag{9.6.109}$$

Verify that it is not only unit length, it is parallel transported along the geodesic,

$$\varepsilon_{\widehat{1}}{}^a \nabla_a \varepsilon_{\widehat{2}}{}^i = 0. \tag{9.6.110}$$

Finally, demonstrate that they are complete by recovering the metric

$$\varepsilon_{\widehat{1}a} \varepsilon_{\widehat{1}b} + \varepsilon_{\widehat{2}a} \varepsilon_{\widehat{2}b} = g_{ab} \tag{9.6.111}$$

and the inverse metric

$$\varepsilon_{\widehat{1}}{}^a \varepsilon_{\widehat{1}}{}^b + \varepsilon_{\widehat{2}}{}^a \varepsilon_{\widehat{2}}{}^b = g^{ab}. \tag{9.6.112}$$

That $\varepsilon_{\widehat{1}}$ is a pure gradient and $\varepsilon_{\widehat{2}}$ is the Hodge dual of one is not a coincidence; all vector fields on the $2-$sphere may be expressed as a superposition of these two forms – see §(9.9) below. □

Problem 9.70. Curl, divergence, and all that The electromagnetism textbook by J.D.Jackson contains on its very last page explicit forms of the gradient and Laplacian of a scalar as well as divergence and curl of a vector – in Cartesian, cylindrical, and spherical coordinates in 3-dimensional flat space. Can you derive them with differential geometric techniques? Note that the vectors there are expressed in an orthonormal basis.

Cartesian coordinates In Cartesian coordinates $\{x^1, x^2, x^3\} \in \mathbb{R}^3$, we have the metric

$$\mathrm{d}\ell^2 = \delta_{ij} \mathrm{d}x^i \mathrm{d}x^j. \tag{9.6.113}$$

Show that the gradient of a scalar ψ is

$$\vec{\nabla}\psi = (\partial_1 \psi, \partial_2 \psi, \partial_3 \psi) = (\partial^1 \psi, \partial^2 \psi, \partial^3 \psi); \tag{9.6.114}$$

the Laplacian of a scalar ψ is

$$\nabla_i \nabla^i \psi = \delta^{ij} \partial_i \partial_j \psi = \left(\partial_1^2 + \partial_2^2 + \partial_3^2\right) \psi; \tag{9.6.115}$$

the divergence of a vector A is

$$\nabla_i A^i = \partial_i A^i; \tag{9.6.116}$$

and the curl of a vector A is

$$(\vec{\nabla} \times \vec{A})^i = \epsilon^{ijk} \partial_j A_k. \tag{9.6.117}$$

Cylindrical coordinates In cylindrical coordinates $\{\rho \geq 0, 0 \leq \phi < 2\pi, z \in \mathbb{R}\}$, employ the following parametrization for the Cartesian components of the 3D Euclidean coordinate vector

$$\vec{x} = (\rho \cos\phi, \rho \sin\phi, z) \tag{9.6.118}$$

to argue that the flat metric is translated from $g_{ij} = \delta_{ij}$ to

$$\mathrm{d}\ell^2 = \mathrm{d}\rho^2 + \rho^2 \mathrm{d}\phi^2 + \mathrm{d}z^2. \tag{9.6.119}$$

Show that the gradient of a scalar ψ is

$$\nabla^{\widehat{\rho}} \psi = \partial_\rho \psi, \qquad \nabla^{\widehat{\phi}} \psi = \frac{1}{\rho} \partial_\phi \psi, \qquad \nabla^{\widehat{z}} \psi = \partial_z \psi; \tag{9.6.120}$$

the Laplacian of a scalar ψ is

$$\nabla_i \nabla^i \psi = \frac{1}{\rho} \partial_\rho \left(\rho \partial_\rho \psi\right) + \frac{1}{\rho^2} \partial_\phi^2 \psi + \partial_z^2 \psi; \tag{9.6.121}$$

the divergence of a vector A is

$$\nabla_i A^i = \frac{1}{\rho} \left(\partial_\rho \left(\rho A^{\widehat{\rho}}\right) + \partial_\phi A^{\widehat{\phi}}\right) + \partial_z A^{\widehat{z}}; \tag{9.6.122}$$

and the curl of a vector A is

$$\begin{aligned} \widetilde{\epsilon}^{\widehat{\rho} jk} \partial_j A_k = \frac{1}{\rho} \partial_\phi A^{\widehat{z}} - \partial_z A^{\widehat{\phi}}, \qquad \widetilde{\epsilon}^{\widehat{\phi} jk} \partial_j A_k = \partial_z A^{\widehat{\rho}} - \partial_\rho A^{\widehat{z}}, \\ \widetilde{\epsilon}^{\widehat{z} jk} \partial_j A_k = \frac{1}{\rho} \left(\partial_\rho \left(\rho A^{\widehat{\phi}}\right) - \partial_\phi A^{\widehat{\rho}}\right). \end{aligned} \tag{9.6.123}$$

Spherical coordinates In spherical coordinates $\{r \geq 0, 0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi\}$ the Cartesian components of the 3D Euclidean coordinate vector reads

$$\vec{x} = (r \sin(\theta) \cos(\phi), r \sin(\theta) \sin(\phi), r \cos(\theta)). \tag{9.6.124}$$

Show that the flat metric is now

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2\right); \tag{9.6.125}$$

the gradient of a scalar ψ is

$$\nabla^{\widehat{r}}\psi = \partial_r\psi, \qquad \nabla^{\widehat{\theta}}\psi = \frac{1}{r}\partial_\theta\psi, \qquad \nabla^{\widehat{\phi}}\psi = \frac{1}{r\sin\theta}\partial_\phi\psi; \tag{9.6.126}$$

the Laplacian of a scalar ψ is

$$\nabla_i\nabla^i\psi = \frac{1}{r^2}\partial_r\left(r^2\partial_r\psi\right) + \frac{1}{r^2\sin\theta}\partial_\theta\left(\sin\theta\cdot\partial_\theta\psi\right) + \frac{1}{r^2(\sin\theta)^2}\partial_\phi^2\psi; \tag{9.6.127}$$

the divergence of a vector A reads

$$\nabla_i A^i = \frac{1}{r^2}\partial_r\left(r^2 A^{\widehat{r}}\right) + \frac{1}{r\sin\theta}\partial_\theta\left(\sin\theta\cdot A^{\widehat{\theta}}\right) + \frac{1}{r\sin\theta}\partial_\phi A^{\widehat{\phi}}; \tag{9.6.128}$$

and the curl of a vector A is given by

$$\widetilde{\epsilon}^{\widehat{r}jk}\partial_j A_k = \frac{1}{r\sin\theta}\left(\partial_\theta(\sin\theta\cdot A^{\widehat{\phi}}) - \partial_\phi A^{\widehat{\theta}}\right), \qquad \widetilde{\epsilon}^{\widehat{\theta}jk}\partial_j A_k = \frac{1}{r\sin\theta}\partial_\phi A^{\widehat{r}} - \frac{1}{r}\partial_r(rA^{\widehat{\phi}}),$$
$$\widetilde{\epsilon}^{\widehat{\phi}jk}\partial_j A_k = \frac{1}{r}\left(\partial_r\left(rA^{\widehat{\theta}}\right) - \partial_\theta A^{\widehat{r}}\right). \tag{9.6.129}$$

□

Problem 9.71. Additional Orthogonal Coordinates Verify the following forms of the metric in flat 3D space, starting from Cartesian coordinates $\mathrm{d}\ell^2 = \mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}$, where $\vec{x}$ are the Cartesian components of the coordinate vector.

Elliptic Cylindrical Coordinates If we choose some fixed length scale $R > 0$ and

$$\vec{x} = \left(R\cosh(\xi)\cos(\phi), R\sinh(\xi)\sin(\phi), z\right); \tag{9.6.130}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = R^2\left(\cosh^2(\xi) - \cos^2(\phi)\right)\left(\mathrm{d}\xi^2 + \mathrm{d}\phi^2\right) + \mathrm{d}z^2. \tag{9.6.131}$$

Parabolic Coordinates If

$$\vec{x} = \left(\frac{a^2 - b^2}{2}, a\cdot b, z\right); \tag{9.6.132}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = \left(a^2 + b^2\right)\left(\mathrm{d}a^2 + \mathrm{d}b^2\right) + \mathrm{d}z^2. \tag{9.6.133}$$

Parabolic Cylindrical Coordinates If

$$\vec{x} = \left(a\cdot b\cdot\cos\phi, a\cdot b\cdot\sin\phi, \frac{a^2 - b^2}{2}\right); \tag{9.6.134}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = \left(a^2 + b^2\right)\left(\mathrm{d}a^2 + \mathrm{d}b^2\right) + (a\cdot b)^2\mathrm{d}\phi^2. \tag{9.6.135}$$

Prolate Spheroidal Coordinates If we choose some fixed length scale $R > 0$ and

$$\vec{x} = R\left(\sinh(\xi)\sin(\theta)\cos(\phi), \sinh(\xi)\sin(\theta)\sin(\phi), \cosh(\xi)\cos(\theta)\right); \tag{9.6.136}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = R^2\left\{\left(\cosh(\xi)^2 - \cos(\theta)^2\right)\left(\mathrm{d}\xi^2 + \mathrm{d}\theta^2\right) + \sin(\theta)^2\sinh(\xi)^2\mathrm{d}\phi^2\right\}. \tag{9.6.137}$$

Oblate Spheroidal Coordinates If we choose some fixed length scale $R > 0$ and

$$\vec{x} = R\left(\cosh(\xi)\sin(\theta)\cos(\phi), \cosh(\xi)\sin(\theta)\sin(\phi), \sinh(\xi)\cos(\theta)\right); \tag{9.6.138}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = R^2\left\{\left(\cosh(\xi)^2 + \cos(\theta)^2 - 1\right)\left(\mathrm{d}\xi^2 + \mathrm{d}\theta^2\right) + \sin(\theta)^2\cosh(\xi)^2\mathrm{d}\phi^2\right\}. \tag{9.6.139}$$

Bispherical Coordinates If we choose some fixed length scale $R > 0$ and

$$\vec{x} = \frac{R}{\cosh(\mu) - \cos(\theta)}\left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \sinh(\mu)\right); \tag{9.6.140}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = \left(\frac{R}{\cosh(\mu) - \cos(\theta)}\right)^2\left(\mathrm{d}\mu^2 + \mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2\right). \tag{9.6.141}$$

Toroidal Coordinates If we choose some fixed length scale $R > 0$ and

$$\vec{x} = \frac{R}{\cosh(\mu) - \cos(\theta)}\left(\sinh(\mu)\cos(\phi), \sinh(\mu)\sin(\phi), \sin(\theta)\right); \tag{9.6.142}$$

the corresponding metric is

$$\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} = \left(\frac{R}{\cosh(\mu) - \cos(\theta)}\right)^2\left(\mathrm{d}\mu^2 + \mathrm{d}\theta^2 + \sinh(\mu)^2\mathrm{d}\phi^2\right). \tag{9.6.143}$$

Conical, Ellipsoidal, and Paraboloidal Coordinates These involve Jacobi elliptic functions cn, sn and dn; see the end of Chapter 5 of Volume 1 of Morse and Feshbach [13].

Curl, divergence, and all that Next, carry out a similar analysis as in Problem (9.70), by computing in the above coordinate systems the gradient and Laplacian on a scalar; and divergence and curl of a vector. Express your answers in the orthonormal basis. □

Problem 9.72. Translation operator in infinite curved space. When discussing the translation operator in, say eq. (5.2.19), we were implicitly assuming that space was flat and translation invariant. In curved space, we could still define a vector space spanned by the position eigenkets $\{|\vec{x}\rangle\}$, where $\vec{x}$ refers to a particular point in space. We also need to define an inner product $\langle\vec{x}|\,\vec{x}'\rangle$; for it to be generally covariant we require that is a coordinate scalar,

$$\langle\vec{x}|\,\vec{x}'\rangle = \frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{|g(\vec{x})g(\vec{x}')|}}. \tag{9.6.144}$$

Argue that any state $|f\rangle$ can now be expressed through the superposition

$$|f\rangle = \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \, |\vec{x}'\rangle \langle \vec{x}' | f\rangle ; \tag{9.6.145}$$

and the completeness relation is therefore

$$\mathbb{I} = \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \, |\vec{x}'\rangle \langle \vec{x}' | . \tag{9.6.146}$$

One way to do so is to apply $\langle \vec{x}|$ on the left from both sides and recover $f(\vec{x}) \equiv \langle \vec{x}| f\rangle$. Next, show that the translation operator in this curved infinite-space context is

$$\mathcal{T}(\vec{\xi}) = \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \left| \vec{x}' + \vec{\xi} \right\rangle \langle \vec{x}' | . \tag{9.6.147}$$

Is this operator unitary? Comment on how translation non-invariance plays a role in the answer to this question. Can you construct the ket-bra operator representation (analogous to eq. (9.6.147)) for the inverse of $\mathcal{T}(\vec{d})$? What happens when $\vec{\xi}$ in eq. (9.6.147) is infinitesimal and satisfies Killing's equation (cf. eq. (11.5.5))? Specifically, show that

$$\mathcal{T}(\vec{\xi})^\dagger \mathcal{T}(\vec{\xi}) = \mathbb{I} + \mathcal{O}\left(\vec{\xi}^2\right) = \mathcal{T}(\vec{\xi}) \mathcal{T}(\vec{\xi})^\dagger ; \tag{9.6.148}$$

provided we identify the $\vec{x}' + \vec{\xi}(\vec{x}')$ as an infinitesimal change-of-coordinates $\vec{x}' \to \vec{x}' + \vec{\xi}(\vec{x}')$. To sum: if $\vec{\xi}$ is a Killing vector, the translation operator acting along $\vec{\xi}$ is unitary up to first order in the infinitesimal displacement. $\square$

9.7 Hypersurfaces

9.7.1 Induced Metrics

There are many physical and mathematical problems where we wish to study some $(N < D)$-dimensional (hyper)surface residing (aka embedded) in a D dimensional ambient space. One way to describe this surface is to first endow it with N coordinates $\{\xi^{\mathrm{I}} | \mathrm{I} = 1, 2, \dots, N\}$, whose indices we will denote with capital letters to distinguish from the D coordinates $\{x^i\}$ parametrizing the ambient space. Then the position of the point $\vec{\xi}$ on this hypersurface in the ambient perspective is given by $\vec{x}(\vec{\xi})$. Distances on this hypersurface can be measured using the ambient metric by restricting the latter on the former, i.e.,

$$g_{ij} \mathrm{d}x^i \mathrm{d}x^j \to g_{ij}\left(\vec{x}(\vec{\xi})\right) \frac{\partial x^i(\vec{\xi})}{\partial \xi^{\mathrm{I}}} \frac{\partial x^j(\vec{\xi})}{\partial \xi^{\mathrm{J}}} \mathrm{d}\xi^{\mathrm{I}} \mathrm{d}\xi^{\mathrm{J}} \equiv H_{\mathrm{IJ}}(\vec{\xi}) \mathrm{d}\xi^{\mathrm{I}} \mathrm{d}\xi^{\mathrm{J}} . \tag{9.7.1}$$

The H_{IJ} is the (induced) metric on the hypersurface.[99]

[99] The Lorentzian signature of curved space*time*s, as opposed to the Euclidean one in curved spaces, complicates the study of hypersurfaces in the former. One has to distinguish between timelike, spacelike and null surfaces. For a pedagogical discussion see Eric Poisson's *A Relativist's Toolkit* – in fact, much of the material in this section is heavily based on its Chapter 3. Note, however, it is not necessary to know General Relativity to study hypersurfaces in curved spacetimes.

Observe that the N vectors

$$\left\{ E^i{}_{\rm A} \partial_{x^i} \equiv \frac{\partial x^i}{\partial \xi^{\rm A}} \partial_i \middle| {\rm A} = 1, 2, \dots, N \right\}, \tag{9.7.2}$$

are tangent to this hypersurface. They form a basis set of tangent vectors at a given point $\vec{x}(\vec{\xi})$, but from the ambient D-dimensional perspective. On the other hand, the $\partial/\partial\xi^{\rm I}$ themselves form a basis set of tangent vectors, from the perspective of an observer confined to live on this hypersurface. Altogether, the tangent vector $E^i{}_{\rm A}$ is a rank-1 tensor (with one upper index i) under ambient space coordinate transformations $\vec{x} = \vec{x}(\vec{x}')$; namely,

$$\frac{\partial x'^j}{\partial x^i} E^i{}_{\rm A} \left(\vec{x}(\vec{\xi}) = \vec{x}(\vec{x}'(\vec{\xi})) \right) = E^{j'}{}_{\rm A} \left(\vec{x}'(\vec{\xi}) \right). \tag{9.7.3}$$

It is also a rank-1 tensor (with one lower index A) under transformations of the hypersurface coordinates $\vec{\xi} = \vec{\xi}(\vec{\xi}')$; namely,

$$\frac{\partial \xi^{\rm A}}{\partial \xi'^{\rm B}} E^i{}_{\rm A} \left(\vec{x}(\vec{\xi}(\vec{\xi}')) \right) = E^i{}_{\rm B'} \left(\vec{x}(\vec{\xi}') \right). \tag{9.7.4}$$

Notationally speaking, eq. (9.7.1) in terms of $\{E^i{}_{\rm B}\}$ reads

$$g_{ij} E^i{}_{\rm A} E^j{}_{\rm B} = H_{\rm AB}. \tag{9.7.5}$$

Example A simple example is provided by the 2-sphere of radius R embedded in 3D flat space. We already know that it can be parametrized by two angles $\xi^{\rm I} \equiv (0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$, such that from the ambient perspective, the sphere is described by

$$x^i(\vec{\xi}) = R(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \qquad \text{(Cartesian components)}. \tag{9.7.6}$$

(Remember R is a fixed quantity here; this amounts to setting $\mathrm{d}r = 0$ in eq. (9.1.3).) The induced metric on the sphere itself, according to eq. (9.7.1), will lead us to the expected result

$$H_{\rm IJ}(\vec{\xi}) \mathrm{d}\xi^{\rm I} \mathrm{d}\xi^{\rm J} = R^2 \left(\mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2 \right). \tag{9.7.7}$$

Remark In the case of the round sphere, it is of course convenient to use the same angular coordinates on the hypersurface as the ones in the ambient space. However, it is important to keep in mind, the choice of hypersurface coordinates $\vec{\xi}$ is, in general, completely independent from that of the ambient ones $\vec{x}$.

Area of 2D surface in 3D flat space A common vector calculus problem is to give some function $f(x, y)$ of two variables, where x and y are to be interpreted as Cartesian coordinates on a flat plane; then proceed to ask what its area is for some specified domain on the (x, y)-plane. We see such a problem can be phrased as a differential geometric one. First, we view f as the z coordinate of some hypersurface embedded in 3-dimensional flat space, so that

$$X^i \equiv (x, y, z) = (x, y, f(x, y)). \tag{9.7.8}$$

The tangent vectors $(\partial X^i/\partial \xi^{\mathrm{I}})$ are

$$\frac{\partial X^i}{\partial x} = (1, 0, \partial_x f), \qquad \frac{\partial X^i}{\partial y} = (0, 1, \partial_y f). \tag{9.7.9}$$

The induced metric, according to eq. (9.7.1), is given by

$$\begin{aligned} H_{\mathrm{IJ}}(\vec{\xi})\mathrm{d}\xi^{\mathrm{I}}\mathrm{d}\xi^{\mathrm{J}} &= \delta_{ij}\left(\frac{\partial X^i}{\partial x}\frac{\partial X^j}{\partial x}(\mathrm{d}x)^2 + \frac{\partial X^i}{\partial y}\frac{\partial X^j}{\partial y}(\mathrm{d}y)^2 + 2\frac{\partial X^i}{\partial x}\frac{\partial X^j}{\partial y}\mathrm{d}x\mathrm{d}y\right), \\ H_{\mathrm{IJ}}(\vec{\xi}) &\doteq \begin{bmatrix} 1+(\partial_x f)^2 & \partial_x f \partial_y f \\ \partial_x f \partial_y f & 1+(\partial_y f)^2 \end{bmatrix}, \qquad \xi^{\mathrm{I}} \equiv (x, y), \end{aligned} \tag{9.7.10}$$

where on the second line the "$\doteq$" means it is "represented by" the matrix to its right – the first row corresponds, from left to right, to the xx, xy components; the second row yx and yy components. Recall that the infinitesimal volume (= 2D area) is given in any coordinate system $\vec{\xi}$ by $\mathrm{d}^2\xi\sqrt{\det H_{\mathrm{IJ}}(\vec{\xi})}$. That means from taking the det of eq. (9.7.10), if the domain on (x, y) is denoted as $\mathfrak{D}$, the corresponding area swept out by f is given by the 2D integral

$$\begin{aligned} \int_{\mathfrak{D}} \mathrm{d}x\mathrm{d}y\sqrt{\det H_{\mathrm{IJ}}(x,y)} &= \int_{\mathfrak{D}} \mathrm{d}x\mathrm{d}y\sqrt{(1+(\partial_x f)^2)(1+(\partial_y f)^2) - (\partial_x f \partial_y f)^2} \\ &= \int_{\mathfrak{D}} \mathrm{d}x\mathrm{d}y\sqrt{1+(\partial_x f(x,y))^2 + (\partial_y f(x,y))^2}. \end{aligned} \tag{9.7.11}$$

Normal to hypersurface Suppose the hypersurface is $(D-1)-$dimensional, sitting in a D dimensional ambient space. Then it could also be described by first identifying a scalar function of the ambient space $f(\vec{x})$ such that some constant-f (i.e., "equi-potential") surface coincides with the hypersurface,

$$f(\vec{x}) = C \equiv \text{constant}. \tag{9.7.12}$$

For example, a 2-sphere of radius R can be defined in Cartesian coordinates $\vec{x}$ as

$$f(\vec{x}) = R^2, \qquad \text{where} \qquad f(\vec{x}) = \vec{x}^2. \tag{9.7.13}$$

Given the function f, we now show that $\mathrm{d}f = 0$ can be used to define a unit normal n^i through

$$n^i \equiv \frac{\nabla^i f}{\sqrt{\nabla^j f \nabla_j f}} = \frac{g^{ik}\partial_k f}{\sqrt{g^{lm}\partial_l f \partial_m f}}. \tag{9.7.14}$$

That n^i is of unit length can be checked by a direct calculation. For n^i to be normal to the hypersurface means, when dotted into the latter's tangent vectors $\{E^i_{\ \mathrm{I}} \equiv \partial x^i/\partial \xi^{\mathrm{I}}\}$ from our previous discussion, it returns zero:

$$E^i_{\ \mathrm{I}} n_i \propto \left.\frac{\partial x^i(\vec{\xi})}{\partial \xi^{\mathrm{I}}}\partial_i f(\vec{x})\right|_{\text{on hypersurface}} = \frac{\partial}{\partial \xi^{\mathrm{I}}} f\left(\vec{x}(\vec{\xi})\right) = \partial_{\mathrm{I}} f(\vec{\xi}) = 0. \tag{9.7.15}$$

That $\partial_{\rm I} f(\vec{\xi}) = 0$ is just a re-statement that f is constant on our hypersurface $\vec{x}(\vec{\xi})$. Using n^i we can also write down the induced metric on the hypersurface as

$$H_{ij} = g_{ij} - n_i n_j. \tag{9.7.16}$$

By induced metric H_{ij} on the hypersurface of one lower dimension than that of the ambient D-space, we mean that the "dot product" of two vectors v^i and w^i, say, is

$$H_{ij} v^i w^j = g_{ij}\, {}_{\parallel}v^i\, {}_{\parallel}w^j; \tag{9.7.17}$$

where $v^i_{\parallel}$ and $w^i_{\parallel}$ are, respectively, v^i and w^i projected along the hyper-surface at hand. In words: $H_{ij} v^i w^i$ is the dot product computed using the ambient metric but with the components of v and w orthogonal to the hypersurface removed; namely,

$$H^i{}_j v^j = v^i - \left(v^j n_j\right) n^j \equiv {}_{\parallel}v^i, \tag{9.7.18}$$

$$H^i{}_j w^j = w^i - \left(w^j n_j\right) n^j \equiv {}_{\parallel}w^i. \tag{9.7.19}$$

Compare this to the Euclidean space expression

$$\vec{v}_{\perp} \equiv \vec{v} - (\vec{v} \cdot \widehat{n})\widehat{n}, \tag{9.7.20}$$

where $\vec{v}_{\perp}$ is now perpendicular to the unit vector $\widehat{n}$. Moreover, that this construction of ${}_{\parallel}v^i$ and ${}_{\parallel}w^i$ yields vectors perpendicular to n^i – i.e., ${}_{\parallel}v^i n_i = 0$ and ${}_{\parallel}w^i n_i = 0$ – is because

$$H_{ij} n^j = \left(g_{ij} - n_i n_j\right) n^j = n_i - n_i = 0. \tag{9.7.21}$$

Since

$$H^i{}_l H^l{}_j = H^i{}_j, \tag{9.7.22}$$

we may therefore verify

$$H_{ij}\, {}_{\parallel}v^i\, {}_{\parallel}w^j = H_{ij} v^i w^j = g_{ij} H^i{}_a H^j{}_b v^a w^b = g_{ij}\, {}_{\parallel}v^i\, {}_{\parallel}w^j. \tag{9.7.23}$$

The dot product between ${}_{\parallel}v^i$ and ${}_{\parallel}w^j$ in the hypersurface geometry H_{ij} is the same as that in the ambient space geometry g_{ij}.

Problem 9.73. 2-Sphere from Embedding Starting from n^i defined in eq. (9.7.14), explain why

$$n_i \mathrm{d}x^i = \frac{\mathrm{d}f}{\sqrt{\langle \mathrm{d}f | \, \mathrm{d}f \rangle}} = \frac{\partial_a f \mathrm{d}x^a}{\sqrt{g^{ij} \partial_i f \partial_j f}}. \tag{9.7.24}$$

For the 2-sphere in 3-dimensional flat space, defined by eq. (9.7.13), calculate the components of the induced metric H_{ij} in eq. (9.7.16) and compare it that in eq. (9.7.7). Hint: compute $\mathrm{d}f$ in spherical coordinates, then consider what is the $-(n_i \mathrm{d}x^i)^2$ occurring in $H_{ij} \mathrm{d}x^i \mathrm{d}x^j = g_{ij} \mathrm{d}x^i \mathrm{d}x^j - (n_i \mathrm{d}x^i)^2$? $\square$

Problem 9.74. Area of 2D surface Consider some 2-dimensional surface parametrized by $\xi^{\rm I} = (\sigma, \rho)$, whose trajectory in D-dimensional geometry g_{ij} is provided by the Cartesian coordinates $\vec{x}(\sigma, \rho)$. What is the formula analogous to eq. (9.7.11), which yields the area of this 2D surface over some domain $\mathfrak{D}$ on the (σ, ρ) plane? Hint: First ask, "what is the 2D induced metric?" Answer:

$$\text{Area} = \int_{\mathfrak{D}} \mathrm{d}\sigma \mathrm{d}\rho \sqrt{(\partial_\sigma \vec{x})^2 (\partial_\rho \vec{x})^2 - (\partial_\sigma \vec{x} \cdot \partial_\rho \vec{x})^2}, \qquad (\partial_{\rm I} \vec{x})^2 \equiv \partial_{\rm I} x^i \partial_{\rm I} x^j g_{ij}; \tag{9.7.25}$$

Eq. (9.7.25) is not too far from the Nambu-Goto action of string theory. □

Problem 9.75. Minimal Area Refer to Problem (9.74). Imagine a 2D sheet, with ambient $D-$coordinates $\vec{x}(\vec{\xi})$ (parametrizing the geometry $g_{ij}(\vec{x})$) and intrinsic 2D ones $\vec{\xi}$, held fixed at some closed loop boundary $\partial\mathfrak{D}$; for instance, a circular ring. Show that the Euler-Lagrange equations describing the minimal area spanned by this 2D surface is

$$\frac{1}{\sqrt{|H|}} \frac{\partial}{\partial \xi^{\rm A}} \left(\sqrt{|H|} H^{\rm AB} \frac{\partial x^i}{\partial \xi^{\rm B}} \right) = -H^{\rm AB} \partial_{\rm A} x^m \partial_{\rm B} x^n \Gamma^i{}_{mn}[g], \tag{9.7.26}$$

$$H_{\rm AB} \equiv \partial_{\rm A} x^m \partial_{\rm B} x^n g_{mn}(\vec{x}). \tag{9.7.27}$$

Be sure to explain how the boundary conditions play a role in this problem. □

2D Laplace Equation from Minimal Area If the ambient space is flat, we may express it in Cartesian coordinates x^i and $g_{ij} = \delta_{ij}$; the Christoffel symbols on the right hand side of eq. (9.7.26) then vanish. This leads to the generally covariant 2D Laplace equation

$$\vec{\nabla}^2 x^i = 0. \tag{9.7.28}$$

Additionally, Problem (9.67) informs us that hypersurface coordinates may be chosen such that $H_{ij}(\xi) = \Omega(\xi)^2 \delta_{ij}$. By invoking eq. (9.6.92), we may now conclude that any 2D hypersurface embedded in $D-$dimensional flat space, with extremal area, must admit a set of intrinsic coordinates $\{\xi^{\rm A}\}$ such that its Cartesian coordinates x^i obey the flat space Laplace equation

$$\left(\partial_{\xi^1}^2 + \partial_{\xi^2}^2\right) x^i = 0. \tag{9.7.29}$$

Induced Tensors, Covariant Derivative on (D–1)–Hypersurface Just as we did so for the metric, given an arbitrary tensor $T_{abc\dots}$ residing in the ambient space, we may compute the corresponding 'induced' tensor on a $(D-1)-$hypersurface by restricting its basis 1−forms $\{\mathrm{d}x^a\}$ to allow variation only on the hypersurface itself: $\mathrm{d}x^a = E^a{}_{\rm A} \mathrm{d}\xi^{\rm A} \equiv (\partial x^i / \partial \xi^{\rm A}) \mathrm{d}\xi^{\rm A}$. This amounts to the projection

$${}_{\parallel}T_{\rm ABC\,\dots} = T_{abc\dots} E^a{}_{\rm A} E^b{}_{\rm B} E^c{}_{\rm C} \dots. \tag{9.7.30}$$

Because $E^i{}_{\rm A}$ is a rank-1 vector, such a contraction produces a scalar ${}_{\parallel}T_{\rm ABC\,\dots}$ with respect to ambient-space coordinate transformations; and a tensor with respect to hypersurface coordinate transformations. Moreover, when the tensor is the metric itself, the right hand side becomes of course the induced metric $H_{\rm AB}$. To preserve covariance on the hypersurface, the ambient space indices (small English alphabets) are moved with the H_{ij}, whereas the hypersurface indices (capital English alphabets) are moved with $H_{\rm IJ}$. The same induced tensor may also be written in the basis

$$ {}_{\parallel}T_{abc\dots} = T_{pqr\dots} H^p{}_a H^q{}_b H^r{}_c \dots . \tag{9.7.31} $$

This ${}_{\parallel}T_{abc\dots}$ now transforms as a tensor under ambient space coordinate transformations; and a scalar under hypersurface coordinate transformations. Note too, the indices of $E^i{}_{\rm I}$ are moved with H_{ij} and $H_{\rm AB}$ and their upper indices counterparts; for instance,

$$ E_{i\rm A} = H_{ij} E^j{}_{\rm I} = g_{ij} E^j{}_{\rm I} \qquad \text{and} \qquad E^{i\rm A} = H^{\rm AB} E^i{}_{\rm B}. \tag{9.7.32} $$

Along similar lines

$$ H_{ij} E^i{}_{\rm I} E^j{}_{\rm J} = (g_{ij} - n_i n_j) E^i{}_{\rm I} E^j{}_{\rm J} = g_{ij} E^i{}_{\rm I} E^j{}_{\rm J} = H_{\rm IJ}; \tag{9.7.33} $$

whereas

$$ H^{\rm IJ} E^i{}_{\rm I} E^j{}_{\rm J} = g^{ij} - n^i n^j = H^{ij}. \tag{9.7.34} $$

That this latter relation is true can be seen by acting $P^i_{\ j} \equiv H^{\rm IJ} E^i{}_{\rm I} E_{j\rm J}$ on all the $\{E^a{}_{\rm A}\}$ and n^a; namely, one may readily verify

$$ P^i_{\ j} E^j{}_{\rm A} = E^i{}_{\rm A} \qquad \text{and} \qquad P^i_{\ j} n^j = 0. \tag{9.7.35} $$

This also motivates its rephrasing as a completeness relation:

$$ H^{\rm IJ} E^i{}_{\rm I} E^j{}_{\rm J} + n^i n^j = g^{ij}. \tag{9.7.36} $$

Next, we may define a covariant derivative of the induced tensor along the hypersurface

$$ D_{\rm A} V_{\rm B} \equiv E^i{}_{\rm A} E^j{}_{\rm B} \nabla_i \, {}_{\parallel}V_j. \tag{9.7.37} $$

Observe that eq. (9.7.37) is not merely the projection of the covariant derivative of the ambient space 1−form, namely $D_{\rm A} V_{\rm B} \neq E^i{}_{\rm A} E^j{}_{\rm B} \nabla_i V_j$, because V_j itself has components that do not lie along the hypersurface. We wish to re-express the right hand side in terms of $V_{\rm A} \equiv E^i{}_{\rm A} V_i = E^i{}_{\rm A} \, {}_{\parallel}V_i$, the 1−form V induced on the hypersurface:

$$ D_{\rm A} V_{\rm B} = E^i{}_{\rm A} \nabla_i \left(E^j{}_{\rm B} \, {}_{\parallel}V_j \right) - \left(E^i{}_{\rm A} \nabla_i E^j{}_{\rm B} \right) E_j{}^{\rm C} V_{\rm C}. \tag{9.7.38} $$

Since $E^j{}_{\rm B} \, {}_{\parallel}V_j = V_{\rm B}$ is now a scalar under ambient space coordinate transformations, $E^i{}_{\rm A} \nabla_i V_{\rm B} = (\partial x^i / \partial \xi^{\rm A}) \partial_{x^i} V_{\rm B} = \partial V_{\rm B} / \partial \xi^{\rm A}$ and we are left with

$$ D_{\rm A} V_{\rm B} = \partial_{\rm A} V_{\rm B} - \gamma^{\rm C}{}_{\rm AB} V_{\rm C}, \tag{9.7.39} $$

where the Christoffel symbols on the hypersurface has been identified as

$$ \gamma^{\rm C}{}_{\rm AB} = E_j{}^{\rm C} \left(E^i{}_{\rm A} \nabla_i E^j{}_{\rm B} \right) \tag{9.7.40} $$
$$ = E_j{}^{\rm C} \left(E^i{}_{\rm A} \partial_i E^j{}_{\rm B} + E^i{}_{\rm A} \Gamma^j{}_{ik} E^k{}_{\rm B} \right) \tag{9.7.41} $$

$$= H_{aj} H^{\mathrm{CK}} \frac{\partial x^a}{\partial \xi^{\mathrm{K}}} \left(\frac{\partial^2 x^j}{\partial \xi^{\mathrm{A}} \partial \xi^{\mathrm{B}}} + \Gamma^j{}_{ik}[g] \frac{\partial x^i}{\partial \xi^{\mathrm{A}}} \frac{\partial x^k}{\partial \xi^{\mathrm{B}}} \right). \tag{9.7.42}$$

Additionally, we may check that this definition of the induced covariant derivative – and hence the associated Christoffel symbols in eq. (9.7.42) – are compatible with H_{ij} itself:

$$D_{\mathrm{A}} H_{\mathrm{BC}} = E^a{}_{\mathrm{A}} E^b{}_{\mathrm{B}} E^c{}_{\mathrm{C}} \nabla_a \left(g_{bc} - n_b n_c \right) \tag{9.7.43}$$

$$= -E^a{}_{\mathrm{A}} E^b{}_{\mathrm{B}} E^c{}_{\mathrm{C}} \left(n_c \nabla_a n_b + n_b \nabla_a n_c \right) = 0. \tag{9.7.44}$$

That implies

$$\gamma^{\mathrm{C}}{}_{\mathrm{AB}} = \frac{1}{2} H^{\mathrm{CK}} \left(\partial_{\mathrm{A}} H_{\mathrm{BK}} + \partial_{\mathrm{B}} H_{\mathrm{AK}} - \partial_{\mathrm{K}} H_{\mathrm{AB}} \right). \tag{9.7.45}$$

Problem 9.76. Show that the definition of the induced covariant derivative of a 1−form would not make sense if the 1−form itself was not first projected onto the $(D-1)$−hypersurface. Specifically, show that

$$E^i{}_{\mathrm{A}} E^j{}_{\mathrm{B}} \nabla_i V_j = D_{\mathrm{A}} V_{\mathrm{B}} - \left(E^i{}_{\mathrm{A}} \nabla_i E^j{}_{\mathrm{B}} \right) n_j (n^k V_k) \tag{9.7.46}$$

$$= D_{\mathrm{A}} V_{\mathrm{B}} + \left(E^i{}_{\mathrm{A}} E^j{}_{\mathrm{B}} \nabla_i n_j \right) (n^k V_k); \tag{9.7.47}$$

where the $D_{\mathrm{A}} V_{\mathrm{B}}$ on the right hand sides is that defined in eq. (9.7.37). (Hint: The second equality comes from recognizing $E^j{}_{\mathrm{B}} n_j = 0$.) Hence, if the component of V_i perpendicular to the hypersurface were *not* projected out beforehand, the resulting 'induced' covariant derivative would still depend on it. Incidentally, the object

$$K_{\mathrm{AB}} \equiv E^i{}_{\mathrm{A}} E^j{}_{\mathrm{B}} \nabla_i n_j \qquad \Leftrightarrow \qquad K_{ab} \equiv H^i{}_a H^j{}_b \nabla_i n_j \tag{9.7.48}$$

is dubbed the *extrinsic curvature* tensor of the $(D-1)$−hypersurface, which characterizes how much 'bending' it suffers from the ambient space point of view – recall the heuristic discussion at the end of §(9.1). □

Problem 9.77. 2-Sphere Christoffels from 3D Flat Ones As an application of eq. (9.7.42), relate the Christoffel symbols of the 2−sphere in eq. (9.5.24) to those of 3D flat space in spherical coordinates described by eq. (9.5.122). (Recall: This relationship has already been revealed in Problem (9.38), albeit in a 'brute force' manner.) Also obtain the same 2-sphere $\{\gamma^{\mathrm{C}}{}_{\mathrm{AB}}\}$ from the 3D flat space metric in Cartesian basis, but with the components of the Cartesian position written in spherical coordinates – namely, $\vec{x} = r(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$. □

Lagrange Multipliers

Differential Forms and Volume Modern integration theory involves differential forms. In D−space with coordinates $\{\vec{x}\}$, one no longer writes $\int f(\vec{x})\sqrt{|g(\vec{x})|}\mathrm{d}^D\vec{x}$, for instance, but rather

$$\int f(\vec{x})\sqrt{|g(\vec{x})|}\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 \wedge \cdots \wedge \mathrm{d}x^D. \tag{9.7.49}$$

The infinitesimal volume $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}$ is now replaced with the D−form (aka volume form)

$$\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|} \equiv \sqrt{|g(\vec{x})|}\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 \wedge \cdots \wedge \mathrm{d}x^D. \tag{9.7.50}$$

More generally, whenever the following $N-$form occurs under an integral sign, we have the definition

$$\underbrace{\mathrm{d}\xi^1 \wedge \mathrm{d}\xi^2 \wedge \dots \wedge \mathrm{d}\xi^{N-1} \wedge \mathrm{d}\xi^N}_{\text{(Differential form notation)}} \equiv \underbrace{\mathrm{d}^N\vec{\xi}}_{\text{Physicists' colloquial math-speak}} . \tag{9.7.51}$$

Here $N \leq D$, where D is the dimension of the ambient space; and we have used $\{\xi^{\mathrm{I}}|\mathrm{I} = 1, \dots, N\}$ instead of $\{\vec{x}\}$ to highlight that a different set of coordinates may be employed when describing a lower dimensional hypersurface embedded in the $D-$space. This needs to be supplemented with the constraint that it is a fully antisymmetric object:

$$\mathrm{d}\xi^{\mathrm{I}_1} \wedge \mathrm{d}\xi^{\mathrm{I}_2} \wedge \dots \wedge \mathrm{d}\xi^{\mathrm{I}_{N-1}} \wedge \mathrm{d}\xi^{\mathrm{I}_N} = \epsilon_{\mathrm{I}_1\dots\mathrm{I}_N}\mathrm{d}\xi^1 \wedge \mathrm{d}\xi^2 \wedge \dots \wedge \mathrm{d}\xi^{N-1} \wedge \mathrm{d}\xi^N. \tag{9.7.52}$$

The set of indices $\{\mathrm{I}_1, \dots, \mathrm{I}_N\}$ is a permutation of $\{1, \dots, N\}$ and $\epsilon_{\mathrm{I}_1\dots\mathrm{I}_N}$ is still the fully anti-symmetric Levi-Civita with $\epsilon_{1\dots N} \equiv 1$; not to be confused with its counterpart $\epsilon_{i_1\dots i_D}$ in the ambient $D-$space. The Jacobian incurred from a change-of-variables, from $\vec{\xi}$ to say $\vec{\xi}'$, comes about through

$$\mathrm{d}^N\vec{\xi} \equiv \mathrm{d}\xi^1 \wedge \dots \wedge \mathrm{d}\xi^N = \frac{\partial \xi^1}{\partial \xi'^{i_1}} \dots \frac{\partial \xi^1}{\partial \xi'^{i_D}} \mathrm{d}\xi'^{i_1} \wedge \dots \wedge \mathrm{d}\xi'^{i_N} \tag{9.7.53}$$

$$= \epsilon_{i_1\dots i_D} \frac{\partial \xi^1}{\partial \xi'^{i_1}} \dots \frac{\partial \xi^1}{\partial \xi'^{i_N}} \mathrm{d}\xi'^1 \wedge \dots \wedge \mathrm{d}\xi'^N \tag{9.7.54}$$

$$= \left(\det \frac{\partial \vec{\xi}}{\partial \vec{\xi}'}\right) \mathrm{d}\xi'^1 \wedge \dots \wedge \mathrm{d}\xi'^N \equiv \left(\det \frac{\partial \vec{\xi}}{\partial \vec{\xi}'}\right) \mathrm{d}^N\vec{\xi}'. \tag{9.7.55}$$

For instance, the volume form of eq. (9.7.50) may now be related to the Levi-Civita (pseudo-)tensor:

$$\begin{aligned}
\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|} &= \sqrt{|g(\vec{x})|}\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \dots \wedge \mathrm{d}x^D \\
&= \frac{1}{D!}\sqrt{|g(\vec{x})|}\epsilon_{i_1\dots i_D}\mathrm{d}x^{i_1} \wedge \mathrm{d}x^{i_2} \wedge \mathrm{d}x^{i_3} \wedge \dots \wedge \mathrm{d}x^{i_D} \\
&= \frac{1}{D!}\widetilde{\epsilon}_{i_1\dots i_D}\mathrm{d}x^{i_1} \wedge \mathrm{d}x^{i_2} \wedge \mathrm{d}x^{i_3} \wedge \dots \wedge \mathrm{d}x^{i_D}.
\end{aligned} \tag{9.7.56}$$

In particular, we may first begin with a locally flat (Cartesian) coordinate system $\{\vec{y}\}$ where $\mathrm{d}^D\vec{y}$ is the infinitesimal volume, and observe that – upon a change-of-variables to an arbitrary system $\{\vec{x}\}$ –

$$\mathrm{d}y^1 \wedge \dots \wedge \mathrm{d}y^D = \left(\det \frac{\partial \vec{y}}{\partial \vec{x}}\right) \mathrm{d}x^1 \wedge \dots \wedge \mathrm{d}x^D. \tag{9.7.57}$$

We have already discussed how going from the locally flat $\vec{y}-$system to an arbitrary $\vec{x}-$system produces an associated Jacobian whose absolute value is the square root of the determinant of the metric in the latter's coordinate basis:

$$\left|\det \frac{\partial \vec{y}}{\partial \vec{x}}\right| = \sqrt{\det g_{ab}(\vec{x})}. \tag{9.7.58}$$

Hence, as least for orientation-preserving coordinate transformations, we recover

$$\mathrm{d}(\text{vol.}) = \mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^D = \sqrt{|g(\vec{x})|}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^D. \tag{9.7.59}$$

The most general $(N \leq D)-$form consists of the superposition of the basis $\{\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_N}\}$, where $\{i_1, \ldots, i_N\}$ are chosen from $\{1, \ldots, D\}$:

$$F \equiv \frac{1}{N!} F_{i_1 \ldots i_N}(\vec{x})\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_N}, \tag{9.7.60}$$

where the components of F are necessarily fully anti-symmetric. If we further choose a $N-$hypersurface described by $\vec{x}(\vec{\xi})$, where $\vec{\xi}$ is its $N-$dimensional intrinsic coordinates,

$$F = \frac{1}{N!} F_{\mathrm{I}'_1 \ldots \mathrm{I}'_N}(\vec{\xi})\mathrm{d}\xi^{\mathrm{I}_1} \wedge \cdots \wedge \mathrm{d}\xi^{\mathrm{I}_N}; \tag{9.7.61}$$

where

$$F_{\mathrm{I}'_1 \ldots \mathrm{I}'_N}\left(\vec{x}(\vec{\xi})\right) = F_{i_1 \ldots i_N}\left(\vec{x}(\vec{\xi})\right) \frac{\partial x^{i_1}}{\partial \xi^{\mathrm{I}_1}} \cdots \frac{\partial x^{i_N}}{\partial \xi^{\mathrm{I}_N}}. \tag{9.7.62}$$

Moreover, we may also write

$$F = F_{1'2'\ldots N'}(\vec{\xi})\mathrm{d}\xi^1 \wedge \cdots \wedge \mathrm{d}\xi^N \equiv F_{1'2'\ldots N'}(\vec{\xi})\mathrm{d}^N\vec{\xi}. \tag{9.7.63}$$

As a physical example, the magnetic field in electromagnetism may be viewed as a $2-$form

$$F = \frac{1}{2} F_{ij}\mathrm{d}x^i \wedge \mathrm{d}x^j; \tag{9.7.64}$$

where $F_{ij} = -F_{ji}$. In the ambient $\vec{x}$ basis, $F_{ij}(\vec{x})$ – as we shall elaborate shortly – correspond to the magnetic flux across the $(i,j)-$plane.

Problem 9.78. Forms span a vector space Verify that the superposition of rank-$(N \leq D)$ differential forms spanned by $\{(1/N!)F_{i_1 \ldots i_N}\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_N}\}$, for arbitrary but fully antisymmetric $\{F_{i_1 \ldots i_N}\}$, forms a vector space. □

More generally, why differential $(N \leq D)-$forms are fundamental to integration theory is because, their fully antisymmetric property allows them to be properly defined as the volume spanned by an $N-$parallelepiped. In a $(D \geq 2)-$dimensional flat space, you might be familiar with the statement that $N \leq D$ linearly independent vectors define a $N-$parallelepiped. Its volume, in turn, is computed through the determinant of the matrix whose columns (or rows) are these vectors. If we now consider the $(N \leq D)-$form built out of N scalar fields $\{\Phi^{\mathrm{I}}|\mathrm{I} = 1, 2, \ldots, N\}$, i.e.,

$$\mathrm{d}\Phi^1 \wedge \cdots \wedge \mathrm{d}\Phi^N, \tag{9.7.65}$$

let us see how it defines an infinitesimal $N-$volume by generalizing the notion of volume-as-determinants. In fact, these scalar fields $\{\Phi^{\mathrm{I}}\}$ can be viewed as coordinates parameterizing some $N-$dimensional sub-space of the ambient $D-$dimensional space. Defining

$$(\mathrm{d}\Phi^{\mathrm{I}})^j \equiv \partial_j \Phi^{\mathrm{I}}\mathrm{d}x^j \qquad (\text{No sum over } j), \qquad 1 \leq \mathrm{I} \leq N, \qquad 1 \leq j \leq D; \tag{9.7.66}$$

we see that the $D-$component object $(\mathrm{d}\Phi^{\mathrm{I}})^j$, for fixed I, is an infinitesimal displacement, if we choose $\vec{y}$ to be a locally flat Cartesian coordinate system – cf. (11.4.17). Starting with the $N=2$ case, we see that

$$(\mathrm{d}\Phi^1)^j \equiv (\partial_{y^1}\Phi^1 \mathrm{d}y^1, \dots, \partial_{y^D}\Phi^1 \mathrm{d}y^D)^{\mathrm{T}} \tag{9.7.67}$$

$$(\mathrm{d}\Phi^2)^j \equiv (\partial_{y^1}\Phi^2 \mathrm{d}y^1, \dots, \partial_{y^D}\Phi^2 \mathrm{d}y^D)^{\mathrm{T}} \tag{9.7.68}$$

span a 2D space; and another set of locally flat Cartesian coordinates $\vec{x}$ may thus be chosen, such that

$$(\mathrm{d}\Phi^1)^k = (\partial_{x^1}\Phi^1 \mathrm{d}x^1, \partial_{x^2}\Phi^1 \mathrm{d}x^2, \vec{0})^{\mathrm{T}} \tag{9.7.69}$$

$$(\mathrm{d}\Phi^2)^k = (\partial_{x^1}\Phi^2 \mathrm{d}x^1, \partial_{x^2}\Phi^2 \mathrm{d}x^2, \vec{0})^{\mathrm{T}}. \tag{9.7.70}$$

By considering the wedge product

$$\mathrm{d}\Phi^1 \wedge \mathrm{d}\Phi^2 = (\partial_i \Phi^1 \mathrm{d}x^i) \wedge (\partial_j \Phi^2 \mathrm{d}x^j) = \partial_i \Phi^1 \partial_j \Phi^2 \mathrm{d}x^i \wedge \mathrm{d}x^j \tag{9.7.71}$$

$$= \det \left[\begin{array}{cc} \partial_{x^1}\Phi^1 & \partial_{x^1}\Phi^2 \\ \partial_{x^2}\Phi^1 & \partial_{x^2}\Phi^2 \end{array} \right] \mathrm{d}x^1 \wedge \mathrm{d}x^2, \tag{9.7.72}$$

we see that it is in fact the 2D area spanned by the parallelogram defined by $(\mathrm{d}\Phi^1)^a$ and $(\mathrm{d}\Phi^2)^b$. Generalizing to the set

$$\{(\mathrm{d}\Phi^{\mathrm{I}})^j | 1 \leq \mathrm{I} \leq N\}, \tag{9.7.73}$$

we see that it is a collection of N infinitesimal displacements; and the wedge product $\mathrm{d}\Phi^1 \wedge \dots \wedge \mathrm{d}\Phi^N$ is simply the volume of the $N-$parallelepiped formed by them, since

$$\begin{aligned} &\mathrm{d}\Phi^1 \wedge \dots \wedge \mathrm{d}\Phi^N \\ &= \epsilon_{\mathrm{J}_1 \dots \mathrm{J}_N} \partial_{x^{\mathrm{J}_1}}\Phi^1 \dots \partial_{x^{\mathrm{J}_N}}\Phi^N \mathrm{d}x^1 \wedge \dots \wedge \mathrm{d}x^N \end{aligned} \tag{9.7.74}$$

$$= \det \left[\begin{array}{ccccc} \partial_{x^1}\Phi^1 & \partial_{x^1}\Phi^2 & \dots & \partial_{x^1}\Phi^{N-1} & \partial_{x^1}\Phi^N \\ \partial_{x^2}\Phi^1 & \partial_{x^2}\Phi^2 & \dots & \partial_{x^2}\Phi^{N-1} & \partial_{x^2}\Phi^N \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \partial_{x^{N-1}}\Phi^1 & \partial_{x^{N-1}}\Phi^2 & \dots & \partial_{x^{N-1}}\Phi^{N-1} & \partial_{x^{N-1}}\Phi^N \\ \partial_{x^N}\Phi^1 & \partial_{x^N}\Phi^2 & \dots & \partial_{x^N}\Phi^{N-1} & \partial_{x^N}\Phi^N \end{array} \right] \mathrm{d}^N \vec{x}. \tag{9.7.75}$$

Even though we worked with a locally flat Cartesian coordinate system $\{\vec{x}\}$ here, so as to aid with the volume interpretation, we may of course expand the wedge product $\mathrm{d}\Phi^1 \wedge \dots \wedge \mathrm{d}\Phi^N$ in any coordinate system we wish. What we are witnessing here is, the anti-symmetric character of the wedge product allows us to generalize the notion of the determinant. Loosely speaking, even though N may not be equal to the dimension of the ambient space, we may still compute the 'determinant' of the 'matrix' whose I^{th} column is the infinitesimal displacement $(\mathrm{d}\Phi^{\mathrm{I}})^a$, because the space perpendicular to these N displacements is automatically discarded by the wedge product – i.e., the matrix in eq. (9.7.75) can always be reduced to a $N \times N$ one.

To sum, $\mathrm{d}\Phi^1 \wedge \dots \wedge \mathrm{d}\Phi^N$ is a coordinate-invariant object to define a volume on a $N \leq D$ dimensional surface. This surface may be embedded in the ambient $D-$space, defined through the 'equi-potential' functions $\{\Phi^{\mathrm{I}} | \mathrm{I} = 1, \dots, N\}$. Or, within certain applications – magnetohydro- and fluid-dynamics, for instance – the associated area of some $2-$form $\mathrm{d}\Phi^1 \wedge \mathrm{d}\Phi^2$ or $3-$form $\mathrm{d}\Phi^1 \wedge \mathrm{d}\Phi^2 \wedge \mathrm{d}\Phi^3$ need not describe space(time) geometric volume itself but the strength of magnetic flux (field lines per area) or particle and/or mass density (substance per volume).

9.7.2 Fluxes, Gauss-Stokes' theorems, Poincaré lemma

Directed surface elements What is the analog of $\mathrm{d}(\vec{\text{Area}})$ from vector calculus? This question is important for the discussion of the curved version of Gauss' theorem, as well as the description of fluxes – rate of flow of, say, a fluid – across surface areas. If we have a $(D-1)$ dimensional hypersurface with induced metric $H_{\mathrm{IJ}}(\xi^{\mathrm{K}})$, determinant $H \equiv \det H_{\mathrm{IJ}}$, and a unit normal n^i to it, then the answer is

$$\mathrm{d}^{D-1}\Sigma_i \equiv \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}n_i\left(\vec{x}(\vec{\xi})\right) \tag{9.7.76}$$

$$= \mathrm{d}^{D-1}\vec{\xi}\,\widetilde{\epsilon}_{ij_1j_2\dots j_{D-1}}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^{j_1}(\vec{\xi})}{\partial \xi^1}\frac{\partial x^{j_2}(\vec{\xi})}{\partial \xi^2}\dots\frac{\partial x^{j_{D-1}}(\vec{\xi})}{\partial \xi^{D-1}}. \tag{9.7.77}$$

The difference between equations (9.7.76) and (9.7.77) is that the first requires knowing the normal vector beforehand, while the second description is purely intrinsic to the hypersurface and can be computed once its parametrization $\vec{x}(\vec{\xi})$ is provided. Also be aware that the choice of orientation of the $\{\xi^{\mathrm{I}}\}$ should be consistent with that of the ambient $\{\vec{x}\}$ and the infinitesimal volume $\mathrm{d}^D\vec{x}\sqrt{|g|}\epsilon_{12\dots D}$.

The $\mathrm{d}^{D-1}\xi\sqrt{|H|}$ is the (scalar) infinitesimal area ($=(D-1)-$volume) and n_i provides the direction. The second equality requires justification. Let's recall the $D-1$ vector fields $\{E^i{}_{\mathrm{I}} \equiv \partial x^i/\partial\xi^{\mathrm{I}} | \mathrm{I} = 1,2,3,\dots,D-1\}$ tangent to the hypersurface.

Problem 9.79. Show that the tensor in eq. (9.7.77),

$$\widetilde{n}_i \equiv \widetilde{\epsilon}_{ij_1j_2\dots j_{D-1}}E^{j_1}{}_1\dots E^{j_{D-1}}{}_{D-1}, \tag{9.7.78}$$

is orthogonal to all the $D-1$ vectors $\{E^i{}_{\mathrm{I}}\}$. Since n_i is the sole remaining direction in the D space, $\widetilde{n}_i$ must be proportional to n_i

$$\widetilde{n}_i = \varphi\cdot n_i. \tag{9.7.79}$$

To find φ we merely have to dot both sides with n^i,

$$\varphi(\vec{\xi}) = \sqrt{|g(\vec{x}(\vec{\xi}))|}\epsilon_{ij_1j_2\dots j_{D-1}}n^i\frac{\partial x^{j_1}(\vec{\xi})}{\partial\xi^1}\dots\frac{\partial x^{j_{D-1}}(\vec{\xi})}{\partial\xi^{D-1}}. \tag{9.7.80}$$

Given a point of the surface $\vec{x}(\vec{\xi})$ we can always choose the coordinates $\vec{x}$ of the ambient space such that, at least in a neighborhood of this point, x^1 refers to the direction orthogonal to the surface and the $\{x^2,x^3,\dots,x^D\}$ lie on the surface itself. Argue that, in this coordinate system, eq. (9.7.14) becomes

$$n^i = \frac{g^{(i)(1)}}{\sqrt{g^{(1)(1)}}}, \tag{9.7.81}$$

and therefore eq. (9.7.80) reads

$$\varphi(\vec{\xi}) = \sqrt{|g(\vec{x}(\vec{\xi}))|}\sqrt{g^{(1)(1)}}. \tag{9.7.82}$$

Cramer's rule (cf. (3.2.30)) from matrix algebra reads: the ij component (the ith row and jth column) of the inverse of a matrix $(A^{-1})_{ij}$ is $((-)^{i+j}/\det A)$ times the determinant of A with the jth row and ith column removed. Use this and the definition of the induced metric to conclude that

$$\varphi(\vec{\xi}) = \sqrt{|H(\vec{\xi})|}, \tag{9.7.83}$$

thereby proving the equality of equations (9.7.76) and (9.7.77). □

Gauss' theorem We are now ready to state (without proof) *Gauss' theorem*. In 3D vector calculus, Gauss tells us the volume integral, over some domain $\mathfrak{D}$, of the divergence of a vector field is equal to the flux of the same vector field across the boundary $\partial\mathfrak{D}$ of the domain. Exactly the same statement applies in a D dimensional ambient curved space with some closed $(D-1)$-dimensional hypersurface that defines $\partial\mathfrak{D}$.

> Let V^i be an arbitrary vector field, and let $\vec{x}(\vec{\xi})$ describe this closed boundary surface so that it has an (outward) directed surface element $\mathrm{d}^{D-1}\Sigma_i$ given by equations (9.7.76) and (9.7.77). Then
>
> $$\int_{\mathfrak{D}} \mathrm{d}^D x \sqrt{|g(\vec{x})|} \nabla_i V^i(\vec{x}) = \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i V^i \left(\vec{x}(\vec{\xi})\right). \tag{9.7.84}$$

Flux Just as in 3D vector calculus, the $\mathrm{d}^{D-1}\Sigma_i V^i$ can be viewed as the flux of some fluid described by V^i across an infinitesimal element of the hypersurface $\partial\mathfrak{D}$.
Remark Gauss' theorem is not terribly surprising if you recognize the integrand as a total derivative,

$$\sqrt{|g|}\nabla_i V^i = \partial_i(\sqrt{|g|}V^i) \tag{9.7.85}$$

(recall eq. (9.5.157)) and therefore it should integrate to become a surface term ($= (D-1)$-dimensional integral). The right hand side of eq. (9.7.84) merely makes this surface integral explicit, in terms of the coordinates $\vec{\xi}$ describing the boundary $\partial\mathfrak{D}$.
Closed surface Note that if you apply Gauss' theorem eq. (9.7.84), on a closed surface such as the sphere, the result is immediately zero. A closed surface is one where there are no boundaries. (For the 2-sphere, imagine starting with the Northern Hemisphere; the boundary is then the equator. By moving this boundary south-wards, i.e., from one latitude line to the next, until it vanishes at the South Pole – our boundary-less surface becomes the 2-sphere.) Since there are no boundaries, the right hand side of eq. (9.7.84) is automatically zero.

Problem 9.80. We may see this directly for the 2-sphere case. The metric on the 2-sphere of radius R is

$$\mathrm{d}\ell^2 = R^2(\mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2), \qquad \theta \in [0,\pi],\ \phi \in [0, 2\pi). \tag{9.7.86}$$

Let V^i be an arbitrary smooth vector field on the 2-sphere. Show explicitly – namely, do the integral – that

$$\int_{\mathbb{S}^2} \mathrm{d}^2 x \sqrt{|g(\vec{x})|} \nabla_i V^i = 0. \tag{9.7.87}$$

Hints: For the ϕ-integral, remember that $\phi = 0$ and $\phi = 2\pi$ refer to the same point, for a fixed θ. And, for regular V^i, $\sin(\theta) \cdot V^r$ is zero at the poles – why? □

Problem 9.81. Hodge dual formulation of Gauss' theorem in D-space. Let us consider the Hodge dual of the vector field in eq. (9.7.84),

$$\widetilde{V}_{i_1 \dots i_{D-1}} = \widetilde{\epsilon}_{i_1 \dots i_{D-1} j} V^j. \tag{9.7.88}$$

First show that

$$\widetilde{\epsilon}^{j i_1 \dots i_{D-1}} \nabla_j \widetilde{V}_{i_1 \dots i_{D-1}} \propto \partial_{[1} \widetilde{V}_{23 \dots D]} \propto \nabla_i V^i. \tag{9.7.89}$$

(Find the proportionality factors.) Then prove the dual formulation of Gauss' theorem:

$$\frac{1}{(D-1)!} \int_{\mathfrak{D}} \mathrm{d}^D x \partial_{[1} \widetilde{V}_{23 \dots D]} = \int_{\partial \mathfrak{D}} \mathrm{d}^{D-1} \xi \widetilde{V}_{i_1 \dots i_{D-1}} \left(\vec{x}(\vec{\xi}) \right) \frac{\partial x^{i_1}(\vec{\xi})}{\partial \xi^1} \dots \frac{\partial x^{i_{D-1}}(\vec{\xi})}{\partial \xi^{D-1}}. \tag{9.7.90}$$

The $\widetilde{V}_{i_1 \dots i_{D-1}} \partial_{\xi^1} x^{i_1} \dots \partial_{\xi^{D-1}} x^{i_{D-1}}$ can be viewed as the original tensor $\widetilde{V}_{i_1 \dots i_{D-1}}$, but projected onto the boundary $\partial \mathfrak{D}$.

In passing, I should point out, what you have shown in eq. (9.7.90), is that the Hodge dual formulation of Gauss' theorem can be written in a compact manner using differential forms notation,

$$\int_{\mathfrak{D}} \mathrm{d} \widetilde{V} = \int_{\partial \mathfrak{D}} \widetilde{V}, \tag{9.7.91}$$

by viewing the fully antisymmetric object $\widetilde{V}$ as a differential $(D-1)$-form. $\square$

Example: Coulomb potential in flat space A basic application of Gauss' theorem is the derivation of the (spherically symmetric) Coulomb potential of a unit point charge in $D \geq 3$ spatial dimensions, satisfying

$$\nabla_i \nabla^i \psi = -\delta^{(D)}(\vec{x} - \vec{x}') \tag{9.7.92}$$

in flat space. Let us consider as domain $\mathfrak{D}$ the sphere of radius r centered at the point charge at $\vec{x}'$. Using spherical coordinates, $\vec{x} = r \widehat{n}(\vec{\xi})$, where $\widehat{n}$ is the unit radial vector emanating from $\vec{x}'$, the induced metric on the boundary $\partial \mathfrak{D}$ is simply the metric of the $(D-1)$-sphere. We now identify in eq. (9.7.84) $V^i = \nabla^i \psi$. The normal vector is simply $n^i \partial_i = \partial_r$, and so Gauss' law using eq. (9.7.76) reads

$$-1 = \int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1} \vec{\xi} \sqrt{|H|} r^{D-1} \partial_r \psi(r). \tag{9.7.93}$$

The $\int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1} \vec{\xi} \sqrt{|H|} = 2\pi^{D/2} / \Gamma(D/2)$ is simply the solid angle subtended by the $(D-1)$-sphere ($\equiv$ volume of the $(D-1)$-sphere of unit radius). So at this point we have

$$\partial_r \psi(r) = -\frac{\Gamma(D/2)}{2\pi^{D/2} r^{D-1}}. \tag{9.7.94}$$

For $D \geq 3$, if we impose the boundary condition $\psi(r \to \infty) = 0$, the unique solution is then

$$\psi(r) = \frac{\Gamma(D/2)}{4((D-2)/2) \pi^{D/2} r^{D-2}} = \frac{\Gamma(\frac{D}{2} - 1)}{4\pi^{D/2} r^{D-2}}. \tag{9.7.95}$$

I have used the Gamma-function identity $\Gamma(z)z = \Gamma(z+1)$. Replacing $r \to |\vec{x} - \vec{x}'|$, we conclude that the Coulomb potential due to a unit strength electric charge is

$$\psi(\vec{x}) = \frac{\Gamma(\frac{D}{2} - 1)}{4\pi^{D/2}|\vec{x} - \vec{x}'|^{D-2}}, \qquad D \geq 3. \tag{9.7.96}$$

It is instructive to also use Gauss' law using eq. (9.7.77).

$$-1 = \int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\vec{\xi} \epsilon_{i_1 \dots i_{D-1} j} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} g^{jk}(\vec{x}(\vec{\xi}))\partial_k \psi(r \equiv \sqrt{\vec{x}^2}). \tag{9.7.97}$$

On the surface of the sphere, we have the completeness relation (cf. (4.3.23)):

$$g^{jk}(\vec{x}(\vec{\xi})) = \delta^{\mathrm{IJ}} \frac{\partial x^j}{\partial \xi^{\mathrm{I}}} \frac{\partial x^k}{\partial \xi^{\mathrm{J}}} + \frac{\partial x^j}{\partial r} \frac{\partial x^k}{\partial r}. \tag{9.7.98}$$

(This is also the coordinate transformation for the inverse metric from Cartesian to Spherical coordinates.) At this point,

$$\begin{aligned} -1 &= \int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\vec{\xi} \epsilon_{i_1 \dots i_{D-1} j} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} \left(\delta^{\mathrm{IJ}} \frac{\partial x^j}{\partial \xi^{\mathrm{I}}} \frac{\partial x^k}{\partial \xi^{\mathrm{J}}} + \frac{\partial x^j}{\partial r} \frac{\partial x^k}{\partial r} \right) \partial_k \psi(r \equiv \sqrt{\vec{x}^2}) \\ &= \int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\vec{\xi} \epsilon_{i_1 \dots i_{D-1} j} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} \frac{\partial x^j}{\partial r} \left(\frac{\partial x^k}{\partial r} \partial_k \psi(r \equiv \sqrt{\vec{x}^2}) \right). \end{aligned} \tag{9.7.99}$$

The Levi-Civita symbol contracted with the Jacobians can now be recognized as simply the determinant of the D-dimensional metric written in spherical coordinates $\sqrt{|g(r, \vec{\xi})|}$. (Note the determinant is positive because of the way we ordered our coordinates.) That is in fact equal to $\sqrt{|H(r, \vec{\xi})|}$ because $g_{rr} = 1$. Whereas $(\partial x^k / \partial r)\partial_k \psi = \partial_r \psi$. We have therefore recovered the previous result using eq. (9.7.76).

Problem 9.82. Coulomb Potential in 2D Use the above arguments to show, the solution of Laplace's equation

$$\nabla_i \nabla^i \psi = -\delta^{(2)} \left(\vec{x} - \vec{x}' \right) \tag{9.7.100}$$

is

$$\psi(\vec{x}) = -\frac{\ln \left(L^{-1} \left| \vec{x} - \vec{x}' \right| \right)}{2\pi}. \tag{9.7.101}$$

Here, L is an arbitrary length scale. Why is there is a restriction $D \geq 3$ in eq. (9.7.96)? Refer to eq. (6.2.62). If we employ "dimensional regularization", by setting $D = 2 - \epsilon$ for $0 < \epsilon \ll 1$, recover eq. (9.7.101) from the finite portion of eq. (9.7.96) in the $\epsilon \to 0$ limit. $\square$

Gauss' Theorem and Integration-By-Parts Gauss' theorem allows us to generalize integration-by-parts in 1D calculus to an arbitrary curved space(time). Specifically, if ψ is a coordinate scalar and V^i a vector,

$$\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}(\nabla_i V^i)\psi = -\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}V^i \nabla_i \psi + \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i V^i \psi; \tag{9.7.102}$$

where $\vec{x}(\vec{\xi})$ describes the boundary of the domain of integration $\mathfrak{D}$, which is denoted by $\partial\mathfrak{D}$; while $\mathrm{d}^{D-1}\Sigma_i = \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H|}n_i$, with H being the determinant of the induced metric on $\partial\mathfrak{D}$ and n_i is the unit normal to it.

The validity of eq. (9.7.102) can be seen by applying Gauss' theorem:

$$\int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i V^i \psi = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\nabla_i\left(V^i\psi\right) \tag{9.7.103}$$

$$= \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}(\nabla_i V^i)\psi + \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}V^i\nabla_i\psi. \tag{9.7.104}$$

Tensor elements A $(D-1)$-dimensional hypersurface embedded in $D-$space has one direction perpendicular to it. A $N \leq D-1$ dimensional one must admit $D-N$ orthogonal directions.

Suppose we have such a $(N < D)$-dimensional domain $\mathfrak{D}$ parametrized by $\{\vec{x}(\xi^{\mathrm{I}})|\mathrm{I} = 1, 2, \dots, N\}$ whose boundary $\partial\mathfrak{D}$ is in turn parametrized by $\{\vec{x}(\theta^{\mathfrak{A}})|\mathfrak{A} = 1, 2, \dots, N-1\}$. We may then define a $(D-N)$-tensor element that generalizes the one in eq. (9.7.77)

$$\mathrm{d}^N\Sigma_{i_1\dots i_{D-N}} \equiv \mathrm{d}^N\xi\ \widetilde{\epsilon}_{i_1\dots i_{D-N}j_1j_2\dots j_N}\left(\vec{x}(\vec{\xi})\right)\frac{\partial x^{j_1}(\vec{\xi})}{\partial\xi^1}\frac{\partial x^{j_2}(\vec{\xi})}{\partial\xi^2}\dots\frac{\partial x^{j_N}(\vec{\xi})}{\partial\xi^N}. \tag{9.7.105}$$

These $\{(\partial x^i/\partial\xi^{\mathrm{I}})\partial_i\}$ are tangent to the $\mathfrak{D}$-hypersurface and therefore, the anti-symmetric character of $\widetilde{\epsilon}$ guarantees that this $\mathrm{d}^N\Sigma_{i_1\dots i_{D-N}}$ is orthogonal to them. For, if we decompose

$$V^i \equiv V^i_\perp + V^{\mathrm{I}}\frac{\partial x^i}{\partial\xi^{\mathrm{I}}}, \qquad \text{with} \qquad g_{ij}V^i_\perp\frac{\partial x^j}{\partial\xi^{\mathrm{I}}} = 0 \qquad \forall\mathrm{I}; \tag{9.7.106}$$

then its contraction with the tensor element will project out all its components parallel to the $\mathfrak{D}-$hypersurface:

$$\mathrm{d}^N\Sigma_{ki_2\dots i_{D-N}}V^j = \mathrm{d}^N\xi\ \widetilde{\epsilon}_{ki_2\dots i_{D-N}j_1j_2\dots j_N}\left(V^k_\perp + V^{\mathrm{I}}\frac{\partial x^k}{\partial\xi^{\mathrm{I}}}\right)\frac{\partial x^{j_1}}{\partial\xi^1}\frac{\partial x^{j_2}}{\partial\xi^2}\dots\frac{\partial x^{j_N}}{\partial\xi^N} \tag{9.7.107}$$

$$= \mathrm{d}^N\xi\ \widetilde{\epsilon}_{ki_2\dots i_{D-N}j_1j_2\dots j_N}V^k_\perp\frac{\partial x^{j_1}}{\partial\xi^1}\frac{\partial x^{j_2}}{\partial\xi^2}\dots\frac{\partial x^{j_N}}{\partial\xi^N}. \tag{9.7.108}$$

Similarly, we may define the surface element on the boundary of $\mathfrak{D}$ to be

$$\mathrm{d}^{N-1}\Sigma_{i_1\dots i_{D-N}k} \equiv \mathrm{d}^{N-1}\theta\ \widetilde{\epsilon}_{i_1\dots i_{D-N}kj_1\dots j_{N-1}}\left(\vec{x}(\vec{\theta})\right)\frac{\partial x^{j_1}(\vec{\theta})}{\partial\theta^1}\frac{\partial x^{j_2}(\vec{\theta})}{\partial\theta^2}\dots\frac{\partial x^{j_{N-1}}(\vec{\theta})}{\partial\theta^{N-1}}. \tag{9.7.109}$$

Again, these $(\partial x^i/\partial\theta^{\mathrm{I}})\partial_i$ are tangent to the $\partial\mathfrak{D}$-hypersurface and therefore, the anti-symmetric character of $\widetilde{\epsilon}$ guarantees that this $\mathrm{d}^{N+1}\Sigma_{i_1\dots i_{D-N+1}}$ is orthogonal to them.

Stokes' theorem[100] In a $(N < D)$-dimensional simply connected subregion $\mathfrak{D}$ of some D-dimensional ambient space, the divergence of a fully antisymmetric

[100]Just like for the Gauss' theorem case, in equations (9.7.105) and (9.7.109), the $\vec{\xi}$ and $\vec{\theta}$ coordinate systems need to be defined with orientations consistent with the ambient $\mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\epsilon_{12\dots D}$ one.

rank $(D-N+1)$ tensor field $B^{i_1\dots i_{D-N}k}$ integrated over the domain $\mathfrak{D}$ can also be expressed as the integral of $B^{i_1\dots i_{D-N}k}$ over its boundary $\partial\mathfrak{D}$. Namely,

$$\int_{\mathfrak{D}} \mathrm{d}^N\Sigma_{i_1\dots i_{D-N}}\nabla_k B^{i_1\dots i_{D-N}k} = \frac{1}{D-N+1}\int_{\partial\mathfrak{D}} \mathrm{d}^{N-1}\Sigma_{i_1\dots i_{D-N}k}B^{i_1\dots i_{D-N}k}, \quad (9.7.110)$$
$$N<D,\ B^{[i_1\dots i_{D-N}k]} = (D-N+1)!B^{i_1\dots i_{D-N}k}.$$

Problem 9.83. Hodge dual formulation of Stokes' theorem. Define

$$\widetilde{B}_{j_1\dots j_{N-1}} \equiv \frac{1}{(D-N+1)!}\widetilde{\epsilon}_{j_1\dots j_{N-1}i_1\dots i_{D-N}k}B^{i_1\dots i_{D-N}k}. \quad (9.7.111)$$

Can you convert eq. (9.7.110) into

$$\frac{1}{(N-1)!}\int_{\mathfrak{D}} \mathrm{d}^N\vec{\xi}\partial_{[i_1}\widetilde{B}_{i_2\dots i_N]}\frac{\partial x^{i_1}}{\partial\xi^1}\dots\frac{\partial x^{i_N}}{\partial\xi^N} = \int_{\partial\mathfrak{D}} \mathrm{d}^{N-1}\vec{\theta}\widetilde{B}_{i_1\dots i_{N-1}}\frac{\partial x^{i_1}}{\partial\theta^1}\dots\frac{\partial x^{i_{N-1}}}{\partial\theta^{N-1}}? \quad (9.7.112)$$

Furthermore, explain why the Jacobians can be "brought inside the derivative".

$$\partial_{[i_1}\widetilde{B}_{i_2\dots i_N]}\frac{\partial x^{i_1}}{\partial\xi^1}\dots\frac{\partial x^{i_N}}{\partial\xi^N} = \frac{\partial x^{i_1}}{\partial\xi^{[1}}\partial_{|i_1|}\left(\frac{\partial x^{i_2}}{\partial\xi^2}\dots\frac{\partial x^{i_N}}{\partial\xi^{N]}}\widetilde{B}_{i_2\dots i_N}\right). \quad (9.7.113)$$

(The $|\cdot|$ around i_1 indicate it is *not* to be part of the anti-symmetrization; only do so for the ξ-indices.) This means we are really taking the exterior derivative of $\widetilde{B}$ induced on $\mathfrak{D}$.

Like for Gauss' theorem, we point out that – by viewing $\widetilde{B}_{j_1\dots j_{N-1}}$ as components of a $(N-1)$-form, the Hodge dual version of Stokes' theorem in eq. (9.7.110) reduces to the simple expression

$$\int_{\mathfrak{D}} \mathrm{d}\widetilde{B} = \int_{\partial\mathfrak{D}} \widetilde{B}, \quad (9.7.114)$$

where $(\mathrm{d}\widetilde{B})_{i_1\dots i_N} \equiv \partial_{[i_1}\widetilde{B}_{i_2\dots i_N]}/(N-1)!$. Note: if the $N-1$ form $\widetilde{B}$ does not depend on the metric, then $\mathrm{d}\widetilde{B}$ does not either. In this form, Stokes' theorem is metric-independent. □

Example: 2D Area as Line Integral As a simple example, let us compute the area A of some domain $\mathfrak{D}$ within a 2D flat space, parametrized by some pseudo-radial coordinate $\rho\in[0,\rho_\star]$ and pseudo-polar angle $\phi\in[0,2\pi)$. Denoting $\xi^i\equiv(\rho,\phi)$, we have

$$A = \int_{\mathfrak{D}} \mathrm{d}x\wedge\mathrm{d}y \quad (9.7.115)$$
$$= \int_{\mathfrak{D}} \frac{\partial x}{\partial\xi^{[1}}\frac{\partial y}{\partial\xi^{2]}}\mathrm{d}^2\vec{\xi} = \frac{1}{2}\int_{\mathfrak{D}}\left(\frac{\partial x}{\partial\xi^{[1}}\frac{\partial y}{\partial\xi^{2]}} - \frac{\partial y}{\partial\xi^{[1}}\frac{\partial x}{\partial\xi^{2]}}\right)\mathrm{d}^2\vec{\xi}. \quad (9.7.116)$$

We have anti-symmetrized the (x,y) coordinates in the last equality so as to match the vector calculus result we shall soon arrive at. In particular, if we recognize

$$\frac{\partial x}{\partial\xi^{[1}}\frac{\partial y}{\partial\xi^{2]}} - \frac{\partial y}{\partial\xi^{[1}}\frac{\partial x}{\partial\xi^{2]}} = \partial_{\xi^{[1}}\left(x\partial_{\xi^{2]}}y - y\partial_{\xi^{2]}}x\right), \quad (9.7.117)$$

then by defining $\widetilde{B}_i \equiv x\partial_{\xi^i}y - y\partial_{\xi^i}x$, we view the integrand as the exterior derivative

$$(\mathrm{d}\widetilde{B})_{ij} = \frac{\partial x}{\partial \xi^{[1}}\frac{\partial y}{\partial \xi^{2]}} - \frac{\partial y}{\partial \xi^{[1}}\frac{\partial x}{\partial \xi^{2]}}. \tag{9.7.118}$$

We may now employ Stokes' theorem in eq. (9.7.114) to express the area A as a line integral on the boundary of $\mathfrak{D}$:

$$A = \frac{1}{2}\int_{\mathfrak{D}} \mathrm{d}\widetilde{B} = \frac{1}{2}\int_{\partial\mathfrak{D}} \widetilde{B} \tag{9.7.119}$$

$$= \frac{1}{2}\int_0^{2\pi} \left(x(\rho_\star,\phi)\partial_\phi x(\rho_\star,\phi) - y(\rho_\star,\phi)\partial_\phi x(\rho_\star,\phi)\right)\mathrm{d}\phi. \tag{9.7.120}$$

This is the vector calculus result employed in eq. (8.2.13).

Relation to 3D vector calculus Stokes' theorem in vector calculus states that the flux of the curl of a vector field $\vec{A}$ over some 2D domain $\mathfrak{D}$ sitting in the ambient 3D flat space, is equal to the line integral of the same vector field along the boundary $\partial\mathfrak{D}$ of the domain. Specifically, if Cartesian coordinates and the ordinary dot product are employed; and $\mathrm{d}^2\vec{a}$ denotes the infinitesimal area element of $\mathfrak{D}$:

$$\int_{\mathfrak{D}} \mathrm{d}^2\vec{a}\cdot\left(\vec{\nabla}\times\vec{A}\right) = \oint_{\partial\mathfrak{D}} \vec{A}\cdot\mathrm{d}\vec{x}. \tag{9.7.121}$$

Because eq. (9.7.110) may not appear, at first sight, to be related to the Stokes' theorem in eq. (9.7.121) from 3D vector calculus, we shall work it out in some detail.

Problem 9.84. Consider some 2D hypersurface $\mathfrak{D}$ residing in a 3D curved space. For simplicity, let us foliate $\mathfrak{D}$ with constant ρ surfaces; let the other coordinate be ϕ, so $\vec{x}(0 \le \rho \le \rho_>, 0 \le \phi \le 2\pi)$ describes a given point on $\mathfrak{D}$ and the boundary $\partial\mathfrak{D}$ is given by the closed loop $\vec{x}(\rho = \rho_>, 0 \le \phi \le 2\pi)$. Let

$$B^{ik} \equiv \widetilde{\epsilon}^{ikj}A_j \tag{9.7.122}$$

for some vector field A^j. This implies in Cartesian coordinates,

$$\nabla_k B^{ik} = \left(\vec{\nabla}\times\vec{A}\right)^i. \tag{9.7.123}$$

Denote $\vec{\xi} = (\rho,\phi)$. Show that Stokes' theorem in eq. (9.7.110) reduces to the $N = 2$ vector calculus case:

$$\int_0^{\rho_>}\mathrm{d}\rho\int_0^{2\pi}\mathrm{d}\phi\sqrt{|H(\vec{\xi})|}\vec{n}\cdot\left(\vec{\nabla}\times\vec{A}\right) = \int_0^{2\pi}\mathrm{d}\phi\frac{\partial\vec{x}(\rho_>,\phi)}{\partial\phi}\cdot\vec{A}(\vec{x}(\rho_>,\phi)). \tag{9.7.124}$$

where the unit normal vector is given by

$$\vec{n} = \frac{(\partial\vec{x}(\vec{\xi})/\partial\rho)\times(\partial\vec{x}(\vec{\xi})/\partial\phi)}{\left|(\partial\vec{x}(\vec{\xi})/\partial\rho)\times(\partial\vec{x}(\vec{\xi})/\partial\phi)\right|}. \tag{9.7.125}$$

Of course, once you've verified Stokes' theorem for a particular coordinate system, you know by general covariance it holds in any coordinate system, i.e.,

$$\int_{\mathfrak{D}} \mathrm{d}^2\xi \sqrt{|H(\vec{\xi})|} n_i \widetilde{\epsilon}^{ijk} \partial_j A_k = \int_{\partial\mathfrak{D}} A_i \mathrm{d}x^i. \tag{9.7.126}$$

Step-by-step guide: Start with eq. (9.7.77), and show that in a Cartesian basis,

$$\mathrm{d}^2\Sigma_i = \mathrm{d}^2\xi \left(\frac{\partial \vec{x}}{\partial \rho} \times \frac{\partial \vec{x}}{\partial \phi} \right)^i. \tag{9.7.127}$$

The induced metric on the 2D domain $\mathfrak{D}$ is

$$H_{\mathrm{IJ}} = \delta_{ij} \partial_{\mathrm{I}} x^i \partial_{\mathrm{J}} x^j. \tag{9.7.128}$$

Work out its determinant. Then work out

$$|(\partial\vec{x}/\partial\rho) \times (\partial\vec{x}/\partial\phi)|^2 \tag{9.7.129}$$

using the identity

$$\widetilde{\epsilon}^{ijk} \widetilde{\epsilon}_{lmk} = \delta^i_l \delta^j_m - \delta^i_m \delta^j_l. \tag{9.7.130}$$

Can you thus relate $\sqrt{|H(\vec{\xi})|}$ to $|(\partial\vec{x}/\partial\rho) \times (\partial\vec{x}/\partial\phi)|$, and thereby verify the left hand side of eq. (9.7.110) yields the left hand side of (9.7.124)?

For the right hand side of eq. (9.7.124), begin by arguing that the boundary (line) element in eq. (9.7.109) becomes

$$\mathrm{d}\Sigma_{ki} = \mathrm{d}\phi\, \widetilde{\epsilon}_{kij} \frac{\partial x^j}{\partial \phi}. \tag{9.7.131}$$

Then use $\widetilde{\epsilon}^{ij_1j_2} \widetilde{\epsilon}_{kj_1j_2} = 2\delta^i_k$ to then show that the right hand side of eq. (9.7.110) is now that of eq. (9.7.124). $\square$

Problem 9.85. Discuss how the tensor element in eq. (9.7.105) transforms under a change of hypersurface coordinates $\vec{\xi} \to \vec{\xi}(\vec{\xi}')$. Do the same for the tensor element in eq. (9.7.109): how does it transforms under a change of hypersurface coordinates $\vec{\theta} \to \vec{\theta}(\vec{\theta}')$? Hint: Are the indices on the left hand side of equations (9.7.105) and (9.7.109) ambient space or hypersurface ones? $\square$

Poincaré Lemma In 3D vector calculus you have learned that a vector $\vec{B}$ is divergence-less everywhere in space iff it is the curl of another vector $\vec{A}$.

$$\vec{\nabla} \cdot \vec{B} = 0 \qquad \Leftrightarrow \qquad \vec{B} = \vec{\nabla} \times \vec{A} \tag{9.7.132}$$

And, the curl of a vector $\vec{B}$ is zero everywhere in space iff it is the gradient of scalar ψ.

$$\vec{\nabla} \times \vec{B} = 0 \qquad \Leftrightarrow \qquad \vec{B} = \vec{\nabla}\psi \tag{9.7.133}$$

Here, we shall see that these statements are special cases of the following.

In an arbitrary D dimensional curved space, let $B_{i_1 \dots i_N}(\vec{x})$ be a fully antisymmetric rank-N tensor field, with $N \leq D$. Then, everywhere within a simply connected region of space,

$$B_{i_1 \dots i_N} = \frac{\partial_{[i_1} C_{i_2 \dots i_N]}}{(N-1)!}, \tag{9.7.134}$$

– i.e., B is the "curl" of a fully antisymmetric rank-$(N-1)$ tensor C – if and only if

$$\partial_{[j} B_{i_1 \dots i_N]} = 0. \tag{9.7.135}$$

In differential form notation, by treating C as a $(N-1)$-form and B as a N-form, Poincaré would read: throughout a simply connected region of space,

$$\mathrm{d}B = 0 \qquad \text{iff} \qquad B = \mathrm{d}C. \tag{9.7.136}$$

Poincaré Example I: Electromagnetism Let us recover the 3D vector calculus statement above, that the divergence-less nature of the magnetic field is equivalent to it being the curl of some vector field. Consider the dual of the magnetic field B^i:

$$\widetilde{B}^{ij} \equiv \widetilde{\epsilon}^{ijk} B_k. \tag{9.7.137}$$

The Poincaré Lemma says $\widetilde{B}_{ij} = \partial_{[i} A_{j]}$ if and only if $\partial_{[k} \widetilde{B}_{ij]} = 0$ everywhere in space. We shall proceed to take the dual of these two conditions. Via eq. (9.6.34), the first is equivalent to

$$\begin{aligned} \widetilde{\epsilon}^{kij} \widetilde{B}_{ij} &= \widetilde{\epsilon}^{kij} \partial_{[i} A_{j]}, \\ &= 2\widetilde{\epsilon}^{kij} \partial_i A_j. \end{aligned} \tag{9.7.138}$$

On the other hand, employing eq. (9.6.34),

$$\widetilde{\epsilon}^{kij} \widetilde{B}_{ij} = \widetilde{\epsilon}^{kij} \widetilde{\epsilon}_{ijl} B^l = 2B^k; \tag{9.7.139}$$

and therefore $\vec{B}$ is the curl of A_i:

$$B^k = \widetilde{\epsilon}^{kij} \partial_i A_j. \tag{9.7.140}$$

While the latter condition $\mathrm{d}\widetilde{B} = 0$ is, again utilizing eq. (9.6.34), equivalent to

$$\begin{aligned} 0 &= \widetilde{\epsilon}^{kij} \partial_k \widetilde{B}_{ij} \\ &= \widetilde{\epsilon}_{kij} \widetilde{\epsilon}^{ijl} \nabla_k B_l = 2\nabla_l B^l. \end{aligned} \tag{9.7.141}$$

That is, the divergence of $\vec{B}$ is zero.

Poincaré Example II A simple application is that of the line integral

$$I(\vec{x}, \vec{x}'; \mathfrak{P}) \equiv \int_{\mathfrak{P}} A_i \mathrm{d}x^i, \tag{9.7.142}$$

where $\mathfrak{P}$ is some path in D-space joining $\vec{x}'$ to $\vec{x}$. Poincaré tells us, if $\partial_{[i} A_{j]} = 0$ everywhere in space, then $A_i = \partial_i \varphi$, the A_i is a gradient of a scalar φ. Then $A_i \mathrm{d}x^i = \partial_i \varphi \mathrm{d}x^i = \mathrm{d}\varphi$, and the integral itself is actually path independent – it depends only on the end points:

$$\int_{\vec{x}'}^{\vec{x}} A_i \mathrm{d}x^i = \int_{\mathfrak{P}} \mathrm{d}\varphi = \varphi(\vec{x}) - \varphi(\vec{x}'), \qquad \text{whenever } \partial_{[i} A_{j]} = 0. \tag{9.7.143}$$

Problem 9.86. A vector is curl-free iff it is a gradient Make a similar translation, from the Poincaré Lemma, to the 3D vector calculus statement that a vector $\vec{B}$ is curl-less ($\vec{\nabla} \times \vec{B} = 0$) if and only if it is a pure gradient ($\vec{B} = \vec{\nabla}\psi$) within the simply connected domain of interest. □

Problem 9.87. Infinitesimally Thin Solenoid Consider the vector potential, written in 3D Cartesian coordinates,

$$A_i \mathrm{d}x^i = \frac{x^1 \mathrm{d}x^2 - x^2 \mathrm{d}x^1}{(x^1)^2 + (x^2)^2}. \tag{9.7.144}$$

Can you calculate

$$F_{ij} = \partial_{[i} A_{j]}? \tag{9.7.145}$$

Consider a 2D surface whose boundary $\partial\mathfrak{D}$ circle around the $(0, 0, -\infty < x^3 < +\infty)$ line once. Can you use Stokes' theorem

$$\frac{1}{2} \int_{\mathfrak{D}} F_{ij} \mathrm{d}x^i \wedge \mathrm{d}x^j = \oint_{\partial\mathfrak{D}} A_i \mathrm{d}x^i \tag{9.7.146}$$

to show that

$$F_{ij} = 2\pi \epsilon_{ij3} \delta(x^1) \delta(x^2)? \tag{9.7.147}$$

Hint: Convert from Cartesian to polar coordinates $(x, y, z) = (r\cos\phi, r\sin\phi, z)$; the line integral on the right hand side of eq. (9.7.126) should simplify considerably.

This problem illustrates the subtlety regarding the "simply connected" requirement of the Poincaré lemma. The magnetic field F_{ij} here describes that of a highly localized solenoid lying along the z-axis; its corresponding vector potential is a pure gradient in any simply connected 3−volume not containing the z-axis, but it is no longer a pure gradient in say a solid torus region encircling (but still not containing) it. □

Problem 9.88. From Arfken *et al* **[18] Problem 4.4.1** In a simply connected region of a generic 3D curved space, show that if U and V are scalars, then

$$\partial_{[i} U \partial_{j]} V = 0 \qquad \text{iff} \qquad U \partial_k V = \partial_k f. \tag{9.7.148}$$

Hint: First, explain why $\partial_{[i} U \partial_{j]} V = \partial_{[i}(U \partial_{j]} V)$.

If the 3D space is flat and if Cartesian coordinates are used, then $\partial_{[i} U \partial_{j]} V$ can be viewed as the cross product $\vec{\nabla} U \times \vec{\nabla} V$. (Why?) This result can then be stated as: $\vec{\nabla} U \times \vec{\nabla} V = 0$ iff $\vec{\nabla} f = U \cdot \vec{\nabla} V$. □

9.8 *Maximally Symmetric Spaces: $D-$Spheres and Hyperboloids

We say that a set of N Killing vectors $\{\vec{\xi}_{\mathrm{I}} | \mathrm{I} = 1, 2, \dots, N\}$ are linearly independent if the solution to $\sum_{\mathrm{I}} a_{\mathrm{I}} \vec{\xi}_{\mathrm{I}} = 0$ for constants $\{a_{\mathrm{I}}\}$ is that all of them need to vanish, $a_{\mathrm{I}} = 0$. This allows us to record two important facts:

- The maximum number of linearly independent Killing vectors in D dimensions is

$$\frac{D(D+1)}{2}. \tag{9.8.1}$$

When a space has the maximum number of linearly independent Killing vectors, we say it is maximally symmetric.

- Maximally symmetric space(time)s are essentially unique up to a constant – its Ricci scalar $\mathcal{R}$. (See Chapter 13 of Weinberg's *Gravitation and Cosmology* [24] for a detailed discussion.)

- The Riemann and Ricci tensors of a maximally symmetric space are built entirely out of the metric and the constant $\mathcal{R}$ via the formulas

$$R_{ijmn} = \frac{\mathcal{R}}{D(D-1)} g_{i[m} g_{n]j} \qquad \text{and} \qquad R_{ij} = \frac{\mathcal{R}}{D} g_{ij}. \tag{9.8.2}$$

Flat space is an example of a maximally symmetric geometry, with zero Ricci scalar. In Cartesian coordinates, $g_{ij} = \delta_{ij}$ – and recalling the results from Problems (9.13) and (9.14) – the D partial derivatives $\{\partial_i\}$ and the $(D^2 - D)/2$ ones $\{x^{[i}\partial^{j]}\}$ altogether consist of $D(D+1)/2$ Killing vectors. On the other hand, the maximally symmetric $D-$space with *positive* Ricci scalars are spheres; while those with *negative* Ricci scalars are hyperboloids.

The uniqueness of maximally symmetric spaces means it does not quite matter *how* one constructs it – all maximally symmetric spaces with the same constant Ricci scalar have, up to coordinate transformations, the same metric. This section can be considered to be an extension of §(9.7), in that we shall follow Weinberg [24] and construct maximally symmetric spaces by viewing them as $D-$dimensional hypersurfaces situated in one higher dimensional flat space(time).

Spheres The $D-$sphere may be be viewed as the surface of constant radius R in $(D+1)-$dimensional flat space:

$$\vec{X}^2 + Z^2 \equiv \delta_{\mathrm{AB}} X^{\mathrm{A}} X^{\mathrm{B}} = R^2, \tag{9.8.3}$$

where $\{X^i | i = 1, 2, \dots, D\}$ and $Z \equiv X^{D+1}$ are Cartesian coordinates in the ambient space. This $D-$sphere remains the same under all rotations $\{\widehat{R}\}$ that leave the origin $\vec{0}$ fixed,

$$X^{\mathrm{A}} \equiv \begin{bmatrix} \vec{X} \\ Z \end{bmatrix} \to \widehat{R} \begin{bmatrix} \vec{X} \\ Z \end{bmatrix}, \tag{9.8.4}$$

$$X^{\mathrm{A}} \to \widehat{R}^{\mathrm{A}}{}_{\mathrm{B}} X^{\mathrm{B}}; \tag{9.8.5}$$

where the $(D+1) \times (D+1)$ matrices are orthogonal,

$$\widehat{R}^{\mathrm{T}} \widehat{R} = \mathbb{I}_{(D+1)\times(D+1)}, \tag{9.8.6}$$

$$\delta_{\mathrm{AB}} \widehat{R}^{\mathrm{A}}{}_{\mathrm{M}} \widehat{R}^{\mathrm{B}}{}_{\mathrm{N}} = \delta_{\mathrm{MN}}. \tag{9.8.7}$$

If $f(\vec{X}, Z)$ is a scalar in this $(D+1)$ flat space, an infinitesimal rotation of $(\vec{X}, Z = X^{D+1})^{\mathrm{T}}$ amounts to

$$f(\vec{X}, Z) \to f\left(\widehat{R} \cdot (\vec{X}, Z)^{\mathrm{T}}\right) \tag{9.8.8}$$

$$\approx f(\vec{X}, Z) + \omega_{\mathrm{AB}} J^{\mathrm{AB}} f(\vec{X}, Z) + \mathcal{O}(\omega^2); \tag{9.8.9}$$

where these capital indices run over $\{1, 2, \dots, D, D+1\}$; ω_{AB} are rotation angles; and the $\widehat{R}$ itself is generated by the Killing vectors

$$J^{\mathrm{AB}} \equiv X^{[\mathrm{A}} \delta^{\mathrm{B}]\mathrm{C}} \partial_{X^{\mathrm{C}}} = X^{[\mathrm{A}} \partial^{\mathrm{B}]}. \tag{9.8.10}$$

In other words, these $\{J^{\mathrm{AB}}\}$ must necessarily be tangent to the $D-$sphere. Since these Js are anti-symmetric under interchange of $\mathrm{A} \leftrightarrow \mathrm{B}$, we have altogether

$$\frac{1}{2}((D+1)^2 - (D+1)) = \frac{1}{2} D(D+1) \tag{9.8.11}$$

Killing vectors, thereby saturating the maximum number allowed in $D-$dimensions.

Let us now solve Z in terms of the rest of the $\vec{X}$,

$$Z = \pm\sqrt{R^2 - \vec{X}^2}, \tag{9.8.12}$$

$$\mathrm{d}Z = \mp \frac{\vec{X} \cdot \mathrm{d}\vec{X}}{\sqrt{R^2 - \vec{X}^2}}. \tag{9.8.13}$$

so that we may then derive the $D-$sphere metric

$$\mathrm{d}\ell^2 = \mathrm{d}\vec{X}^2 + \mathrm{d}Z^2 \tag{9.8.14}$$

$$= \mathrm{d}\vec{X}^2 + \frac{\left(R^{-1}\vec{X} \cdot \mathrm{d}\vec{X}\right)^2}{1 - \vec{X}^2/R^2}; \tag{9.8.15}$$

or

$$g_{ij} = \delta_{ij} + \frac{X^i X^j / R^2}{1 - \vec{X}^2/R^2}; \tag{9.8.16}$$

with determinant

$$\det g_{ij} = \frac{1}{1 - \vec{X}^2/R^2}. \tag{9.8.17}$$

Note that the $\vec{X}^2 = R^2 - Z^2$ constraint in eq. (9.8.3) implies $\vec{X}^2/R \leq 1$; and, hence, the metric determinant is guaranteed to be positive.

The inverse metric, on the other hand, is

$$g^{ij} = \delta^{ij} - \frac{X^i X^j}{R^2}. \tag{9.8.18}$$

Of course, in this form, the metric and its inverse are only describing half the sphere at a time, depending on which sign ($\pm$) we choose for Z in eq. (9.8.12). (To describe the entire $D-$sphere except perhaps its North Pole using a single set of coordinates, recall Problem (9.12).) However, these $\{X^i\}$ coordinates allow us to readily compute the Christoffel symbols of eq. (9.8.15),

$$\Gamma^i{}_{mn} = \frac{X^i}{R^2} g_{mn}. \tag{9.8.19}$$

By a direct calculation, the Ricci scalar is indeed a positive constant:

$$\mathcal{R} = \frac{D(D-1)}{R^2}. \tag{9.8.20}$$

Next, let us choose the affinely parametrized geodesic Lagrangian to be 1/2; or, equivalently,

$$g_{mn}\dot{X}^m\dot{X}^n = 1, \tag{9.8.21}$$

where each overdot denotes a derivative with respect to the affine parameter λ. This constraint, as you will show in Problem (9.90), leads to the world function on the $D-$sphere. When utilized in the geodesic equation $\ddot{X}^i + \Gamma^i{}_{mn}\dot{X}^m\dot{X}^n = 0$, eq. (9.8.21) converts it into that of the simple harmonic oscillator:

$$\ddot{X}^i + \frac{X^i}{R^2} = 0. \tag{9.8.22}$$

The solution linking $\vec{X}(\lambda = 0) = \vec{X}_0$ to $\vec{X}(\lambda = \ell) = \vec{X}_1$ is given by

$$\vec{X}(0 \leq \lambda \leq \ell) = \vec{X}_0 \cos(\lambda/R) + \frac{\vec{X}_1 - \vec{X}_0\cos(\ell/R)}{\sin(\ell/R)}\sin(\lambda/R). \tag{9.8.23}$$

Here $\ell = \arccos(\vec{X}_0 \cdot \vec{X}_1)$; and one may readily check that $\vec{X}(\lambda)^2 = 1$.

Killing Vectors on $D-$Sphere When viewed as vectors in flat $(D+1)-$space, the $J^{\text{AB}} = X^{[\text{A}}\delta^{\text{B}]\text{C}}\partial_{\text{C}}$ may be readily verified to be Killing. Lowering its indices, we verify– for fixed AB – the $\nabla_{\{\text{E}}J^{\text{AB}}{}_{\text{F}\}} = 0$ form of the Killing equation:

$$\partial_{\{\text{E}}\left(X^{[\text{A}}\delta^{\text{B}]}{}_{\text{F}\}}\right) = \delta^{[\text{A}}_{\{\text{E}}\delta^{\text{B}]}{}_{\text{F}\}} = 0. \tag{9.8.24}$$

What is somewhat less obvious is the fact that these J^{AB}s are tangent to the radius$-R$ sphere embedded in $(D+1)-$space. For that, let us temporarily promote R to a radial coordinate in the $(D+1)-$space and define

$$X^{\text{A}} \equiv \left(\vec{X}, Z\right)^{\text{T}} \equiv R \cdot \widehat{r}(\vec{\theta}), \tag{9.8.25}$$

where $\vec{\theta}$ is now the set of D angles parametrizing the $D-$sphere. These J^{AB}s with lower indices, again for fixed AB, now become

$$X^{[\text{A}}\mathrm{d}X^{\text{B}]} = R\widehat{r}^{[\text{A}}\widehat{r}^{\text{B}]}\mathrm{d}R + R^2\widehat{r}^{[\text{A}}\partial_{\theta^{\text{C}}}\widehat{r}^{\text{B}]}\mathrm{d}\theta^{\text{C}} \tag{9.8.26}$$

$$= R^2 \widehat{r}^{[\mathrm{A}} \partial_{\theta^{\mathrm{C}}} \widehat{r}^{\mathrm{B}]} \mathrm{d}\theta^{\mathrm{C}}. \tag{9.8.27}$$

There are no $\mathrm{d}R$ terms – no displacement in the radial direction. As far as the J^{AB}s are concerned the radius R is indeed constant.

In terms of the $\{X^i\}$, the Killing 1−forms are $X^{[i}\mathrm{d}X^{j]}$ and $X^i\mathrm{d}Z - Z\mathrm{d}X^i$, for $i \in \{1, 2, \dots, D\}$. The latter reads

$$X^i \mathrm{d}Z - Z\mathrm{d}X^i = \mp X^i \frac{\vec{X} \cdot \mathrm{d}\vec{X}}{\sqrt{R^2 - \vec{X}^2}} \mp \sqrt{R^2 - \vec{X}^2}\mathrm{d}X^i. \tag{9.8.28}$$

The Killing vectors may be obtained from their 1−form cousins by replacing

$$\mathrm{d}X^i \to g^{ij}\partial_j = \delta^{ij}\partial_j - \frac{X^i X^j}{R^2}\partial_j. \tag{9.8.29}$$

We have

$$X^{[i}\mathrm{d}X^{j]} \to X^{[i}\delta^{j]k}\partial_k - \frac{X^{[i}X^{j]}X^k}{R^2}\partial_k \tag{9.8.30}$$

$$= X^{[i}\partial^{j]} \tag{9.8.31}$$

and

$$\mp \frac{X^i\mathrm{d}Z - Z\mathrm{d}X^i}{R} \to \frac{X^i X^j/R^2}{\sqrt{1 - \frac{\vec{X}^2}{R^2}}}\left(\delta^{jk} - \frac{X^j X^k}{R^2}\right)\partial_k + \sqrt{1 - \frac{\vec{X}^2}{R^2}}\left(\delta^{ik} - \frac{X^i X^k}{R^2}\right)\partial_k \tag{9.8.32}$$

$$= \sqrt{1 - \vec{X}^2/R^2} \cdot \partial^i. \tag{9.8.33}$$

Notice, as the radius of curvature goes to infinity, $R \to \infty$, these Killing vectors in equations (9.8.31) and (9.8.33) transform respectively into rotation and translation generators in flat space.

Problem 9.89. Determinant Apply $\partial_i \ln \sqrt{|g|} = \Gamma^l{}_{il}$ in eq. (9.5.156) to eq. (9.8.19), and derive eq. (9.8.17). □

Problem 9.90. Affine Parameter and D-Sphere World Function Let $\widehat{r}_1$ and $\widehat{r}_2$ be the unit radial vectors in the ambient $(D+1)$−space, pointing from $\vec{0}$ to the surface of the D−sphere. Assuming their tips both lie within the same hemisphere, explain why the λ employed in eq. (9.8.21) is related to the cosine of the angle between them:

$$\widehat{r}_1 \cdot \widehat{r}_2 = \cos(\ell/R). \tag{9.8.34}$$

In fact, deduce that the world function itself is simply

$$\sigma(\widehat{r}_1, \widehat{r}_2) = \frac{R^2}{2}\arccos(\widehat{r}_1 \cdot \widehat{r}_2). \tag{9.8.35}$$

Hint: Starting from the integral representation of Synge's world function, explain why ℓ is the arc length. □

Problem 9.91. Non-Uniqueness and Topology Notice these solutions in eq. (9.8.23) have a period of $2\pi R$; and, hence, there are infinite number of geodesics joining any pair of points. Can you explain why?

Also, when $\ell = \pi R$, the $\sin(\ell/R)$ in eq. (9.8.23) renders it ill defined. Could you explain what is happening? How are the $\vec{X}_0$ and $\vec{X}_1$ related when $\ell = \pi R$? And, what are the geodesics joining them? $\square$

Problem 9.92. Three Sphere in Spherical Coordinates Explain why, in 4D, the following spherical coordinates can be used parametrize the surface in eq. (9.8.3):

$$(\vec{X}, Z) = R\left(\sin\left[\frac{\rho}{R}\right]\sin[\theta]\cos[\phi], \sin\left[\frac{\rho}{R}\right]\sin[\theta]\sin[\phi], \sin\left[\frac{\rho}{R}\right]\cos[\theta], \cos\left[\frac{\rho}{R}\right]\right). \tag{9.8.36}$$

Explain why the relevant domain of these variables are

$$0 \leq \frac{\rho}{R}, \theta \leq \pi \qquad \text{and} \qquad 0 \leq \phi < 2\pi. \tag{9.8.37}$$

Verify that, in terms of (ρ, θ, ϕ), eq. (9.8.15) becomes

$$\mathrm{d}\ell^2 = \mathrm{d}\rho^2 + R^2 \sin^2\left[\frac{\rho}{R}\right]\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right). \tag{9.8.38}$$

Both equations (9.8.15) and (9.8.38) appear in the context of positively curved maximally symmetric spatial subspaces of an expanding universe. $\square$

Problem 9.93. D-Dimensional Hyperboloid Starting from the hyperbolic equation

$$\vec{X}^2 - W^2 \equiv \eta_{\mathrm{AB}} X^{\mathrm{A}} X^{\mathrm{B}} = -R^2, \tag{9.8.39}$$

$$X^{\mathrm{A}} \equiv (W, \vec{X}); \tag{9.8.40}$$

in $D+1$ dimensional 'spacetime'

$$\mathrm{d}\ell^2 = \mathrm{d}\vec{X}^2 - \mathrm{d}W^2 = \eta_{\mathrm{AB}} \mathrm{d}X^{\mathrm{A}} \mathrm{d}X^{\mathrm{B}}; \tag{9.8.41}$$

show that the induced metric on one half of the hyperboloid can be expressed as

$$\mathrm{d}\ell^2 = \mathrm{d}\vec{X}^2 - \frac{\left(R^{-1}\vec{X}\cdot\mathrm{d}\vec{X}\right)^2}{1 + \vec{X}^2/R^2}, \tag{9.8.42}$$

$$g_{ij} = \delta_{ij} - \frac{X^i X^j/R^2}{1 + \vec{X}^2/R^2}; \tag{9.8.43}$$

and its determinant is

$$\det g_{ij} = \frac{1}{1 + \vec{X}^2/R^2}. \tag{9.8.44}$$

Next, show that the inverse metric and Christoffel symbols are

$$g^{ij} = \delta^{ij} + \frac{X^i X^j}{R^2} \tag{9.8.45}$$

and

$$\Gamma^i{}_{mn} = -\frac{X^i}{R^2} g_{mn}. \tag{9.8.46}$$

Let λ be the affine parameter for the geodesic such that

$$g_{mn} \dot{X}^m \dot{X}^n = 1 \tag{9.8.47}$$

and

$$\vec{X}[\lambda = 0] = \vec{X}_0 \qquad \text{and} \qquad \vec{X}[\lambda = \tau] = \vec{X}_1. \tag{9.8.48}$$

Demonstrate that the geodesic solutions are then given by

$$\vec{X}[0 \leq \lambda \leq \tau] = \vec{X}_0 \cosh(\lambda/R) + \frac{\vec{X}_1 - \vec{X}_0 \cosh(\tau/R)}{\sinh(\tau/R)} \sinh(\lambda/R). \tag{9.8.49}$$

Finally, show that the Ricci scalar is indeed a negative constant:

$$\mathcal{R} = -\frac{D(D-1)}{R^2}. \tag{9.8.50}$$

Killing Vectors on D Hyperboloid Explain why the $D(D+1)/2$ Killing vectors are the $D(D-1)/2$ vectors

$$\left\{ X^{[i} \delta^{j]k} \partial_k = X^{[i} \partial^{j]} \right\}. \tag{9.8.51}$$

plus the D vectors

$$\left\{ \sqrt{1 + \vec{X}^2/R^2} \cdot \partial_i \right\}. \tag{9.8.52}$$

Notice how equations (9.8.51) and (9.8.52) respectively recover rotation and translation generators in flat $D-$space as $R \to \infty$. $\square$

Problem 9.94. Three Hyperboloid in Hyperbolic-Spherical Coordinates Explain why, in 4D, the following hyperbolic-spherical coordinates can be used parametrize the surface in eq. (9.8.39):

$$(\vec{X}, Z) = R \left(\sinh\left[\frac{\rho}{R}\right] \sin[\theta] \cos[\phi], \sinh\left[\frac{\rho}{R}\right] \sin[\theta] \sin[\phi], \sinh\left[\frac{\rho}{R}\right] \cos[\theta], \cosh\left[\frac{\rho}{R}\right] \right). \tag{9.8.53}$$

Explain why the relevant domain of these variables are

$$\rho \geq 0, \qquad 0 \leq \theta \leq \pi \qquad \text{and} \qquad 0 \leq \phi < 2\pi. \tag{9.8.54}$$

Verify that, in terms of (ρ, θ, ϕ), eq. (9.8.15) becomes

$$d\ell^2 = d\rho^2 + R^2 \sinh^2\left[\frac{\rho}{R}\right] \left(d\theta^2 + \sin^2\theta d\phi^2 \right). \tag{9.8.55}$$

Both equations (9.8.43) and (9.8.55) appear in the context of negatively curved maximally symmetric spatial subspaces of an expanding universe. $\square$

9.9 *Helmholtz Decomposition of Vectors on $\mathbb{S}^2$

Basis Vector Fields We have seen that the angular spherical harmonics $\{Y_\ell^m(\widehat{x})\}$ are a complete set of functions on the $2-$sphere. Let us now employ them to build the following complete set of vector fields on the same $2-$sphere, by taking their gradients (which yields a curl-free vector) and the associated Hodge duals (which produces a divergence-free vector):

$$\left\{\nabla^i Y_\ell^m(\theta,\phi) = g^{ij}\partial_j Y_\ell^m \qquad \text{and} \qquad \widetilde{\epsilon}^{ij}\nabla_j Y_\ell^m(\theta,\phi) = \widetilde{\epsilon}^{ij}\partial_j Y_\ell^m(\theta,\phi)\right\}, \tag{9.9.1}$$

where $\{Y_\ell^m|\ell = 0,1,2,\dots, m = -\ell, -\ell+1,\dots,\ell-1,\ell\}$ are the angular spherical harmonics; and the covariant derivative ∇, inverse metric g^{ij}, and the covariant Levi-Civita tensor are all built out of the metric

$$g_{ij}\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2, \tag{9.9.2}$$

$$\theta \in [0,\pi], \qquad \phi \in (0,2\pi]. \tag{9.9.3}$$

That is, for any V^i there must be coefficients $\{A_\ell^m\}$ and $\{B_\ell^m\}$ corresponding to, respectively, its curl-free and divergence-free parts:

$$V^i(\theta,\phi) = -\sum_{\ell=1}^{\infty}\sum_{m=-\ell}^{+\ell}\frac{1}{\ell(\ell+1)}\left(A_\ell^m\cdot\nabla^i Y_\ell^m - B_\ell^m\cdot\widetilde{\epsilon}^{ij}\nabla_j Y_\ell^m\right). \tag{9.9.4}$$

[101]These constants $\{A_\ell^m\}$ and $\{B_\ell^m\}$ are uniquely given by

$$A_\ell^m = \int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{x}}\nabla_i V^i(\widehat{x})\overline{Y_\ell^m(\widehat{x})}, \tag{9.9.5}$$

$$B_\ell^m = \int_{\mathbb{S}^2}\mathrm{d}\Omega_{\widehat{x}}\widetilde{\epsilon}^{ij}\nabla_i V_j(\widehat{x})\overline{Y_\ell^m(\widehat{x})}. \tag{9.9.6}$$

Proof Taking the divergence of both sides of eq. (9.9.4), recalling the eigenvalue equation $\nabla_i\nabla^i Y_\ell^m = -\ell(\ell+1)Y_\ell^m$ and recognizing $\nabla_i(\widetilde{\epsilon}^{ij}\partial_j Y_\ell^m) = 0$,

$$\nabla_i V^i(\theta,\phi) = -\sum_{\substack{1\leq\ell\leq+\infty\\-\ell\leq m\leq+\ell}}(\ell(\ell+1))^{-1}A_\ell^m\nabla_i\nabla^i Y_\ell^m(\theta,\phi) \tag{9.9.7}$$

$$= \sum_{\substack{1\leq\ell\leq+\infty\\-\ell\leq m\leq+\ell}}A_\ell^m Y_\ell^m(\theta,\phi). \tag{9.9.8}$$

Taking the curl on both sides of eq. (9.9.4), and recognizing the curl of a gradient is zero $(\widetilde{\epsilon}_{ij}\nabla^i\nabla^j = 0)$ while $\widetilde{\epsilon}_{ij}\widetilde{\epsilon}^{jk} = -\delta_i^k$,

$$\widetilde{\epsilon}_{ij}\nabla^i V^j(\theta,\phi) = -\sum_{\substack{1\leq\ell\leq+\infty\\-\ell\leq m\leq+\ell}}(\ell(\ell+1))^{-1}(-)^2 B_\ell^m\delta_i^k\nabla^i\nabla_k Y_\ell^m(\theta,\phi) \tag{9.9.9}$$

[101]You may wonder if it is possible to have constant vector fields, since these $\{\nabla^i Y_\ell^m\}$ and $\{\widetilde{\epsilon}^{ij}\nabla_j Y_\ell^m\}$ are necessarily (θ,ϕ)-dependent. The answer is provided by the hairy ball theorem (aka hedgehog theorem): every continuous vector field on the $2-$sphere must necessarily vanish somewhere. For instance, ∂_θ appears to be a constant vector field, with unit length $g_{\theta\theta} = 1$ everywhere. However, as it is moved along a longitude line across the North or South pole, it changes direction abruptly and is therefore not continuous at these two locations.

$$= \sum_{\substack{1\leq\ell\leq+\infty \\ -\ell\leq m\leq+\ell}} B_\ell^m Y_\ell^m(\theta,\phi). \tag{9.9.10}$$

We have shown the *consistency* of the prescription in eq. (9.9.4) with its divergence and curl. But how do we know we have captured the full content of V^i? Suppose instead

$$V^i(\theta,\phi) = -\sum_{\substack{1\leq\ell\leq+\infty \\ -\ell\leq m\leq+\ell}} (\ell(\ell+1))^{-1}\left(A_\ell^m \nabla^i Y_\ell^m(\theta,\phi) - B_\ell^m \widetilde{\epsilon}^{ij}\partial_j Y_\ell^m(\theta,\phi)\right) + W^i, \tag{9.9.11}$$

where W^i is *defined* to be the difference between the exact V^i and the summation on the right hand side. The proof is complete once we can show $W^i = 0$. First, we may take the curl and divergence on both sides to deduce that

$$\widetilde{\epsilon}^{ij}\nabla_i W_j = 0 = \nabla_i W^i. \tag{9.9.12}$$

By the Poincaré lemma, the first equality implies $W_j = \partial_j\varphi$ for some scalar φ. The second equality then implies that $\nabla_i\nabla^i\varphi = 0$. But the solution to $\nabla_i\nabla^i\varphi = 0$ on a 2−sphere is $\varphi = \varphi_0 =$ constant and therefore $W^i = \nabla^i\varphi_0 = 0$.

To sum: any smooth vector field V^i on the 2−sphere may be decomposed into a gradient plus a dual gradient

$$V^i = \nabla^i\psi_1 + \widetilde{\epsilon}^{ij}\nabla_j\psi_2. \tag{9.9.13}$$

In particular, these gradients and dual gradients are acting on the superposition of (the complete set of) angular spherical harmonics – respectively, the first and second terms of eq. (9.9.4), with unique coefficients given in equations (9.9.5) and (9.9.6).

Corollary Given 4 scalar fields on the 2−sphere – A_1, A_2, B_1 and B_2 – if they obey

$$\nabla^i A_1 + \widetilde{\epsilon}^{ij}\nabla_j B_1 = \nabla^i A_2 + \widetilde{\epsilon}^{ij}\nabla_j B_2, \tag{9.9.14}$$

then both $A_1 - A_2$ and $B_1 - B_2$ are (different) constants. To see this result, simply take the curl and divergence on both sides; and again recall the solution of $\nabla_i\nabla^i\varphi = 0$ is a constant.

Problem 9.95. Orthonormal basis Define the basis vector fields in eq. (9.9.1) as

$$A^i_{\ell,m} \equiv \nabla^i Y_\ell^m/\sqrt{\ell(\ell+1)} \qquad \text{and} \qquad B^i_{\ell,m} \equiv \widetilde{\epsilon}^{ij}\partial_j Y_\ell^m/\sqrt{\ell(\ell+1)}. \tag{9.9.15}$$

Next, define the inner product between vector fields U^i and W^i as

$$\langle U|\, W\rangle \equiv \int_{\mathbb{S}^2} \mathrm{d}\Omega_{\widehat{x}} \overline{U^i} W_i. \tag{9.9.16}$$

Show that all the $\{A^i_{\ell,m}\}$ are orthogonal to all the $\{B^i_{\ell,m}\}$; i.e., $\langle A_\ell^m |\, B_{\ell'}^{m'}\rangle = 0$. Then show that the $\{A^i_{\ell,m}\}$ are themselves orthonormal; and so is $\{B^i_{\ell,m}\}$. Hint: Integrate-by-parts. $\square$

Problem 9.96. Most general divergence-free vector $W^i(r,\theta,\phi)$ in flat 3D In 3D flat space, the metric in spherical coordinates may be written as

$$g_{ij}\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}r^2 + r^2 H_{\mathrm{IJ}}\mathrm{d}x^{\mathrm{I}}\mathrm{d}x^{\mathrm{J}}, \tag{9.9.17}$$

$$H_{\mathrm{IJ}} \mathrm{d}x^{\mathrm{I}} \mathrm{d}x^{\mathrm{J}} = \mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2. \tag{9.9.18}$$

Show that the most general divergence-less vector W^i, obeying $\nabla_i W^i = 0$, takes the following form. For arbitrary scalar fields $\varphi(r, \theta, \phi)$ and $\psi(r, \theta, \phi)$, its radial component reads

$$W^r = -\frac{1}{r^2} \vec{\nabla}^2_{\mathbb{S}^2} \psi; \tag{9.9.19}$$

while its angular components are

$$W^{\mathrm{I}} = \frac{1}{r^2} \left(\widetilde{\epsilon}^{\mathrm{IJ}} \partial_{\mathrm{J}} \varphi + H^{\mathrm{IJ}} \partial_{\mathrm{J}} \partial_r \psi \right). \tag{9.9.20}$$

Here, I and J run over $\{\theta, \phi\}$; $H^{\mathrm{IJ}} = \mathrm{diag}(1, 1/\sin(\theta)^2)$ is the inverse metric on the 2−sphere; and $\vec{\nabla}^2_{\mathbb{S}^2}$ and $\widetilde{\epsilon}_{\mathrm{IJ}} = \sqrt{\det H_{\mathrm{AB}}} \epsilon_{\mathrm{IJ}}$ are respectively its Laplacian and Levi-Civita tensor.

Since the curl of a vector field is automatically divergence-free, we expect $\widetilde{\epsilon}^{ijk} \partial_j W_k$ to take the same form as W^i itself. Show that

$$\widetilde{\epsilon}^{rjk} \partial_j W_k = -\frac{1}{r^2} \vec{\nabla}^2_{\mathbb{S}^2} \varphi; \tag{9.9.21}$$

and, for $\mathrm{I} \in \{\theta, \phi\}$,

$$\widetilde{\epsilon}^{\mathrm{I}jk} \partial_j W_k = \frac{1}{r^2} \left(\widetilde{\epsilon}^{\mathrm{IJ}} \partial_{\mathrm{J}} \left(-\frac{\vec{\nabla}^2_{\mathbb{S}^2} \psi}{r^2} - \partial_r^2 \psi \right) + H^{\mathrm{IJ}} \partial_{\mathrm{J}} \partial_r \varphi \right). \tag{9.9.22}$$

In other words, taking the curl of the most general divergence-free W^i amounts to replacing its ψ with φ; and its φ with

$$-\frac{\vec{\nabla}^2_{\mathbb{S}^2} \psi}{r^2} - \partial_r^2 \psi. \tag{9.9.23}$$

Hints: For the first part, first recall the Poincaré lemma to re-write W^i as a curl of some vector field U_i. Then perform a Helmholtz decomposition on the angular components of U_i. You may find it useful to recognize, the 3D Levi-Civita tensor $\widetilde{\epsilon}_{ijk}$ is related to the 2-sphere one via $\widetilde{\epsilon}_{r\mathrm{IJ}} = r^2 \widetilde{\epsilon}_{\mathrm{IJ}}$. (Why is it true?) □

Problem 9.97. Vector Eigenfunctions Show that the $\{\nabla^i Y_\ell^m\}$ and $\{\widetilde{\epsilon}^{ij} \nabla_j Y_\ell^m\}$ are degenerate vector eigenfunctions of the Laplacian $\vec{\nabla}^2 \equiv \nabla_l \nabla^l$ on the unit sphere:

$$\vec{\nabla}^2 \nabla^i Y_\ell^m = (1 - \ell(\ell+1)) \nabla^i Y_\ell^m, \tag{9.9.24}$$

$$\vec{\nabla}^2 \left(\widetilde{\epsilon}^{ij} \nabla_j Y_\ell^m \right) = (1 - \ell(\ell+1)) \widetilde{\epsilon}^{ij} \nabla_j Y_\ell^m. \tag{9.9.25}$$

This degeneracy is broken by including the parity transformation, defined by $(\theta, \phi) \to (\pi - \theta, \phi + \phi)$ – can you explain why?

Hint: For some scalar f, first show that $\vec{\nabla}^2 \nabla_a f = (\nabla_a \vec{\nabla}^2 + R_{al} \nabla^l) f$. □

Problem 9.98. Completeness Relation Explain why

$$\sum_{\ell \geq 1} \sum_{-\ell \leq m \leq \ell} \frac{\nabla^i Y_\ell^m(\theta, \phi) \overline{\nabla_{j'} Y_\ell^m(\theta', \phi')} + \widetilde{\epsilon}^{ia} \nabla_a Y_\ell^m(\theta, \phi) \overline{\widetilde{\epsilon}_{j'b'} \nabla^{b'} Y_\ell^m(\theta', \phi')}}{\ell(\ell+1)}$$

$$= \delta^i{}_j \frac{\delta(\theta - \theta') \delta(\phi - \phi')}{\sqrt[4]{H(\theta, \phi) \cdot H(\theta', \phi')}} \tag{9.9.26}$$

is the completeness relation of vector spherical harmonics. □

9.10 ⋆Non-Relativistic Lagrangian Mechanics

Lagrangian Dynamics: General Coordinates In flat $D-$space with time coordinate t, the non-relativistic kinetic energy of a particle with mass m written in Cartesian coordinates $\{\vec{z}(t)\}$ is

$$T \equiv \frac{m}{2}\dot{\vec{z}}^2 = \frac{m}{2}\delta_{ij}\dot{z}^i\dot{z}^j. \tag{9.10.1}$$

If we perform a $t-$independent coordinate transformation $\vec{z}(t) \to \vec{z}(\vec{q}(t))$, so that $\mathrm{d}z^i/\mathrm{d}t = (\partial z^i/\partial q^a)(\mathrm{d}q^a/\mathrm{d}t)$, then we see that the kinetic energy per unit mass is in fact the Lagrangian of geodesic motion in some metric:

$$T/m = \frac{1}{2}g_{ij}(\vec{q})\dot{q}^i\dot{q}^j, \tag{9.10.2}$$

$$g_{ij}(\vec{q}) = \frac{\partial\vec{z}}{\partial q^i}\cdot\frac{\partial\vec{z}}{\partial q^j}. \tag{9.10.3}$$

If the particle also experiences a potential energy per unit mass of U, the corresponding Lagrangian L_0 will then be

$$L_0 = \frac{1}{2}g_{ij}(\vec{q})\dot{q}^i\dot{q}^j - U(\vec{q}); \tag{9.10.4}$$

where U is to be treated as a scalar under spatial coordinate transformations. As long as U depends only on the $\vec{q}$ and not on its derivatives, Newton's second law – acceleration equals negative gradient of the potential energy per unit mass – arises from applying the Euler-Lagrange equations:

$$\frac{D^2q^i}{\mathrm{d}t^2} \equiv \ddot{q}^i + \Gamma^i{}_{ab}\dot{q}^a\dot{q}^b = -\nabla^i U. \tag{9.10.5}$$

Constraints If we further subject our point mass to N scalar constraints of the form

$$G_{\mathrm{I}}(\vec{q}) = 0, \tag{9.10.6}$$

for $\mathrm{I} = 1, 2, \dots, N$; the modified Lagrangian now reads instead

$$L = \frac{1}{2}g_{ij}(\vec{q})\dot{q}^i\dot{q}^j - U(\vec{q}) - \Lambda^{\mathrm{I}}(t)G_{\mathrm{I}}(\vec{q}). \tag{9.10.7}$$

In $D-$space, 1 constraint would define a $D-1$ dimensional hypersurface; 2 constraints would define a $D-2$ dimensional hypersurface; and so on, until we have $D-1$ constraints defining a 1D line. Any more constraints than $D-1$ would be overly restrictive. For the case of N constraints, we have a $D-N$ dimensional surface.

We may apply the Euler-Lagrange equations to eq. (9.10.7).

$$\frac{D^2q^i}{\mathrm{d}t^2} = -\nabla^i\left(U(\vec{q}) + \lambda^{\mathrm{I}}(t)G_{\mathrm{I}}(\vec{q})\right); \tag{9.10.8}$$

These D equations are then to be solved together with the N ones in (9.10.6); as well as $2D$ appropriate boundary conditions. As long as U and $\{G_{\mathrm{I}}\}$ do not depend on time, the total conserved energy per unit mass E/m of the particle on the $(D-N)-$surface is

$$\frac{E}{m} = \left(\frac{1}{2}g_{ij}(\vec{q})\dot{q}^i\dot{q}^j + U(\vec{q})\right)_{\{G_{\mathrm{I}}=0\}} . \tag{9.10.9}$$

Now, suppose it possible to erect a coordinate system enveloping our $(D-N)-$surface, such that $\vec{x}_\parallel$ are 'parallel' to it and $\vec{x}_\perp$ are 'perpendicular'. The surface itself is parametrized as the $\vec{x}_\perp \equiv \vec{x}_{\perp,0}$ (constant) surface:

$$\vec{q}\left(\vec{x}_\parallel, \vec{x}_\perp\right) = \vec{q}\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right). \tag{9.10.10}$$

This $(D-N)-$surface must be a simultaneous solution of the 'equi-potential' conditions

$$G_{\mathrm{I}}(\vec{x}_\parallel, \vec{x}_{\perp,0}) = 0, \qquad\qquad \mathrm{I} = 1, \dots, N; \tag{9.10.11}$$

so we must have

$$\partial_{x^i_\parallel} G_{\mathrm{I}}\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right) = 0. \tag{9.10.12}$$

By assumption, our coordinate system is orthogonal, so the metric now reads

$$g_{ij}\mathrm{d}x^i\mathrm{d}x^j = g^\parallel_{ij}(\vec{x}_\parallel, \vec{x}_\perp)\mathrm{d}x^i_\parallel\mathrm{d}x^j_\parallel + g^\perp_{ij}(\vec{x}_\parallel, \vec{x}_\perp)\mathrm{d}x^i_\perp\mathrm{d}x^j_\perp \tag{9.10.13}$$

and the Lagrangian in eq. (9.10.7) becomes

$$L = \frac{1}{2}g^\parallel_{ij}(\vec{x}_\parallel, \vec{x}_\perp)\dot{x}^i_\parallel\dot{x}^j_\parallel + \frac{1}{2}g^\perp_{ij}(\vec{x}_\parallel, \vec{x}_\perp)\dot{x}^i_\perp\dot{x}^j_\perp - U(\vec{x}_\parallel, \vec{x}_\perp) - \Lambda^{\mathrm{I}}(t)G_{\mathrm{I}}(\vec{x}_\parallel, \vec{x}_\perp). \tag{9.10.14}$$

If we obtain from eq. (9.10.14) its Euler-Lagrangian equations, then proceed to apply the constraints in eq. (9.10.11), all the $\dot{\vec{x}}_\perp$, $\ddot{\vec{x}}_\perp$ and higher derivatives would vanish. For instance, the only first derivatives in $D^2(x_\parallel, x_\perp)^i/\mathrm{d}t^2$ are the $\dot{\vec{x}}_\parallel$s. We will discover that, for i evaluated on the $x^i_\parallel$ component,

$$\Gamma^{i_\parallel}{}_{ab}\dot{x}^a_\parallel\dot{x}^b_\parallel = \Gamma^i{}_{ab}[g_\parallel]\dot{x}^a_\parallel\dot{x}^b_\parallel, \tag{9.10.15}$$

where $\Gamma^i{}_{ab}[g_\parallel]$ is the Christoffel symbol built entirely out of $g^\parallel_{ab}$ evaluated on the constraint hypersurface:

$$\Gamma^i{}_{ab}[g_\parallel] = \frac{1}{2}g^{ij}_\parallel\left(\partial_{x^a_\parallel}g^\parallel_{bj} + \partial_{x^b_\parallel}g^\parallel_{aj} - \partial_{x^j_\parallel}g^\parallel_{ab}\right)_{\vec{x}_\perp=\vec{x}_{\perp,0}} . \tag{9.10.16}$$

Also, for i evaluated on the $x^i_\perp$ component,

$$\Gamma^{i_\perp}{}_{ab}\dot{x}^a_\parallel\dot{x}^b_\parallel = -\frac{1}{2}g^{ij}_\perp\partial_{x^j_\perp}g^\parallel_{ab}\cdot\dot{x}^a_\parallel\dot{x}^b_\parallel. \tag{9.10.17}$$

Collecting these results, eq. (9.10.8) is transformed into parallel-to-$(D-N)-$surface components

$$\frac{D^2 x_\parallel^i}{\mathrm{d}t^2} = \ddot{x}_\parallel^i + \Gamma^i{}_{ab}[g_\parallel]\dot{x}_\parallel^a \dot{x}_\parallel^b = -g_\parallel^{ij}\partial_{x_\parallel^j} U\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right); \tag{9.10.18}$$

and perpendicular-to-$(D-N)-$surface components

$$\frac{D^2 x_\perp^i}{\mathrm{d}t^2} = -\frac{1}{2} g_\perp^{ij}\partial_{x_\perp^j} g_{ab}^\parallel \cdot \dot{x}_\parallel^a \dot{x}_\parallel^b = -g_\perp^{ij}\partial_{x_\perp^j}\left(U\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right) + \Lambda^{\mathrm{I}} G_{\mathrm{I}}\left(\vec{x}_\parallel, \vec{x}_\perp = \vec{x}_{\perp,0}\right)\right). \tag{9.10.19}$$

This tells us the normal force N^i per unit mass is

$$N^i = -\Lambda^{\mathrm{I}}\nabla^{i_\perp} G_{\mathrm{I}} = g_\perp^{ij}\partial_{x_\perp^j} U\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right) - \frac{1}{2} g_\perp^{ij}\partial_{x_\perp^j} g_{ab}^\parallel\left(\vec{x}_\parallel, \vec{x}_\perp = \vec{x}_{\perp,0}\right)\cdot \dot{x}_\parallel^a \dot{x}_\parallel^b. \tag{9.10.20}$$

To reiterate the key finding in eq. (9.10.18):

> Geometrically speaking, non-relativistic motion on a constraint hypersurface reduces to accelerated motion driven by an external force projected along it.

Remarks Notice, the parallel components in eq. (9.10.18) are a closed set of equations that may be solved once appropriate boundary conditions are provided. In fact, if we did not need to solve for the normal force in eq. (9.10.20), it suffices then to simply reduce the problem to that encoded by the already-constrained Lagrangian

$$L_c \equiv \frac{1}{2} g_{ab}^\parallel\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right)\dot{x}_\parallel^a \dot{x}_\parallel^b - U\left(\vec{x}_\parallel, \vec{x}_{\perp,0}\right). \tag{9.10.21}$$

If we do wish to obtain the normal force in eq. (9.10.20), observe that once $\vec{x}_\parallel$ is known by solving eq. (9.10.18), the former can be determined right away.

Finally, if the $N \times N$ object $M^i_{\mathrm{I}} \equiv \nabla^{i_\perp} G_{\mathrm{I}}$ in eq. (9.10.20) is invertible, by finding its inverse, we may then compute the individual Lagrange multipliers $\{\Lambda^{\mathrm{I}}\}$.

Problem 9.99. Verify equations (9.10.15) and (9.10.17) by starting from the definition of the Christoffel symbol $\Gamma^i{}_{ab}$ of the full metric. □

Additional problems relating differential geometry and classical mechanics
The following problems assume some familiarity with the material in §(8.3) and §(8.4). In particular, $\{\cdot,\cdot\}$ denotes the Poisson bracket.

Problem 9.100. (Non-)Uniqess of Lagrangian in Higher Dimensions Prove that the Lagrangian $L(\lambda, \vec{q}, \dot{\vec{q}})$ is unique up to an additive total derivative. Let $\vec{q}$ reside in arbitrary dimensions D. Specifically, if L_1 and L_2 give the same equations

$$\frac{\partial L_1}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_1}{\partial \dot{q}^i} = \frac{\partial L_2}{\partial q^i} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L_2}{\partial \dot{q}^i}; \tag{9.10.22}$$

then

$$L_1 - L_2 \equiv \Delta L = \frac{\mathrm{d}}{\mathrm{d}t} F\left(\lambda, \vec{q}\right). \tag{9.10.23}$$

This F depends only on λ and $\vec{q}$, but not on $\dot{\vec{q}}$ or its higher derivatives.

Hint: You may argue that, upon applying the Euler-Lagrange operator on ΔL, the coefficients of $\ddot{q}^i$ and $\dot{q}^i$ must be zero for all $i = 1, \dots, D$. □

Problem 9.101. Canonical Transformations In $D-$dimensions, consider the following '1-form' built out of the pairs of generalized coordinates $\vec{q}$ and $\vec{Q}$; and its conjugate momentum $\vec{p}$ and $\vec{P}$:

$$\Delta_{\mathrm{H}} \equiv p_i \mathrm{d}q^i - P_i \mathrm{d}Q^i. \tag{9.10.24}$$

Prove that it is a pure gradient, namely

$$\Delta_{\mathrm{H}} = \partial_{q^i}\Sigma(\vec{q},\vec{p})\mathrm{d}q^i + \partial_{p_i}\Sigma(\vec{q},\vec{p})\mathrm{d}p_i = \mathrm{d}\Sigma, \tag{9.10.25}$$

iff the Poisson bracket in eq. (8.4.43) is satisfied.

Hint: This problem is the $D > 1$ generalization of the $D = 1$ analysis (using the 2D Poincaré lemma) performed after Problem (8.35). □

Problem 9.102. Infinitesimal Canonical Transformations Prove that the most general infinitesimal canonical transformation

$$(\vec{q},\vec{p}) \to (\vec{Q},\vec{P}) \equiv (\vec{q}+\delta\vec{q},\vec{p}+\delta\vec{p}); \tag{9.10.26}$$

is given by

$$\delta q^i = \partial_{p_i}\left(A(\vec{q},\vec{p}) + C_p(\vec{p})\right), \tag{9.10.27}$$

$$\delta p_i = -\partial_{q^i}\left(A(\vec{q},\vec{p}) - C_q(\vec{q})\right); \tag{9.10.28}$$

Hints: $\{Q^i, Q^j\} = 0$ plus the Poincaré lemma should allow you to deduce $\delta q^i = \partial_{p_i}\Phi(\vec{q},\vec{p})$ for some arbitrary Φ; and similarly $\{P_i, P_j\} = 0$ should lead you to $\delta p_i = \partial_{q^i}\Psi$ for some arbitrary Ψ. Finally, $\{Q^i, P_j\} = \delta^i_j$ will relate Φ and Ψ. □

Problem 9.103. Volume form in phase space In the Hamiltonian formulation of classical mechanics, the state of the system is described by its generalized coordinate $\vec{q}$ and the corresponding conjugate momentum $\vec{p}$. We may define its corresponding infinitesimal phase space volume to be

$$\mathrm{d}^{2D}\widetilde{V} \equiv \prod_{i=1}^{D} \mathrm{d}q^i \wedge \mathrm{d}p_i. \tag{9.10.29}$$

Prove that $\mathrm{d}^{2D}\widetilde{V}$ is invariant (up to an overall sign) – namely,

$$\prod_{i=1}^{D} \mathrm{d}q^i \wedge \mathrm{d}p_i = (\pm)\prod_{i=1}^{D} \mathrm{d}Q^i \wedge \mathrm{d}P_i \tag{9.10.30}$$

– under canonical transformations

$$\vec{r} \equiv (\vec{q},\vec{p}) \to \vec{r}\left(\vec{R} \equiv (\vec{Q},\vec{P})\right) \tag{9.10.31}$$

obeying eq. (8.4.33).

By exploiting the results from Problem (8.36), explain why $\mathrm{d}^{2D}\widetilde{V}$ remains the same over an infinitesimal time evolution $t \to t + \mathrm{d}t$.

Hint: You may find Problem (3.8) useful in showing

$$(\det \partial\vec{R}/\partial\vec{r})^2 = (\det\{Q^i, P_j\})^2 = 1, \tag{9.10.32}$$

where $\{\cdot,\cdot\}$ are Poisson brackets. □

10 Differential Geometry of Flat Spacetimes

In this and the following chapter, we move on to differential geometry in flat and curved space*time*s. It would be help the reader to be somewhat familiar with the discussion in §(9) – in many instances, I will simply bring over the results from there to the spacetime context. In this chapter, I begin with a discussion of Lorentz/Poincaré symmetry in flat spacetime, since it is fundamental to both Special and General Relativity.

10.1 Independence of c under change of inertial frames, Geodesics

Spacetime Metric The key difference between Newtonian classical dynamics and Special Relativistic ones is their notion of an inertial frame. Special Relativity replaces the absolute time and flat space $g_{ij} = \delta_{ij}$ notion of an inertial frame with flat (aka Minkowski) *spacetime*. In Cartesian coordinates

$$x^\mu \equiv (x^0, x^i) \equiv (t, \vec{x}) \tag{10.1.1}$$

the 'square' of the distance between x^μ and $x^\mu + \mathrm{d}x^\mu$, oftentimes dubbed the 'line element' – the analog of the flat space Pythagorean theorem $\mathrm{d}\ell^2 = \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}$ – is:

$$\begin{aligned} \mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu &= (\mathrm{d}x^0)^2 - \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x} \\ &= (\mathrm{d}t)^2 - \delta_{ij}\mathrm{d}x^i \mathrm{d}x^j; \end{aligned} \tag{10.1.2}$$

where the coefficients of the quadratic forms is dubbed the Minkowski metric tensor, which in turn reads

$$\eta_{\mu\nu} \doteq \mathrm{diag}[1, -1, \dots, -1]. \tag{10.1.3}$$

Unlike the usual Pythagoras' theorem, we see that the 'square' of the infinitesimal spacetime distance can be either positive $\mathrm{d}s^2 > 0$, when $\mathrm{d}t^2 > \mathrm{d}\vec{x}^2$ ('timelike'); negative $\mathrm{d}s^2 < 0$, when $\mathrm{d}t^2 < \mathrm{d}\vec{x}^2$ ('spacelike'); or zero $\mathrm{d}s^2 = 0$, when $\mathrm{d}t^2 = \mathrm{d}\vec{x}^2$ (null). We will witness the consequences of this indefinite metric throughout the rest of this book.

The inverse metric $\eta^{\mu\nu}$ is simply the matrix inverse, $\eta^{\alpha\sigma}\eta_{\sigma\beta} = \delta^\alpha_\beta$; it is numerically equal to the flat metric itself:

$$\eta^{\mu\nu} \doteq \mathrm{diag}\left[1, -1, \dots, -1\right], \tag{10.1.4}$$

$$g^{\mu\nu}\partial_\mu \otimes \partial_\nu = \eta^{\mu\nu}\partial_\mu \otimes \partial_\nu \tag{10.1.5}$$

$$= \partial_t \otimes \partial_t - \delta^{ij}\partial_i \otimes \partial_j. \tag{10.1.6}$$

Strictly speaking we should be writing eq. (10.1.2) in the 'dimensionally-correct' form

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}; \tag{10.1.7}$$

where c is the speed of light and $[\mathrm{d}s^2] = \text{Length}^2$; the Cartesian coordinates themselves should read $x^\mu = (ct, \vec{x})$. The derivatives ought to read

$$\partial_0 \equiv \frac{1}{c}\frac{\partial}{\partial t} \qquad \text{and} \qquad \partial_i \equiv \frac{\partial}{\partial x^i}. \tag{10.1.8}$$

However, as explained in §(D), since the speed of light shows up frequently in relativity and gravitational physics, it is often advantageous to set $c = 1$, which in turn means all speeds are measured using c as the base unit. ($v = 0.23$ would mean $v = 0.23c$, for instance.) We shall do so throughout this section.

Notice too, we have switched from Latin/English alphabets in §(9), say

$$i, j, k, \dots \in \{1, 2, \dots, D\} \tag{10.1.9}$$

to Greek ones

$$\mu, \nu, \dots \in \{0, 1, 2, \dots, D \equiv d - 1\}; \tag{10.1.10}$$

the former run over the spatial coordinates while the latter over time (0th) and space $(1, \dots, D)$. Also note that the opposite 'mostly plus' sign convention $\eta_{\mu\nu} = \text{diag}[-1, +1, \dots, +1]$ is equally valid and, in fact, more popular in the contemporary physics literature.

Constancy of c One of the primary motivations that led Einstein to recognize eq. (10.1.2) as the proper geometric setting to describe physics, is the realization that the speed of light c is constant in all inertial frames. In modern physics, the latter is viewed as a consequence of spacetime translation and Lorentz symmetry, as well as the null character of the trajectories swept out by photons. Specifically, we shall soon derive below, the most general coordinate transformation $x = x(x')$ the preserves the form of the metric, namely

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \eta_{\mu\nu}\mathrm{d}x'^\mu \mathrm{d}x'^\nu, \tag{10.1.11}$$

is given by

$$x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu \tag{10.1.12}$$

for transformation matrices $\{\Lambda^\alpha{}_\beta\}$ satisfying

$$\Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu \eta_{\alpha\beta} = \eta_{\mu\nu}, \tag{10.1.13}$$

and constant vectors $\{a^\mu\}$. These Λ's are known as Lorentz transformations – consisting of spatial rotations and parity flips, time reversals, as well as Lorentz boosts that mix space and time coordinates – while the constant a^μ describe translations in spacetime. The physical interpretation is that the frames parametrized by $\{x^\mu = (t, \vec{x})\}$ and $\{x'^\mu = (t', \vec{x}')\}$ are *inertial* frames. We shall witness below, compact bodies with no external forces acting on them will sweep out (straight line) geodesics. To interpret physical phenomenon taking place in one frame from the other frame's perspective, one would first have to figure out how to relate x and x'. Moreover, because both time and space are no longer 'absolute' when switching between inertial frames, it is important to highlight that both the space and time coordinates – aka the spacetime event – are required to properly specify the location of a material body in a given inertial frame. The failure to do so, together with the continued but misguided adherence to the Galilean intuition of Newtonian mechanics, is what gives rise to the apparent 'paradoxes' of Special Relativity.

Within the flat spacetime geometry of eq. (10.1.11), the photon is further postulated to follow null geodesics $\mathrm{d}s^2 = 0$. Note that, once $\mathrm{d}s^2 = 0$ in some coordinate system $\{x^\mu\}$, it must hold in some other coordinate system $\{x'^\mu\}$ related by eq. (10.1.12). Let x^μ and x'^μ be the spacetime Cartesian coordinates of a single photon in these two different Lorentz frames. Invoking its null character, we have $(\mathrm{d}x^0)^2 = \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}$ and $(\mathrm{d}x'^0)^2 = \mathrm{d}\vec{x}' \cdot \mathrm{d}\vec{x}'$. This in turn tells us the speeds in both frames are unity:

$$\frac{|\mathrm{d}\vec{x}|}{\mathrm{d}x^0} = \frac{|\mathrm{d}\vec{x}'|}{\mathrm{d}x'^0} = 1. \tag{10.1.14}$$

A more thorough (and hence deeper) justification would be to recognize, it is the sign difference between the 'time' part and the 'space' part of the metric in eq. (10.1.2) – together with its Lorentz invariance – that gives rise to the wave equations obeyed by the photon. Equation (10.1.14) then follows as a consequence for its wavefront. More generally, the main tenets of Special Relativity are as follows.

> **Special Relativity** Ignoring the effects of gravity, fundamental physical laws must remain the same no matter the inertial frame $\{x^\mu\}$, whose geometry is in turn described by the interval $\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$. More precisely, the equations of fundamental physics must take the same form when switching from one arbitrary inertial frame $\{x^\mu\}$ to another $\{x'^\mu\}$−frame related via the spacetime symmetry transformations of eq. (10.1.12). □

Extremal Intervals Just as extremized lengths $\mathrm{d}\ell = \int \sqrt{\mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}}$ yield straight lines in flat space, let us witness that extremizing the interval s yields a straight line in spacetime. For a given path in spacetime $z^\mu(\lambda_0 \leq \lambda \leq \lambda_1)$ such that its end points are fixed

$$z^\mu(\lambda = \lambda_0) = x' \qquad \text{and} \qquad z^\mu(\lambda = \lambda_1) = x; \tag{10.1.15}$$

we may compute the net interval via the integral

$$s = \int_{\lambda_0}^{\lambda_1} \mathrm{d}\lambda |\eta_{\mu\nu}\dot{z}^\mu \dot{z}^\nu|^{1/2}, \qquad \dot{z}^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\lambda. \tag{10.1.16}$$

(Note that this definition is in fact independent of the choice of the 1D coordinate λ; namely, if we transform $\lambda = \lambda(\lambda')$, the above formula would take exactly the same form, with $\lambda \to \lambda'$.) By extremizing s; applying the Euler-Lagrange equations we find that

$$\frac{1}{|\eta_{\alpha\beta}\dot{z}^\alpha \dot{z}^\beta|^{1/2}} \frac{\mathrm{d}U^\mu}{\mathrm{d}\lambda} = 0 \tag{10.1.17}$$

where we have defined the covariant 'd−velocity'

$$U^\mu \equiv \frac{\dot{z}^\mu}{|\eta_{\sigma\rho}\dot{z}^\sigma \dot{z}^\rho|^{1/2}}. \tag{10.1.18}$$

In differential geometry, paths that extremize length or the interval are known as *geodesics*. In flat space(time), geodesics are straight lines.

Problem 10.1. Verify that the d−velocity in eq. (10.1.18) remains the same under a change of the 1D coordinate, $\lambda = \lambda(\lambda')$. □

Problem 10.2. Straight Lines Argue that it is always possible to find a 1D coordinate λ' such that the integrand in eq. (10.1.16) can be rendered a constant C; in particular,

$$\mathrm{d}\lambda' C = \mathrm{d}\lambda |\eta_{\mu\nu}\dot{z}^\mu(\lambda)\dot{z}^\nu(\lambda)|^{1/2}. \tag{10.1.19}$$

This sort of 1D coordinate that renders the square root factor, and hence $\eta_{\mu\nu}\dot{z}^\mu(\lambda)\dot{z}^\nu(\lambda)$, constant is known as an *affine parameter*; and (up to constant re-scalings) for timelike $\dot{z}^\mu$ would turn out to be its proper time and for spacelike $\dot{z}^\mu$ its proper length. (Hint: Derive the relevant differential equation.) Explain why that implies eq. (10.1.17) can always be transformed into

$$\frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\lambda'^2} = 0. \tag{10.1.20}$$

By incorporating the boundary conditions in eq. (10.1.15), deduce that the extremized interval yields a straight line joining $x \leftrightarrow x'$; namely

$$z^\mu(\lambda'_0 \leq \lambda' \leq \lambda'_1) = x'^\mu + \frac{\lambda' - \lambda'_0}{\lambda'_1 - \lambda'_0}(x^\mu - x'^\mu) \tag{10.1.21}$$

and associated interval is

$$s(x \leftrightarrow x') = \left|(x - x')^2\right|^{1/2}, \tag{10.1.22}$$

$$(x - x')^2 \equiv \eta_{\alpha\beta}(x^\alpha - x'^\alpha)(x^\beta - x'^\beta). \tag{10.1.23}$$

Explain why, if λ were not affine, the solution reads instead

$$z^\mu(\lambda_0 \leq \lambda \leq \lambda_1) = x'^\mu + f(\lambda)(x^\mu - x'^\mu), \tag{10.1.24}$$

where f is an arbitrary monotonically increasing function of λ subject to the boundary conditions $f(\lambda_0) = 0$ and $f(\lambda_1) = 1$. That the solution for $z^\mu(\lambda)$ depends on an arbitrary function f, despite being an ordinary differential equation (albeit a nonlinear one in eq. (10.1.17)), is because of the invariance of the action in eq. (10.1.16) under change of the 1D coordinate λ.

Observe that no assumption was made on whether $\dot{z}^\mu$ was timelike ($\eta_{\mu\nu}\dot{z}^\mu\dot{z}^\nu > 0$), spacelike ($\eta_{\mu\nu}\dot{z}^\mu\dot{z}^\nu < 0$) or null ($\eta_{\mu\nu}\dot{z}^\mu\dot{z}^\nu = 0$). On the other hand, as we shall argue below, material bodies sweep out timelike worldlines z^μ; this $\mathrm{d}z^\mu/\mathrm{d}\lambda'^2 = 0$, when applied to such cases, can be viewed as the force-free 'Newton's second law' in Special Relativity. Moreover, the $\lambda' = \tau$ may be identified with the proper time of the material body. □

Synge World Function In the General Relativity literature, half the 'square' of the geodesic distance linking $x \leftrightarrow x'$ is known as Sygne's world function $\sigma(x, x')$. In the Minkowski spacetime of Special Relativity, and in terms of the affine parameter λ', it is given by the integral

$$\sigma(x, x') = \frac{1}{2}(\lambda'_1 - \lambda'_0)\int_{\lambda'_0}^{\lambda'_1} \eta_{\mu\nu}\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda'}\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda'}\mathrm{d}\lambda'. \tag{10.1.25}$$

Extremizing $\sigma(x, x')$ for fixed end points x and x' readily yields eq. (10.1.20). According to eq. (10.1.21),

$$\dot{z}^\mu = \frac{x^\mu - x'^\mu}{\lambda'_1 - \lambda'_0}. \tag{10.1.26}$$

This returns – compare against eq. (10.1.22) –

$$\sigma(x, x') = \frac{1}{2}(x - x')^2 = \frac{1}{2}\left((t - t')^2 - |\vec{x} - \vec{x}'|^2\right), \tag{10.1.27}$$

where $x^0 \equiv t$ and $x'^0 = t'$; confirming that $\sigma(x, x')$ is indeed the 'square' of the interval.

Cartesian Spacetime Displacement Vectors

10.2 Poincaré Symmetry and Lorentz Tensors

Spacetime Symmetries We have defined Poincaré transformations[102] $x(x')$ to be the set of all coordinate transformations that leave the flat spacetime metric invariant (cf. eq. (10.1.11)). Poincaré and Lorentz symmetries play fundamental roles in our understanding of both classical relativistic physics and quantum theories of elementary particle interactions; hence, this motivates us to *derive* it. As we will now proceed to demonstrate, the most general invertible Poincaré transformation is in fact the one in eq. (10.1.12). [103]

Now, under a coordinate transformation, eq. (10.1.11) reads

$$\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}\mathrm{d}x'^\alpha \mathrm{d}x'^\beta = \eta_{\alpha'\beta'}\mathrm{d}x'^\alpha \mathrm{d}x'^\beta. \tag{10.2.1}$$

Let us differentiate both sides of eq. (10.2.1) with respect to x'^σ.

$$\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} + \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial^2 x^\nu}{\partial x'^\sigma \partial x'^\beta} = 0. \tag{10.2.2}$$

Next, consider symmetrizing $\sigma\alpha$ and anti-symmetrizing $\sigma\beta$.

$$2\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} + \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial^2 x^\nu}{\partial x'^\sigma \partial x'^\beta} + \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\sigma}\frac{\partial^2 x^\nu}{\partial x'^\alpha \partial x'^\beta} = 0 \tag{10.2.3}$$

$$\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} - \eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\beta \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\sigma} = 0 \tag{10.2.4}$$

Since partial derivatives commute, the second term from the left of eq. (10.2.2) vanishes upon anti-symmetrization of $\sigma\beta$. Adding equations (10.2.3) and (10.2.4) hands us

$$3\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} + \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial^2 x^\nu}{\partial x'^\sigma \partial x'^\beta} = 0. \tag{10.2.5}$$

Finally, subtracting eq. (10.2.2) from eq. (10.2.5) produces

$$2\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} = 0. \tag{10.2.6}$$

Because we have assumed Poincaré transformations are invertible, we may contract both sides with $\partial x'^\beta / \partial x^\kappa$.

$$\eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}\frac{\partial x'^\beta}{\partial x^\kappa} = \eta_{\mu\nu}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha}\delta^\nu_\kappa = 0. \tag{10.2.7}$$

[102]Poincaré transformations are also sometimes known as inhomogeneous Lorentz transformations.

[103]This argument can be found in Weinberg [24].

Finally, we contract both sides with $\eta^{\kappa\rho}$:

$$\eta_{\mu\kappa}\eta^{\kappa\rho}\frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} = \frac{\partial^2 x^\rho}{\partial x'^\sigma \partial x'^\alpha} = 0. \tag{10.2.8}$$

In words: since the second x'-derivative of x has to vanish, the transformation from x to x' can at most go linearly as x'; it cannot involve higher powers of x'. This implies the form in eq. (10.1.12). Plugging eq. (10.1.12) the latter into eq. (10.2.1), we recover the necessary definition of the Lorentz transformation in eq. (10.1.13).

> **Poincaré Transformations** The most general invertible coordinate transformations that leave the Cartesian Minkowski metric invariant involve the spacetime-constant Lorentz transformations $\{\Lambda^\mu{}_\alpha\}$ of eq (10.1.13) plus constant spacetime translations.

(Homogeneous) Lorentz Transformations form a Group[104] If $\Lambda^\mu{}_\alpha$ and $\Lambda'^\mu{}_\alpha$ denotes different Lorentz transformations, then notice the composition

$$\Lambda''^\mu{}_\alpha \equiv \Lambda^\mu{}_\sigma \Lambda'^\sigma{}_\alpha \tag{10.2.9}$$

is also a Lorentz transformation. For, keeping in mind the fundamental definition in eq. (10.1.13), we may directly compute

$$\begin{aligned}\Lambda''^\mu{}_\alpha \Lambda''^\nu{}_\beta \eta_{\mu\nu} &= \Lambda^\mu{}_\sigma \Lambda'^\sigma{}_\alpha \Lambda^\nu{}_\rho \Lambda'^\rho{}_\beta \eta_{\mu\nu} \\ &= \Lambda'^\sigma{}_\alpha \Lambda'^\rho{}_\beta \eta_{\sigma\rho} = \eta_{\alpha\beta}.\end{aligned} \tag{10.2.10}$$

To summarize:

> The set of all Lorentz transformations $\{\Lambda^\mu{}_\alpha\}$ satisfying eq. (10.1.13), together with the composition law in eq. (10.2.9) for defining successive Lorentz transformations, form a *Group*.

Proof Let $\Lambda^\mu{}_\alpha$, $\Lambda'^\mu{}_\alpha$ and $\Lambda''^\mu{}_\alpha$ denote distinct Lorentz transformations.

- *Closure* Above, we have just verified that applying successive Lorentz transformations yields another Lorentz transformation; for e.g., $\Lambda^\mu{}_\sigma \Lambda'^\sigma{}_\nu$ and $\Lambda^\mu{}_\sigma \Lambda'^\sigma{}_\rho \Lambda''^\rho{}_\nu$ are Lorentz transformations.

- *Associativity* Because applying successive Lorentz transformations amount to matrix multiplication, and since the latter is associative, that means Lorentz transformations are associative:

$$\Lambda \cdot \Lambda' \cdot \Lambda'' = \Lambda \cdot (\Lambda' \cdot \Lambda'') = (\Lambda \cdot \Lambda') \cdot \Lambda''. \tag{10.2.11}$$

- *Identity* $\delta^\mu{}_\alpha$ is the identity Lorentz transformation:

$$\delta^\mu{}_\sigma \Lambda^\sigma{}_\nu = \Lambda^\mu{}_\sigma \delta^\sigma{}_\nu = \Lambda^\mu{}_\nu, \tag{10.2.12}$$

  and

$$\delta^\mu{}_\alpha \delta^\nu{}_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}. \tag{10.2.13}$$

[104] Refer to §(B) for the defining axioms of a Group.

Problem 10.3. Explain why eq. (10.1.13) is equivalent to the matrix equation

$$\Lambda^{\mathrm{T}} \cdot \eta \cdot \Lambda = \eta. \tag{10.2.14}$$

Hint: What are $\eta_{\mu\nu}\Lambda^{\nu}{}_{\beta}$ and $A^{\nu}{}_{\beta}B_{\nu\gamma}$ in matrix notation? □

- *Inverse* Let us take the determinant of both sides of eq. (10.2.14), the matrix equivalent of eq. (10.1.13). Recalling $\det A^{\mathrm{T}} = \det A$ for any matrix A,

$$(\det \Lambda)^2 = 1 \qquad \Rightarrow \qquad \det \Lambda = \pm 1. \tag{10.2.15}$$

Since the determinant of Λ is strictly non-zero, what eq. (10.2.15) teaches us is that Λ is always invertible: Λ^{-1} is guaranteed to exist. What remains is to check that, if Λ is a Lorentz transformation, so is Λ^{-1}. Starting with the matrix form of eq. (10.2.14), and utilizing $(\Lambda^{-1})^{\mathrm{T}} = (\Lambda^{\mathrm{T}})^{-1}$,

$$\Lambda^{\mathrm{T}}\eta\Lambda = \eta \tag{10.2.16}$$

$$(\Lambda^{\mathrm{T}})^{-1}\Lambda^{\mathrm{T}}\eta\Lambda\Lambda^{-1} = (\Lambda^{\mathrm{T}})^{-1} \cdot \eta \cdot \Lambda^{-1} \tag{10.2.17}$$

$$\eta = (\Lambda^{-1})^{\mathrm{T}} \cdot \eta \cdot \Lambda^{-1}. \tag{10.2.18}$$

Problem 10.4. (d+1) × (d+1) Poincaré Transformation Matrix Prove that the Poincaré transformation in eq. (10.1.12) also defines a group. To systemize the discussion, first promote the spacetime coordinates to $d+1$ dimensional objects:

$$x^{\mathfrak{A}} = (x^\mu, 1)^{\mathrm{T}} \qquad \text{and} \qquad x'^{\mathfrak{A}} = (x'^\mu, 1)^{\mathrm{T}}, \tag{10.2.19}$$

with $\mathfrak{A} = 0, 1, 2, \dots, d-1, d$. Then define the matrix

$$\Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a] = \begin{bmatrix} \Lambda^\mu{}_\nu & a^\mu \\ 0\dots 0 & 1 \end{bmatrix}; \tag{10.2.20}$$

namely, its upper left $d \times d$ block is simply the Lorentz transformation $\Lambda^\mu{}_\nu$; while its rightmost column is $(a^\mu, 1)^{\mathrm{T}}$ and its bottom row is $(0 \dots 0\ 1)$.

First check that $x^{\mathfrak{A}} = \Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a]x'^{\mathfrak{B}}$ is equivalent to eq. (10.1.12). You should also find that

$$\Pi[\Lambda_1, a_1] \cdot \Pi[\Lambda_2, a_2] = \Pi[\Lambda_1 \cdot \Lambda_2, \Lambda_1 \cdot a_2 + a_1] \tag{10.2.21}$$

and therefore its inverse is

$$\Pi[\Lambda, a]^{-1} = \Pi[\Lambda^{-1}, -\Lambda^{-1} \cdot a]. \tag{10.2.22}$$

Next, we observe that the full Poincaré transformation may be implemented in two steps, first by a Lorentz transformation, followed by a spacetime translation:

$$x^\mu \to \Lambda^\mu{}_\nu x^\nu \to \Lambda^\mu{}_\nu x^\nu + a^\mu. \tag{10.2.23}$$

Verify that the general element is indeed a Lorentz transformation followed by a spacetime translation:

$$\Pi[\mathbb{I}, a] \cdot \Pi[\Lambda, 0] = \Pi[\Lambda, a]. \tag{10.2.24}$$

Then proceed to explain why these set of matrices $\{\Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a]\}$ for different Lorentz transformations Λ and translation vectors a, with the usual rules of matrix multiplication, together define a group. □

Vector fields and Lorentzian 'inner product' We have successfully proven that, the form of the metric (or, 'interval') in eq. (10.1.11) is preserved iff the inertial frames x and x' are related via the Poincaré transformations in equations (10.1.12) and (10.1.13). Along similar lines we may readily conclude that

$$\dot{z}^2 \equiv \eta_{\alpha\beta}\dot{z}^\alpha\dot{z}^\beta = \eta_{\alpha\beta}\Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu\dot{z}'^\mu\dot{z}'^\nu = \eta_{\alpha\beta}\dot{z}'^\mu\dot{z}'^\nu \tag{10.2.25}$$

is also Lorentz invariant. Now, a given trajectory $z^\mu(\lambda)$ in spacetime may always be associated with a vector field by simply defining $v^\mu(z) = \mathrm{d}z^\mu/\mathrm{d}\lambda$. On the other hand, given a vector field in spacetime $v^\mu(x)$, we may in turn define $\dot{z}^\mu$ at an arbitrary location $x \equiv z(\lambda)$ via the same relation $v^\mu(z) = \mathrm{d}z^\mu/\mathrm{d}\lambda$; allowing us to integrate this first order ODE to obtain the trajectory $z(\lambda)$. Hence, by interpreting λ as a fictitious time, we may always identify a vector field v^μ as *tangent* to a flow of some kind.

For two distinct vectors v and w, we may further view the Lorentz transformation Λ as a linear operator that preserves their *Lorentzian* inner product; if $v' = \Lambda \cdot v$ and $w' = \Lambda \cdot w$, their inner product yields

$$v' \cdot w' \equiv \eta_{\alpha\beta}v'^\alpha w'^\beta = \eta_{\alpha\beta}\Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu v^\mu w^\nu \tag{10.2.26}$$

$$= \eta_{\mu\nu}v^\mu w^\nu = v \cdot w. \tag{10.2.27}$$

Moreover, because the metric itself is symmetric $\eta_{\alpha\beta} = \eta_{\beta\alpha}$, the inner product is symmetric too:

$$v \cdot w = \eta_{\alpha\beta}v^\alpha w^\beta = \eta_{\beta\alpha}w^\beta v^\alpha = w \cdot v. \tag{10.2.28}$$

This Lorentzian inner product invariance is very much analogous to rotations in $\mathbb{R}^D$ being the linear transformations that preserve the Euclidean inner product between spatial vectors: $\vec{v} \cdot \vec{w} = \vec{v}' \cdot \vec{w}'$ for all $\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I}_{D\times D}$, where $\vec{v}' \equiv \widehat{R} \cdot \vec{v}$ and $\vec{w}' \equiv \widehat{R} \cdot \vec{w}$. Just as the interval $\mathrm{d}s^2$ can be timelike, spacelike or null; so can these tangent vectors v^μ. Since $v^\mu = \mathrm{d}z^\mu/\mathrm{d}\lambda$, we may say

- v^μ is timelike if $v^\mu v^\nu \eta_{\mu\nu} = (v^0)^2 - \vec{v}^2 > 0$, this amounts to $\mathrm{d}s^2 > 0$.
- v^μ is spacelike if $v^\mu v^\nu \eta_{\mu\nu} = (v^0)^2 - \vec{v}^2 < 0$, this amounts to $\mathrm{d}s^2 < 0$.
- v^μ is null if $v^\mu v^\nu \eta_{\mu\nu} = (v^0)^2 - \vec{v}^2 = 0$, this amounts to $\mathrm{d}s^2 = 0$.

As we have just seen $v^2 \equiv v \cdot v$ is invariant under the change of inertial frames – hence, in this sense, the timelike, spacelike, or null character of tangent vectors are in fact intrinsic property of a given v^μ.

Problem 10.5. Inverse Lorentz The inverse Lorentz transformation Λ^{-1} may be constructed explicitly. Moving the indices on the Λ's with the metric, so for example,

$$\Lambda^\mu{}_\alpha\eta_{\mu\nu} = \Lambda_{\nu\alpha}. \tag{10.2.29}$$

First explain how to go from eq. (10.1.13) to

$$\Lambda_\sigma{}^\alpha\Lambda^\sigma{}_\beta = \delta^\alpha{}_\beta \tag{10.2.30}$$

and deduce the inverse Lorentz transformation

$$\left(\Lambda^{-1}\right)^{\alpha}{}_{\beta} = \Lambda_{\beta}{}^{\alpha} = \eta_{\beta\nu}\eta^{\alpha\mu}\Lambda^{\nu}{}_{\mu}. \tag{10.2.31}$$

Alternatively, from eq. (10.2.14), explain why

$$\left(\eta \cdot \Lambda^{\mathrm{T}} \cdot \eta\right) \Lambda = \mathbb{I}_{d\times d}. \tag{10.2.32}$$

This allows the inverse to be read off from the left hand side – can you see why? Throughout this problem, recall the inverse always exists because $\det \Lambda = \pm 1$. $\square$

Problem 10.6. Explain why

$$\Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta}\eta^{\alpha\beta} = \eta^{\mu\nu} \tag{10.2.33}$$

and

$$\Lambda_{\alpha}{}^{\mu}\Lambda_{\beta}{}^{\nu}\eta_{\mu\nu} = \eta_{\alpha\beta}, \qquad \Lambda_{\alpha}{}^{\mu}\Lambda_{\beta}{}^{\nu}\eta^{\alpha\beta} = \eta^{\mu\nu}. \tag{10.2.34}$$

Hint: Start from eq. (10.2.18). $\square$

Jacobians and Spacetime Tensors Note that under the Poincaré transformation in eq. (10.1.12),

$$\frac{\partial x^{\alpha}}{\partial x'^{\beta}} = \Lambda^{\alpha}{}_{\beta}, \tag{10.2.35}$$
$$\frac{\partial x'^{\alpha}}{\partial x^{\beta}} = \Lambda_{\beta}{}^{\alpha}. \tag{10.2.36}$$

This implies

$$\mathrm{d}x^{\alpha} = \Lambda^{\alpha}{}_{\beta}\mathrm{d}x'^{\beta}, \tag{10.2.37}$$
$$\frac{\partial}{\partial x^{\alpha}} \equiv \partial_{\alpha} = \Lambda_{\alpha}{}^{\beta}\partial_{\beta'} \equiv \Lambda_{\alpha}{}^{\beta}\frac{\partial}{\partial x'^{\beta}}; \tag{10.2.38}$$

and the volume measure itself is Lorentz invariant (up to an overall sign)

$$\mathrm{d}^d x = (\det\Lambda)\mathrm{d}^d x' = (\pm)\mathrm{d}^d x'. \tag{10.2.39}$$

Spacetime *scalars* are functions of spacetime that return a unique value f for a given point x in spacetime. Under a change of coordinates $x = x(x')$, i.e., where x and x' refer to the same spacetime point, the scalar field $f(x)$ becomes

$$f'(x') \equiv f(x(x')), \tag{10.2.40}$$

where the prime on the left hand side is not a derivative but merely reminds us that f is now evaluated in the x' coordinate system. On the other hand, spacetime *tensors* are objects involving the superposition of basis tangent vectors $\{\partial_{\mu}\}$ and/or 1−forms $\{\mathrm{d}x^{\mu}\}$; for instance,

$$T \equiv T^{\mu\nu\dots}{}_{\alpha\beta\dots}(x)\partial_{x^{\mu}} \otimes \partial_{x^{\nu}} \otimes \dots \otimes \mathrm{d}x^{\alpha} \otimes \mathrm{d}x^{\beta} \otimes \dots. \tag{10.2.41}$$

(Here and throughout the rest of this book, x is a short-hand for the coordinates x^μ.) For a vector $V^\mu(x)$,

$$V^\alpha(x)\partial_{x^\alpha} = V^\alpha(x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu)\Lambda_\alpha{}^\beta \partial_{x'^\beta} \tag{10.2.42}$$

$$\equiv V'^\beta(x')\partial_{x'^\beta}. \tag{10.2.43}$$

The last line is a definition, in that $V'^\beta(x')$ is the $\beta-$th component of the vector V in the x' coordinate system because it is the coefficient of $\partial_{x'^\beta}$. Using the inverse relationship between $\Lambda_\alpha{}^\beta$ and $\Lambda^\alpha{}_\beta$, we recognize the transformation rule

$$V^\alpha(x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu) = \Lambda^\alpha{}_\beta V'^\beta(x'). \tag{10.2.44}$$

To sum: we view the tensor T, V, etc. as intrinsically meaningful, and hence invariant under change-of-coordinates; whereas it is the *coefficients* of the basis partial derivatives $\{\partial_\mu\}$ and/or infinitesimal displacements $\{\mathrm{d}x^\mu\}$ that define their components in that particular frame.

Problem 10.7. Lorentz 1-Form Show that a $1-$form $A_\mu(x)\mathrm{d}x^\mu$ components transform as

$$A_\alpha(x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu) = \Lambda_\alpha{}^\beta A'_\beta(x'), \tag{10.2.45}$$

where A'_β are the $1-$form components in the $x'-$system. Hint: You should first find that $A_\alpha(x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu)\Lambda^\alpha{}_\beta \mathrm{d}x'^\beta = A'_\beta(x')\mathrm{d}x'^\beta$. □

Higher Rank Tensors By performing a coordinate transformation on their basis tangent vectors and $1-$forms, one may deduce the generalization of equations (10.2.44) and (10.2.45) to the general Lorentz tensor to be

$$T^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}(x = \Lambda\cdot x' + a) = \Lambda^{\mu_1}{}_{\nu_1}\Lambda^{\mu_2}{}_{\nu_2}\dots\Lambda_{\alpha_1}{}^{\beta_1}\Lambda_{\alpha_2}{}^{\beta_2}\dots T'^{\nu_1\nu_2\dots}{}_{\beta_1\beta_2\dots}(x'). \tag{10.2.46}$$

We see that the metric $\eta_{\alpha\beta}$ and its inverse $\eta^{\alpha\beta}$ are not only rank-2 tensors, they are *invariant* ones, reflecting the Poincaré symmetry of the geometry.

Problem 10.8. Superpositions and Tensor Products If $T^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}$ and $U^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}$ are Lorentz tensors and provided they have the same index structure, explain why their superposition

$$A\cdot T^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots} + B\cdot U^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}, \tag{10.2.47}$$

for scalars A and B, forms a Lorentz tensor.

If $T^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}$ and $V^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}$, where their index structures are now independent, explain why their product

$$T^{\mu_1\mu_2\dots}{}_{\alpha_1\alpha_2\dots}\cdot U^{\nu_1\nu_2\dots}{}_{\beta_1\beta_2\dots} \tag{10.2.48}$$

is also a Lorentz tensor. □

Why is the 'Lorentz tensor' an important concept? The key reason is that: requiring the equations of physics to transform as Lorentz tensors ensure they take the same form in all inertial frames in the Special Relativistic sense of Minkowski spacetime. If, say, we know that $A^{\mu\nu}(x) = B^{\mu\nu}(x)$ in some inertial frame $\{x^\mu\}$; and $A^{\mu\nu}$ and $B^{\mu\nu}$ are both Lorentz tensors, we may perform the Poincaré transformation to deduce

$$\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta A'^{\alpha\beta}(x') = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta B'^{\alpha\beta}(x'). \tag{10.2.49}$$

Since these Lorentz transformations are invertible, we may immediately deduce

$$A'^{\alpha\beta}(x') = B'^{\alpha\beta}(x'). \tag{10.2.50}$$

A simple example is provided by the force-free relativistic 'Newton's 2nd law' in eq. (10.1.17). We already know that $\eta_{\mu\nu}\mathrm{d}z^\mu \mathrm{d}z^\nu$ is invariant under change-of-frames $z^\mu = \Lambda^\mu{}_\nu z'^\nu + a^\mu$ (which implies $\mathrm{d}z^\mu = \Lambda^\mu{}_\nu \mathrm{d}z'^\nu$); and therefore so is $\eta_{\mu\nu}\dot{z}^\mu \dot{z}^\nu \equiv \dot{z}^2$. Hence, we must have

$$\frac{1}{|\dot{z}'^2|^{1/2}} \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\dot{z}'^\mu}{|\dot{z}'^2|^{1/2}}\right) = 0. \tag{10.2.51}$$

and the solution transformed from eq. (10.1.24) is

$$z'^\mu(\lambda) = \Lambda^\mu{}_\nu \left(x'^\nu + f(\lambda)(x^\nu - x'^\nu)\right). \tag{10.2.52}$$

The Lorentz covariance of this discussion is, in turn, guaranteed by the Lorentz invariance of the action principle encoded within eq. (10.1.16).

Moving indices Comparing equations (10.2.37) and (10.2.38) as well as equations (10.2.44) and (10.2.45) inform us that upper indices transform 'oppositely' from lower ones. This implies, for instance, that the contraction of two indices forms a Lorentz scalar; for,

$$A_\mu(x) B^\mu(x) = A'_\alpha(x') \Lambda_\mu{}^\alpha \Lambda^\mu{}_\beta B'^\beta(x') \tag{10.2.53}$$
$$= A'_\alpha(x') \delta^\alpha_\beta B'^\beta(x') = A'_\alpha(x') B'^\alpha(x'). \tag{10.2.54}$$

By contracting tensor indices with either the metric tensor $\eta_{\alpha\beta}$ or its inverse $\eta^{\alpha\beta}$, we may not only 'move' tensor indices from up to down or vice versa; we may do so while maintaining their character as Lorentz tensors. For example,

$$v^\mu = \eta^{\mu\nu} v_\nu, \qquad v_\mu = \eta_{\mu\nu} v^\nu; \tag{10.2.55}$$
$$T_{\mu\nu} = \eta_{\mu\alpha}\eta_{\nu\beta} T^{\alpha\beta}, \qquad T^{\mu\nu} = \eta^{\mu\alpha}\eta^{\nu\beta} T_{\alpha\beta}. \tag{10.2.56}$$

Moreover, $v^\mu = \eta^{\mu\nu} v_\nu$ in fact implies $v_\mu = \eta_{\mu\nu} v^\nu$ and vice versa; simply by contracting both sides with the metric or its inverse. Moreover, when contracting a pair of up and down indices, it does not matter which one is up or which one is down:

$$A_\mu B^\mu = A^\nu \eta_{\nu\mu} B^\mu = A^\nu B_\nu = A_\mu \eta^{\mu\nu} B_\nu. \tag{10.2.57}$$

Due to the 'mostly minus' sign convention for $\eta_{\mu\nu}$, the explicit relations in a Cartesian basis are

$$v^0 = v_0 \qquad \text{and} \qquad v^i = -v_i. \tag{10.2.58}$$

It suffices to demonstrate the transformation properties for a rank-1 object, since indices transform independently. If, say, $v_\mu(x)$ is a Lorentz $1-$form, then $v^\mu \equiv \eta^{\mu\nu} v_\nu$ must be a Lorentz vector because

$$v^\mu(x) = \eta^{\mu\nu} \Lambda_\nu{}^\alpha v'_\alpha(x') \tag{10.2.59}$$
$$= \eta^{\mu\nu} \Lambda_\nu{}^\alpha \eta_{\alpha\beta} v'^\beta(x') = \Lambda^\mu{}_\nu v'^\nu(x'). \tag{10.2.60}$$

Problem 10.9. If v^μ is a Lorentz vector, show that $v_\mu = \eta_{\mu\nu} v^\nu$ must transform as a Lorentz $1-$form. $\square$

Problem 10.10. Cartesian Derivatives as Lorentz Tensors By defining

$$\partial^\mu \equiv \eta^{\mu\nu} \partial_\nu, \tag{10.2.61}$$

show that ∂^μ indeed transforms as a Lorentz vector under the Poincaré transformation in eq. (10.1.12). Use this result or otherwise to demonstrate that the scalar wave operator

$$\Box_x \equiv \eta^{\mu\nu} \partial_\mu \partial_\nu \equiv \partial_x^2 \tag{10.2.62}$$

is Lorentz invariant: $\eta^{\mu\nu} \partial_\mu \partial_\nu = \eta^{\mu\nu} \partial_{\mu'} \partial_{\nu'}$, where $\partial_\mu \equiv \partial/\partial x^\mu$ and $\partial_{\mu'} \equiv \partial/\partial x'^\mu$. $\square$

Problem 10.11. Explain why the spacetime 'divergence' $\partial_\mu J^\mu$ is a Lorentz scalar. Many *local* conservation laws – electric charge, fermion number, mass-energy, etc. – are expressed in the Lorentz invariant $\partial_\mu J^\mu = 0$ relation, where J^μ is the 'flow' or 'current' associated with the conserved quantity. $\square$

Lorentz Covariance of Wave Equations Fundamental physics involves field theories, both at the classical and quantum level. For example, in a given inertial frame $\{x^\mu\}$, consider the following *wave equation* involving a scalar field $\varphi(x)$:

$$\eta^{\mu\nu} \partial_\mu \partial_\nu \varphi(x) \equiv \partial_x^2 \varphi(x) = 0. \tag{10.2.63}$$

In terms of the time derivative ∂_t and the spatial Laplacian $\vec{\nabla}^2 = \delta^{ij} \partial_i \partial_j$,

$$\left(\partial_t^2 - \vec{\nabla}_{\vec{x}}^2\right) \varphi(t, \vec{x}) = 0. \tag{10.2.64}$$

If physics were the same in all inertial frames, we expect that in a different inertial frame $\{x'^\mu\}$, with primed indices referring to these primed coordinates (e.g., $\partial_{\mu'} \equiv \partial/\partial x'^\mu$),

$$\eta^{\mu\nu} \partial_{\mu'} \partial_{\nu'} \varphi(x') \equiv \partial_{x'}^2 \varphi(x') = 0, \tag{10.2.65}$$
$$\left(\partial_{t'}^2 - \vec{\nabla}_{\vec{x}'}^2\right) \varphi(t', \vec{x}') = 0. \tag{10.2.66}$$

For the wave equations (10.2.63) and (10.2.65) to be consistent with each other, we must be able to relate them through a Lorentz transformations. Problem (10.10) shows this is indeed the case, because the wave operator itself is Lorentz invariant. Moreover, if $\varphi(x)$ solves the homogeneous wave equation, then

$$\varphi'(x') \equiv \varphi(x^\alpha = \Lambda^\alpha{}_\beta x'^\beta + a^\alpha) \tag{10.2.67}$$

must solve the corresponding wave equation in the x'-frame. There is no need to re-solve the wave equation in the $x'-$frame.

That the wave equation in eq. (10.2.63) is linear – the superposition of solutions is another solution – means we may view

$$\varphi(x) = \int \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \widetilde{\varphi}(k) e^{-ik\cdot x} \tag{10.2.68}$$

to be a superposition of plane waves $\{\exp(-ik \cdot x)\}$, where

$$k \cdot x \equiv k_\mu x^\mu = k_0 t - \delta_{ij} k^i x^j = k_0 t - \vec{k} \cdot \vec{x}; \tag{10.2.69}$$

i.e., it admits a spacetime Fourier decomposition. The first derivative of the plane wave is

$$\partial_\mu e^{-ik\cdot x} = e^{-ik\cdot x} \partial_\mu (-ik_\alpha x^\alpha) = e^{-ik\cdot x}(-ik_\alpha \delta^\alpha_\mu) = e^{-ik\cdot x}(-ik_\mu). \tag{10.2.70}$$

The wave equation for a single basis plane wave therefore returns

$$0 = \partial_x^2 e^{-ik\cdot x} = \partial_\mu \partial^\mu e^{-ik\cdot x} \tag{10.2.71}$$

$$= e^{-ik\cdot x}(-ik_\mu)(-ik^\mu) = e^{-ik\cdot x}(-k^2), \tag{10.2.72}$$

$$k^2 \equiv k_\mu k^\mu = \omega^2 - \vec{k}^2, \qquad \omega \equiv k_0; \tag{10.2.73}$$

where we have employed the customary notation for the zeroth component of the wave vector as angular frequency; so that the time-dependent portion of the solution is $e^{-i\omega t}$. The crucial relative sign difference between the ω^2 term and the $\vec{k}^2 \equiv \vec{k} \cdot \vec{k}$ terms arises because of the Minkowski metric. If all the signs were the same, the only solution would have been $k_\mu = 0$. But, instead, we have the null dispersion relation

$$k^2 = 0 \qquad \Leftrightarrow \qquad \omega^2 = |\vec{k}|^2. \tag{10.2.74}$$

The same basis plane wave solution in the $x'-$frame is

$$e^{-ik\cdot x} = \exp\left(-ik_\mu \left(\Lambda^\mu{}_\nu x'^\nu + a^\mu\right)\right) = e^{-ik'\cdot x'} e^{-ik\cdot a}, \tag{10.2.75}$$

where we have identified the Lorentz-transformed wave vector to be

$$k'_\nu \equiv k_\mu \Lambda^\mu{}_\nu. \tag{10.2.76}$$

The null dispersion relation still holds: $k^2 = 0 = k'^2$. That k_μ needs to be null means we *should not* be integrating over all possible k_μ when performing the Fourier decomposition of φ. Rather, to maintain this null constraint in a Lorentz invariant manner, we shall introduce a Dirac $\delta-$function:

$$\varphi(x) = \int \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} (2\pi)\delta(k^2) \widetilde{\varphi}(k) e^{-ik\cdot x} \tag{10.2.77}$$

$$= \int \frac{\mathrm{d}k_0 \mathrm{d}^D \vec{k}}{(2\pi)^D} \left(\frac{\delta(k_0 - |\vec{k}|)}{2|\vec{k}|} + \frac{\delta(k_0 + |\vec{k}|)}{2|\vec{k}|} \right) \widetilde{\varphi}(k) e^{-ik\cdot x} \tag{10.2.78}$$

$$\equiv \int \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left(\widetilde{\varphi}^+(\vec{k}) e^{i|\vec{k}|\left(\widehat{k}\cdot\vec{x}-t\right)} + \widetilde{\varphi}^-(\vec{k}) e^{i|\vec{k}|\left(\widehat{k}\cdot\vec{x}+t\right)} \right); \tag{10.2.79}$$

$$\widetilde{\varphi}^+(\vec{k}) = \frac{\widetilde{\varphi}(k_0 = |\vec{k}|, \vec{k})}{2|\vec{k}|} = \overline{\widetilde{\varphi}^-(\vec{k})}. \tag{10.2.80}$$

In quantum field theory, these $\widetilde{\varphi}^\pm$ become *operators*; for a fixed wave vector $\vec{k}$, they are analogous to the raising and lowering operators of the quantum simple harmonic oscillator.

Problem 10.12. Massive Scalar Wave Equation Consider instead the massive scalar wave equation

$$\left(\partial^2 + m^2\right) \varphi = 0. \tag{10.2.81}$$

Show that the plane wave basis still solve this wave equation, but the dispersion relation now reads

$$k^2 = m^2 \qquad \Leftrightarrow \qquad \omega^2 = \vec{k}^2 + m^2. \tag{10.2.82}$$

If m is mass, put back the $\hbar$s and cs in the above dispersion relation. In quantum mechanics, energy associated with angular frequency ω is $E = \hbar\omega$ while momentum associated with the wave vector $\vec{k}$ is $\vec{p} = \hbar\vec{k}$;[105] from these, obtain the famous Einstein formula

$$E^2 = \vec{p}^2 c^2 + (mc^2)^2. \tag{10.2.83}$$

Hint: You may assume $[k_\mu]$ is inverse length, since $[x^\mu]$ is length and $k \cdot x$ being the exponent is dimensionless. □

Problem 10.13. Lorentz Covariance of Higher Rank Wave Equations Consider the (Cartesian coordinate basis) wave equations

$$\partial^2 A_\mu(x) = 0 = \partial^2 H_{\mu\nu}(x). \tag{10.2.84}$$

Explain why, if $A_\mu(x)$ and $H_{\mu\nu}(x)$ are solutions to the wave equation in this $x-$inertial frame, then if we define a different inertial frame

$$x^\alpha = \Lambda^\alpha{}_\beta x'^\beta, \tag{10.2.85}$$

where $\Lambda^\alpha{}_\beta$ is a Lorentz transformation, the transformed fields

$$A'_{\alpha'}(x') \equiv A_\mu(x^\sigma = \Lambda^\sigma{}_\rho x'^\rho)\Lambda^\mu{}_\alpha, \tag{10.2.86}$$

$$\text{and} \qquad H'_{\alpha'\beta'}(x') \equiv H_{\mu\nu}(x^\sigma = \Lambda^\sigma{}_\rho x'^\rho)\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta \tag{10.2.87}$$

must necessarily solve the corresponding wave equation in this $x'-$frame. Be sure to write down the $x'-$frame wave equations. □

[105] Recall: When the left hand side of Schrödinger's equation $i\hbar\partial_t |\Psi\rangle = H|\Psi\rangle$ involves a single frequency, $|\Psi\rangle \propto e^{-i\omega t}$; upon differentiation $\hbar\omega|\Psi\rangle = H|\Psi\rangle$. When spatial translation symmetry is present, $-i\hbar\vec{\nabla}_{\vec{x}}$ is the Hermitian linear momentum operator with eigenfunctions $\{e^{i\vec{k}\cdot\vec{x}}\}$ and eigenvalues $\{\vec{p} = \hbar\vec{k}\}$.

Levi-Civita Pseudo-Tensor Defining $\epsilon_{012\dots D} \equiv 1$, the Levi-Civita symbol in Minkowski is in fact a pseudo-tensor because

$$\Lambda^{\mu_0}{}_{\nu_0} \dots \Lambda^{\mu_D}{}_{\nu_D} \epsilon_{\mu_0\mu_1\dots\mu_D} = \epsilon_{\nu_0\dots\nu_D} \det \Lambda. \tag{10.2.88}$$

(Recall equations (3.2.1) and (3.2.10).) It is invariant if $\det \Lambda = 1$; whereas it is odd under Lorentz transformation if $\det \Lambda = -1$. Moreover, we may raise its indices and discover

$$\epsilon^{\mu_0\dots\mu_D} \equiv \eta^{\mu_0\nu_0} \dots \eta^{\mu_D\nu_D} \epsilon_{\nu_0\nu_1\dots\nu_D} = \epsilon_{\mu_0\dots\mu_D}(-)^D \tag{10.2.89}$$

because $\det \eta^{\mu\nu} = (-)^D$ in our 'mostly minus' convention. In particular,

$$\epsilon^{012\dots D} = (-)^D. \tag{10.2.90}$$

Problem 10.14. Partial Contractions Prove the following counterpart to eq. (10.2.88):

$$\Lambda^{\mu_0}{}_{\nu_0} \dots \Lambda^{\mu_D}{}_{\nu_D} \epsilon^{\nu_0\nu_1\dots\nu_D} = (\det \Lambda) \epsilon^{\mu_0\mu_1\dots\mu_D}. \tag{10.2.91}$$

Next, for $0 < N < D$, prove the following relations:

$$\Lambda^{\mu_{N+1}}{}_{\nu_{N+1}} \dots \Lambda^{\mu_D}{}_{\nu_D} \epsilon_{\alpha_0\alpha_1\dots\alpha_N\mu_{N+1}\dots\mu_D} = (\det \Lambda) \Lambda_{\alpha_0}{}^{\nu_0} \Lambda_{\alpha_1}{}^{\nu_1} \dots \Lambda_{\alpha_N}{}^{\nu_N} \epsilon_{\nu_0\dots\nu_D} \tag{10.2.92}$$

$$\Lambda^{\mu_{N+1}}{}_{\nu_{N+1}} \dots \Lambda^{\mu_D}{}_{\nu_D} \epsilon^{\alpha_0\alpha_1\dots\alpha_N\nu_{N+1}\dots\nu_D} = (\det \Lambda) \Lambda_{\mu_0}{}^{\alpha_0} \dots \Lambda_{\mu_N}{}^{\alpha_N} \epsilon^{\mu_0\mu_1\dots\mu_D}. \tag{10.2.93}$$

Hints: Contract both sides of equations (10.2.88) and (10.2.91) with appropriate Λ's. $\square$

Curvilinear Coordinates We close this section by emphasizing, while Cartesian coordinates are important in elucidating the Poincaré symmetry of flat (Minkowski) spacetime, physics (and geometry) should of course not depend on the coordinates used to compute its observables. To this end, let us re-write some of the objects we have encountered thus far in arbitrary coordinates. If x^α are Cartesian while y^α denote some other (possibly curvilinear) spacetime coordinates, and the transformation $x(y)$ is given, then the line element in eq. (10.1.2) now becomes

$$\mathrm{d}s^2 = \eta_{\alpha\beta} \frac{\partial x^\alpha}{\partial y^\mu} \frac{\partial x^\beta}{\partial y^\nu} \mathrm{d}y^\mu \mathrm{d}y^\nu \equiv g_{\mu'\nu'}(y) \mathrm{d}y^\mu \mathrm{d}y^\nu. \tag{10.2.94}$$

If $g^{\mu'\nu'}$ and $\sqrt{|g(y)|}$ are respectively the inverse metric and the square root of the determinant of $g_{\mu'\nu'}$, the scalar wave operator now reads

$$\Box_y \varphi = \frac{\partial_{y^\mu} \left(\sqrt{|g(y)|} g^{\mu'\nu'} \partial_{y^\nu} \varphi \right)}{\sqrt{|g(y)|}}. \tag{10.2.95}$$

The tangent vector along some trajectory $x^\mu(\lambda') = x^\mu(y(\lambda'))$ reads

$$\frac{\mathrm{d}}{\mathrm{d}\lambda'} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda'} \partial_{x^\mu} = \frac{\mathrm{d}y^\mu}{\mathrm{d}\lambda'} \partial_{y^\mu} \tag{10.2.96}$$

– which does not involve the spacetime metric at all – whereas its covariant acceleration is

$$\frac{D^2}{\mathrm{d}\lambda'^2} = \left(\frac{\mathrm{d}^2 y^\alpha}{\mathrm{d}\lambda'^2} + \frac{\partial y^\alpha}{\partial x^\sigma} \frac{\partial x^\sigma}{\partial y^\mu \partial y^\nu} \frac{\mathrm{d}y^\mu}{\mathrm{d}\lambda'} \frac{\mathrm{d}y^\nu}{\mathrm{d}\lambda'} \right) \frac{\partial}{\partial y^\alpha}. \tag{10.2.97}$$

Problem 10.15. Drive eq. (10.2.97) from the general formula for the covariant acceleration $a^\mu = \ddot{y}^\alpha + \Gamma^\alpha{}_{\mu\nu}\dot{y}^\mu\dot{y}^\nu$, where the overdot represents the derivative with respect to λ'. Show that the coefficients of ∂_{y^α} are Lorentz invariant. $\Box$

Problem 10.16. World Function in Arbitrary Coordinates Starting from eq. (10.1.25), we may deduce the world function in arbitrary coordinates $\{y^\alpha\}$ to be

$$\sigma(x, x') = (\lambda'_2 - \lambda'_1) \int_{\lambda'_1}^{\lambda'_2} L_g \mathrm{d}\lambda', \tag{10.2.98}$$

$$L_g \equiv \frac{1}{2}\eta_{\alpha\beta}\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial x^\beta}{\partial y^\nu}\frac{\mathrm{d}y^\mu}{\mathrm{d}\lambda'}\frac{\mathrm{d}y^\nu}{\mathrm{d}\lambda'}, \tag{10.2.99}$$

where λ' is an affine parameter. By extremizing the world function with the end points x and x' held fixed – i.e., by applying the Euler-Lagrange equations to the geodesic Lagrangian L_g – show that this leads to $D^2/\mathrm{d}\lambda'^2 = 0$, where the covariant acceleration is the one in eq. (10.2.97). $\Box$

10.3 Lorentz Transformations: Construction

10.3.1 Two Dimensions

What is the most general Lorentz transformation? Let us begin in $(1+1)D \equiv 2D$, where this commonly employed lingo '$D+1$' means D space plus one time dimension. In 2D, there are no spatial rotations – only space translations – to speak of, but we shall see that time and space may be mixed. Parametrize the Lorentz transformation matrix in the fundamental relation of equations (10.1.13) and (10.2.14) as

$$\Lambda = \begin{bmatrix} a & p \\ b & q \end{bmatrix}, \tag{10.3.1}$$

where all parameters are real because Λ is real. Inserting this into eq. (10.2.14),

$$\Lambda^{\mathrm{T}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \Lambda = \begin{bmatrix} a^2 - b^2 & ap - bq \\ ap - bq & p^2 - q^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{10.3.2}$$

The diagonal equations $a^2 - b^2 = 1 = q^2 - p^2$ describes two hyperbolas, whose solutions are therefore

$$(a, b) = (\sigma_a \cosh\xi, \sinh\xi), \tag{10.3.3}$$

$$(p, q) = (\sinh\xi', \sigma_q \cosh\xi'), \tag{10.3.4}$$

where $\xi \in \mathbb{R}$ and the $\sigma_{a,q}$ can be either $+1$ or -1. These hyperbola equations also indicate $V^\mu \equiv (a, b)$ is a unit timelike vector, $V^2 \equiv V^\mu V^\nu \eta_{\mu\nu} = +1$; and $W^\mu \equiv (p, q)$ is a unit spacelike vector, $W^2 \equiv W^\mu W^\nu \eta_{\mu\nu} = -1$. We could have inserted a $\sigma_{b,p} = \pm 1$ in front of $\sinh\xi$ and $\sinh\xi'$, but since sinh runs over the reals, any positive or negative signs out front may be accounted for by choosing the right sign for its argument (ξ or ξ'); namely, $\sigma'' \sinh\xi'' = \sinh(\sigma'' \cdot \xi'')$. Whenever $\sigma_a\sigma_q = +1$, the off diagonal term equation after factoring out $-\sigma_q$ now reads

$$\sinh(\xi - \xi') = 0, \tag{10.3.5}$$

implying $\xi' = \xi$. The $\sigma_a \sigma_q = 1$ condition means $(\sigma_a, \sigma_q) = (+1, +1)$ or $(\sigma_a, \sigma_q) = (-1, -1)$. The former yields

$$\Lambda = \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix}; \tag{10.3.6}$$

whereas the latter, upon replacing $\xi \to -\xi$ (recall ξ is arbitrary), leads to

$$\Lambda = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix} = \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{10.3.7}$$

Whenever $\sigma_a \sigma_q = -1$, the off diagonal equation reduces to

$$\sinh(\xi + \xi') = 0, \tag{10.3.8}$$

implying $\xi' = -\xi$. The $\sigma_a \sigma_q = -1$ condition means $(\sigma_a, \sigma_q) = (+1, -1)$ or $(\sigma_a, \sigma_q) = (-1, +1)$. The former yields, upon replacing $\xi \to -\xi$,

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix}; \tag{10.3.9}$$

whereas the latter leads to

$$\Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix}. \tag{10.3.10}$$

Problem 10.17. Starting from equations (10.3.3) and (10.3.4), verify equations (10.3.5) through (10.3.10). Finally, explain why the most general 2×2 matrix obeying equations (10.1.13) and (10.2.14) – i.e., the most general 2D Lorentz transformation – is therefore

$$\Lambda = \begin{bmatrix} \sigma_T & 0 \\ 0 & \sigma_P \end{bmatrix} \cdot \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix}, \tag{10.3.11}$$

where σ_T and σ_P can be independently either $+1$ or -1. □

When $\sigma_T = 1 = \sigma_P$ in eq. (10.3.11) we obtain the continuous *Lorentz boost*

$$\Lambda^\mu{}_\nu(\xi) = \begin{bmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{bmatrix}, \tag{10.3.12}$$

where the real ξ is known as *rapidity*. When $\xi = 0$ in eq. (10.3.11) and $\sigma_T = -1 = -\sigma_P$, we obtain the discrete time-reversal transformation

$$\Lambda = \begin{bmatrix} -1 & 0 \\ 0 & +1 \end{bmatrix}, \tag{10.3.13}$$

since $\Lambda^\mu{}_\nu(x^0, x^1)^\nu = (-x^0, x^1)$. And when $\xi = 0$ in eq. (10.3.11) and $\sigma_T = 1 = -\sigma_P$, we obtain the space-reversal (aka parity) one

$$\Lambda = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{10.3.14}$$

since $\Lambda^\mu{}_\nu(x^0, x^1)^\nu = (x^0, -x^1)$. Upon recalling that the product of Lorentz transformations yield another Lorentz transformation, we gather:

The most general 2D Lorentz transformation consists of (products of) time reversals, parity flips, and continuous Lorentz boosts parameterized by a real ξ.

Lorentz boost The Lorentz boost in eq. (10.3.12) is what mixes space and time – neither time reversal, parity flips, nor spatial rotations (in higher dimensions) do so – and this is what revolutionized humanity's understanding of spacetime at the beginning of the 1900's. Inspired by the fact that the speed of light is the same in all inertial frames, Einstein discovered *Special Relativity*, that the space and time coordinates of one frame have to become intertwined when being transformed to those in another frame. We will turn this around later when discussing Maxwell's equations: the constancy of the speed of light in all inertial frames is in fact a consequence of the Lorentz covariance of the former.

Problem 10.18. Rapidities add Using the Lorentz boost in eq. (10.3.12), show that

$$\Lambda^{\mu}{}_{\sigma}(\xi)\Lambda^{\sigma}{}_{\nu}(\xi') = \Lambda^{\mu}{}_{\nu}(\xi + \xi'). \tag{10.3.15}$$

That is, two consecutive Lorentz boosts with rapidities ξ and ξ' is equivalent to a single with rapidity $\xi + \xi'$. It should also be recognized, equations (10.3.12) and (10.3.15) are analogous to equations (5.4.6) and (5.4.11), with rapidity ξ being the analog to the rotation angle θ. □

Problem 10.19. Exponential form Show that the Lorentz boost in eq. (10.3.12) can be put in the exponential form

$$\Lambda^{\mu}{}_{\nu}(\xi) = \exp\left(\xi \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right). \tag{10.3.16}$$

This is the parallel to eq. (5.4.40).

Hints: You may decompose the matrix occurring in the exponent in terms of the Pauli matrices of (3.2.18), followed by employing the result in eq. (3.2.26). □

Problem 10.20. Lorentz Boosts, Dilatations and Null Coordinates in 2D If $x^\mu = (x^0, x^1)$ are Cartesian coordinates in 2D Minkowski, i.e., with $\mathrm{d}s^2 = (\mathrm{d}x^0)^2 - (\mathrm{d}x^1)^2$, and if we define the (null) coordinates

$$x^{\pm} \equiv x^0 \pm x^1; \tag{10.3.17}$$

then show that under Lorentz boosts in eq. (10.3.12)

$$x^\alpha \to \Lambda^\alpha{}_\beta x^\beta, \tag{10.3.18}$$

the new (null) coordinates transform as

$$x^{\pm} \to e^{\pm\xi} x^{\pm}. \tag{10.3.19}$$

Also show that the Minkowski metric in these coordinates reads

$$\mathrm{d}s^2 = \mathrm{d}x^+ \mathrm{d}x^-, \tag{10.3.20}$$

so that the Lorentz transformation in eq. (10.3.19) leaves it invariant. Further demonstrate that the wave operator in these coordinates read

$$\Box\psi \equiv |g|^{-1/2}\partial_\mu\left(|g|^{1/2} g^{\mu\nu}\partial_\nu\psi\right) = 4\partial_+\partial_-\psi. \tag{10.3.21}$$

Can you see why these coordinates are dubbed 'null'? Hint: What happens when $\mathrm{d}s^2 = 0$? □

To understand the meaning of the rapidity ξ, let us consider applying it to an arbitrary 2D vector U^μ.

$$U' \equiv \Lambda \cdot U = \left[\begin{array}{c} U^0 \cosh\xi + U^1 \sinh\xi \\ U^1 \cosh\xi + U^0 \sinh\xi \end{array} \right]. \tag{10.3.22}$$

Lorentz Boost: Timelike Case For a timelike vector U^μ, obeying $U^2 \equiv \eta_{\mu\nu} U^\mu U^\nu > 0$, we will now show that it is *not* possible to boost to a frame where its time component is zero; but, for the same reasons, it is *always* possible to Lorentz boost to its 'rest frame', namely $U'^\mu = \Lambda^\mu{}_\nu U^\nu = (U'^0, \vec{0})$.

To obtain $U'^0 = 0$ in the 'boosted' frame, eq. (10.3.22) says

$$U^0 \cosh\xi + U^1 \sinh\xi = 0. \tag{10.3.23}$$

Since $\cosh\xi \geq 1$, this in turn is equivalent to

$$\tanh\xi = -U^0/U^1. \tag{10.3.24}$$

However, timelike $U^2 > 0 \Rightarrow (U^0)^2 > (U^1)^2 \Rightarrow |U^0/U^1| > 1$. Thus it is not possible to find a finite ξ such that $U'^0 = 0$, since tanh within $[-1, +1]$ while $-U^0/U^1$ is either strictly less than -1 or greater than $+1$. On the other hand, if we wish to set $U'^1 = 0$ in the 'boosted' frame,

$$U^1 \cosh\xi + U^0 \sinh\xi = 0; \tag{10.3.25}$$

which in turn is equivalent to

$$\tanh\xi = -U^1/U^0. \tag{10.3.26}$$

Thus, we may always solve uniquely for the ξ that would set the spatial component to zero, since $|U^1/U^0|$ lies within the open interval $(-1, +1)$ by the timelike assumption.

Next, recall that tangent vectors may be interpreted as the derivative of the spacetime coordinates with respect to some parameter λ, namely $U^\mu \equiv \mathrm{d}x^\mu/\mathrm{d}\lambda$. Therefore

$$-\tanh\xi = \frac{U^1}{U^0} = \frac{\mathrm{d}x^1}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}x^0} = \frac{\mathrm{d}x^1}{\mathrm{d}x^0} \equiv v < 1 \tag{10.3.27}$$

is the velocity associated with U^μ in the frame $\{x^\mu\}$. Starting from $\tanh\xi = -v$, some algebra would then hand us

$$\cosh(\xi) = \gamma \equiv \frac{1}{\sqrt{1-v^2}}, \tag{10.3.28}$$

$$\sinh(\xi) = -\gamma \cdot v = -\frac{v}{\sqrt{1-v^2}}, \tag{10.3.29}$$

$$\Lambda^\mu{}_\nu(v) = \left[\begin{array}{cc} \gamma & -\gamma\cdot v \\ -\gamma\cdot v & \gamma \end{array} \right]. \tag{10.3.30}$$

(A short calculation would verify $\Lambda(v)\cdot\Lambda(-v) = \mathbb{I}$.) This in turn yields

$$U'^\alpha = \left(\mathrm{sgn}(U^0)\sqrt{\eta_{\mu\nu}U^\mu U^\nu}, 0\right)^{\mathrm{T}}; \tag{10.3.31}$$

leading us to interpret the $\Lambda^\mu{}_\nu$ we have found in eq. (10.3.30) as the boost that bring observers to the frame where the flow associated with U^μ is 'at rest'. If $U^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$, where τ is proper time, then $\eta_{\mu\nu}U^\mu U^\nu = 1$. Additionally, we shall witness below how the factor $\gamma = 1/(1-v^2)^{1/2}$ yields dramatic differences from ordinary Newtonian kinematics when the speeds involved approach unity because it can grow without bound, $\gamma(v \to 1) \to \infty$.

Problem 10.21. 2D Timelike Vectors Since U^0 can never be zero, argue that an arbitrary timlike vector can, for an appropriate $\xi_0 \in \mathbb{R}$, be expressed as

$$U^\mu = \sqrt{U^\alpha U_\alpha}\,(\sigma_t \cosh \xi_0, \sinh \xi_0)\,, \tag{10.3.32}$$

where $\sigma_t = \pm 1$ may be viewed as the direction of the timelike 'flow' – either towards the future (+) or the past (−). Under an arbitrary Lorentz boost $\Lambda(\xi)$, verify

$$\Lambda(\xi)\cdot U = \sigma_t \sqrt{U^\alpha U_\alpha}\,(\cosh(\xi + \sigma_t \xi_0), \sinh(\xi + \sigma_t \xi_0))\,. \tag{10.3.33}$$

Notice the sign of the zeroth component never changes under a continuous Lorentz boost – it is a Lorentz invariant – a fact that could already be glimpsed from eq. (10.3.31), since it must be possible to boost into the rest frame the same timelike vector but starting from different inertial frames. On the other hand, the sign of the space component is not a Lorentz invariant. □

Proper Time The Lorentz-invariant infinitesimal interval $\mathrm{d}s = |\dot{Z}^2|^{1/2}\mathrm{d}\lambda$ for timelike $\dot{Z}^\mu$ may now be evaluated in the 'rest frame', where $\dot{Z}'^\mu = (\dot{Z}'^0, 0) = \Lambda^\mu{}_\nu \dot{Z}^\nu$:

$$\mathrm{d}s = \left|(\mathrm{d}Z'^0/\mathrm{d}\lambda)^2\right|^{1/2} \mathrm{d}\lambda = \mathrm{d}Z'^0, \tag{10.3.34}$$

as long as $Z'^0(\lambda)$ a monotonically increasing function of λ. If we define *proper time* to be the time measured by the clock at rest with the trajectory, $\mathrm{d}Z'^i/\mathrm{d}\lambda = 0$, we see that $\mathrm{d}s$ is in fact its infinitesimal version.

For the timelike vector field $U^\mu = \mathrm{d}Z^\mu/\mathrm{d}\lambda$, its associated Lorentz-invariant infinitesimal proper time is given by $\mathrm{d}\tau = |U^2|^{1/2}\mathrm{d}\lambda = |\dot{Z}^2|^{1/2}\mathrm{d}\lambda$.

Lorentz Boost: Spacelike case For a spacelike vector U^μ, obeying $U^2 \equiv \eta_{\mu\nu}U^\mu U^\nu < 0$, we will now show, it is always possible to find a Lorentz boost so that $U'^\mu = \Lambda^\mu{}_\nu U^\nu = (0, U'^1)$ has no time components – hence the term 'spacelike'. This can correspond, for instance, to the vector joining two spatial locations within a macroscopic body at a given time.

Suppose U were spacelike in 2D, $U^2 < 0 \Rightarrow (U^0)^2 < (U^1)^2 \Rightarrow |U^1/U^0| = |\mathrm{d}x^1/\mathrm{d}x^0| \equiv |v| > 1$. Then, recalling eq. (10.3.22), it is not possible to find a finite ξ such that $U'^1 = 0$, because that would amount to solving $\tanh\xi = -U^1/U^0$, but tanh lies between −1 and +1 whereas $-U^1/U^0 = -v$ is either less than −1 or greater than +1. On the other hand, it is certainly possible to have $U'^0 = 0$. Simply do $\tanh\xi = -U^0/U^1 = -1/v$. Similar algebra to the timelike case then hands us

$$\cosh\xi = \left(1 - v^{-2}\right)^{-1/2} = \frac{|v|}{\sqrt{v^2-1}}, \tag{10.3.35}$$

$$\sinh\xi = -(1/v)\left(1 - v^{-2}\right)^{-1/2} = -\frac{\mathrm{sgn}(v)}{\sqrt{v^2-1}}, \tag{10.3.36}$$

$$U'^{\alpha} = \left(0, \mathrm{sgn}(v)\sqrt{-\eta_{\mu\nu}U^{\mu}U^{\nu}}\right)^{\mathrm{T}}, \qquad v \equiv \frac{U^1}{U^0} > 1. \tag{10.3.37}$$

We may interpret U'^{μ} and U^{μ} as infinitesimal vectors joining the same pair of spacetime points but in their respective frames. Specifically, U'^{μ} are the components in the frame where the pair lies on the same constant-time surface ($U'^0 = 0$); while U^{μ} are the components in a boosted frame.

Problem 10.22. 2D Spacelike Vectors Since U^1 can never be zero, argue that an arbitrary spacelike vector can, for an appropriate $\xi_0 \in \mathbb{R}$, be expressed as

$$U^{\mu} = \sqrt{-U^{\alpha}U_{\alpha}}\left(\sinh\xi_0, \sigma_s\cosh\xi_0\right)^{\mathrm{T}}, \tag{10.3.38}$$

where $\sigma_s = \pm 1$ may be viewed as the direction of the spacelike 'flow' – either towards the positive (+) or the negative (−) 1−direction. Under an arbitrary Lorentz boost $\Lambda(\xi)$, verify

$$\Lambda(\xi)\cdot U = \sigma_s\sqrt{-U^{\alpha}U_{\alpha}}\left(\sinh(\xi+\sigma_t\xi_0), \cosh(\xi+\sigma_t\xi_0)\right)^{\mathrm{T}}. \tag{10.3.39}$$

Notice the sign of the space component never changes under a continuous Lorentz boost – it is a Lorentz invariant – a fact that could already be glimpsed from eq. (10.3.31), since it must be possible to boost into the constant-time frame the same spacelike vector but starting from different inertial frames. On the other hand, the sign of the time component is not a Lorentz invariant. □

Proper Length We may now evaluate the Lorentz-invariant infinitesimal interval $\mathrm{d}s = |\dot{Z}^2|^{1/2}\mathrm{d}\lambda$ in the frame where $\dot{Z}'^{\mu}$ is purely spatial. As long as $Z'^{i}(\lambda)$ is a monotonically increasing function of λ,

$$\mathrm{d}s = \left|-(\mathrm{d}Z'^1/\mathrm{d}\lambda)^2\right|^{1/2}\mathrm{d}\lambda = \mathrm{d}Z'^1. \tag{10.3.40}$$

Defining *proper length* to be the distance swept out by the trajectory in the frame where it is purely spatial, $\mathrm{d}Z'^0/\mathrm{d}\lambda = 0$, we may summarize:

> For the spacelike vector field $U^{\mu} = \mathrm{d}Z^{\mu}/\mathrm{d}\lambda$, its associated Lorentz-invariant infinitesimal proper length is given by $\mathrm{d}\ell = |U^2|^{1/2}\mathrm{d}\lambda = |\dot{Z}^2|^{1/2}\mathrm{d}\lambda$.

Lorentz Boost: Null (aka lightlike) case Finally, let U^{μ} be null, obeying $U^2 = \eta_{\mu\nu}U^{\mu}U^{\nu} = 0$. In 2D, that means $(U^0)^2 = (U^1)^2$, which in turn implies

$$U^{\mu} = \omega(1, \pm 1)^{\mathrm{T}} \tag{10.3.41}$$

for some real number ω. Upon a Lorentz boost, eq. (10.3.22) tells us it becomes re-scaled by $e^{\pm\xi}$:

$$U' \equiv \Lambda \cdot U = \omega\begin{bmatrix}\cosh\xi \pm \sinh\xi \\ \sinh\xi \pm \cosh\xi\end{bmatrix} = \omega\begin{bmatrix}1 \\ \pm 1\end{bmatrix}e^{\pm\xi}. \tag{10.3.42}$$

(This result is related to Problem (10.20).) Neither the sign of the time nor space components can ever be flipped by a continuous Lorentz boost.

As we shall see below, if $U^\mu = \omega(1, \pm 1)^{\mathrm{T}}$ and $U'^\mu = \omega'(1, \pm 1)^{\mathrm{T}}$ describe the $d-$momentum of a photon in, respectively, the $\{x^\mu\}$ and $\{x'^\mu\}$ inertial frames, so that $|\omega|$ and $|\omega'|$ are the corresponding frequencies the ratio

$$\frac{U'^0}{U^0} = \frac{\omega'}{\omega} = \exp(\pm\xi) \tag{10.3.43}$$

describes photon's red- or blue-shift in the $x'-$frame relative to the $x-$frame. Notice it is not possible to set either the time or space component to zero, unless $\xi \to \pm\infty$.

Chronology-in-time From equations (10.3.33) and (10.3.43), we may collect the following important fact:

> The sign of the time components (i.e., positive or negative) of timelike and null vectors is a Lorentz invariant in that it cannot be altered by a continuous Lorentz boost.

On the other hand, eq. (10.3.39) informs us:

> The sign of the time components (i.e., positive or negative) of spacelike vectors is *not* a Lorentz invariant in that it *can* be altered by a continuous Lorentz boost.

We shall argue below that these features regarding timelike and null versus spacelike vectors persist in higher dimensions.

Coordinate Axes Overlaid Let us take a moment to place the $x'^\mu = (t', x')$ frame Cartesian axes on top of the $x^\mu = (t, x)$ ones. Firstly, note that the $x'-$axis and the lines parallel to it are simply the constant $t'-$lines:

$$0 = \mathrm{d}t' = \mathrm{d}t \cosh\xi + \mathrm{d}x \sinh\xi \tag{10.3.44}$$

$$\frac{\mathrm{d}t}{\mathrm{d}x} = -\tanh\xi. \tag{10.3.45}$$

Whereas, the the $t'-$axis and the lines parallel to it are simply the constant $x'-$lines:

$$0 = \mathrm{d}x' = \mathrm{d}t \sinh\xi + \mathrm{d}x \cosh\xi \tag{10.3.46}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\tanh\xi. \tag{10.3.47}$$

Hence, the angle that the x'-axis makes with the $x-$axis is the same as the angle the t'-axis makes with the $t-$axis. These orthogonal $(t', x')-$axes therefore do not 'look' orthogonal when overlaid on their (t, x) counterparts; appearances are no longer a good source of intuition here. Furthermore, as ξ grows larger these $(t', x')-$axes tend to $\pm\pi/4$ lines, where they become null. The rather un-intuitive character of these results is due to the geometry being that of *flat spacetime*, not flat space.

Constant Velocity Motion Below, we shall argue that massive bodies must move at subluminal speeds $|v| < 1$ in an inertial frame. As such, let us consider the situation where there is a family of observers moving at a constant velocity v relative to another family of observers – where each family of observers are at rest with respect to their respective inertial frames – the former parametrized with (t, x) and the latter (t', x'). If the $(t', x')-$frame is moving at velocity v with respect to the $(t, x)-$frame, from eq. (10.3.30), we gather

$$\left[\begin{array}{c} t' \\ x' \end{array}\right] = \Lambda(v) \cdot (t, x)^{\mathrm{T}} = \gamma(v) \left[\begin{array}{c} t - vx \\ x - vt \end{array}\right] \tag{10.3.48}$$

and

$$\left[\begin{array}{c} t \\ x \end{array}\right] = \Lambda(-v) \cdot (t', x')^{\mathrm{T}} = \gamma(v) \left[\begin{array}{c} t' + vx' \\ x' + vt' \end{array}\right], \tag{10.3.49}$$

where $\gamma(v) = (1 - v^2)^{-1/2} = \gamma(-v)$. Any two points in spacetime $(t_{1,2}, x_{1,2})$ corresponding to $(t'_{1,2}, x'_{1,2})$ have displacement vectors from $1 \to 2$ given by

$$\Delta'^{\mu} \equiv (t'_2 - t'_1, x'_2 - x'_1)^{\mu} = \gamma(v) \left[\begin{array}{c} t_2 - t_1 - v(x_2 - x_1) \\ x_2 - x_1 - v(t_2 - t_1) \end{array}\right] \tag{10.3.50}$$

and

$$\Delta^{\mu} \equiv (t_2 - t_1, x_2 - x_1)^{\mu} = \gamma(v) \left[\begin{array}{c} t'_2 - t'_1 + v(x'_2 - x'_1) \\ x'_2 - x'_1 + v(t'_2 - t'_1) \end{array}\right]. \tag{10.3.51}$$

Length Contraction (Draw a figure.) Suppose we place a 'ruler' of proper length L at rest in the $(t, x)-$frame, so that it may be parametrized by the worldsheet

$$z^{\mu} = (t \in \mathbb{R}, a \leq x \leq L + a)^{\mathrm{T}}, \tag{10.3.52}$$

for some real a. The same worldsheet in the $(t', x')-$frame is

$$z'^{\mu} = (t', x')^{\mathrm{T}} = \gamma(v) \left[\begin{array}{c} t - vx \\ x - vt \end{array}\right]. \tag{10.3.53}$$

The displacement vector from the left end of the ruler at $(t_1, x_1 = a)$ to some other (arbitrary) point on it at $(t_2, a < x_2 \leq L + a)$ but in the $(t', x')-$frame is

$$\Delta'^{\mu} = (t'_2 - t'_1, x'_2 - x'_1)^{\mathrm{T}} \tag{10.3.54}$$

$$= \gamma(v) \left[\begin{array}{c} t_2 - t_1 - v(x_2 - a) \\ x_2 - a - v(t_2 - t_1) \end{array}\right]. \tag{10.3.55}$$

The spatial extent of the ruler in the $(t', x')-$frame at a constant time is gotten by setting $\Delta'^0 = 0$; implying $t_2 - t_1 = v(x_2 - a)$ and

$$\Delta'^{\mu} = \gamma(v)(0, (1 - v^2)(x_2 - a))^{\mathrm{T}} \tag{10.3.56}$$

$$= \sqrt{1 - v^2} \left[\begin{array}{c} 0 \\ x_2 - a \end{array}\right]. \tag{10.3.57}$$

Since $x_2 - a$ runs from 0 to L over the spatial trajectory of the rule at fixed time, we see that the ruler in the $(t', x')-$frame has length

$$L' = \sqrt{1 - v^2} L < L. \tag{10.3.58}$$

This is known as *length contraction.*

Example: Moving Pole-in-Barn A classical 'paradox' involving length contraction goes as follows. A barn and a pole each has proper length L. The pole is moving at constant positive velocity $0 < v < 1$ relative to the barn. Would the pole fit inside the barn? From the perspective of the barn, the pole would be length contracted – the faster it is moving, the more easily it ought to fit. On the other hand, from the pole's perspective, the barn would be length contracted and therefore it would not fit. Whether or not the pole fits in the barn, however, is a question of *simultaneity*: do the ends of the pole reside between the two ends of the barn *at the same time*? And since, for a pair of spacelike separated events such as the ends of the pole/barn, their chronology is not Lorentz invariant – different inertial frames would view one occurring before the other, or vice versa – there really is no paradox. (Draw a figure.)

In the rest ($x'-$)frame of the pole, it can be parametrized as a *worldsheet*

$$Z_p'^\mu = \left(t' \in \mathbb{R}, 0 \leq z' \leq L\right)^{\mathrm{T}}. \tag{10.3.59}$$

For a fixed time t', its spatial trajectory runs from the $z' = 0$ end of the pole to the $z' = L$ end. Notice the tangent vectors

$$\frac{\partial Z_p'^\mu}{\partial t'} = (1, 0) \qquad \text{and} \qquad \frac{\partial Z_p'^\mu}{\partial z'} = (0, 1) \tag{10.3.60}$$

are respectively timelike and spacelike. In the rest ($x-$)frame of the barn, the pole worldsheet is instead

$$Z_p^\mu = \gamma(v) \left[\begin{array}{c} t' + vz' \\ z' + vt' \end{array} \right]. \tag{10.3.61}$$

The worldsheet of the barn in this $x-$frame is

$$Z_b^\mu = \left(t \in \mathbb{R}, z_0 \leq z \leq z_0 + L\right)^{\mathrm{T}} \tag{10.3.62}$$

for some constant z_0; while in the rest ($x'-$)frame of the pole,

$$Z_b'^\mu = \gamma(v) \left[\begin{array}{c} t - vz \\ z - vt \end{array} \right]. \tag{10.3.63}$$

For a fixed time $t = \gamma(t' + vz')$, the spatial trajectory of the right-moving pole in the rest frame of the barn is

$$Z_p^1 = \gamma z' + v(t - \gamma v z') \tag{10.3.64}$$

$$= \sqrt{1 - v^2} z' + vt \in [vt, \sqrt{1 - v^2} L + vt]. \tag{10.3.65}$$

When the $z' = L$ (right) end of the pole just touches the $z = z_0 + L$ (right) end of the barn at $t = (L/v)(1 - \sqrt{1 - v^2}) + z_0/v$, the $z' = 0$ (left) end of the pole lies to the right of the $z = z_0$ (left) end of the barn –

$$Z_p^1(z' = 0) = L(1 - \sqrt{1 - v^2}) + z_0 > z_0 \tag{10.3.66}$$

– namely, the pole fits inside the barn. On the other hand, in the rest ($x'-$)frame of the pole, and for a fixed time $t' = \gamma(t - vz)$, the spatial trajectory of the left-moving barn is

$$Z_b'^1 = \sqrt{1-v^2}z - vt' \in [\sqrt{1-v^2}z_0 - vt', \sqrt{1-v^2}(z_0 + L) - vt']. \tag{10.3.67}$$

When the $z = z_0$ (left) end of the barn just touches the $z' = 0$ (left) end of the pole at $t' = (z_0/v)\sqrt{1-v^2}$, the $z = z_0 + L$ (right) end of the barn lies to the left of the $z' = L$ (right) end of the pole –

$$Z_b'^1(z = L) = L\sqrt{1-v^2} < L \tag{10.3.68}$$

– namely, the pole does not fit inside the barn.

Time Dilation Consider a pair of twins born at each and every x at time $t = 0$ in an inertial frame (t, x). For $t > 0$, the first twin stays at rest at spatial location x, sweeping out a wordline $(t \geq 0, x)$. The second twin goes on a subluminal constant velocity motion, sweeping out a worldline $(t, x + vt)$. In the frame (t', x') at rest with the second twin, we have

$$(t', x') = \gamma(v)(t - vx, x - vt); \tag{10.3.69}$$

whereas in the frame (t, x) at rest with the first twin, we have

$$(t, x) = \gamma(v)(t' + vx', x' + vt'). \tag{10.3.70}$$

For the first twin the displacement vector joining his worldline from t_1 to t_2 is given by $\Delta^\mu = (t_2 - t_1, 0)$ in the rest frame (t, x). On the other hand, the same displacement according to the $(t', x')-$frame would demand $\Delta^1 = 0 \Rightarrow x'_2 - x'_1 = -v(t'_2 - t'_1)$ and therefore

$$\Delta^\mu = (t_2 - t_1, 0)^{\mathrm{T}} = \sqrt{1-v^2}\begin{bmatrix} t'_2 - t'_1 \\ 0 \end{bmatrix}. \tag{10.3.71}$$

Since the first twin is at rest in the $(t, x)-$frame, $t_2 - t_1$ is his proper time elapsed, and

$$t_2 - t_1 = \sqrt{1-v^2}(t'_2 - t'_1) < t'_2 - t'_1. \tag{10.3.72}$$

For the second twin the displacement vector joining his worldline from t'_1 to t'_2 is given by $\Delta'^\mu = (t'_2 - t'_1, 0)$ in the rest frame (t', x'). On the other hand, the same displacement according to the $(t, x)-$frame would demand $\Delta'^1 = 0 \Rightarrow x_2 - x_1 = v(t_2 - t_1)$ and therefore

$$\Delta'^\mu = (t'_2 - t'_1, 0)^{\mathrm{T}} = \sqrt{1-v^2}\begin{bmatrix} t_2 - t_1 \\ 0 \end{bmatrix}. \tag{10.3.73}$$

Since the second twin is at rest in the $(t', x')-$frame, $t'_2 - t'_1$ is his proper time elapsed, and

$$t'_2 - t'_1 = \sqrt{1-v^2}(t_2 - t_1) < t_2 - t_1. \tag{10.3.74}$$

These results are known as relativistic *time dilation* – the clocks attached to bodies moving at a constant velocity with respect to a given observer appear to run slower than her own proper time. Note that equations (10.3.72) and (10.3.74) do not contradict each other; because they are not referring to the same physical situation. In the former, the first twin at rest in the $(t, x)-$frame is watching a family of twins passing by at a constant velocity v, and comparing her clock with theirs. In the latter, second twin at rest in the $(t', x')-$frame is watching a different family of twins passing by at a constant velocity $-v$, and comparing her clock with theirs. Below, we shall ask a different question: for a given pair of twins born at the same spacetime location (t_1, x), with one twin staying at rest and the second going on a trip before returning to meet the first twin at some future spacetime point (t_2, x), which twin has aged more?

10.3.2 Higher Dimensions

Moving on to higher dimensions, we first return to the 00 component of (10.1.13):

$$(\Lambda^0{}_0)^2 - \delta_{ij}\Lambda^i{}_0\Lambda^j{}_0 = 1, \tag{10.3.75}$$

which tells us either $\Lambda^0{}_0 \geq +1$ or $\Lambda^0{}_0 \leq -1$ because

$$(\Lambda^0{}_0)^2 = 1 + \sum_{i=1}^{D}(\Lambda^i{}_0)^2 \geq 1. \tag{10.3.76}$$

By viewing $\Lambda^i{}_0 = v \cdot \widehat{v}^i$ as a spatial vector, where $\widehat{v}$ obeys $\widehat{v}^2 = 1$, this is an equation describing the hyperbola involving its magnitude and $\Lambda^0{}_0$. The solution is $\Lambda^0{}_0 = \sigma_0 \cosh\xi$ and $\Lambda^i{}_0 = \sinh\xi\widehat{v}^i$ for σ_0 either $= 1$ or -1; but this can always be implemented by writing $\Lambda = \text{diag}[\sigma_0, \mathbb{I}] \cdot \Lambda'$, where $\Lambda'^0_{\ 0} = \cosh\xi$ and $\Lambda'^i_{\ 0} = \sinh\xi\widehat{v}^i$. (Be sure to recognize $\text{diag}[\sigma_0, \mathbb{I}]$ to be a Lorentz transformation too.) Hence, we choose to express

$$\Lambda'^0_{\ 0} = \cosh\xi \tag{10.3.77}$$
$$\Lambda'^i_{\ 0} = \sinh\xi\widehat{v}^i. \tag{10.3.78}$$

The $0i$ components involving Λ' read

$$\Lambda'^0_{\ 0}\Lambda'^0_{\ i} = \delta_{mn}\Lambda'^m_{\ 0}\Lambda'^n_{\ i}, \tag{10.3.79}$$
$$\Lambda'^0_{\ i} = \tanh\xi\widehat{v}^l\Lambda'^l_{\ i}. \tag{10.3.80}$$

Finally, the ij components are

$$\Lambda'^0_{\ i}\Lambda'^0_{\ j} - \delta_{mn}\Lambda'^m_{\ i}\Lambda'^n_{\ j} = -\delta_{ij}, \tag{10.3.81}$$
$$\left(\delta^{mn} - \tanh^2\xi\widehat{v}^m\widehat{v}^n\right)\Lambda'^m_{\ i}\Lambda'^n_{\ j} = \delta_{ij}. \tag{10.3.82}$$

If we choose a rotation matrix $\widehat{R}$ such that $\widehat{v}^i = \widehat{R}^{ij}\delta^j_1$, and recall $1 - \tanh^2\xi = \text{sech}^2\xi$, then

$$\left(\delta^m_1\delta^n_1\text{sech}^2\xi + \delta^{mn} - \delta^m_1\delta^n_1\right)(\widehat{R}^{pm}\Lambda'^p_{\ i})(\widehat{R}^{qn}\Lambda'^q_{\ j}) = \delta_{ij}. \tag{10.3.83}$$

Problem 10.23. Explain why eq. (10.3.83) may be phrased as the matrix equation

$$\widehat{O}'^{\text{T}}\widehat{O}' = \mathbb{I}, \tag{10.3.84}$$

where

$$\widehat{O}' = \begin{bmatrix} \text{sech}\,\xi & \vec{0}^{\text{T}} \\ \vec{0} & \mathbb{I}_{(D-1)\times(D-1)} \end{bmatrix} \cdot \widehat{R}^{\text{T}} \cdot \widehat{\Lambda}_{D\times D}, \tag{10.3.85}$$

where $\widehat{\Lambda}_{D\times D}$ is the $D \times D$ sub-block of $\Lambda^\mu{}_\nu$, containing its space-space components. $\square$

From equations (10.3.84) and (10.3.85), the following object must therefore be an orthogonal matrix $\widehat{O}'$:

$$\left(\delta_1^a \delta_1^l \operatorname{sech} \xi + \delta^{al} - \delta_1^a \delta_1^l\right) \widehat{R}^{pl} \Lambda^p{}_b \equiv \widehat{O}'^a{}_b. \tag{10.3.86}$$

Inverting this relation or the matrix version in eq. (10.3.85) yields

$$\widehat{\Lambda} = \widehat{R} \cdot \operatorname{diag}(\cosh \xi, \mathbb{I}_{(D-1)\times(D-1)}) \cdot \widehat{R}^{\mathrm{T}} \widehat{R} \cdot \widehat{O}', \tag{10.3.87}$$

$$\Lambda^a{}_b = \widehat{R}^{am} \left(\delta_1^m \delta_1^l \cosh \xi + \delta^{ml} - \delta_1^m \delta_1^l\right) \widehat{R}^{il} \widehat{R}^{ik} \widehat{O}'^k_b \tag{10.3.88}$$

$$= \left(\widehat{v}^a \widehat{v}^l \cosh \xi + \delta^{al} - \widehat{v}^a \widehat{v}^l\right) \widehat{O}^l{}_b, \tag{10.3.89}$$

where $\widehat{O}$ is an arbitrary orthogonal matrix since $\widehat{O}'$ was one. This in turn implies

$$\Lambda^0{}_i = \sinh \xi \widehat{v}^l \widehat{O}^l{}_i. \tag{10.3.90}$$

At this point, we may gather

$$\Lambda^\mu{}_\nu = \begin{bmatrix} \sigma_0 & 0 \\ 0 & \mathbb{I}_{D\times D} \end{bmatrix} \cdot \Lambda_B \cdot \Lambda_O; \tag{10.3.91}$$

where the pure boost is

$$\Lambda_B = \begin{bmatrix} \cosh \xi & \sinh \xi \cdot \widehat{v}^{\mathrm{T}} \\ \sinh \xi \cdot \widehat{v} & \cosh \xi \widehat{v}\widehat{v}^{\mathrm{T}} + \mathbb{I}_{D\times D} - \widehat{v}\widehat{v}^{\mathrm{T}} \end{bmatrix} \tag{10.3.92}$$

and the spatial rotations and/or parity flips is

$$\Lambda_O = \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{O} \end{bmatrix}, \qquad \widehat{O}^{\mathrm{T}} \widehat{O} = \mathbb{I}. \tag{10.3.93}$$

We highlight that the $\mathbb{I} - \widehat{v}\widehat{v}^{\mathrm{T}}$ in eq. (10.3.92) is a projector, in that when applied to an arbitrary spatial vector $\vec{a}$, it returns a vector perpendicular to $\widehat{v}$. Specifically, because $(\mathbb{I} - \widehat{v}\widehat{v}^{\mathrm{T}})\widehat{v} = \vec{0}$, we have the orthogonality condition

$$\widehat{v}^{\mathrm{T}} \left(\mathbb{I} - \widehat{v}\widehat{v}^{\mathrm{T}}\right) \widehat{a} \equiv \widehat{v} \cdot \vec{a}_\perp = 0. \tag{10.3.94}$$

This teaches us that a given pure boost parametrized by $\widehat{v}$ really mingles the time component and the spatial ones parallel to $\widehat{v}$; the spatial components perpendicular to $\widehat{v}$ remain un-touched.

Next, taking the determinant of both sides of (10.2.14), and recalling $\det A^{\mathrm{T}} = \det A$ and $\det(A \cdot B) = \det A \det B$ for matrices A and B, one gathers the determinant of the Lorentz transformation matrix Λ is either $+1$ or -1 because

$$(\det \Lambda)^2 = 1. \tag{10.3.95}$$

Observe that the identity $\Lambda = \mathbb{I}$ transformation has $\Lambda^0{}_0 = 1$ and unit determinant. Hence, any Lorentz transformation that can be tuned to $\mathbb{I}$ by adjusting its continuous parameters must have $\Lambda^0{}_0 \geq 1$ and $\det \Lambda = 1$, since the regions $\Lambda^0{}_0 \leq -1$ or $\det \Lambda = -1$ are disjoint from the former and would therefore require a *discontinuous* variation in Λ to bridge them. Additionally, notice the time reversal transformation

$$\Lambda_T \equiv \text{diag}(-1, +1, \dots, +1), \tag{10.3.96}$$

and the parity flip of the ith axis, defined as the identity matrix with its $(\mu, \nu) = (i, i)$ component replaced with -1 – i.e.,

$$\Lambda_P(i) = \text{diag}(+1, \dots, +1, -1, +1, \dots, +1), \tag{10.3.97}$$

where the -1 occurs at the ith component – each has determinant -1:

$$\det \Lambda_T = \det \Lambda_P(i) = -1. \tag{10.3.98}$$

Pure Boost: Timelike and Spacelike Cases Generalizing the 2D analysis, we now specialize the pure boost in eq. (10.3.92) to the 'timelike' and 'spacelike' cases. The former involves the unit timelike vector

$$U^\mu = \gamma(\vec{v})(1, \vec{v}), \tag{10.3.99}$$

$$\gamma(\vec{v}) \equiv \left(1 - \vec{v}^2\right)^{-1/2}; \tag{10.3.100}$$

whereby we identify $\cosh\xi = 1/\sqrt{1 - \vec{v}^2}$ and $\sinh\xi \cdot \widehat{v}^i = -v^i/\sqrt{1 - \vec{v}^2}$:

$$\Lambda_B(\vec{v}^2 < 1) = \begin{bmatrix} \gamma & -\gamma \cdot \vec{v}^{\mathrm{T}} \\ -\gamma \cdot \vec{v} & \gamma \widehat{v}\widehat{v}^{\mathrm{T}} + \mathbb{I}_{D\times D} - \widehat{v}\widehat{v}^{\mathrm{T}} \end{bmatrix}, \tag{10.3.101}$$

$$\Lambda_B(\vec{v}^2 < 1) \cdot U = \left(1, \vec{0}\right)^{\mathrm{T}}. \tag{10.3.102}$$

The latter involves the unit spacelike vector

$$U'^\mu = (1, \vec{v})/\sqrt{\vec{v}^2 - 1}, \tag{10.3.103}$$

whereby we identify $\cosh\xi = |\vec{v}|/\sqrt{\vec{v}^2 - 1}$ and $\sinh\xi \cdot \widehat{v}^i = -\widehat{v}^i/\sqrt{\vec{v}^2 - 1}$:

$$\Lambda_B(\vec{v}^2 > 1) = \begin{bmatrix} |\vec{v}|/\sqrt{\vec{v}^2 - 1} & -\widehat{v}^{\mathrm{T}}/\sqrt{\vec{v}^2 - 1} \\ -\widehat{v}/\sqrt{\vec{v}^2 - 1} & \frac{|\vec{v}|}{\sqrt{\vec{v}^2 - 1}} \widehat{v}\widehat{v}^{\mathrm{T}} + \mathbb{I}_{D\times D} - \widehat{v}\widehat{v}^{\mathrm{T}} \end{bmatrix}, \tag{10.3.104}$$

$$\Lambda_B(\vec{v}^2 > 1) \cdot U' = (0, \widehat{v})^{\mathrm{T}}. \tag{10.3.105}$$

Even though the pure boost matrix in eq. (10.3.101) is a special case of the more general eq. (10.3.92), it is the most commonly used one because of its association with boosting into the inertial frame at rest with a timelike vector field – i.e., eq. (10.3.102) – which in turn can be associated with the 'flow of time' experienced by a material body.

Problem 10.24. Verify equations (10.3.102) and (10.3.105). □

Most General Change-of-Inertial Frames As we shall argue below, massive bodies follow timelike trajectories. If we associate changing inertial frames with switching from the rest frame of one material observer into another, we may therefore use the timelike boost in eq. (10.3.101) in eq. (10.3.91) to deduce the most general change-of-frames $x^\mu = \Lambda^\mu{}_\nu x'^\nu$ is

$$\begin{bmatrix} t \\ \vec{x} \end{bmatrix} = \begin{bmatrix} \sigma_0 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \mathbb{I}_{D\times D} \end{bmatrix} \begin{bmatrix} \gamma & -\gamma \cdot \vec{v}^{\mathrm{T}} \\ -\gamma \cdot \vec{v} & \gamma \widehat{v}\widehat{v}^{\mathrm{T}} + \mathbb{I}_{D\times D} - \widehat{v}\widehat{v}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R} \end{bmatrix} \begin{bmatrix} t' \\ \vec{x}' \end{bmatrix} \tag{10.3.106}$$

$$= \begin{bmatrix} \sigma_0 \gamma(\vec{v}) \left(t' - \vec{v}^{\mathrm{T}} \cdot \widehat{R} \cdot \vec{x}' \right) \\ \gamma(\vec{v}) \left(\widehat{v} \left(\widehat{v}^{\mathrm{T}} \cdot \widehat{R} \cdot \vec{x}' \right) - \vec{v} \cdot t' \right) + \left(\widehat{R} \cdot \vec{x}' \right)_\perp \end{bmatrix}, \tag{10.3.107}$$

$$\widehat{R}^{\mathrm{T}} \widehat{R} = \mathbb{I}, \qquad \sigma_0 \in \{-1, +1\}, \qquad \widehat{v} \equiv \vec{v}/|\vec{v}|, \qquad \left(\widehat{R} \cdot \vec{x} \right)_\perp^i \equiv \left(\delta^{ij} - \widehat{v}^i \widehat{v}^j \right) \left(\widehat{R} \cdot \vec{x} \right)^j.$$

The factor $\gamma(\vec{v}) = 1/\sqrt{1 - \vec{v}^2}$ tends to unity for non-relativistic speeds $|\vec{v}| \ll 1$; whereas it blows up as $|\vec{v}|$ approaches unity. It is largely responsible for the many dramatic kinematic effects in Special Relativity.

Problem 10.25. Non-Relativistic Limit and Galilean Transformation Put back the c's and take the $c \to \infty$ limit to deduce the Galilean transformation of Newtonian mechanics:

$$t = \sigma_0 t' \qquad \text{and} \qquad \vec{x} = \widehat{R} \cdot \vec{x}' - \vec{v} \cdot t', \tag{10.3.108}$$

which *do not* contain any factors of c. □

Global versus Local Lorentz Transformations

Exponential Form For the rest of this derivation, let us focus on the case where Λ is parametrized by a set of continuous parameters such that they may tuned to yield $\Lambda = \mathbb{I}$. As we shall see, the continuous parameters in Λ tell us how large/small a rotation and/or boost is being performed. As discussed in the §(5), such a matrix that is continuously connected to the identify may write expressible in an exponential form:

$$\Lambda^\mu{}_\nu = \left(e^X \right)^\mu{}_\nu = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \left(X^\ell \right)^\mu{}_\nu . \tag{10.3.109}$$

Because we are moving indices with the metric $\eta_{\alpha\beta}$ – for e.g., $\eta_{\mu\nu} X^\mu{}_\alpha = X_{\nu\alpha}$ – the position of the indices on any object (upper or lower) is important. In particular, the Lorentz transformation itself $\Lambda^\mu{}_\nu$ has one upper and one lower index; and this means the X in $\Lambda = e^X$ must, too, have one upper and one lower index – for instance, the n-th term in the Taylor series reads:

$$\frac{1}{n!} X^\mu{}_{\sigma_1} X^{\sigma_1}{}_{\sigma_2} X^{\sigma_2}{}_{\sigma_3} \dots X^{\sigma_{n-2}}{}_{\sigma_{n-1}} X^{\sigma_{n-1}}{}_\nu . \tag{10.3.110}$$

If we use the defining relation in eq. (10.1.13), but consider it for small X only,

$$\left(\delta^\mu{}_\alpha + X^\mu{}_\alpha + \mathcal{O}\left(X^2\right) \right) \eta_{\mu\nu} \left(\delta^\nu{}_\beta + X^\nu{}_\beta + \mathcal{O}\left(X^2\right) \right) \tag{10.3.111}$$

$$= \eta_{\alpha\beta} + \delta^\mu{}_\alpha \eta_{\mu\nu} X^\nu{}_\beta + X^\mu{}_\alpha \eta_{\mu\nu} \delta^\nu{}_\beta + \mathcal{O}\left(X^2\right)$$

$$= \eta_{\alpha\beta} + X_{\alpha\beta} + X_{\beta\alpha} + \mathcal{O}\left(X^2\right) = \eta_{\alpha\beta}. \tag{10.3.112}$$

The order$-X$ terms will vanish iff $X_{\alpha\beta}$ itself (with both lower indices) or $X^{\alpha\beta}$ (with both upper indices) is anti-symmetric:

$$X_{\alpha\beta} = -X_{\beta\alpha}. \tag{10.3.113}$$

This situation is very similar to the one encountered in §(5.5), where the exponent of the rotation matrix is also anti-symmetric, except here the exponent is anti-symmetric only after its left index has been lowered. The most general $X_{\alpha\beta}$ is, hence, the superposition of the $(d^2-d)/2$ basis anti-symmetric $d \times d$ matrices. Denoting $-i(\widehat{J}^{\mu\nu})_{\alpha\beta}$ as the $(\alpha,\beta)-$component of the (μ,ν)th basis – with $+1$ filling the (μ,ν) upper-triangular component; -1 filling the (ν,μ)th lower triangular component; and 0 filling the rest of its components – followed by raising its left index with the inverse metric, we obtain in $(1+1)D$ the sole basis matrix

$$-i(\widehat{J}^{01})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \tag{10.3.114}$$

in $(2+1)D$ the three basis matrices

$$-i(\widehat{J}^{01})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{02})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{12})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}; \tag{10.3.115}$$

in $(3+1)D$ the six basis matrices

$$\begin{aligned} -i(\widehat{J}^{01})^{\alpha}{}_{\beta} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{02})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{03})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}; \\ -i(\widehat{J}^{12})^{\alpha}{}_{\beta} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{13})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, -i(\widehat{J}^{23})^{\alpha}{}_{\beta} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}; \end{aligned} \tag{10.3.116}$$

and so on. We gather that

$$\left(\exp\left(-i\xi\widehat{J}^{0i}\right)\right)^{\alpha}{}_{\beta} \tag{10.3.117}$$

describes a Lorentz boost along the ith spatial axis with rapidity $\xi \in \mathbb{R}$; whereas

$$\left(\exp\left(-i\theta\widehat{J}^{ij}\right)\right)^{\alpha}{}_{\beta} \tag{10.3.118}$$

describes a counterclockwise rotation of the $(i,j)-$plane for positive angle θ. You may further verify that our definition may be summarized as

$$-i\left(\widehat{J}^{\mu\nu}\right)^{\alpha}{}_{\beta} = \eta^{\mu\alpha}\delta^{\nu}{}_{\beta} - \eta^{\nu\alpha}\delta^{\mu}{}_{\beta} = +i\left(\widehat{J}^{\nu\mu}\right)^{\alpha}{}_{\beta}. \tag{10.3.119}$$

Because of the anti-symmetry under swapping of the labels $\mu \leftrightarrow \nu$, the superposition of these $-iJ$'s to form the most general X may be obtained using an anti-symmetric parameters-matrix $\{\omega_{\mu\nu}\}$, since any symmetric portion would drop out:

$$X^\alpha{}_\beta = -\frac{i}{2}\omega_{\mu\nu}(\widehat{J}^{\mu\nu})^\alpha{}_\beta, \tag{10.3.120}$$

$$\omega_{\mu\nu} = -\omega_{\nu\mu} \in \mathbb{R}. \tag{10.3.121}$$

These ω's need to be real because $\Lambda^\alpha{}_\beta$ is real. The general Lorentz transformation continuously connected to the identity is therefore

$$\Lambda^\alpha{}_\beta = \left(\exp\left(-\frac{i}{2}\omega_{\mu\nu}\widehat{J}^{\mu\nu}\right)\right)^\alpha{}_\beta. \tag{10.3.122}$$

If we define the boost generator as

$$\widehat{K}^i \equiv \widehat{J}^{0i} \tag{10.3.123}$$

and rename $\omega_{0i} \equiv \xi_i$, a pure boost can be expressed as

$$\Lambda(\vec{\xi})^\mu{}_\nu = \exp\left(-i\xi_i K^i\right). \tag{10.3.124}$$

A pure rotation, on the other hand, must take the form

$$\left(\exp\left(-\frac{i}{2}\omega_{mn}\widehat{J}^{mn}\right)\right)^\mu{}_\nu = \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}}_{d-1} \\ \vec{0}_{d-1} & \widehat{R} \end{bmatrix}, \tag{10.3.125}$$

where $\vec{0}_{d-1}$ is a column with $d-1$ zeroes; and $\widehat{R}$ is a $(d-1)\times(d-1)$ real orthogonal matrix.

A comparison with the choice of basis of rotation generators in eq. (5.5.43) is in order. Because the exponent here has one upper and one lower index, we have deliberately chosen our signs so that $-i(\widehat{J}^{ij})^a{}_b[\text{here}] = -i(\widehat{J}^{ij})_{ab}[(5.5.43)]$. This ensures that the Lorentz transformation $(\exp(-i\theta\widehat{J}^{ij}))^\alpha{}_\beta$ describes a counterclockwise rotation on the $(i,j)-$plane for a positive angle θ, just like $\exp(-i\theta\widehat{J}^{ij})$ in §(5.5).

Problem 10.26. Lorentz Transformations from Generators Explain why the $(2+1)D$ Lorentz transformations arising from exponentials of the generators in eq. (10.3.115) are the boosts

$$\left(\exp\left(-i\xi\widehat{J}^{01}\right)\right)^\mu{}_\nu = \begin{bmatrix} \cosh\xi & \sinh\xi & 0 \\ \sinh\xi & \cosh\xi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{10.3.126}$$

$$\left(\exp\left(-i\xi\widehat{J}^{02}\right)\right)^\mu{}_\nu = \begin{bmatrix} \cosh\xi & 0 & \sinh\xi \\ 0 & 1 & 0 \\ \sinh\xi & 0 & \cosh\xi \end{bmatrix}, \tag{10.3.127}$$

and rotation

$$\left(\exp\left(-i\theta\widehat{J}^{12}\right)\right)^\mu{}_\nu = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}; \tag{10.3.128}$$

while those from the $(3+1)D$ generators in eq. (10.3.116) are the boosts

$$\left(\exp\left(-i\xi\widehat{J}^{01}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} \cosh\xi & \sinh\xi & 0 & 0 \\ \sinh\xi & \cosh\xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{10.3.129}$$

$$\left(\exp\left(-i\xi\widehat{J}^{02}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} \cosh\xi & 0 & \sinh\xi & 0 \\ 0 & 1 & 0 & 0 \\ \sinh\xi & 0 & \cosh\xi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{10.3.130}$$

$$\left(\exp\left(-i\xi\widehat{J}^{03}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} \cosh\xi & 0 & 0 & \sinh\xi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\xi & 0 & 0 & \cosh\xi \end{bmatrix}, \tag{10.3.131}$$

and the rotations

$$\left(\exp\left(-i\theta\widehat{J}^{12}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{10.3.132}$$

$$\left(\exp\left(-i\theta\widehat{J}^{13}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & 0 & -\sin\theta \\ 0 & 0 & 1 & 0 \\ 0 & \sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{10.3.133}$$

$$\left(\exp\left(-i\theta\widehat{J}^{23}\right)\right)^{\mu}{}_{\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{bmatrix}. \tag{10.3.134}$$

Hint: These generators are really acting on a 2D subspace of Minkowski spacetime. □

Problem 10.27. Anti-Hermitian Boost, Hermitian Rotation generators Prove that the boost generators $K^i \equiv \widehat{J}^{0i}$ are anti-hermitian; while the spatial rotation generators $\widehat{J}^{ab}$ are hermitian. □

Problem 10.28. Pure Boost Along ith Cartesian Axis Show that the pure boost along the ith Cartesian axis can be expressed as

$$\left(\exp\left(-i\xi K^i\right)\right)^{\mu}{}_{\nu} = \delta^{\mu}_{0}\delta^{0}_{\nu}\cosh\xi + \delta^{\mu}_{i}\delta^{i}_{\nu}(\cosh\xi - 1) + (\delta^{\mu}_{0}\delta^{i}_{\nu} + \delta^{\mu}_{i}\delta^{0}_{\nu})\sinh\xi + \sum_{\sigma\neq 0}\delta^{\mu}_{\sigma}\delta^{\sigma}_{\nu}. \tag{10.3.135}$$

Hint: Plug in all possible (μ,ν)s and check that it yields the correct matrix. □

Transforming the generators Let us now consider

$$\Lambda^{\mu}{}_{\sigma}\left(\exp\left(-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}^{\alpha\beta}\right)\right)^{\sigma}{}_{\rho}\Lambda_{\nu}{}^{\rho} \equiv \left(\exp\left(-\frac{i}{2}\omega'_{\alpha\beta}\widehat{J}^{\alpha\beta}\right)\right)^{\mu}{}_{\nu}. \tag{10.3.136}$$

Each term in the resulting Taylor series contains products of

$$-\frac{i}{2}\omega_{\alpha\beta}\Lambda^{\mu}{}_{\sigma}\left(\widehat{J}^{\alpha\beta}\right)^{\sigma}{}_{\rho}\Lambda_{\nu}{}^{\rho} = \frac{1}{2}\omega_{\alpha\beta}\Lambda^{\mu}{}_{\sigma}\left(\eta^{\sigma[\alpha}\delta^{\beta]}_{\rho}\right)\Lambda_{\nu}{}^{\rho} \tag{10.3.137}$$

$$= \omega_{\alpha\beta}\Lambda^{\mu\alpha}\Lambda_{\nu}{}^{\beta}. \tag{10.3.138}$$

On the other hand

$$-\frac{i}{2}\omega'_{\alpha\beta}(\widehat{J}^{\alpha\beta})^{\mu}{}_{\nu} = \frac{1}{2}\omega'_{\alpha\beta}\eta^{\mu[\alpha}\delta^{\beta]}_{\nu} \tag{10.3.139}$$

$$= \frac{1}{2}\omega'^{\mu}{}_{\nu} + \frac{1}{2}\omega'{}_{\nu}{}^{\mu}. \tag{10.3.140}$$

Lowering the μ index reveals

$$\omega'_{\mu\nu} = \omega_{\alpha\beta}\Lambda_{\mu}{}^{\alpha}\Lambda_{\nu}{}^{\beta}, \tag{10.3.141}$$

which implies

$$\Lambda^{\mu}{}_{\sigma}\left(\exp\left(-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}^{\alpha\beta}\right)\right)^{\sigma}{}_{\rho}\Lambda_{\nu}{}^{\rho} = \left(\exp\left(-\frac{i}{2}\omega_{\sigma\rho}\Lambda_{\alpha}{}^{\sigma}\Lambda_{\beta}{}^{\rho}\widehat{J}^{\alpha\beta}\right)\right)^{\mu}{}_{\nu} \tag{10.3.142}$$

$$\equiv \left(\exp\left(-\frac{i}{2}\omega_{\sigma\rho}\widehat{J}'^{\sigma\rho}\right)\right)^{\mu}{}_{\nu}. \tag{10.3.143}$$

If we view this as a change-of-basis induced by the Lorentz transformation $\Lambda^{\mu}{}_{\nu}$, we see that the basis generators themselves transform as rank-2 Lorentz tensors:

$$\widehat{J}'^{\alpha\beta} = \widehat{J}^{\sigma\rho}\Lambda_{\sigma}{}^{\alpha}\Lambda_{\rho}{}^{\beta}. \tag{10.3.144}$$

Lorentz Boost: General Case A general Lorentz boost $\exp(-i\xi_i K^i)$ is parametrized by the $d-1$ component ξ_i. This suggests that the general boost is parametrized by a spatial velocity. To see this, we employ a pure rotation

$$\Lambda(\widehat{R})^{\mu}{}_{\nu} \equiv \begin{bmatrix} 1 & 0 \\ 0 & \widehat{R}^{a}{}_{b}\delta^{\mu}_{a}\delta^{b}_{\nu} \end{bmatrix}, \tag{10.3.145}$$

whose inverse is

$$\Lambda(\widehat{R})_{\mu}{}^{\nu} \equiv \begin{bmatrix} 1 & 0 \\ 0 & \widehat{R}^{a}{}_{b}\delta^{a}_{\mu}\delta^{\nu}_{b} \end{bmatrix}. \tag{10.3.146}$$

Then, according to eq. (10.3.144), the transformed boost generator can be computed as

$$\widehat{J}'^{\alpha\beta} \equiv \widehat{J}^{0j}\Lambda_{0}{}^{\alpha}\Lambda_{j}{}^{\beta} = \widehat{J}^{0j}\delta^{\alpha}_{0}(\widehat{R}^{\mathrm{T}})_{j}{}^{\beta} \tag{10.3.147}$$

$$= \widehat{J}^{0j}\delta^{\alpha}_{0}\delta^{\beta}_{l}\widehat{R}^{l}{}_{j} = \delta^{\alpha}_{0}\delta^{\beta}_{l}\widehat{R}^{l}{}_{j}\widehat{K}^{j} \tag{10.3.148}$$

$$\widehat{K}'^{l} = \widehat{R}^{l}{}_{j}K^{j}. \tag{10.3.149}$$

In words: the boost generator remains a boost generator, but rotated by the $(d-1)\times(d-1)$ rotation matrix $\widehat{R}$. But according to eq. (10.3.135), we already know to boost along the ith Cartesian axis. For some *fixed* i, by defining the spatial vector

$$\widehat{n}^j \equiv \widehat{R}^i{}_j, \tag{10.3.150}$$

– which we may readily verify is of unit length

$$\vec{n}^2 = \widehat{R}^i{}_j \widehat{R}^i{}_j = \delta^{ii} \qquad \text{(no sum over } i\text{)} \tag{10.3.151}$$

– we may assert

$$\xi \widehat{n} \cdot \vec{K} \equiv \xi n^j K^j = \xi \delta^i_l \widehat{R}^l{}_j K^j \tag{10.3.152}$$

and

$$\left(\exp\left(-i\xi\widehat{n}\cdot\vec{K}\right)\right)^\mu{}_\nu = \Lambda(\widehat{R})^\mu{}_\sigma \left(\exp\left(-i\xi K^i\right)\right)^\sigma{}_\rho \Lambda(\widehat{R})_\nu{}^\rho. \tag{10.3.153}$$

The relevant contractions are

$$\Lambda(\widehat{R})^\mu{}_\sigma \delta^\sigma_0 = \delta^\mu_0, \tag{10.3.154}$$
$$\Lambda(\widehat{R})^\mu{}_\sigma \delta^\sigma_i = \delta^\mu_a \widehat{R}^a{}_i; \tag{10.3.155}$$

and

$$\Lambda(\widehat{R})_\nu{}^\rho \delta^0_\rho = \delta^0_\nu, \tag{10.3.156}$$
$$\Lambda(\widehat{R})_\nu{}^\rho \delta^i_\rho = \delta^a_\nu \widehat{R}^a{}_i. \tag{10.3.157}$$

YZ: To be continued

> Non-compact groups (i.e., with infinite volume) such as the Lorentz group (where its rapidity runs over $\mathbb{R}$) *do not* yield finite dimensional unitary representations. Compact groups (i.e., with finite volume) such as the rotation group (where rotation angles run over a finite interval) *do* yield finite dimensional unitary representations.

Problem 10.29. Show, by a direct calculation, that $\exp(-i\xi K^1)$ and $\exp(-i\xi K^2)$ do indeed yield the boosts in equations (**??**) and (**??**) respectively. Show that $\exp(-i\theta\widehat{J}^{12})$ does indeed yield the rotation in eq. (**??**). Hint: You may write $K^j = i\,|0\rangle\,\langle j| + i\,|j\rangle\,\langle 0|$ and use a fictitious Hilbert space where $\langle\mu|\,\nu\rangle = \delta^{\mu\nu}$ and $(K^j)^\mu{}_\nu = \langle\mu\,|K^j|\,\nu\rangle$; then compute the Taylor series of $\exp(-i\xi K^j)$. $\square$

Problem 10.30. We have only seen that eq. (10.3.119) generates individual boosts and rotations in $(2+1)$D. Explain why eq. (10.3.119) does in fact generalize to the generators of boosts and rotations in all dimensions $d \geq 3$. Hint: See previous problem. $\square$

Problem 10.31. The determinant is a tool that can tell us there are certain Lorentz transformations that are disconnected from the identity – examples are

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} -\cosh\xi & \sinh\xi \\ -\sinh\xi & \cosh\xi \end{bmatrix}. \tag{10.3.158}$$

You can explain why these are disconnected from $\mathbb{I}$? $\square$

Problem 10.32. Lorentz Transforming the Generators Consider the following composition of Lorentz transformations matrices which act on Cartesian coordinate vectors $\{x^\mu\}$:

$$\Lambda^\mu{}_\sigma \left(\exp\left[-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}^{\alpha\beta}\right]\right)^\sigma{}_\rho (\Lambda^{-1})^\rho{}_\nu = \left(\exp\left[-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}''^{\alpha\beta}\right]\right)^\mu{}_\nu . \tag{10.3.159}$$

Show that

$$-\frac{i}{2}\omega_{\alpha\beta}\Lambda^\mu{}_\sigma \left(\widehat{J}^{\alpha\beta}\right)^\sigma{}_\rho \Lambda_\nu{}^\rho = -\frac{i}{2}\omega_{\alpha\beta}\Lambda_\sigma{}^\alpha\Lambda_\rho{}^\beta \left(\widehat{J}^{\sigma\rho}\right)^\mu{}_\nu \tag{10.3.160}$$

and explain why this, in turn, implies the generator on the right hand side in eq. (10.3.159) obeys the matrix equation

$$\widehat{J}''^{\alpha\beta} = \widehat{J}^{\sigma\rho}\Lambda_\sigma{}^\alpha\Lambda_\rho{}^\beta; \tag{10.3.161}$$

i.e., the (α,β) labels for the generators transform as rank-2 Lorentz tensors under the change-of-basis transformation in eq. (10.3.159). Hint: Use eq. (10.3.119). $\square$

This result in equations (10.3.159) and (10.3.161) is also consistent with the manifestly Lorentz covariant Lie algebra of eq. (10.6.8). Furthermore, if we now construct a representation of the Lorentz group – convert all the Λs into its corresponding linear operator acting on some vector space – eq. (10.3.159) will become

$$D(\Lambda)\exp\left[-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}^{\alpha\beta}\right] D(\Lambda^{-1}) = \exp\left[-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}''^{\alpha\beta}\right]; \tag{10.3.162}$$

where the operator generator transforms as

$$\widehat{J}''^{\alpha\beta} = \widehat{J}^{\sigma\rho}\Lambda_\sigma{}^\alpha\Lambda_\rho{}^\beta, \tag{10.3.163}$$

for some Lorentz transformation matrix Λ whose (μ,ν) component is $\Lambda^\mu{}_\nu$; with $\Lambda_\mu{}^\nu$ being its inverse in that $\Lambda^\mu{}_\alpha\Lambda_\mu{}^\beta = \delta_\alpha{}^\beta$.

Rotations Let us recall that any spatial vector v^i may be rotated to point along the $1-$axis while preserving its Euclidean length. That is, there is always a $\widehat{R}$, obeying $\widehat{R}^{\mathrm{T}}\widehat{R} = \mathbb{I}$ such that

$$\widehat{R}^i{}_j v^j \doteq \pm\, |\vec{v}|(1,0,\dots,0)^{\mathrm{T}}, \qquad |\vec{v}| \equiv \sqrt{\delta_{ij}v^i v^j}. \tag{10.3.164}$$

[106]Conversely, since $\widehat{R}$ is necessarily invertible, any spatial vector v^i can be obtained by rotating it from $|\vec{v}|(1,\vec{0}^{\mathrm{T}})$. Moreover, in $D+1$ notation, these rotation matrices can be written as

$$\widehat{R}^\mu{}_\nu \doteq \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R}^i{}_j \end{bmatrix} \tag{10.3.165}$$

$$\widehat{R}^0{}_\nu v^\nu = v^0, \tag{10.3.166}$$

[106]This $\widehat{R}$ is not unique: for example, by choosing another rotation matrix $\widehat{R}''$ that only rotates the space orthogonal to v^i, $\widehat{R}\widehat{R}''\vec{v}$ and $\widehat{R}\vec{v}$ both yield the same result.

$$\widehat{R}^i{}_\nu v^\nu = \widehat{R}^i{}_j v^j = (\pm|\vec{v}|, 0, \dots, 0)^{\mathrm{T}}. \tag{10.3.167}$$

These considerations tell us, if we wish to study Lorentz transformations that are *not* rotations, we may reduce their study to the $(1+1)$D case. To see this, we first observe that

$$\Lambda \begin{bmatrix} v^0 \\ v^1 \\ \vdots \\ v^D \end{bmatrix} = \Lambda \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R} \end{bmatrix} \begin{bmatrix} v^0 \\ \pm|\vec{v}| \\ \vec{0} \end{bmatrix}. \tag{10.3.168}$$

And if the result of this matrix multiplication yields non-zero spatial components, namely $(v'^0, v'^1, \dots, v'^D)^{\mathrm{T}}$, we may again find a rotation matrix $\widehat{R}'$ such that

$$\Lambda \begin{bmatrix} v^0 \\ v^1 \\ \vdots \\ v^D \end{bmatrix} = \begin{bmatrix} v'^0 \\ v'^1 \\ \vdots \\ v'^D \end{bmatrix} = \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R}' \end{bmatrix} \begin{bmatrix} v'^0 \\ \pm|\vec{v}'| \\ \vec{0} \end{bmatrix}. \tag{10.3.169}$$

At this point, we have reduced our study of Lorentz transformations to

$$\begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R}'^{\mathrm{T}} \end{bmatrix} \Lambda \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R} \end{bmatrix} \begin{bmatrix} v^0 \\ v^1 \\ \vec{0} \end{bmatrix} \equiv \Lambda' \begin{bmatrix} v^0 \\ v^1 \\ \vec{0} \end{bmatrix} = \begin{bmatrix} v'^0 \\ v'^1 \\ \vec{0} \end{bmatrix}. \tag{10.3.170}$$

Because Λ was arbitrary so is Λ', since one can be gotten from another via rotations.

Time Reversal and Parity Flips Suppose the time component of the vector v^μ were negative ($v^0 < 0$), we may write it as

$$\begin{bmatrix} -|v^0| \\ \vec{v} \end{bmatrix} = \widehat{T} \begin{bmatrix} |v^0| \\ \vec{v} \end{bmatrix}, \qquad \widehat{T} \equiv \begin{bmatrix} -1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \mathbb{I}_{D\times D} \end{bmatrix}; \tag{10.3.171}$$

where $\widehat{T}$ is the time reversal matrix since it reverses the sign of the time component of the vector. You may readily check that $\widehat{T}$ itself is a Lorentz transformation in that it satisfies $\widehat{T}^{\mathrm{T}}\eta\widehat{T} = \eta$.

Problem 10.33. Parity flip of the ith axis Suppose we wish to flip the sign of the ith spatial component of the vector, namely $v^i \to -v^i$. You can probably guess, this may be implemented via the diagonal matrix with all entries set to unity, except the ith component – which is set instead to -1.

$${}_i\widehat{P}^\mu{}_\nu v^\nu = v^\mu, \qquad \mu \neq i, \tag{10.3.172}$$

$${}_i\widehat{P}^i{}_\nu v^\nu = -v^i, \tag{10.3.173}$$

$${}_i\widehat{P} \equiv \mathrm{diag}[1, 1, \dots, 1, \underbrace{-1}_{(i+1)\text{th component}}, 1, \dots, 1]. \tag{10.3.174}$$

Define the rotation matrix $\widehat{R}^\mu{}_\nu$ such that it leaves all the axes orthogonal to the 1st and ith invariant, namely

$$\widehat{R}^\mu{}_\nu \widehat{e}^\nu_\ell = \widehat{e}^\nu_\ell, \tag{10.3.175}$$

$$\widehat{e}^{\mu}_{\ell} \equiv \delta^{\mu}_{\ell}, \qquad \ell \neq 1, i; \tag{10.3.176}$$

while rotating the $(1, i)$-plane clockwise by $\pi/2$:

$$\widehat{R} \cdot \widehat{e}_1 = -\widehat{e}_i, \qquad \widehat{R} \cdot \widehat{e}_i = +\widehat{e}_1. \tag{10.3.177}$$

Now argue that

$${}_i\widehat{P} = \widehat{R}^{\mathrm{T}} \cdot {}_1\widehat{P} \cdot \widehat{R}. \tag{10.3.178}$$

Is ${}_i\widehat{P}$ a Lorentz transformation? $\square$

As an important aside, we may generalize the two-dimensional Lorentz boost in eq. (10.3.30) to $D-$dimensions. One way to do it, is to simply append to the 2D Lorentz-boost matrix a $(D-2)\times(D-2)$ identity matrix (that leaves the $2-$ through $D-$spatial components unaltered) in a block diagonal form:

$$\Lambda^{\mu}{}_{\nu} \overset{?}{\doteq} \begin{bmatrix} \gamma & -\gamma \cdot v & 0 \\ -\gamma \cdot v & \gamma & 0 \\ 0 & 0 & \mathbb{I}_{(D-2)\times(D-2)} \end{bmatrix}. \tag{10.3.179}$$

But this is not doing much: we are still only boosting in the $1-$direction. What if we wish to boost in v^i direction, where v^i is now some arbitrary spatial vector? To this end, we may promote the $(0,1)$ and $(1,0)$ components of eq. (10.3.30) to the spatial vectors $\Lambda^0{}_i$ and $\Lambda^i{}_0$ parallel to v^i. Whereas the $(1,1)$ component of eq. (10.3.30) is to be viewed as acting on the 1D space parallel to v^i, namely the operator $v^i v^j/\vec{v}^2$. (As a check: When $v^i = v(1, \vec{0})$, $v^i v^j/\vec{v}^2 = \delta^i_1 \delta^j_1$.) The identity operator acting on the orthogonal $(D-2)\times(D-2)$ space, i.e., the analog of $\mathbb{I}_{(D-2)\times(D-2)}$ in eq. (10.3.179), is $\Pi^{ij} = \delta^{ij} - v^i v^j/\vec{v}^2$. (Notice: $\Pi^{ij} v^j = (\delta^{ij} - v^i v^j/\vec{v}^2)v^j = 0$.) Altogether, the Lorentz boost in the v^i direction is given by

$$\Lambda^{\mu}{}_{\nu}(\vec{v}) \doteq \begin{bmatrix} \gamma & -\gamma v^i \\ -\gamma v^i & \gamma \frac{v^i v^j}{\vec{v}^2} + \left(\delta^{ij} - \frac{v^i v^j}{\vec{v}^2}\right) \end{bmatrix}, \qquad \vec{v}^2 \equiv \delta_{ab} v^a v^b. \tag{10.3.180}$$

It may be worthwhile to phrase this discussion in terms of the Cartesian coordinates $\{x^\mu\}$ and $\{x'^\mu\}$ parametrizing the two inertial frames. What we have shown is that the Lorentz boost in eq. (10.3.180) describes

$$U'^{\mu} = \Lambda^{\mu}{}_{\nu}(\vec{v}) U^{\nu}, \qquad U^{\mu} = \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\lambda}; \tag{10.3.181}$$

$$U'^{\mu} = \frac{\mathrm{d}x'^{\mu}}{\mathrm{d}\lambda} = \left(\mathrm{sgn}(U^0)\sqrt{\eta_{\mu\nu} U^{\mu} U^{\nu}}, 0\right)^{\mathrm{T}}. \tag{10.3.182}$$

λ is the intrinsic 1D coordinate parametrizing the worldlines, and by definition does not alter under Lorentz boost. The above statement is therefore equivalent to

$$\mathrm{d}x'^{\mu} = \Lambda^{\mu}{}_{\nu}(\vec{v}) \mathrm{d}x^{\nu}, \tag{10.3.183}$$

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu}(\vec{v}) x^{\nu} + a^{\mu}, \tag{10.3.184}$$

where the spacetime translation a^μ shows up here as integration constants.

Problem 10.34. Rotating from Boost Along 1−Axis If we define the rotation matrix $\widehat{R}$ as the one which rotates $\widehat{e}_1$, the unit vector along the 1−axis, to $\widehat{v} \equiv \vec{v}/v$; namely,

$$\widehat{R} \cdot \widehat{e}_1 = \widehat{v}; \tag{10.3.185}$$

show that eq. (10.3.180) can be written as

$$\Lambda^\mu{}_\nu(\vec{v}) \doteq \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R}_{D\times D} \end{bmatrix} \cdot \begin{bmatrix} \gamma & -\gamma \cdot v & \vec{0}^{\mathrm{T}} \\ -\gamma \cdot v & \gamma & \vec{0}^{\mathrm{T}} \\ \vec{0} & \vec{0} & \mathbb{I}_{D\times D} \end{bmatrix} \cdot \begin{bmatrix} 1 & \vec{0}^{\mathrm{T}} \\ \vec{0} & \widehat{R}^{\mathrm{T}}_{D\times D} \end{bmatrix}. \tag{10.3.186}$$

Explain why this amounts to rotating the boost along the 1−axis to the one along the $\vec{v}/v$ direction. □

Problem 10.35. Lorentz boost in $(D+1)$−dimensions If $v^\mu \equiv (1, v^i)$, check via a direction calculation that the $\Lambda^\mu{}_\nu$ in eq. (10.3.180) produces a $\Lambda^\mu{}_\nu v^\nu$ that has no non-trivial spatial components. Also check that eq. (10.3.180) is, in fact, a Lorentz transformation. What is $\Lambda^\mu{}_\sigma(\vec{v})\Lambda^\sigma{}_\nu(-\vec{v})$?

> **Summary** Our analysis of the group of matrices $\{\Lambda\}$ obeying $\Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu \eta_{\alpha\beta} = \eta_{\mu\nu}$ reveals that these Lorentz transformations consists of: time reversals, parity flips, spatial rotations and Lorentz boosts. (The first two are discrete and the last two are continuous transformations.) A timelike vector can always be Lorentz-boosted so that all its spatial components are zero; while a spacelike vector can always be Lorentz-boosted so that its time component is zero.

Problem 10.36. Rotation Group and the Rest Frame In the rest frame of a timelike vector k^μ, we may put

$$k^\mu = \left(\omega, \vec{0}\right). \tag{10.3.187}$$

Prove that the most general Lorentz transformation $\Lambda^\mu{}_\nu$ that leaves k^μ invariant – namely, $\Lambda^\mu{}_\nu k^\nu = k^\mu$ – is a spatial rotation.

Hint: First show that invariance of k implies $\Lambda^0{}_0 = 1$ and $\Lambda^i{}_0 = 0$. Then utilize the Lorentz group condition $\Lambda^{\mathrm{T}}\eta\Lambda = \eta$; it may help to decompose it into 00, $0i$ and ij components. □

Problem 10.37. Invariance Group of Null Vectors Any null vector k^μ may, upon an appropriate spatial rotation, time-reversal, and/or parity flip, be phrased in the form

$$\Lambda'^\mu{}_\nu k^\nu \equiv k'^\mu = \omega(1, 1, \vec{0}). \tag{10.3.188}$$

Prove that the most general Lorentz transformation that leaves k' invariant, namely $\Lambda^\mu{}_\nu k'^\nu = k'^\mu$, is given by

$$\Lambda \equiv \exp\left(i\xi_{i_\perp}\left(\widehat{J}^{i_\perp 0} - \widehat{J}^{i_\perp 1}\right) - \frac{i}{2}\omega_{a_\perp b_\perp}\widehat{J}^{a_\perp b_\perp}\right), \tag{10.3.189}$$

where the basis generators can be found in eq. (10.3.119), $\{\xi_{i_\perp}\}$ are arbitrary real parameters, and the indices $\{i_\perp, a_\perp, b_\perp\}$ run over $\{2, 3, \dots, D = d-1\}$, the spatial directions orthogonal to the 1−axis. The $\{\omega_{a_\perp b_\perp}\}$, in particular, parametrize rotations of the axes perpendicular to the 1−direction – there is nothing to rotate, since the 2 through D components are zero.

Hints: If $\Lambda = \exp X$ leaves k' invariant, first argue why $X \cdot k' = 0$; i.e., solve for the linear combination of Lorentz generators where $(1, 1, \vec{0})$ is a null eigenvector. □

Problem 10.38. Null, Spacelike vs. Timelike Do null vectors span a vector space? Simiarly, do spacelike or timelike vectors span a vector space? Hint: Check for closure. □

10.4 Kinematics in Minkowski Spacetime (aka Special Relativity)

Light Cones and Causality Previously, in 2D Minkowski, we have shown that timelike vectors can always be Lorentz boosted to yield zero spatial components; whereas spacelike vectors can always boosted to yield zero time components. Let us now extend these results to the arbitrary dimensional case, by first applying a rotation. Specifically, for *any* vector U^μ, it is always possible to find a rotation $\Lambda(\widehat{R})$ such that

$$\Lambda(\widehat{R}) \cdot U = \left(U^0, \widehat{R} \cdot \vec{U}\right)^{\mathrm{T}} = \left(U^0, U'^1, \vec{0}\right)^{\mathrm{T}}. \tag{10.4.1}$$

This immediately reduces the analysis to a 2D one; whereby we may simply invoke our previous result to assert the following.

> **Proper Time** If $U^\mu = \mathrm{d}Z^\mu/\mathrm{d}\lambda$ were a timelike vector, a Lorentz transformation Λ may always be found such that the spatial components of the resulting $U' = \Lambda \cdot U$ all vanish. Moreover, this implies the Lorentz-invariant infinitesimal interval $\mathrm{d}s = |U^2|^{1/2}\mathrm{d}\lambda = \mathrm{d}\tau$ is the infinitesimal proper time τ associated with the trajectory $Z^\mu(\lambda)$; in fact, $\lambda = \tau$ iff $U^\mu U_\mu = 1$. If $(t, \vec{x})$ parametrize an inertial frame, the time coordinate t may also be viewed as the proper time of the family of observers at rest with it: $\mathrm{d}\vec{Z}/\mathrm{d}t = \vec{0}$.
>
> **Proper Length** If $U^\mu = \mathrm{d}Z^\mu/\mathrm{d}\lambda$ were a spacelike vector, a Lorentz transformation Λ may always be found such that the time component of the resulting $U' = \Lambda \cdot U$ vanish. Moreover, this implies the Lorentz-invariant infinitesimal interval $\mathrm{d}s = |U^2|^{1/2}\mathrm{d}\lambda = \mathrm{d}\ell$ is the infinitesimal proper length ℓ associated with the trajectory $Z^\mu(\lambda)$; in fact, $\lambda = \tau$ iff $U^\mu U_\mu = -1$.
>
> **Chronology-in-time** The sign of the time components of a timelike or null vector is a Lorentz invariant in that it cannot be flipped via a continuous Lorentz boost. Whereas, the sign of the time component of a spacelike vector is not a Lorentz invariant; it depends on the choice of inertial frame.

Referring to Synge's world function in eq. (10.1.27), we may define the Cartesian displacement vector

$$\Delta^\mu(x' \to x) \equiv x^\mu - x'^\mu \tag{10.4.2}$$

and recognize half the square of the geodesic distance between x' and x to be re-expressible as the Lorentz invariant inner product of Δ with itself:

$$\sigma(x, x') = \frac{1}{2}\Delta^\alpha \Delta_\alpha. \tag{10.4.3}$$

For timelike Δ we must have $\sigma > 0$; for spacelike Δ, $\sigma < 0$; and for null Δ, $\sigma = 0$. For a fixed x', its *light cone* is defined the locus of all points x connected to it via a null geodesic – namely, where $\sigma(x, x') = 0$, which translates to $(t - t')^2 = |\vec{x} - \vec{x}'|^2$. The *forward light cone* of x' refers to the portion of the light cone in its future, where $t > t'$,

$$\Delta^\mu = |\vec{x} - \vec{x}'| \left(1, \widehat{\Delta}\right); \tag{10.4.4}$$

whereas the *backward light cone* of x' refers to the portion of the light cone in its past, where $t < t'$,

$$\Delta^\mu = |\vec{x} - \vec{x}'| \left(-1, \widehat{\Delta}\right); \tag{10.4.5}$$

for arbitrary unit spatial vector $\widehat{\Delta}$. For a fixed elapsed time $t - t' = \pm|\vec{x} - \vec{x}'|$, these light cones are really the surfaces of the sphere with radius $|\vec{x} - \vec{x}'|$. Moreover, under a Lorentz transformation Λ, the displacement $x - x'$ would be altered to $\Delta' \equiv \Lambda \cdot (x - x')$; but it will not only still remain on the null cone $\Delta'^\alpha \Delta'_\alpha = 0$, it will do so on the forward one if Δ^μ were so and on the backward one if Δ^μ were so, since the sign of the time component of a null vector is Lorentz invariant.

On the other hand, the timelike case where for fixed $\sigma(x, x') > 0$ defines a Lorentz invariant pair of disjoint hyperboloids,

$$(\Delta^0)^2 - \vec{\Delta}^2 = 2\sigma > 0 \tag{10.4.6}$$

– one for $\Delta^0 > 0$ and the other for $\Delta^0 < 0$. Under a Lorentz transformation Λ, the displacement $x - x'$ would be altered to $\Delta' \equiv \Lambda \cdot (x - x')$; but it will not only still remain on this disjoint hyperboloid $\Delta'^\alpha \Delta'_\alpha = 2\sigma$, it will do so on the forward one if Δ^μ were so and on the backward one if Δ^μ were so, since the sign of the time component of a timelike vector is Lorentz invariant.

Whereas, the spacelike case defined by fixed $\sigma(x, x') < 0$ is instead a continuous hyperboloid wrapping around the time (Δ^0-)axis,

$$\vec{\Delta}^2 - (\Delta^0)^2 = 2|\sigma|. \tag{10.4.7}$$

Under a Lorentz transformation Λ, the displacement $x - x'$ would be altered to $\Delta' \equiv \Lambda \cdot (x - x')$, but it will still remain on this continuous hyperboloid $\Delta'^\alpha \Delta'_\alpha = -2|\sigma|$. We see the light cone $\sigma = 0$ delineates the 'boundary' between the timelike and spacelike cases.

These considerations are important for the notion of *causality*; the hypothesis that 'cause' must precede the 'effect'. More specifically, if a signal is generated by some source at x' propagates along a geodesic – for flat spacetime, a straight line – and is then responsible for some process to take place at x; we expect the former must occur before the latter, namely $t > t'$. For this to hold for all possible physical processes, we therefore require that such signals may only be either timelike or null; and cannot be spacelike. This is because the time component of a spacelike vector is not a Lorentz invariant: $\Delta^0 = t - t'$ can in fact be reversed in sign, and therefore 'effect' can precede 'cause' simply by changing inertial frames. We shall see explicitly that the solution of the wave equations that underlie much of classical field theories in fundamental physics does obey this form of causality: the waves generated by a source cannot propagate outside its light cone.

Problem 10.39. Timelike Geodesics and Extremal Proper Time For a timelike trajectory $z^\mu = (t, \vec{z}(t))$, by extremizing its proper time

$$\tau(1 \to 2) \equiv \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \sqrt{\eta_{\mu\nu} \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda}} = \int_{t_1}^{t_2} \mathrm{d}t \sqrt{1 - \dot{\vec{z}}^2} \tag{10.4.8}$$

while holding fixed its end points $z^\mu(\lambda_{1,2}) = z^\mu(t_{1,2}) = (t_{1,2}, \vec{z}_{1,2})$, show that this recovers the non-affine geodesic equation in eq. (10.1.17). Explain why the general solution to this extremal path yields the velocity

$$U^\mu \equiv \frac{\mathrm{d}z^\mu/\mathrm{d}t}{\sqrt{(\mathrm{d}z/\mathrm{d}t)^2}} = \frac{(1, \vec{v})}{\sqrt{1 - \vec{v}^2}}, \tag{10.4.9}$$

where $\vec{v}$ is constant and subject to the constraint $\vec{v}^2 < 1$. (This discussion is actually the special case of equations (10.1.17) and (10.1.18).) Finally, derive the conserved Noether charges corresponding to time-translation symmetry

$$U^0 = \left(1 - \vec{v}^2\right)^{-1/2}; \tag{10.4.10}$$

and space-translation symmetry

$$U^i = v^i \cdot \left(1 - \vec{v}^2\right)^{-1/2}. \tag{10.4.11}$$

These are, of course, the components of U^μ above. □

Instead of extremizing τ, we may equivalently extremize the following 'point particle' action

$$S_{\mathrm{pp}} \equiv -m \int_{t_1}^{t_2} \mathrm{d}t \sqrt{g_{\mu'\nu'} \dot{z}^\mu \dot{z}^\nu} \tag{10.4.12}$$

$$= -m \int_{t_1}^{t_2} \mathrm{d}t \sqrt{1 - \dot{\vec{z}}^2}, \tag{10.4.13}$$

where the first line holds for arbitrary coordinates, the second for the Cartesian ones at hand, and m is the mass associated with the point particle sweeping out $z^\mu = (t, \vec{z})$. For small velocities $\dot{\vec{z}}^2 \ll 1$, we may Taylor expand the square root and obtain

$$S_{\mathrm{pp}} \equiv \int_{t_1}^{t_2} \mathrm{d}t \left(-m + \frac{m}{2} \dot{\vec{z}}^2 + \mathcal{O}(\dot{\vec{z}}^4)\right), \tag{10.4.14}$$

which recovers the non-relativistic action involving kinetic energy $(m/2)\dot{\vec{z}}^2$ up to the constant shift $-m$. Because the point mass action $-m \int \mathrm{d}\tau$ does not depend on spacetime coordinates, only on their first derivatives, the spacetime translation symmetry implies there is an associated conserved Noether current – its energy-momentum vector

$$P^\mu \equiv -\frac{\partial}{\partial \dot{z}^\mu}\left(-m\sqrt{(\mathrm{d}z/\mathrm{d}\lambda)^2}\right) \tag{10.4.15}$$

$$= m \cdot U^\mu = \frac{m}{\sqrt{1-\vec{v}^2}}\,(1,\vec{v})\,, \tag{10.4.16}$$

$$v^i \equiv \mathrm{d}z^i/\mathrm{d}t. \tag{10.4.17}$$

To corroborate this interpretation, we take the non-relativistic limit:

$$P^\mu = \left(m + \frac{m}{2}\vec{v}^2 + \mathcal{O}(v^4), m\cdot\vec{v} + \mathcal{O}(v^3)\right). \tag{10.4.18}$$

If the total action of the system consists of $N > 1$ bodies –

$$S \equiv -\sum_{\mathrm{I}=1}^{N} m_{\mathrm{I}} \int \mathrm{d}\tau_{\mathrm{I}} \qquad\qquad \mathrm{d}\tau_{\mathrm{I}} = \sqrt{(\mathrm{d}z_{\mathrm{I}}/\mathrm{d}t)^2}\mathrm{d}t \tag{10.4.19}$$

– the total linear momentum is also conserved:

$$P^\mu_{\mathrm{tot}} \equiv \sum_{\mathrm{I}=1}^{N} P^\mu_{\mathrm{I}}. \tag{10.4.20}$$

Because $P^0 \to \infty$ as the particle speed approaches unity, we see that it takes an infinite amount of energy to accelerate a massive particle to the speed of light. The enormous amount of energy expenditure needed to boost a massive particle to near-unity speeds is why it is extremely costly to build a high-energy particle accelerator. Moreover, from $U^\mu U_\mu = 1$, we have the Einstein formula

$$P^2 \equiv P^\mu P_\mu = E^2 - \vec{P}^2 = m^2. \tag{10.4.21}$$

General Timelike Worldlines More generally, because we have just argued that massive bodies are expected to yield timelike – never null – trajectories we may therefore associate the unit timelike vector $U^\mu = (1,\vec{v})/\sqrt{1-\vec{v}^2}$ with the velocity tangent to their worldlines; where $\vec{v} = \mathrm{d}\vec{z}/\mathrm{d}t$ is not necessrily constant but still subject to the subluminal condition $\vec{v}^2 < 1$. For $U^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\lambda$, recall the infinitesimal proper time is $\mathrm{d}\tau = \mathrm{d}\lambda\sqrt{U^\sigma U_\sigma} = \mathrm{d}\lambda$; this confirms

$$U^\mu \partial_{z^\mu} = \frac{(1, \mathrm{d}\vec{z}/\mathrm{d}t)^\mu}{\sqrt{1-(\mathrm{d}\vec{z}/\mathrm{d}t)^2}}\frac{\partial}{\partial z^\mu} = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}, \frac{\mathrm{d}\vec{z}}{\mathrm{d}\tau}\right)^\mu \frac{\partial}{\partial z^\mu} = \frac{\mathrm{d}}{\mathrm{d}\tau}. \tag{10.4.22}$$

The

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1-(\mathrm{d}\vec{z}/\mathrm{d}t)^2}} = 1 + \frac{1}{2}(\mathrm{d}\vec{z}/\mathrm{d}t)^2 + \mathcal{O}(v^4) \tag{10.4.23}$$

relation between proper time and inertial frame time t means, in the non-relativistic limit $(\mathrm{d}\vec{z}/\mathrm{d}t)^2 \ll 1$ the two are nearly the same. The second derivative is

$$\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\sqrt{1-(\mathrm{d}\vec{z}/\mathrm{d}t)^2}} \tag{10.4.24}$$

$$= \frac{(\mathrm{d}\vec{z}/\mathrm{d}t) \cdot (\mathrm{d}^2\vec{z}/\mathrm{d}t^2)}{(1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2)^2}, \tag{10.4.25}$$

which vanishes if $c \to \infty$. The 'proper acceleration' may be expressed in terms of $t-$derivatives

$$\frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\tau^2} = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \frac{\mathrm{d}^2 z^\mu}{\mathrm{d}t^2} + \frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2}\frac{\mathrm{d}z^\mu}{\mathrm{d}t} \tag{10.4.26}$$

$$= \frac{1}{1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2}\left(0, \frac{\mathrm{d}^2\vec{z}}{\mathrm{d}t^2}\right) + \frac{(\mathrm{d}\vec{z}/\mathrm{d}t) \cdot (\mathrm{d}^2\vec{z}/\mathrm{d}t^2)}{(1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2)^2}\left(1, \frac{\mathrm{d}\vec{z}}{\mathrm{d}t}\right). \tag{10.4.27}$$

In the local rest frame of $U^\mu = (1, \mathrm{d}\vec{z}/\mathrm{d}t)/\sqrt{1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2}$, i.e., $U' \equiv \Lambda \cdot U = (1, \vec{0})$, the first derivative $\mathrm{d}\vec{z}'/\mathrm{d}t'$ is zero and therefore the local acceleration in the same rest frame becomes

$$\frac{\mathrm{d}^2 z'^\mu}{\mathrm{d}\tau^2} = \left(0, \frac{\mathrm{d}^2\vec{z}'}{\mathrm{d}t'^2}\right). \tag{10.4.28}$$

This implies the acceleration of a timelike worldline is spacelike: $(\mathrm{d}^2 z/\mathrm{d}\tau^2)^2 < 0$.

Problem 10.40. Explain why $V^\mu \equiv (1, \mathrm{d}\vec{z}/\mathrm{d}t)^\mu = (\mathrm{d}/\mathrm{d}t)(t, \vec{z})^\mu$ is *not* a Lorentz vector. The $\sqrt{1 - v^2}$ in the denominator of U^μ is vital in ensuring its Lorentz covariance. ☐

Problem 10.41. Proper time to inertial time Explain why the 'inverse' relation of $\mathrm{d}\tau = \sqrt{1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2}\mathrm{d}t$ is

$$\mathrm{d}t = \sqrt{1 + (\mathrm{d}\vec{z}/\mathrm{d}\tau)^2}\mathrm{d}\tau, \tag{10.4.29}$$

and may be integrated to obtain t in terms of τ. Hint: Consider 'squaring' the relation $\mathrm{d}\vec{z}/\mathrm{d}\tau = (1 - (\mathrm{d}\vec{z}/\mathrm{d}t)^2)^{-1/2}\mathrm{d}\vec{z}/\mathrm{d}t$. ☐

Problem 10.42. Spacelike Geodesics and Extremal Proper Length For a spacelike trajectory $z^\mu = (t, \vec{z}(t))$, by extremizing its proper length

$$\ell \equiv \int_{t_1}^{t_2} \mathrm{d}t\sqrt{\dot{\vec{z}}^2 - 1} \tag{10.4.30}$$

while holding fixed its end points $z^\mu(t_{1,2}) = (t_{1,2}, \vec{z}_{1,2})$, show that this recovers the non-affine geodesic equation in eq. (10.1.17). Explain why the general solution to this extremal path yields the velocity

$$U^\mu \equiv \frac{\mathrm{d}z^\mu/\mathrm{d}t}{\sqrt{-(\mathrm{d}\vec{z}/\mathrm{d}t)^2}} = \frac{(1, \vec{v})}{\sqrt{\vec{v}^2 - 1}}, \tag{10.4.31}$$

where $\vec{v}$ is constant and subject to the constraint $\vec{v}^2 > 1$. ☐

Collisions and Decays In elementary particle interactions, the conservation of energy-momentum plays a key role in determining the 'before' and 'after' of a given collision or decay process.

Newton's 2nd Law: Relativistic Version Since we already know $U^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\tau$ and $\mathrm{d}U^\mu/\mathrm{d}\tau \equiv \mathrm{d}^2 z^\mu/\mathrm{d}\tau^2$ are Lorentz $d-$vectors that reduce to, respectively, the Newtonian velocity and acceleration in the non-relativistic limit, we may therefore propose the relativistic version of Newton's second law to be

$$m\frac{\mathrm{d}U^\mu}{\mathrm{d}\tau} = m\frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\tau^2} = f^\mu, \tag{10.4.32}$$

where the $d-$force f^μ needs to be specified for a given physical setup. However, unlike the non-relativistic version, this f^α needs to obey the following constraint.

Problem 10.43. Orthogonality Starting with $U^\alpha U_\alpha = 1$ or otherwise, show that the relativistic force f^μ must be orthogonal to the velocity U^μ; namely, derive the relation

$$U \cdot f = 0. \tag{10.4.33}$$

Furthermore, explain why, in the instantaneous rest frame of the mass m the zeroth component of the force must be zero. □

Problem 10.44. Constant Co-Moving Force (Acceleration)

Time Dilation In an inertial frame $x^\mu = (t, \vec{x})$, the infinitesimal proper time of a given timelike trajectory $z^\mu(t) = (t, \vec{z}(t))$ according to eq. (10.4.8) is

$$\mathrm{d}\tau = \mathrm{d}t\sqrt{1 - \dot{\vec{z}}^2}. \tag{10.4.34}$$

By the timelike assumption $\dot{\vec{z}}^2 < 1$, we deduce $\mathrm{d}\tau < \mathrm{d}t$. That proper time of a moving body ticks slower than the 'rest frame' observers' t is known as *time dilation*. This appears to lead to apparent 'paradoxes' such as the following. Consider two families of observers, both at rest in their respective inertial frames, but with one set moving at a constant velocity with respect to the other. In the rest frame of the first set of observers, the second set of observers' proper times are ticking slower than theirs. But from the second set of observers' viewpoint, the first set of observers' proper times are ticking slower than theirs. Since time elapsed is not necessarily a Lorentz invariant, this is not a contradiction. To sharpen this comparison, we consider two twins born at the same spacetime point $x'^\mu = (t', \vec{x}')$ in some inertial frame. The first twin remains at the same location in space, as described by the worldline $z^\mu_{\mathrm{I}} = (s, \vec{x}')$; with zero spatial velocity. The other twin goes on a trip $z^\mu_{\mathrm{II}}(s) = (s, \vec{z}_{\mathrm{II}}(s))$, say, traveling to some distant planet and back. At some point in the future, $t > t'$, the two twins meet at $x^\mu = (t, \vec{x}')$. For the first twin, his proper time elapsed is

$$\Delta\tau_{\mathrm{I}} = \int_{t'}^{t} \sqrt{1 - \vec{0}^2}\mathrm{d}s = t - t'; \tag{10.4.35}$$

while for the second twin, his elapsed proper time is

$$\Delta\tau_{\mathrm{II}} = \int_{t'}^{t} \sqrt{1 - \dot{\vec{z}}_{\mathrm{II}}^2}\mathrm{d}s < t - t'; \tag{10.4.36}$$

because the square root factor must be less than unity for parts of his journey and, moreover, never exceeds it. The conclusion is that the second twin ages more slowly than the first who remained at rest in the inertial frame where they were born into. Because the second twin has to return to the same spatial location, he cannot be in an inertial frame for the entire journey; for e.g., turning around would require acceleration. More generally, if both twins had gone off on their separate journeys before ending back together at $(t, \vec{x}')$, their worldlines $z^\mu_{\mathrm{I,II}}(s)$ in the inertial frame would have to be specified and the $\Delta\tau_{\mathrm{I,II}} = \int_{t'}^t (1 - \dot{\vec{z}}^2_{\mathrm{I,II}})^{1/2} \mathrm{d}s$ integrals for both of them would need to be evaluated before we can predict who aged more.

Longest time, shortest length Incidentally, this analysis tells us that extremizing proper time yields the *longest* time joining timelike separated spacetime points $x' \leftrightarrow x$. We already know that the geodesic in flat spacetime is a straight line. If it is a timelike geodesic, we may boost to the frame where $x - x'$ has only time components, where $x - x' \to (t - t', \vec{0})$. Then, given any other path joining $x' \leftrightarrow x$, we have just argued that its $\Delta\tau$ is necessarily shorter than $t - t'$. A similar analysis, on the other hand, for the geodesic connecting spacelike separated points x and x' would yield the conclusion that extremizing proper length returns the *shortest* length.

Example In an inertial frame $\{x^\mu\}$, a particle with constant $d-$velocity $\mathrm{d}z^\mu/\mathrm{d}\tau = (1, \vec{v})/\sqrt{1 - \vec{v}^2}$ travels from $z^\mu(t') = (t', \vec{x}')$ to $z^\mu(t) = (t, \vec{x}' + (t - t')\vec{v})$. The trajectory itself is

$$z^\mu(t' \leq s \leq t) = (s, \vec{x}' + (s - t')\vec{v}) \,. \tag{10.4.37}$$

To the observers at rest in the inertial frame, this process took proper time $t - t'$. The coordinate velocity is $\mathrm{d}z^\mu/\mathrm{d}t = (1, \vec{v})$; and the moving particle's proper time elapsed by

$$\Delta\tau = \int_{t'}^t \sqrt{1 - \vec{v}^2} \mathrm{d}s = (t - t')\sqrt{1 - \vec{v}^2}. \tag{10.4.38}$$

Problem 10.45. Cosmic rays and muons In their rest frame, muons have an average lifetime of approximately 2.2×10^{-6} seconds. Suppose they are generated in the upper atmosphere, roughly 10 kilometers above the ground. What is the minimum speed of these cosmic ray muons, relative to the Earth, if we were to be able to observe them at the ground level? □

Length Contraction In the inertial $x-$frame, the displacement vector joining spacetime location z_1 to z_2 is

$$\Delta^\mu = z_2^\mu - z_1^\mu = (t_2 - t_1, \vec{z}_2 - \vec{z}_2). \tag{10.4.39}$$

In the $x'-$frame moving at $\vec{v} = \mathrm{d}\vec{x}/\mathrm{d}t$ relative to the $x-$frame, the components are now

$$\Delta'^\mu = \begin{bmatrix} \gamma(\vec{v}) & -\gamma(\vec{v}) \cdot \vec{v}^{\mathrm{T}} \\ -\gamma(\vec{v}) \cdot \vec{v} & \gamma(\vec{v})\widehat{v}\widehat{v}^{\mathrm{T}} + \mathbb{I}_{D\times D} - \widehat{v}\widehat{v}^{\mathrm{T}} \end{bmatrix} \cdot \Delta \tag{10.4.40}$$

$$= \begin{bmatrix} \gamma(\vec{v})\left(t_2 - t_1 - \vec{v} \cdot (\vec{z}_2 - \vec{z}_1)\right) \\ (z_2 - z_1)^i_\perp + \gamma(\vec{v})\left(\widehat{v}^i \widehat{v} \cdot (\vec{z}_2 - \vec{z}_1) - v^i(t_2 - t_1)\right) \end{bmatrix}, \tag{10.4.41}$$

$$(z_2 - z_1)^i_\perp \equiv \left(\delta^{ij} - \widehat{v}^i \widehat{v}^j\right)(z_2 - z_1)^j. \tag{10.4.42}$$

Consider an extended body at rest in the $x-$frame. By choosing $\vec{z}_1$ to lie within the body – say, its center-of-mass – we may explore the spatial extent of the body by defining the displacement vector from it to any other point $\vec{z}_2$ also within the same body; i.e., $\Delta^\mu = (0, \vec{z}_2 - \vec{z}_1)$. In the x'−frame, we then have

$$\Delta'^\mu = \begin{bmatrix} -\gamma(\vec{v})\vec{v} \cdot (\vec{z}_2 - \vec{z}_1) \\ (z_2 - z_1)^i_\perp + \gamma(\vec{v})\widehat{v}^i\widehat{v} \cdot (\vec{z}_2 - \vec{z}_1) \end{bmatrix}. \tag{10.4.43}$$

Unsurprisingly, the same displacement in this x'−frame has non-zero displacement in time. However, what is usually meant by shape, size, etc. of an extended body is the range of the spatial displacement vector at constant time $\Delta'^0 = 0$. This means we necessarily need to consider Δ^μ with non-zero time components; with the constraint that Δ^μ is spacelike. Setting to zero Δ'^0 in eq. (10.4.41) means $t_2 - t_1 = \vec{v} \cdot (\vec{z}_2 - \vec{z}_1)$. In turn,

$$\Delta'^\mu \to \begin{bmatrix} 0 \\ (z_2 - z_1)^i_\perp + \sqrt{1 - \vec{v}^2}\widehat{v} \cdot (\vec{z}_2 - \vec{z}_1) \end{bmatrix}. \tag{10.4.44}$$

An extended body in an inertial frame moving relative to it is 'length contracted' along the direction of motion. In this sense, the notion of perfectly rigid bodies is incompatible with the Lorentz invariance of flat spacetime.

Velocity Addition In a given x−frame, it is certainly legitimate to differentiate with respect to some affine parameter λ the relative Cartesian position $\Delta^\mu \equiv z^\mu_{\mathrm{I}} - z^\mu_{\mathrm{II}}$ between two point particles at z^μ_{I} and z^μ_{II}; and dub $\mathrm{d}\Delta^\mu/\mathrm{d}\lambda$ their 'relative velocity'. However, if

$$U^\mu_{\mathrm{I}} = \frac{(1, \vec{v}_{\mathrm{I}})}{\sqrt{1 - \vec{v}^2_{\mathrm{I}}}} = \frac{\mathrm{d}z^\mu_{\mathrm{I}}}{\mathrm{d}\tau_{\mathrm{I}}} \tag{10.4.45}$$

$$\text{and} \qquad U^\mu_{\mathrm{II}} = \frac{(1, \vec{v}_{\mathrm{II}})}{\sqrt{1 - \vec{v}^2_{\mathrm{II}}}} = \frac{\mathrm{d}z^\mu_{\mathrm{II}}}{\mathrm{d}\tau_{\mathrm{II}}} \tag{10.4.46}$$

are the timelike d−velocities of two worldlines in an inertial frame $\{x^\mu\}$, and if we wish to deduce how, say, velocity-II would appear in the rest frame of velocity-I, which we shall define to be the x'−frame, the answer is *not* given by such a $\mathrm{d}\Delta^\mu/\mathrm{d}\lambda$. Equations (10.3.101) and (10.3.102) teach us that, instead,

$$\frac{\mathrm{d}z'^\mu_{\mathrm{II}}}{\mathrm{d}\tau_{\mathrm{II}}} \equiv \begin{bmatrix} \gamma(\vec{v}_{\mathrm{I}}) & -\gamma(\vec{v}_{\mathrm{I}}) \cdot \vec{v}^{\mathrm{T}}_{\mathrm{I}} \\ -\gamma(\vec{v}_{\mathrm{I}}) \cdot \vec{v}_{\mathrm{I}} & \gamma(\vec{v}_{\mathrm{I}})\widehat{v}_{\mathrm{I}}\widehat{v}^{\mathrm{T}}_{\mathrm{I}} + \mathbb{I}_{D\times D} - \widehat{v}_{\mathrm{I}}\widehat{v}^{\mathrm{T}}_{\mathrm{I}} \end{bmatrix} \cdot U_{\mathrm{II}} \tag{10.4.47}$$

$$= \gamma(\vec{v}_{\mathrm{II}}) \begin{bmatrix} \gamma(\vec{v}_{\mathrm{I}})\left(1 - \vec{v}_{\mathrm{I}} \cdot \vec{v}_{\mathrm{II}}\right) \\ \vec{v}_{\mathrm{II}\perp} - \gamma(\vec{v}_{\mathrm{I}})\left(\vec{v}_{\mathrm{I}} - \widehat{v}_{\mathrm{I}}\left(\widehat{v}_{\mathrm{I}} \cdot \vec{v}_{\mathrm{II}}\right)\right) \end{bmatrix}. \tag{10.4.48}$$

are the d−velocity components of the U^μ_{II} but now in the inertial frame at rest with U^μ_{I} with the components of $\vec{v}_{\mathrm{II}}$ perpendicular to $\vec{v}_{\mathrm{I}}$ given by

$$\vec{v}_{\mathrm{II}\perp} = \vec{v}_{\mathrm{II}} - \widehat{v}_{\mathrm{I}}\left(\widehat{v}_{\mathrm{I}} \cdot \vec{v}_{\mathrm{II}}\right), \tag{10.4.49}$$

$$\widehat{v}_{\mathrm{II}} \equiv \vec{v}_{\mathrm{II}}/|\vec{v}_{\mathrm{II}}|. \tag{10.4.50}$$

When all speeds are highly subluminal ($\vec{v}^2_{\mathrm{I}}, \vec{v}^2_{\mathrm{II}} \ll 1$) the result reduces to

$$\frac{\mathrm{d}z'^\mu_{\mathrm{II}}}{\mathrm{d}\tau_{\mathrm{II}}} \approx (1, \vec{v}_{\mathrm{II}} - \vec{v}_{\mathrm{I}}). \tag{10.4.51}$$

Dividing the spatial components by the zeroth components returns the inertial frame coordinate velocity,

$$\frac{\mathrm{d}\vec{z}_{\mathrm{II}}'}{\mathrm{d}\tau_{\mathrm{II}}}\frac{\mathrm{d}\tau_{\mathrm{II}}}{\mathrm{d}z'^0} = \frac{\mathrm{d}\vec{z}_{\mathrm{II}}'}{\mathrm{d}t'} \equiv \vec{v}_{\mathrm{II}}' = \frac{\vec{v}_{\mathrm{II}\perp} - \gamma(\vec{v}_{\mathrm{I}})\left(\vec{v}_{\mathrm{I}} - \widehat{v}_{\mathrm{I}}(\widehat{v}_{\mathrm{I}} \cdot \vec{v}_{\mathrm{II}})\right)}{\gamma(\vec{v}_{\mathrm{I}})(1 - \vec{v}_{\mathrm{I}} \cdot \vec{v}_{\mathrm{II}})}. \tag{10.4.52}$$

Problem 10.46. (1+1)D In (1+1)D argue that eq. (10.4.52) reduces to

$$v_{\mathrm{II}}' = \frac{v_{\mathrm{II}} - v_{\mathrm{I}}}{1 - v_{\mathrm{I}} v_{\mathrm{II}}}. \tag{10.4.53}$$

Set $v_{\mathrm{II}} = \pm 1$. Does the answer make sense? □

Problem 10.47. Stellar Aberration Let $\mathrm{d}z^\mu/\mathrm{d}\lambda \equiv k^\mu = \omega(1, \widehat{n})$, with $\widehat{n}^2 = 1$, be the null vector associated with a photon signal from a very distant star – i.e., so that parallax effects can be ignored – as seen from the rest frame of the Sun. If $U^\mu = (1, \vec{v})/\sqrt{1 - \vec{v}^2} = \gamma(\vec{v})(1, \vec{v})$ is the 4−velocity of the Earth as it orbits the Sun, show that the components of the null vector in the instantaneous frame of the Earth is

$$k'^\mu = \omega \begin{bmatrix} \gamma(\vec{v})\left(1 - \vec{v} \cdot \widehat{n}\right) \\ \widehat{n}_\perp - \gamma(\vec{v})\left(\vec{v} - \widehat{v}\left(\widehat{v} \cdot \widehat{n}\right)\right) \end{bmatrix} \tag{10.4.54}$$

$$\widehat{n}_\perp = \widehat{n} - \widehat{v}\left(\widehat{v} \cdot \widehat{n}\right), \tag{10.4.55}$$

$$\widehat{v} \equiv \vec{v}/|\vec{v}|. \tag{10.4.56}$$

Verify that k'^0 can also be gotten from the Lorentz inner product

$$\omega' \equiv k'^0 = k \cdot U = \omega \frac{1 - \vec{v} \cdot \widehat{n}}{\sqrt{1 - \vec{v}^2}}. \tag{10.4.57}$$

Compare the angular positions of the star as seen by an observer on Earth's surface, at opposite sides of the Earth's orbit. If $\Delta\theta$ denotes the change in the angular position of the star, in particular, demonstrate that

$$\cos\Delta\theta = \frac{1 - 2\vec{v}^2 + \vec{v}^2(\widehat{v} \cdot \widehat{n})^2}{1 - \vec{v}^2(\widehat{v} \cdot \widehat{n})^2} \tag{10.4.58}$$

$$= 1 - \frac{1}{2}\left(2\sqrt{1 - (\widehat{v} \cdot \widehat{n})^2}|\vec{v}|\right)^2 + \mathcal{O}(\vec{v}^4). \tag{10.4.59}$$

(Hint: On the opposite side of the orbit, $\widehat{v}$ must be replaced with $-\widehat{v}$.) Read off $\Delta\theta$ if the non-relativistic $\vec{v}^2 \ll 1$ limit holds. □

Relativistic Beaming If instead of the Earth, we have a relativistic observer moving at speeds $v \sim 1$ relative to some inertial frame, so that $\gamma(\vec{v}) \to \infty$. The spatial components in eq. (10.4.54) would be dominated by

$$\vec{k}' \to -\omega\gamma(\vec{v})\vec{v}\left(1 - \frac{\vec{v} \cdot \widehat{n}}{\vec{v}^2}\right), \tag{10.4.60}$$

indicating stars in the sky are now bunched up in the region where the observer is moving towards. Additionally, the $\gamma(\vec{v})$ in $k'^0 = \omega'$ (cf. (10.4.54)) implies there will also be extreme blue-shifting of their light.

Doppler Frequency Shifts The frequency associated with an oscillating field such as the photon can be extracted from its phase. If we employ the JWKB approximation $\varphi \sim Ae^{iS}$ and U^μ is the unit timelike tangent vector of an observer's worldline, the signal's frequency according to the observer is the angle swept out by its phase per unit proper time:

$$\omega_{\text{obs}} \equiv \left|\frac{\mathrm{d}S}{\mathrm{d}\tau}\right| = \left|U^\mu \partial_\mu S\right|. \tag{10.4.61}$$

For φ obeying the wave equation so that $\varphi \sim e^{-ik\cdot x}$ with some null k_μ (obeying $k^2 = 0$) and the observer at rest in the inertial frame, $U^\mu \partial_\mu = \partial_t$, we obtain $|U \cdot k| = \partial_t(\omega \cdot t - \vec{k} \cdot \vec{x}) = \omega$ as expected. For general $\mathrm{d}/\mathrm{d}\tau = U^\mu \partial_\mu$,

$$\omega_{\text{obs}} = |U^\mu \partial_\mu (k_\sigma x^\sigma)| = |U^\sigma k_\sigma|. \tag{10.4.62}$$

The result in eq. (10.4.57) demonstrates the consistency of this interpretation of observed frequency by computing a given null vector in the instantaneous rest frame of an observer. The utility of formula (10.4.62) is that it holds even in curved spacetimes, provided the wavelength of the signal is significantly shorter that those of the geometric curvature, so that the JWKB approximation scheme may be applied.

Cosmic Microwave Background Photons In the rest frame of the universe we reside in, an observer would see a nearly uniform bath of cosmic microwave background (CMB) photons at temperature ~ 2.73 K. This means no matter the observation direction $\widehat{r}$, the observer would detect photons of the same average frequency ω. On the other hand, if our Earth were not in the rest frame of the cosmos, but moving at $\vec{v}$ relative to it, the observed frequency according to eq. (10.4.57) is[107]

$$\frac{\omega'}{\omega} = \frac{1 + \vec{v} \cdot \widehat{r}}{\sqrt{1 - \vec{v}^2}}, \tag{10.4.63}$$

where we have identified $\widehat{n} = -\widehat{r}$, since the photon is moving towards the observer. If the Earth were moving towards the photon, $\widehat{v} \cdot \widehat{r} = 1$, we find a *blue*-shift

$$\frac{\omega'}{\omega} \to \sqrt{\frac{1+v}{1-v}} > 1; \tag{10.4.64}$$

where $v \equiv |\vec{v}|$. If the Earth were moving in a direction perpendicular to the photon's, $\widehat{v} \cdot \widehat{r} = 0$, there is still a blue-shift

$$\frac{\omega'}{\omega} \to \frac{1}{\sqrt{1 - \vec{v}^2}} > 1. \tag{10.4.65}$$

If the Earth were moving away from the photon, $\widehat{v} \cdot \widehat{r} = -1$, we find a *red*-shift

$$\frac{\omega'}{\omega} \to \sqrt{\frac{1-v}{1+v}} < 1. \tag{10.4.66}$$

[107] See Muller, Richard A. "*The cosmic background radiation and the new aether drift.*" Scientific American 238.5 (1978): 64-80.

Problem 10.48. Show that the moving observer sees no Doppler shift along the circle-in-the-sky centered at the $\widehat{r} = \widehat{v}$ direction, with angular radius

$$\arccos\left(\frac{\sqrt{1-v^2}-1}{v}\right). \tag{10.4.67}$$

Take the non-relativistic $v \ll 1$ and ultra-relativistic $v \sim 1$ limits and describe the resulting appearances of the CMB sky. $\square$

Rotating Frames in (3+1)D Let us now consider a spatial frame $\{\vec{x}'\}$ rotating along the unit axis $\widehat{n}$ with angular frequency ω in an inertial frame $\{x^\mu\}$, and discover the relativistic versions of the 'fictitious forces' in Newtonian mechanics. Starting from eq. (5.6.20), we may relate

$$\begin{aligned} x^i &= \widehat{R}_{\widehat{n}}(\omega t)^{ij} x'^j && (10.4.68) \\ &= \widehat{n}^i\,(\widehat{n}\cdot\vec{x}') + \cos(\omega t) x'^i_\perp + \sin(\omega t)(\widehat{n}\times\vec{x}')^i; && (10.4.69) \end{aligned}$$

where the components of $\vec{x}'$ perpendicular to $\widehat{n}$ are

$$\vec{x}'_\perp = \vec{x}' - \widehat{n}\,(\widehat{n}\cdot\vec{x}')\,. \tag{10.4.70}$$

The corresponding infinitesimal spatial displacement is

$$\begin{aligned} \mathrm{d}x^i = \widehat{n}^i\,(\widehat{n}\cdot\mathrm{d}\vec{x}') + \cos(\omega t)\mathrm{d}x'^i_\perp + \sin(\omega t)(\widehat{n}\times\mathrm{d}\vec{x}')^i \\ - \omega\mathrm{d}t\left(\sin(\omega t)x'^i_\perp - \cos(\omega t)(\widehat{n}\times\vec{x}')^i\right). \end{aligned} \tag{10.4.71}$$

and the line element $\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}\vec{x}^2$ is converted into

$$\begin{aligned} \mathrm{d}s^2 &= \left(1 - (\vec{\omega}\times\vec{x}')^2\right)\mathrm{d}t^2 - 2(\vec{\omega}\times\vec{x}')\cdot\mathrm{d}\vec{x}'\mathrm{d}t - \mathrm{d}\vec{x}'^2, && (10.4.72) \\ \vec{\omega} &\equiv \omega\widehat{n}. && (10.4.73) \end{aligned}$$

You may notice the coefficient of $\mathrm{d}t^2$ goes from positive to negative as $(\vec{\omega}\times\vec{x}')^2$ increases. In fact, when all terms of the line element takes the same sign, we no longer have spacetime but only space. This apparent pathology arises because, for fixed $\vec{x}'$, the $\vec{\omega}\times\vec{x}'$ is its coordinate velocity tangent to the circle of radius $|\vec{x}'_\perp|$. When $(\vec{\omega}\times\vec{x}')^2$ equals to or exceeds unity, motion has become null or even superluminal (spacelike). As we have argued, such a motion is not possible for massive bodies; and as such we shall assume for the rest of the discussion that $|\vec{\omega}\times\vec{x}'| < 1$.

Problem 10.49. Verify that eq. (10.4.72) follows from eq. (10.4.71). Hints: You may find eq. (9.6.60) useful; and also recognize $(\widehat{n}\times\mathrm{d}\vec{x}')\cdot\vec{x}' = -(\widehat{n}\times\vec{x}')\cdot\mathrm{d}\vec{x}'$ – do you see why? $\square$

If the line element is $g_{\mu'\nu'}\mathrm{d}y^\mu\mathrm{d}y^\nu$ the associated Lagrangian for geodesic motion of some trajectory $z'^\mu(\tau)$ is $L_g = (1/2)g_{\mu'\nu'}\dot{z}'^\mu\dot{z}'^\nu$, where the overdot is a derivative with respect to proper time τ. For the case at hand,

$$L_g = \frac{1}{2}\left(1 - (\vec{\omega}\times\vec{z}')^2\right)\dot{t}^2 - (\vec{\omega}\times\vec{z}')\cdot\dot{\vec{z}}'\dot{t} - \frac{1}{2}\dot{\vec{z}}'^2. \tag{10.4.74}$$

The factor $(\vec{\omega} \times \vec{z}')^2$ is also equal to $\omega^2 \vec{z}'^2 - (\vec{\omega} \cdot \vec{z}')^2$.

The use of proper time τ, an affine parameter, means the Lagrangian itself must be $1/2$.

$$1 = \left(1 - (\vec{\omega} \times \vec{z}')^2\right) \dot{t}^2 - 2(\vec{\omega} \times \vec{z}') \cdot \dot{\vec{z}}' \dot{t} - \dot{\vec{z}}'^2 \tag{10.4.75}$$

Since L_g does not dependent on t, only on its derivative $\dot{t}$,

$$\frac{\partial L_g}{\partial \dot{t}} = \dot{t}\left(1 - (\vec{\omega} \times \vec{z}')^2\right) - (\vec{\omega} \times \vec{z}') \cdot \dot{\vec{z}}' = \text{constant}. \tag{10.4.76}$$

Additionally, a direct calculation returns

$$\frac{\partial L_g}{\partial \dot{z}'^i} = -\dot{t}(\vec{\omega} \times \vec{z}')^i - \dot{x}'^i, \tag{10.4.77}$$

$$-\frac{\mathrm{d}}{\mathrm{d}\tau} \frac{\partial L_g}{\partial \dot{z}'^i} = \ddot{x}'^i + \ddot{t}(\vec{\omega} \times \vec{z}')^i + \dot{t}(\vec{\omega} \times \dot{\vec{z}}')^i; \tag{10.4.78}$$

and

$$-\frac{\partial L_g}{\partial z'^i} = -\dot{t}^2 \left(\vec{\omega} \times (\vec{\omega} \times \vec{z}')\right)^i + \dot{t}(\dot{\vec{z}}' \times \vec{\omega})^i. \tag{10.4.79}$$

Altogether, we find that (Cartesian) spatial acceleration in a rotating frame is driven by three different 'fictitious forces per unit mass':

$$\ddot{\vec{z}}' = -\ddot{t}(\vec{\omega} \times \vec{z}') - \dot{t}^2 \vec{\omega} \times (\vec{\omega} \times \vec{z}') - 2\dot{t}(\vec{\omega} \times \dot{\vec{z}}'), \tag{10.4.80}$$

where the double cross product can also be written as

$$\vec{\omega} \times (\vec{\omega} \times \vec{z}') = \vec{\omega}(\vec{\omega} \cdot \vec{z}') - \omega^2 \vec{z}'. \tag{10.4.81}$$

Finally, multiplying throughout by the particle's mass m, and taking the non-relativistic limit where $\tau \approx t$, so that $\mathrm{d}t/\mathrm{d}\tau \approx 1$, $\mathrm{d}^2 t/\mathrm{d}\tau^2 \approx 0$, and

$$m \frac{\mathrm{d}^2 \vec{z}'}{\mathrm{d}t^2} = \vec{F}_{\text{Coriolis}} + \vec{F}_{\text{Centrifugal}}; \tag{10.4.82}$$

with the Coriolis and centrifugal forces taking, respectively, the forms

$$\vec{F}_{\text{Coriolis}} = -2m\vec{\omega} \times \frac{\mathrm{d}\vec{z}'}{\mathrm{d}t} \tag{10.4.83}$$

and

$$\vec{F}_{\text{Centrifugal}} = -m\vec{\omega} \times (\vec{\omega} \times \vec{z}'). \tag{10.4.84}$$

Problem 10.50. Non-Relativistic Rotating Frame in (3+1)D [108]In this problem, we shall explore a different route to derive the fictitious forces in eq. (10.4.82); by exploiting

[108]The above discussion and this problem is a response to Kuan Nan Lin's question regarding how to use differential geometry to describe rotating frames in classical mechanics.

the point particle action in eq. (10.4.12) but using the inertial frame time t as the non-affine parameter. Begin by explaining the relation between proper time and inertial frame time t:

$$\mathrm{d}\tau = \sqrt{g_{\mu\nu}\frac{\mathrm{d}z'^\mu}{\mathrm{d}t}\frac{\mathrm{d}z'^\mu}{\mathrm{d}t}}\mathrm{d}t \tag{10.4.85}$$

$$= \left(1 - (\vec{\omega}\times\vec{z}')^2 - 2(\vec{\omega}\times\vec{z}')\cdot\mathrm{d}\vec{z}'/\mathrm{d}t - (\mathrm{d}\vec{z}'/\mathrm{d}t)^2\right)^{1/2}\mathrm{d}t. \tag{10.4.86}$$

Next, argue that the non-relativistic limit of the point particle action is

$$S_{\rm pp} = -m\int_{t=t_1}^{t=t_2}\mathrm{d}\tau = \int_{t_1}^{t_2}\left(-m + L_{\rm NR} + \mathcal{O}(v^3)\right)\mathrm{d}t; \tag{10.4.87}$$

where the $\mathcal{O}(v^2)$ Lagrangian for the rotating frame is

$$L_{\rm NR} = \frac{m}{2}\left(\frac{\mathrm{d}\vec{z}'}{\mathrm{d}t}\right)^2 + \frac{m}{2}\left(\vec{\omega}\times\vec{z}'\right)^2 + m(\vec{\omega}\times\vec{z}')\cdot\frac{\mathrm{d}\vec{z}'}{\mathrm{d}t}. \tag{10.4.88}$$

Here, by v^n we mean there are n factors of speed $\omega|\vec{z}'_\perp|$ or $|\mathrm{d}\vec{z}'/\mathrm{d}t|$ in the Lagrangian. Extremize this point particle action and recover the above Coriolis and centrifugal forces. □

Transmission of null signals between timelike worldlines As a first approximation, the primary means we experience the world and the universe at large is through the detection of null signals emitted or reflected from material bodies at a distance. To this end, let us model this by considering two given timelike worldlines in flat spacetime, $z^\mu_{1,2}$, in an inertial frame with Cartesian coordinates $(t,\vec{x})$. Let the first worldline be emitting a geodesic null signal at time t_1, which then hits the second some time t_2 later; namely, the geodesic separation between these two events must be zero $\sigma(z_1(t_1), z_2(t_2)) = 0$. Since the signal is null, and we are assuming causality holds, this translates to the retarded time constraint

$$t_2 - t_1 = |\vec{z}_2(t_2) - \vec{z}_1(t_1)|. \tag{10.4.89}$$

This can be solved for t_2 and, hence, $\vec{z}_2(t_2)$ – as long as $(t_1, \vec{z}_1(t_1))$ is known.

An example is the transmission of null signals between two timelike geodesics, $z^\mu_{1,2}$, which we may parametrize as

$$z^\mu_{1,2}(t) = (0, \vec{z}'_{1,2}) + t(1, \vec{v}_{1,2}), \tag{10.4.90}$$

where $\vec{z}'_{1,2}$ are their respective 'initial' positions and $\mathrm{d}\vec{z}_{1,2}/\mathrm{d}t = \vec{v}_{1,2}$ constant coordinate velocities. The retarded time constraint reduces to

$$t_2 - t_1 = \left|\vec{z}'_2 - \vec{z}'_1 + t_2\vec{v}_2 - t_1\vec{v}_1\right|. \tag{10.4.91}$$

Some algebra would uncover, the positive solution to $t_2 - t_1$ is

$$t_2 - t_1 = \frac{\sqrt{C_1^2\left(1-\vec{v}_2^2\right) + C_2^2} + C_2}{1-\vec{v}_2^2}, \tag{10.4.92}$$

$$C_1^2 \equiv (\vec{z}'_2 - \vec{z}'_1 + t_1(\vec{v}_2 - \vec{v}_1))^2, \tag{10.4.93}$$

$$C_2 \equiv (\vec{z}_2' - \vec{z}_1' + t_1(\vec{v}_2 - \vec{v}_1)) \cdot \vec{v}_2. \tag{10.4.94}$$

Since the coordinate velocities are constant, up to an arbitrary additive shift, the inertial frame times can be converted into the proper times $\tau_{1,2}$ of the two geodesics:

$$t_{1,2} = \frac{\tau_{1,2}}{\sqrt{1 - \vec{v}_{1,2}^2}}. \tag{10.4.95}$$

Furthermore, suppose the first geodesic emits a sharp null pulse at regular (proper time) intervals in his rest frame; namely, $\tau_1 = n \cdot \tau_0$ for integer n and constant τ_0. Then, the proper time $\tau_{2,n}$ when the nth signal is received by the second worldline is given by

$$\tau_{2,n} = \sqrt{\frac{1 - \vec{v}_2^2}{1 - \vec{v}_1^2}} n \cdot \tau_0 + \frac{\sqrt{C_{1,n}^2 (1 - \vec{v}_2^2) + C_{2,n}^2} + C_{2,n}}{\sqrt{1 - \vec{v}_2^2}}, \tag{10.4.96}$$

$$C_{1,n}^2 \equiv \left(\vec{z}_2' - \vec{z}_1' + \frac{n \cdot \tau_0}{\sqrt{1 - \vec{v}_1^2}} (\vec{v}_2 - \vec{v}_1) \right)^2, \tag{10.4.97}$$

$$C_{2,n} \equiv \left(\vec{z}_2' - \vec{z}_1' + \frac{n \cdot \tau_0}{\sqrt{1 - \vec{v}_1^2}} (\vec{v}_2 - \vec{v}_1) \right) \cdot \vec{v}_2. \tag{10.4.98}$$

According to the second worldline, his proper time between the nth signal and the $(n+1)$th one is therefore

$$\begin{aligned} \tau_{2,n+1} - \tau_{2,n} = &\frac{\sqrt{1 - \vec{v}_2^2}}{\sqrt{1 - \vec{v}_1^2}} \tau_0 \\ &+ \frac{1}{\sqrt{1 - \vec{v}_2^2}} \left(\sqrt{C_{1,n+1}^2 (1 - \vec{v}_2^2) + C_{2,n+1}^2} - \sqrt{C_{1,n}^2 (1 - \vec{v}_2^2) + C_{2,n}^2} + C_{2,n+1} - C_{2,n} \right). \end{aligned} \tag{10.4.99}$$

One can see that, even for constant velocity motion, the comparison of clocks attached to geodesic observers by sending null signals between them readily yields complicated answers. Moreover, this sort of proper time comparison is often said to be analogous to the Doppler frequency shift. However – from this explicit calculation – this interpretation does not hold in general. Instead, it is only when the period τ_0 is short enough, when compared to the other length scales at hand, that a parallel can be drawn.

Problem 10.51. Verify that eq. (10.4.91) does lead to eq. (10.4.92). □

Problem 10.52. Roemer, Jupiter's satellites, and c In this problem, we shall explore a toy problem that captures the essence of Ole Roemer's 1676 measurement of the speed of light c; assuming the rest frame of the Sun is an inertial one. For simplicity, let us assume the Earth sweeps out a circular orbit on the $(1,2)-$plane centered at the Sun's spatial location:

$$Z^\mu = (t, \ell_E \cos(\omega_E t) \widehat{e}_1 + \ell_E \sin(\omega_E t) \widehat{e}_2), \tag{10.4.100}$$

where ω_E is the orbital angular frequency and $\widehat{e}_{1,2}$ are mutually orthonormal spatial basis vectors. Instead of Jupiter and its satellites, let us consider a beacon – an astrophysical 'light house' – placed at a fixed spatial location on the same 2D plane containing Earth's orbit,

$$L^\mu = (t, \ell_J \cos\psi \widehat{e}_1 + \ell_J \sin\psi \widehat{e}_2), \tag{10.4.101}$$

where the spatial radius ℓ_J is equal to the radius of Jupiter's orbit around the Sun. This beacon broadcasts an isotropic pulse of null rays periodically, with period τ; namely, at times

$$t_{i,n} \equiv n \cdot \tau, \qquad n \in \mathbb{Z}. \tag{10.4.102}$$

Assume that $\omega_E \tau$ is an extremely small angle; use, for example, the orbital period of Io, one of Jupiter's satellites, to justify this.

Show that, for the detector attached to the Earth, its proper time elapsed Δs over the inertial frame time interval $t \in [t_1, t_2]$ is given by

$$\Delta s = \sqrt{1 - (\omega_E \ell_E)^2}(t_2 - t_1). \tag{10.4.103}$$

Notice the square root factor is the relativistic time dilation due to the motion of the Earth; it is constant even though the Earth's rest frame is not an inertial one, because we have assumed the Earth is moving at constant speed $v = \omega_E \ell_E$.

Next, show that the receipt inertial frame time $t_{f,n}$ of the pulse emitted at $t_{i,n}$ obeys

$$t_{f,n} - t_{i,n} = \sqrt{\ell_E^2 + \ell_J^2}\sqrt{1 - 2\frac{\ell_E \ell_J}{\ell_E^2 + \ell_J^2}\cos(\omega_E t_{f,n} - \psi)}. \tag{10.4.104}$$

Explain why $\omega_E t_{f,n} - \psi$ is the angle between Earth's position at receipt time and the beacon's, as measured from the Sun. Do you see, when $\omega_E t_{f,n} - \psi = \pm\pi/2$, the Earth is moving directly away $(+)$ and towards $(-)$ the beacon?

We may now deduce the observed period of the beacon τ', according to the Earth-bound detector, to be simply the proper time interval in eq. (10.4.103) between the receipt of consecutive signals,

$$\tau' = \sqrt{1 - (\omega_E \ell_E)^2}(t_{f,n+1} - t_{f,n}). \tag{10.4.105}$$

Verify $\omega_E \ell_E \sim 10^{-4}$ and $\ell_E/\ell_J \sim 1/5$; hence the former may be neglected relative to the latter. Expand up to first order in ℓ_E/ℓ_J and zeroth order in $\omega_E \ell_E$ to show that when the Earth-Sun-Jupiter system forms a straight line is space, the observed period is

$$\tau_0' \approx \tau; \tag{10.4.106}$$

whereas when the Earth-Sun-Jupiter system forms a right-angle triangle in space, the observed periods are

$$\tau_\pm' \approx \tau \left(1 \pm \frac{\omega_E \ell_E}{c}\right), \tag{10.4.107}$$

where c has been restored, and the $+$ sign is when the Earth is moving away from Jupiter and $-$ sign is when the Earth is moving towards Jupiter. Hint: Taylor expand the square root in eq. (10.4.104) and compute $t_{f,n+1} - t_{f,n}$. You should find

$$\tau' \approx \tau \left(1 + \frac{\omega_E \ell_E}{c}\sin\left(\omega_E t_{f,n} - \psi\right)\right). \tag{10.4.108}$$

Deduce that

$$c = 2(\omega_E \ell_E) \frac{\tau_0}{\tau'_+ - \tau'_-}. \tag{10.4.109}$$

Does this formula take into account the relativistic time dilation due to Earth's motion relative to the Sun?

Provided the orbital parameters of Earth and Jupiter are already known to Roemer, we see the speed of light may be determined by recording the three observed orbital periods of Io, one of Jupiter's satellites – when the Earth-Sun-Jupiter forms a straight line in space, and when it forms a right triangle. □

10.5 Poincaré Transformations: Including Spacetime Translations

Poincaré: Including Spacetime Translations For the rest of this section, we shall consider the inclusion of (constant) spacetime translations, in addition to Lorentz transformations. The full Poincaré transformation may be implemented in two steps, first by a Lorentz transformation, followed by a spacetime translation:

$$x^\mu \to \Lambda^\mu{}_\nu x^\nu \to \Lambda^\mu{}_\nu x^\nu + a^\mu. \tag{10.5.1}$$

Specifically, under the following change from inertial frame $\{x^\mu\}$ to $\{x'^\mu\}$ via

$$x^\mu \equiv \Lambda^\mu{}_\nu x'^\nu + a^\mu, \tag{10.5.2}$$

for Lorentz transformations $\Lambda^\mu{}_\nu$ and constant spacetime displacement vector a^μ; the Minkowski spacetime geometry is left invariant:

$$\eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \eta_{\mu\nu} \mathrm{d}x'^\mu \mathrm{d}x'^\nu. \tag{10.5.3}$$

This implies the associated linear operator must take the form

$$D(\Lambda, a) = \exp\left(-ia^\gamma P_\gamma\right) \exp\left(-\frac{i}{2}\omega_{\mu\nu} J^{\mu\nu}\right), \tag{10.5.4}$$

where the $e^{-ia\cdot P}$ implements spacetime translations; while the $e^{-(i/2)\omega\cdot J}$ Lorentz transformations.

Problem 10.53. $(d+1)\times(d+1)$ **Poincaré Transformation Matrix** Prove that the Poincaré transformation in eq. (10.1.12) also defines a group. To systemize the discussion, first promote the spacetime coordinates to $d+1$ dimensional objects:

$$x^{\mathfrak{A}} = (x^\mu, 1)^{\mathrm{T}} \qquad \text{and} \qquad x'^{\mathfrak{A}} = (x'^\mu, 1)^{\mathrm{T}}, \tag{10.5.5}$$

with $\mathfrak{A} = 0, 1, 2, \dots, d-1, d$. Then define the matrix

$$\Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a] = \begin{bmatrix} \Lambda^\mu{}_\nu & a^\mu \\ 0\dots 0 & 1 \end{bmatrix}; \tag{10.5.6}$$

namely, its upper left $d\times d$ block is simply the Lorentz transformation $\Lambda^\mu{}_\nu$; while its rightmost column is $(a^\mu, 1)^{\mathrm{T}}$ and its bottom row is $(0\dots 0\ 1)$.

First check that $x^{\mathfrak{A}} = \Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a] x'^{\mathfrak{B}}$ is equivalent to eq. (10.1.12). You should also find that

$$\Pi[\Lambda_1, a_1] \cdot \Pi[\Lambda_2, a_2] = \Pi[\Lambda_1 \cdot \Lambda_2, \Lambda_1 \cdot a_2 + a_1] \tag{10.5.7}$$

and therefore its inverse is

$$\Pi[\Lambda, a]^{-1} = \Pi[\Lambda^{-1}, -\Lambda^{-1} \cdot a]. \tag{10.5.8}$$

Moreover, verify that the general element is indeed a Lorentz transformation followed by a spacetime translation:

$$\Pi[\mathbb{I}, a] \cdot \Pi[\Lambda, 0] = \Pi[\Lambda, a]. \tag{10.5.9}$$

Then proceed to explain why these set of matrices $\{\Pi^{\mathfrak{A}}{}_{\mathfrak{B}}[\Lambda, a]\}$ for different Lorentz transformations Λ and translation vectors a, with the usual rules of matrix multiplication, together define a group. $\square$

10.6 Representations of the Poincaré Group: Generalities

Physical applications often require not only the direct study of spacetime symmetries, but how they impact the quantum states describing physical systems. That is, we need to be able to implement spatial rotations, Lorentz boosts, parity flips, time reversals, and spacetime translations on linear algebraic systems. This is the subject of *group representation theory*, where group elements are converted into linear operators acting on some vector space in such a manner as to preserve the fundamental multiplication rules of the original group. That is, if $g_{1,2,3}$ are group elements and $g_1 \cdot g_2 = g_3$, and if $D(g)$ is the linear operator corresponding to the element g, we must have

$$D(g_1) D(g_2) = D(g_3) = D(g_1 \cdot g_2). \tag{10.6.1}$$

We first write

$$D(\Lambda, a) = \exp(-i a^\gamma P_\gamma) \exp\left(-\frac{i}{2} \omega_{\mu\nu} J^{\mu\nu}\right), \tag{10.6.2}$$

where, now, $\{P_\gamma\}$ and $J^{\mu\nu}$ are linear operators acting on some vector space. In order to preserve the group multiplication rules, we shall demand that these operators obey the same commutation relations that the original group generators do. The rationale for this requirement is, in turn, due to the Baker-Campbell-Hausdorff formula. It tells us the product of exponentials is given by

$$e^X e^Y = \exp\left(X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] - \frac{1}{12}[Y, [X, Y]] + \dots\right), \tag{10.6.3}$$

$$[A, B] \equiv AB - BA; \tag{10.6.4}$$

where the exponent on the right hand involves sums of commutators $[\cdot, \cdot]$, commutators of commutators, commutators of commutators of commutators, etc.

If the generic form of the Lorentz transformation in eq. (10.3.109) holds, we would expect that the product of two Lorentz transformations to yield the same exponential form:

$$\exp\left(-\frac{i}{2}a_{\mu\nu}J^{\mu\nu}\right)\exp\left(-\frac{i}{2}b_{\alpha\beta}J^{\alpha\beta}\right)=\exp\left(-\frac{i}{2}c_{\delta\gamma}J^{\delta\gamma}\right). \tag{10.6.5}$$

Comparison with eq. (5.1.46) tells us, in order for the product of two Lorentz transformations to return the exponential form on the right hand side, the commutators of the generators $\{J^{\mu\nu}\}$ ought to return linear combinations of the generators. This way, higher commutators will continue to return further linear combinations of the generators, which then guarantees the form on the right hand side of eq. (10.6.5). More specifically, according to eq. (5.1.46), the first commutator would yield

$$\begin{aligned}
e^{-\frac{i}{2}a_{\mu\nu}J^{\mu\nu}}e^{-\frac{i}{2}b_{\mu\nu}J^{\mu\nu}} &= \exp\Bigg[-\frac{i}{2}(a_{\mu\nu}+b_{\mu\nu})J^{\mu\nu}+\frac{1}{2}\left(-\frac{i}{2}\right)^2 a_{\mu\nu}b_{\alpha\beta}\left[J^{\mu\nu},J^{\alpha\beta}\right] \\
&\qquad +\frac{1}{12}\left(-\frac{i}{2}\right)^3 a_{\sigma\rho}a_{\mu\nu}b_{\alpha\beta}\left[J^{\sigma\rho},\left[J^{\mu\nu},J^{\alpha\beta}\right]\right]+\dots\Bigg] &(10.6.6)\\
&= \exp\Bigg[-\frac{i}{2}(a_{\mu\nu}+b_{\mu\nu})J^{\mu\nu}+\frac{1}{2}\left(-\frac{i}{2}\right)^2 a_{\mu\nu}b_{\alpha\beta}Q^{\mu\nu\alpha\beta}{}_{\kappa\xi}J^{\kappa\xi} \\
&\qquad +\frac{1}{12}\left(-\frac{i}{2}\right)^3 a_{\sigma\rho}a_{\mu\nu}b_{\alpha\beta}Q^{\mu\nu\alpha\beta}{}_{\kappa\xi}Q^{\sigma\rho\kappa\xi}{}_{\omega\lambda}J^{\omega\lambda}+\dots\Bigg], &(10.6.7)
\end{aligned}$$

for appropriate complex numbers $\{Q^{\mu\nu\alpha\beta}{}_{\lambda\tau}\}$.

This is precisely what occurs. The commutation relations between generators of a general Lie group is known as its Lie algebra. For the Lorentz generators, a direct computation using eq. (10.3.119) would return:

Lie Algebra for $\mathbf{SO}_{D,1}$

$$\left[J^{\mu\nu},J^{\rho\sigma}\right]=i\left(\eta^{\nu\rho}J^{\mu\sigma}-\eta^{\mu\rho}J^{\nu\sigma}+\eta^{\mu\sigma}J^{\nu\rho}-\eta^{\nu\sigma}J^{\mu\rho}\right). \tag{10.6.8}$$

Problem 10.54. Verify eq. (10.6.8) using eq. (10.3.119). □

Problem 10.55. Generators form a vector space Remember that linear operators acting on a Hilbert space themselves form a vector space. Consider a collection of linearly independent linear operators $\{L_1,L_2,\dots,L_N\}$. Suppose they are closed under commutation, namely

$$[L_{\mathrm{A}},L_{\mathrm{B}}]=\sum_{\mathrm{C}=1}^{N}\chi_{\mathrm{ABC}}L_{\mathrm{C}}; \tag{10.6.9}$$

for any A and B; and the $\{\chi_{\mathrm{ABC}}\}$ here are (complex) numbers. Prove that these N operators form a vector space. □

Problem 10.56. Position Representation Let $|x\rangle$ be the spacetime position representation, so that (cf. (10.6.2))

$$D(\Lambda, a)\,|x\rangle \equiv |\Lambda \cdot x + a\rangle\,. \tag{10.6.10}$$

Prove that this definition yields a unitary $D(\Lambda, a)$ provided

$$\langle x|\, x'\rangle = \delta^{(d)}(x - x'). \tag{10.6.11}$$

For an arbitrary state consider the infinitesimal version of $\langle f|\, D(\Lambda, a)\,|x\rangle = \langle f|\,\Lambda \cdot x + a\rangle$; and deduce the following position spacetime representations of the Poincaré generators:

$$\langle x\,|P_\mu|\, f\rangle = -i\partial_\mu \,\langle x|\, f\rangle\,, \tag{10.6.12}$$
$$\langle x\,|J^{\mu\nu}|\, f\rangle = i x^{[\mu}\partial^{\nu]}\,\langle x|\, f\rangle\,. \tag{10.6.13}$$

Hint: Use Taylor expansion and eq. (10.3.119) for the infinitesimal form of Λ. □

We may arrive at eq. (10.5.2) in two steps: (I) translation by a^μ, namely $x^\mu \to x^\mu + a^\mu$; followed by (II) Lorentz transformation by $\Lambda^\mu{}_\nu$, namely $x^\mu + a^\mu \to \Lambda^\mu{}_\nu x^\nu + a^\mu$. Altogether, the Lie Algebra of these operators are

$$[P_\mu, P_\nu] = 0, \tag{10.6.14}$$
$$[J^{\mu\nu}, P^\rho] = -i\eta^{\rho[\mu}P^{\nu]}, \tag{10.6.15}$$
$$[J^{\mu\nu}, J^{\rho\sigma}] = i\left(\eta^{\nu\rho}J^{\mu\sigma} - \eta^{\mu\rho}J^{\nu\sigma} + \eta^{\mu\sigma}J^{\nu\rho} - \eta^{\nu\sigma}J^{\mu\rho}\right). \tag{10.6.16}$$

Problem 10.57. $\mathrm{SO}_{D,1}$ Lie Algebra from Position Representation Use eq. (10.6.13) to recover the Lie algebra of the Lorentz group in eq. (10.6.8) □

Lie Algebra for Poincaré Group From eq. (10.6.12) we see that the momentum operators P_μ commute among themselves. Since we have already obtained the Lie algebra for the Lorentz group, what remains is the commutation relations the momentum operator P_μ and the Lorentz generators. A short calculation tells us

$$\left[J^{\alpha\beta}, P_\mu\right] = \left[x^{[\alpha}\partial^{\beta]}, \partial_\mu\right] \tag{10.6.17}$$
$$= \left(x^{[\alpha}\left[\partial^{\beta]}, \partial_\mu\right] + \left[x^{[\alpha}, \partial_\mu\right]\partial^{\beta]}\right) \tag{10.6.18}$$
$$= -i\delta^{[\alpha}{}_\mu\left(-i\partial^{\beta]}\right). \tag{10.6.19}$$

We gather, the Lie algebra for the Poincaré group is eq. (10.6.8) together with

$$[P_\mu, P_\nu] = 0 \qquad \text{and} \qquad \left[J^{\alpha\beta}, P_\mu\right] = i\delta^{[\alpha}{}_\mu P^{\beta]}. \tag{10.6.20}$$

Transforming the Poincaré Generators Exploiting the position spacetime basis, we may deduce how the Poincaré group generators themselves transform under a Poincaré transformation. Keeping in mind equations (10.5.7) and (10.5.8), we consider

Problem 10.58. Infinite Dimensional Unitary Representation of Lorentz Consider the space spanned by the eigenkets of 'on-shell' momentum $\{|k\rangle\}$, with $k_\alpha k^\alpha \equiv k^2 = m^2 \geq 0$. First, explain why the following integration measures are Lorentz invariant:

$$\frac{\mathrm{d}^{D+1}k}{(2\pi)^D}\Theta(\pm k_0)\delta\left(k^2 - m^2\right) \tag{10.6.21}$$

and

$$\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D(2E_{\vec{k}})}, \qquad \text{with} \qquad E_{\vec{k}} \equiv \sqrt{\vec{k}^2 + m^2}. \tag{10.6.22}$$

Next, explain why the inner product

$$\langle k|\, k'\rangle = 2E_{\vec{k}}(2\pi)^D\delta^{(D)}\left(\vec{k} - \vec{k}'\right) \tag{10.6.23}$$

is Lorentz invariant. Finally, if the Lorentz transformation operator $D(\Lambda_\alpha{}^\beta)$ is defined as

$$D(\Lambda)\,|k\rangle = |k'\rangle\,, \qquad \text{with} \qquad k'_\alpha \equiv \Lambda_\alpha{}^\beta k_\beta; \tag{10.6.24}$$

explain why $D(\Lambda)$ is unitary. □

Vector and Tensor Operators Suppose V^μ is a Hermitian vector operator with eigenstate $|v\rangle$. Then

$$D(\Lambda, 0)\,|v\rangle = |\Lambda \cdot v\rangle \tag{10.6.25}$$

means

$$D(\Lambda, 0)^{-1}V^\mu D(\Lambda, 0)\,|v\rangle = \Lambda^\mu{}_\nu v^\nu D(\Lambda, 0)^{-1}\,|\Lambda \cdot v\rangle \tag{10.6.26}$$

$$= \Lambda^\mu{}_\nu V^\nu\,|v\rangle\,. \tag{10.6.27}$$

Since $|v\rangle$ was arbitrary we therefore see that

$$D(\Lambda, 0)^{-1}V^\mu D(\Lambda, 0) = \Lambda^\mu{}_\nu V^\nu. \tag{10.6.28}$$

We may now take eq. (10.6.28) to *define* a vector operator in Minkowski spacetime. The corresponding definition for a Lorentz tensor operator would be

$$D(\Lambda, 0)^{-1}T^{\mu_1\dots\mu_A}{}_{\nu_1\dots\nu_B}D(\Lambda, 0) = \Lambda^{\mu_1}{}_{\alpha_1}\dots\Lambda^{\mu_A}{}_{\alpha_A}\Lambda_{\nu_1}{}^{\beta_1}\dots\Lambda_{\nu_B}{}^{\beta_B}T^{\alpha_1\dots\alpha_A}{}_{\beta_1\dots\beta_B}. \tag{10.6.29}$$

From these relations, scalar operators may be readily constructed by contracting tensor or vector operators – for example,

$$V^\mu V_\mu, \qquad T_{\mu\nu}V^\mu V^\nu, \qquad T^{\mu\nu}T_{\mu\nu}, \qquad \text{etc.} \tag{10.6.30}$$

are scalars.

Problem 10.59. Explain why, if V^μ is a vector operator, it must obey

$$\left[J^{\alpha\beta}, V^\mu\right] = -i\eta^{\mu[\alpha}V^{\beta]}. \tag{10.6.31}$$

Hint: This is the infinitesimal version of eq. (10.6.28).

From the Lie algebra in eq. (10.6.15) and (10.6.16), explain why P^μ and $J^{\alpha\beta}$ are, respectively, vector and tensor operators. Why does that, in turn, imply that $P^2 \equiv P^\nu P_\nu$ commutes with $J^{\alpha\beta}$?

$$\left[J^{\alpha\beta}, P^2\right] = 0 \tag{10.6.32}$$

□

10.7 Representations of the Lorentz and Poincaré Groups: (3+1)D

10.7.1 $\mathrm{SO}_{3,1}$ Lie Algebra

As far as we can tell, the world we live in has 3 space and 1 time dimensions. Let us now work out the Lie Algebra in eq. (10.6.8) more explicitly. Denoting the boost generator as

$$K^i \equiv J^{0i} \tag{10.7.1}$$

and the rotation generators as

$$J^i \equiv \frac{1}{2}\epsilon^{imn} J^{mn} \qquad \Leftrightarrow \qquad \epsilon^{imn} J^i \equiv J^{mn}; \tag{10.7.2}$$

with $\epsilon^{123} = \epsilon_{123} \equiv 1$. The generic Lorentz transformation continuously connected to the identity is the exponential

$$D(\vec{\xi}, \vec{\theta}) = \exp\left(-i\xi_j K^j - i\theta_j J^j\right). \tag{10.7.3}$$

At this point, these $\{D(\vec{\xi}, \vec{\theta})\}$ are not necessarily the 4×4 matrices obeying $\Lambda^{\mathrm{T}}\eta\Lambda = \eta$. Rather, they are simply linear operators, and their generators merely need to obey the same commutation relations in eq. (10.6.8).

We may compute from eq. (10.6.8) that

$$[J^m, J^n] = i\epsilon^{mnl} J^l, \tag{10.7.4}$$
$$[K^m, J^n] = i\epsilon^{mnl} K^l, \tag{10.7.5}$$
$$[K^m, K^n] = -i\epsilon^{mnl} J^l. \tag{10.7.6}$$

Problem 10.60. $\mathbf{SU}(2)_+ \otimes \mathbf{SU}(2)_-$ Let us next define

$$J^i_+ \equiv \frac{1}{2}\left(J^i + iK^i\right), \tag{10.7.7}$$
$$J^i_- \equiv \frac{1}{2}\left(J^i - iK^i\right). \tag{10.7.8}$$

Use equations (10.7.4) through (10.7.6) to show that

$$\left[J^i_+, J^j_+\right] = i\epsilon^{ijk} J^k_+, \tag{10.7.9}$$
$$\left[J^i_-, J^j_-\right] = i\epsilon^{ijk} J^k_-, \tag{10.7.10}$$
$$\left[J^i_+, J^j_-\right] = 0. \tag{10.7.11}$$

Equations (10.7.9) and (10.7.10) tell us the $J^i_\pm$ obey the same algebra as the angular momentum ones in eq. (10.7.4); and eq. (10.7.11) says the two sets $\{\vec{J}_+\}$ and $\{\vec{J}_-\}$ commute. $\square$

Equations (10.7.9), (10.7.10), and (10.7.11) indicate, the Lie Algebra of $\mathrm{SO}_{3,1}$ is a pair of SU_2 or SO_3 ones. This in turn informs us, we may take the simultaneous observables

$$(\vec{J}^2_+, J^3_+) \qquad \text{and} \qquad (\vec{J}^2_-, J^3_-), \tag{10.7.12}$$

and therefore label the eigenstates by $|j_+, j_-\rangle$, where the $j_\pm$ can be integer or half integer and

$$\vec{J}_+^2 |j_+, j_-\rangle = j_+(j_+ + 1) |j_+, j_-\rangle \tag{10.7.13}$$

$$\vec{J}_-^2 |j_+, j_-\rangle = j_-(j_- + 1) |j_+, j_-\rangle . \tag{10.7.14}$$

(The $-j_\pm \leq m_\pm \leq j_\pm$ labels have been suppressed to save clutter.) More simply, we say that these states fall in the (j_+, j_-) representation. Since $J^i = J_+^i + J_-^i$, we may identity the total angular momentum j as lying between $|j_+ - j_-|$ and $j_+ + j_-$ in integer steps.

10.7.2 $\text{SL}_{2,\mathbb{C}}$ Spinors: $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2})$ representations

In the 4D Minkowski spacetime we reside in, it turns out the fundamental objects that encodes Lorentz covariance, that physical laws should take the same form in all inertial frames, are Weyl *spinors*. This is because of the need to describe spin$-1/2$ fermions – leptons and quarks – in Nature. Moreover, ordinary spacetime vectors and tensors can also be built out of them. We will see how they arise from studying the group of Special Linear 2-dimensional operators, $\text{SL}_{2,\mathbb{C}}$.

We begin by collecting the results in Problems (3.12) and (4.28) as well as the 'Pauli matrices from their algebra' discussion in §(4.3.2). Next, we will define the group $\text{SL}_{2,\mathbb{C}}$, and proceed to describe its connection to the Lorentz group. The two inequivalent 2×2 representations of $\text{SL}_{2,\mathbb{C}}$ will be constructed and the discussion will culminate in the derivation of the Dirac equation, which is not only Lorentz but also parity covariant.

Basic Properties of $\{\sigma^\mu\}$ A basis set of orthonormal 2×2 complex matrices is provided by $\{\sigma^\mu | \mu = 0, 1, 2, 3\}$, the 2×2 identity matrix

$$\sigma^0 \equiv \mathbb{I}_{2\times2} \tag{10.7.15}$$

together with the Hermitian Pauli matrices $\{\sigma^i\}$. The $\{\sigma^i | i = 1, 2, 3\}$ may be viewed as arising from the algebra

$$\sigma^i \sigma^j = \delta^{ij} \mathbb{I}_{2\times2} + i\epsilon^{ijk} \sigma^k, \tag{10.7.16}$$

which immediately implies the respective anti-commutator and commutator results:

$$\{\sigma^i, \sigma^j\} = 2\delta^{ij} \qquad \text{and} \qquad [\sigma^i, \sigma^j] = 2i\epsilon^{ijk}\sigma^k. \tag{10.7.17}$$

Dividing the equality on the right side of eq. (10.7.17) by 4, followed by taking the complex conjugate, we infer the $\{\sigma^i/2\}$ and $\{-(\sigma^i)^*/2\}$ obey the SO_3 and SU_2 Lie Algebra

$$\left[\frac{\sigma^a}{2}, \frac{\sigma^b}{2}\right] = i\epsilon^{abc}\frac{\sigma^c}{2}, \tag{10.7.18}$$

$$\left[\frac{-(\sigma^a)^*}{2}, \frac{-(\sigma^b)^*}{2}\right] = i\epsilon^{abc}\frac{-(\sigma^c)^*}{2}. \tag{10.7.19}$$

As a result of eq. (10.7.16), the Pauli matrices have eigenvalues ±1, namely

$$\sigma^i |\pm; i\rangle = \pm |\pm; i\rangle ; \tag{10.7.20}$$

and thus -1 determinant (i.e., product of eigenvalues) and zero trace (i.e., sum of eigenvalues):

$$\det \sigma^i = -1, \qquad \mathrm{Tr}\ \sigma^i = 0. \tag{10.7.21}$$

An equivalent way of writing eq. (10.7.16) is to employ arbitrary *complex* vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$. Denoting $\vec{a} \cdot \vec{\sigma} \equiv a_i \sigma^i$,

$$(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma}) = \vec{a} \cdot \vec{b} + i(\vec{a} \times \vec{b}) \cdot \vec{\sigma}, \qquad (\vec{a} \times \vec{b})^k = \epsilon^{ijk} a_i b_j. \tag{10.7.22}$$

We may multiply by $(\vec{c} \cdot \vec{\sigma})$ from the right on both sides:

$$(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma})(\vec{c} \cdot \vec{\sigma}) = i(\vec{a} \times \vec{b}) \cdot \vec{c} + \left\{ (\vec{c} \cdot \vec{b})\vec{a} - (\vec{c} \cdot \vec{a})\vec{b} + (\vec{a} \cdot \vec{b})\vec{c} \right\} \cdot \vec{\sigma}. \tag{10.7.23}$$

Problem 10.61. Verify eq. (10.7.23). □

In the representation where σ^3 is diagonal,

$$\sigma^0 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \sigma^1 \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma^2 \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma^3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{10.7.24}$$

The inner product of $\{\sigma^\mu\}$ is provided by the matrix trace,

$$\langle \sigma^\mu | \, \sigma^\nu \rangle \equiv \frac{1}{2} \mathrm{Tr} \left[\sigma^\mu \sigma^\nu \right] = \delta^{\mu\nu}. \tag{10.7.25}$$

Since the $\{\sigma^\mu\}$ form a basis, any 2×2 complex matrix A may be obtained as a superposition $A = q_\mu \sigma^\mu$ by choosing the appropriate complex parameters $\{q_\mu\}$. In addition, we will utilize

$$\bar{\sigma}^\mu \equiv \left(\mathbb{I}_{2\times 2}, -\sigma^i \right) = \sigma_\mu. \tag{10.7.26}$$

[109]These $\{\bar{\sigma}^\mu\}$ form a orthogonal basis as well; i.e., $\langle \bar{\sigma}^\mu | \, \bar{\sigma}^\nu \rangle = (1/2)\mathrm{Tr} \left[\bar{\sigma}^\mu \bar{\sigma}^\nu \right] = \delta^{\mu\nu}$. We also need the 2D Levi-Civita symbol ϵ. Since ϵ is real and antisymmetric,

$$\epsilon^\dagger = \epsilon^{\mathrm{T}} = -\epsilon, \tag{10.7.27}$$

a direct calculation would reveal it is also unitary:

$$\epsilon \cdot \epsilon^\dagger = -\epsilon^2 = \mathbb{I}. \tag{10.7.28}$$

The non-zero components are

$$\epsilon^{12} = \epsilon_{12} = 1, \qquad \text{and} \qquad \epsilon^{21} = \epsilon_{21} = -1. \tag{10.7.29}$$

According to eq. (10.7.16), because $\sigma^i \sigma^i = \mathbb{I}$ (for fixed i) that implies σ^i is its own inverse. We may then invoke eq. (3.2.8) to state

$$(\sigma^i)^{-1} = \sigma^i = -\frac{\epsilon (\sigma^i)^{\mathrm{T}} \epsilon}{\det \sigma^i} = \frac{\epsilon (\sigma^i)^{\mathrm{T}} \epsilon^\dagger}{\det \sigma^i} = \frac{\epsilon^\dagger (\sigma^i)^{\mathrm{T}} \epsilon}{\det \sigma^i}. \tag{10.7.30}$$

[109]Caution: The over-bar on $\bar{\sigma}$ is not complex conjugation.

Since ϵ is real, $\det \sigma^i = -1$ (cf. eq. (10.7.21)), and σ^i is Hermitian, we may take the complex conjugate on both sides and deduce

$$(\sigma^i)^* = \epsilon \cdot \sigma^i \cdot \epsilon = \epsilon^\dagger \left(-\sigma^i\right) \epsilon = \epsilon \left(-\sigma^i\right) \epsilon^\dagger. \tag{10.7.31}$$

Since $\epsilon^2 = -\mathbb{I}$, we may multiply both sides by ϵ on the left and right, and deduce the two basis employed in equations (10.7.18) and (10.7.19) are related via a change-of-basis induced by ϵ:

$$\epsilon \cdot (\sigma^i)^* \cdot \epsilon = \epsilon^\dagger \cdot (-\sigma^i)^* \cdot \epsilon = \epsilon \cdot (-\sigma^i)^* \cdot \epsilon^\dagger = \sigma^i. \tag{10.7.32}$$

Problem 10.62. Verify that, for *complex* p_μ,

$$\bar{\sigma}^\mu \sigma^\nu p_\mu p_\nu = \sigma^\mu \bar{\sigma}^\nu p_\mu p_\nu = p^2 \cdot \mathbb{I}_{2\times 2}; \tag{10.7.33}$$

where $p^2 \equiv p_\mu p^\mu$. This result will be useful in analyzing the dispersion relations arising from the Majorana, Weyl, and Dirac equations. □

Problem 10.63. Using the notation in eq. (10.7.26), explain why

$$\epsilon \cdot (\bar{\sigma}^\mu)^* \cdot \epsilon^\dagger = \epsilon^\dagger \cdot (\bar{\sigma}^\mu)^* \cdot \epsilon = \sigma^\mu, \tag{10.7.34}$$

$$\epsilon \cdot (\sigma^\mu)^* \cdot \epsilon^\dagger = \epsilon^\dagger \cdot (\sigma^\mu)^* \cdot \epsilon = \bar{\sigma}^\mu; \tag{10.7.35}$$

and therefore

$$\epsilon \cdot \bar{\sigma}^\mu \cdot \epsilon^\dagger = \epsilon^\dagger \cdot \bar{\sigma}^\mu \cdot \epsilon = (\sigma^\mu)^*, \tag{10.7.36}$$

$$\epsilon \cdot \sigma^\mu \cdot \epsilon^\dagger = \epsilon^\dagger \cdot \sigma^\mu \cdot \epsilon = (\bar{\sigma}^\mu)^*. \tag{10.7.37}$$

Hint: Remember the properties of ϵ and σ^0. □

Because $(\sigma^\mu)^2 = \mathbb{I}$ and $\sigma^\mu / \det \sigma^\mu = \bar{\sigma}^\mu = (\mathbb{I}, -\sigma^i) = \sigma_\mu$, eq. (3.2.8) informs us

$$\sigma^\mu = -\epsilon \cdot (\bar{\sigma}^\mu)^{\mathrm{T}} \cdot \epsilon = -\epsilon \cdot (\sigma_\mu)^{\mathrm{T}} \cdot \epsilon \tag{10.7.38}$$

$$= \epsilon^\dagger \cdot (\bar{\sigma}^\mu)^{\mathrm{T}} \cdot \epsilon = \epsilon^\dagger \cdot (\sigma_\mu)^{\mathrm{T}} \cdot \epsilon. \tag{10.7.39}$$

Remember, $\bar{\sigma}^\mu = \sigma_\mu$ because lowering the spatial components costs a minus sign. Likewise, because $(\bar{\sigma}^\mu)^2 = \mathbb{I}$ and $\bar{\sigma}^\mu / \det \bar{\sigma}^\mu = \sigma^\mu = \bar{\sigma}_\mu$, eq. (3.2.8) informs us

$$\bar{\sigma}^\mu = -\epsilon \cdot (\sigma^\mu)^{\mathrm{T}} \cdot \epsilon = -\epsilon \cdot (\bar{\sigma}_\mu)^{\mathrm{T}} \cdot \epsilon \tag{10.7.40}$$

$$= \epsilon^\dagger \cdot (\sigma^\mu)^{\mathrm{T}} \cdot \epsilon = \epsilon^\dagger \cdot (\bar{\sigma}_\mu)^{\mathrm{T}} \cdot \epsilon. \tag{10.7.41}$$

Exponential of Pauli Matrices For any complex $\{\psi_i\}$, we have from eq. (3.2.26),

$$\exp\left(-\frac{i}{2}\psi_i \sigma^i\right) = \cos\left(\frac{|\vec{\psi}|}{2}\right) - i\frac{\vec{\psi}\cdot\vec{\sigma}}{|\vec{\psi}|}\sin\left(\frac{|\vec{\psi}|}{2}\right), \tag{10.7.42}$$

$$\vec{\psi}\cdot\vec{\sigma} \equiv \psi_j \sigma^j, \qquad |\vec{\psi}| \equiv \sqrt{\psi_i \psi_i}. \tag{10.7.43}$$

One may readily check that its inverse is

$$\left(\exp\left(-\frac{i}{2}\psi_i\sigma^i\right)\right)^{-1} = \exp\left(+\frac{i}{2}\psi_i\sigma^i\right) = \cos\left(\frac{|\vec{\psi}|}{2}\right) + i\frac{\vec{\psi}\cdot\vec{\sigma}}{|\vec{\psi}|}\sin\left(\frac{|\vec{\psi}|}{2}\right). \tag{10.7.44}$$

When $\psi_i \equiv \theta_i$ is purely real, eq. (10.7.42) retains the same form. Whereas, if $\psi_i \equiv i\xi_i$ is purely imaginary, $|\vec{\psi}| = i|\vec{\xi}|$ and eq. (10.7.42) becomes the following Hermitian object:

$$\exp\left(\frac{1}{2}\xi_i\sigma^i\right) = \cosh\left(\frac{|\vec{\xi}|}{2}\right) + \frac{\vec{\xi}\cdot\vec{\sigma}}{|\vec{\xi}|}\sinh\left(\frac{|\vec{\xi}|}{2}\right). \tag{10.7.45}$$

[110]Observe that the relation in eq. (3.2.26) is basis independent; namely, if we found a different representation of the Pauli matrices

$$\sigma'^i = U\sigma^i U^{-1} \qquad \Leftrightarrow \qquad U^{-1}\sigma'^i U = \sigma^i \tag{10.7.46}$$

then the algebra in eq. (10.7.16) and the exponential result in eq. (10.7.42) would respectively become

$$U^{-1}\sigma'^i U U^{-1}\sigma'^j U = U^{-1}\left(\delta^{ij} + i\epsilon^{ijk}\sigma'^k\right)U, \tag{10.7.47}$$

$$\sigma'^i\sigma'^j = \delta^{ij} + i\epsilon^{ijk}\sigma'^k \tag{10.7.48}$$

and

$$\exp\left(-\frac{i}{2}\psi_i U^{-1}\sigma'^i U\right) = U^{-1}\exp\left(-\frac{i}{2}\psi_i\sigma'^i\right)U = U^{-1}\left(\cos\left(\frac{|\vec{\psi}|}{2}\right) - i\frac{\psi_j\sigma'^j}{|\vec{\psi}|}\sin\left(\frac{|\vec{\psi}|}{2}\right)\right)U,$$

$$\exp\left(-\frac{i}{2}\psi_i\sigma'^i\right) = \cos\left(\frac{|\vec{\psi}|}{2}\right) - i\frac{\psi_j\sigma'^j}{|\vec{\psi}|}\sin\left(\frac{|\vec{\psi}|}{2}\right). \tag{10.7.49}$$

Helicity Eigenstates The Pauli matrices divided by 2, $\{\sigma^i/2|i=1,2,3\}$, are associated with spin$-1/2$ systems. The *helicity operator*, in turn, is defined as the Hermitian object

$$\widehat{p}\cdot\vec{\sigma} \equiv \frac{p_i}{|\vec{p}|}\sigma^i \tag{10.7.50}$$

for real p_i and $|\vec{p}| \equiv \sqrt{\delta^{ij}p_i p_j}$. It may be diagonalized as

$$\frac{p_i}{|\vec{p}|}\left(\sigma^i\right)_{\mathrm{A}\dot{\mathrm{B}}} = \xi^+_{\mathrm{A}}\overline{\xi^+}_{\dot{\mathrm{B}}} - \xi^-_{\mathrm{A}}\overline{\xi^-}_{\dot{\mathrm{B}}}; \tag{10.7.51}$$

$$\frac{p_i}{|\vec{p}|}\sigma^i\xi^\pm = \pm\xi^\pm. \tag{10.7.52}$$

[110]We will take the $\sqrt{\cdot}$ in the definition of $|\vec{\psi}|$ to be the positive square root. However, note that, since sine is an odd power series, $\sin[|\vec{\psi}|/2]/|\vec{\psi}|$ is an even power series and there is actually no ambiguity.

[111]In the representation of the Pauli matrices in eq. (10.7.24), the unit norm *helicity eigenstates* are, up to overall phases,

$$\xi^{+}{}_{\mathrm{A}}(\vec{p}) = \left(e^{-i\phi_p}\cos\left[\frac{\theta_p}{2}\right], \sin\left[\frac{\theta_p}{2}\right]\right)^{\mathrm{T}} \tag{10.7.53}$$

$$= \frac{1}{\sqrt{2|\vec{p}|}}\left(\frac{\sqrt{|\vec{p}|+p_3}\sqrt{|\vec{p}|^2-p_3^2}}{p_1+ip_2}, \sqrt{|\vec{p}|-p_3}\right)^{\mathrm{T}} \tag{10.7.54}$$

and

$$\xi^{-}{}_{\mathrm{A}}(\vec{p}) = \left(-e^{-i\phi_p}\sin\left[\frac{\theta_p}{2}\right], \cos\left[\frac{\theta_p}{2}\right]\right)^{\mathrm{T}} \tag{10.7.55}$$

$$= \frac{1}{\sqrt{2|\vec{p}|}}\left(-\frac{\sqrt{|\vec{p}|-p_3}\sqrt{|\vec{p}|^2-p_3^2}}{p_1+ip_2}, \sqrt{|\vec{p}|+p_3}\right)^{\mathrm{T}}. \tag{10.7.56}$$

Note that we have switched to spherical coordinates in momentum space, namely

$$p_i \equiv p\left(\sin\theta_p\cos\phi_p, \sin\theta_p\sin\phi_p, \cos\theta_p\right). \tag{10.7.57}$$

Also notice, under parity $\vec{p} \to -\vec{p}$; or in spherical coordinates,

$$\phi_p \to \phi_p + \pi \qquad \text{and} \qquad \theta_p \to \pi - \theta_p, \tag{10.7.58}$$

the helicity eigenstates in equations (10.7.53) and (10.7.55) transform into each other:

$$\xi^+ \to \xi^- \qquad \text{and} \qquad \xi^- \to \xi^+. \tag{10.7.59}$$

Equivalently,

$$\xi^{\pm}(-\vec{p}) = \xi^{\mp}(\vec{p}). \tag{10.7.60}$$

These orthonormal eigenstates $\xi^\pm$ of the Hermitian $p_i\sigma^i$, in equations (10.7.53) and (10.7.55), span the 2D complex vector space, so their completeness relation is

$$\mathbb{I}_{\mathrm{A}\dot{\mathrm{B}}} = \xi^+_{\mathrm{A}}\overline{\xi^+}_{\dot{\mathrm{B}}} + \xi^-_{\mathrm{A}}\overline{\xi^-}_{\dot{\mathrm{B}}}; \tag{10.7.61}$$

Therefore, $p_\mu\sigma^\mu = p_0\mathbb{I} + p_i\sigma^i$ itself must be p_0 times of eq. (10.7.61) plus $|\vec{p}|$ times of eq. (10.7.51).

$$p_\mu\left(\sigma^\mu\right)_{\mathrm{A}\dot{\mathrm{B}}} \equiv p_{\mathrm{A}\dot{\mathrm{B}}} = \lambda_+\xi^+_{\mathrm{A}}\overline{\xi^+}_{\dot{\mathrm{B}}} + \lambda_-\xi^-_{\mathrm{A}}\overline{\xi^-}_{\dot{\mathrm{B}}}, \tag{10.7.62}$$

$$\lambda_\pm \equiv p_0 \pm |\vec{p}| \tag{10.7.63}$$

The notation $p_{\mathrm{A}\dot{\mathrm{B}}}$ swaps one Lorentz index μ (on p_μ) for two 'spinor' indices $\mathrm{A}\dot{\mathrm{B}}$. Let us note that both μ and $\mathrm{A}\dot{\mathrm{B}}$ has the necessary 4 components to describe a vector – this suggests we are not losing any information when switching between these two descriptions. As we will show shortly, indeed, this is merely a novel change-of-basis.

[111]This dotted/un-dotted notation, due to van der Waerden, will be explained shortly.

Problem 10.64. Eigensystem of $\exp(\vec{\zeta}\cdot\vec{\sigma}/2)$ Argue that the helicity eigenstates in equations (10.7.53) and (10.7.55) are eigenstates of the Hermitian operator $\exp(\sigma^i\zeta_i/2)$ in eq. (10.7.45), whenever $\zeta_i = |\vec{\zeta}|p_i/|\vec{p}| \equiv \zeta\widehat{p}_i$. Specifically, show that

$$\exp\left(\frac{\zeta}{2}\widehat{p}_i\sigma^i\right)\xi^{\pm}(\vec{p}) = e^{\pm\zeta/2}\xi^{\pm}(\vec{p}), \qquad \zeta \equiv |\vec{\zeta}|. \tag{10.7.64}$$

Below, we shall discover that $\exp(\vec{\zeta}\cdot\vec{\sigma}/2)$ corresponds to Lorentz boosts. □

Problem 10.65. Parity Transformation Show that the helicity eigenstates may be converted into each other – up to overall phases – via the relations

$$\epsilon^\dagger(\xi^{\pm})^* = e^{i\delta}\xi^{\mp} \tag{10.7.65}$$

where δ is an arbitrary real number. Note that, there is no matrix U such that $U\xi^{\pm} = e^{i\delta}\xi^{\mp}$; namely, no parity transformation linear operator exists – the complex conjugation is a critical ingredient in this parity conversion process. □

Completeness of $\{\sigma^\mu\}$: Spacetime vs. Spinor Indices Since the $\{\sigma^\mu\}$ form an orthonormal basis, they must admit some form of the completeness relation in eq. (4.3.23). Now, according to eq. (10.7.25), $c_\mu\sigma^\mu \Leftrightarrow c_\mu = (1/2)\mathrm{Tr}\left[(c_\nu\sigma^\nu)\sigma^\mu\right]$ for any complex coefficients $\{c_\nu\}$. (We will not distinguish between dotted and un-dotted indices for now.) Consider the Ath row Bth column of the matrix $c_\nu\sigma^\nu$:

$$c_\mu(\sigma^\mu)_{\mathrm{AB}} = \sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\mathrm{Tr}\left[(\sigma^\mu)^\dagger c_\nu\sigma^\nu\right] \tag{10.7.66}$$

$$= \sum_{1\leq\mathrm{C,D}\leq2}\left(\sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\overline{(\sigma^\mu)^{\mathrm{T}}}_{\mathrm{DC}}\right)c_\nu(\sigma^\nu)_{\mathrm{CD}} \tag{10.7.67}$$

$$= \sum_{1\leq\mathrm{C,D}\leq2}\left(\sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\overline{(\sigma^\mu)}_{\mathrm{CD}}\right)c_\nu(\sigma^\nu)_{\mathrm{CD}}. \tag{10.7.68}$$

We may view the terms in the parenthesis on the last line as an operator that acts on the operator $c_\nu\sigma^\nu$. But since c_ν was arbitrary, it must act on each and every σ^ν to return σ^ν, since the left hand side is $c_\nu\sigma^\nu$. And because the $\{\sigma^\nu\}$ are the basis kets of the space of operators acting on a 2D complex vector space, the term in parenthesis must be the identity.

$$\sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\overline{(\sigma^\mu)}_{\mathrm{CD}} = \sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}(\sigma^\mu)^{\mathrm{T}}_{\mathrm{CD}} = \delta^{\mathrm{C}}_{\mathrm{A}}\delta^{\mathrm{D}}_{\mathrm{B}} \tag{10.7.69}$$

In the second equality we have used the Hermitian nature of σ^μ to deduce $(\sigma^\mu)^\dagger_{\mathrm{AB}} = \overline{(\sigma^\mu)^{\mathrm{T}}}_{\mathrm{AB}} = (\sigma^\mu)_{\mathrm{AB}} \Leftrightarrow (\sigma^\mu)^{\mathrm{T}}_{\mathrm{AB}} = (\sigma^\mu)^*_{\mathrm{AB}}$. If we further employ $(\sigma^\mu)^* = \epsilon\cdot\bar{\sigma}_\mu\cdot\epsilon^\dagger = \epsilon\cdot\bar{\sigma}_\mu\cdot\epsilon^{\mathrm{T}}$ in eq. (10.7.36) within the leftmost expression,

$$\sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\overline{(\sigma^\mu)}_{\mathrm{CD}} = \sum_{0\leq\mu\leq3}\frac{1}{2}(\sigma^\mu)_{\mathrm{AB}}\epsilon^{\mathrm{CM}}\epsilon^{\mathrm{DN}}(\bar{\sigma}_\mu)_{\mathrm{MN}}. \tag{10.7.70}$$

If we now restore the dotted notation on the right index, so that

$$\epsilon^{\mathrm{CM}}\epsilon^{\dot{\mathrm{D}}\dot{\mathrm{N}}}(\bar{\sigma}_\mu)_{\mathrm{M}\dot{\mathrm{N}}} = (\bar{\sigma}_\mu)_{\mathrm{M}\dot{\mathrm{N}}}\epsilon^{\mathrm{MC}}\epsilon^{\dot{\mathrm{N}}\dot{\mathrm{D}}} \equiv (\bar{\sigma}_\mu)^{\mathrm{C}\dot{\mathrm{D}}}, \tag{10.7.71}$$

then eq. (10.7.69), with Einstein summation in force, becomes

$$\frac{1}{2}\sigma^\mu{}_{\mathrm{A}\dot{\mathrm{B}}}\bar{\sigma}_\mu{}^{\mathrm{C}\dot{\mathrm{D}}} = \delta^{\mathrm{C}}_{\mathrm{A}}\delta^{\dot{\mathrm{D}}}_{\dot{\mathrm{B}}}. \tag{10.7.72}$$

Next, consider

$$\begin{aligned}
(\sigma^\mu)_{\mathrm{M}\dot{\mathrm{N}}}(\sigma_\nu)^{\mathrm{M}\dot{\mathrm{N}}} &= (\sigma^\mu)_{\mathrm{M}\dot{\mathrm{N}}}\epsilon^{\dot{\mathrm{M}}\dot{\mathrm{A}}}\epsilon^{\dot{\mathrm{N}}\dot{\mathrm{B}}}(\sigma_\nu)_{\mathrm{A}\dot{\mathrm{B}}} = (\sigma^\mu)^{\mathrm{T}}_{\dot{\mathrm{N}}\mathrm{M}}\epsilon^{\dot{\mathrm{M}}\dot{\mathrm{A}}}(\sigma_\nu)_{\mathrm{A}\dot{\mathrm{B}}}(\epsilon^{\mathrm{T}})^{\dot{\mathrm{B}}\dot{\mathrm{N}}} &(10.7.73)\\
&= \mathrm{Tr}\left[(\sigma^\mu)^{\mathrm{T}}\cdot\epsilon\cdot\sigma_\nu\cdot\epsilon^\dagger\right] = \mathrm{Tr}\left[\epsilon^\dagger\cdot(\sigma^\mu)^{\mathrm{T}}\cdot\epsilon\cdot\sigma_\nu\right] &(10.7.74)\\
&= \mathrm{Tr}\left[\bar{\sigma}^\mu\sigma_\nu\right] = \mathrm{Tr}\left[\sigma_\mu\sigma_\nu\right], &(10.7.75)
\end{aligned}$$

where eq. (10.7.41) was used in the last line. Invoking the orthonormality of the $\{\sigma^\mu\}$ in eq. (10.7.25),

$$\frac{1}{2}(\sigma^\mu)_{\mathrm{M}\dot{\mathrm{N}}}(\sigma_\nu)^{\mathrm{M}\dot{\mathrm{N}}} \equiv \frac{1}{2}(\sigma^\mu)_{\mathrm{M}\dot{\mathrm{N}}}\epsilon^{\mathrm{MA}}\epsilon^{\dot{\mathrm{N}}\dot{\mathrm{B}}}(\sigma_\nu)_{\mathrm{A}\dot{\mathrm{B}}} = \delta^\mu{}_\nu. \tag{10.7.76}$$

Equation (10.7.76) tell us we may view the spacetime Lorentz index μ and the pair of spinor indices $\mathrm{A}\dot{\mathrm{B}}$ as different basis for describing tensors. For example, we may now switch between the momentum p_μ and $p_{\mathrm{A}\dot{\mathrm{B}}}$ via:

$$p_\mu\sigma^\mu{}_{\mathrm{A}\dot{\mathrm{B}}} = p_{\mathrm{A}\dot{\mathrm{B}}} \qquad \Leftrightarrow \qquad p_\mu = \frac{1}{2}\sigma_\mu{}^{\mathrm{A}\dot{\mathrm{B}}}p_{\mathrm{A}\dot{\mathrm{B}}}, \tag{10.7.77}$$

where the relation on the right is a direct consequence of eq. (10.7.76),

$$p_\mu = \frac{1}{2}\sigma_\mu{}^{\mathrm{A}\dot{\mathrm{B}}}\sigma^\nu{}_{\mathrm{A}\dot{\mathrm{B}}}p_\nu = \delta^\nu_\mu p_\nu = p_\mu. \tag{10.7.78}$$

That a spacetime vector p_μ may be described as a rank-2 spinor object is a good reason to turn now to the study of the $\mathrm{SL}_{2,\mathbb{C}}$ group. It shall reveal how spinors are more fundamental Lorentz covariant objects than spacetime vectors/tensors; since the latter are merely a special case of the former.

$\mathrm{SL}_{2,\mathbb{C}}$ and Levi-Civita 'Metric' The $\mathrm{SL}_{2,\mathbb{C}}$ group refers to the set of complex 2×2 matrices $\{L\}$ which have unit determinant – the 'Special' in the SL. Employing eq. (3.2.9) or (3.2.10), we may express this definition as

$$\begin{aligned}
\mathrm{SL}_{2,\mathbb{C}}:\ L_{\mathrm{I}}{}^{\mathrm{A}}L_{\mathrm{J}}{}^{\mathrm{B}}\epsilon_{\mathrm{AB}} &= \epsilon_{\mathrm{IJ}}, \qquad L \in 2\times 2 \text{ complex matrices}; &(10.7.79)\\
\epsilon_{12} \equiv \epsilon^{12} &\equiv 1. &(10.7.80)
\end{aligned}$$

Since $\det L^{\mathrm{T}} = \det L$, this definition is equivalent to

$$\epsilon^{\mathrm{AB}}L_{\mathrm{A}}{}^{\mathrm{I}}L_{\mathrm{B}}{}^{\mathrm{J}} = \epsilon^{\mathrm{IJ}}, \qquad \epsilon^{12} \equiv 1. \tag{10.7.81}$$

Since the Levi-Civita ϵ is real, we may take the complex conjugation of these equations and deduce

$$(L^*)_{\mathrm{I}}{}^{\mathrm{A}}(L^*)_{\mathrm{J}}{}^{\mathrm{B}}\epsilon_{\mathrm{AB}} = \epsilon_{\mathrm{IJ}} \qquad \text{and} \qquad \epsilon^{\mathrm{AB}}(L^*)_{\mathrm{A}}{}^{\mathrm{I}}(L^*)_{\mathrm{B}}{}^{\mathrm{J}} = \epsilon^{\mathrm{IJ}}. \tag{10.7.82}$$

The 2D Levi-Civita symbol is invariant under $\mathrm{SL}_{2,\mathbb{C}}$ transformations. Moreover, comparison with the definition of Lorentz invariance, $\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}$, suggests we may view ϵ_{AB} as the 'metric' and $\epsilon^{\mathrm{BA}} = (\epsilon^\dagger)_{\mathrm{AB}}$ as the inverse 'metric'. Above, we have already moved the indices of σ^μ and $\bar{\sigma}^\mu$ with the 2D Levi-Civita; below, we will justify why the indices of the Weyl spinors may be moved with it too.

Problem 10.66. $\mathbf{SL}_{2,\mathbb{C}}$ Group Structure Prove that $\mathrm{SL}_{2,\mathbb{C}}$ in fact forms a group. See Appendix (B) for the axioms of a Group. □

Problem 10.67. Use eq. (3.2.8) to argue that, for L belonging to the $\mathrm{SL}_{2,\mathbb{C}}$ group, its inverse obeys

$$L^{-1} = \epsilon^\dagger \cdot L^{\mathrm{T}} \cdot \epsilon = \epsilon \cdot L^{\mathrm{T}} \cdot \epsilon^\dagger. \tag{10.7.83}$$

and therefore

$$L = \epsilon^\dagger \cdot (L^{-1})^{\mathrm{T}} \cdot \epsilon = \epsilon^\dagger \cdot (L^{\mathrm{T}})^{-1} \cdot \epsilon. \tag{10.7.84}$$

Why do these further imply the following?

$$L^* = \epsilon^\dagger \cdot (L^{-1})^\dagger \cdot \epsilon = \epsilon^\dagger \cdot (L^\dagger)^{-1} \cdot \epsilon \tag{10.7.85}$$

$$(L^{-1})^* = \epsilon^\dagger \cdot L^\dagger \cdot \epsilon = \epsilon^\dagger \cdot L^\dagger \cdot \epsilon. \tag{10.7.86}$$

Since $\epsilon^\dagger \epsilon = \mathbb{I}$ – i.e., Levi-Civita ϵ is unitary – these results teach us, L is equivalent to $(L^{\mathrm{T}})^{-1} = (L^{-1})^{\mathrm{T}}$; and hence L^{-1} is equivalent to L^{T}; whereas $(L^{-1})^\dagger = (L^\dagger)^{-1}$ is equivalent to L^*. Furthermore, we shall see below that L and L^* are *inequivalent*. □

Construction of L We have discussed in §(5.2), any operator that is continuously connected to the identity can be written in the form $\exp X$. Since L has unit determinant (cf. (10.7.102)), let us focus on the case where it is continuously connected to the identity whenever it does depend on a set of complex parameters $\{\psi_\mu\}$, say:

$$L = e^{X(\psi)}. \tag{10.7.87}$$

Now, if we use eq. (5.6.152), $\det e^X = e^{\mathrm{Tr}[X]}$, we find that

$$\det L = e^{\mathrm{Tr}\, X(\psi)} = 1. \tag{10.7.88}$$

This implies

$$\mathrm{Tr} X(\psi) = 2\pi i n, \qquad n = 0, \pm 1, \pm 2, \dots . \tag{10.7.89}$$

Recalling that the $\{\sigma^\mu\}$ form a complete set, we may express

$$X(\psi) = -(i/2)\psi_\mu \sigma^\mu, \tag{10.7.90}$$

with four arbitrary complex parameters $\{\psi_\mu\}$; and using the trace properties in eq. (10.7.21), we see that Tr $X(q) = -iq_0 = 2\pi i n$. Since this $-(i/2)q_0\sigma^0 = i\pi n\mathbb{I}_{2\times 2}$, which commutes with all the other Pauli matrices, we have at this point

$$L = (-)^n e^{-(i/2)\psi_j\sigma^j}. \tag{10.7.91}$$

For even n, we see that $\psi_j = 0$ would return the identity whereas when n is odd, we may recall eq. (10.7.42) to first write

$$-\mathbb{I}_{2\times 2} = \exp(-i\pi\widehat{p}_j\sigma^j) \tag{10.7.92}$$

and therefore L is just the product of two exponentials

$$(-)^n e^{-(i/2)\psi_j\sigma^j} = \exp(-i\pi\widehat{p}_j\sigma^j)\exp(-(i/2)\psi_j\sigma^j). \tag{10.7.93}$$

To sum:

> The most general unit determinant 2×2 complex matrix that is continuously connected to the identity is given by eq. (10.7.42) for arbitrary complex $\{\psi_j\}$.

Problem 10.68. $\mathrm{SL}_{2,\mathbb{C}}$ generators Consider the infinitesimal $\mathrm{SL}_{2,\mathbb{C}}$ transformation

$$L_{\mathrm{A}}{}^{\mathrm{B}} = \delta_{\mathrm{A}}{}^{\mathrm{B}} + \omega_{\mathrm{A}}{}^{\mathrm{B}}. \tag{10.7.94}$$

Show that, by viewing ϵ^{AB} and $\omega_{\mathrm{A}}{}^{\mathrm{B}}$ as matrices,

$$\epsilon \cdot \omega = (\epsilon \cdot \omega)^{\mathrm{T}}. \tag{10.7.95}$$

From this, deduce

$$\omega_{\mathrm{A}}{}^{\mathrm{B}} = \begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix}, \tag{10.7.96}$$

where α, β, and γ are arbitrary complex parameters. Notice, not only is $\omega_{\mathrm{A}}{}^{\mathrm{B}}$ traceless, it contains 6 real parameters – in accordance to the 3 directions for boosts plus the 3 directions for rotations. $\square$

Lorentz Transformations and $\mathrm{SL}_{2,\mathbb{C}}$ Let us now turn to the key goal of this section: the connection between the $\mathrm{SL}_{2,\mathbb{C}}$ group and the Lorentz group. To this end, we first consider the following. If $p_\mu \equiv (p_0, p_1, p_2, p_3)$ is a real 4-component momentum vector, one would find that the determinant of $p_\mu\sigma^\mu$ yields the Lorentz invariant p^2:

$$\det p_\mu\sigma^\mu = \eta^{\mu\nu}p_\mu p_\nu \equiv p^2. \tag{10.7.97}$$

[112]If we exploited the representation in eq. (10.7.24),

$$p_\mu\sigma^\mu = \begin{bmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{bmatrix}, \tag{10.7.98}$$

[112]Although we are concerned with the full Lorentz group here, note that $\det p_i\sigma^i = -\vec{p}^2$; so one may also use Pauli matrices to analyze representations of the rotation group alone, i.e., all transformations that leave $\vec{p}^2$ invariant.

from which eq. (10.7.97) may then be readily verified. Furthermore, if we now multiply a 2×2 complex matrix L to the left and $L^\dagger$ to the right of the matrix $p_\mu \sigma^\mu$, namely

$$p_\mu \sigma^\mu \to L \cdot p_\mu \sigma^\mu \cdot L^\dagger \tag{10.7.99}$$

– this transformation preserves the Hermitian nature of $p_\mu \sigma^\mu$ for real p_μ – then its determinant will transform as

$$p^2 = \det(p_\mu \sigma^\mu) \to \det(L \cdot p_\mu \sigma^\mu \cdot L^\dagger) = |\det L|^2 \, p^2. \tag{10.7.100}$$

To merely preserve the Lorentz 'dot product', namely $p^2 \to p^2$, we may choose the determinant to be a phase; i.e., $\det L = e^{i\delta}$ for $\delta \in [0, 2\pi)$. However, eq. (10.7.79) would then read

$$L_{\mathrm{I}}{}^{\mathrm{A}} L_{\mathrm{J}}{}^{\mathrm{B}} \epsilon_{\mathrm{AB}} = \epsilon_{\mathrm{IJ}} \cdot e^{i\delta}; \tag{10.7.101}$$

spoiling the 'metric' analogy with $\Lambda^{\mathrm{T}} \eta \Lambda = \eta$. Motivated by the above $\mathrm{SL}_{2,\mathbb{C}}$ considerations, therefore, let us choose

$$\det L = 1, \tag{10.7.102}$$

so that

$$p^2 \to p^2 \qquad \text{and} \qquad L \cdot \epsilon \cdot L^{\mathrm{T}} = \epsilon. \tag{10.7.103}$$

On the other hand, since $\{\sigma^\mu\}$ forms a complete set of 2×2 matrices, $L \cdot \sigma^\mu \cdot L^\dagger$ must return the superposition

$$L \cdot \sigma^\mu \cdot L^\dagger = \sigma^\nu \Lambda_\nu{}^\mu, \tag{10.7.104}$$

where $\Lambda_\nu{}^\mu$ is, for the moment, merely a linear transformation. However,

$$\det \sigma^\mu p_\mu = \det(L \cdot (\sigma^\mu p_\mu) \cdot L^\dagger) = \det(\sigma^\mu p'_\mu), \tag{10.7.105}$$

with

$$p'_\mu \equiv \Lambda_\mu{}^\nu p_\nu, \tag{10.7.106}$$

must be simultaneously equal to both p^2 and p'^2. Thus, $\Lambda_\nu{}^\mu$ is not merely a linear transformation, it must be a Lorentz transformation. In other words: the group of $\mathrm{SL}_{2,\mathbb{C}}$ matrices $\{L\}$ implement Lorentz transformations $\{\Lambda_\nu{}^\mu\}$ via eq. (10.7.104).

We may now systematically construct these L's and their generators. Specifically, for a given generator of the Lorentz transformation matrix – recall eq. (10.3.119) –

$$(\widehat{J}^{\alpha\beta})_\mu{}^\nu = i\delta_\mu^{[\alpha} \eta^{\beta]\nu}, \tag{10.7.107}$$

we wish to solve for the $\{\psi_j\}$ in $L = \exp(-i\omega\vec{\psi} \cdot \vec{\sigma}/2)$ such that, for arbitrary $\omega \in \mathbb{R}$, eq. (10.7.104) is satisfied:

$$\exp\left(-i\omega\psi_j \frac{\sigma^j}{2}\right) \sigma^\mu \exp\left(i\omega\psi_j^* \frac{\sigma^j}{2}\right) = \sigma^\nu \left(\exp\left(-i\omega\widehat{J}^{\alpha\beta}\right)\right)_\nu{}^\mu. \tag{10.7.108}$$

(Since ψ_j is arbitrary it has been re-scaled by ω for technical convenience.) We may Taylor expand both sides to obtain

$$\sigma^\mu - i\omega\left(\psi_j \frac{\sigma^j}{2}\sigma^\mu - \sigma^\mu \psi_j^* \frac{\sigma^j}{2}\right) + \dots = \sigma^\mu - i\omega\sigma^\nu \left(\widehat{J}^{\alpha\beta}\right)_\nu^{\ \ \mu} + \dots, \tag{10.7.109}$$

$$\psi_j \frac{\sigma^j}{2}\sigma^\mu - \sigma^\mu \psi_j^* \frac{\sigma^j}{2} = -i\eta^{\mu[\alpha}\sigma^{\beta]}; \tag{10.7.110}$$

where, because the ψ's are complex, the left hand side at first order in ω is *not* a commutator. Denoting the Re and Im portions of ψ as ψ^{R} and ψ^{I} respectively, one finds – for $\mu = 0, k$:

$$i\psi_\ell^{\mathrm{I}}\sigma^\ell = -i\eta^{0[\alpha}\sigma^{\beta]}, \tag{10.7.111}$$

$$i\psi_k^{\mathrm{I}}\sigma^0 + i\psi_\ell^{\mathrm{R}}\epsilon^{\ell kj}\sigma^j = -i\eta^{k[\alpha}\sigma^{\beta]}; \tag{10.7.112}$$

where eq. (3.2.23) was employed to reach the second line. For boosts, let us further set $\alpha\beta = 0a$:

$$i\psi_\ell^{\mathrm{I}}\sigma^\ell = -i\eta^{00}\sigma^a, \tag{10.7.113}$$

$$i\psi_k^{\mathrm{I}}\sigma^0 + i\psi_\ell^{\mathrm{R}}\epsilon^{\ell kj}\sigma^j = i\eta^{ka}\sigma^0. \tag{10.7.114}$$

Since the $\{\sigma^\mu\}$ form an orthogonal basis, the unique solution is $\psi_\ell^{\mathrm{I}} = -\delta_\ell^a = \eta^{a\ell}$ and $\psi_\ell^{\mathrm{R}} = 0$:

$$K_{\mathrm{L}}^a = -i\frac{\sigma^a}{2}. \tag{10.7.115}$$

For rotations, we now put $\alpha\beta = ab$:

$$i\psi_\ell^{\mathrm{I}}\sigma^\ell = 0, \tag{10.7.116}$$

$$i\psi_k^{\mathrm{I}}\sigma^0 + i\psi_\ell^{\mathrm{R}}\epsilon^{\ell kj}\sigma^j = i\delta^{k[a}\delta^{b]j}\sigma^j. \tag{10.7.117}$$

Again by virtue of the $\{\sigma^\mu\}$ being an orthogonal basis, the unique solution is $\psi_\ell^{\mathrm{I}} = 0$ and $\psi_\ell^{\mathrm{R}}\epsilon^{\ell kj} = \delta^{k[a}\delta^{b]j}$. The latter is equivalent to $\psi_\ell^{\mathrm{R}} = \epsilon_{\ell ab}$ and we therefore obtain

$$J_{\mathrm{L}}^{ab} = \epsilon_{\ell ab}\frac{\sigma^\ell}{2} \qquad \text{and} \qquad J_{\mathrm{L}}^c = \frac{\sigma^c}{2}. \tag{10.7.118}$$

Problem 10.69. Generators of $\mathrm{SL}_{2,\mathbb{C}}$: Second Set Suppose we replaced the basis $\{\sigma^\mu\}$ with the alternate one $\{\bar{\sigma}^\mu\}$ and instead solve for the $\{\psi_j\}$ in

$$\exp\left(-i\omega\psi_j\frac{\sigma^j}{2}\right)\bar{\sigma}^\mu \exp\left(i\omega\psi_j^*\frac{\sigma^j}{2}\right) = \bar{\sigma}^\nu\left(\exp\left(-i\omega\widehat{J}^{\alpha\beta}\right)\right)_\nu^{\ \ \mu}. \tag{10.7.119}$$

Show that the a-th boost generator is now

$$K_{\mathrm{R}}^a = i\frac{\sigma^a}{2}; \tag{10.7.120}$$

while the rotation generators remains the same

$$J_{\mathrm{R}}^{ab} = \epsilon_{\ell ab}\frac{\sigma^\ell}{2} \qquad \text{and} \qquad J_{\mathrm{R}}^c = \frac{\sigma^c}{2}. \tag{10.7.121}$$

Be sure to check that $\det \bar{\sigma}^\mu p_\mu = p^2$; so $\{\bar{\sigma}^\mu\}$ may indeed be used as an alternative basis to define the $\mathrm{SL}_{2,\mathbb{C}}$ group.

Finally: Notice both the $\vec{K}_{\mathrm{L}}$ and $\vec{K}_{\mathrm{R}}$ boost generators are anti-Hermitian; whereas the rotation ones $\vec{J}_{\mathrm{L}}$ and $\vec{J}_{\mathrm{R}}$ are Hermitian. $\square$

Let us now clarify the 'L' and 'R' subscripts occurring on these generators. Inserting our solutions for boosts and rotations into equations (10.7.7) and (10.7.8), we find

$$J^a_{+,\mathrm{L}} = \frac{1}{2}\left(J^a_{\mathrm{L}} + iK^a_{\mathrm{L}}\right) \tag{10.7.122}$$

$$= \frac{1}{2}\left(\frac{\sigma^a}{2} + i\frac{-i\sigma^a}{2}\right) = \frac{\sigma^a}{2}, \tag{10.7.123}$$

$$J^a_{-,\mathrm{L}} = \frac{1}{2}\left(J^a_{\mathrm{L}} - iK^a_{\mathrm{L}}\right) \tag{10.7.124}$$

$$= \frac{1}{2}\left(\frac{\sigma^a}{2} - i\frac{-i\sigma^a}{2}\right) = 0. \tag{10.7.125}$$

These not only satisfies the $\mathrm{SO}_{3,1}$ Lie algebra $[J^a_\pm, J^b_\pm] = i\epsilon^{abc}J^c_\pm$ and $[J^a_+, J^b_-] = 0$ in equations (10.7.9), (10.7.10) and (10.7.11); it informs us that we have obtained the $(j_+, j_-) = (1/2, 0)$ 'left-handed' representation. Similarly,

$$J^a_{+,\mathrm{R}} = \frac{1}{2}\left(J^a_{\mathrm{R}} + iK^a_{\mathrm{R}}\right) \tag{10.7.126}$$

$$= \frac{1}{2}\left(\frac{\sigma^a}{2} + i\frac{i\sigma^a}{2}\right) = 0, \tag{10.7.127}$$

$$J^a_{-,\mathrm{R}} = \frac{1}{2}\left(J^a_{\mathrm{R}} - iK^a_{\mathrm{R}}\right) \tag{10.7.128}$$

$$= \frac{1}{2}\left(\frac{\sigma^a}{2} - i\frac{i\sigma^a}{2}\right) = \frac{\sigma^a}{2}. \tag{10.7.129}$$

These not only satisfies the $\mathrm{SO}_{3,1}$ Lie algebra $[J^a_\pm, J^b_\pm] = i\epsilon^{abc}J^c_\pm$ and $[J^a_+, J^b_-] = 0$ in equations (10.7.9), (10.7.10) and (10.7.11); it informs us that we have obtained the $(j_+, j_-) = (0, 1/2)$ 'right-handed' representation.

Problem 10.70. Left and Right: Covariant Form Show that the 'left' and 'right' handed generators may be re-expressed using the relativistic notation

$$J^{\mu\nu}_{\mathrm{L}} \equiv \frac{i}{4}\sigma^{[\mu}\bar{\sigma}^{\nu]} \qquad \text{and} \qquad J^{\mu\nu}_{\mathrm{R}} \equiv \frac{i}{4}\bar{\sigma}^{[\mu}\sigma^{\nu]}. \tag{10.7.130}$$

That is, verify $J^{0a}_{\mathrm{L}} = K^a_{\mathrm{L}}$, $J^{0a}_{\mathrm{R}} = K^a_{\mathrm{R}}$, and $J^{ab}_{L,R} = \epsilon^{abc}\sigma^c/2$. □

Because these non-Hermitian generators are † of each other,

$$(J^{\mu\nu}_{\mathrm{L}})^\dagger = -\frac{i}{4}\bar{\sigma}^{[\nu}\sigma^{\mu]} = J^{\mu\nu}_{\mathrm{R}} \tag{10.7.131}$$

– which in turn means $(e^{-(i/2)\omega\cdot J_{\mathrm{L}}})^\dagger = e^{+(i/2)\omega\cdot J_{\mathrm{R}}}$ – we may now write

$$\exp\left(-\frac{i}{2}\omega_{\alpha\beta}J^{\alpha\beta}_{\mathrm{L}}\right)(\sigma^\mu p_\mu)\exp\left(\frac{i}{2}\omega_{\sigma\rho}J^{\sigma\rho}_{\mathrm{R}}\right) = \sigma^\nu p'_\nu, \tag{10.7.132}$$

$$\exp\left(-\frac{i}{2}\omega_{\alpha\beta}J^{\alpha\beta}_{\mathrm{R}}\right)(\bar{\sigma}^\mu p_\mu)\exp\left(\frac{i}{2}\omega_{\sigma\rho}J^{\sigma\rho}_{\mathrm{L}}\right) = \bar{\sigma}^\nu p'_\nu; \tag{10.7.133}$$

where the Lorentz-transformed momentum is

$$p'_\nu = \Lambda_\nu{}^\mu p_\mu, \tag{10.7.134}$$

$$\Lambda_\nu{}^\mu = \left(\exp\left(-\frac{i}{2}\omega_{\alpha\beta}\widehat{J}^{\alpha\beta}\right)\right)_\nu{}^\mu. \tag{10.7.135}$$

From here onward, we shall always denote the left handed $\mathrm{SL}_{2,\mathbb{C}}$ as L. In terms of the real rapidities $\{\xi_\ell\}$ and rotation angles $\{\theta_\ell\}$, the left-handed transformation is

$$L \equiv \exp\left(-\frac{i}{2}\omega_{\alpha\beta}J_{\mathrm{L}}^{\alpha\beta}\right) = \exp\left(-\xi_\ell\frac{\sigma^\ell}{2} - i\theta_\ell\frac{\sigma^\ell}{2}\right) \equiv e^{-\vec{\xi}\cdot\frac{\vec{\sigma}}{2}-i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}}; \tag{10.7.136}$$

and the right handed one is

$$(L^\dagger)^{-1} = (L^{-1})^\dagger = \exp\left(-\frac{i}{2}\omega_{\alpha\beta}J_{\mathrm{R}}^{\alpha\beta}\right) = \exp\left(\xi_\ell\frac{\sigma^\ell}{2} - i\theta_\ell\frac{\sigma^\ell}{2}\right) \equiv e^{\vec{\xi}\cdot\frac{\vec{\sigma}}{2}-i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}}. \tag{10.7.137}$$

When the rotation angles are zero, the resulting pure boost matrices are Hermitian ones. In particular, observe the left handed pure boost along $\vec{\xi}$ is the same as the right handed one in the opposite direction $-\vec{\xi}$:

$$L(-\vec{\xi},\vec{\theta}=\vec{0}) = (L^{-1}(\vec{\xi},\vec{\theta}=\vec{0}))^\dagger. \tag{10.7.138}$$

When the rapidity is zero, on the other hand, the left and right handed rotation transformations are not only identical, they become unitary ones:

$$L(\vec{\xi}=\vec{0},\vec{\theta}) = (L^{-1}(\vec{\xi}=\vec{0},\vec{\theta}))^\dagger. \tag{10.7.139}$$

Furthermore, these 'left' and 'right' handed $\mathrm{SL}_{2,\mathbb{C}}$ transformations may be related by complex conjugation and a similarity transformation. Recalling equations (10.7.31) and (10.7.32),

$$\epsilon\cdot\overline{L}\cdot\epsilon^\dagger = (L^\dagger)^{-1} = (L^{-1})^\dagger \qquad \text{and} \qquad \epsilon\cdot(L^{-1})^{\mathrm{T}}\cdot\epsilon^\dagger = L. \tag{10.7.140}$$

However, they *cannot* be related by a similarity transformation alone – there is no U such that

$$U\cdot L\cdot U^{-1} = (L^\dagger)^{-1}. \tag{10.7.141}$$

Hence, the 'left' and 'right' handed representations of $\mathrm{SL}_{2,\mathbb{C}}$ are said to be inequivalent; though we have just witnessed, the complex conjugate of the left handed group element *is* equivalent to the right handed one, and vice versa.

Let us now turn to the explicit implementation of rotations and Lorentz boosts through the above relation

$$L\cdot\sigma^\mu p_\mu\cdot L^\dagger = \sigma^\nu\Lambda_\nu{}^\mu p_\mu \equiv \sigma^\mu p'_\mu. \tag{10.7.142}$$

Rotations Set $\vec{\xi}=0$ in eq. (10.7.136) and focus on the case

$$\theta_j\sigma^j \to \theta\sigma^k \tag{10.7.143}$$

for a fixed $1 \leq k \leq 3$; so that eq. (10.7.42) informs us

$$L = \exp\left(-\frac{i}{2}\theta\sigma^k\right) = \cos(\theta/2) - i\sigma^k \sin(\theta/2). \tag{10.7.144}$$

Eq. (10.7.142), in turn, now reads

$$\begin{aligned} p_\mu\sigma^\mu &\to L \cdot p_\mu\sigma^\mu \cdot L^\dagger \\ &= e^{-(i/2)\theta\sigma^k} p_0 e^{(i/2)\theta\sigma^k} + \left(\cos(\theta/2) - i\sigma^k \sin(\theta/2)\right) p_i\sigma^i \left(\cos(\theta/2) + i\sigma^k \sin(\theta/2)\right) \\ &= p_0 + p'_i\sigma^i. \end{aligned} \tag{10.7.145}$$

If $k = 1$, we have p_i rotated on the $(2,3)$ plane:

$$p'_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}^j{}_i p_j. \tag{10.7.146}$$

If $k = 2$, we have p_i rotated on the $(1,3)$ plane:

$$p'_i = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}^j{}_i p_j. \tag{10.7.147}$$

If $k = 3$, we have p_i rotated on the $(1,2)$ plane:

$$p'_i = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}^j{}_i p_j. \tag{10.7.148}$$

Problem 10.71. Verify eq. (10.7.145) for any one of the $k = 1, 2, 3$. □

Boosts Next, we set $\vec{\theta} = 0$ in eq. (10.7.136) and focus on the case

$$\xi_j\sigma^j \to \xi\sigma^k, \tag{10.7.149}$$

again for a fixed $k = 1, 2, 3$. Again invoking eq. (10.7.42),

$$L = \exp\left(\frac{1}{2}\xi\sigma^k\right) = \cosh(\xi/2) + \sigma^k \sinh(\xi/2). \tag{10.7.150}$$

Eq. (10.7.142) is now

$$\begin{aligned} p_\mu\sigma^\mu &\to L \cdot p_\mu\sigma^\mu \cdot L^\dagger \\ &= \left(\cosh(\xi/2) + \sigma^k \sinh(\xi/2)\right) p_\mu\sigma^\mu \left(\cosh(\xi/2) + \sigma^k \sinh(\xi/2)\right) \\ &= p'_\mu\sigma^\mu. \end{aligned} \tag{10.7.151}$$

If $k = 1$, we have p_μ boosted in the 1−direction:

$$p'_\mu = \begin{bmatrix} \cosh\xi & \sinh\xi & 0 & 0 \\ \sinh\xi & \cosh\xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_\mu^{\ \ \nu} p_\nu. \tag{10.7.152}$$

If $k = 2$, we have p_μ boosted in the 2−direction:

$$p'_\mu = \begin{bmatrix} \cosh\xi & 0 & \sinh\xi & 0 \\ 0 & 1 & 0 & 0 \\ \sinh\xi & 0 & \cosh\xi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_\mu^{\ \ \nu} p_\nu. \tag{10.7.153}$$

If $k = 3$, we have p_μ boosted in the 3−direction:

$$p'_\mu = \begin{bmatrix} \cosh\xi & 0 & 0 & \sinh\xi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\xi & 0 & 0 & \cosh\xi \end{bmatrix}_\mu^{\ \ \nu} p_\nu. \tag{10.7.154}$$

Problem 10.72. Verify eq. (10.7.151) for any one of the $k = 1, 2, 3$. □

Problem 10.73. Spatial Rotation Around Axis Use $\mathrm{SL}_{2,\mathbb{C}}$ transformations to re-derive eq. (5.6.20); namely, show that the general rotation matrix describing counter-clockwise rotation by angle θ around the unit vector $\widehat{n}$ is given by

$$\widehat{R}_i^{\ j} = \cos\theta\,(\delta_{ij} - \widehat{n}_i\widehat{n}_j) + \sin\theta\cdot\epsilon_{ikj}\widehat{n}_k + \widehat{n}_i\widehat{n}_j. \tag{10.7.155}$$

(Hints: Both equations (10.7.22) and (10.7.23) may prove useful.) Further verify that, if the unit vector is parametrized using spherical coordinates,

$$\widehat{n}_i(\alpha,\beta) \equiv (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)\,, \tag{10.7.156}$$

the general rotation matrix then reads

$$\begin{aligned} &\widehat{R}(\theta\cdot\widehat{n}(\alpha,\beta)) \qquad\qquad (10.7.157) \\ &= \begin{bmatrix} \frac{1}{4}c_\theta\left(3 - 2s_\alpha^2 c_{2\beta} + c_{2\alpha}\right) + s_\alpha^2 c_\beta^2 & s_\alpha^2 s_{2\beta} s_{\theta/2}^2 - c_\alpha s_\theta & s_\alpha s_\beta s_\theta + s_{2\alpha} c_\beta s_{\theta/2}^2 \\ s_\alpha^2 s_{2\beta} s_{\theta/2}^2 + c_\alpha s_\theta & \frac{1}{4}c_\theta\left(3 + 2s_\alpha^2 c_{2\beta} + c_{2\alpha}\right) + s_\alpha^2 s_\beta^2 & s_{2\alpha} s_\beta s_{\theta/2}^2 - s_\alpha c_\beta s_\theta \\ s_{2\alpha} c_\beta s_{\theta/2}^2 - s_\alpha s_\beta s_\theta & s_{2\alpha} s_\beta s_{\theta/2}^2 + s_\alpha c_\beta s_\theta & s_\alpha^2 c_\theta + c_\alpha^2 \end{bmatrix}; \end{aligned}$$

where s and c are, respectively, sine and cosine. (This rotation matrix has been derived in Problem eq. (5.72) – refer to its eq. (5.6.16) – using different methods.) You may check this result recovers the expected form of the rotation matrix when the rotation axes are the 1−, 2−, and 3−Cartesian ones. □

Problem 10.74. General Lorentz Boost Use $\mathrm{SL}_{2,\mathbb{C}}$ transformations to show that the general Lorentz boost along the unit spatial vector $\widehat{n}$ with rapidity ξ – $\xi_\ell \equiv \xi \cdot \widehat{n}_\ell$ – is given by

$$\Lambda_\mu{}^\nu = \delta_\mu^0 \delta_0^\nu \cosh\xi + \delta_\mu^\ell \delta_\ell^\nu + \left(\delta_\mu^\ell \widehat{n}_\ell \delta_0^\nu + \delta_\mu^0 \widehat{n}_\ell \delta_\ell^\nu\right)\sinh\xi + \left(\delta_\mu^\ell \widehat{n}_\ell \widehat{n}_m \delta_m^\nu\right)(\cosh\xi - 1). \tag{10.7.158}$$

(Hints: Both equations (10.7.22) and (10.7.23) may prove useful.) Moreover, show that the general boost transformation reduces to the timelike case

$$\Lambda_\nu{}^\mu(\vec{v}) \doteq \begin{bmatrix} \gamma & \gamma v^i \\ \gamma v^i & \gamma\frac{v^i v^j}{\vec{v}^2} + \left(\delta^{ij} - \frac{v^i v^j}{\vec{v}^2}\right) \end{bmatrix}, \tag{10.7.159}$$

$$\gamma \equiv (1-\vec{v}^2)^{-1/2}, \qquad \vec{v}^2 \equiv \delta_{ab} v^a v^b \tag{10.7.160}$$

– recall eq. (10.3.180) – provided $\tanh|\xi| = |\vec{v}| < 1$. □

Dotted and Undotted: Weyl Spinors A left-handed *Weyl spinor* ξ_{A} is an object that transforms under Lorentz transformations as $\xi \to L \cdot \xi$, where L is an element of the $(j_+, j_-) = (1/2, 0)$ representation of $\mathrm{SL}_{2,\mathbb{C}}$. In index notation,

$$\xi_{\mathrm{A}}(p) \to \xi'_{\mathrm{A}}(p'_\mu = \Lambda_\mu{}^\nu p_\nu) \equiv L_{\mathrm{A}}{}^{\mathrm{B}} \xi_{\mathrm{B}}(p). \tag{10.7.161}$$

If the spinor transforms not as $\xi \to L \cdot \xi$ but as $\xi \to L^* \cdot \xi$, the right-handed $(j_+, j_-) = (0, 1/2)$ representation of $\mathrm{SL}_{2,\mathbb{C}}$, we put a dot on the relevant indices:

$$\xi_{\dot{\mathrm{A}}} \to \xi'_{\dot{\mathrm{A}}}(p'_\mu = \Lambda_\mu{}^\nu p_\nu) \equiv (L^*)_{\dot{\mathrm{A}}}{}^{\dot{\mathrm{B}}} \xi_{\dot{\mathrm{B}}}(p). \tag{10.7.162}$$

Of course, since ξ was arbitrary, taking the complex conjugate of eq. (10.7.161) brings us to the form in eq. (10.7.162); and vice versa. Hence, we may also denote

$$(\xi_{\mathrm{A}})^* = (\xi^*)_{\dot{\mathrm{A}}} \qquad \text{and} \qquad (\xi_{\dot{\mathrm{A}}})^* = (\xi^*)_{\mathrm{A}}. \tag{10.7.163}$$

As we shall prove shortly, L and L^* are *inequivalent* – i.e., there is no change-of-basis U such that $ULU^{-1} = L^*$. However, L^* is equivalent to the right-handed transformation in eq. (10.7.137). This dotted versus undotted notation essentially distinguishes between the left versus right handed spinors. For instance, we may now express eq. (10.7.62) as

$$L_{\mathrm{A}}{}^{\mathrm{M}} (L^*)_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{N}}} p_{\mathrm{M}\dot{\mathrm{N}}} = (\sigma^\nu)_{\mathrm{A}\dot{\mathrm{B}}} \Lambda_\nu{}^\mu p_\mu \equiv p'_{\mathrm{A}\dot{\mathrm{B}}} \tag{10.7.164}$$

$$= \lambda_+ \xi'^{+}_{\mathrm{A}} \overline{\xi'^{+}_{\dot{\mathrm{B}}}} + \lambda_- \xi'^{-}_{\mathrm{A}} \overline{\xi'^{-}_{\dot{\mathrm{B}}}}; \tag{10.7.165}$$

where the 'new' but un-normalized eigenvectors and eigenvalues are

$$\xi'^{\pm}{}_{\mathrm{A}}\left(p'_\mu = \Lambda_\mu{}^\nu p_\nu\right) = L_{\mathrm{A}}{}^{\mathrm{B}} \xi^{\pm}{}_{\mathrm{B}}(p_\mu), \tag{10.7.166}$$

$$(\overline{\xi'^{\pm}})_{\dot{\mathrm{A}}}\left(p'_\mu = \Lambda_\mu{}^\nu p_\nu\right) = (L^*)_{\dot{\mathrm{A}}}{}^{\dot{\mathrm{B}}} (\overline{\xi^{\pm}})_{\dot{\mathrm{B}}}(p_\mu); \tag{10.7.167}$$

with the old eigenvalues

$$\lambda_\pm \equiv p_0 \pm |\vec{p}|. \tag{10.7.168}$$

The dotted spinors transform as L^* while the undotted ones as L. Furthermore, as already alluded to, in the context of $p_\mu \sigma^\mu$, these $\xi^\pm$ are also helicity eigenstates of $p_i \sigma^i$.

If we normalize the spinors to unity

$$\zeta'^{\pm}_{\mathrm{A}} \equiv \xi'^{\pm}_{\mathrm{A}} \left| \left(\xi'^{\pm}\right)^\dagger \xi'^{\pm} \right|^{-\frac{1}{2}}; \tag{10.7.169}$$

then eq. (10.7.164) now reads

$$L_{\mathrm{A}}{}^{\mathrm{M}} (L^*)_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{N}}} p_{\mathrm{M}\dot{\mathrm{N}}} = p'_{\mathrm{A}\dot{\mathrm{B}}} = \lambda'_+ \zeta'^{+}_{\mathrm{A}} \overline{\zeta'^{+}_{\dot{\mathrm{B}}}} + \lambda'_- \zeta'^{-}_{\mathrm{A}} \overline{\zeta'^{-}_{\dot{\mathrm{B}}}}; \tag{10.7.170}$$

with the new eigenvalues

$$\lambda'_\pm \equiv p'_0 \pm |\vec{p}'|. \tag{10.7.171}$$

Problem 10.75. Rotating Spinors Derive the helicity eigenstates in equations (10.7.53) and (10.7.55) (up to overall phases) by rotating the eigenstate of $\widehat{e}_3 \cdot \vec{\sigma} = \sigma^3$; see eq. (10.7.24). Hint: The generic radial vector $\widehat{r} = (\sin\theta_p \cos\phi_p, \sin\theta_p \sin\phi_p, \cos\theta_p)$ can be obtained from $\widehat{e}_3$, the unit vector along the 3−axis, by first rotating it around the $\widehat{e}_2$ axis by θ_p; followed by rotating it around the $\widehat{e}_3$ axis by ϕ_p. □

Problem 10.76. Boosting Massive Spinors In this problem, we will study the spinor pure-boost operator

$$L_{\mathrm{B}} \equiv \exp\left(-\xi_i \sigma^i/2\right) \tag{10.7.172}$$

using the results from eq. (10.7.64) and Problem (10.74).

Positive Energy Let p_μ be a 4−momentum obeying massive dispersion relation with positive energy:

$$p_\mu = \left(\sqrt{p^2+m^2}, p_i\right). \tag{10.7.173}$$

Throughout, with $p \equiv |\vec{p}| \geq 0$ and $m > 0$, let us set

$$\xi_i \equiv -\operatorname{arctanh}\left[\frac{p}{\sqrt{p^2+m^2}}\right] \cdot \widehat{p}_i, \tag{10.7.174}$$

$$\widehat{p}_i \equiv p_i/p. \tag{10.7.175}$$

First, show that

$$L_{\mathrm{B}} \left(m \cdot \sigma^0\right) L_{\mathrm{B}}^\dagger = m \cdot L_{\mathrm{B}}^2 = p_\nu \sigma^\nu \equiv \sigma \cdot p \tag{10.7.176}$$

and

$$(L_{\mathrm{B}}^{-1})^\dagger \left(m \cdot \sigma^0\right) L_{\mathrm{B}}^{-1} = m \cdot (L_{\mathrm{B}}^{-1})^2 = p_\nu \bar{\sigma}^\nu \equiv \bar{\sigma} \cdot p. \tag{10.7.177}$$

We see that L_{B} boosts the 4−momentum $(m, \vec{0})$ ‘at rest’ to a generic p_μ with positive energy.

Next, explain why, when viewed as a matrices,

$$L_{\mathrm{B}} = \sqrt{\sigma \cdot p/m}, \tag{10.7.178}$$

$$(L_{\mathrm{B}}^{-1})^\dagger = L_{\mathrm{B}}^{-1} = \sqrt{\bar{\sigma} \cdot p/m}; \tag{10.7.179}$$

where all the square roots are the positive one. Namely: since both $\sigma \cdot p$, and $\bar{\sigma} \cdot p$ are Hermitian, and since they commute (why?), they must be simultaneously diagonalized by the unit norm helicity eigenstates ($\xi^\dagger \xi = 1$) in eq. (10.7.53) and (10.7.55):

$$\sqrt{\sigma \cdot p} = \sqrt{\lambda_+} \xi^+ (\xi^+)^\dagger + \sqrt{\lambda_-} \xi^- (\xi^-)^\dagger, \tag{10.7.180}$$

$$\sqrt{\bar{\sigma} \cdot p} = \sqrt{\lambda_-} \xi^+ (\xi^+)^\dagger + \sqrt{\lambda_+} \xi^- (\xi^-)^\dagger, \tag{10.7.181}$$

$$\sqrt{\lambda_\pm} = \sqrt{\sqrt{p^2 + m^2} \pm p}. \tag{10.7.182}$$

The eigenvalues of $\sqrt{m} \cdot L_{\mathrm{B}}$ are the positive square roots of the eigenvalues of $\sigma \cdot p$; while those of $\sqrt{m} \cdot L_{\mathrm{B}}^{-1}$ are the positive square roots of those of $\bar{\sigma} \cdot p$.

Negative Energy Now, let p'_μ be a 4−momentum obeying massive dispersion relation with negative energy:

$$p'_\mu = \left(-\sqrt{p^2 + m^2}, -p_i \right) = -p_\mu. \tag{10.7.183}$$

[113]With the same L_{B} in eq. (10.7.176), show that

$$L_{\mathrm{B}} \left(-m \cdot \sigma^0 \right) L_{\mathrm{B}}^\dagger = -\sigma \cdot p = \sigma \cdot p' \tag{10.7.184}$$

and

$$(L_{\mathrm{B}}^{-1})^\dagger \left(-m \cdot \sigma^0 \right) L_{\mathrm{B}}^{-1} = -\bar{\sigma} \cdot p = \bar{\sigma} \cdot p'. \tag{10.7.185}$$

We see that L_{B} boosts a 4−momentum $(-m, \vec{0})$ 'at rest' to a generic p'_μ with negative energy. □

(1/2,0)⊗(0,1/2) as a vector Building upon our understanding of dotted versus undotted spinors, we may now recognize eq. (10.7.164) as the tensor product representation $(\frac{1}{2}, 0) \otimes (0, \frac{1}{2})$ giving rise to the spacetime vector.

$$\begin{aligned} &\left(\exp\left[-\left(\vec{\xi} + i\vec{\theta} \right) \cdot \frac{\vec{\sigma}}{2} \right]_{\left(\frac{1}{2},0\right)} \right)_{\mathrm{A}}{}^{\mathrm{M}} \left(\overline{\exp\left[-\left(\vec{\xi} + i\vec{\theta} \right) \cdot \frac{\vec{\sigma}}{2} \right]_{\left(0,\frac{1}{2}\right)}} \right)_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{N}}} \sigma^\mu{}_{\mathrm{M}\dot{\mathrm{N}}} \\ &= \sigma^\nu{}_{\mathrm{A}\dot{\mathrm{B}}} \Lambda_\nu{}^\mu(\vec{\xi}, \vec{\theta}) = \sigma^\nu{}_{\mathrm{A}\dot{\mathrm{B}}} \left(\exp\left(-i\xi_\ell \widehat{J}^{0\ell} - i\theta_\ell \widehat{J}^\ell \right) \right)_\nu{}^\mu . \end{aligned} \tag{10.7.186}$$

Contracting both sides with a momentum vector returns

$$L_{\mathrm{A}}{}^{\mathrm{M}} (L^*)_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{N}}} p_{\mathrm{M}\dot{\mathrm{N}}} = p'_{\mathrm{A}\dot{\mathrm{B}}}, \tag{10.7.187}$$

$$p'_\mu = \Lambda_\mu{}^\nu p_\nu. \tag{10.7.188}$$

[113]The negative sign in front of the spatial components is for technical convenience. Since p_i is arbitrary, there is no loss of generality; though p'_μ should be interpreted as the negative energy *and* parity-flipped version of p_μ.

Raising and Lowering Spinor Indices Since the Levi-Civita ϵ acts like a 'metric', in the sense that $\epsilon_{\mathrm{AB}} L^{\mathrm{A}}{}_{\mathrm{M}} L^{\mathrm{B}}{}_{\mathrm{N}} = \epsilon_{\mathrm{MN}}$ is analogous to $\eta_{\alpha\beta}\Lambda^{\alpha}{}_{\mu}\Lambda^{\beta}{}_{\nu} = \eta_{\mu\nu}$, let us define the raising of a spinor index using the Levi-Civita metric $\epsilon_{\mathrm{AB}} = \epsilon^{\mathrm{AB}}$ via the relations

$$\xi^{\mathrm{A}} \equiv \xi_{\mathrm{B}}\epsilon^{\mathrm{BA}} \qquad \text{and} \qquad \xi^{\dot{\mathrm{A}}} \equiv \xi_{\dot{\mathrm{B}}}\epsilon^{\dot{\mathrm{B}}\dot{\mathrm{A}}}. \tag{10.7.189}$$

Lowering an index is accomplished through

$$\xi_{\mathrm{A}} \equiv \epsilon_{\mathrm{AB}}\xi^{\mathrm{B}} \qquad \text{and} \qquad \xi_{\dot{\mathrm{A}}} \equiv \epsilon_{\dot{\mathrm{A}}\dot{\mathrm{B}}}\xi^{\dot{\mathrm{B}}}. \tag{10.7.190}$$

It is important to remember, we contract with the right index of ϵ when lowering an index; whereas we do so with the left index of ϵ when raising. This is to ensure a consistent result – for e.g.,

$$\xi_{\mathrm{A}} = \epsilon_{\mathrm{AB}}\xi^{\mathrm{B}} = \epsilon_{\mathrm{AB}}\epsilon^{\mathrm{CB}}\xi_{\mathrm{C}} = \left(\epsilon\cdot\epsilon^{\dagger}\right)_{\mathrm{A}}{}^{\mathrm{C}}\xi_{\mathrm{C}} = \delta_{\mathrm{A}}{}^{\mathrm{C}}\xi_{\mathrm{C}}. \tag{10.7.191}$$

We may even move the indices of ϵ; for instance, keeping in mind $\epsilon^2 = -\mathbb{I}$,

$$\epsilon_{\mathrm{AB}} = \epsilon_{\mathrm{AM}}\epsilon_{\mathrm{BN}}\epsilon^{\mathrm{MN}} \tag{10.7.192}$$
$$= -\delta^{\mathrm{N}}_{\mathrm{A}}\epsilon_{\mathrm{BN}} = -\epsilon_{\mathrm{BA}}. \tag{10.7.193}$$

Problem 10.77. Show that, for arbitrary spinors ξ and ζ,

$$\xi\cdot\zeta \equiv \xi^{\mathrm{A}}\zeta_{\mathrm{A}} = -\xi_{\mathrm{A}}\zeta^{\mathrm{A}} = -\zeta\cdot\xi. \tag{10.7.194}$$

Hence, unlike the spacetime tensor case, a pair of contracted up-down indices may be swapped in their positions – but at the cost of a minus sign. □

Transformation of upper indices We have defined the transformation of lower spinor indices through equations (10.7.161) and (10.7.162). For the upper indices, let us use $(L^{-1})^{\mathrm{T}} = \epsilon^{\dagger}\cdot L\cdot\epsilon$ to deduce,

$$\xi^{\mathrm{A}} \to \left(L_{\mathrm{B}}{}^{\mathrm{C}}\xi_{\mathrm{C}}\right)\epsilon^{\mathrm{BA}} \tag{10.7.195}$$
$$= \left(\epsilon^{\mathrm{BA}}L_{\mathrm{B}}{}^{\mathrm{C}}\epsilon_{\mathrm{CD}}\right)\xi^{\mathrm{D}} = \xi^{\mathrm{B}}(L^{-1})_{\mathrm{B}}{}^{\mathrm{A}}, \tag{10.7.196}$$

where $(L^{-1})_{\mathrm{B}}{}^{\mathrm{A}} \equiv ((L^{-1})^{\mathrm{T}})^{\mathrm{A}}{}_{\mathrm{B}}$.

Problem 10.78. Upper Dotted Index Show that, under a Lorentz transformation,

$$\xi^{\dot{\mathrm{A}}} \to ((L^{\dagger})^{-1})^{\dot{\mathrm{A}}}{}_{\dot{\mathrm{B}}}\xi^{\dot{\mathrm{B}}} = \xi^{\dot{\mathrm{B}}}((L^{*})^{-1})_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{A}}}. \tag{10.7.197}$$

Hint: Remember $(L^{\dagger})^{-1} = \epsilon\cdot L^{*}\cdot\epsilon^{\dagger}$. □

These results tell us, just like in tensor calculus, the upper and lower indices transform oppositely. This means a Lorentz scalar is formed when a pair of upper and lower indices are contracted; for instance, $\xi^{\mathrm{A}}\zeta_{\mathrm{A}}$ or $\xi^{\dot{\mathrm{A}}}\zeta_{\dot{\mathrm{A}}}$. For, under Lorentz transformations,

$$\xi^{\mathrm{A}}\zeta_{\mathrm{A}} \to \xi^{\mathrm{B}}(L^{-1})_{\mathrm{B}}{}^{\mathrm{A}}L_{\mathrm{A}}{}^{\mathrm{C}}\zeta_{\mathrm{C}} \tag{10.7.198}$$

$$= \xi^{\mathrm{B}} \delta_{\mathrm{B}}{}^{\mathrm{C}} \zeta_{\mathrm{C}}. \tag{10.7.199}$$

Likewise,

$$\xi^{\dot{\mathrm{A}}} \zeta_{\dot{\mathrm{A}}} \to \xi^{\dot{\mathrm{B}}} ((L^*)^{-1})_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{A}}} L^*{}_{\dot{\mathrm{A}}}{}^{\dot{\mathrm{C}}} \zeta_{\dot{\mathrm{C}}} \tag{10.7.200}$$

$$\to \xi^{\dot{\mathrm{B}}} \delta_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{C}}} \zeta_{\dot{\mathrm{C}}}. \tag{10.7.201}$$

Alternately, we may also see that

$$\xi \cdot \zeta \equiv \xi_{\mathrm{A}} \epsilon^{\mathrm{AB}} \zeta_{\mathrm{B}} \to \epsilon^{\mathrm{IJ}} L_{\mathrm{I}}{}^{\mathrm{A}} L_{\mathrm{J}}{}^{\mathrm{B}} \xi_{\mathrm{A}} \zeta_{\mathrm{B}} \tag{10.7.202}$$

$$= (\det L) \epsilon^{\mathrm{AB}} \xi_{\mathrm{A}} \zeta_{\mathrm{B}} = \xi \cdot \zeta. \tag{10.7.203}$$

The second equality is due to the defining condition of the $\mathrm{SL}_{2,\mathbb{C}}$ group, $\det L = 1$, as expressed in eq. (10.7.256). Likewise,

$$\epsilon^{\dot{\mathrm{A}}\dot{\mathrm{B}}} \xi_{\dot{\mathrm{A}}} \zeta_{\dot{\mathrm{B}}} \to \epsilon^{\dot{\mathrm{I}}\dot{\mathrm{J}}} \overline{L_{\dot{\mathrm{I}}}{}^{\dot{\mathrm{A}}} L_{\dot{\mathrm{J}}}{}^{\dot{\mathrm{B}}}} \xi_{\dot{\mathrm{A}}} \zeta_{\dot{\mathrm{B}}} = \epsilon^{\dot{\mathrm{A}}\dot{\mathrm{B}}} \xi_{\dot{\mathrm{A}}} \zeta_{\dot{\mathrm{B}}}. \tag{10.7.204}$$

Note that the scalar product between a dotted and un-dotted spinor $\epsilon^{\mathrm{AB}} \xi_{\dot{\mathrm{A}}} \eta_{\mathrm{B}}$ would not, in general, be an invariant because its transformation will involve both L and L^*.

Problem 10.79. Parity Transformation: Index Notation Explain why, the parity transformation formula in eq. (10.7.65) translates, in index notation, into

$$\overline{(\xi^{\pm})^{\mathrm{A}}} = (\xi^{\pm})^{\dot{\mathrm{A}}} = (\overline{\xi^{\pm}})_{\dot{\mathrm{B}}} \epsilon^{\dot{\mathrm{B}}\dot{\mathrm{A}}} = e^{i\delta} (\xi^{\mp})_{\mathrm{A}} \tag{10.7.205}$$

and

$$(\xi^{\pm})^{\mathrm{A}} = (\xi^{\pm})_{\mathrm{B}} \epsilon^{\mathrm{BA}} = e^{-i\delta} (\overline{\xi^{\mp}})_{\dot{\mathrm{A}}}; \tag{10.7.206}$$

where δ is an arbitrary real number and the $(\xi^{\pm})_{\mathrm{A}}$ are the A-th components of the positive $(+)$ and negative $(-)$ helicity eigenstates. □

Problem 10.80. Starting from $(\sigma \cdot p)_{\mathrm{A}\dot{\mathrm{B}}} = p_{\mathrm{A}\dot{\mathrm{B}}}$ or otherwise, explain why the AB components of $\bar{\sigma} \cdot p \equiv \bar{\sigma}^\mu p_\mu$ should be denoted as

$$p^{\dot{\mathrm{A}}\mathrm{B}} \equiv (\bar{\sigma} \cdot p)^{\dot{\mathrm{A}}\mathrm{B}} = p_\mu (\bar{\sigma}^\mu)^{\dot{\mathrm{A}}\mathrm{B}}. \tag{10.7.207}$$

Moreover, demonstrate that

$$(\bar{\sigma}^\mu)^{\dot{\mathrm{E}}\mathrm{F}} L_{\dot{\mathrm{E}}}{}^{\dot{\mathrm{A}}} L_{\mathrm{F}}{}^{\mathrm{B}} = (\bar{\sigma}^\nu)^{\dot{\mathrm{A}}\mathrm{B}} \Lambda_\nu{}^\mu. \tag{10.7.208}$$

If we view $\sigma \cdot p$ as matrix, it is equal to $\lambda_+ \xi^+ (\xi^+)^\dagger + \lambda_- \xi^- (\xi^-)^\dagger$, for eigenvalues $\lambda_\pm = p_0 \pm |\vec{p}|$ and associated unit-norm eigenvectors $\xi^\pm$. We also have $\bar{\sigma} \cdot p = \lambda_- \xi^+ (\xi^+)^\dagger + \lambda_+ \xi^- (\xi^-)^\dagger$ – why? From this, demonstrate that

$$(\xi^{\pm})^{\dot{\mathrm{A}}} (\xi^{\pm})^{\mathrm{B}} = (\xi^{\mp})_{\mathrm{A}} (\xi^{\mp})_{\dot{\mathrm{B}}}. \tag{10.7.209}$$

(This can also be obtained from equations (10.7.205) and (10.7.206) above.) Also explain why

$$p_{\mathrm{A}\dot{\mathrm{B}}} p^{\dot{\mathrm{B}}\mathrm{C}} = p^2 \cdot \delta_{\mathrm{A}}{}^{\mathrm{C}}. \tag{10.7.210}$$

Hints: From Problem (10.63), how is $(\sigma^\mu)^*$ related to $\bar{\sigma}^\mu$? How does $(L^{-1})^\dagger (\bar{\sigma} \cdot p) L^{-1}$ transform? □

Problem 10.81. Spacetime Vectors from Bi-Spinors If $L\sigma^\mu L^\dagger = \sigma^\nu \Lambda_\nu{}^\mu$, derive the following transformation rules:

$$\xi^{\mathrm{A}} \sigma^\mu{}_{\mathrm{A}\dot{\mathrm{B}}} \zeta^{\dot{\mathrm{B}}} \to \Lambda^\mu{}_\nu \xi^{\mathrm{A}} \sigma^\nu{}_{\mathrm{A}\dot{\mathrm{B}}} \zeta^{\dot{\mathrm{B}}}, \tag{10.7.211}$$

$$\xi_{\dot{\mathrm{A}}} (\bar{\sigma}^\mu)^{\dot{\mathrm{A}}\mathrm{B}} \zeta_{\mathrm{B}} \to \Lambda^\mu{}_\nu \xi_{\dot{\mathrm{A}}} (\bar{\sigma}^\nu)^{\dot{\mathrm{A}}\mathrm{B}} \zeta_{\mathrm{B}}. \tag{10.7.212}$$

In matrix form, they read instead

$$\rho^\dagger \sigma^\mu \rho \to \Lambda^\mu{}_\nu \rho^\dagger \sigma^\nu \rho, \tag{10.7.213}$$

$$\lambda^\dagger \bar{\sigma}^\mu \lambda \to \Lambda^\mu{}_\nu \lambda^\dagger \bar{\sigma}^\nu \lambda; \tag{10.7.214}$$

where λ is a left-handed spinor while ρ is a right-handed one. Can you explain why

$$\rho^\dagger \sigma^\mu \lambda, \qquad \lambda^\dagger \sigma^\mu \rho, \qquad \rho^\dagger \bar{\sigma}^\mu \lambda \qquad \text{and} \qquad \lambda^\dagger \bar{\sigma}^\mu \rho \tag{10.7.215}$$

do not transform properly as Lorentz tensors? □

Problem 10.82. Prove that

$$\frac{1}{2}\left(\rho_{\mathrm{I}}^\dagger \sigma^\mu \rho_{\mathrm{II}}\right)\left(\lambda_{\mathrm{I}}^\dagger \bar{\sigma}_\mu \lambda_{\mathrm{II}}\right) = \left(\rho_{\mathrm{I}}^\dagger \lambda_{\mathrm{II}}\right)\left(\lambda_{\mathrm{I}}^\dagger \rho_{\mathrm{II}}\right). \tag{10.7.216}$$

Observe that this is a *Lorentz invariant* relation – provided the ρ's transform as right handed spinors and the λ's as left handed ones. □

Parity Thus far we have only studied continuous Lorentz transformations – boosts and rotations – implemented using $\mathrm{SL}_{2,\mathbb{C}}$ transformations $\{L\}$ via $L\sigma^\mu L^\dagger = \sigma^\nu \Lambda_\nu{}^\mu$. Let us turn to the parity operation, defined by the swapping of the spatial components of momentum with its negative:

$$p_i \leftrightarrow -p_i. \tag{10.7.217}$$

This amounts to swapping

$$\sigma^\mu p_\mu \leftrightarrow \bar{\sigma}^\mu p_\mu \tag{10.7.218}$$

which is equivalent to swapping

$$p_{\mathrm{A}\dot{\mathrm{B}}} \leftrightarrow p^{\dot{\mathrm{A}}\mathrm{B}}. \tag{10.7.219}$$

Moreover, remember that both $p_{\mathrm{A}\dot{\mathrm{B}}}$ and $p^{\dot{\mathrm{A}}\mathrm{B}}$ involves the sum over complete set of spinors:

$$p_{\mathrm{A}\dot{\mathrm{B}}} = \lambda_+ (\xi^+)_{\mathrm{A}} (\xi^+)_{\dot{\mathrm{B}}} + \lambda_- (\xi^-)_{\mathrm{A}} (\xi^-)_{\dot{\mathrm{B}}} \tag{10.7.220}$$

and

$$p^{\dot{\mathrm{A}}\mathrm{B}} = \lambda_+ (\xi^+)^{\dot{\mathrm{A}}} (\xi^+)^{\mathrm{B}} + \lambda_- (\xi^-)^{\dot{\mathrm{A}}} (\xi^-)^{\mathrm{B}} \tag{10.7.221}$$

$$= \lambda_+ (\xi^-)_{\mathrm{A}} (\xi^-)_{\dot{\mathrm{B}}} + \lambda_- (\xi^+)_{\mathrm{A}} (\xi^+)_{\dot{\mathrm{B}}}. \tag{10.7.222}$$

This tells us, at the level of spinors, parity is implemented (up to multiplicative phase factors) – recall, too, equations (10.7.205) and (10.7.206) – via the swap

$$\xi_{\mathrm{A}} \leftrightarrow \xi^{\dot{\mathrm{A}}}. \tag{10.7.223}$$

Summary: $\mathbf{SL}_{2,\mathbb{C}}$ group elements With the convention

$$\Lambda = \exp(-i\vec{\xi}\cdot\vec{K} - i\vec{\theta}\cdot\vec{J}) \equiv \exp\left(-i\xi_\ell \widehat{K}^\ell - i\theta_\ell \widehat{J}^\ell\right), \tag{10.7.224}$$

we summarize the $\mathrm{SL}_{2,\mathbb{C}}$ group as follows.

Chiral Left For the $(j_+, j_-) = (1/2, 0)$ representation of the Lorentz group, the corresponding $\mathrm{SL}_{2,\mathbb{C}}$ generators are $J_{\mathrm{L}}^{\mu\nu} \equiv \frac{i}{4}\sigma^{[\mu}\bar{\sigma}^{\nu]}$. For real boost $\{\xi_j\}$ and rotation $\{\theta_j\}$ parameters,

$$\xi_{\mathrm{A}} \to \xi'_{\mathrm{A}} \equiv L_{\mathrm{A}}{}^{\mathrm{B}}\xi_{\mathrm{B}}, \tag{10.7.225}$$

$$\xi^{\mathrm{A}} \to \xi'^{\mathrm{A}} \equiv \xi^{\mathrm{B}}(L^{-1})_{\mathrm{B}}{}^{\mathrm{A}}; \tag{10.7.226}$$

where

$$L\left(\vec{\xi}, \vec{\theta}\right) = \exp\left(-\frac{i}{2}\omega_{\mu\nu}J_{\mathrm{L}}^{\mu\nu}\right) \tag{10.7.227}$$

$$= \exp\left(-\vec{\xi}\cdot\frac{\vec{\sigma}}{2} - i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}\right); \tag{10.7.228}$$

$$K_{\mathrm{L}}^i = -\frac{i}{2}\sigma^i \qquad \text{and} \qquad J_{\mathrm{L}}^i = \frac{1}{2}\sigma^i. \tag{10.7.229}$$

The left handed $\mathrm{SL}_{2,\mathbb{C}}$ transformation itself is

$$L(\sigma^\mu p_\mu)L^\dagger = \sigma^\mu \Lambda_\mu{}^\nu p_\nu \equiv \sigma^\mu p'_\mu. \tag{10.7.230}$$

Chiral Right Whereas, the 'chiral right' inequivalent $(j_+, j_-) = (0, 1/2)$ representation of the Lorentz has corresponding $\mathrm{SL}_{2,\mathbb{C}}$ generators $J_{\mathrm{R}}^{\mu\nu} \equiv \frac{i}{4}\bar{\sigma}^{[\mu}\sigma^{\nu]}$; and is provided by the transformation

$$\xi^{\dot{\mathrm{A}}} \to \xi'^{\dot{\mathrm{A}}} \equiv ((L^\dagger)^{-1})^{\dot{\mathrm{A}}}{}_{\dot{\mathrm{B}}}\xi^{\dot{\mathrm{B}}} \tag{10.7.231}$$

$$= \xi^{\dot{\mathrm{B}}}((L^*)^{-1})_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{A}}}, \tag{10.7.232}$$

$$\xi_{\dot{\mathrm{A}}} \to \xi'_{\dot{\mathrm{A}}} \equiv (L^*)_{\dot{\mathrm{A}}}{}^{\dot{\mathrm{B}}}\xi_{\dot{\mathrm{B}}}; \tag{10.7.233}$$

where

$$\overline{L\left(\vec{\xi}, \vec{\theta}\right)} = \epsilon^\dagger \exp\left(-\frac{i}{2}\omega_{\mu\nu}J_{\mathrm{R}}^{\mu\nu}\right)\epsilon \tag{10.7.234}$$

$$= \epsilon^\dagger \exp\left(\vec{\xi}\cdot\frac{\vec{\sigma}}{2} - i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}\right)\epsilon \tag{10.7.235}$$

$$= \epsilon^\dagger \left(L\left(\vec{\xi}, \vec{\theta}\right)^\dagger\right)^{-1}\epsilon = \epsilon^\dagger \left(L\left(\vec{\xi}, \vec{\theta}\right)^{-1}\right)^\dagger \epsilon, \tag{10.7.236}$$

$$K^i_{\mathrm{R}} = \frac{i}{2}\sigma^i \qquad \text{and} \qquad J^i_{\mathrm{R}} = \frac{1}{2}\sigma^i. \tag{10.7.237}$$

And, for the same $p'_\mu = \Lambda_\mu{}^\nu p_\nu$ as the above 'right-handed' transformation,

$$(L^{-1})^\dagger(\bar{\sigma}^\mu p_\mu)L^{-1} = \bar{\sigma}^\nu p'_\nu. \tag{10.7.238}$$

YZ: Older stuff below – ignore.

for some linear transformation $\Lambda^\mu{}_\nu$. Now, the determinant transforms as

$$\det(\sigma^\nu p_\nu) \to \det p_\mu(L\cdot\sigma^\mu\cdot L^\dagger) \tag{10.7.239}$$
$$= \det p'_\mu\sigma^\mu = p'^2 = p^2, \tag{10.7.240}$$
$$p'_\nu \equiv p_\nu\Lambda^\mu{}_\nu. \tag{10.7.241}$$

But $p'^2 = (\eta^{\alpha\beta}M_\alpha{}^\mu M_\beta{}^\nu)p_\mu p_\nu = p^2$ means $M_\nu{}^\mu$ must be a Lorentz transformation.

$$L\cdot\sigma^\mu p_\mu\cdot L^\dagger = \sigma^\nu\Lambda_\nu{}^\mu p_\mu \equiv \sigma^\mu p'_\mu. \tag{10.7.242}$$

In other words,

$$p'_\mu = \Lambda_\mu{}^\nu p_\nu, \tag{10.7.243}$$
$$L(\vec{\xi},\vec{\theta})\cdot\sigma^\mu\cdot L(\vec{\xi},\vec{\theta})^\dagger = \sigma^\nu\Lambda_\nu{}^\mu(\vec{\xi},\vec{\theta}). \tag{10.7.244}$$

Problem 10.83. Explain why eq. (10.7.244) implies

$$L^{-1}\sigma^\mu(L^{-1})^\dagger = \Lambda^\mu{}_\nu(\vec{\xi},\vec{\theta})\sigma^\nu, \tag{10.7.245}$$
$$\Lambda^\alpha{}_\nu\Lambda_\beta{}^\nu = \delta^\alpha{}_\beta. \tag{10.7.246}$$

Next, explain why eq. (10.7.244) also leads immediately to

$$L^*(\sigma^\mu)^*L^{\mathrm{T}} = (\sigma^\nu)^*\Lambda_\nu{}^\mu(\vec{\xi},\vec{\theta}). \tag{10.7.247}$$

Then demonstrate that

$$(L^{-1})^\dagger\bar{\sigma}^\mu L^{-1} = \bar{\sigma}^\nu\Lambda_\nu{}^\mu(\vec{\xi},\vec{\theta}); \tag{10.7.248}$$

and

$$L^\dagger\bar{\sigma}^\mu L = \Lambda^\mu{}_\nu\bar{\sigma}^\nu = \bar{\sigma}^\nu\Lambda_\nu{}^\mu(-\vec{\xi},-\vec{\theta}). \tag{10.7.249}$$

Next, employ eq. (**??**) to explain why

$$L^\dagger(\vec{\xi},\vec{\theta}) = L(\vec{\xi},-\vec{\theta}), \tag{10.7.250}$$
$$(L^{-1})^\dagger(\vec{\xi},\vec{\theta}) = L(-\vec{\xi},\vec{\theta}). \tag{10.7.251}$$

Why do these imply the following?

$$L^\dagger\sigma^\mu L = \sigma^\nu\Lambda_\nu{}^\mu(\vec{\xi},-\vec{\theta}), \tag{10.7.252}$$

$$L^{-1}\bar{\sigma}^{\mu}(L^{-1})^{\dagger} = \bar{\sigma}^{\nu}\Lambda_{\nu}{}^{\mu}(\vec{\xi}, -\vec{\theta}). \tag{10.7.253}$$

$$(L^{-1})^{\dagger}\sigma^{\mu}L^{-1} = \sigma^{\nu}\Lambda_{\nu}{}^{\mu}(-\vec{\xi}, \vec{\theta}), \tag{10.7.254}$$
$$L\bar{\sigma}^{\mu}L^{\dagger} = \bar{\sigma}^{\nu}\Lambda_{\nu}{}^{\mu}(-\vec{\xi}, \vec{\theta}). \tag{10.7.255}$$

Notice the equation pairs (10.7.245) and (10.7.249) lead to the same 4×4 Lorentz transformation $\Lambda_{\nu}{}^{\mu}(-\vec{\xi}, -\vec{\theta})$; whereas, equation pairs (10.7.245) and (10.7.249) lead to the same 4×4 Lorentz transformation $\Lambda_{\nu}{}^{\mu}(\vec{\xi}, -\vec{\theta})$. These will play key roles in understanding how the Dirac equation transforms under Lorentz transformations. □

Rotations, Boosts, and $\mathbf{SL}_{2,\mathbb{C}}$ We have discovered that the group of 2×2 matrices $\{L\}$ continuously connected to the identity obeying

$$\epsilon^{\mathrm{AB}}L_{\mathrm{A}}{}^{\mathrm{I}}L_{\mathrm{B}}{}^{\mathrm{J}} = \epsilon^{\mathrm{IJ}} \qquad \Leftrightarrow \qquad L_{\mathrm{A}}{}^{\mathrm{I}}L_{\mathrm{B}}{}^{\mathrm{J}}\epsilon_{\mathrm{IJ}} = \epsilon_{\mathrm{AB}} \tag{10.7.256}$$

implements Lorentz transformations

$$L_{\mathrm{A}}{}^{\mathrm{I}}\overline{L_{\dot{\mathrm{B}}}{}^{\dot{\mathrm{J}}}}\sigma^{\mu}{}_{\mathrm{I}\dot{\mathrm{J}}} = \sigma^{\nu}{}_{\mathrm{A}\dot{\mathrm{B}}}\Lambda_{\nu}{}^{\mu}. \tag{10.7.257}$$

In terms of matrix multiplication,

$$L\sigma^{\mu}L^{\dagger} = \sigma^{\nu}\Lambda_{\nu}{}^{\mu}; \tag{10.7.258}$$

where the $\Lambda_{\nu}{}^{\mu}$ is the 4×4 Lorentz transformations parametrized by $\{\vec{\xi}, \vec{\theta}\}$ satisfying eq. (10.1.13). Other equivalent forms are

$$L^{-1}\sigma^{\mu}(L^{-1})^{\dagger} = \Lambda^{\mu}{}_{\nu}\sigma^{\nu}, \tag{10.7.259}$$
$$L^{*}(\sigma^{\mu})^{*}L^{\mathrm{T}} = (\sigma^{\nu})^{*}\Lambda_{\nu}{}^{\mu}, \tag{10.7.260}$$
$$(L^{-1})^{\dagger}\bar{\sigma}^{\mu}L^{-1} = \bar{\sigma}^{\nu}\Lambda_{\nu}{}^{\mu}, \tag{10.7.261}$$
$$L^{\dagger}\bar{\sigma}^{\mu}L = \Lambda^{\mu}{}_{\nu}\bar{\sigma}^{\nu}. \tag{10.7.262}$$

□

L and L^* are inequivalent L and its complex conjugate $\overline{L} = L^*$ are not equivalent transformations once Lorentz boosts are included; i.e., once $\vec{\xi} \neq 0$. To see this, we first recall, for any Taylor-expandable function f, $Uf(A)U^{-1} = f(UAU^{-1})$ for arbitrary operators A and invertible U. Remembering the form of L in (**??**), let us consider

$$UL^{*}U^{-1} = \exp\left(\frac{1}{2}U\left(\xi_j + i\theta_j\right)(\sigma^j)^{*}U^{-1}\right). \tag{10.7.263}$$

Suppose it were possible to find a change-of-basis such that L^* becomes L in eq. (**??**), that means we must have for a given j,

$$U \cdot \rho_j e^{-i\vartheta_j}(\sigma^j)^{*}U^{-1} = \rho_j e^{i\vartheta_j}\sigma^j, \tag{10.7.264}$$

$$\rho_j e^{i\vartheta_j} \equiv \xi_j - i\theta_j, \tag{10.7.265}$$

$$\rho_j = \sqrt{\xi_j^2 + \theta_j^2}, \qquad \tan\vartheta_j = -\frac{\theta_j}{\xi_j}. \tag{10.7.266}$$

Taking the determinant on both sides of the first line, for a fixed j,

$$\det\left[e^{-2i\vartheta_j}(\sigma^j)^*\right] = \det\left[\sigma^j\right] \tag{10.7.267}$$

$$e^{-4i\vartheta_j}\overline{\det\left[\sigma^j\right]} = -e^{-4i\vartheta_j} = \det\left[\sigma^j\right] = -1. \tag{10.7.268}$$

(We have used $\det\sigma^i = -1$.) The only situation L may be mapped to L^* and vice versa through a change-of-basis occurs when $\vartheta_j = 2\pi n/4 = \pi n/2$ for integer n. For even n, this corresponds to pure boosts, because

$$\xi_j - i\theta_j = \rho_j e^{i\frac{\pi}{2}n} = \pm\rho_j. \tag{10.7.269}$$

For odd n, this corresponds to pure rotations, because

$$\xi_j - i\theta_j = \rho_j e^{i\frac{\pi}{2}n} = \pm i\rho_j. \tag{10.7.270}$$

In fact, using $\epsilon(\sigma^i)^*\epsilon^\dagger = -\sigma^i$ in eq. (10.7.32),

$$\epsilon\cdot\overline{L[\vec{\xi}=0]}\cdot\epsilon^\dagger = \epsilon e^{+(i/2)\theta_j(\sigma^j)^*}\epsilon^\dagger = e^{+(i/2)\theta_j\epsilon(\sigma^j)^*\epsilon^\dagger} \tag{10.7.271}$$

$$= e^{-(i/2)\theta_j\sigma^j} = L[\vec{\xi}=0]. \tag{10.7.272}$$

However, as we shall show below, there is no transformation U that could bring a pure boost L^* back to L:

$$U\cdot L[\vec{\theta}=\vec{0}]^*\cdot U^{-1} = U\cdot e^{\frac{1}{2}\xi_j(\sigma^j)^*}\cdot U^{-1} \neq e^{\frac{1}{2}\xi_j\sigma^j} = L[\vec{\theta}=\vec{0}]. \tag{10.7.273}$$

To sum: only the complex conjugate of a pure rotation may be mapped into the same pure rotation.

Chiral Right, Chiral Left, and Vectors from $\mathbf{SL}_{2,\mathbb{C}}$ That L and L^* are generically inequivalent transformations is why the former corresponds to un-dotted indices and the latter to dotted ones in eq. (10.7.164) – the notation helps distinguishes between them. To understand this distinction further, let us recall the convention in eq. (10.7.3) for the generic Lorentz transformation, and write

$$L = \exp\left(-i\xi_j i\frac{\sigma^j}{2} - i\theta_j\frac{\sigma^j}{2}\right); \tag{10.7.274}$$

and by referring to generic Lorentz transformation in eq. (10.7.3), we may identify the boost and rotation generators as, respectively,

$$K_{\mathrm{R}}^i = i\frac{\sigma^i}{2} \qquad \text{and} \qquad J_{\mathrm{R}}^i = \frac{\sigma^i}{2}. \tag{10.7.275}$$

In this representation, therefore, the Lie algebra in equations (10.7.7) and (10.7.8) reads

$$J_+^i = \frac{1}{4}\left(\sigma^i + i^2\sigma^i\right) = 0 \tag{10.7.276}$$

$$J^i_- = \frac{1}{4}\left(\sigma^i - i^2\sigma^i\right) = \frac{\sigma^i}{2}. \tag{10.7.277}$$

Since the Pauli matrices have eigenvalues ± 1, the J^i_+ generators describe $j_+ = 0$; whereas the J^i_- ones spin $j_- = 1/2$. We therefore label this is as the $(j_+, j_-) = (0, 1/2)$ representation. It is dubbed by quantum field theory texts [35] as a 'chiral right spinor' transformation.

As for the L^*, we may express it as

$$L^* = \exp\left(-i\xi_j(-i)\frac{-(\sigma^j)^*}{2} - i\theta_j\frac{-(\sigma^j)^*}{2}\right) \tag{10.7.278}$$

and again referring to eq. (10.7.3),

$$K^i = -i\frac{-(\sigma^i)^*}{2} \qquad \text{and} \qquad J^i = \frac{-(\sigma^i)^*}{2}. \tag{10.7.279}$$

In this case, we may compute the Lie algebra in equations (10.7.7) and (10.7.8):

$$J^i_+ = \frac{1}{4}\left(-(\sigma^i)^* + i^2(\sigma^i)^*\right) = -\frac{(\sigma^i)^*}{2} \tag{10.7.280}$$

$$J^i_- = \frac{1}{4}\left(-(\sigma^i)^* - i^2(\sigma^i)^*\right) = 0. \tag{10.7.281}$$

This is the $(j_+, j_-) = (1/2, 0)$ representation; and is dubbed by quantum field theory texts [35] as a 'chiral left spinor' transformation. We may also recall eq. (10.7.31) and discover that eq. (10.7.279) is equivalent to

$$K^i = \epsilon^\dagger\left(-\frac{i}{2}\sigma^i\right)\epsilon \qquad \text{and} \qquad J^i = \epsilon^\dagger\left(\frac{1}{2}\sigma^i\right)\epsilon; \tag{10.7.282}$$

which in turn implies we must also obtain an equivalent $(j_+, j_-) = (1/2, 0)$ representation using

$$K^i_{\mathrm{L}} = -\frac{i}{2}\sigma^i \qquad \text{and} \qquad J^i_{\mathrm{L}} = \frac{1}{2}\sigma^i. \tag{10.7.283}$$

At this point, eq. (10.7.282) applied to eq. (10.7.278) hands us

$$L^* = \epsilon^\dagger \exp\left(-\frac{1}{2}\left(\vec{\xi} + i\vec{\theta}\right)\cdot\vec{\sigma}\right)\epsilon \tag{10.7.284}$$

$$= \epsilon^\dagger (L^\dagger)^{-1}\epsilon, \tag{10.7.285}$$

where the second equality follows from the hermicity of the $\{\sigma^i\}$ and the fact that $(\exp(q_i\sigma^i))^{-1} = \exp(-q_i\sigma^i)$. For later use, we employ the notation in eq. (10.7.26) and record here that eq. (10.7.283) may be obtained through

$$J^{\mu\nu}_{\mathrm{L}} \equiv \frac{i}{4}\sigma^{[\mu}\bar{\sigma}^{\nu]}, \tag{10.7.286}$$

$$J^{0i}_{\mathrm{L}} = \frac{i}{4}\left(\sigma^0(-)\sigma^i - \sigma^i\sigma^0\right) \tag{10.7.287}$$

$$= -\frac{i}{2}\sigma^i = K^i_{\mathrm{L}}; \tag{10.7.288}$$

$$J^{ab}_{\mathrm{L}} = \frac{i}{4}\left(\sigma^a(-)\sigma^b - \sigma^b(-)\sigma^a\right) \tag{10.7.289}$$

$$= -\frac{i}{4}[\sigma^a, \sigma^b] = \frac{1}{2}\epsilon^{abc}\sigma^c = \epsilon^{abc}J^c_{\mathrm{L}}. \tag{10.7.290}$$

This is consistent with equations (10.7.1) and (10.7.2). Similarly, eq. (10.7.279) may be obtained through

$$J^{\mu\nu}_{\mathrm{R}} \equiv \frac{i}{4}\bar{\sigma}^{[\mu}\sigma^{\nu]}, \tag{10.7.291}$$

$$J^{0i}_{\mathrm{R}} = \frac{i}{4}\left(\sigma^0\sigma^i - (-)\sigma^i\sigma^0\right) \tag{10.7.292}$$

$$= +\frac{i}{2}\sigma^i = K^i_{\mathrm{R}}; \tag{10.7.293}$$

$$J^{ab}_{\mathrm{R}} = \frac{i}{4}\left((-)\sigma^a\sigma^b - (-)\sigma^b\sigma^a\right) \tag{10.7.294}$$

$$= -\frac{i}{4}[\sigma^a, \sigma^b] = \frac{1}{2}\epsilon^{abc}\sigma^c = \epsilon^{abc}J^c_{\mathrm{R}}. \tag{10.7.295}$$

10.7.3 Majorana, Weyl and Dirac Equations

This section is the sequel to the previous one. Here, I will use the $\mathrm{SL}_{2,\mathbb{C}}$ ingredients introduced in §(10.7.2) to build the Majorana, Weyl, and Dirac equations and their solutions. These, in turn, describe the (semi-classical limits) of spin$-1/2$ fermions.

The primary guiding principle for building a relativistic equation is: it must take the same form in all inertial frames. For instance, if

$$\eta_{\alpha\beta}\mathrm{d}x^\alpha\mathrm{d}x^\beta = \eta_{\alpha\beta}\mathrm{d}x'^\alpha\mathrm{d}x'^\beta; \tag{10.7.296}$$

then we have seen that the homogeneous wave equation for a relativistic scalar is

$$\eta^{\mu\nu}\partial_{x^\mu}\partial_{x^\nu}\varphi(x) = 0 = \eta^{\mu\nu}\partial_{x'^\mu}\partial_{x'^\nu}\varphi(x'). \tag{10.7.297}$$

We now turn to constructing such relativistic wave equations for spin$-1/2$ systems. In momentum spacetime, we have found two inequivalent Lorentz transformations, the 'left-handed' $\lambda \to \lambda'(\Lambda \cdot p) = L\lambda(p)$ and the 'right-handed' $\rho(p) \to \rho'(\Lambda \cdot p) = (L^{-1})^\dagger\rho(p)$. Let $\Lambda_\alpha{}^\beta(\vec{\xi}, \vec{\theta})$ be the Lorentz transformation occurring in eq. (10.7.258). Then consider, for instance, 'right-handed' Lorentz transforming

$$(\bar{\sigma}^\nu p_\nu)\lambda(p) \to (L^{-1})^\dagger(\bar{\sigma}^\nu p_\nu)L^{-1}L\lambda(p) \tag{10.7.298}$$

$$= (\bar{\sigma}^\nu p'_\nu)\lambda'(p'), \tag{10.7.299}$$

where the Lorentz-transformed momentum is

$$p'_\alpha = \Lambda_\alpha{}^\beta(\vec{\xi}, \vec{\theta})p_\beta. \tag{10.7.300}$$

In index notation,

$$p^{\dot{A}B}\lambda_B \to ((L^\dagger)^{-1})^{\dot{A}}{}_{\dot{C}}p^{\dot{C}B}\lambda_B \tag{10.7.301}$$

$$= p'^{\dot{A}'B'}\lambda_{B'}; \tag{10.7.302}$$

where the primed indices denote, for e.g.,

$$\xi_{B'} \equiv L_B{}^C\xi_C \qquad \text{and} \qquad \xi^{C'} \equiv \xi^B(L^{-1})_B{}^C. \tag{10.7.303}$$

Problem 10.84. Explain why

$$(\sigma^\nu p_\nu)\rho(p) \to L(\sigma^\nu p_\nu)\rho(p) \tag{10.7.304}$$

becomes

$$L(\sigma^\nu p_\nu)\rho(p) = (\sigma^\nu p'_\nu)\rho'(p'), \tag{10.7.305}$$

$$\rho'(p') \equiv (L^\dagger)^{-1}\cdot\rho(p). \tag{10.7.306}$$

Also explain why these steps expressed in index notation read

$$p_{A\dot{B}}\rho^{\dot{B}} \to L_A{}^C p_{C\dot{B}}\rho^{\dot{B}} \tag{10.7.307}$$

$$= p'_{A'\dot{B}'}\rho'^{\dot{B}'}. \tag{10.7.308}$$

where p' and p are related by $p'_\alpha = \Lambda_\alpha{}^\beta(\vec{\xi},\vec{\theta})p_\beta$. □

Altogether, these considerations may be summarized as follows

> Applying $(\sigma\cdot p)$ to a right-handed $(0,1/2)$ spinor converts it into a left handed $(1/2,0)$ one. Whereas, applying $(\bar{\sigma}\cdot p)$ to a left-handed $(1/2,0)$ spinor converts it into a right-handed $(0,1/2)$ one.

From the four types of objects we have studied thus far, built out of an arbitrary spinor ζ, namely

$$p^{\dot{A}B}\zeta_B, \qquad p_{A\dot{B}}\zeta^{\dot{B}}, \qquad \zeta^{\dot{A}}, \qquad \text{and} \qquad \zeta_A, \tag{10.7.309}$$

we may now proceed to form Lorentz covariant equations involving them. (Since upper and lower indices are related via contraction with ϵ, we do not consider them distinct 'types' for the discussion at hand.) The terms involving momentum p are needed if we wish to form *differential* equations in position spacetime.

Left and Right Handed Spinors in Spacetime Before we proceeding, let us first answer the following question. If we view $\lambda'(p'_\alpha = \Lambda_\alpha{}^\beta p_\beta) \equiv L\lambda(p)$ as the left handed spinor in the $p'-$frame, what is the corresponding transformation in position spacetime? Starting from the integral of all such momentum-space left-handed spinors within the $\{x^\mu\}$ frame,

$$\psi_L(x) \equiv \int\frac{\mathrm{d}^4p}{(2\pi)^4}\lambda(p)e^{-ip_\alpha x^\alpha}, \tag{10.7.310}$$

we now consider applying L from the left, to discover it generates the position spacetime left handed Weyl spinor in the new $\{x'^\mu \equiv \Lambda^\mu{}_\nu x^\nu\}$ frame:

$$L\psi_L(x) = \int \frac{\mathrm{d}^4 p'}{(2\pi)^4} \lambda'(p'_\alpha = \Lambda_\alpha{}^\beta p_\beta) e^{-ip'_\mu \Lambda^\mu{}_\nu x^\nu} \tag{10.7.311}$$

$$\equiv \psi'_L \left(x'^\alpha = \Lambda^\alpha{}_\beta x^\beta\right). \tag{10.7.312}$$

In other words, the left handed spinor in the $x'-$frame is related to that in the $x-$frame through

$$\psi'_L(x') = L\psi_L\left(x^\alpha = x'^\beta \Lambda_\beta{}^\alpha\right). \tag{10.7.313}$$

Similarly, we may deduce the right handed spinor in the $x'-$frame is related to that in the $x-$frame through

$$\psi'_R(x') = (L^{-1})^\dagger \psi_R\left(x^\alpha = x'^\beta \Lambda_\beta{}^\alpha\right). \tag{10.7.314}$$

Incidentally, this also means

$$\partial_{x^\alpha} = \Lambda^\beta{}_\alpha \partial_{x'^\beta}. \tag{10.7.315}$$

Majorana Equation To involve one and only one spinor, we may now form the *Majorana* equation (in momentum space):

$$p^{\dot{\mathrm{A}}\mathrm{B}} \lambda_{\mathrm{B}} = m \cdot \lambda^{\dot{\mathrm{A}}}, \tag{10.7.316}$$

where m carry physical dimension of mass because p does. In matrix notation,

$$(\bar{\sigma} \cdot p)\lambda = m \cdot \epsilon^\dagger \cdot \lambda^*. \tag{10.7.317}$$

We may also form

$$p_{\mathrm{A}\dot{\mathrm{B}}} \rho^{\dot{\mathrm{B}}} = m \cdot \rho_{\mathrm{A}}, \tag{10.7.318}$$

whose matrix form is

$$(\sigma \cdot p)\rho = m \cdot \epsilon \cdot \rho^*. \tag{10.7.319}$$

Here and below, we will first derive the spinor wave equations in momentum space; then transform to position spacetime as follows. Multiply all spinors in momentum space with the basis spacetime plane wave $e^{-ip\cdot x} \equiv e^{-ip_\alpha x^\alpha}$. For the Majorana equations at hand, this translates to

$$(\bar{\sigma} \cdot p)(\zeta e^{-ip\cdot x}) = i\bar{\sigma}^\mu \partial_\mu (\zeta e^{-ip\cdot x}) = 0 \tag{10.7.320}$$

and

$$(\sigma \cdot p)(\zeta e^{-ip\cdot x}) = i\sigma^\mu \partial_\mu (\zeta e^{-ip\cdot x}) = 0. \tag{10.7.321}$$

Since we are dealing with linear equations, $\zeta e^{-ip\cdot x}$ may be superposed over all momentum modes to obtain the general position spacetime equations. This amounts to replacing $\zeta(p)e^{-ip\cdot x} \to \psi(x) \equiv \psi(t, \vec{x})$:

$$i\partial^{\dot{A}B}\psi_B(x) = m \cdot \psi^{\dot{A}}(x), \tag{10.7.322}$$

$$\text{or} \qquad (i\bar{\sigma} \cdot \partial)\psi(x) = m \cdot \epsilon^\dagger \cdot \psi(x)^*; \tag{10.7.323}$$

and

$$i\partial_{A\dot{B}}\psi^{\dot{B}}(x) = m \cdot \psi_A(x), \tag{10.7.324}$$

$$\text{or} \qquad (i\sigma \cdot \partial)\psi(x) = m \cdot \epsilon \cdot \psi^*(x). \tag{10.7.325}$$

Problem 10.85. Majorana Dispersion Relations Show that both equations (10.7.316) and (10.7.318) imply

$$p^2 \equiv p_\alpha p^\alpha = m^2; \tag{10.7.326}$$

i.e., they do indeed obey the dispersion relations of a massive particle. Hint: Apply $(\sigma \cdot p)$ to both sides of eq. (10.7.317); and $(\bar{\sigma} \cdot p)$ to both sides of eq. (10.7.319). □

Problem 10.86. Charge Conjugation Show that eq. (10.7.316) can be converted into eq. (10.7.318) and vice versa – i.e., the solution to one can be used to obtain the solution to the other. In this sense, the two forms of Majorana equations are not independent.

Hint: Start with either eq. (10.7.316) or (10.7.318) and take its complex conjugate. □

Weyl Equations The massless limits ($m = 0$) of the Majorana equations immediately yield the *Weyl* equations:

$$p^{\dot{A}B}\zeta_B = 0 = (\bar{\sigma} \cdot p)\zeta \tag{10.7.327}$$

and

$$p_{A\dot{B}}\zeta^{\dot{B}} = 0 = (\sigma \cdot p)\zeta. \tag{10.7.328}$$

The position spacetime versions are

$$i\partial^{\dot{A}B}\psi_B(x) = 0 = (i\bar{\sigma} \cdot \partial)\psi(x) \tag{10.7.329}$$

and

$$i\partial_{A\dot{B}}\psi^{\dot{B}}(x) = 0 = (i\sigma \cdot \partial)\psi(x). \tag{10.7.330}$$

Problem 10.87. Solutions to Weyl Equation In this problem, we shall obtain the solutions to the Weyl equation.

First, show that the appropriate dispersion relation is the zero mass one:

$$p^2 \equiv p_\alpha p^\alpha = 0. \tag{10.7.331}$$

This implies $p_0 = \pm|\vec{p}| \equiv \pm p$.

Method I Explain why, for massless dispersion relations; and for positive energies $p_0 = |\vec{p}| \equiv p$, the momentum in written spinor basis takes a factorized form:

$$\sigma \cdot p = 2p \cdot \zeta^+ (\zeta^+)^\dagger, \tag{10.7.332}$$

$$p_{\mathrm{A\dot{B}}} = 2p \cdot \zeta^+{}_{\mathrm{A}} \overline{\zeta^+}{}_{\mathrm{\dot{B}}}, \tag{10.7.333}$$

where $(p_i \sigma^i)\zeta^\pm = \pm|\vec{p}|\zeta^\pm$. Whereas for negative energies, $p_0 = -|\vec{p}| \equiv -p$, explain why the factorized form is:

$$\sigma \cdot p = -2p \cdot \zeta^- (\zeta^-)^\dagger, \tag{10.7.334}$$

$$p_{\mathrm{A\dot{B}}} = -2p \cdot \zeta^-{}_{\mathrm{A}} \overline{\zeta^-}{}_{\mathrm{\dot{B}}}. \tag{10.7.335}$$

Since eigenstates of a Hermitian operator corresponding to distinct eigenvalues are necessarily orthogonal, $(\zeta^+)^\dagger \zeta^- = 0 = (\zeta^-)^\dagger \zeta^+$, explain why the (unit-norm) positive and negative energy solutions to eq. (10.7.328) are, respectively, the negative and positive helicity states:

$$\zeta(p_0 = \pm p) = \zeta^\mp. \tag{10.7.336}$$

Alternatively, you may also invoke eq. (10.7.65) to make the argument using index notation.

Method II Next, suppose the momentum p_i were pointing strictly along the 3−axis, so that

$$\sigma \cdot p = \pm p_3 \cdot \sigma^0 + p_3 \cdot \sigma^3. \tag{10.7.337}$$

Choose a basis such that σ^3 is diagonal, and show that, for $p_0 = \pm p \equiv \pm|p_3|$:

$$\zeta^- = (0,1)^{\mathrm{T}}, \tag{10.7.338}$$

$$\zeta^+ = (1,0)^{\mathrm{T}}. \tag{10.7.339}$$

This shows that helicity-down and helicity-up states correspond to positive and energy solutions. Moreover, there is no way a positive energy solution to be helicity-up; nor a negative energy solution to be helicity-down – heuristically speaking, one cannot outrun a massless particle, to flip the sign of momentum in $p_i \sigma^i$.

Now, rotate these $\zeta^\pm(p_i = p_3 \cdot \delta_i^3)$ to the generic $\zeta^\mp(\vec{p})$; see Problem (10.75). □

Dirac Equation and Parity If we now instead allow two distinct spinors λ and ρ, the Majorana equations may be promoted into the following pair:

$$p^{\mathrm{\dot{A}B}} \lambda_{\mathrm{B}} = m_1 \cdot \rho^{\mathrm{\dot{A}}}, \tag{10.7.340}$$

$$p_{\mathrm{A\dot{B}}} \rho^{\mathrm{\dot{B}}} = m_2 \cdot \lambda_{\mathrm{A}}, \tag{10.7.341}$$

where the masses $m_{1,2}$ may in principle be distinct ($m_1 \neq m_2$). However, if we now apply the parity operation, swapping $\rho_{\mathrm{A}} \leftrightarrow \lambda^{\mathrm{\dot{A}}}$ and $\vec{p} \leftrightarrow -\vec{p}$, we discover equations (10.7.340) and (10.7.341) becomes transformed into

$$p_{\mathrm{A\dot{B}}} \rho^{\mathrm{\dot{B}}} = m_1 \cdot \lambda_{\mathrm{A}}, \tag{10.7.342}$$

$$p^{\dot{\mathrm{A}}\mathrm{B}}\lambda_{\mathrm{B}} = m_2 \cdot \rho^{\dot{\mathrm{A}}}. \tag{10.7.343}$$

This tells us, if we also wish to produce a pair of parity invariant equations, then $m_1 = m_2$. Upon doing so, $m_1 = m_2 \equiv m$, we arrive at the Lorentz-covariant and parity invariant version of equations (10.7.340) and (10.7.341):

$$p^{\dot{\mathrm{A}}\mathrm{B}}\lambda_{\mathrm{B}} = m \cdot \rho^{\dot{\mathrm{A}}}, \tag{10.7.344}$$

$$p_{\mathrm{A}\dot{\mathrm{B}}}\rho^{\dot{\mathrm{B}}} = m \cdot \lambda_{\mathrm{A}}. \tag{10.7.345}$$

The matrix form of Dirac's equations is

$$(\bar{\sigma} \cdot p)\lambda(p) = m \cdot \rho(p) \qquad \text{and} \qquad (\sigma \cdot p)\rho(p) = m \cdot \lambda(p). \tag{10.7.346}$$

If we put $\widetilde{\psi}(p)e^{-ip\cdot x} = (\lambda(p), \rho(p))^{\mathrm{T}} e^{-ip\cdot x}$ and integrate over these arbitrary momentum to obtain the position spacetime expression, we would obtain the Dirac equation – first derived by the theoretical physicist P.A.M. Dirac via different arguments –

$$(i\bar{\sigma} \cdot \partial)\psi_L(x) = m \cdot \psi_R(x) \qquad \text{and} \qquad (i\sigma \cdot \partial)\psi_R(x) = m \cdot \psi_L(x); \tag{10.7.347}$$

where

$$\psi_L(x) = \int \frac{\mathrm{d}^4 p}{(2\pi)^4}\lambda(p)e^{-ip\cdot x}, \tag{10.7.348}$$

$$\psi_R(x) = \int \frac{\mathrm{d}^4 p}{(2\pi)^4}\rho(p)e^{-ip\cdot x}. \tag{10.7.349}$$

Dirac Matrices Define the 4×4 Dirac matrices $\{\gamma^\mu | \mu = 0, 1, 2, 3\}$, in what is known as the chiral basis, as follows.

$$\gamma^\mu \equiv \begin{bmatrix} 0_{2\times 2} & \sigma^\mu \\ \bar{\sigma}^\mu & 0_{2\times 2} \end{bmatrix} \tag{10.7.350}$$

Then, upon defining the 4−component Dirac spinor

$$\psi(x) \equiv (\psi_L(x), \psi_R(x))^{\mathrm{T}}, \tag{10.7.351}$$

eq. (10.7.347) may be re-packaged into a form closer to what Dirac himself wrote down:

$$0 = (i\gamma^\mu\partial_\mu - m)\,\psi \tag{10.7.352}$$

$$= \begin{bmatrix} -m & i\sigma^\mu\partial_\mu \\ i\bar{\sigma}^\mu\partial_\mu & -m \end{bmatrix} \begin{bmatrix} \psi_L(x) \\ \psi_R(x) \end{bmatrix}. \tag{10.7.353}$$

Moreover, these Dirac matrices satisfy the anti-commutation relations (aka *Clifford Algebra*)

$$\{\gamma^\mu, \gamma^\nu\} \equiv \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}\mathbb{I}. \tag{10.7.354}$$

This property may be exploited to show that the Dirac solution obeys the massive wave equation, by applying $-i\gamma^\nu\partial_\nu - m$:

$$0 = (-i\gamma^\nu\partial_\nu - m)(i\gamma^\mu\partial_\mu - m)\psi \tag{10.7.355}$$

$$= (\gamma^\nu \gamma^\mu \partial_\nu \partial_\mu + m^2)\psi \tag{10.7.356}$$

$$= \left(\frac{1}{2}\{\gamma^\nu, \gamma^\mu\}\partial_\nu \partial_\mu + m^2\right)\psi = (\partial^2 + m^2)\psi, \tag{10.7.357}$$

where in the third equality, the product of the Dirac matrices can be symmetrized because the derivatives commute.

Problem 10.88. Generators of Lorentz Verify the anti-commutation relations in eq. (10.7.354). Then, verify that the 'left' and 'right' handed generators in eq. (10.7.130) may be recovered via

$$J_D^{\mu\nu} \equiv \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right] = \begin{bmatrix} J_{\mathrm{L}}^{\mu\nu} & 0 \\ 0 & J_{\mathrm{R}}^{\mu\nu} \end{bmatrix}. \tag{10.7.358}$$

If $\psi'(x')$ is the Dirac spinor in the $x'-$ frame and $\psi(x)$ is that in the $x-$frame, as related by $x'^\alpha = \Lambda^\alpha{}_\beta x^\beta$, explain why the Lorentz transformations of the position spacetime left and right handed spinors – see equations (10.7.313) and (10.7.314) – is equivalent to following Lorentz transformation of the Dirac spinor:

$$\psi'(x') = \exp\left(-\frac{i}{2}\omega_{\alpha\beta}J_D^{\alpha\beta}\right)\psi(x^\alpha = x'^\beta \Lambda_\beta{}^\alpha). \tag{10.7.359}$$

and

$$\bar{\psi}'(x') \equiv \psi'^\dagger(x')\gamma^0 = \bar{\psi}\left(x^\alpha = x'^\beta \Lambda_\beta{}^\alpha\right)\exp\left(\frac{i}{2}\omega_{\alpha\beta}J_D^{\alpha\beta}\right). \tag{10.7.360}$$

Use these transformation rules to explain why sandwiching Dirac matrices in between $\bar{\psi}$ and ψ – for instance,

$$\bar{\psi}\gamma^\mu\psi, \qquad \bar{\psi}[\gamma^\mu, \gamma^\nu]\psi, \qquad \text{etc.} \tag{10.7.361}$$

– yield (Minkowski) spacetime tensors. By writing $\bar{\psi}\gamma^\mu\psi$ in terms of its left and right handed 2$-$component spinors, compare the discussion here to that in Problem (10.81).

Finally, we record here (without proof) that in *any* spacetime dimension $D + 1 \geq 2$, if matrices $\{\gamma^\mu\}$ may be found satisfying eq. (10.7.354), then $J_D^{\mu\nu} \equiv \frac{i}{4}\left[\gamma^\mu, \gamma^\nu\right]$ would satisfy the $\mathrm{SO}_{D,1}$ Lie Algebra. This, in turn, implies the $D + 1$ dimensional Minkowski spacetime Dirac equation $(i\gamma^\mu\partial_\mu - m)\psi = 0$ remains relativistically covariant. $\square$

Problem 10.89. Dirac Equation: Solutions Let us first consider the solutions 'at rest':

$$q_\mu = \pm\left(m, \vec{0}\right). \tag{10.7.362}$$

Show that

$$\lambda = \pm\rho \equiv \zeta, \tag{10.7.363}$$

or equivalently,

$$\widetilde{\psi} = \left(\zeta, \pm\zeta\right)^{\mathrm{T}}, \tag{10.7.364}$$

for arbitrary spinor ζ, is the general solution for the Dirac equation 'at rest'. Next, we shall move on to solve the positive (+) and negative (−) solutions

$$\widetilde{\psi}^{\pm}(p)e^{\mp ip_\alpha x^\alpha} = (\lambda(p), \rho(p))^{\mathrm{T}} \exp\left(\mp i(\sqrt{p^2+m^2}t + p_j x^j)\right), \tag{10.7.365}$$

$$p_\mu \equiv (\sqrt{p^2+m^2}, p_i). \tag{10.7.366}$$

Now, the boosted version of eq. (10.7.346) takes the covariant form

$$(L^{-1})^\dagger(\bar{\sigma}\cdot q)(L^{-1})L\lambda(q) = m(L^{-1})^\dagger\rho(q) \tag{10.7.367}$$

$$L(\sigma\cdot q)L^\dagger(L^\dagger)^{-1}\rho(q) = mL\lambda(q). \tag{10.7.368}$$

You should find, for arbitrary spinor ζ, the general solution for positive (+) and negative (−) energies:

$$\widetilde{\psi}^{\pm}(p) = \left(\sqrt{\sigma\cdot p}\cdot\zeta, \pm\sqrt{\bar{\sigma}\cdot p}\cdot\zeta\right)^{\mathrm{T}}, \tag{10.7.369}$$

where, if $\xi^\pm$ obeys $(p_i\sigma^i)\xi^\pm = \pm p\cdot\xi^\pm$ for $p \equiv |\vec{p}|$,

$$\sqrt{\sigma\cdot p} = \sqrt{\sqrt{p^2+m^2}+p}\cdot\xi^+(\xi^+)^\dagger + \sqrt{\sqrt{p^2+m^2}-p}\cdot\xi^-(\xi^-)^\dagger, \tag{10.7.370}$$

$$\sqrt{\bar{\sigma}\cdot p} = \sqrt{\sqrt{p^2+m^2}-p}\cdot\xi^+(\xi^+)^\dagger + \sqrt{\sqrt{p^2+m^2}+p}\cdot\xi^-(\xi^-)^\dagger. \tag{10.7.371}$$

The arbitrary position spacetime solution therefore involves a sum over the basis spinors $\xi^1 \equiv (1,0)^{\mathrm{T}}$ and $\xi^2 \equiv (0,1)^{\mathrm{T}}$ as well as an integral over all wavelengths:

$$\psi(x) = \int \frac{\mathrm{d}^3\vec{p}}{(2\pi)^3}\frac{1}{\sqrt{2p_0}}\sum_{s\in\{1,2\}}\left(a^s_{\vec{p}}u^s_{\vec{p}}e^{-ip\cdot x} + b^s_{\vec{p}}v^s_{\vec{p}}e^{+ip\cdot x}\right); \tag{10.7.372}$$

$$u^s_{\vec{p}} \equiv \left(\sqrt{\sigma\cdot p}\xi^s, \sqrt{\bar{\sigma}\cdot p}\xi^s\right)^{\mathrm{T}}, \tag{10.7.373}$$

$$v^s_{\vec{p}} \equiv \left(\sqrt{\sigma\cdot p}\xi^s, -\sqrt{\bar{\sigma}\cdot p}\xi^s\right)^{\mathrm{T}}; \tag{10.7.374}$$

for appropriate complex coefficients $\{a^s_{\vec{p}}, b^s_{\vec{p}}\}$.

Hint: Problem (10.76) should be useful. □

Problem 10.90. Scalars, Vectors, and Tensors from Spinors If λ and ρ are, respectively, left and right handed spinors, explain why

$$\lambda^\dagger\rho \qquad \text{and} \qquad \rho^\dagger\lambda \tag{10.7.375}$$

are spacetime scalars while $\lambda^\dagger\lambda$ and $\rho^\dagger\rho$ are *not*. Furthermore, explain why

$$\lambda^\dagger\bar{\sigma}^\mu\lambda \qquad \text{and} \qquad \rho^\dagger\sigma^\mu\rho \tag{10.7.376}$$

transform as spacetime vectors while $\rho^\dagger\bar{\sigma}^\mu\rho$, $\rho^\dagger\bar{\sigma}^\mu\lambda$, $\lambda^\dagger\sigma^\mu\lambda$, and $\lambda^\dagger\sigma^\mu\rho$ are not. □

Vectors

Parity and Clifford Algebra

Charge Conjugation

Time Reversal

10.7.4 Poincaré: Lorentz and Space-Time Translations

Let us begin by recording the commutation relations between momentum P^μ and the boost and angular momentum operators.

$$\left[K^i, P^0\right] = -iP^i, \qquad \left[K^i, P^j\right] = -i\delta^{ij}P^0, \tag{10.7.377}$$

$$\left[J^a, P^0\right] = 0, \qquad \left[J^c, P^i\right] = i\epsilon^{cib}P^b. \tag{10.7.378}$$

Equations (10.7.377) and (10.7.378) together with equations (10.7.4), (10.7.5), and (10.7.6) altogether form the Lie algebra of the (3+1)D Poincaré group.

Problem 10.91. Verify (10.7.377) and (10.7.378) from the Lie algebra $[J^{\alpha\beta}, P^\mu] = -i\eta^{\mu[\alpha}P^{\beta]}$. □

Casimir Operators Just like for the rotation group, we now search for a maximal set of mutually commuting Hermitian operators. We first seek operators that commutes with *all* the generators of the Poincaré group $\{J^{\alpha\beta}, P^\mu\}$. This is similar to seeking the analog of $(1/2)J^{ab}J^{ab}$ of the rotation group. To form such operators, they must be Lorentz scalars, since any Lorentz tensor operator *would not* commute with $J^{\alpha\beta}$. We may consider contracting the generators, where the non-zero quadratic ones are

$$J^2 \equiv \frac{1}{2}J^{\alpha\beta}J_{\alpha\beta} \qquad \text{and} \qquad P^2 \equiv P_\alpha P^\alpha; \tag{10.7.379}$$

since $J^{\alpha\beta}P_\alpha P_\beta = 0$ by anti-symmetry. However, while J^2 and P^2 are both Lorentz scalars; only P^2 is a Poincaré scalar.

Problem 10.92. Verify

$$[J^2, P^\mu] = -i\left\{J^\mu{}_\nu, P^\nu\right\} \neq 0, \tag{10.7.380}$$

where $\{\cdot,\cdot\}$ is the anti-commutator. □

To obtain a cubic spacetime scalar, we would need either zero or two P's since we need to contract even number of indices. But we cannot have, correspondingly, three J's or a single J due to its anti-symmetry. To form a quartic scalar operator, we first introduce the *Pauli-Lubanski vector*,

$$W^\mu \equiv \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}P_\nu J_{\alpha\beta} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}J_{\alpha\beta}P_\nu, \tag{10.7.381}$$

where the order of P and J is immaterial due to eq. (10.6.15) and the anti-symmetric character of ϵ. This vector commutes with momentum:

$$\left[W^\mu, P^\lambda\right] = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}\left[J_{\alpha\beta}, P^\lambda\right]P_\nu \tag{10.7.382}$$

$$= \epsilon^{\mu\nu\alpha\beta}(-i)\delta^\lambda{}_\alpha P_\beta P_\nu = 0, \tag{10.7.383}$$

by the anti-symmetry of the Levi-Civita. On the other hand, since W^μ is a vector,

$$\left[J^{\alpha\beta}, W^\mu\right] = -i\eta^{\mu[\alpha}W^{\beta]}. \tag{10.7.384}$$

From these two facts, we gather that

$$W^2 \equiv W^\mu W_\mu \tag{10.7.385}$$

is therefore a Poincaré scalar.

Problem 10.93. Verify that

$$W^2 = J_{\alpha\sigma} J_\beta{}^\sigma P^\alpha P^\beta - J^2 P^2 \tag{10.7.386}$$

and that it indeed commutes with the generators of the Poincaré group: $[W^2, J^{\alpha\beta}] = 0 = [W^2, P^\mu]$. Be sure to explain why the order of the operators in eq. (10.7.386) does not matter. □

Problem 10.94. Space-Time Decomposition Verify that

$$W^0 = P^i J^i \equiv \vec{P} \cdot \vec{J} = \vec{J} \cdot \vec{P}, \tag{10.7.387}$$

$$W^i = P^0 J^i + (\vec{K} \times \vec{P})^i = P^0 J^i - (\vec{P} \times \vec{K})^i, \tag{10.7.388}$$

provided we define the cross product in the usual manner, $(\vec{K} \times \vec{P})^i = \epsilon^{ijk} K^j P^k$.

Physically speaking, the angular momentum J_{ab} occurring within W^0 must be the 'intrinsic' one, not the orbital $L^{ij} = X^i P_j - X^j P_i$ one, since this latter operator will cancel out due to the anti-symmetric character of the Levi-Civita symbol: $\epsilon^{0ijk} L_{ij} P_k = 0$. □

Since P_μ commutes with itself and with $\{P^2, W^2\}$, we now seek out the simultaneous eigenstates of the set $\{P_\mu, P^2, W^2\}$. (We cannot do so for $\{J^{\alpha\beta}, P^2, W^2\}$ since $J^{\alpha\beta}$ does not commute with itself.) This procedure is analogous to the discussion for the rotation group, where we sought out the simultaneous eigenstates of $\{\vec{J}^2, J^3\}$.

Mass If $|k\rangle$ were the eigenstate of the momentum operator P_μ, obeying

$$P_\mu |k\rangle = k_\mu |k\rangle, \tag{10.7.389}$$

then for timelike k_μ we may associate the eigenvalue of P^2 with mass-squared; namely,

$$P^2 |k\rangle = k^\mu k_\mu |k\rangle = M^2 |k\rangle. \tag{10.7.390}$$

For null momenta k_μ, we have

$$P^2 |k\rangle = 0; \tag{10.7.391}$$

and for spacelike k_μ,

$$P^2 |k\rangle = -M^2 |k\rangle, \qquad M \in \mathbb{R}. \tag{10.7.392}$$

In physical applications, and for the timelike and null cases, the zeroth component P^0 is the Hamiltonian, which governs the time evolution of the physical system at hand.

Note that, due to the relativistic invariance of $k^2 \equiv k_\mu k^\mu$ providing a constraint between the zeroth and spatial components of momentum, it is not legitimate to normalize the states as

$$\langle k | k' \rangle = \delta^{(4)}(k - k'), \tag{10.7.393}$$

since, the $\delta-$functions imposing $\vec{k} = \vec{k}'$ would immediately render the $\delta-$function for the zeroth component to be singular: $\delta(k^0 - k'^0) = \delta(0)$. Instead, we impose

$$\langle k | \, k' \rangle = 2E_{\vec{k}} \delta^{(3)}(\vec{k} - \vec{k}'), \tag{10.7.394}$$

where $E_{\vec{k}} \equiv \sqrt{\vec{k}^2 + m^2}$, at least for timelike or null k_μ. **YZ: Explain.**

Relativistic Spin: Massive Case For timelike k_μ, we may now choose to work in its 'rest frame' where

$$\Lambda^\mu{}_\nu k^\nu \equiv k'^\mu = (M, \vec{0}). \tag{10.7.395}$$

(Any negative signs in front of M can be removed by the time-reversal operation.) More specifically, we shall assume it is possible to find a $D(\Lambda) \equiv D(\Lambda, 0)$ such that (up to a phase factor)

$$D(\Lambda) \, |k\rangle = |k'\rangle \, . \tag{10.7.396}$$

We may thus consider

$$W^2 \, |k\rangle = D(\Lambda)^{-1} D(\Lambda) W^2 D(\Lambda)^{-1} D(\Lambda) \, |k\rangle \tag{10.7.397}$$

$$= e^{i\delta_{\mathrm{M}}} D(\Lambda)^{-1} W^2 \, |k'\rangle = e^{i\delta_{\mathrm{M}}} D(\Lambda^{-1}) W^2 \, |k'\rangle \, , \tag{10.7.398}$$

where we have exploited the Lorentz invariance of W^2 in the second equality. Since this forms holds for *any* timelike k, we see that diagonalizing W^2 can always be carried out in the rest frame: $W^2 \, |k'\rangle = w^2 \, |k'\rangle$ means $W^2 \, |k\rangle = w^2 \, |k\rangle$. Since the order of the operators J and P do not matter in W^2, we first examine

$$W^\mu \, |k'\rangle = \frac{1}{2} \epsilon^{\mu 0 \alpha\beta} J_{\alpha\beta} M \, |k'\rangle \tag{10.7.399}$$

$$= M \delta^\mu_\ell J^\ell \, |k'\rangle \, , \tag{10.7.400}$$

with $J^\ell \equiv (1/2) \epsilon^{\ell mn} J^{mn}$ and $\epsilon^{123} \equiv 1$. When acting on $|k'\rangle$, we may therefore identify W^2 with the square of the angular momentum operator from the rotation group:

$$W^2 \, |k'\rangle = -M^2 \vec{J}^2 \, |k'\rangle \, . \tag{10.7.401}$$

According to Problem (10.36), the most general Lorentz transformation that leaves $k' = (M, \vec{0})$ invariant is the group of spatial rotations. That is, for any $\widehat{R}^{\mathrm{T}} \widehat{R} = \mathbb{I}$, we must have (up to a phase factor)

$$D(\widehat{R}) \, |k'\rangle \propto |k'\rangle \, ; \tag{10.7.402}$$

namely, the rotation generators $J^a = (1/2) \epsilon^{abc} J^{bc}$ when acting on $|k'\rangle$ cannot alter k', though it can mix other degrees of freedom within the state. Even though the generators of rotation do not commute amongst themselves, within this $k'-$frame, we may thus simultaneously diagonalize $\{W^2 \propto \vec{J}^2, J^3\}$ as was done for the rotation group. The eigenstate of W^2 with momentum k' is thus labeled by (ℓ, m), where $\ell \in \{0, 1/2, 1, 3/2, 2, \dots\}$ and $-\ell \leq m \leq \ell$ in integer steps – namely, a massive spin ℓ state has $2\ell + 1$ degrees of freedom.

$$W^2 \, |k', \ell, m\rangle = -M^2 \ell(\ell+1) \, |k', \ell, m\rangle \tag{10.7.403}$$

$$J^3 |k', \ell, m\rangle = m |k', \ell, m\rangle . \tag{10.7.404}$$

The general eigenstate of W^2 and P^2 for the massive case is therefore

$$P^2 |k, M, \ell\rangle = M^2 |k, M, \ell\rangle , \tag{10.7.405}$$
$$W^2 |k, M, \ell\rangle = -M^2 \ell(\ell+1) |k, M, \ell\rangle ; \tag{10.7.406}$$

where

$$\left|k^\mu = k'^\nu \Lambda_\nu{}^\mu \equiv (\Lambda^{-1}k')^\mu, M, \ell\right\rangle = D(\Lambda^{-1}) \sum_{-\ell \leq m \leq m} \chi_m \left|k' = (M, \vec{0}), \ell, m\right\rangle , \tag{10.7.407}$$

$$\sum_{-\ell \leq m \leq \ell} |\chi_m|^2 = 1. \tag{10.7.408}$$

Relativistic Spin: Null Case If k^μ were null, we may choose a coordinate system such that its spatial component lie along the $1-$axis:

$$\Lambda^\mu{}_\nu k^\nu \equiv k'^\mu = \omega(1, 1, 0, 0). \tag{10.7.409}$$

(Any $-$ signs in the $1-$component may be removed by a parity flip operation.) We shall assume there is a $D(\Lambda) \equiv D(\Lambda, 0)$ such that (up to a phase factor) $D(\Lambda) |k\rangle = |k'\rangle$ and, like the timelike case, consider

$$W^2 |k\rangle = D(\Lambda^{-1}) W^2 |k'\rangle . \tag{10.7.410}$$

Since this relation holds for all null k's, we see that diagonalizing W^2 can always be done in the $k'-$frame: $W^2 |k'\rangle = w^2 |k'\rangle$ means $W^2 |k\rangle = w^2 |k\rangle$. Using eq. (10.7.386),

$$\begin{aligned} W^2 |k'\rangle &= \omega^2 \left(J_{0i} J^{0i} - J_{1\sigma} J^{1\sigma} - J_{0i} J^{1i} + J_{1i} J^{0i}\right) |k'\rangle & (10.7.411) \\ &= \omega^2 \left(-J^{0i_\perp} J^{0i_\perp} - J^{1i_\perp} J^{1i_\perp} + J^{0i_\perp} J^{1i_\perp} + J^{1i_\perp} J^{0i_\perp}\right) |k'\rangle & (10.7.412) \\ &= -\omega^2 M^{i_\perp} M^{i_\perp} |k'\rangle , & (10.7.413) \end{aligned}$$

where $i_\perp$ runs over $\{2, 3\}$ and

$$M^{i_\perp} \equiv J^{0i_\perp} - J^{1i_\perp}. \tag{10.7.414}$$

From Problem (10.37), we may recognize the operator version of eq. (10.3.189):

$$\exp(-i\xi_{i_\perp} M^{i_\perp} - i\phi J^{23}) |k'\rangle \propto |k'\rangle ; \tag{10.7.415}$$

i.e., neither $M^{i_\perp}$ nor J^{23} may change k' when acting upon $|k'\rangle$, though they may mix other degrees of freedom occurring within the state. Within this $k'-$frame, we may thus simultaneously diagonalize $\{W^2 \propto M^{i_\perp} M^{i_\perp}, J^{23}\}$.

Problem 10.95. 2D Euclidean Group Prove that the three operators $\{M^{i_\perp}, J^{23}\}$ obey the Lie algebra of the 2D Euclidean group in eq. (5.4.97), provided we identify

$$M^{i_\perp} \leftrightarrow P_a[(5.4.97)] \qquad \text{and} \qquad J^{23} \leftrightarrow J[(5.4.97)]. \tag{10.7.416}$$

That is, you should find, with $\epsilon^{23} \equiv 1$,

$$\left[J^{23}, M^{i_\perp}\right] = i\epsilon^{i_\perp j_\perp} M^{j_\perp} \qquad \text{and} \qquad \left[M^{i_\perp}, M^{j_\perp}\right] = 0. \tag{10.7.417}$$

□

From the study of the 2D Euclidean group, we recall that the compatible observables are $M^2 \equiv M^{i_\perp} M^{i_\perp}$, whose eigenvalue $\lambda^2 \geq 0$ is a non-negative real number; and J^{23}, whose eigenvalue is an integer $n \in \mathbb{Z}$ for periodic boundary conditions and odd multiple of $1/2$, $n \in \{\pm 1/2, \pm 3/2, \pm 5/2, \dots\}$, for anti-periodic boundary conditions. We may interpret $J^{23} = J^1 = \widehat{k} \cdot \vec{J}$ as the helicity operator – spin projected along spatial momentum; and this n denotes its eigenvalue. On the other hand, λ^2 corresponds to 'continuous spin'; though as of this writing there is no evidence for its existence in Nature.

$$J^{23} \left|k', n, \lambda\right\rangle = n \left|k', n, \lambda\right\rangle , \tag{10.7.418}$$

$$M^{i_\perp} M^{i_\perp} \left|k', n, \lambda\right\rangle = \lambda^2 \left|k', n, \lambda\right\rangle , \qquad \lambda \in \mathbb{R}, \tag{10.7.419}$$

$$W^2 \left|k', n, \lambda\right\rangle = -\omega^2 \lambda^2 \left|k', n, \lambda\right\rangle \tag{10.7.420}$$

$$P^2 \left|k', n, \lambda\right\rangle = 0. \tag{10.7.421}$$

11 Differential Geometry of Curved Spacetimes

In this Chapter, I cover curved spacetime differential geometry proper from §(11.1) through §(11.4), focusing on issues not well developed in §(9). These three sections, together with §(9), are intended to form the first portion – the *kinematics* of curved space(time)s[114] – of a course on gravitation and physics in curved spacetimes. Following that, §(11.8) contains somewhat specialized content regarding the expansion of geometric quantities off some fixed 'background' geometry; and finally, in §(11.9) we compile conformal transformation properties of geometric objects.

11.1 Curved Metrics, Orthonormal Frames, Volume

Curved Spacetime, Spacetime Volume The generalization of the 'distance-squared' between x^μ to $x^\mu + \mathrm{d}x^\mu$, from the Minkowski to the curved case, is the following "line element":

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{11.1.1}$$

where x is simply shorthand for the spacetime coordinates $\{x^\mu\}$, which we emphasize may no longer be Cartesian. Because in a curved spacetime $g_{\mu\nu}$ can no longer be brought to the form $\eta_{\mu\nu}$ – a fact we shall examine in more detail below – note that this implies a global Lorentz inertial frame no longer exists. Much of Special Relativity no longer applies in a curved spacetime.

We need to demand that $g_{\mu\nu}$ be real, symmetric, and has 1 positive eigenvalue associated with the one 'time' coordinate and $(d-1)$ negative ones for the spatial coordinates.[115] The infinitesimal spacetime volume continues to take the form

$$\mathrm{d}(\text{vol.}) = \mathrm{d}^d x \sqrt{|g(x)|}, \tag{11.1.2}$$

where $|g(x)| = |\det g_{\mu\nu}(x)|$ is now the absolute value of the determinant of the metric $g_{\mu\nu}$.

Orthonormal Basis Cartesian coordinates play a basic but special role in interpreting physics in both flat Euclidean space δ_{ij} and flat Minkowski spacetime $\eta_{\mu\nu}$: they parametrize time durations and spatial distances in orthogonal directions – i.e., every increasing tick mark along a given Cartesian axis corresponds directly to a measurement of increasing length or time in that direction. This is generically not so, say, for coordinates in curved space(time) because the notion of what constitutes a 'straight line' is significantly more subtle there; or even spherical coordinates $(r \geq 0, 0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$ in flat 3D space – for the latter, only the radial coordinate r corresponds to actual distance (from the origin).

Therefore, just like the curved space case, to interpret physics in the neighborhood of some spacetime location x^μ, we introduce an orthonormal basis $\{\varepsilon^{\widehat{\mu}}{}_\alpha\}$ through the 'diagonalization' process:

$$g_{\mu\nu}(x) = \eta_{\alpha\beta}\varepsilon^{\widehat{\alpha}}{}_\mu(x)\varepsilon^{\widehat{\beta}}{}_\nu(x). \tag{11.1.3}$$

[114] As opposed to the dynamics of spacetime, which involves studying General Relativity, Einstein's field equations for the metric, and its applications.

[115] The opposite sign convention is more popular these days: one negative eigenvalue of $g_{\mu\nu}$ associated with time; and $(d-1)$ positive ones with space. Both sign conventions are usually equally valid; but see [36] for exotic exceptions.

By defining $\varepsilon^{\widehat{\alpha}} \equiv \varepsilon^{\widehat{\alpha}}{}_{\mu} \mathrm{d}x^{\mu}$, the analog to achieving a Cartesian-like expression for the spacetime metric is

$$\mathrm{d}s^2 = \left(\varepsilon^{\widehat{0}}\right)^2 - \sum_{i=1}^{D} \left(\varepsilon^{\widehat{i}}\right)^2 = \eta_{\mu\nu} \varepsilon^{\widehat{\mu}} \varepsilon^{\widehat{\nu}}. \tag{11.1.4}$$

This means under a local Lorentz transformation – i.e., for all

$$\Lambda^{\mu}{}_{\alpha}(x) \Lambda^{\nu}{}_{\beta}(x) \eta_{\mu\nu} = \eta_{\alpha\beta}, \tag{11.1.5}$$

$$\varepsilon'^{\widehat{\mu}}(x) = \Lambda^{\mu}{}_{\alpha}(x) \varepsilon'^{\widehat{\alpha}}(x) \tag{11.1.6}$$

– the metric remains the same:

$$\mathrm{d}s^2 = \eta_{\mu\nu} \varepsilon^{\widehat{\mu}} \varepsilon^{\widehat{\nu}} = \eta_{\mu\nu} \varepsilon'^{\widehat{\mu}} \varepsilon'^{\widehat{\nu}}. \tag{11.1.7}$$

By viewing $\widehat{\varepsilon}$ as the matrix with the αth row and μth column given by $\varepsilon^{\widehat{\alpha}}{}_{\mu}$, the determinant of the metric $g_{\mu\nu}$ can be written as

$$\det g_{\mu\nu}(x) = (\det \widehat{\varepsilon})^2 \det \eta_{\mu\nu}. \tag{11.1.8}$$

The infinitesimal spacetime volume in eq. (11.1.2) now can be expressed as

$$\mathrm{d}^d x \sqrt{|g(x)|} = \mathrm{d}^d x \det \widehat{\varepsilon} \tag{11.1.9}$$

$$= \varepsilon^{\widehat{0}} \wedge \varepsilon^{\widehat{1}} \wedge \dots \wedge \varepsilon^{\widehat{d-1}}. \tag{11.1.10}$$

The second equality follows because

$$\begin{aligned} \varepsilon^{\widehat{0}} \wedge \dots \wedge \varepsilon^{\widehat{d-1}} &= \varepsilon^{\widehat{0}}{}_{\mu_1} \mathrm{d}x^{\mu_1} \wedge \dots \wedge \varepsilon^{\widehat{d-1}}{}_{\mu_d} \mathrm{d}x^{\mu_d} \\ &= \epsilon_{\mu_1 \dots \mu_d} \varepsilon^{\widehat{0}}{}_{\mu_1} \dots \varepsilon^{\widehat{d-1}}{}_{\mu_d} \mathrm{d}x^0 \wedge \dots \wedge \mathrm{d}x^{d-1} = (\det \widehat{\varepsilon}) \mathrm{d}^d x. \end{aligned} \tag{11.1.11}$$

Of course, that $g_{\mu\nu}$ may be 'diagonalized' follows from the fact that $g_{\mu\nu}$ is a real symmetric matrix:

$$g_{\mu\nu} = \sum_{\alpha,\beta} O^{\alpha}{}_{\mu} \lambda_{\alpha} \eta_{\alpha\beta} O^{\beta}{}_{\nu} = \sum_{\alpha,\beta} \varepsilon^{\widehat{\alpha}}{}_{\mu} \eta_{\alpha\beta} \varepsilon^{\widehat{\beta}}{}_{\nu}, \tag{11.1.12}$$

where all $\{\lambda_{\alpha}\}$ are positive by assumption, so we may take their positive root:

$$\varepsilon^{\widehat{\alpha}}{}_{\mu} = \sqrt{\lambda_{\alpha}} O^{\alpha}{}_{\mu}, \qquad \{\lambda_{\alpha} > 0\}, \qquad (\text{No sum over } \alpha). \tag{11.1.13}$$

That $\varepsilon^{\widehat{0}}{}_{\mu}$ acts as 'standard clock' and $\{\varepsilon^{\widehat{i}}{}_{\mu} | i = 1, 2, \dots, D\}$ act as 'standard rulers' is because they are of unit length:

$$g^{\mu\nu} \varepsilon^{\widehat{\alpha}}{}_{\mu} \varepsilon^{\widehat{\beta}}{}_{\nu} = \eta^{\alpha\beta}. \tag{11.1.14}$$

The $\widehat{\cdot}$ on the index indicates it is to be moved with the flat metric, namely

$$\varepsilon^{\widehat{\alpha}}{}_{\mu} = \eta^{\alpha\beta}\varepsilon_{\widehat{\beta}\mu} \qquad \text{and} \qquad \varepsilon_{\widehat{\alpha}\mu} = \eta_{\alpha\beta}\varepsilon^{\widehat{\beta}}{}_{\mu}; \tag{11.1.15}$$

while the spacetime index is to be moved with the spacetime metric

$$\varepsilon^{\widehat{\alpha}\mu} = g^{\mu\nu}\varepsilon^{\widehat{\alpha}}{}_{\nu} \qquad \text{and} \qquad \varepsilon^{\widehat{\alpha}}{}_{\mu} = g_{\mu\nu}\varepsilon^{\widehat{\alpha}\nu}. \tag{11.1.16}$$

In other words, we view $\varepsilon_{\widehat{\alpha}}{}^{\mu}$ as the μth spacetime component of the αth vector field in the basis set $\{\varepsilon_{\widehat{\alpha}}{}^{\mu}|\alpha = 0, 1, 2, \dots, D \equiv d-1\}$. We may elaborate on the interpretation that $\{\varepsilon^{\widehat{\alpha}}{}_{\mu}\}$ act as 'standard clock/rulers' as follows. For a test (scalar) function $f(x)$ defined throughout spacetime, the rate of change of f along $\varepsilon_{\widehat{0}}$ is

$$\langle \mathrm{d}f|\, \varepsilon_{\widehat{0}}\rangle = \varepsilon_{\widehat{0}}{}^{\mu}\partial_\mu f \equiv \frac{\mathrm{d}f}{\mathrm{d}y^0}; \tag{11.1.17}$$

whereas that along $\varepsilon_{\widehat{i}}$ is

$$\langle \mathrm{d}f|\, \varepsilon_{\widehat{i}}\rangle = \varepsilon_{\widehat{i}}{}^{\mu}\partial_\mu f \equiv \frac{\mathrm{d}f}{\mathrm{d}y^i}; \tag{11.1.18}$$

where y^0 and $\{y^i\}$ are to be viewed as 'time' and 'spatial' parameters along the integral curves of $\{\varepsilon_{\widehat{\mu}}{}^{\alpha}\}$. That these are Cartesian-like can now be expressed as

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}y^\mu}\right| \left.\frac{\mathrm{d}}{\mathrm{d}y^\nu}\right\rangle = \varepsilon_{\widehat{\mu}}{}^{\alpha}\varepsilon_{\widehat{\nu}}{}^{\beta}\,\langle \partial_\alpha|\,\partial_\beta\rangle = \varepsilon_{\widehat{\mu}}{}^{\alpha}\varepsilon_{\widehat{\nu}}{}^{\beta}g_{\alpha\beta} = \eta_{\mu\nu}. \tag{11.1.19}$$

It is worth reiterating that the first equalities of eq. (11.1.12) are really assumptions, in that the definitions of curved spaces include assuming all the eigenvalues of the metric are positive whereas that of curved spacetimes include assuming all but one eigenvalue is negative.[116]

Problem 11.1. Orthonormal Frames in Kerr-Schild Spacetimes A special class of geometries, known as *Kerr-Schild* spacetimes, take the following form.

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + Hk_\mu k_\nu \tag{11.1.20}$$

Many of the known black hole spacetimes can be put in this form; and in such a context, $\bar{g}_{\mu\nu}$ usually refers to flat or de Sitter spacetime.[117] The k_μ is null with respect to $\bar{g}_{\mu\nu}$, i.e.,

$$\bar{g}_{\alpha\beta}k^\alpha k^\beta = 0, \tag{11.1.21}$$

and we shall move its indices with $\bar{g}_{\mu\nu}$.

[116]In $d-$spacetime dimensions, with our sign convention in place, if there were n 'time' directions and $(d-n)$ 'spatial' ones, then this carries with it the assumption that $g_{\mu\nu}$ has n positive eigenvalues and $(d-n)$ negative ones.

[117]See Gibbons et al. [37] arXiv: hep-th/0404008. The special property of Kerr-Schild coordinates is that Einstein's equations become *linear* in these coordinates.

Verify that the inverse metric is

$$g^{\mu\nu} = \bar{g}^{\mu\nu} - H k^\mu k^\nu, \tag{11.1.22}$$

where $\bar{g}^{\mu\sigma}$ is the inverse of $\bar{g}_{\mu\sigma}$, namely $\bar{g}^{\mu\sigma}\bar{g}_{\sigma\nu} \equiv \delta^\mu{}_\nu$. Suppose the orthonormal frame fields are known for $\bar{g}_{\mu\nu}$, namely

$$\bar{g}_{\mu\nu} = \eta_{\alpha\beta}\bar{\varepsilon}^{\widehat{\alpha}}{}_\mu \bar{\varepsilon}^{\widehat{\beta}}{}_\nu; \tag{11.1.23}$$

verify that the orthonormal frame fields are

$$\varepsilon^{\widehat{\alpha}}{}_\mu = \bar{\varepsilon}^{\widehat{\alpha}}{}_\sigma \left(\delta^\sigma{}_\mu + \frac{1}{2} H k^\sigma k_\mu\right). \tag{11.1.24}$$

Can you explain why k^μ is also null with respect to the full metric $g_{\mu\nu}$? □

Commutators and Coordinates Note that the $\{\mathrm{d}/\mathrm{d}y^\mu\}$ in eq. (11.1.19) do not, generically, commute. For instance, acting on a scalar function,

$$\begin{aligned}\left[\frac{\mathrm{d}}{\mathrm{d}y^\mu}, \frac{\mathrm{d}}{\mathrm{d}y^\nu}\right] f(x) &= \left(\frac{\mathrm{d}}{\mathrm{d}y^\mu}\frac{\mathrm{d}}{\mathrm{d}y^\nu} - \frac{\mathrm{d}}{\mathrm{d}y^\nu}\frac{\mathrm{d}}{\mathrm{d}y^\mu}\right) f(x) && (11.1.25)\\ &= \left(\varepsilon_{\widehat{\mu}}{}^\alpha \partial_\alpha \varepsilon_{\widehat{\nu}}{}^\beta - \varepsilon_{\widehat{\nu}}{}^\alpha \partial_\alpha \varepsilon_{\widehat{\mu}}{}^\beta\right) \partial_\beta f(x) \neq 0. && (11.1.26)\end{aligned}$$

More generally, for any two vector fields V^μ and W^μ, their commutator is

$$\begin{aligned}[V, W]^\mu &= V^\sigma \nabla_\sigma W^\mu - W^\sigma \nabla_\sigma V^\mu && (11.1.27)\\ &= V^\sigma \partial_\sigma W^\mu - W^\sigma \partial_\sigma V^\mu. && (11.1.28)\end{aligned}$$

(Can you explain why the covariant derivatives can be replaced with partial ones?) A theorem in differential geometry[118] tells us:

> A set of $1 < N \leq d$ vector fields $\{\mathrm{d}/\mathrm{d}\xi^\mu\}$ form a coordinate basis in the $d-$dimensional space(time) they inhabit, if and only if they commute.

To elaborate: if these N vector fields commute, we may integrate them to find a $N-$dimensional coordinate grid within the $d-$dimensional spacetime. Conversely, we are already accustomed to the fact that the partial derivatives with respect to the coordinates of space(time) do, of course, commute amongst themselves. When $N = d$, and if $[\mathrm{d}/\mathrm{d}y^\mu, \mathrm{d}/\mathrm{d}y^\nu] = 0$ in eq. (11.1.19), we would not only have found coordinates $\{y^\mu\}$ for our spacetime, we would have found this spacetime is a flat one.

What are coordinates? At this juncture, it is perhaps important to clarify what a coordinate system is. For instance, if we had in 2D $[\mathrm{d}/\mathrm{d}y^0, \mathrm{d}/\mathrm{d}y^1] \neq 0$, this means it is not possible to vary the 'coordinate' y^0 (i.e., along the integral curve of $\mathrm{d}/\mathrm{d}y^0$) without holding the 'coordinate' y^1 fixed; or, it is not possible to hold y^0 fixed while moving along the integral curve of $\mathrm{d}/\mathrm{d}y^1$. More generally, in a $d-$dimensional space(time), if x^μ is a coordinate parametrizing space(time), then it must be possible to vary it while keeping fixed the rest of its counterparts $\{x^\nu | \nu \neq \mu\}$.

[118] See, for instance, Schutz [23] for a pedagogical discussion.

11.2 Timelike, Spacelike vs. Null Vectors

A fundamental difference between (curved) space versus spacetime, is that the former involves strictly positive distances while the latter – because of the $\eta_{00} = +1$ for orthonormal 'time' versus $\eta_{ii} = -1$ for the ith orthonormal space component – involves positive, Zero, and negative 'distance-squared'.

With our 'mostly minus' sign convention (cf. eq. (10.1.2)), a vector v^μ is:

- *Time-like* if $v^2 \equiv \eta_{\mu\nu} v^{\widehat{\mu}} v^{\widehat{\nu}} > 0$. We have seen in §(10.1): if $v^2 > 0$, it is always possible to find a Lorentz transformation Λ (cf. eq. (10.1.13)) such that $\Lambda^\mu{}_\alpha v^{\widehat{\alpha}} = (v'^{\widehat{0}}, \vec{0})$. In flat spacetime, if $\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu > 0$ then this result indicates it is always possible to find an inertial frame where $\mathrm{d}s^2 = \mathrm{d}t'^2$: hence the phrase 'timelike'. (Also see Problem (10.50).)

  More generally, for a timelike trajectory $z^\mu(\lambda)$ in curved spacetime – i.e.,

$$g_{\mu\nu}(\mathrm{d}z^\mu/\mathrm{d}\lambda)(\mathrm{d}z^\nu/\mathrm{d}\lambda) > 0, \tag{11.2.1}$$

  we may identify

$$\mathrm{d}\tau \equiv \mathrm{d}\lambda \sqrt{g_{\mu\nu}(z(\lambda)) \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda}} \tag{11.2.2}$$

  as the (infinitesimal) *proper time*, the time read by the watch of an observer whose worldline is $z^\mu(\lambda)$.

  Suppose the timelike trajectory *were* – it need not always be – in 'free-fall', i.e., obeying the geodesic equation. Below, the resulting Fermi normal coordinate expansion of equations (11.4.8) through (11.4.10) teaches us, along the timelike worldline of a freely-falling observer the geometry becomes flat, i.e., $g_{\mu\nu} \to \eta_{\mu\nu}$; $z^0 = s = \tau$ is the proper time; and $\dot{z}^i = 0$: altogether, we thus recover the above statement that $g_{\mu\nu}\mathrm{d}z^\mu \mathrm{d}z^\nu = \eta_{00}(\mathrm{d}z^0)^2 = (\mathrm{d}\tau)^2$.

  More generally, using the orthonormal frame fields in eq. (11.1.12),

$$\mathrm{d}\tau = \mathrm{d}\lambda \sqrt{\eta_{\alpha\beta} \frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda} \frac{\mathrm{d}z^{\widehat{\beta}}}{\mathrm{d}\lambda}}, \qquad \frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda} \equiv \varepsilon^{\widehat{\alpha}}{}_\mu(z(\lambda)) \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda}. \tag{11.2.3}$$

  Since $v^{\widehat{\mu}} \equiv \mathrm{d}z^{\widehat{\mu}}/\mathrm{d}\lambda$ is assumed to be timelike, it must be possible to find a local Lorentz transformation $\Lambda^\mu{}_\nu(z)$ such that $\Lambda^\mu{}_\nu v^{\widehat{\nu}} = (v'^{\widehat{0}}, \vec{0})$. Assuming $\mathrm{d}\lambda > 0$,

$$\begin{aligned} \mathrm{d}\tau &= \mathrm{d}\lambda \sqrt{\eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda} \frac{\mathrm{d}z^{\widehat{\beta}}}{\mathrm{d}\lambda}}, \\ &= \mathrm{d}\lambda \sqrt{\left(\frac{\mathrm{d}z'^{\widehat{0}}}{\mathrm{d}\lambda}\right)^2} = |\mathrm{d}z'^{\widehat{0}}|. \end{aligned} \tag{11.2.4}$$

  The generalization of the discussion in Problem (10.50) to timelike trajectories $z^\mu(\tau)$ in generic curved spacetimes is as follows. If τ refers to its proper time and $u^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\tau$, then u^0 cannot be arbitrary but is related to the proper spatial velocity $\vec{u}$ via

$$g_{00}(u^0)^2 + 2g_{0i}u^0u^i + g_{ij}u^iu^j = +1. \tag{11.2.5}$$

Multiplying throughout by $1/(u^0)^2 = (\mathrm{d}\tau/\mathrm{d}x^0)^2$,

$$g_{00} + 2g_{0i}\frac{\mathrm{d}\tau}{\mathrm{d}x^0}\frac{\mathrm{d}z^i}{\mathrm{d}\tau} + g_{ij}\left(\frac{\mathrm{d}\tau}{\mathrm{d}x^0}\right)^2\frac{\mathrm{d}z^i}{\mathrm{d}\tau}\frac{\mathrm{d}z^j}{\mathrm{d}\tau} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}x^0}\right)^2 \tag{11.2.6}$$

$$g_{\mu\nu}\frac{\mathrm{d}z^\mu}{\mathrm{d}x^0}\frac{\mathrm{d}z^\nu}{\mathrm{d}x^0} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}x^0}\right)^2. \tag{11.2.7}$$

Furthermore, if the trajectory is moving forward in time, then $u^0 = \mathrm{d}x^0/\mathrm{d}\tau > 0$ and the positive square root is to be chosen:

$$\frac{\mathrm{d}\tau}{\mathrm{d}x^0} = +\sqrt{g_{\mu\nu}\frac{\mathrm{d}z^\mu}{\mathrm{d}x^0}\frac{\mathrm{d}z^\nu}{\mathrm{d}x^0}}. \tag{11.2.8}$$

- *Space-like* if $v^2 \equiv \eta_{\mu\nu}v^{\widehat{\mu}}v^{\widehat{\nu}} < 0$. We have seen in §(10.1): if $v^2 < 0$, it is always possible to find a Lorentz transformation Λ such that $\Lambda^\mu{}_\alpha v^{\widehat{\alpha}} = (0, v'^{\widehat{i}})$. In flat spacetime, if $\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu < 0$ then this result indicates it is always possible to find an inertial frame where $\mathrm{d}s^2 = -\mathrm{d}\vec{x}'^2$: hence the phrase 'spacelike'.

  More generally, for a spacelike trajectory $z^\mu(\lambda)$ in curved spacetime – i.e., $g_{\mu\nu}(\mathrm{d}z^\mu/\mathrm{d}\lambda)(\mathrm{d}z^\nu/\mathrm{d}\lambda) < 0$, we may identify

$$\mathrm{d}\ell \equiv \mathrm{d}\lambda\sqrt{\left|g_{\mu\nu}(z(\lambda))\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda}\right|} \tag{11.2.9}$$

  as the (infinitesimal) *proper length*, the distance read off some measuring rod whose trajectory is $z^\mu(\lambda)$. (As a check: when $g_{\mu\nu} = \eta_{\mu\nu}$ and $\mathrm{d}t = 0$, i.e., the rod is lying on the constant$-t$ surface, then $\mathrm{d}\ell = |\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}|^{1/2}$.) Using the orthonormal frame fields in eq. (11.1.12),

$$\mathrm{d}\ell = \mathrm{d}\lambda\sqrt{\left|\eta_{\alpha\beta}\frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda}\frac{\mathrm{d}z^{\widehat{\beta}}}{\mathrm{d}\lambda}\right|}, \qquad \frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda} \equiv \varepsilon^{\widehat{\alpha}}{}_\mu\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda}. \tag{11.2.10}$$

  Furthermore, since $v^{\widehat{\mu}} \equiv \mathrm{d}z^{\widehat{\mu}}/\mathrm{d}\lambda$ is assumed to be spacelike, it must be possible to find a local Lorentz transformation $\Lambda^\mu{}_\nu(z)$ such that $\Lambda^\mu{}_\nu v^{\widehat{\nu}} = (0, v'^{\widehat{i}})$; assuming $\mathrm{d}\lambda > 0$,

$$\mathrm{d}\ell = \mathrm{d}\lambda\sqrt{\eta_{\mu\nu}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta\frac{\mathrm{d}z^{\widehat{\alpha}}}{\mathrm{d}\lambda}\frac{\mathrm{d}z^{\widehat{\beta}}}{\mathrm{d}\lambda}} = |\mathrm{d}\vec{z}'|\,; \tag{11.2.11}$$

$$\mathrm{d}\vec{z}'^{\widehat{i}} \equiv \Lambda^i{}_\mu\varepsilon^{\widehat{\mu}}{}_\nu\mathrm{d}z^\nu. \tag{11.2.12}$$

- *Null* if $v^2 \equiv \eta_{\mu\nu}v^{\widehat{\mu}}v^{\widehat{\nu}} = 0$. We have already seen, in flat spacetime, if $\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = 0$ then $|\mathrm{d}\vec{x}|/\mathrm{d}x^0 = |\mathrm{d}\vec{x}'|/\mathrm{d}x'^0 = 1$ in all inertial frames.

It is physically important to reiterate: one of the reasons why it is important to make such a distinction between vectors, is because it is *not possible* to find a Lorentz transformation that would linearly transform one of the above three types of vectors into another different type – for e.g., it is not possible to Lorentz transform a null vector into a time-like one (a photon has no 'rest frame'); or a time-like vector into a space-like one; etc. This is because their Lorentzian 'norm-squared'

$$v^2 \equiv \eta_{\mu\nu} v^{\widehat{\mu}} v^{\widehat{\nu}} = \eta_{\alpha\beta} \Lambda^{\alpha}{}_{\mu} \Lambda^{\beta}{}_{\nu} v^{\widehat{\mu}} v^{\widehat{\nu}} = \eta_{\alpha\beta} v'^{\widehat{\alpha}} v'^{\widehat{\beta}} \tag{11.2.13}$$

has to be invariant under all Lorentz transformations $v'^{\widehat{\alpha}} \equiv \Lambda^{\alpha}{}_{\mu} v^{\widehat{\mu}}$. This in turn teaches us: if v^2 were positive, it has to remain so; likewise, if it were zero or negative, a Lorentz transformation cannot alter this attribute.

Proper times and Gravitational Time Dilation Consider two observers sweeping out their respective timelike worldlines in spacetime, $y^\mu(\lambda)$ and $z^\mu(\lambda)$. If we use the time coordinate of the geometry to parameterize their trajectories, their proper times – i.e., the time read by their watches – are given by

$$\mathrm{d}\tau_y \equiv \mathrm{d}t \sqrt{g_{\mu\nu}(y(t)) \dot{y}^\mu \dot{y}^\nu}, \qquad \dot{y}^\mu \equiv \frac{\mathrm{d}y^\mu}{\mathrm{d}t}; \tag{11.2.14}$$

$$\mathrm{d}\tau_z \equiv \mathrm{d}t \sqrt{g_{\mu\nu}(z(t)) \dot{z}^\mu \dot{z}^\nu}, \qquad \dot{z}^\mu \equiv \frac{\mathrm{d}z^\mu}{\mathrm{d}t}. \tag{11.2.15}$$

In flat spacetime, clocks that are synchronized in one frame are no longer synchronized in a different frame – chronology is not a Lorentz invariant. We see that, in curved spacetime, the infinitesimal *passage* of proper time measured by observers at the same 'coordinate time' t depends on their spacetime locations:

$$\frac{\mathrm{d}\tau_y}{\mathrm{d}\tau_z} = \sqrt{\frac{g_{\mu\nu}(y(t)) \dot{y}^\mu \dot{y}^\nu}{g_{\alpha\beta}(z(t)) \dot{z}^\alpha \dot{z}^\beta}}. \tag{11.2.16}$$

Physically speaking, eq. (11.2.16) *does not*, in general, yield the ratio of proper times measured by observers at two different locations. (Drawing a spacetime diagram here helps.) To do so, one would have to specify the trajectories of both $y^\mu(\lambda_1 \leq \lambda \leq \lambda_2)$ and $z^\mu(\lambda'_1 \leq \lambda' \leq \lambda'_2)$, before the integrals $\Delta\tau_1 \equiv \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \sqrt{g_{\mu\nu} \dot{y}^\mu \dot{y}^\nu}$ and $\Delta\tau_2 \equiv \int_{\lambda'_1}^{\lambda'_2} \mathrm{d}\lambda' \sqrt{g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu}$ are evaluated and compared. On the other hand, in the weak field limit – and as the following Problem demonstrates – eq. (11.2.16) *can* often be used to compare the elapse of proper times.

Problem 11.2. Near Earth Proper Times and Distances The spacetime geometry around the Earth itself can be approximated by the line element

$$\mathrm{d}s^2 = \left(1 - \frac{r_{s,E}}{r}\right) \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - r_{s,E}/r} - r^2 \left(\mathrm{d}\theta^2 + \sin(\theta)^2 \mathrm{d}\phi^2\right), \tag{11.2.17}$$

where t is the time coordinate and (r, θ, ϕ) are analogs of the spherical coordinates. Whereas $r_{s,E}$ is known as the Schwarzschild radius of the Earth, and depends on the Earth's mass M_{E} through the expression

$$r_{s,E} \equiv 2 G_{\mathrm{N}} M_{\mathrm{E}}. \tag{11.2.18}$$

Find the 4−beins (i.e., orthonormal frame fields) of the geometry in eq. (11.2.17). Then find the numerical value of $r_{s,E}$ in eq. (11.2.18) and take the ratio $r_{s,E}/R_{\mathrm{E}}$, where R_{E} is the radius of the Earth. Explain why this means we may – for practical purposes – expand the metric in eq. (11.2.18) as

$$\mathrm{d}s^2 = \left(1 - \frac{r_{s,E}}{r}\right)\mathrm{d}t^2 - \mathrm{d}r^2\left(1 + \frac{r_{s,E}}{r} + \left(\frac{r_{s,E}}{r}\right)^2 + \left(\frac{r_{s,E}}{r}\right)^3 + \dots\right) \\ - r^2\left(\mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2\right). \tag{11.2.19}$$

Since we are not in flat spacetime, the (t, r, θ, ϕ) are no longer subject to the same interpretation. However, use your computation of $r_{s,E}/R_{\mathrm{E}}$ to *estimate* the error incurred if we do continue to interpret t and r as though they measured time and radial distances, with respect to a frame centered at the Earth's core.

Taipei 101[119] Consider placing one clock at the base of the Taipei 101 tower and another at its tip. Denoting the time elapsed at the base of the tower as $\Delta\tau_{\mathrm{B}}$; that at the tip as $\Delta\tau_{\mathrm{T}}$; and assuming for simplicity the Earth is a perfect sphere – show that eq. (11.2.16) translates to

$$\frac{\Delta\tau_{\mathrm{B}}}{\Delta\tau_{\mathrm{T}}} = \sqrt{\frac{g_{00}(R_{\mathrm{E}})}{g_{00}(R_{\mathrm{E}} + h_{101})}} \approx 1 + \frac{1}{2}\left(\frac{r_{s,E}}{R_{\mathrm{E}} + h_{101}} - \frac{r_{s,E}}{R_{\mathrm{E}}}\right). \tag{11.2.20}$$

Here, R_{E} is the radius of the Earth and h_{101} is the height of the Taipei 101 tower. Notice the right hand side is related to the difference in the Newtonian gravitational potentials at the top and bottom of the tower.

In actuality, both clocks are in motion, since the Earth is rotating. Can you estimate what is the error incurred from assuming they are at rest? First arrive at eq. (11.2.20) analytically, then plug in the relevant numbers to compute the numerical value of $\Delta\tau_{\mathrm{B}}/\Delta\tau_{\mathrm{T}}$. Does the clock at the base of Taipei 101 or that on its tip tick more slowly?

Global Positioning Satellites This gravitational time dilation is an effect that needs to be accounted for when setting up a network of Global Positioning Satellites (GPS); for details, see Ashby [38]. □

11.3 Connections, Curvature, Geodesics

Connections and Christoffel Symbols The partial derivative on a scalar φ is a rank-1 tensor, so we shall simply define the covariant derivative acting on φ to be

$$\nabla_\alpha \varphi = \partial_\alpha \varphi. \tag{11.3.1}$$

Because the partial derivative itself cannot yield a tensor once it acts on tensor, we need to introduce a connection $\Gamma^\mu{}_{\alpha\beta}$, i.e.,

$$\nabla_\sigma V^\mu = \partial_\sigma V^\mu + \Gamma^\mu{}_{\sigma\rho} V^\rho. \tag{11.3.2}$$

Under a coordinate transformation of the partial derivatives and V^μ, say going from x to x',

$$\partial_\sigma V^\mu + \Gamma^\mu{}_{\sigma\rho} V^\rho = \frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial x^\mu}{\partial x'^\nu}\partial_{\lambda'} V^{\nu'} + \left(\frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial^2 x^\mu}{\partial x'^\lambda x'^\nu} + \Gamma^\mu{}_{\sigma\rho}\frac{\partial x^\rho}{\partial x'^\nu}\right) V^{\nu'}. \tag{11.3.3}$$

[119] Two closely related experiments: (I) Pound and Rebka and (II) Gravity Probe A.

On the other hand, if $\nabla_\sigma V^\mu$ were to transform as a tensor,

$$\partial_\sigma V^\mu + \Gamma^\mu{}_{\sigma\rho} V^\rho = \frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial x^\mu}{\partial x'^\nu}\partial_{\lambda'} V^{\nu'} + \frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial x^\mu}{\partial x'^\tau}\Gamma^{\tau'}{}_{\lambda'\nu'} V^{\nu'}. \tag{11.3.4}$$

[120]Since $V^{\nu'}$ is an arbitrary vector, we may read off its coefficient on the right hand sides of equations (11.3.3) and (11.3.4), and deduce the connection has to transform as

$$\frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial^2 x^\mu}{\partial x'^\lambda x'^\nu} + \Gamma^\mu{}_{\sigma\rho}(x)\frac{\partial x^\rho}{\partial x'^\nu} = \frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial x^\mu}{\partial x'^\tau}\Gamma^{\tau'}{}_{\lambda'\nu'}(x'). \tag{11.3.5}$$

Moving all the Jacobians onto the connection written in the $\{x^\mu\}$ frame,

$$\Gamma^{\tau'}{}_{\kappa'\nu'}(x') = \frac{\partial x'^\tau}{\partial x^\mu}\frac{\partial^2 x^\mu}{\partial x'^\kappa x'^\nu} + \frac{\partial x'^\tau}{\partial x^\mu}\Gamma^\mu{}_{\sigma\rho}(x)\frac{\partial x^\sigma}{\partial x'^\kappa}\frac{\partial x^\rho}{\partial x'^\nu}. \tag{11.3.6}$$

All connections have to satisfy this non-tensorial transformation law. On the other hand, if we found an object that transforms according to eq. (11.3.6), and if one employs it in eq. (11.3.2), then the resulting $\nabla_\alpha V^\mu$ would transform as a tensor.

Product rule Because covariant derivatives should reduce to partial derivatives in flat Cartesian coordinates, it is natural to require the former to obey the usual product rule. For any two tensors T_1 and T_2, and suppressing all indices,

$$\nabla(T_1 T_2) = (\nabla T_1)T_2 + T_1(\nabla T_2). \tag{11.3.7}$$

Problem 11.3. Covariant Derivative on 1-form Let us take the covariant derivative of a 1-form:

$$\nabla_\alpha V_\mu = \partial_\alpha V_\mu + \Gamma'^\sigma{}_{\alpha\mu} V_\sigma. \tag{11.3.8}$$

Can you prove that this connection is negative of the vector one in eq. (11.3.2)?

$$\Gamma'^\sigma{}_{\alpha\mu} = -\Gamma^\sigma{}_{\alpha\mu}, \tag{11.3.9}$$

where $\Gamma^\sigma{}_{\alpha\mu}$ is the connection in eq. (11.3.2) – if we define the covariant derivative of a scalar to be simply the partial derivative acting on the same, i.e.,

$$\nabla_\alpha \left(V^\mu W_\mu\right) = \partial_\alpha \left(V^\mu W_\mu\right)? \tag{11.3.10}$$

You should assume the product rule holds, namely $\nabla_\alpha (V^\mu W_\mu) = (\nabla_\alpha V^\mu) W_\mu + V^\mu (\nabla_\alpha W_\mu)$. Expand these covariant derivatives in terms of the connections and argue why this leads to eq. (11.3.9). □

Suppose we found two such connections, ${}_{(1)}\Gamma^\tau{}_{\kappa\nu}(x)$ and ${}_{(2)}\Gamma^\tau{}_{\kappa\nu}(x)$. Notice their difference does transform as a tensor because the first term on the right hand side involving the Hessian $\partial^2 x/\partial x'\partial x'$ cancels out:

$${}_{(1)}\Gamma^{\tau'}{}_{\kappa'\nu'}(x') - {}_{(2)}\Gamma^{\tau'}{}_{\kappa'\nu'}(x') = \frac{\partial x'^\tau}{\partial x^\mu}\left({}_{(1)}\Gamma^\mu{}_{\sigma\rho}(x) - {}_{(2)}\Gamma^\mu{}_{\sigma\rho}(x)\right)\frac{\partial x^\sigma}{\partial x'^\kappa}\frac{\partial x^\rho}{\partial x'^\nu}. \tag{11.3.11}$$

[120]All un-primed indices represent tensor components in the x-system; while all primed indices those in the x' system.

Now, any connection can be decomposed into its symmetric and antisymmetric parts in the following sense:

$$\Gamma^{\mu}{}_{\alpha\beta} = \frac{1}{2}\Gamma^{\mu}{}_{\{\alpha\beta\}} + \frac{1}{2}\Gamma^{\mu}{}_{[\alpha\beta]}. \tag{11.3.12}$$

This is, of course, mere tautology. However, let us denote

$$_{(1)}\Gamma^{\mu}{}_{\alpha\beta} \equiv \frac{1}{2}\Gamma^{\mu}{}_{\alpha\beta} \qquad \text{and} \qquad _{(2)}\Gamma^{\mu}{}_{\alpha\beta} \equiv \frac{1}{2}\Gamma^{\mu}{}_{\beta\alpha}; \tag{11.3.13}$$

so that

$$\frac{1}{2}\Gamma^{\mu}{}_{[\alpha\beta]} = {}_{(1)}\Gamma^{\mu}{}_{\alpha\beta} - {}_{(2)}\Gamma^{\mu}{}_{\alpha\beta} \equiv T^{\mu}{}_{\alpha\beta}. \tag{11.3.14}$$

We then see that this anti-symmetric part of the connection is in fact a tensor. It is the symmetric part $(1/2)\Gamma^{\mu}{}_{\{\alpha\beta\}}$ that does not transform as a tensor. *For the rest of these notes, by $\Gamma^{\mu}{}_{\alpha\beta}$ we shall always mean a symmetric connection.* This means our covariant derivative would now read

$$\nabla_{\alpha}V^{\mu} = \partial_{\alpha}V^{\mu} + \Gamma^{\mu}{}_{\alpha\beta}V^{\beta} + T^{\mu}{}_{\alpha\beta}V^{\beta}. \tag{11.3.15}$$

As is common within the physics literature, we proceed to set to zero the torsion term: $T^{\mu}{}_{\alpha\beta} \to 0$. If we further impose the metric compatibility condition,

$$\nabla_{\mu}g_{\alpha\beta} = 0, \tag{11.3.16}$$

then we have already seen in §(9) this (together with the zero torsion assumption) implies

$$\Gamma^{\mu}{}_{\alpha\beta} = \frac{1}{2}g^{\mu\sigma}\left(\partial_{\alpha}g_{\beta\sigma} + \partial_{\beta}g_{\alpha\sigma} - \partial_{\sigma}g_{\alpha\beta}\right). \tag{11.3.17}$$

[121]**Parallel Transport and Riemann Tensor** Along a curve $z^{\mu}(\lambda)$ such that one end is $z^{\mu}(\lambda=\lambda_1) = x'^{\mu}$ and the other end is $z^{\mu}(\lambda=\lambda_2) = x^{\mu}$, we may parallel transport some vector V^{α} from x' to x, i.e., over a finite range of the $\lambda-$parameter, by exponentiating the covariant derivative along $z^{\mu}(\lambda)$. If $V^{\alpha}(x' \to x)$ is the result of this parallel transport – not to be confused with $V^{\alpha}(x)$, which is simply V^{α} evaluated at x – we have

$$V^{\alpha}\left(x' \overset{z(\lambda)}{\to} x\right) = \exp\left[(\lambda_2-\lambda_1)\dot{z}^{\mu}(\lambda)\nabla_{\mu}\right]V^{\alpha}(x')\big|_{\lambda=\lambda_1}. \tag{11.3.18}$$

This is the covariant derivative analog of the Taylor expansion of a scalar function – where, translation by a constant spacetime vector a^{μ} may be implemented as

$$f(x^{\mu}+a^{\mu}) = \exp\left(a^{\nu}\partial_{\nu}\right)f(x^{\mu}). \tag{11.3.19}$$

[121]Note that if we were to relax both the zero torsion and metric compatibility conditions, this amounts to introducing two new tensors: $(1/2)\Gamma^{\mu}{}_{[\alpha\beta]} = T^{\mu}{}_{\alpha\beta}$ and $\nabla_{\mu}g_{\alpha\beta} = Q_{\mu\alpha\beta}$. If they are of any physical relevance, we would need to introduce dynamics for them: namely, what sort of partial differential equations do $T^{\mu}{}_{\alpha\beta}$ and $Q_{\mu\alpha\beta}$ obey; and, what are they sourced by?

Eq. (11.3.18) is also consistent with the discussion leading up to eq. (9.5.23), which in the curved spacetime context would be: a spacetime tensor $T^{\mu_1\cdots\mu_N}$ is invariant under parallel transport along some curve whose tangent vector is v^μ, whenever

$$v^\sigma \nabla_\sigma T^{\mu_1\cdots\mu_N} = 0 \tag{11.3.20}$$

along the entire curve. For, once $\dot{z}^\mu(\lambda_1)$ in eq. (11.3.18) is identified with v^μ, if eq. (11.3.20) is satisfied then

$$\exp\left[(\lambda_2 - \lambda_1) v^\mu(x') \nabla_\mu\right] V^\alpha(x') = V^\alpha(x'), \tag{11.3.21}$$

since the first covariant-derivative – and hence all higher ones – in the exp-Taylor series must yield zero.

To elucidate the definition of geometric curvature as the failure of tensors to remain invariant under parallel transport, we may now attempt to parallel transport a vector V^α around a closed parallelogram defined by the tangent vectors A and B. We shall soon see how the Riemann tensor itself emerges from such an analysis.

Let the 4 sides of this parallelogram have infinitesimal affine parameter length ϵ. We will now start from one of its 4 corners, which we will denote as x. V^α will be parallel transported from x to $x + \epsilon A$; then to $x + \epsilon A + \epsilon B$; then to $x + \epsilon A + \epsilon B - \epsilon A = x + \epsilon B$; and finally back to $x + \epsilon B - \epsilon B = x$. Let us first work out the parallel transport along the 'side' A using eq. (11.3.18). Denoting $\nabla_A \equiv A^\mu \nabla_\mu$, $\nabla_B \equiv B^\mu \nabla_\mu$, etc.,

$$\begin{aligned} V^\alpha(x \to x + \epsilon A) &= \exp(\epsilon \nabla_A) V^\alpha(x), \\ &= V^\alpha(x) + \epsilon \nabla_A V^\alpha(x) + \frac{\epsilon^2}{2} \nabla_A^2 V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right). \end{aligned} \tag{11.3.22}$$

We then parallel transport this result from $x + \epsilon A$ to $x + \epsilon A + \epsilon B$.

$$\begin{aligned} &V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B) \\ &= \exp(\epsilon \nabla_B) \exp(\epsilon \nabla_A) V^\alpha(x), \\ &= V^\alpha(x) + \epsilon \nabla_A V^\alpha(x) + \frac{\epsilon^2}{2} \nabla_A^2 V^\alpha(x) \\ &\qquad + \epsilon \nabla_B V^\alpha(x) + \epsilon^2 \nabla_B \nabla_A V^\alpha(x) \\ &\qquad + \frac{\epsilon^2}{2} \nabla_B^2 V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right) \\ &= V^\alpha(x) + \epsilon \left(\nabla_A + \nabla_B\right) V^\alpha(x) + \frac{\epsilon^2}{2} \left(\nabla_A^2 + \nabla_B^2 + 2\nabla_B \nabla_A\right) V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right). \end{aligned} \tag{11.3.23}$$

Pressing on, we now parallel transport this result from $x + \epsilon A + \epsilon B$ to $x + \epsilon B$.

$$\begin{aligned} &V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x + \epsilon B) \\ &= \exp(-\epsilon \nabla_A) \exp(\epsilon \nabla_B) \exp(\epsilon \nabla_A) V^\alpha(x), \\ &= V^\alpha(x) + \epsilon \left(\nabla_A + \nabla_B\right) V^\alpha(x) + \frac{\epsilon^2}{2} \left(\nabla_A^2 + \nabla_B^2 + 2\nabla_B \nabla_A\right) V^\alpha(x) \\ &\qquad - \epsilon \nabla_A V^\alpha(x) - \epsilon^2 \left(\nabla_A^2 + \nabla_A \nabla_B\right) V^\alpha(x) \end{aligned}$$

$$+\frac{\epsilon^2}{2}\nabla_A^2 V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right)$$
$$= V^\alpha(x) + \epsilon\nabla_B V^\alpha(x) + \epsilon^2\left(\frac{1}{2}\nabla_B^2 + \nabla_B\nabla_A - \nabla_A\nabla_B\right)V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right). \tag{11.3.24}$$

Finally, we parallel transport this back to $x + \epsilon B - \epsilon B = x$.

$$\begin{aligned}
&V^\alpha(x \to x+\epsilon A \to x+\epsilon A+\epsilon B \to x+\epsilon B \to x) \\
&= \exp(-\epsilon\nabla_B)\exp(-\epsilon\nabla_A)\exp(\epsilon\nabla_B)\exp(\epsilon\nabla_A)V^\alpha(x), \\
&= V^\alpha(x) + \epsilon\nabla_B V^\alpha(x) + \epsilon^2\left(\frac{1}{2}\nabla_B^2 + \nabla_B\nabla_A - \nabla_A\nabla_B\right)V^\alpha(x) \\
&\qquad - \epsilon\nabla_B V^\alpha(x) - \epsilon^2\nabla_B^2 V^\alpha(x) \\
&\qquad + \frac{\epsilon^2}{2}\nabla_B^2 V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right) \\
&= V^\alpha(x) + \epsilon^2\left(\nabla_B\nabla_A - \nabla_A\nabla_B\right)V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right).
\end{aligned} \tag{11.3.25}$$

[122]We have arrived at the central characterization of *local* geometric curvature. By parallel transporting a vector around an infinitesimal parallelogram, we see the parallel transported vector differs from the original one by the commutator of covariant derivatives with respect to the two tangent vectors defining the parallelogram. In the same vein, their difference is also proportional to the area of this parallogram, i.e., it scales as $\mathcal{O}\left(\epsilon^2\right)$ for infinitesimal ϵ.

$$V^\alpha(x \to x+\epsilon A \to x+\epsilon A+\epsilon B \to x+\epsilon B \to x) - V^\alpha(x) \tag{11.3.26}$$
$$= \epsilon^2\left[\nabla_B, \nabla_A\right]V^\alpha(x) + \mathcal{O}\left(\epsilon^3\right),$$
$$[\nabla_B, \nabla_A] \equiv \nabla_B\nabla_A - \nabla_A\nabla_B. \tag{11.3.27}$$

We shall proceed to calculate the commutator in a coordinate basis.

$$\begin{aligned}
[\nabla_A, \nabla_B]V^\mu &\equiv A^\sigma\nabla_\sigma\left(B^\rho\nabla_\rho V^\mu\right) - B^\sigma\nabla_\sigma\left(A^\rho\nabla_\rho V^\mu\right) \\
&= \left(A^\sigma\nabla_\sigma B^\rho - B^\sigma\nabla_\sigma A^\rho\right)\nabla_\rho V^\mu + A^\sigma B^\rho[\nabla_\sigma, \nabla_\rho]V^\mu.
\end{aligned} \tag{11.3.28}$$

Let us tackle the two groups separately. Firstly,

$$\begin{aligned}
[A, B]^\rho\nabla_\rho V^\mu &\equiv \left(A^\sigma\nabla_\sigma B^\rho - B^\sigma\nabla_\sigma A^\rho\right)\nabla_\rho V^\mu \\
&= \left(A^\sigma\partial_\sigma B^\rho + \Gamma^\rho{}_{\sigma\lambda}A^\sigma B^\lambda - B^\sigma\partial_\sigma A^\rho - \Gamma^\rho{}_{\sigma\lambda}B^\sigma A^\lambda\right)\nabla_\rho V^\mu \\
&= \left(A^\sigma\partial_\sigma B^\rho - B^\sigma\partial_\sigma A^\rho\right)\nabla_\rho V^\mu.
\end{aligned} \tag{11.3.29}$$

Next, we need $A^\sigma B^\rho[\nabla_\sigma, \nabla_\rho]V^\mu = A^\sigma B^\rho(\nabla_\sigma\nabla_\rho - \nabla_\rho\nabla_\sigma)V^\mu$. The first term is

$$A^\sigma B^\rho\nabla_\sigma\nabla_\rho V^\mu = A^\sigma B^\rho\left(\partial_\sigma\nabla_\rho V^\mu - \Gamma^\lambda{}_{\sigma\rho}\nabla_\lambda V^\mu + \Gamma^\mu{}_{\sigma\lambda}\nabla_\rho V^\lambda\right)$$

[122]The careful reader may complain, we should have evaluated the covariant derivatives at the various corners of the parallelogram – namely, $\exp(-\epsilon\nabla_{B(x+\epsilon A(x)+\epsilon B(x+\epsilon A(x))-\epsilon A(x+\epsilon A(x)+\epsilon B(x+\epsilon A(x))))})$ $\exp(-\epsilon\nabla_{A(x+\epsilon A(x)+\epsilon B(x+\epsilon A(x)))})$ $\exp(\epsilon\nabla_{B(x+\epsilon A(x))})\exp(\epsilon\nabla_{A(x)})$ – rather than all at x, as we have done here. Note that this would not have altered the lowest order results, i.e., the $\epsilon^2[\nabla_B, \nabla_A]V^\alpha$, since evaluating at the corners will multiply the extant terms by $(1 + \mathcal{O}(\epsilon))$.

$$
\begin{aligned}
&= A^\sigma B^\rho \left(\partial_\sigma \left(\partial_\rho V^\mu + \Gamma^\mu{}_{\rho\lambda} V^\lambda\right) - \Gamma^\lambda{}_{\sigma\rho} \left(\partial_\lambda V^\mu + \Gamma^\mu{}_{\lambda\omega} V^\omega\right) + \Gamma^\mu{}_{\sigma\lambda} \left(\partial_\rho V^\lambda + \Gamma^\lambda{}_{\rho\omega} V^\omega\right)\right) \\
&= A^\sigma B^\rho \Big\{\partial_\sigma \partial_\rho V^\mu + \partial_\sigma \Gamma^\mu{}_{\rho\lambda} V^\lambda + \Gamma^\mu{}_{\rho\lambda} \partial_\sigma V^\lambda - \Gamma^\lambda{}_{\sigma\rho} \left(\partial_\lambda V^\mu + \Gamma^\mu{}_{\lambda\omega} V^\omega\right) \\
&\qquad + \Gamma^\mu{}_{\sigma\lambda} \left(\partial_\rho V^\lambda + \Gamma^\lambda{}_{\rho\omega} V^\omega\right) \Big\}. \qquad (11.3.30)
\end{aligned}
$$

Swapping $(\sigma \leftrightarrow \rho)$ within the parenthesis $\{\dots\}$ and subtract the two results, we gather

$$
\begin{aligned}
A^\sigma B^\rho [\nabla_\sigma, \nabla_\rho] V^\mu &= A^\sigma B^\rho \Big\{\partial_{[\sigma} \Gamma^\mu{}_{\rho]\lambda} V^\lambda + \Gamma^\mu{}_{\lambda[\rho} \partial_{\sigma]} V^\lambda - \Gamma^\lambda{}_{[\sigma\rho]} \left(\partial_\lambda V^\mu + \Gamma^\mu{}_{\lambda\omega} V^\omega\right) \\
&\qquad + \Gamma^\mu{}_{\lambda[\sigma} \partial_{\rho]} V^\lambda + \Gamma^\mu{}_{\lambda[\sigma} \Gamma^\lambda{}_{\rho]\omega} V^\omega \Big\} \qquad (11.3.31) \\
&= A^\sigma B^\rho \left(\partial_{[\sigma} \Gamma^\mu{}_{\rho]\omega} + \Gamma^\mu{}_{\lambda[\sigma} \Gamma^\lambda{}_{\rho]\omega}\right) V^\omega. \qquad (11.3.32)
\end{aligned}
$$

Notice we have used the symmetry of the Christoffel symbols $\Gamma^\mu{}_{\alpha\beta} = \Gamma^\mu{}_{\beta\alpha}$ to arrive at this result. Since A and B are arbitrary, let us observe that the commutator of covariant derivatives acting on a vector field is not a different operator, but rather an algebraic operation:

$$
\begin{aligned}
[\nabla_\mu, \nabla_\nu] V^\alpha &= R^\alpha{}_{\beta\mu\nu} V^\beta, \qquad (11.3.33) \\
R^\alpha{}_{\beta\mu\nu} &\equiv \partial_{[\mu} \Gamma^\alpha{}_{\nu]\beta} + \Gamma^\alpha{}_{\sigma[\mu} \Gamma^\sigma{}_{\nu]\beta} \qquad (11.3.34) \\
&= \partial_\mu \Gamma^\alpha{}_{\nu\beta} - \partial_\nu \Gamma^\alpha{}_{\mu\beta} + \Gamma^\alpha{}_{\sigma\mu} \Gamma^\sigma{}_{\nu\beta} - \Gamma^\alpha{}_{\sigma\nu} \Gamma^\sigma{}_{\mu\beta}. \qquad (11.3.35)
\end{aligned}
$$

Inserting the results in equations (11.3.29) and (11.3.32) into eq. (11.3.28) – we gather, for arbitrary vector fields A and B:

$$
\left([\nabla_A, \nabla_B] - \nabla_{[A,B]}\right) V^\mu = R^\mu{}_{\nu\alpha\beta} V^\nu A^\alpha B^\beta. \qquad (11.3.36)
$$

Moreover, we may return to eq. (11.3.26) and re-express it as

$$
\begin{aligned}
&V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x + \epsilon B \to x) - V^\alpha(x) \qquad (11.3.37) \\
&\quad = \epsilon^2 \left(R^\alpha{}_{\beta\mu\nu}(x) V^\beta(x) B^\mu(x) A^\nu(x) + \nabla_{[B,A]} V^\alpha(x)\right) + \mathcal{O}\left(\epsilon^3\right). \qquad (11.3.38)
\end{aligned}
$$

When $A = \partial_\mu$ and $B = \partial_\nu$ are coordinate basis vectors themselves, $[A, B] = [\partial_\mu, \partial_\nu] = 0$, and eq. (11.3.36) then coincides with eq. (11.3.33). Earlier, we have already mentioned: if $[A, B] = 0$, the vector fields A and B can be integrated to form a local 2D coordinate system; while if $[A, B] \neq 0$, they cannot form a good coordinate system. Hence the failure of parallel transport invariance due to the $\nabla_{[A,B]}$ term in eq. (11.3.37) is really a measure of the coordinate-worthiness of A and B; whereas it is the Riemann tensor term that appears to tell us something about the intrinsic local curvature of the geometry itself.

Covariantly Constant Vector Fields What we have just argued is, there are no covariantly constant vector fields unless the Riemann tensor is zero. That is,

$$
\nabla_\mu V_\nu = 0 \qquad (11.3.39)
$$

has a solution if and only if Riemann is zero.

Problem 11.4. Index Symmetries of the Riemann tensor Explain why, if a tensor $\Sigma_{\alpha\beta}$ is antisymmetric in one coordinate system, it has to be anti-symmetric in any other coordinate system. Similarly, explain why, if $\Sigma_{\alpha\beta}$ is symmetric in one coordinate system, it has to be symmetric in any other coordinate system. Compute the Riemann tensor in a locally flat coordinate system[123] and show that

$$R_{\alpha\beta\mu\nu} = \frac{1}{2}\left(\partial_\beta \partial_{[\mu} g_{\nu]\alpha} - \partial_\alpha \partial_{[\mu} g_{\nu]\beta}\right). \tag{11.3.40}$$

From this result, argue that Riemann has the following symmetries:

$$R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}, \qquad R_{\mu\nu\alpha\beta} = -R_{\nu\mu\alpha\beta}, \qquad R_{\mu\nu\alpha\beta} = -R_{\mu\nu\beta\alpha}. \tag{11.3.41}$$

Additionally, prove the following Bianchi identity:

$$R_{\mu[\nu\alpha\beta]} = 0. \tag{11.3.42}$$

Also show that the symmetry in eq. (11.3.42) is equivalent to the following – you may find eq. (9.6.33) useful here:

$$R_{\mu\nu\alpha\beta} = R_{\mu[\beta\alpha]\nu}. \tag{11.3.43}$$

These index symmetries indicate the components of the Riemann tensor are not all (algebraically) independent. (Bonus: Can you show, these are the *only* index symmetries of Riemann?) Below, we shall see there are additional differential relations (aka "Bianchi identities") between various components of the Riemann tensor.

Finally, use these symmetries to show that

$$[\nabla_\alpha, \nabla_\beta] V_\nu = -R^\mu{}_{\nu\alpha\beta} V_\mu. \tag{11.3.44}$$

Hint: Start with $[\nabla_\alpha, \nabla_\beta](g_{\nu\sigma} V^\sigma)$. □

Assuming equations (11.3.41) and (11.3.42) are the *only* index symmetries of the Riemann tensor, we may now prove that it has $d^2(d^2-1)/12$ algebraically independent components in d space(time) dimensions. First view Riemann as a 2-index object $R_{\mu\nu\alpha\beta} \equiv R_{\mathrm{AB}}$, where A and B refer respectively to the pair $\mu\nu$ and $\alpha\beta$. In this notation, $R_{\mathrm{AB}} = R_{\mathrm{BA}}$ because of the $R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}$ in eq. (11.3.41). Any $M \times M$ symmetric square matrix has $M + (M^2 - M)/2 = M(M+1)/2$ independent components. In our case, M is the number of independent A (or B) indices, which in turn – because $\mu\nu$ (or $\alpha\beta$) are anti-symmetric – is equal to the number of independent components of a $d \times d$ anti-symmetric matrix; namely, $M = (d^2 - d)/2 = d(d-1)/2$. At this point, we have gathered Riemann has at most $M(M+1)/2 = (1/8)d(d-1)(d(d-1)+2)$ algebraically independent components. Finally, let us observe eq. (11.3.42) implies we have additional $\binom{d}{4} = d(d-1)(d-2)(d-3)/24$ constraint algebraic equations. This then allows us to conclude Riemann itself has, indeed,

$$\frac{1}{8} d(d-1)(d(d-1)+2) - \binom{d}{4} = \frac{d^2(d^2-1)}{12} \tag{11.3.45}$$

[123] See equations (11.4.8) through (11.4.10) below.

algebraically independent components. To prove this final assertion, note that, if any pair of indices $\nu\alpha\beta$ within the $[\dots]$ of eq. (11.3.42) are the same, we obtain a trivial $0 = 0$. Next, if we consider, say, $0 = R_{\mu[\mu\alpha\beta]}$ – since $\nu\alpha\beta$ are already fully anti-symmetric, it suffices to set only one of them equal to μ. Then, $0 = R_{\mu\mu[\alpha\beta]} - 2R_{\mu\alpha\mu\beta} - 2R_{\mu\beta\alpha\mu}$, where μ is not summed over and equations (9.6.33) and (11.3.41) were employed. Since $R_{\mu\mu\alpha\beta} = 0$ by eq. (11.3.41), we therefore have $R_{\mu\alpha\mu\beta} = -R_{\mu\beta\alpha\mu}$, which is already covered by eq. (11.3.41) itself. These considerations teach us: eq. (11.3.42) does not yield an independent identity from those in eq. (11.3.41) whenever any pair of indices are repeated (but not summed over). Since all 4 indices in eq. (11.3.42) must be distinct for it to be an independent constraint, it remains to be shown that the *order* of $\mu\nu\alpha\beta$ is immaterial. Again, because $\nu\alpha\beta$ are already fully anti-symmetric, we only need to consider swapping, say, μ with ν in eq. (11.3.42). But, eq. (11.3.42) is equivalent to eq. (11.3.43). By eq. (11.3.41) $R_{\mu\nu\alpha\beta} = -R_{\nu\mu\alpha\beta}$ and by a direct calculation $R_{\mu[\beta\alpha]\nu} = R_{\nu[\beta\alpha]\mu}$. In other words, the identity $R_{\mu\nu\alpha\beta} = R_{\mu[\beta\alpha]\nu}$ is entirely equivalent to $R_{\nu\mu\alpha\beta} = R_{\nu[\beta\alpha]\mu}$ and therefore the order of $\mu\nu\alpha\beta$ in eq. (11.3.42) does not matter.

Ricci tensor and scalar Because of the symmetries of Riemann in eq. (11.3.41), we have $g^{\alpha\beta}R_{\alpha\beta\mu\nu} = -g^{\alpha\beta}R_{\beta\alpha\mu\nu} = -g^{\beta\alpha}R_{\beta\alpha\mu\nu} = 0$; and likewise, $R_{\alpha\beta\mu}{}^{\mu} = 0$. In fact, the Ricci tensor is defined as the sole distinct and non-zero contraction of Riemann:

$$R_{\mu\nu} \equiv R^{\sigma}{}_{\mu\sigma\nu}. \tag{11.3.46}$$

This is a symmetric tensor, $R_{\mu\nu} = R_{\nu\mu}$, because of eq. (11.3.41); for,

$$R_{\mu\nu} = g^{\sigma\rho}R_{\sigma\mu\rho\nu} = g^{\rho\sigma}R_{\rho\nu\sigma\mu} = R_{\nu\mu}. \tag{11.3.47}$$

Its contraction yields the Ricci scalar

$$\mathcal{R} \equiv g^{\mu\nu}R_{\mu\nu}. \tag{11.3.48}$$

Problem 11.5. Commutator of covariant derivatives on higher rank tensor Prove that

$$\begin{aligned}
&[\nabla_\mu, \nabla_\nu]T^{\alpha_1\dots\alpha_N}{}_{\beta_1\dots\beta_M} \\
&= R^{\alpha_1}{}_{\sigma\mu\nu}T^{\sigma\alpha_2\dots\alpha_N}{}_{\beta_1\dots\beta_M} + R^{\alpha_2}{}_{\sigma\mu\nu}T^{\alpha_1\sigma\alpha_3\dots\alpha_N}{}_{\beta_1\dots\beta_M} + \dots + R^{\alpha_N}{}_{\sigma\mu\nu}T^{\alpha_1\dots\alpha_{N-1}\sigma}{}_{\beta_1\dots\beta_M} \\
&- R^{\sigma}{}_{\beta_1\mu\nu}T^{\alpha_1\dots\alpha_N}{}_{\sigma\beta_2\dots\beta_M} - R^{\sigma}{}_{\beta_2\mu\nu}T^{\alpha_1\dots\alpha_N}{}_{\beta_1\sigma\beta_3\dots\beta_M} - \dots - R^{\sigma}{}_{\beta_M\mu\nu}T^{\alpha_1\dots\alpha_N}{}_{\beta_1\dots\beta_{M-1}\sigma}.
\end{aligned} \tag{11.3.49}$$

Also verify that

$$[\nabla_\alpha, \nabla_\beta]\varphi = 0, \tag{11.3.50}$$

where φ is a scalar. □

Problem 11.6. Bianchi identities II If $[A, B] \equiv AB - BA$, can you show that the differential operator

$$[\nabla_\alpha, [\nabla_\beta, \nabla_\delta]] + [\nabla_\beta, [\nabla_\delta, \nabla_\alpha]] + [\nabla_\delta, [\nabla_\alpha, \nabla_\beta]] \tag{11.3.51}$$

is actually zero? (Hint: Just expand out the commutators.) Why does that imply

$$\nabla_{[\alpha}R^{\mu\nu}{}_{\beta\delta]} = 0? \tag{11.3.52}$$

Using this result, show that

$$\nabla_\sigma R^{\sigma\beta}{}_{\mu\nu} = \nabla_{[\mu} R^\beta{}_{\nu]}. \tag{11.3.53}$$

The *Einstein tensor* is defined as

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R}. \tag{11.3.54}$$

From eq. (11.3.53) can you show the divergence-less property of the Einstein tensor, i.e.,

$$\nabla^\mu G_{\mu\nu} = \nabla^\mu \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} \right) = 0? \tag{11.3.55}$$

This is an important property when understanding Einstein's equations of General Relativity, with a non-zero cosmological constant Λ,

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G_{\rm N} T_{\mu\nu}, \tag{11.3.56}$$

where $T_{\mu\nu}$ encodes the energy-momentum-stress-shear of matter. By employing eq. (11.3.55) and metric compatibility, we see that taking the divergence of eq. (11.3.56) leads us to the conservation of energy-momentum-shear-stress: $\nabla^\mu T_{\mu\nu} = 0$. □

Remark: Christoffel vs. Riemann Before moving on to geodesics, I wish to emphasize the basic facts that, given a space(time) metric:

> Non-zero Christoffel symbols do not imply non-zero space(time) curvature. Non-trivial space(time) curvature at a given spacetime location does not imply non-trivial Christoffel symbols.

The confusion that Christoffel symbols are somehow intrinsically tied to *curved* space(time)s is likely linked to the fact that one often encounters them for the first time while taking a course on General Relativity. Note, however, that while the Christoffel symbols of flat space(time) in Cartesian coordinates are trivial; they become non-zero when written in spherical coordinates – recall Problem (9.38). On the other hand, in a locally flat or Fermi-Normal-Coordinate system – see equations (9.2.1) in the previous Chapter; and (11.4.8)–(11.4.10) below – the Christoffel symbols vanish at $\vec{y}_0$ in the former and along the freely falling geodesic $y^\alpha = (\tau, \vec{y})$ in the latter.

Geodesics As already noted, even in flat spacetime, $\mathrm{d}s^2$ is not positive-definite (cf. (10.1.2)), unlike its purely spatial counterpart. Therefore, when computing the distance along a line in spacetime $z^\mu(\lambda)$, with boundary values $z(\lambda_1) \equiv x'$ and $z(\lambda_2) \equiv x$, we need to take the square root of its absolute value:

$$s = \int_{\lambda_1}^{\lambda_2} \left| g_{\mu\nu}\left(z(\lambda)\right) \frac{\mathrm{d}z^\mu(\lambda)}{\mathrm{d}\lambda} \frac{\mathrm{d}z^\nu(\lambda)}{\mathrm{d}\lambda} \right|^{1/2} \mathrm{d}\lambda. \tag{11.3.57}$$

A geodesic in curved spacetime that joins two points x and x' is a path that extremizes the distance between them. Using an affine parameter to describe the geodesic, i.e., using a λ such that $\sqrt{|g_{\mu\nu}\dot{z}^\mu\dot{z}^\nu|}$ = constant, this amounts to imposing the principle of stationary action on *Synge's world function* (recall eq. (9.5.49)):

$$\sigma(x,x') \equiv \frac{1}{2}(\lambda_2-\lambda_1)\int_{\lambda_1}^{\lambda_2} g_{\alpha\beta}\left(z(\lambda)\right)\frac{\mathrm{d}z^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}z^\beta}{\mathrm{d}\lambda}\mathrm{d}\lambda, \tag{11.3.58}$$

$$z^\mu(\lambda_1) = x'^\mu, \qquad z^\mu(\lambda_2) = x^\mu. \tag{11.3.59}$$

When evaluated on geodesics, eq. (11.3.58) is half the square of the geodesic distance between x and x'. The curved spactime geodesic equation in affine-parameter form which follows from eq. (11.3.58), is

$$\frac{D^2 z^\mu}{\mathrm{d}\lambda^2} \equiv \frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}z^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}z^\beta}{\mathrm{d}\lambda} = 0. \tag{11.3.60}$$

Problem 11.7. Choice of 'units' for affine parameter Show that eq. (11.3.60) takes the same form under re-scaling and constant shifts of the parameter λ. That is, if

$$\lambda = a\lambda' + b, \tag{11.3.61}$$

for constants a and b, then eq. (11.3.60) becomes

$$\frac{D^2 z^\mu}{\mathrm{d}\lambda'^2} \equiv \frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\lambda'^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}z^\alpha}{\mathrm{d}\lambda'}\frac{\mathrm{d}z^\beta}{\mathrm{d}\lambda'} = 0. \tag{11.3.62}$$

For the timelike and spacelike cases, this is telling us that proper time and proper length are respectively only defined up to an overall re-scaling and an additive shift. In other words, both the base units and its 'zero' may be altered at will. □

The discussion in §(9.5) had already informed us, the Lagragian associated with eq. (11.3.58),

$$L_{\mathrm{g}} \equiv \frac{1}{2}g_{\mu\nu}(z(\lambda))\dot{z}^\mu\dot{z}^\nu, \qquad \dot{z}^\mu \equiv \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda}, \tag{11.3.63}$$

not only oftentimes provides a more efficient means of computing the Christoffel symbols, it is a constant of motion. Unlike the curved space case, however, this Lagrangian L_{g} can now be positive, zero, or negative. Because the affine parameter is only defined up to a constant shift and re-scaling, we have for $\lambda \equiv a\lambda'$ ($a \equiv$ constant),

$$L_g[\lambda] = \frac{1}{2}g_{\mu\nu}(z(\lambda))\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda}\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda} = \frac{1}{2}g_{\mu\nu}(z(\lambda'))\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda'}\frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda'}\frac{1}{a^2} = \frac{L_g[\lambda']}{a^2}. \tag{11.3.64}$$

By choosing a appropriately, we may thus deduce the following.

- If $\dot{z}^\mu$ is timelike, then by choosing the affine parameter to be proper time $\mathrm{d}\lambda\sqrt{g_{\mu\nu}\dot{z}^\mu\dot{z}^\nu} = \mathrm{d}\tau$, we see that the Lagrangian is then set to $L_{\mathrm{g}} = 1/2$.
- If $\dot{z}^\mu$ is spacelike, then by choosing the affine parameter to be proper length $\mathrm{d}\lambda\sqrt{|g_{\mu\nu}\dot{z}^\mu\dot{z}^\nu|} = \mathrm{d}\ell$, we see that the Lagrangian is then set to $L_{\mathrm{g}} = -1/2$.

- If $\dot{z}^\mu$ is null, then the Lagrangian is zero: $L_{\rm g} = 0$. Since both sides of eq. (11.3.64) will remain zero under re-scaling, there is always a freedom to rescale the affine parameter by a constant:

$$L_g[\lambda] = 0 = L_g[\lambda'], \tag{11.3.65}$$

whenever $\lambda = (\text{constant}) \times \lambda'$.

Max or Min? A timelike path may be approximated as a series of jagged *null* paths. (Drawing a figure here would help.) This indicates there cannot be a non-zero lower bound to the proper time between two fixed spacetime events, since we may simply deform the timelike path closer and closer to these jagged null ones and hence approach (from above) zero proper time.[124] As long as the geodesic is unique, an extremum cannot be an inflection point because that would mean the proper time has no maximum; but along a timelike path $z^\mu(\lambda)$ in a metric $g_{\mu\nu}$, with spacetime coordinates x^μ and orthonormal frame fields defined through $g_{\mu\nu} = \eta_{\alpha\beta}\varepsilon^{\widehat{\alpha}}{}_\mu \varepsilon^{\widehat{\beta}}{}_\nu$, the proper time must be bounded by

$$\int \mathrm{d}\tau = \int \sqrt{(\mathrm{d}z^{\widehat{0}})^2 - \delta_{ij}\mathrm{d}z^{\widehat{i}}\mathrm{d}z^{\widehat{j}}} \leq \int |\mathrm{d}z^{\widehat{0}}|, \qquad \mathrm{d}z^{\widehat{\mu}} \equiv \varepsilon^{\widehat{\mu}}{}_\alpha \mathrm{d}z^\alpha. \tag{11.3.66}$$

Therefore, at least locally:[125]

A timelike extremum must be a maximum proper time.

A spacelike path cannot, in fact, be approximated as a series of jagged null paths. (Drawing a figure here would help.) But any spacelike path can be increased in length by simply adding more wiggles to it, say. As long as the geodesic is unique, an inflection point should not exist, since that would mean the proper length can approach zero for any two end points – a statement that cannot be true even in flat spacetime. Therefore, at least locally:[126]

A spacelike extremum must be a minimum length.

Haniltonian Dynamics of Geodesics In §(9.5), we also delineated an alternate but equivalent Hamiltonian formulation for geodesic motion. The conjugate momentum p_μ to the coordinate z^μ is

$$p_\mu \equiv \frac{\partial L_g}{\partial \dot{z}^\mu} = g_{\mu\nu}\dot{z}^\nu. \tag{11.3.67}$$

The Hamiltonian is

$$H(z,p) = \frac{1}{2} g^{\alpha\beta}\left(z(\lambda)\right) p_\alpha(\lambda) p_\beta(\lambda); \tag{11.3.68}$$

[124] A version of this argument may be found in Carroll's lecture notes [27].

[125] Global topology matters. Minkowski spacetime may be 'compactified' in time by identifying $(0,\vec{x})$ with $(T,\vec{x})$; i.e., time is now periodic, with period T. The geodesics linking $(0,\vec{x})$ to $(T,\vec{x})$ are $z^\mu(0 \leq \lambda \leq 1) = (0,\vec{x}) + \lambda(T,\vec{0})$ and $z^\mu(0 \leq \lambda \leq 1) = (0,\vec{x})$.

[126] Globally, topology matters. For instance, on a 2−sphere, the geodesic joining two points is not unique because it can either be the smaller or larger arc. In this case, the extremums are, respectively, the local minimum and maximum.

and the associated Hamilton's equations are

$$\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda} = \frac{\partial H}{\partial p_\mu} = g^{\mu\nu} p_\nu, \tag{11.3.69}$$

$$\frac{\mathrm{d}p_\mu}{\mathrm{d}\lambda} = -\frac{\partial H}{\partial z^\mu} = -\frac{1}{2}(\partial_\mu g^{\alpha\beta}) p_\alpha p_\beta. \tag{11.3.70}$$

Together, equations (11.3.69) and (11.3.70) are equivalent to eq. (11.3.60).

Example In flat spacetime, the Hamiltonian would read

$$H = \frac{1}{2}\eta^{\alpha\beta} p_\alpha p_\beta. \tag{11.3.71}$$

Since $\eta^{\alpha\beta}$ is a constant matrix, we infer from equations (11.3.69) and (11.3.70) the conservation of linear momentum:

$$\dot{z}^\mu = \eta^{\mu\nu} p_\nu = p^\mu, \tag{11.3.72}$$

$$\dot{p}_\mu = 0. \tag{11.3.73}$$

Formal solution to geodesic equation We may re-write eq. (11.3.60) into an integral equation by simply integrating both sides with respect to the affine parameter λ:

$$v^\mu(\lambda) = v^\mu(\lambda_1) - \int_{z(\lambda_1)}^{z(\lambda)} \Gamma^\mu{}_{\alpha\beta} v^\alpha \mathrm{d}z^\beta; \tag{11.3.74}$$

where $v^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\lambda$; the lower limit is $\lambda = \lambda_1$; and we have left the upper limit indefinite. The integral on the right hand side can be viewed as an integral operator acting on the tangent vector at $v^\alpha(z(\lambda))$. By iterating this equation infinite number of times – akin to the Born series expansion in quantum mechanics – it is possible to arrive at a formal (as opposed to explicit) solution to the geodesic equation.

Problem 11.8. Synge's World Function In Minkowski Verify that Synge's world function (cf. (11.3.58)) in Minkowski spacetime is

$$\bar{\sigma}(x, x') = \frac{1}{2}(x - x')^2 \equiv \frac{1}{2}\eta_{\mu\nu}(x - x')^\mu (x - x')^\nu, \tag{11.3.75}$$

$$(x - x')^\mu \equiv x^\mu - x'^\mu. \tag{11.3.76}$$

Explain the Poincaré invariance of the world function.

Hint: If we denote the geodesic $z^\mu(0 \leq \lambda \leq 1)$ joining x' to x in Minkowski spacetime, verify that the solution is

$$z^\mu(0 \leq \lambda \leq 1) = x'^\mu + \lambda(x - x')^\mu. \tag{11.3.77}$$

This is, of course, the 'constant velocity' solution of classical kinematics if we identify λ as a fictitious time. □

Problem 11.9. Geodesic Vector Fields Let $v^\mu(x)$ be a vector field defined throughout a given spacetime. Show that the geodesic equation (11.3.60) follows from

$$v^\sigma \nabla_\sigma v^\mu = 0, \tag{11.3.78}$$

i.e., v^μ is parallel transported along itself – provided we recall the 'velocity flow' interpretation of a vector field:

$$v^\mu\left(z(s)\right) = \frac{\mathrm{d}z^\mu}{\mathrm{d}s}. \tag{11.3.79}$$

Parallel transport preserves norm-squared The metric compatibility condition in eq. (11.3.16) obeyed by the covariant derivative ∇_α can be thought of as the requirement that the norm-squared $v^2 \equiv g_{\mu\nu}v^\mu v^\nu$ of a geodesic vector (v^μ subject to eq. (11.3.78)) be preserved under parallel transport. Can you explain this statement using the appropriate equations?

Non-affine form of geodesic equation Suppose instead

$$v^\sigma \nabla_\sigma v^\mu = \kappa v^\mu, \tag{11.3.80}$$

where κ is constant. This is the more general form of the geodesic equation, where the parameter λ is not an affine one. Nonetheless, by considering the quantity $v^\sigma\nabla_\sigma(v^\mu/(v_\nu v^\nu)^p)$, for some real number p, show how eq. (11.3.80) can be transformed into the form in eq. (11.3.78); that is, identify an appropriate v'^μ such that

$$v'^\sigma \nabla_\sigma v'^\mu = 0. \tag{11.3.81}$$

You should comment on how this re-scaling fails when v^μ is null.

Starting from the finite distance integral

$$s \equiv \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda' \sqrt{|g_{\mu\nu}(z(\lambda'))\dot{z}^\mu \dot{z}^\nu|}, \qquad \dot{z}^\mu \equiv \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda'}, \tag{11.3.82}$$

$$z^\mu(\lambda_1) = x', \qquad z^\mu(\lambda_2) = x; \tag{11.3.83}$$

show that demanding s be extremized leads to the non-affine geodesic equation

$$\ddot{z}^\mu + \Gamma^\mu{}_{\alpha\beta}\dot{z}^\alpha \dot{z}^\beta = \dot{z}^\mu \frac{\mathrm{d}}{\mathrm{d}\lambda'} \ln\sqrt{g_{\alpha\beta}\dot{z}^\alpha\dot{z}^\beta}. \tag{11.3.84}$$

In the next problem, you shall discover how to convert eq. (11.3.84) into its affinely parametrized form. □

Problem 11.10. Relation Between Affine and Non-Affine Parameters Starting from the non-affine form in eq. (11.3.80) of the geodesic equation

$$\frac{D^2 z^\mu}{\mathrm{d}\lambda'^2} \equiv \frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\lambda'^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}z^\alpha}{\mathrm{d}\lambda'}\frac{\mathrm{d}z^\beta}{\mathrm{d}\lambda'} = \kappa(\lambda')\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda'}, \tag{11.3.85}$$

show that the change-of-variables $\lambda' = \lambda'(\lambda)$ yields

$$\frac{D^2 z^\mu}{\mathrm{d}\lambda'^2} = \left(\frac{\mathrm{d}\lambda}{\mathrm{d}\lambda'}\right)^2 \frac{D^2 z^\mu}{\mathrm{d}\lambda^2} + \frac{\mathrm{d}^2\lambda}{\mathrm{d}\lambda'^2}\frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda}. \tag{11.3.86}$$

If we wish to recover the affinely parametrized geodesic equation $D^2 z^\mu/\mathrm{d}\lambda^2 = 0$ in eq. (11.3.60) argue that

$$\frac{\mathrm{d}^2\lambda}{\mathrm{d}\lambda'^2} = \kappa(\lambda')\frac{\mathrm{d}\lambda}{\mathrm{d}\lambda'}; \tag{11.3.87}$$

and, in turn, the solution is

$$\lambda(\lambda') = \lambda(\lambda'_0) + \frac{\mathrm{d}\lambda'(\lambda'_0)}{\mathrm{d}\lambda'}\int_{\lambda'_0}^{\lambda'}\mathrm{d}s' \exp\left(\int_{\lambda'_0}^{s'}\kappa(s)\mathrm{d}s\right), \tag{11.3.88}$$

for some 'initial' λ'_0. (Note that this transformation from non-affine λ' to affine λ in eq. (11.3.88) *does not* assume $\mathrm{d}z^\mu/\mathrm{d}\lambda$ is timelike, spacelike, or null.) For the specific case in eq. (11.3.84), explain why

$$\lambda(\lambda') = \lambda(\lambda' = \lambda'_0) + \frac{\mathrm{d}\lambda(\lambda' = \lambda'_0)}{\mathrm{d}\lambda}\int_{\lambda'_0}^{\lambda'}\mathrm{d}s\frac{\sqrt{g_{\mu\nu}(z(s))\dot{z}^\mu(s)\dot{z}^\nu(s)}}{\sqrt{g_{\alpha\beta}(z(\lambda'_0))\dot{z}^\alpha(\lambda'_0)\dot{z}^\beta(\lambda'_0)}}. \tag{11.3.89}$$

□

Geodesic Vector Fields in Cosmology An elementary example of a geodesic vector field occurs in cosmology. There is evidence that we live in a universe described by the following metric at the very largest length scales:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}. \tag{11.3.90}$$

Let us demonstrate that

$$U^\mu = \delta^\mu_0 \tag{11.3.91}$$

is in fact a timelike geodesic vector field. Firstly,

$$g_{\mu\nu}U^\mu U^\nu = g_{00} = 1 > 0. \tag{11.3.92}$$

Next, keeping in mind eq. (11.3.91), we compute

$$U^\mu\nabla_\mu U^\alpha = \nabla_0 U^\alpha = \partial_0\delta^\alpha_0 + \Gamma^\alpha{}_{00} \tag{11.3.93}$$

$$= \frac{1}{2}g^{\alpha\sigma}\left(\partial_0 g_{0\sigma} + \partial_0 g_{0\sigma} - \partial_\sigma g_{00}\right) = g^{\alpha 0}\partial_0 g_{00} = 0. \tag{11.3.94}$$

The interpretation is that $U^\mu = \delta^\mu_0$ is tangent to the worldlines of observers 'at rest' with the expanding universe, since the spatial velocities are zero. Furthermore, we may infer that (cf. eq. (9.7.16))

$$H_{\mu\nu} = g_{\mu\nu} - U_\mu U_\nu \tag{11.3.95}$$

is the metric orthogonal to U^μ itself; namely,

$$H_{\mu\nu}U^\nu = U_\mu - U_\mu(U_\nu U^\nu) = 0 \tag{11.3.96}$$

because eq. (11.3.92) tells us $U_\nu U^\nu = 1$. The space orthogonal to U_μ reads

$$\mathrm{d}\ell^2 = -H_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = -(\mathrm{d}t^2 - a^2\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} - (U_\mu\mathrm{d}x^\mu)^2) = a(t)^2\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}, \tag{11.3.97}$$

as $(U_\mu\mathrm{d}x^\mu)^2 = (\delta^0_\mu\mathrm{d}x^\mu)^2 = \mathrm{d}t^2$. It is expanding/contracting, with relative $t-$dependent size governed by $a(t)$.

Problem 11.11. Geodesic Flow Preserves Timelike, Spacelike or Null Character Let v^α be a geodesic vector field. Prove that, if v^α is timelike, null, or spacelike at a given spacetime location z, it remains timelike, null, or spacelike along the entire integral curve passing through z. Hint: Compute $v^\sigma \nabla_\sigma v^2 \equiv v^\sigma \nabla_\sigma (g_{\alpha\beta} v^\alpha v^\beta)$. You should find that this result hold for both affinely and non-affinely parametrized v^α. $\square$

Problem 11.12. Null Geodesics and Weyl Transformations Suppose two geometries $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$ are related via a Weyl transformation

$$g_{\mu\nu}(x) = \Omega(x)^2 \bar{g}_{\mu\nu}(x). \tag{11.3.98}$$

We note that, as long as $\Omega \neq 0$, then the null constraint $\bar{g}_{\mu\nu} q^\mu q^\nu = 0$ is satisfied with respect to $\bar{g}_{\mu\nu}$ iff the constraint $g_{\mu\nu} q^\mu q^\nu = \Omega^2 \bar{g}_{\mu\nu} q^\mu q^\nu = 0$ is satisfied with respect to its Weyl-transformed counterpart $g_{\mu\nu}$. This suggests the null geodesics in $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$ are related.

Consider the null geodesic equation in the geometry $g_{\mu\nu}(x)$,

$$k^\sigma \nabla_\sigma k^\mu = 0, \qquad g_{\mu\nu} k^\mu k^\nu = 0 \tag{11.3.99}$$

where ∇ is the covariant derivative with respect to $g_{\mu\nu}$; as well as the null geodesic equation in $\bar{g}_{\mu\nu}(x)$,

$$\bar{k}^\sigma \overline{\nabla}_\sigma \bar{k}^\mu = 0, \qquad \bar{g}_{\mu\nu} \bar{k}^\mu \bar{k}^\nu = 0; \tag{11.3.100}$$

where $\overline{\nabla}$ is the covariant derivative with respect to $\bar{g}_{\mu\nu}$. Show that

$$\bar{k}^\mu = \Omega^2 \cdot k^\mu. \tag{11.3.101}$$

Hint: First show that the Christoffel symbol $\overline{\Gamma}^\mu{}_{\alpha\beta}[\bar{g}]$ built solely out of $\bar{g}_{\mu\nu}$ is related to $\Gamma^\mu{}_{\alpha\beta}[g]$ built out of $g_{\mu\nu}$ through the relation

$$\Gamma^\mu{}_{\alpha\beta}[g] = \bar{\Gamma}^\mu{}_{\alpha\beta}[\bar{g}] + \delta^\mu_{\{\beta} \overline{\nabla}_{\alpha\}} \ln \Omega - \bar{g}_{\alpha\beta} \overline{\nabla}^\mu \ln \Omega. \tag{11.3.102}$$

Then remember to use the constraint $g_{\mu\nu} k^\mu k^\nu = 0 = \bar{g}_{\mu\nu} \bar{k}^\mu \bar{k}^\nu$.

A spacetime is said to be conformally flat if it takes the form

$$g_{\mu\nu}(x) = \Omega(x)^2 \eta_{\mu\nu}. \tag{11.3.103}$$

Solve the null geodesic equation explicitly in such a spacetime. $\square$

Problem 11.13. Shapiro Time Delay in Static Newtonian Spacetimes As a simple application of Synge's world function, let us consider an isolated (non-relativistic) astrophysical system centered at $\vec{x} = 0$. We shall assume its gravity is weak, and may be described by a static Newtonian potential Φ, through the metric

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\Phi(\vec{x}) \delta_{\mu\nu}. \tag{11.3.104}$$

Within 4D Linearized General Relativity, we will find that the Newtonian potential is sourced by the astrophysical energy density ρ via Poisson's equation:

$$\vec{\nabla}^2 \Phi(\vec{x}) = 4\pi G_{\mathrm{N}} \rho(\vec{x}). \tag{11.3.105}$$

In §(12) below, we shall solve this equation through the Euclidean Green's function.

$$\Phi(\vec{x}) = -G_{\mathrm{N}} \int_{\mathbb{R}^3} \frac{\rho(\vec{x}')}{|\vec{x} - \vec{x}'|} \mathrm{d}^3\vec{x}'. \tag{11.3.106}$$

Let us shoot a beam of light from one side of the astrophysical system to opposite side, through its central region where Φ is non-trivial. Assume the emitter and receiver are at rest, respectively at $\vec{x} = \vec{x}_e$ and $\vec{x} = \vec{x}_r$; and they are far away enough that Φ is negligible, so that to a good approximation, the global time t refers to their proper times. Our primary goal is to compute the elapsed time between receipt t_r and emission t_e.

In §(11.8.1) below, we develop perturbation theory off a flat background spacetime. In particular, we show that the world function up to first order in perturbation takes the form in eq. (11.8.88). Exploit it to show that, by virtue of being a null signal,

$$T^2 = R^2 - 2(T^2 + R^2) \int_0^1 \Phi\left(\vec{x}_e + \lambda(\vec{x}_r - \vec{x}_e)\right) \mathrm{d}\lambda + \mathcal{O}\left(\Phi^2\right). \tag{11.3.107}$$

where

$$T \equiv t_r - t_e \qquad \text{and} \qquad R \equiv |\vec{x}_r - \vec{x}_e|. \tag{11.3.108}$$

According to eq. (11.3.107), T^2 goes as R^2 plus an order Φ correction. Therefore, replacing the T^2 on the right hand side of eq. (11.3.107) with R^2 would incur an error of order Φ^2. Explain why the time elapsed $T = t_r - t_e$ is thus

$$T = R\left(1 - 2\int_0^1 \Phi\left(\vec{x}_e + \lambda(\vec{x}_r - \vec{x}_e)\right) \mathrm{d}\lambda\right) + \mathcal{O}\left(\Phi^2\right). \tag{11.3.109}$$

Why is this a time *delay*? Hint: What sign is the gravitational potential Φ? You may notice this is a time *delay*, because energy density is strictly *positive*!

This Shapiro time delay was first measured in practice by bouncing radio waves from Earth off Mercury and Venus during their superior conjunctions; see [40, 41, 42]. To date, the most precise Shapiro time-delay measurement is from the Doppler tracking of the Cassini spacecraft; see §4.1.2 of [39]. □

Problem 11.14. Verify that the non-zero Christoffel symbols of eq. (11.3.104) are given by

$$\Gamma^t{}_{ti} = \Gamma^t{}_{it} = \partial_i\Phi, \tag{11.3.110}$$

$$\Gamma^j{}_{ji} = \Gamma^j{}_{ij} = -\partial_i\Phi \qquad \text{(No sum over } j\text{)} \tag{11.3.111}$$

$$\Gamma^x{}_{tt} = \Gamma^x{}_{yy} = \Gamma^x{}_{zz} = \partial_x\Phi, \tag{11.3.112}$$

$$\Gamma^y{}_{tt} = \Gamma^y{}_{xx} = \Gamma^y{}_{zz} = \partial_y\Phi, \tag{11.3.113}$$

$$\Gamma^z{}_{tt} = \Gamma^z{}_{xx} = \Gamma^z{}_{yy} = \partial_z\Phi, \tag{11.3.114}$$

up to first order in the Newtonian potential Φ. □

Problem 11.15. Fermat's Principle and Effective Refractive Index Consider the static weak-field Newtonian spacetime

$$\mathrm{d}s^2 = (1 + 2\Phi(\vec{x}))\,\mathrm{d}t^2 - (1 - 2\Phi(\vec{x}))\,\mathrm{d}\vec{x}^2, \tag{11.3.115}$$
$$|\Phi| \ll 1. \tag{11.3.116}$$

If an over-dot denotes a derivative with respect to an affine parameter λ, explain why we may exploit the $t-$independence of the metric to assert

$$(1 + 2\Phi)\dot{t} = 1. \tag{11.3.117}$$

Use this constraint to derive the null geodesic equation, accurate up to linear order in Φ:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left((1 - 2\Phi(\vec{x}))\,\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\right) = \partial_i\,(1 - 2\Phi) + \mathcal{O}(\Phi^2). \tag{11.3.118}$$

Now, consider instead the (coordinate) time elapsed between emission and detection of a given photon. Explain why, up to first order in Φ,

$$\Delta t = \int |\dot{\vec{x}}|\,(1 - 2\Phi(\vec{x}))\,\mathrm{d}\lambda' + \mathcal{O}(\Phi^2); \tag{11.3.119}$$

where $|\dot{\vec{x}}| \equiv |(\mathrm{d}\vec{x}/\mathrm{d}\lambda')^2|^{1/2}$. Fermat's principle states that the null path in such a static spacetime is one where the elapsed time Δt is minimized. Show that, in fact, minimizing Δt leads to eq. (11.3.118). Explain how $\mathrm{d}\lambda$ is related to $\mathrm{d}\lambda'$.

Refractive Index Compare equations (8.1.67) and (11.3.118). What is the effective refractive index of the static Newtonian spacetime at hand? $\square$

Synchronous Frame In a small enough region of spacetime, we may consider a family of freely-falling (timelike) observers and employ their proper time and spatial coordinates to be the spacetime coordinates of the metric itself. More specifically, let $Z^\mu = (\tau, Z^i)$ be the trajectory of a free-falling observer. Since we are trying to use its spatial coordinates to label spatial locations within the spacetime, Z^i itself may be parametrized by some coordinates $\vec{y}$, i.e., $Z^i = Z^i(\vec{y})$, in order to distinguish one observer's world line from another. On the other hand, let us assume the proper times of these observers may be synchronized. Therefore, the family is described by $Z^\mu = (\tau, Z^i(\vec{y}))$, where Z^i is τ (proper time) independent because, by virtue of being the spatial coordinate, it must necessarily be 'at rest'. And since τ is the proper time,

$$1 = g_{\mu\nu}\frac{\mathrm{d}Z^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}Z^\nu}{\mathrm{d}\tau} = g_{00}. \tag{11.3.120}$$

Moreover, since these are free falling observers, the geodesic equation must hold.

$$\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}Z^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}Z^\beta}{\mathrm{d}\tau} = \Gamma^\mu{}_{00} = 0 \tag{11.3.121}$$

Expanding out the Christoffel symbol,

$$2\partial_0 g_{0\mu} = \partial_\mu g_{00}. \tag{11.3.122}$$

When $\mu = 0$, the relation is a trivial $0 = 0$. When $\mu = i$, on the other hand, we deduce g_{0i} must be τ independent. At this point, we may gather

$$\mathrm{d}s^2 = \mathrm{d}\tau^2 + 2g_{0j}(\vec{y})\mathrm{d}y^j\mathrm{d}\tau + g_{ij}(\tau,\vec{y})\mathrm{d}y^i\mathrm{d}y^j \tag{11.3.123}$$

$$= \left(\mathrm{d}\left(\tau + \int^{\vec{y}} g_{0j}(\vec{y}')\mathrm{d}y'^j\right)\right)^2 + \left(g_{ij}(\tau,\vec{y}) - g_{0i}(\vec{y})g_{0j}(\vec{y})\right)\mathrm{d}y^i\mathrm{d}y^j. \tag{11.3.124}$$

By defining

$$t \equiv \tau + \int^{\vec{y}} g_{0j}(\vec{y}')\mathrm{d}y'^j, \tag{11.3.125}$$

$$H_{ij} \equiv g_{0i}(\vec{y})g_{0j}(\vec{y}) - g_{ij}(\tau,\vec{y}), \tag{11.3.126}$$

we have the metric put in the 'synchronous gauge', defined as the form of the metric where its 00 component is one and its $0i$ components are zero:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - H_{ij}(t,\vec{y})\mathrm{d}y^i\mathrm{d}y^j. \tag{11.3.127}$$

Problem 11.16. Synchronous Implies Freely Falling Now, starting from the 'synchronous gauge' metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 + g_{ij}(t,\vec{y})\mathrm{d}y^i\mathrm{d}y^j, \tag{11.3.128}$$

show that co-moving worldlines

$$Z^\mu = (t, \vec{Z}), \tag{11.3.129}$$

– where $\vec{Z}$ are t independent – satisfy the geodesic equation. Hence, together with the above discussion, we may summarize:

> The metric may be expressed in 'synchronous gauge' if and only if we use the (synchronized) proper times and spatial coordinates of some family of free-falling observers to label the spacetime events of the geometry itself.

The freedom to choose coordinates, means there are d choices in a $d-$dimensional spacetime. Notice, this is precisely the same number of components as $g_{\mu 0} = g_{0\mu}$ – namely, choosing the coordinates to be those of the free falling observer allows us to set $g_{00} = 1$ and $g_{0i} = 0 = g_{i0}$. $\Box$

11.4 Equivalence Principles and Geometry-Induced Tidal Forces

Weak Equivalence Principle, 'Free-Fall' and Gravity as a Non-Force The universal nature of gravitation – how it appears to act in the same way upon all material bodies independent of their internal composition – is known as the Weak Equivalence Principle. As we will see, the basic reason why the weak equivalence principle holds is because *everything* experiences the same spacetime $g_{\mu\nu}$.

Within non-relativistic physics, the acceleration of some mass M_1 located at $\vec{x}_1$, due to the Newtonian gravitational 'force' exerted by some other mass M_2 at $\vec{x}_2$, is given by

$$M_1\frac{\mathrm{d}^2\vec{x}_1}{\mathrm{d}t^2} = -\widehat{n}\frac{G_{\mathrm{N}}M_1M_2}{|\vec{x}_1 - \vec{x}_2|^2}, \qquad \widehat{n} \equiv \frac{\vec{x}_1 - \vec{x}_2}{|\vec{x}_1 - \vec{x}_2|}. \tag{11.4.1}$$

Strictly speaking the M_1 on the left hand side is the 'inertial mass', a characterization of the resistance – so to speak – of any material body to being accelerated by an external force. While the M_1 on the right hand side is the 'gravitational mass', describing the strength to which the material body interacts with the gravitational 'force'. Viewed from this perspective, the equivalence principle is the assertion that the inertial and gravitational masses are the same, so that the resulting motion does not depend on them:

$$\frac{\mathrm{d}^2\vec{x}_1}{\mathrm{d}t^2} = -\widehat{n}\frac{G_{\mathrm{N}}M_2}{|\vec{x}_1 - \vec{x}_2|^2}. \tag{11.4.2}$$

Similarly, the acceleration of body 2 due to the gravitational force exerted by body 1 is independent of M_2:

$$\frac{\mathrm{d}^2\vec{x}_2}{\mathrm{d}t^2} = +\widehat{n}\frac{G_{\mathrm{N}}M_1}{|\vec{x}_1 - \vec{x}_2|^2}. \tag{11.4.3}$$

This *Weak Equivalence Principle*[127] is one of the primary motivations that led Einstein to recognize gravitation as the manifestation of curved spacetime. The reason why inertial mass appears to be equal to its gravitational counterpart, is because material bodies now follow geodesics $z^\mu(\tau)$ in curved spacetimes:

$$a^\mu \equiv \frac{D^2 z^\mu}{\mathrm{d}\tau^2} \equiv \frac{\mathrm{d}^2 z^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}z^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}z^\beta}{\mathrm{d}\tau} = 0; \tag{11.4.4}$$

with the timelike constraint

$$g_{\mu\nu}\left(z(\lambda)\right)\frac{\mathrm{d}z^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}z^\nu}{\mathrm{d}\tau} > 0; \tag{11.4.5}$$

so that their motion only depends on the curved geometry itself and does not depend on their own mass. From this point of view, gravity is no longer a force. Now, if there *were* an external non-gravitational force f^μ, then the covariant Newton's second law for a system of mass M would read: $MD^2z^\mu/\mathrm{d}\tau^2 = f^\mu$.

Note that, strictly speaking, this "gravity-induced-dynamics-as-geodesics" is actually an idealization that applies for material bodies with no internal structure and whose proper sizes are very small compared to the length scale(s) associated with the geometric curvature itself. In reality, all physical systems have internal structure – non-trivial quadrupole moments, spin/rotation, etc. – and may furthermore be large enough that their full dynamics require detailed analysis to understand properly.

Newton vs. Einstein Observe that the Newtonian gravity of eq. (11.4.1) in an instantaneous force, in that the force on body 1 due to body 2 (or, vice versa) changes immediately when body 2 starts changing its position $\vec{x}_2$ – even though it is located at a finite distance away. However, Special Relativity tells us there ought to be an ultimate speed limit in Nature, i.e., no physical effect/information can travel faster than c. This apparent inconsistency between Newtonian gravity and Einstein's Special Relativity is of course a driving motivation that led

[127] See Will [39] arXiv: 1403.7377 for a review on experimental tests of various versions of the Equivalence Principle and other aspects of General Relativity. See also the Eöt-Wash Group.

Einstein to General Relativity. As we shall see shortly, by postulating that the effects of gravitation are in fact the result of residing in a curved spacetime, the Lorentz symmetry responsible for Special Relativity is recovered in any local "freely-falling" frame.

Massless particles Finally, this dynamics-as-geodesics also led Einstein to realize – if gravitation does indeed apply universally – that massless particles such as photons, i.e., electromagnetic waves, must also be influenced by the gravitational field too. This is a significant departure from Newton's law of gravity in eq. (11.4.1), which may lead one to suspect otherwise, since $M_{\text{photon}} = 0$. It is possible to justify this statement in detail, but we shall simply assert here – to leading order in the JWKB approximation (i.e., in the high frequency limit) photons in fact sweep out geodesics $z^\mu(\lambda)$ in curved spacetimes

$$a^\mu \equiv \frac{D^2 z^\mu}{\mathrm{d}\lambda^2} = 0; \tag{11.4.6}$$

subject to the *null* constraint – compare against eq. (11.4.5) –

$$g_{\mu\nu}\left(z(\lambda)\right) \frac{\mathrm{d}z^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}z^\nu}{\mathrm{d}\lambda} = 0. \tag{11.4.7}$$

Locally flat coordinates, Einstein Equivalence Principle & Symmetries We now come to one of the most important features of curved spacetimes. In the neighborhood of a timelike geodesic $y^\mu = (\tau, \vec{y})$, one may choose *Fermi normal coordinates* $x^\mu \equiv (\tau, \vec{x})$ such that spacetime appears flat up to distances of $\mathcal{O}(1/|\max R_{\mu\nu\alpha\beta}(y = (\tau, \vec{y}))|^{1/2})$; namely, $g_{\mu\nu} = \eta_{\mu\nu}$ plus corrections that begin at quadratic order in the displacement $\vec{x} - \vec{y}$:

$$g_{00}(\tau, \vec{x}) = 1 - R_{0a0b}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}\left((x-y)^3\right), \tag{11.4.8}$$

$$g_{0i}(\tau, \vec{x}) = -\frac{2}{3} R_{0aib}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}\left((x-y)^3\right), \tag{11.4.9}$$

$$g_{ij}(\tau, \vec{x}) = \eta_{ij} - \frac{1}{3} R_{iajb}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}\left((x-y)^3\right). \tag{11.4.10}$$

Here $x^0 = \tau$ is the time coordinate, and is also the proper time of the observer with the trajectory $y^\mu(\tau) = (\tau, \vec{y})$. (The $\vec{y}$ are fixed spatial coordinates; they do not depend on τ.) Suppose you were placed inside a closed box, so you cannot tell what's outside. Then provided the box is small enough, you will not be able to distinguish between being in "free-fall" in a gravitational field versus being in a completely empty Minkowski spacetime.[128]

As already alluded to in the 'Newton vs. Einstein' discussion above, just as the rotation and translation symmetries of flat Euclidean space carried over to a small enough region of curved spaces – the FNC expansion of equations (11.4.8) through (11.4.10) indicates that, within the spacetime neighborhood of a freely-falling observer, any curved spacetime is Lorentz and spacetime-translation symmetric.

Summary Physically speaking, in a freely falling frame $\{x^\mu\}$ – i.e., centered along a timelike geodesic at $x = y$ – physics in a curved spacetime is the same as

[128] The primary difference between eq. (9.2.1) and equations (11.4.8)-(11.4.10), apart from the fact that the former deals with curved spaces and the latter with curved spacetimes, is that the former only expresses the metric as a flat one at a single point, whereas the latter does so along the entire geodesic.

that in flat Minkowski spacetime up to corrections that go at least as

$$\epsilon_{\mathrm{E}} \equiv \left(\frac{\text{Length or inverse mass scale of system}}{\text{Length scale of the spacetime geometric curvature}}\right)^2 . \tag{11.4.11}$$

In particular, since the Christoffel symbols on the world line vanishes, the geodesic y^μ itself obeys the free-particle version of Newton's 2nd law: $\mathrm{d}^2 y^\mu/\mathrm{d}s^2 = 0$.

More generally, because material bodies (with mass > 0) sweep out geodesics according to eq. (11.4.4), they all fall at the same rate – independent of their gravitational or inertial masses. To quip: "acceleration is zero, gravity is not a force."

This is the essence of the equivalence principle that lead Einstein to recognize curved spacetime to be the setting to formulate his General Theory of Relativity.

Problem 11.17. In this problem, we will understand why we may always choose the frame where the spatial components $\vec{y}$ are time (i.e., $\tau-$)independent.

First use the geodesic equation obeyed by y^α to conclude $\mathrm{d}y^\alpha/\mathrm{d}\tau$ are constants. If τ refers to the proper time of the freely falling observer at $y^\alpha(\tau)$, then explain why

$$\eta_{\alpha\beta}\frac{\mathrm{d}y^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}y^\beta}{\mathrm{d}\tau} = 1. \tag{11.4.12}$$

Since this is a Lorentz invariant condition, $\{y^\alpha\}$ can be Lorentz boosted $y^\alpha \to \Lambda^\alpha{}_\mu y^\mu$ to the rest frame such that

$$\frac{\mathrm{d}y^\alpha}{\mathrm{d}s} \to \Lambda^\alpha{}_\mu \frac{\mathrm{d}y^\mu}{\mathrm{d}s} = \left(1, \vec{0}\right); \tag{11.4.13}$$

where the $\{\Lambda^\alpha{}_\mu\}$ themselves are time-independent. In other words, one can always find a frame where $\dot{y}^i = 0$; i.e., y^i are $\tau-$independent.

To sum: in the co-moving frame of the freely falling observer $y^\alpha(\tau)$, the only τ dependence in equations (11.4.8), (11.4.9) and (11.4.10) occur in the Riemann tensor. □

Problem 11.18. Verify that the coefficients in front of the Riemann tensor in equations (11.4.8), (11.4.9) and (11.4.10) are independent of the spacetime dimension. That is, starting with

$$g_{00}(x) = 1 - A \cdot R_{0a0b}(\tau) \cdot (x-y)^a (x-y)^b + \mathcal{O}\left((x-y)^3\right), \tag{11.4.14}$$

$$g_{0i}(x) = -B \cdot R_{0aib}(\tau) \cdot (x-y)^a (x-y)^b + \mathcal{O}\left((x-y)^3\right), \tag{11.4.15}$$

$$g_{ij}(x) = \eta_{ij} - C \cdot R_{iajb}(\tau) \cdot (x-y)^a (x-y)^b + \mathcal{O}\left((x-y)^3\right), \tag{11.4.16}$$

where A, B, C are unknown constants, recover the Riemann tensor at $x = y$. Hint: the calculation of R_{0ijk} and R_{abij} may require the Bianchi identity $R_{0[ijk]} = 0$.

Note: This problem is not meant to be a derivation of the Fermi normal expansion in equations (11.4.8), (11.4.9), and (11.4.10) – for that, see Poisson [22] §1.6 – but merely a consistency check. □

Fermi versus Riemann Normal Coordinates The Riemann normal coordinate system $\{y^\alpha\}$ version of eq. (9.2.1) but in curved spacetimes reads

$$g_{\mu\nu}(y \to y_0) = \eta_{\mu\nu} - \frac{1}{3}R_{\mu\alpha\nu\beta}(y_0) \cdot (y-y_0)^\alpha (y-y_0)^\beta + \mathcal{O}\left((y-y_0)^3\right). \tag{11.4.17}$$

This is to be contrasted with equations (11.4.8), (11.4.9), and (11.4.10). The latter holds along the entire 'free-falling' geodesic; where eq. (11.4.17) only holds in the neighborhood around $y \approx y_0$. In particular, the Riemann tensor in eq. (11.4.17) should be viewed as a constant; while the Riemann in equations (11.4.8), (11.4.9), and (11.4.10) is a function of time, since curvature can change along the geodesic.

Problem 11.19. Gravitational force in a (3+1)D weak gravitational field Consider the following metric written in a (3+1)D Cartesian $x^\mu = (t, \vec{x})$ coordinates.

$$g_{\mu\nu}(t,\vec{x})\mathrm{d}x^\mu \mathrm{d}x^\nu = (1 + 2\Phi(t,\vec{x}))\,\mathrm{d}t^2 - (1 - 2\Phi(t,\vec{x}))\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x} + 2A_i(t,\vec{x})\mathrm{d}t\mathrm{d}x^i, \tag{11.4.18}$$

where $|\Phi|, |A_i| \ll 1$. Starting from the non-affine form of the action principle for the test mass M moving in a bound orbit,

$$\begin{aligned} S_{\mathrm{pp}} &= -M\int_{t_1}^{t_2}\mathrm{d}t\sqrt{g_{\mu\nu}v^\mu v^\nu}, \qquad\qquad v^\mu \equiv \dot{z}^\mu \equiv \frac{\mathrm{d}z^\mu}{\mathrm{d}t} \\ &= -M\int_{t_1}^{t_2}\mathrm{d}t\sqrt{1 - \vec{v}^2 + 2\Phi(1+\vec{v}^2) + 2A_i v^i}, \qquad\qquad \vec{v}^2 \equiv \delta_{ij}\dot{z}^i\dot{z}^j; \end{aligned} \tag{11.4.19}$$

show that up to first order in Φ, A_i, and v^i, the geodesic equation reads

$$\ddot{z}^i = F^i_{\ \nu}(1,\vec{v})^\nu = E^i + (\vec{v}\times\vec{B})^i + 3(\partial_t\Phi)v^i, \tag{11.4.20}$$

with an overdot representing a $t-$derivative and

$$E^i = F^{i0}, \qquad B^i = -\frac{1}{2}\epsilon_{ijk}F_{jk}; \tag{11.4.21}$$

where the rank-2 field strength built out of $A_\mu \equiv (\Phi, A_i)$ is

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{11.4.22}$$

Notice the resemblance of eq. (11.4.23) to the Lorentz force law of electromagnetism whenever Φ is $t-$independent – namely,

$$\ddot{x}^i = E^i + (\vec{v}\times\vec{B})^i. \tag{11.4.23}$$

These are the equations of *gravitoelectromagnetism.* Note, however, this expansion in formal in the sense that, while Φ and A_i may both be 'small'; Φ may still be significantly larger than A_i in, for instance, the case of linearized Einstein's equations sourced by a slowly rotating material body; i.e., do not use these equations without taking into account your specific physical application. □

Newtonian Gravity, Frame Dragging In eq. (11.4.23), when A_i is $t-$indepndent, $F^{i0} = -\partial_i\Phi$, and Newton's law of gravitation is captured within the 'electric field' term $\vec{E}$; i.e., force per unit mass is negative gradient of the gravitational potential Φ. On the other hand, the presence of the 'magnetic field' term $\vec{B}$ – as the curl of A_i – has no counterpart in Newtonian gravity, even in this non-relativistic limit. Moreover, $\vec{v}\times\vec{B}$ generates a force perpendicular to the velocity; and, in the case of linearized Einstein's equations sourced by a slowly rotating body, means a free-falling body would be 'dragged along' by the rotation effects contained within A_i – hence, the term 'frame dragging'.

Geodesic Deviation and Tidal Forces We now turn to the derivation of the *geodesic deviation* equation. Consider two geodesics that are infinitesimally close-by. Let both of them be parametrized by λ, so that we may connect one geodesic to the other at the same λ via an infinitesimal vector ξ^μ. We will denote the tangent vector to one of geodesics to be U^μ, such that

$$U^\sigma \nabla_\sigma U^\mu = 0. \tag{11.4.24}$$

Furthermore, we will assume that $[U, \xi] = 0$, i.e., U and ξ may be integrated to form a 2D coordinate system in the neighborhood of this pair of geodesics. Then, the acceleration of the deviation vector becomes

$$\begin{aligned} U^\alpha \nabla_\alpha \left(U^\beta \nabla_\beta \xi^\mu\right) &= U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^\mu \\ &= \nabla_U \nabla_U \xi^\mu = -R^\mu{}_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta. \end{aligned} \tag{11.4.25}$$

As its name suggests, this equation tells us how the deviation vector ξ^μ joining two infinitesimally displaced geodesics is accelerated by the presence of spacetime curvature through the Riemann tensor. If spacetime were flat, the acceleration will be zero: two initially parallel geodesics will remain so.

Moreover, for a small but macroscopic system, if U^μ is a timelike vector tangent to, say, the geodesic trajectory of its center-of-mass, the geodesic deviation equation (11.4.25) then describes *tidal forces* acting on it – via Newton's second law. In other words, the relative acceleration between the 'particles' that comprise the system – induced by spacetime curvature – would compete with the system's internal forces.[129] That the Riemann tensor can be viewed as the source of tidal forces, complements its closely related geometric role as the measure of the non-invariance of parallel transport of vectors around an infinitesimal closed loop.

Derivation of eq. (11.4.25) We start by noting $[\xi, U] = (\xi^\alpha \partial_\alpha U^\mu - U^\alpha \partial_\alpha \xi^\mu)\partial_\mu = 0$ translates to

$$\nabla_U \xi = \nabla_\xi U; \tag{11.4.26}$$

because $\nabla_\xi U^\mu = \xi^\sigma \partial_\sigma U^\mu + \Gamma^\mu{}_{\sigma\kappa} \xi^\sigma U^\kappa$ and $\nabla_U \xi^\mu = U^\sigma \partial_\sigma \xi^\mu + \Gamma^\mu{}_{\sigma\kappa} \xi^\sigma U^\kappa$; i.e., the Christoffel terms cancel due to the symmetry $\Gamma^\mu{}_{\alpha\beta} = \Gamma^\mu{}_{\beta\alpha}$. We then start with the geodesic equation $\nabla_U U^\mu = 0$ and act ∇_ξ upon it.

$$\nabla_\xi \nabla_U U^\mu = 0 \tag{11.4.27}$$

$$\nabla_U \underbrace{\nabla_\xi U^\mu}_{=\nabla_U \xi^\mu} + \left[\nabla_\xi, \nabla_U\right] U^\mu = 0 \tag{11.4.28}$$

$$\nabla_U \nabla_U \xi^\mu = -R^\mu{}_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta \tag{11.4.29}$$

On the last line, we have exploited the assumption that $[U, \xi] = 0$ to say $\left[\nabla_\xi, \nabla_U\right] U^\mu = R^\mu{}_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta$ – recall eq. (11.3.36).

[129] The first gravitational wave detectors were in fact based on measuring the tidal squeezing and stretching of solid bars of aluminum. They are known as "Weber bars", named after their inventor Joseph Weber.

Problem 11.20. Alternate Derivation of Geodesic Deivation Equation A less geometric but equally valid manner to derive eq. (11.4.25) is to appeal to the very definition of geodesic deviation. Suppose $y^\mu(\tau)$ and $y^\mu(\tau)+\xi^\mu(\tau)$ are nearby geodesics. That means the latter obeys the geodesic equation

$$\frac{\mathrm{d}^2(y^\mu(\tau)+\xi^\mu(\tau))}{\mathrm{d}\tau^2}+\Gamma^\mu{}_{\alpha\beta}(y+\xi)\frac{\mathrm{d}(y^\alpha(\tau)+\xi^\alpha(\tau))}{\mathrm{d}\tau}\frac{\mathrm{d}(y^\beta(\tau)+\xi^\beta(\tau))}{\mathrm{d}\tau}=0. \tag{11.4.30}$$

If the components ξ^μ may be considered 'small,' expand the above up to linear order in ξ^μ and show that

$$\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}\tau^2}+2\Gamma^\mu{}_{\alpha\beta}(y)\frac{\mathrm{d}\xi^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}y^\beta}{\mathrm{d}\tau}+\xi^\sigma\partial_\sigma\Gamma^\mu{}_{\alpha\beta}(y)\frac{\mathrm{d}y^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}y^\beta}{\mathrm{d}\tau}=0. \tag{11.4.31}$$

Now proceed to demonstrate that equations (11.4.25) and (11.4.31) are equivalent. □

Problem 11.21. Geodesic Deviation and FNC Argue that all the Christoffel symbols $\Gamma^\alpha{}_{\mu\nu}$ evaluated along the free-falling geodesic in equations (11.4.8)-(11.4.10), namely when $x=y$, vanish. Then argue that all the time derivatives of the Christoffel symbols vanish along y too: $\partial_\tau^{n\geq 1}\Gamma^\alpha{}_{\mu\nu}=0$. (Hints: Recall from Problem (9.34) that, specifying the first derivatives of the metric is equivalent to specifying the Christoffel symbols. Why is $\partial_\tau^{n\geq 1}g_{\alpha\beta}(x=y)=0$? Why is $\partial_\tau^{n\geq 1}\partial_i g_{\alpha\beta}(x=y)=0$?) Why does this imply, denoting $U^\mu\equiv \mathrm{d}y^\mu/\mathrm{d}\tau$, the geodesic equation

$$U^\nu\nabla_\nu U^\mu=\frac{\mathrm{d}U^\mu}{\mathrm{d}\tau}=0? \tag{11.4.32}$$

Next, evaluate the geodesic deviation equation in these Fermi Normal Coordinates (FNC) system. Specifically, show that

$$U^\alpha U^\beta\nabla_\alpha\nabla_\beta\xi^\mu=\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}\tau^2}=-R^\mu{}_{0\nu0}\xi^\nu. \tag{11.4.33}$$

Why does this imply, *if* the deviation vector is purely spatial at a given $s=s_0$, specifically $\xi^0(\tau_0)=0=\mathrm{d}\xi^0/\mathrm{d}\tau_0$, then it remains so for all time? (Hint: In an FNC system and on the world line of the free-falling observer, $R^0{}_{0\alpha\beta}=R_{00\alpha\beta}$. What do the (anti)symmetries of the Riemann tensor say about the right hand side?) □

Problem 11.22. A Common Error Eq. (11.4.33) says that the acceleration of the deviation vector within the FNC system is simply the ordinary one: i.e.,

$$U^\alpha U^\beta\nabla_\alpha\nabla_\beta\xi^\mu=\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}\tau^2}. \tag{11.4.34}$$

Thus, eq. (11.4.33) yields an intuitive interpretation, that a pair of nearby freely falling observers would sense there is a force acting between them (provided by the Riemann tensor), as though they were in flat spacetime. However, it appears to be a common error for gravitation textbooks to assert that eq. (11.4.34) holds more generally than in a FNC system, particularly when discussing how gravitational waves distort the proper distances between pairs of nearby free-falling test masses.

To this end, let us assume the metric at hand has been put in the synchronous gauge, defined to be the coordinate system where $g_{00} = g^{00} = 1$ and $g_{0i} = g^{0i} = 0$. Moreover, assume the spatial metric is slightly perturbed from the Euclidean one; namely,

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \mathrm{d}\tau^2 - (\delta_{ij} - h_{ij}(\tau, \vec{x}))\,\mathrm{d}x^i \mathrm{d}x^j, \qquad |h_{ij}| \ll 1, \qquad U^\mu = \delta^\mu_0. \tag{11.4.35}$$

Show that eq. (11.4.34) is no longer true; but up to first order in h_{ij} it reads instead

$$U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^\mu = \ddot{\xi}^\mu + \eta^{\mu j}\dot{\xi}^k \dot{h}_{jk} + \frac{1}{2}\eta^{\mu j}\xi^k \ddot{h}_{jk} + \mathcal{O}(h^2), \tag{11.4.36}$$

where all the overdot(s) are partial derivative(s) with respect to proper time τ. □

Problem 11.23. Tidal forces due to mass monopole of isolated body In this problem we will consider sprinkling test masses initially at rest on the surface of an imaginary sphere of very small radius r_ϵ, whose center is located far from that of a static isolated body whose stress tensor is dominated by its mass density $\rho(\vec{x})$. We will examine how these test masses will respond to the gravitational tidal forces exerted by ρ.

Assume that the weak field metric generated by ρ is given by eq. (11.4.18); it is possible to justify this statement by using the linearized Einstein's equations. Show that the vector field

$$U^\mu(t, \vec{x}) \equiv \delta^\mu_0(1 - \Phi(\vec{x})) - t\delta^\mu_i \partial_i \Phi(\vec{x}) \tag{11.4.37}$$

is a timelike geodesic up to linear order in the Newtonian potential Φ. This U^μ may be viewed as the tangent vector to the worldline of the observer who was released from rest in the $(t, \vec{x})$ coordinate system at $t = 0$. (To ensure this remains a valid perturbative solution we shall also assume $t/r \ll 1$.) Let $\xi^\mu = (\xi^0, \vec{\xi})$ be the deviation vector whose spatial components we wish to interpret as the small displacement vector joining the center of the imaginary sphere to its surface. Use the above U^α to show that – up to first order in Φ – the right hand sides of its geodesic deviation equations are

$$U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^0 = 0, \tag{11.4.38}$$

$$U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^i = R_{i0j0}\xi^j; \tag{11.4.39}$$

where the linearized Riemann tensor reads

$$R_{i0j0} = -\partial_i \partial_j \Phi(\vec{x}). \tag{11.4.40}$$

(Hint: You may find the results in Problem (11.14) useful.) Assuming that the monopole contribution dominates,

$$\Phi(\vec{x}) \approx \Phi(r) = -\frac{G_{\mathrm{N}} M}{r} = -\frac{r_s}{2r}, \tag{11.4.41}$$

show that these tidal forces have strengths that scale as $1/r^3$ as opposed to the $1/r^2$ forces of Newtonian gravity itself – specifically, you should find

$$R_{i0j0} \approx -\left(\delta^{ij} - \widehat{r}^i \widehat{r}^j\right)\frac{\Phi'(r)}{r} - \widehat{r}^i \widehat{r}^j \Phi''(r), \qquad \widehat{r}^i \equiv \frac{x^i}{r}, \tag{11.4.42}$$

so that the result follows simply from counting the powers of $1/r$ from $\Phi'(r)/r$ and $\Phi''(r)$. By setting $\vec{\xi}$ to be (anti-)parallel and perpendicular to the radial direction $\widehat{r}$, argue that the test masses lying on the radial line emanating from the body centered at $\vec{x} = \vec{0}$ will be *stretched apart* while the test masses lying on the plane perpendicular to $\widehat{r}$ will be *squeezed together*. (Hint: You should be able to see that $\delta^{ij} - \widehat{r}^i \widehat{r}^j$ is the Euclidean space orthogonal to $\widehat{r}$.)

The shape of the Earth's ocean tides can be analyzed in this manner by viewing the Earth as 'falling' in the gravitational fields of the Moon and the Sun. $\square$

Evolution of Distortion Tensor That U and ξ commutes in eq. (11.4.26) – the integral curves of the two vector fields can be used to form a local 2D coordinate system – can be interpreted as the first order evolution equation of the deviation vector ξ along the geodesic U, as driven by the 'distortion tensor' $\Sigma^\alpha{}_\beta \equiv \nabla_\beta U^\alpha = U^\alpha{}_{;\beta}$:

$$U^\nu \nabla_\nu \xi^\mu = \Sigma^\mu{}_\nu \xi^\nu. \tag{11.4.43}$$

We may thus study the evolution of this distortion tensor itself, along U.

$$U^\gamma \nabla_\gamma \Sigma^\alpha{}_\beta = U^\gamma \nabla_\gamma \nabla_\beta U^\alpha \tag{11.4.44}$$

$$= U^\gamma [\nabla_\gamma \nabla_\beta] U^\alpha + U^\gamma \nabla_\beta \nabla_\gamma U^\alpha \tag{11.4.45}$$

$$= U^\gamma R^\alpha{}_{\sigma\gamma\beta} U^\sigma + \nabla_\beta (U^\gamma \nabla_\gamma U^\alpha) - (\nabla_\beta U^\gamma)(\nabla_\gamma U^\alpha). \tag{11.4.46}$$

Using the geodesic equation $U\nabla U = 0$ and the antisymmetry $R^\alpha{}_{\sigma\gamma\beta} = -R^\alpha{}_{\sigma\beta\gamma}$, we gather

$$U^\gamma \nabla_\gamma \Sigma^\alpha{}_\beta = -R^\alpha{}_{\mu\beta\nu} U^\mu U^\nu - \Sigma^\alpha{}_\gamma \Sigma^\gamma{}_\beta. \tag{11.4.47}$$

Suppose we have $D \equiv d - 1$ such vectors $\{\xi_{\rm I}\}$, $\mathrm{I} \in \{1, \dots, D\}$, where we assume they are orthogonal to U, and commute with U and with themselves. The (infinitesimal) volume of spacetime perpendicular to U and spanned by these ξs is thus

$$V \equiv \widetilde{\epsilon}_{\mu\nu_1\dots\nu_D} U^\mu \xi_1^{\nu_1} \dots \xi_D^{\nu_D}. \tag{11.4.48}$$

Taking into account the geodesic character of U and the covariantly constant nature of $\widetilde{\epsilon}$, the time evolution of this 'local' volume is

$$\dot{V} \equiv U^\sigma \nabla_\sigma V = \widetilde{\epsilon}_{\mu\nu_1\dots\nu_D} U^\mu \Sigma^{\nu_1}{}_{\sigma_1} \xi_1^{\sigma_1} \dots \xi_D^{\nu_D} + \dots + \widetilde{\epsilon}_{\mu\nu_1\dots\nu_D} U^\mu \xi_1^{\nu_1} \dots \Sigma^{\nu_D}{}_{\sigma_D} \xi_D^{\sigma_D}. \tag{11.4.49}$$

Because these U and ξs commute, they may be integrated to yield their respective local coordinates $\{y^0, y^1, \dots, y^D\}$, and

$$V \equiv \widetilde{\epsilon}_{0\ 1\ 2\dots\ D} \tag{11.4.50}$$

$$= \sqrt{|g(y)|} \tag{11.4.51}$$

$$\dot{V} = \widetilde{\epsilon}_{0\ 1\ 2\ \dots D} \Sigma^1{}_1 + \dots + \widetilde{\epsilon}_{0\ 1\dots D} \Sigma^D{}_D \tag{11.4.52}$$

$$= \sqrt{|g(y)|} \Sigma^\sigma{}_\sigma. \tag{11.4.53}$$

YZ: Check this argument again. Dividing the second line by the first, and recognizing $\Sigma^\sigma{}_\sigma = \nabla \cdot U$, we see that all dependence on the ξs drop out:

If U is an affinely parametrized geodesic, its divergence $\nabla_\mu U^\mu$ measures $\mathrm{d}\ln V/\mathrm{d}\lambda$, the $\lambda-$rate of fractional change of spacetime volume perpendicular to U itself.

Now, if U were timelike, we may normalize $U^2 = 1$ and the metric orthogonal to it is

$$H_{\mu\nu} = g_{\mu\nu} - U_\mu U_\nu, \tag{11.4.54}$$

since $H_{\mu\nu}U^\nu = U_\mu - U_\nu(U^2) = 0$. Its trace is $H^\sigma{}_\sigma = d-1$, telling us that the space orthogonal to a timelike vector is one dimension lower than the ambient spacetime.

Now, if U were spacelike, we may normalize $U^2 = -1$ and the metric orthogonal to it is

$$H_{\mu\nu} = g_{\mu\nu} + U_\mu U_\nu, \tag{11.4.55}$$

since $H_{\mu\nu}U^\nu = U_\mu + U_\nu(U^2) = 0$. Its trace is $H^\sigma{}_\sigma = d-1$, telling us that the spacetime orthogonal to a spacelike vector is one dimension lower than the ambient spacetime.

On the other hand, if U were actually null, the space orthogonal to it would be *two* dimensions lower than the ambient spacetime. To see this, simply examine Minkowski spacetime itself, with coordinate system $(t, z, \vec{x}_\perp)$, where t is time and $(z, \vec{x}_\perp)$ are Cartesian spatial coordinates. Suppose we consider the following null rays,

$$y^\alpha_\pm(\lambda \geq 0) = \left(t_0 + \lambda, z_0 \pm \lambda, \vec{0}\right), \tag{11.4.56}$$

describing a photon emitted at $(t = t_0, z = z_0, \vec{x}_\perp = \vec{0})$ and moving strictly in the positive $(+)$ or negative $(-)$ z direction. Then, if we compute the sum and difference of the (t, z) coordinates; namely,

$$y^t_+ - y^z_+ = t_0 - z_0, \tag{11.4.57}$$
$$y^t_+ + y^z_+ = t_0 + z_0 + 2\lambda; \tag{11.4.58}$$

and

$$y^t_- + y^z_- = t_0 + z_0, \tag{11.4.59}$$
$$y^t_- - y^z_- = t_0 - z_0 + 2\lambda; \tag{11.4.60}$$

this tells us the null signal propagating strictly along the positive z direction has constant t minus z components and that along the negative z direction has constant t plus z components. This motivates us define

$$x^\pm \equiv t \pm z. \tag{11.4.61}$$

The flat metric becomes

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}z^2 - \mathrm{d}\vec{x}^2_\perp \tag{11.4.62}$$
$$= \mathrm{d}x^+\mathrm{d}x^- - \mathrm{d}\vec{x}^2_\perp. \tag{11.4.63}$$

For the null signal along the $+z$ direction ($-z$ direction), we set the corresponding $x^- =$ constant ($x^+ =$ constant); i.e., $\mathrm{d}x^- = 0$ ($\mathrm{d}x^+ = 0$), and the induced metric becomes

$$\mathrm{d}s^2 \to -\mathrm{d}\vec{x}^2_\perp. \tag{11.4.64}$$

Since any metric is locally flat, this dimension reduction by two must be a generic phenomenon. In any case, to construct a metric $H_{\mu\nu}$ perpendicular to a null U, we need to find another null vector N such that $N \cdot U = 1$ – for instance, in the above flat spacetime example, if $U_\mu \mathrm{d}x^\mu = \mathrm{d}x^-/\sqrt{2}$ then we may choose $N_\mu \mathrm{d}x^\mu = \mathrm{d}x^+/\sqrt{2}$ (and vice versa). The geometry orthogonal to U is then given by

$$H_{\mu\nu} = g_{\mu\nu} - U_\mu N_\nu - U_\nu N_\mu. \tag{11.4.65}$$

(Note that the prescription in eq. (11.4.54) does not work when $U^2 = 0$ because $(g_{\mu\nu} - U_\mu U_\nu)U^\mu = U_\mu$.) For instance, in the above flat spacetime example,

$$H_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu - 2(U_\mu \mathrm{d}x^\mu)(N_\nu \mathrm{d}x^\nu) \tag{11.4.66}$$

$$= \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu - \mathrm{d}x^-\mathrm{d}x^+ = -\mathrm{d}\vec{x}_\perp^2. \tag{11.4.67}$$

For the general case, we may verify directly that $H_{\mu\nu}$ is orthogonal to both U and N,

$$H_{\mu\nu}U^\nu = U_\mu - U_\mu(N \cdot U) - (U^2)N_\mu \tag{11.4.68}$$

$$= U_\mu - U_\mu = 0 \tag{11.4.69}$$

and

$$H_{\mu\nu}N^\nu = N_\mu - U_\mu(N^2) - (N \cdot U)N_\mu \tag{11.4.70}$$

$$= N_\mu - N_\mu = 0. \tag{11.4.71}$$

At this point, we may return to eq. (11.4.47) and decompose $\Sigma_{\alpha\beta} = \nabla_\beta U_\alpha$ into its irreducible components. Define the divergence (*expansion*) to be

$$\Sigma^\sigma{}_\sigma = \nabla_\sigma U^\sigma \equiv \theta; \tag{11.4.72}$$

the anti-symmetric (*rotation*) piece to be

$$\frac{1}{2}\Sigma_{[\alpha\beta]} = \frac{1}{2}\nabla_{[\beta}U_{\alpha]} \equiv \omega_{\alpha\beta}; \tag{11.4.73}$$

and symmetric-traceless (*shear*) portions of $\Sigma_{\alpha\beta}$ – for U timelike – to be

$$\frac{1}{2}\Sigma_{\{\alpha\beta\}} - \frac{\theta}{d-1}H_{\alpha\beta} \equiv \sigma_{\alpha\beta} \tag{11.4.74}$$

whereas for U null

$$\frac{1}{2}\Sigma_{\{\alpha\beta\}} - \frac{\theta}{d-2}H_{\alpha\beta} \equiv \sigma_{\alpha\beta}. \tag{11.4.75}$$

In other words,

$$\nabla_\beta U_\alpha \equiv \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \frac{\theta}{d-1}H_{\alpha\beta}, \qquad \text{(Timelike/Spacelike)} \tag{11.4.76}$$

$$\equiv \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \frac{\theta}{d-2}H_{\alpha\beta}, \qquad \text{(Null)}. \tag{11.4.77}$$

This leads us to the following pair of time-evolution equations for the expansion $\theta = \nabla \cdot U$.

Problem 11.24. Raychaudhuri Equations Take the trace of eq. (11.4.47) to derive the Raychaudhuri equation, describing the evolution of the expansion of geodesic congruence:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{d-1} - \sigma^{\alpha\beta}\sigma_{\alpha\beta} + \omega^{\alpha\beta}\omega_{\alpha\beta} - R_{\alpha\beta}U^\alpha U^\beta, \qquad \text{(Timelike } U\text{)} \tag{11.4.78}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{d-2} - \sigma^{\alpha\beta}\sigma_{\alpha\beta} + \omega^{\alpha\beta}\omega_{\alpha\beta} - R_{\alpha\beta}U^\alpha U^\beta, \qquad \text{(Null } U\text{)}. \tag{11.4.79}$$

Hint: You should first argue that $\omega^{\alpha\beta}\sigma_{\alpha\beta} = \omega^{\alpha\beta}H_{\alpha\beta} = \sigma^{\alpha\beta}\sigma_{\alpha\beta} = 0$. □

Problem 11.25. Divergence vs Trace Explain why the trace of $\Sigma_{\alpha\beta} \equiv \nabla_\beta U_\alpha$ using either the full inverse metric $g^{\alpha\beta}$ or the orthogonal one $H^{\alpha\beta}$ both yields the divergence $\nabla_\sigma U^\sigma$; i.e., show that

$$g^{\alpha\beta}\nabla_\beta U_\alpha = H^{\alpha\beta}\nabla_\beta U_\alpha = \nabla_\sigma U^\sigma. \tag{11.4.80}$$

You should find that this result holds for U timelike, spacelike and null. □

Problem 11.26. Projector Properties Use the orthogonality properties

$$H_{\mu\nu}U^\nu = 0 \qquad \text{(Timelike and Spacelike)}, \tag{11.4.81}$$

$$H_{\mu\nu}U^\nu = 0 = H_{\mu\nu}N^\nu \qquad \text{(Null)} \tag{11.4.82}$$

to explain the projector property

$$H^\mu{}_\sigma H^\sigma{}_\nu = H^\mu{}_\nu. \tag{11.4.83}$$

What is $H_{\mu\nu}H^\mu{}_\alpha H^\nu{}_\beta$? □

Geometric Meaning of Ricci Tensor Having discussed at some length the meaning of the Riemann tensor, we may now ask: Is there a geometric meaning to its trace, the Ricci tensor in eq. (11.3.47)? One such geometric meaning can be found within the Raychaudhuri equations (11.4.78) and (11.4.79), which describe the rate of expansion or contraction of a bundle of integral curves; which in turn describes the rate of change of spacetime volume along these geodesic trajectories. Another (related) perspective is its relation to the local volume of spacetime relative that of Minkowski. For, we may identify in equations (11.4.17),

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}; \tag{11.4.84}$$

where

$$h_{\mu\nu}(y) = -\frac{1}{3}R_{\mu\alpha\nu\beta}(y_0)\cdot(y-y_0)^\alpha(y-y_0)^\beta + \mathcal{O}\left((y-y_0)^3\right). \tag{11.4.85}$$

This in turn implies, from eq. (11.3.47),

$$h(y) \equiv \eta^{\alpha\beta}(y)h_{\alpha\beta}(y) = -\frac{1}{3}R_{\alpha\beta}(y_0)\cdot(y-y_0)^\alpha(y-y_0)^\beta + \mathcal{O}\left((y-y_0)^3\right). \tag{11.4.86}$$

At this point, we may invoke the spacetime version of the discussion leading up to eq. (9.5.154), to deduce the infinitesimal spacetime volume element around $y = y_0$ is given by

$$\mathrm{d}^d y\sqrt{|g(y \approx y_0)|} = \mathrm{d}^d y \left(1 - \frac{1}{6} R_{\alpha\beta}(y_0) \cdot (y - y_0)^\alpha (y - y_0)^\beta + \mathcal{O}\left((y - y_0)^3\right)\right). \qquad (11.4.87)$$

This teaches us: the Ricci tensor controls the growth/shrinking of volume, relative to that in flat spacetime, as one follows the congruence of vectors $(y - y_0)^\alpha$ emanating from some fixed location y_0.[130]

Interlude Let us pause to summarize the physics we have revealed thus far.

> In a curved spacetime, where $g_{\mu\nu} \neq \eta_{\mu\nu}$, no global Lorentz inertial frame exists. The collective motion of a system of mass M sweeps out a timelike geodesic – recall equations (11.3.60), (11.3.78), and (11.3.84) – whose dynamics is actually independent of M as long as its internal structure can be neglected. In the co-moving frame of an observer situated within this same system, physical laws appear to be the same as that in Minkowski spacetime up to distances of order $1/|\max R_{\widehat{\alpha}\widehat{\beta}\widehat{\mu}\widehat{\nu}}|^{1/2}$. However, once the finite size of the physical system is taken into account, one would find tidal forces exerted upon it due to spacetime curvature itself – this is described by the geodesic deviation eq. (11.4.33).

11.5 Symmetries and Geodesics

Killing Vectors A geometry is said to enjoy an isometry – or, symmetry – when we perform a coordinate transformation induced by the following infinitesimal displacement

$$x^\mu \to x^\mu + \xi^\mu(x) \qquad (11.5.1)$$

and find that the geometry is unchanged

$$g_{\mu\nu}(x) \to g_{\mu\nu}(x) + \mathcal{O}\left(\xi^2\right). \qquad (11.5.2)$$

Generically, under the infinitesimal transformation of eq. (11.5.1),

$$g_{\mu\nu}(x) \to g_{\mu\nu}(x) + \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu. \qquad (11.5.3)$$

where

$$\nabla_{\{\mu}\xi_{\nu\}} = \xi^\sigma \partial_\sigma g_{\mu\nu} + g_{\sigma\{\mu}\partial_{\nu\}}\xi^\sigma. \qquad (11.5.4)$$

If an isometry exists along the integral curve of ξ^μ, it has to obey Killing's equation – recall equations (9.3.5) and (9.3.6) –

$$\nabla_{\{\mu}\xi_{\nu\}} = \xi^\sigma \partial_\sigma g_{\mu\nu} + \partial_{\{\mu}\xi^\sigma g_{\nu\}\sigma} = 0. \qquad (11.5.5)$$

In fact, by exponentiating the infinitesimal coordinate transformation, it is possible to show that – if ξ^μ is a Killing vector (i.e., it satisfies eq. (11.5.5)), then an isometry exists along its integral curve. In other words,

---

[130] A shorter version of this discussion may be found on Wikipedia. A closely related explanation of the meaning of Einstein's equation (11.3.56) using the Raychaudhuri equation may be found in Baez and Bunn [30].

A spacetime geometry enjoys an isometry (aka symmetry) along the integral curve of ξ^μ iff it obeys $\nabla_{\{\mu}\xi_{\nu\}} = \nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = 0$.

In a $d-$dimensional spacetime, there are at most $d(d+1)/2$ Killing vectors. A spacetime that has $d(d+1)/2$ Killing vectors is called *maximally symmetric*. (See Weinberg [24] for a discussion.)

Problem 11.27. Conserved quantities along geodesics (I of II) ∘ If p_μ denotes the 'momentum' variable of a geodesic

$$p_\mu \equiv \frac{\partial L_{\text{g}}}{\partial \dot{z}^\mu} = g_{\mu\nu}\left(z(\lambda)\right)\frac{\mathrm{d}z^\nu(\lambda)}{\mathrm{d}\lambda}, \tag{11.5.6}$$

where L_{g} is defined in eq. (11.3.63), and if ξ^μ is a Killing vector of the same geometry $\nabla_{\{\alpha}\xi_{\beta\}} = 0$, show that

$$\xi^\mu(z(\lambda))p_\mu(\lambda) = g_{\alpha\beta}\dot{z}^\alpha(\lambda)\xi^\beta(z(\lambda)) \tag{11.5.7}$$

is a constant along the geodesic $z^\mu(\lambda)$. Hint: See Problem (9.29).

(II of II) ∘ The vector field version of this result goes as follows.

If the geodesic equation $v^\sigma\nabla_\sigma v^\mu = 0$ holds, and if ξ^μ is a Killing vector, then $\xi_\nu v^\nu$ is conserved along the integral curve of v^μ.

Can you demonstrate the validity of this statement? □

Second Derivatives of Killing Vectors Now let us also consider the second derivatives of ξ^μ. In particular, we will now explain why

$$\nabla_\alpha\nabla_\beta\xi_\delta = R^\lambda{}_{\alpha\beta\delta}\xi_\lambda. \tag{11.5.8}$$

Consider

$$\begin{aligned}
0 &= \nabla_\delta\nabla_{\{\alpha}\xi_{\beta\}} && (11.5.9)\\
&= [\nabla_\delta,\nabla_\alpha]\xi_\beta + \nabla_\alpha\nabla_\delta\xi_\beta + [\nabla_\delta,\nabla_\beta]\xi_\alpha + \nabla_\beta\nabla_\delta\xi_\alpha && (11.5.10)\\
&= -R^\lambda{}_{\beta\delta\alpha}\xi_\lambda - \nabla_\alpha\nabla_\beta\xi_\delta - R^\lambda{}_{\alpha\delta\beta}\xi_\lambda - \nabla_\beta\nabla_\alpha\xi_\delta && (11.5.11)
\end{aligned}$$

Because Bianchi says $0 = R^\lambda{}_{[\alpha\beta\delta]} \Rightarrow R^\lambda{}_{\alpha\beta\delta} = R^\lambda{}_{\beta\alpha\delta} + R^\lambda{}_{\delta\beta\alpha}$.

$$\begin{aligned}
0 &= -R^\lambda{}_{\beta\delta\alpha}\xi_\lambda - \nabla_\alpha\nabla_\beta\xi_\delta + \left(R^\lambda{}_{\beta\alpha\delta} + R^\lambda{}_{\delta\beta\alpha}\right)\xi_\lambda - \nabla_\beta\nabla_\alpha\xi_\delta && (11.5.12)\\
0 &= -2R^\lambda{}_{\beta\delta\alpha}\xi_\lambda - \nabla_{\{\beta}\nabla_{\alpha\}}\xi_\delta - [\nabla_\beta,\nabla_\alpha]\xi_\delta && (11.5.13)\\
0 &= -2R^\lambda{}_{\beta\delta\alpha}\xi_\lambda - 2\nabla_\beta\nabla_\alpha\xi_\delta && (11.5.14)
\end{aligned}$$

This proves eq. (11.5.8).

Commutators of Killing Vectors Next, we will show that

The commutator of 2 Killing vectors is also a Killing vector.

Let U and V be Killing vectors. If $\xi \equiv [U, V]$, we need to verify that

$$\nabla_{\{\alpha}\xi_{\beta\}} = \nabla_{\{\alpha}[U, V]_{\beta\}} = 0. \tag{11.5.15}$$

More explicitly, let us compute:

$$\begin{aligned}
&\nabla_\alpha(U^\mu\nabla_\mu V_\beta - V^\mu\nabla_\mu U_\beta) + (\alpha \leftrightarrow \beta) \\
&= \nabla_\alpha U^\mu\nabla_\mu V_\beta - \nabla_\alpha V^\mu\nabla_\mu U_\beta + U^\mu\nabla_\alpha\nabla_\mu V_\beta - V^\mu\nabla_\alpha\nabla_\mu U_\beta + (\alpha \leftrightarrow \beta) \\
&= -\nabla_\mu U_\alpha\nabla^\mu V_\beta + \nabla_\mu V_\alpha\nabla^\mu U_\beta + U^\mu\nabla_{[\alpha}\nabla_{\mu]}V_\beta + U^\mu\nabla_\mu\nabla_\alpha V_\beta \\
&\qquad\qquad - V^\mu\nabla_{[\alpha}\nabla_{\mu]}U_\beta - V^\mu\nabla_\mu\nabla_\alpha U_\beta + (\alpha \leftrightarrow \beta) \\
&= -U^\mu R^\sigma{}_{\beta\alpha\mu}V_\sigma + V^\mu R^\sigma{}_{\beta\alpha\mu}U_\sigma + (\alpha \leftrightarrow \beta) \\
&= -U^{[\mu}V^{\sigma]}R_{\sigma\{\beta\alpha\}\mu} = 0.
\end{aligned}$$

The $(\alpha \leftrightarrow \beta)$ means we are taking all the terms preceding it and swapping $\alpha \leftrightarrow \beta$. Moreover, we have repeatedly used the Killing equations $\nabla_\alpha U_\beta = -\nabla_\beta U_\alpha$ and $\nabla_\alpha V_\beta = -\nabla_\beta V_\alpha$.

Problem 11.28. Killing Vectors in Minkowski In Minkowski spacetime $g_{\mu\nu} = \eta_{\mu\nu}$, with Cartesian coordinates $\{x^\mu\}$, use eq. (11.5.8) to argue that the most general Killing vector takes the form

$$\xi_\mu = \ell_\mu + \omega_{\mu\nu}x^\nu, \tag{11.5.16}$$

for constant ℓ_μ and $\omega_{\mu\nu}$. (Hint: Think about Taylor expansions; use eq. (11.5.8) to show that the 2nd and higher partial derivatives of ξ_δ are zero.) Then use the Killing equation (11.5.5) to infer that

$$\omega_{\mu\nu} = -\omega_{\nu\mu}. \tag{11.5.17}$$

The ℓ_μ corresponds to infinitesimal spacetime translation and the $\omega_{\mu\nu}$ to infinitesimal Lorentz boosts and rotations. Explain why this implies the following are the Killing vectors of flat spacetime:

$$\partial_\mu \qquad \text{(Generators of spacetime translations)} \tag{11.5.18}$$

and

$$x^{[\mu}\partial^{\nu]} \qquad \text{(Generators of Lorentz boosts or rotations).} \tag{11.5.19}$$

There are d distinct ∂_μ's and (due to their antisymmetry) $(1/2)(d^2-d)$ distinct $x^{[\mu}\partial^{\nu]}$'s. Therefore there are a total of $d(d+1)/2$ Killing vectors in Minkowski – i.e., it is maximally symmetric. □

It might be instructive to check our understanding of rotation and boosts against the 2D case we have worked out earlier via different means. Up to first order in the rotation angle θ, the 2D rotation matrix in eq. (**??**) reads

$$\widehat{R}^i{}_j(\theta) = \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix} + \mathcal{O}\left(\theta^2\right). \tag{11.5.20}$$

In other words, $\widehat{R}^i{}_j(\theta) = \delta_{ij} - \theta\epsilon_{ij}$, where ϵ_{ij} is the Levi-Civita symbol in 2D with $\epsilon_{12} \equiv 1$. Applying a rotation of the 2D Cartesian coordinates x^i upon a test (scalar) function f,

$$f(x^i) \to f\left(\widehat{R}^i{}_j x^j\right) = f\left(x^i - \theta\epsilon_{ij}x^j + \mathcal{O}\left(\theta^2\right)\right) \tag{11.5.21}$$

$$= f(\vec{x}) - \theta\epsilon_{ij}x^j\partial_i f(\vec{x}) + \mathcal{O}\left(\theta^2\right). \tag{11.5.22}$$

Since θ is arbitrary, the basic differential operator that implements an infinitesimal rotation of the coordinate system on any Minkowski scalar is

$$-\epsilon_{ij}x^j\partial_i = x^1\partial_2 - x^2\partial_1. \tag{11.5.23}$$

This is the 2D version of eq. (11.5.19) for rotations. As for 2D Lorentz boosts, eq. (**??**) tells us

$$\Lambda^\mu{}_\nu(\xi) = \begin{bmatrix} 1 & \xi \\ \xi & 1 \end{bmatrix} + \mathcal{O}\left(\xi^2\right). \tag{11.5.24}$$

(This ξ is known as *rapidity*.) Here, we have $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \xi \cdot \epsilon^\mu{}_\nu$, where $\epsilon_{\mu\nu}$ is the Levi-Civita tensor in 2D Minkowski with $\epsilon_{01} \equiv 1$. Therefore, to implement an infinitesimal Lorentz boost on the Cartesian coordinates within a test (scalar) function $f(x^\mu)$, we do

$$f(x^\mu) \to f\left(\Lambda^\mu{}_\nu x^\nu\right) = f\left(x^\mu + \xi\epsilon^\mu{}_\nu x^\nu + \mathcal{O}\left(\xi^2\right)\right) \tag{11.5.25}$$

$$= f(x) - \xi\epsilon_{\nu\mu}x^\nu\partial^\mu f(x) + \mathcal{O}\left(\xi^2\right). \tag{11.5.26}$$

Since ξ is arbitrary, to implement a Lorentz boost of the coordinate system on any Minkowski scalar, the appropriate differential operator is

$$\epsilon_{\mu\nu}x^\mu\partial^\nu = x^0\partial^1 - x^1\partial^0; \tag{11.5.27}$$

which again is encoded within eq. (11.5.19).

Problem 11.29. Lie Algebra from Killing Vectors Verify that Lie Algebra of $\mathrm{SO}_{D,1}$ in (10.6.8) is recovered if we exploit eq. (11.5.19) to define

$$J^{\mu\nu} = i(x^\mu\partial^\nu - x^\nu\partial^\mu), \tag{11.5.28}$$

where $\partial^\mu \equiv \eta^{\mu\nu}\partial_\nu$. This tells us

$$f(x) \to \exp(-(i/2)\omega_{\mu\nu}J^{\mu\nu})f(x) \tag{11.5.29}$$

under a Lorentz boost or rotation. □

Problem 11.30. Co-moving Observers & Rulers In Cosmology We live in a universe that, at the very largest length scales, is described by the following spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}; \tag{11.5.30}$$

where $a(t)$ describes the relative size of the universe. Enumerate as many constants-of-motion as possible of this geometry. (Hint: Focus on the spatial part of the metric and try to draw a connection with the previous problem.)

In this cosmological context, a co-moving observer is one that does not move spatially, i.e., $\mathrm{d}\vec{x} = 0$. Solve the geodesic swept out by such an observer.

Galaxies A and B are respectively located at $\vec{x}$ and $\vec{x}'$ at a fixed cosmic time t. What is their spatial distance on this constant t slice of spacetime? □

Problem 11.31. Killing identities involving Ricci Prove the following results. If ξ^μ is a Killing vector and $R_{\alpha\beta}$ and $\mathcal{R}$ are the Ricci tensor and scalar respectively, then

$$\nabla^\alpha \left(R_{\alpha\beta}\xi^\beta \right) = 0 = \xi^\alpha \nabla^\beta R_{\alpha\beta} \qquad \text{and} \qquad \xi^\alpha \nabla_\alpha \mathcal{R} = 0. \tag{11.5.31}$$

Hints: First use eq. (11.5.8) to show that

$$\Box \xi_\delta = -R^\lambda{}_\delta \xi_\lambda, \tag{11.5.32}$$

$$\Box \equiv g^{\alpha\beta}\nabla_\alpha\nabla_\beta = \nabla_\alpha\nabla^\alpha. \tag{11.5.33}$$

Then take the divergence on both sides, and commute the covariant derivatives until you obtain the term $\Box\nabla^\delta\xi_\delta$ – what is $\nabla^\delta\xi_\delta$ equal to? You may also need to employ the Einstein tensor Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$ to infer that $\xi^\alpha\nabla_\alpha\mathcal{R} = 0$. $\Box$

11.5.1 Schwarzschild – the non-rotating Black Hole

As an example of how symmetries may be exploited to study geodesic motion, let us now study the physically important Schwarzschild spacetime, the metric of a non-rotating black hole:

$$\mathrm{d}s^2 = \left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{r_s}{r}} - r^2\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right). \tag{11.5.34}$$

Notice this is a rotationally invariant geometry. Below, we will justify why it is therefore always possible to assume that motion is taking place only on the equatorial plane $\theta = \pi/2$. The associated geodesic Lagrangian is therefore

$$L_g = \frac{1}{2}\left\{\left(1 - \frac{r_s}{r}\right)t'^2 - \frac{r'^2}{1 - \frac{r_s}{r}} - r^2\phi'^2\right\}. \tag{11.5.35}$$

Problem 11.32. Ricci-flat Solution Verify that the Schwarzschild metric in eq. (11.5.34) is Ricci flat: $R_{\mu\nu} = 0$? Why is $R_{\mu\nu} = 0$ iff $G_{\mu\nu} = 0$? In General Relativity, this result implies the Schwarzschild geometry is a vacuum (i.e., matter-free) solution. $\Box$

Problem 11.33. Christoffel Symbols Verify that the non-zero Christoffel symbols of the Schwarzschild metric in eq. (11.5.34) are

$$\Gamma^t{}_{tr} = \Gamma^t{}_{rt} = \frac{r_s/(2r^2)}{1 - r_s/r}, \tag{11.5.36}$$

$$\Gamma^r{}_{tt} = \left(1 - \frac{r_s}{r}\right)\frac{r_s}{2r^2}, \qquad \Gamma^r{}_{rr} = -\frac{r_s/(2r^2)}{1 - r_s/r}, \tag{11.5.37}$$

$$\Gamma^r{}_{\theta\theta} = -(r - r_s), \qquad \Gamma^r{}_{\phi\phi} = -(r - r_s)\sin^2\theta, \tag{11.5.38}$$

$$\Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \frac{1}{r}, \qquad \Gamma^\theta{}_{\phi\phi} = -\sin\theta\cdot\cos\theta, \tag{11.5.39}$$

$$\Gamma^\phi{}_{r\phi} = \Gamma^\phi{}_{\phi r} = \frac{1}{r}, \qquad \Gamma^\phi{}_{\theta\phi} = \Gamma^\phi{}_{\phi\theta} = \cot\theta. \tag{11.5.40}$$

$\Box$

Null geodesics in Schwarzschild Let us begin with null geodesics, where the Lagrangian is zero.

$$\left(1 - \frac{r_s}{r}\right) t'^2 - \frac{r'^2}{1 - \frac{r_s}{r}} - r^2 \phi'^2 = 0 \tag{11.5.41}$$

The prime is the derivative with respect to some affine parameter λ.

Let us observe that the Schwarzschild metric in eq. (11.5.34) is $t-$ and $\phi-$independent, corresponding respectively to time-translation and rotational symmetry. Their conserved quantities are energy ε, where

$$\varepsilon \equiv \frac{\partial L_g}{\partial t'} = \left(1 - \frac{r_s}{r}\right) t'; \tag{11.5.42}$$

and angular momentum ℓ, where

$$\ell \equiv -\frac{\partial L_g}{\partial \phi'} = r^2 \phi'. \tag{11.5.43}$$

The Lagrangian being zero, as well as energy and angular momentum conservation, together yields

$$t' = \frac{\varepsilon}{1 - r_s/r}, \tag{11.5.44}$$

$$\phi' = \frac{\ell}{r^2}, \tag{11.5.45}$$

and

$$\frac{1}{2} r'^2 + V = \frac{\varepsilon^2}{2}. \tag{11.5.46}$$

We have defined

$$V \equiv \frac{1}{2}\left(\frac{\ell}{r_s}\right)^2 \left(\left(\frac{r_s}{r}\right)^2 - \left(\frac{r_s}{r}\right)^3\right). \tag{11.5.47}$$

We may view eq. (11.5.46) as the conservation of kinetic energy $(1/2)r'^2$ plus potential energy V. Defining

$$\alpha \equiv \frac{\ell}{r_s} \qquad \text{and} \qquad \frac{r}{r_s} \equiv \rho, \tag{11.5.48}$$

we may re-express the potential as

$$V(\rho) = \frac{\alpha^2}{2}\left(\frac{1}{\rho^2} - \frac{1}{\rho^3}\right); \tag{11.5.49}$$

which has a derivative

$$V'(\rho) = \frac{3\alpha^2}{2\rho^3}\left(\frac{1}{\rho} - \frac{2}{3}\right). \tag{11.5.50}$$

This tells us there is only one turning point, which is located at

$$r = \frac{3}{2} r_s. \tag{11.5.51}$$

The potential energy V goes to $-\infty$ as $r \to 0$; $V(r = r_s) = 0$; and V goes to 0 as $r \to \infty$. Whereas, at $r = (3/2)r_s$, we have

$$V(r/r_s = 3/2) = \frac{2}{27}\alpha^2 > 0. \tag{11.5.52}$$

Hence eq. (11.5.51) is the global maximum and is an unstable equilibrium. For $\varepsilon^2/2 < (2/27)\alpha^2$, any null ray moving towards the black hole from a large distance will eventually turn around and fly off to infinity. For $\varepsilon^2/2 > (2/27)\alpha^2$, if the zero mass particle were moving outward it will fly off to infinity; whereas if it were moving inward it will eventually plunge into the black hole.

If $\varepsilon^2/2 = (2/27)\alpha^2$, the photon can now orbit at the radius in eq. (11.5.51). This is known as the *photon ring*, whose existence in turn leads to the formation of the 'shadow' of a black hole.

> The only bound lightlike orbit is the circular one at $r = \frac{3}{2}r_s$. It is unstable – any small perturbations would cause it to either plunge into the black hole or escape to infinity.

These photon trajectories are stark manifestations of the effect of strong relativistic gravity. While in Newtonian gravity, light has no mass and therefore cannot be affected by gravity; if gravity is in fact the manifestation of curved spacetime, we see that (high frequency) light can in fact make circles around our central black hole; or, become deflected by it. Below, you will solve the light deflection angle due to a weak source of gravity such as the Sun; this was in fact one of the "classic" tests of General Relativity.

YZ: Gravitational lensing. Kruskal-Szekeres. Riemann in orthonormal frame. Penrose diagram; (t,r) vs (u,v). Horizon. Schutz: photon emission from collapsing star; horizon generated by null rays. Dragging of inertial frames. Gyroscope precession/Lense-Thirring?

Problem 11.34. Re-scaling the Affine Parameter Explain why it is possible to re-scale the affine parameter such that we may choose $\varepsilon = 1$ in eq. (11.5.44); so that equations (11.5.44), (11.5.45), and (11.5.46) become

$$t' = (1 - r_s/r)^{-1}, \tag{11.5.53}$$

$$\phi' = \frac{\ell}{r^2}; \tag{11.5.54}$$

and

$$\frac{1}{2}r'^2 + \frac{1}{2}V = \frac{1}{2}, \tag{11.5.55}$$

$$V \equiv \alpha^2 \left(\left(\frac{r_s}{r}\right)^2 - \left(\frac{r_s}{r}\right)^3 \right). \tag{11.5.56}$$

Hint: Remember $L_g = 0$ for null geodesics. □

Problem 11.35. Equatorial Plane In this problem, we will justify why geodesic motion in Schwarzschild may always be taken to be confined solely on the equatorial plane $\theta = \pi/2$. First show that, upon the radial coordinate transformation

$$r^2 \to \left(1 + \frac{r_s}{4r}\right)^4 r^2, \tag{11.5.57}$$

the Schwarzschild geometry in eq. (11.5.34) takes the *isotropic* form

$$\mathrm{d}s^2 = \left(\frac{1 - \frac{r_s}{4r}}{1 + \frac{r_s}{4r}}\right)^2 \mathrm{d}t^2 - \left(1 + \frac{r_s}{4r}\right)^4 \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}, \qquad r \equiv \sqrt{\vec{x} \cdot \vec{x}}. \tag{11.5.58}$$

(Notice, this metric is manifestly invariant under

$$x^i \to \widehat{R}^i{}_j x^j, \tag{11.5.59}$$

as long as the matrix $\widehat{R}$ is constant and orthogonal, $\widehat{R}^i{}_a \widehat{R}^j{}_b \delta_{ij} = \delta_{ab}$.) Next, verify following rotation-generating Killing vectors

$$\xi^\mu \equiv \delta^\mu_k \left(x^{[i} \delta^{j]k}\right) \tag{11.5.60}$$

(for fixed i and j) indeed satisfy the Killing equation $\nabla_{\{\mu} \xi_{\nu\}} = 0$. Then, show that, along a geodesic $z^\mu(\lambda)$, the 'angular momentum'

$$L^{ij} \equiv \left(1 + \frac{r_s}{4|\vec{z}|}\right)^4 z^{[i} \mathrm{d}z^{j]}/\mathrm{d}\lambda \tag{11.5.61}$$

is a constant of notion; where $|\vec{z}| \equiv (\vec{z} \cdot \vec{z})^{1/2}$. Explain why, if $(i,j) \neq (1,2)$, it always possible to find a rotation matrix $\widehat{R}$ such that, under $\vec{x} \to \widehat{R} \cdot \vec{x}$, the 'new' (i,j) becomes $(1,2)$. Why does that tell us, if $\vec{z}$ and $\mathrm{d}\vec{z}/\mathrm{d}\lambda$ lies on the $(1,2)-$plane, it shall remain so for all time? (Hint: You may find it easier to first contract eq. (11.5.61) with the Levi-Civita symbol ϵ^{ijk} and interpret the result in a fictitious Euclidean 3−space.) $\square$

Problem 11.36. Light Deflection Due To Static Mass Monopole in 4D In General Relativity the weak field metric generated by an isolated system, of total mass M, is dominated by its mass monopole and hence goes as $1/r$ (i.e., its Newtonian potential)

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\Phi\delta_{\mu\nu} = \eta_{\mu\nu} - \frac{r_s}{r}\delta_{\mu\nu}, \tag{11.5.62}$$

where we assume $|\Phi| = r_s/r \ll 1$ and

$$r_s \equiv 2G_\mathrm{N} M. \tag{11.5.63}$$

Now, the metric of an isolated static non-rotating black hole – i.e., the Schwarzschild black hole – in isotropic coordinates was given in eq. (11.5.58).

The $r_s \equiv 2G_\mathrm{N}M$ here is the Schwarzschild radius; any object falling behind $r < r_s$ will not be able to return to the $r > r_s$ region unless it is able to travel faster than light.

Expand this metric in eq. (11.5.58) up to first order r_s/r and verify this yields eq. (11.5.62). We may therefore identify eq. (11.5.62) as either the metric due to the monopole moment of some static mass density $\rho(\vec{x})$ or the far field limit $r_s/r \ll 1$ of the Schwarzschild black hole.

Statement of Problem: Now consider shooting a beam of light from afar, and by solving the appropriate null geodesic equations, figure out how much angular deflection $\Delta\varphi$ it suffers due to the presence of a mass monopole. Express the answer $\Delta\varphi$ in terms of the coordinate radius of closest approach r_0. We shall see that the symmetries of the time-independent and rotation-invariant geometry of eq. (11.5.62) will prove very useful to this end.

Step-by-step Guide: First, write down the affine-parameter form of the Lagrangian L_g for geodesic motion in eq. (11.5.62) in spherical coordinates

$$\vec{x} = r\left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)\right). \tag{11.5.64}$$

Spherical Symmetry and θ Because of the spherical symmetry of the problem, we may always assume that all geodesic motion takes place on the equatorial plane:

$$\theta = \frac{\pi}{2}. \tag{11.5.65}$$

'Energy' Conservation Proceed to use the $t-$independence of the metric, together with the invariance of the null geodesic Lagrangian L_g under constant re-scaling of its affine parameter λ, to argue that λ itself can always be chosen such that

$$\dot{t} = \left(1 - \frac{r_s}{r}\right)^{-1}. \tag{11.5.66}$$

Angular Momentum conservation Next, use the $\phi-$independence of the metric to show that angular momentum conservation $-\partial L_g/\partial\dot{\phi} \equiv \ell$ (constant) yields

$$\dot{\phi} = \frac{\ell}{r^2}\left(1 + \frac{r_s}{r}\right)^{-1}. \tag{11.5.67}$$

We are primarily interested in the trajectory as a function of angle, so we may eliminate all $\dot{r} \equiv \mathrm{d}r/\mathrm{d}\lambda$ as

$$\dot{r} = \frac{\mathrm{d}\phi}{\mathrm{d}\lambda}r'(\phi) = \frac{\ell}{r^2}\left(1 + \frac{r_s}{r}\right)^{-1} r'(\phi), \tag{11.5.68}$$

where eq. (11.5.67) was employed in the second equality. At this point, by utilizing equations (11.5.65), (11.5.66), (11.5.67) and (11.5.68), verify that the geodesic Lagrangian now takes the form

$$L_g = \frac{1}{2}\left(\frac{r}{r - r_s} - \frac{\ell^2}{r^2(1 + r_s/r)}\left(1 + \left(\frac{r'(\phi)}{r}\right)^2\right)\right). \tag{11.5.69}$$

Closest approach vs angular momentum If r_0 is the coordinate radius of closest approach, which we shall assume is appreciably larger than the Schwarzschild radius $r_0 \gg r_s$, that means $r'(\phi) = 0$ when $r = r_0$. Show that

$$\ell = r_0\sqrt{\frac{r_0 + r_s}{r_0 - r_s}}. \tag{11.5.70}$$

An ODE Since null geodesics render $L_g = 0$, utilize eq. (11.5.70) in eq. (11.5.69), and proceed to show that – to first order in r_s –

$$\frac{\mathrm{d}\phi}{\mathrm{d}r} = \frac{1}{\sqrt{r^2 - r_0^2}} \left(\frac{r_0}{r} + \frac{r_s}{r + r_0} \right) + \mathcal{O}\left(r_s^2\right). \tag{11.5.71}$$

By integrating from infinity $r = \infty$ to closest approach $r = r_0$ and then out to infinity again $r = \infty$, show that the angular deflection is

$$\Delta\varphi = \frac{2r_s}{r_0}. \tag{11.5.72}$$

Note that, if the photon were undeflected, the total change in angle $(\int_{r=\infty}^{r_0} \mathrm{d}r + \int_{r_0}^{\infty} \mathrm{d}r)(\mathrm{d}\phi/\mathrm{d}r)$ would be π. Therefore, the total *deflection* angle is

$$\Delta\varphi = 2\left|\int_{r=\infty}^{r_0} \frac{\mathrm{d}\phi}{\mathrm{d}r}\mathrm{d}r\right| - \pi. \tag{11.5.73}$$

Physical vs Coordinate Radius Even though r_0 is the coordinate radius of closest approach, in a weakly curved spacetime dominated by the monopole moment of the central object, estimate the error incurred if we set r_0 to be the *physical* radius of closest approach. What is the angular deflection due to the Sun, if a beam of light were to just graze its surface?

Remark For further help on this problem, consult §8.5 of Weinberg [24]. □

Schwarzschild Horizon Let us now study strictly-radial motion to gain some understanding of the meaning of r_s. From eq. (11.5.41), and assuming $r \neq r_s$,

$$\left(\frac{\mathrm{d}t}{\mathrm{d}r}\right)^2 = \left(1 - \frac{r_s}{r}\right)^{-2}. \tag{11.5.74}$$

Far away from the black hole, $r \to \infty$, we recover the light cone in flat spacetime; i.e., $\mathrm{d}t/\mathrm{d}r$ describes the 45-degree slopes on the t versus r spacetime diagram:

$$\lim_{r\to\infty} \frac{\mathrm{d}t}{\mathrm{d}r} = \pm 1. \tag{11.5.75}$$

Problem 11.37. Massive Particles in Schwarzschild In this problem we will explore the geodesics of massive particles around a Schwarzschild black hole in eq. (11.5.34).

By specializing to the equatorial plane $\theta = \pi/2$ explain why, when evaluated on the geodesic solutions, we may choose

$$\frac{1}{2}\left\{ \left(1 - \frac{r_s}{r}\right) t'(\lambda)^2 - \frac{r'(\lambda)^2}{1 - \frac{r_s}{r}} - r^2\phi'(\lambda)^2 \right\} = \frac{1}{2}, \tag{11.5.76}$$

$$\left(1 - \frac{r_s}{r}\right) t'(\lambda) = \varepsilon, \qquad r^2\phi'(\lambda) = \ell; \tag{11.5.77}$$

where ε and ℓ are constants-of-motion respectively associated with time-translation and rotational symmetry.

From these relations, deduce the conservation of 'kinetic' $r'^2/2$ plus 'potential energy' $V/2$,

$$\frac{1}{2}r'(\lambda)^2 + \frac{1}{2}V = \frac{\varepsilon^2 - 1}{2} \tag{11.5.78}$$

$$V \equiv \frac{\ell^2}{r_s^2}\left(\left(\frac{r_s}{r}\right)^2 - \left(\frac{r_s}{r}\right)^3\right) - \frac{r_s}{r}. \tag{11.5.79}$$

(Note that the geodesic equation for $r(\lambda)$ is *not* 'Newton's 2nd law' $r'' = -(1/2)\partial_r V$, because the geodesic Lagrangian does not take the non-relativistic classical mechanics form $r'^2/2 - V/2$.) Since ℓ is arbitrary at this point, let us redefine

$$\ell/r_s \equiv \alpha \geq 0 \tag{11.5.80}$$

$$V \equiv \alpha^2\left(\left(\frac{r_s}{r}\right)^2 - \left(\frac{r_s}{r}\right)^3\right) - \frac{r_s}{r}. \tag{11.5.81}$$

Show that the turning points of V, i.e., satisfying $V'(r_\pm) = 0$, as well as the zeroth and second derivatives of V at $r_\pm$ are

$$\frac{r_\pm}{r_s} = \alpha\left(\alpha \pm \sqrt{\alpha^2 - 3}\right), \tag{11.5.82}$$

$$\frac{1}{2}V(r_\pm) = -\frac{1 \pm \sqrt{\alpha^2-3}\left(\alpha \pm \sqrt{\alpha^2-3}\right)}{2\alpha\left(\alpha \pm \sqrt{\alpha^2-3}\right)^3}, \tag{11.5.83}$$

$$\frac{1}{2}\left(V(r_-) - V(r_+)\right) = \frac{2}{27}\frac{(\alpha^2-3)^{\frac{3}{2}}}{\alpha} \geq 0, \tag{11.5.84}$$

$$\frac{1}{2}V''(r_\pm) = \pm\frac{\sqrt{\alpha^2-3}}{\alpha^3\left(\alpha - \sqrt{\alpha^2-3}\right)^4}. \tag{11.5.85}$$

For the moment, let us examine the case where $\alpha > \sqrt{3}$. The second derivative results tell us $V(r_-)$ is a local maximum and $V(r_+)$ is a local minimum. The minimum $V(r_+)$ is strictly negative; whereas the maximum $V(r_-)$ is negative for $\sqrt{3} \leq \alpha < 2$. (Can you show the latter statement regarding $V(r_-)$?) Moreover, since $V(r_-) - V(r_+) \geq 0$ that means the maximum is always higher than the minimum. Therefore, we see from eq. (11.5.81) that V is zero at $r/r_s = +\infty$; goes negative as we approach the black hole until the minimum at r_+ is reached before climbing up to the maximum at r_-; then as r grows even smaller, V plunges to $-\infty$.

Now, if $\alpha = \sqrt{3}$, the two $r_\pm$ merges to become a single $r_\pm = 3r_s$. The $(1/2)V = -1/18$ and $V'' = 0$ there. The potential $V/2$ therefore starts off at zero at $r = \infty$, goes negative as one approaches the black hole, reaches an inflection point at $r_\pm = 3r_s$, before plunging to negative infinity as $r/r_s \to 0$.

Can you make representative plots of $V/2$?

Physically, we may therefore divide orbits of massive particles in the follow manner.

- $\alpha = \sqrt{3}$: If $-1 \leq \varepsilon^2 - 1 < 0$ (or, in other words, $0 \leq \varepsilon^2 < 1$) the particle is bound to the black hole. Even if it were initially moving outwards, it will turn around and plunge into the black hole. If $\varepsilon^2 \geq 1$, on the other hand, if the particle were moving inwards it will plunge into the black hole; but if it were initially moving outwards instead, it will fly off to infinity.

- $\sqrt{3} < \alpha < 2$: If $V(r_+) \leq \varepsilon^2 - 1 < V(r_-)$ and $r > r_-$ the particle is bound to the black hole, but will not plunge into the black hole – can you figure out the minimum and maximum r of the orbit as a function of ε^2? If, on the other hand, $V(r_-) < \varepsilon^2 - 1 < 0$, then even if the particle were moving outward at first, it would turn around and plunge into the black hole. And if $\varepsilon^2 > 1$, if the particle were moving outward, it will fly off to infinity; whereas if it were moving inwards it will plunge into the black hole.
- $\alpha \geq 2$: If $V(r_+) \leq \varepsilon^2 - 1 < 0$ and $r_- < r < r_+$, the particle is bound to the black hole, but will not plunge into the black hole – can you figure out the minimum and maximum r of the orbit as a function of ε^2? If $V(r_-) > \varepsilon^2 - 1 > 0$ and $r > r_-$, then even if the particle were moving inward at first, it will turn around and fly out to infinity. (This is the relativistic analog of the hyperbolic unbound orbit in Newtonian gravity.) And if $\varepsilon^2 - 1 > V(r_-)$, if the particle were moving outward, it will fly off to infinity; whereas if it were moving inwards it will plunge into the black hole.

Innermost Stable Circular Orbit (ISCO) For $\alpha > \sqrt{3}$, since $V(r_+)$ is a minimum, a stable circular orbit is described by $\varepsilon^2 - 1 = V(r_+)$. On the other hand, since $V(r_-)$ is a maximum, an unstable circular orbit exists where $\varepsilon^2 - 1 = V(r_-)$. But when $\alpha = \sqrt{3}$, the two $r_\pm$ merges into $r_\pm = 3r_s$ and this point becomes an inflection point where $V'' = 0$. This situation allows for a marginally stable circular orbit at $r = 3r_s$ when $V = -1/9 = \varepsilon^2 - 1$. Explain why $r/r_s = 3$ is called the ISCO by showing that r_+/r_s is a strictly increasing function of α – i.e., as α is decreased r_+ is moved inwards until it merges with r_- when $\alpha = \sqrt{3}$. □

Problem 11.38. Precession of Perihelion at 1PN In this problem we shall take the weak field limit of eq. (11.5.58), the Schwarzschild black hole in isotropic coordinates, to study the precession of planetary orbits. Roughly speaking, once we include the $1/r^2$ General Relativistic corrections to the $1/r$ Newtonian potential, planetary orbits are no longer fixed ellipses. Rather, their (quasi-)elliptical orbits begin to rotate – aka 'precess'.

To begin, we first record the Keplarian solution of a test mass moving on the $(1,2)$ plane around a much larger one M, which generates a purely Newtonian potential. Replacing $G_{\rm N}(M_1 + M_2) \to r_s/2$ and $\phi_0 \to 0$ in equations (8.3.58), (8.3.75), (8.3.78), and (8.3.81),

$$r(\phi) = \sqrt[3]{\frac{r_s}{2\Omega^2}} \cdot \frac{1 - e^2}{1 + e \cdot \cos(\phi)}, \tag{11.5.86}$$

$$\dot{\phi}(t) = \Omega \frac{(1 + e \cdot \cos(\phi))^2}{(1 - e^2)^{3/2}}, \tag{11.5.87}$$

where the Cartesian position vector is $\vec{x} = r(\cos\phi, \sin\phi, 0)$ and Ω is the orbital angular frequency. We have arranged the perihelion (point of orbit nearest to the central mass M) to be at $\phi = 0$.

First verify that the unit-norm version of the Laplace-Runge-Lenz vector in eq. (8.3.81) reads

$$\widehat{A} \equiv \frac{\vec{A}}{|\vec{A}|} \tag{11.5.88}$$

$$= \left(\frac{2r^2\dot{\phi}\left(\dot{r}\sin(\phi) + r\dot{\phi}\cos(\phi)\right) - r_s\cos(\phi)}{\sqrt{4r^3\dot{\phi}^2\left(r\dot{r}^2 + r^3\phi'^2 - r_s\right) + r_s^2}}, \frac{2r^2\dot{\phi}\left(r\dot{\phi}\sin(\phi) - \dot{r}\cos(\phi)\right) - r_s\sin(\phi)}{\sqrt{4r^3\dot{\phi}^2\left(r\dot{r}^2 + r^3\dot{\phi}^2 - r_s\right) + r_s^2}}, 0 \right);$$

and, when evaluated on equations (11.5.86) and (11.5.87), yield

$$\widehat{A} = (1, 0, 0). \tag{11.5.89}$$

That is, $\widehat{A}$ is not only constant, it points towards the perihelion. Hence, if the precession of the perihelion may be obtained by simply computing the time rate of change in the unit-norm version of the Laplace-Runge-Lenz vector. Geometrically, we see that $\mathrm{d}\widehat{A}/\mathrm{d}t$ is perpendicular to $\widehat{A}$, since $\widehat{A}^2 = 1$ implies $\widehat{A} \cdot \mathrm{d}\widehat{A}/\mathrm{d}t = 0$. Moreover, the unit vector $(0, 0, 1)^{\mathrm{T}} \times \widehat{A}$ lies on the $(1, 2)$ plane and is perpendicular to $\widehat{A}$ too. We may thus deduce, the infinitesimal counter-clockwise angle swept out by the precession of perihelion is $\mathrm{d}\psi = ((0, 0, 1)^{\mathrm{T}} \times \widehat{A}) \cdot \mathrm{d}\widehat{A}$; and therefore

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = ((0, 0, 1)^{\mathrm{T}} \times \widehat{A}) \cdot \frac{\mathrm{d}\widehat{A}}{\mathrm{d}t}. \tag{11.5.90}$$

Next, keeping in mind the Newtonian gravity relation, that $\dot{\vec{x}}^2$ must scale as the Newtonian potential $r_s/(2r)$, show that the non-affine geodesic Lagrangian expanded to one order beyond Newton is

$$\begin{aligned} L'_g &\equiv 1 - \sqrt{g_{\mu\nu}(\mathrm{d}x^\mu/\mathrm{d}t)(\mathrm{d}x^\nu/\mathrm{d}t)} \\ &= \frac{1}{2}\left(\dot{\vec{x}}^2 + \frac{r_s}{r}\right) + \frac{v^2}{8}\left(\dot{\vec{x}}^4 + \frac{6r_s}{r}\dot{\vec{x}}^2 - \frac{r_s^2}{r^2}\right) + \mathcal{O}\left((r_s/r)^3\right); \end{aligned} \tag{11.5.91}$$

where v^2 is a fictitious parameter (to be set to unity at the end of the calculations below), to remind us the second group of terms scale as r_s/r or $\dot{\vec{x}}^2$ relative to the first. Use this 'first post-Newtonian' (1PN) Lagrangian to compute eq. (11.5.90), followed by evaluating the result on the Newtonian solutions in equations (11.5.86) and (11.5.87). You should obtain

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \left(\frac{r_s}{2}\right)^{2/3} \Omega^{5/3}(e\cos(\phi) + 1)^2 \frac{(e^2 - 3)\cos(\phi) + e(3 - 5\cos(2\phi))}{e\left(1 - e^2\right)^{5/2}}. \tag{11.5.92}$$

Note that this result would be *zero* if the Newtonian equations-of-motion were used, so the result must already be order v^2; this is why it is legitimate to evaluate them on equations (11.5.86) and (11.5.87) – the relativistic corrections to these Newtonian solutions would only yield higher corrections to $\mathrm{d}\psi/\mathrm{d}t$ itself.

Integrate this result over one orbit and show that

$$\Delta\psi \equiv \int_{1\text{ period}} \frac{\mathrm{d}\psi}{\mathrm{d}t}\mathrm{d}t = \int_0^{2\pi} \frac{\mathrm{d}\psi}{\mathrm{d}t}\frac{\mathrm{d}\phi}{\dot{\phi}} \tag{11.5.93}$$

$$= \frac{3\pi}{1 - e^2}\sqrt[3]{2(\Omega \cdot r_s)^2} = \frac{6\pi}{1 - e^2}(\Omega \cdot G_{\mathrm{N}} M)^{2/3}. \tag{11.5.94}$$

Plug in the relevant parameters for the planet Mercury – what do you obtain? The answer should be very close to 43 arc seconds per century. □

Precession of gyroscope spin Imagine allowing a small gyroscope to 'free-fall' in a curved spacetime – what determines the evolution of its spin vector S^μ? If the spacetime were actually flat, then we expect S^μ to simply be constant in time: $U^\mu \partial_\mu S^\alpha = \mathrm{d}S^\mu/\mathrm{d}\tau = 0$, where U^μ is tangent to the gyroscope's geodesic world line. The straightforward generalization to a curved spacetime is the postulate

$$U^\mu \nabla_\mu S^\alpha = \frac{\mathrm{d}S^\alpha}{\mathrm{d}\tau} + U^\mu \Gamma^\alpha{}_{\mu\beta} S^\beta = 0. \tag{11.5.95}$$

Note that S^α is associated with a spatial vector – the rotation of our gyroscope. (In higher dimensions S^α will have to be associated with a rank-2 tensor.) In flat spacetime, we would have $S^\alpha = (0, \vec{S})$, for instance. Whereas in curved spacetime we would instead demand

$$S^\alpha U_\alpha = 0. \tag{11.5.96}$$

We may readily check that this constraint, once imposed, will be preserved under evolution along the gyroscope's world line,

$$\nabla_U (S \cdot U) = U^\alpha \nabla_U S_\alpha + S^\alpha \nabla_U U_\alpha = 0. \tag{11.5.97}$$

Furthermore, if we compute

$$\nabla_U (S \cdot S) = 2 S^\alpha \nabla_U S_\alpha = 0, \tag{11.5.98}$$

we see that the magnitude of S^α – i.e., how fast the gyroscope is spinning – does not change under parallel transport. Hence, if we are merely interested in where the spin is pointing, not how fast it is spinning, we would further normalize it to

$$S^2 \equiv S^\alpha S_\alpha = -1. \tag{11.5.99}$$

In the following problem, let us solve the parallel transport eq. (11.5.95) subject to the constraints equations (11.5.96) and (11.5.99) for gyroscope placed in a circular orbit around the Schwarzschild black hole.

Problem 11.39. Let S^α be associated with a gyroscope's normalized spin vector. If the gyroscope itself is 'free-falling' in eq. (11.5.34) along a circular orbit at coordinate radius r_0, demonstrate that its parallel transported spin vector precesses only on the plane containing the orbit itself. Moreover, after one orbit, the spin vector has precessed by an angle

$$\Delta\phi = 2\pi \left(1 - \sqrt{1 - \frac{3 r_s}{2 r_0}}\right). \tag{11.5.100}$$

The spin component perpendicular to the orbit remains constant.

Remarks This precession of the gyroscope's spin may be attributed to its interaction between the gyroscope's orbital angular momentum around the central mass. The interaction between the gyroscope spin and rotational spin of the central mass would give rise to additional precession effects on the former.

Hints: By spherical symmetry, first solve the radial orbit on the $\theta = \pi/2$ plane and show that the geodesic trajectory Z^μ of the gyroscope is given by – for fixed initial time t_0 and azimuth angle ϕ_0 –

$$Z^\mu(\tau) = \left(t_0 + \tau \sqrt{\frac{r_0/r_s}{(r_0/r_s) - 3/2}}, r_0, \frac{\pi}{2}, \phi_0 + \frac{\tau}{r_0 \sqrt{2(r_0/r_s) - 3}} \right); \tag{11.5.101}$$

with coordinate velocity

$$U^\mu = \left(\sqrt{\frac{r_0/r_s}{(r_0/r_s) - 3/2}}, 0, 0, \frac{1}{r_0\sqrt{2(r_0/r_s) - 3}} \right). \tag{11.5.102}$$

Next, show that $\Gamma^\theta{}_{\alpha\beta} U^\alpha S^\beta = 0$ and explain why S^θ is constant.

Then, compute the other three eigenvectors V^μ_{I} (where $\mathrm{I} \in \{1, 2, 3\}$) of the 'matrix' $U^\alpha \Gamma^\mu{}_{\alpha\nu}$; and use it to solve

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \left(S^{\mathrm{I}} V^\mu_{\mathrm{I}} \right) = -U^\alpha \Gamma^\mu{}_{\alpha\beta} \left(S^{\mathrm{I}} V^\beta_{\mathrm{I}} \right). \tag{11.5.103}$$

After ensuring S^α is real and imposing the constraints $S^2 = -1$ and $S \cdot U = 0$, you should arrive at the solutions

$$S^t = \frac{1}{\sqrt{1 - r_s/r_0}} \sqrt{\frac{1 - (S^\theta_0)^2}{2(r_0/r_s) - 3}} \cos\left(\frac{\tau}{\sqrt{2}(r_0/r_s)^{3/2} r_s} + \varphi \right), \tag{11.5.104}$$

$$S^r = \sqrt{\left(1 - \frac{r_s}{r_0}\right) \left(1 - (S^\theta_0)^2\right)} \sin\left(\frac{\tau}{\sqrt{2}(r_0/r_s)^{3/2} r_s} + \varphi \right), \tag{11.5.105}$$

$$S^\theta = \frac{S^\theta_0}{r_0}, \tag{11.5.106}$$

$$S^\phi = \frac{1}{r_0} \sqrt{\frac{((r_0/r_s) - 1)\left(1 - (S^\theta_0)^2\right)}{(r_0/r_s) - 3/2}} \cos\left(\frac{\tau}{\sqrt{2}(r_0/r_s)^{3/2} r_s} + \varphi \right). \tag{11.5.107}$$

where S^θ_0 is constant and φ is an arbitrary phase.

Explain why eq. (11.5.101) implies the orbital period is

$$T_O = 2\pi r_0 \sqrt{2(r_0/r_s) - 3}; \tag{11.5.108}$$

whereas equations (11.5.104)–(11.5.107) imply the time it takes for the spin to return to its original configuration is.

$$T_S = 2\pi\sqrt{2}(r_0/r_s)^{3/2} r_s. \tag{11.5.109}$$

Comment on whether T_O or T_S larger, and what that says about the direction of spin precession – with or against the orbital motion. Explain why, if $\varepsilon^{\widehat{\phi}}{}_\mu$ is the orthogonal frame field in the $\theta-$direction, the $H_{\mu\nu}$ in

$$-H_{\mu\nu} = g_{\mu\nu} - U_\mu U_\nu + \varepsilon^{\widehat{\phi}}{}_\mu \varepsilon^{\widehat{\phi}}{}_\nu \tag{11.5.110}$$

is the spatial metric orthogonal to both the worldline of the gyroscope and to the $\theta-$direction. Exploit this spatial metric to explain why the precession angle is gotten by comparing the spin vector with itself after one orbit:

$$\cos[\Delta\phi] = \frac{S^\mu(\lambda = T_O) S^\nu(\lambda = 0) H_{\mu\nu}(\theta = \pi/2)}{S^\alpha(\lambda) S^\beta(\lambda) H_{\alpha\beta}(\theta = \pi/2)}; \tag{11.5.111}$$

with the λ in the denominator being arbitrary. After computing the right hand side, you should be able to read off $\Delta\phi$ in eq. (11.5.100) from the result. □

Killing Tensors A rank$-N$ Killing tensor $K_{\mu_1 \dots \mu_N}$ is a fully symmetric object that satisfies

$$\nabla_{\{\mu} K_{\nu_1 \dots \nu_N\}} = 0. \tag{11.5.112}$$

A motivation for this definition is the following. If v^α is an affinely parametrized geodesic vector field (such that $v^\sigma \nabla_\sigma v^\alpha = 0$), then a conserved quantity along its integral curve may be constructed from $K_{\mu_1 \dots \mu_N}$ via the prescription

$$K_{\mu_1 \dots \mu_N} v^{\mu_1} \dots v^{\mu_N}. \tag{11.5.113}$$

For, we may compute

$$v^\sigma \nabla_\sigma \left(K_{\mu_1 \dots \mu_N} v^{\mu_1} \dots v^{\mu_N} \right) = v^\sigma \left(\nabla_\sigma K_{\mu_1 \dots \mu_N} \right) v^{\mu_1} \dots v^{\mu_N} \tag{11.5.114}$$

$$= \frac{1}{(N+1)!} v^\sigma \left(\nabla_{\{\sigma} K_{\mu_1 \dots \mu_N\}} \right) v^{\mu_1} \dots v^{\mu_N} = 0. \tag{11.5.115}$$

A simple example of a rank-2 Killing tensor is the metric $g_{\alpha\beta}$ itself. Since we are assuming ∇ to be metric compatible, Problem (11.11) tells us $v^2 \equiv g_{\alpha\beta} v^\alpha v^\beta$ is in fact constant if v^α is an affinely parametrized geodesic.

Problem 11.40. Cosmological Killing Tensor For the spatially flat cosmology of

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}, \tag{11.5.116}$$

verify that

$$K_{\mu\nu} = a(t)^2 \left(g_{\mu\nu} - \delta^0_\mu \delta^0_\nu \right) \tag{11.5.117}$$

is a Killing tensor. □

Problem 11.41. Carter Constant and the Kerr Metric

11.6 ⋆Symmetries and Spacetimes

Maximally symmetric *spaces* were constructed in §(9.8). In this section, we shall extend the construction to spacetimes with maximally symmetric *sub-spaces*; as well as maximally symmetric *spacetimes*.

11.6.1 Spacetimes with Maximally Symmetric Subspaces

General Ansatz Suppose $g_{\mathrm{AB}}(y) \mathrm{d}y^{\mathrm{A}} \mathrm{d}y^{\mathrm{B}}$ were a maximally symmetric space, and suppose the subspace of some spacetime $g_{\alpha\beta}(x = (y, z))$ contained $g_{\mathrm{AB}}(y)$. Then, in general, it must take the form

$$\mathrm{d}s^2 = H_{\mathrm{M'N'}}(z) \mathrm{d}z^{\mathrm{M}} \mathrm{d}z^{\mathrm{N}} + F(z) g_{\mathrm{AB}}(y) \mathrm{d}y^{\mathrm{A}} \mathrm{d}y^{\mathrm{B}}, \tag{11.6.1}$$

where the full coordinate system $\{x^\mu\}$ is the union of the $\{y^{\mathrm{A}}\}$ and $\{z^{\mathrm{A}}\}$ ones. The key applications we shall consider are to cosmology and the time-independent and spherically symmetric spacetimes.

4D Cosmology To model the geometry of our universe on the largest length scales, cosmologists usually assume, at a fixed observer time t, that space itself is maximally symmetric. Equivalently, at a fixed time t, the universe is homogeneous (same geometry everywhere in space) and isotropic (rotation-symmetric at every point in space). From §(9.8), we saw that a maximally symmetric 3−space is either a sphere or a hyperboloid or flat space itself. Exploiting the ansatz in eq. (11.6.1), we have

$$\mathrm{d}s^2 = A(t)\mathrm{d}t^2 - B(t)\left(\mathrm{d}\vec{x}^2 \mp \frac{(R^{-1}\vec{x}\cdot\mathrm{d}\vec{x})^2}{1 \pm \vec{x}^2/R^2}\right), \tag{11.6.2}$$

where the $+$ sign is for the negatively curved 3−subspace and $-$ is for the negatively curved one; or,

$$\mathrm{d}s^2 = A(t)\mathrm{d}t^2 - B(t)\mathrm{d}\vec{x}^2 \tag{11.6.3}$$

for the flat one. In order for the $\mathrm{d}t^2$ term to describe the time-time component of the metric, and the rest of the terms to describe the spatial geometry, we need $A, B > 0$.

Starting with the spatially flat metric, we may readily redefine $t \to \int^t \mathrm{d}t/\sqrt{A(t)}$, so that

$$A(t)\mathrm{d}t^2 \to \mathrm{d}t^2; \tag{11.6.4}$$

and upon redefining $B(t) \to a(t)^2$, we arrive at

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2\mathrm{d}\vec{x}^2 \qquad \text{(Spatially flat cosmology)}. \tag{11.6.5}$$

Problem 11.42. Positive Spatial Curvature Explain why a 4D universe with a positively curved maximally symmetric space may be expressed as

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}t^2 - a(t)^2\left(\mathrm{d}\vec{x}^2 + \frac{(R^{-1}\vec{x}\cdot\mathrm{d}\vec{x})^2}{1 - \vec{x}^2/R^2}\right) & (11.6.6)\\ &= \mathrm{d}t^2 - a(t)^2\left(\mathrm{d}r^2 + R^2\sin^2(r/R)\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right)\right), & (11.6.7)\end{aligned}$$

where the second equality employs the spherical coordinates $\vec{x} = R\sin(r/R)\widehat{r}(\theta,\phi)$, with $0 \leq r/R, \theta \leq \pi$ and $\phi \in [0, 2\pi)$. □

Problem 11.43. Negative Spatial Curvature Explain why a 4D universe with a negatively curved maximally symmetric space may be expressed as

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}t^2 - a(t)^2\left(\mathrm{d}\vec{x}^2 - \frac{(R^{-1}\vec{x}\cdot\mathrm{d}\vec{x})^2}{1 + \vec{x}^2/R^2}\right) & (11.6.8)\\ &= \mathrm{d}t^2 - a(t)^2\left(\mathrm{d}r^2 + R^2\sinh^2(r/R)\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right)\right), & (11.6.9)\end{aligned}$$

where the second equality employs the hyperbolic-spherical coordinates $\vec{x} = R\sinh(r/R)\widehat{r}(\theta,\phi)$, with $r \geq 0$, $0 \leq \theta \leq \pi$ and $\phi \in [0, 2\pi)$. □

Problem 11.44. Empty Universe with Positive Cosmological Constant For $\Lambda > 0$, solve

$$G^{\mu}{}_{\nu}[g] = \Lambda\delta^{\mu}{}_{\nu}, \tag{11.6.10}$$

where $G^{\mu}{}_{\nu}$ is Einstein's tensor, with zero, positive, and negative spatial curvature – i.e., solve for the $a(t)$ in equations (11.6.5), (11.6.6) (or (11.6.7)), and (11.6.8) (or (11.6.9)). You should find

$$a(t) = \exp\left[\pm\sqrt{\frac{\Lambda}{3}}(t - t_0)\right], \qquad \text{(Flat)}, \tag{11.6.11}$$

$$= \frac{1}{R}\sqrt{\frac{3}{\Lambda}}\cosh\left[\sqrt{\frac{\Lambda}{3}}(t - t_0)\right], \qquad \text{(Positive)}, \tag{11.6.12}$$

$$= \frac{1}{R}\sqrt{\frac{3}{\Lambda}}\sinh\left[\sqrt{\frac{\Lambda}{3}}(t - t_0)\right], \qquad \text{(Negative)}. \tag{11.6.13}$$

In the next section we shall discover these solutions correspond to a *maximally symmetric* spacetime; specifically, de Sitter spacetime with $\Lambda > 0$. Notice, when $\Lambda = 0$, flat spacetime is recovered. □

Problem 11.45. 4D Spherically Symmetric Metric By demanding a 4D metric to be spherically symmetric argue that it must take the form

$$\mathrm{d}s^2 = A(\tau, r)\mathrm{d}\tau^2 - B(\tau, r)\mathrm{d}r^2 - 2F(\tau, r)\mathrm{d}\tau\mathrm{d}r - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right), \tag{11.6.14}$$

$$\theta \in [0, \pi], \qquad \phi \in [0, 2\pi). \tag{11.6.15}$$

(Be sure to explain why the factor multiplying the angular terms can be made independent of the time τ.) Next, argue that it may be put into the diagonal form

$$\mathrm{d}s^2 = P(t, r)\mathrm{d}t^2 - Q(t, r)\mathrm{d}r^2 - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right). \tag{11.6.16}$$

To ensure this is a spacetime, i.e., with 1 positive and 3 negative eigenvalues, we need both P and Q to be positive. This result is the starting point for studying the geometry sourced by any spherically symmetric body, including that of the non-rotating black hole.

To go from eq. (11.6.14) to eq. (11.6.16), you may wish to consider

$$\tau \equiv t + G(t, r), \tag{11.6.17}$$

where G itself satisfies the first order differential equation

$$\partial_r G = \frac{F(t + G, r)}{A(t + G, r)}. \tag{11.6.18}$$

□

11.6.2 Maximally Symmetric Spacetimes

In d dimensional spacetime, there can be at most $d(d+1)/2$ Killing vectors. A spacetime that has $d(d+1)/2$ Killing vectors is said to be maximally symmetric. Flat spacetime is such an example, with d translation generators $\{\partial_\mu\}$, and $(d^2 - d)/2$ Lorentz transformation generators $\{x^{[\mu}\partial^{\nu]}\}$ yielding a total of $(d^2 + d)/2$ Killing vectors. Its Ricci scalar is zero. In this section, we shall construct maximally symmetric spacetimes with non-zero Ricci scalars. Because maximally symmetric spacetimes are unique, it does not matter *how* we do it.

We first record that, maximally symmetric spacetimes must necessarily have constant Ricci scalars; i.e., $\partial_\mu \mathcal{R} = 0$. Their Riemann and Ricci tensors are then, respectively,

$$R_{\alpha\beta\mu\nu} = \frac{\mathcal{R}}{d(d-1)} g_{\alpha[\mu} g_{\nu]\beta} \qquad \text{and} \qquad R_{\alpha\beta} = \frac{\mathcal{R}}{d} g_{\alpha\beta}. \tag{11.6.19}$$

Negative Ricci Scalar: de Sitter Spacetime We shall witness that the $d-$dimensional negatively curved maximally symmetric spacetime may be viewed as a hyperboloid situated in $(d+1)-$dimensional Minkowski spacetime,

$$-\eta_{\mathrm{AB}} X^{\mathrm{A}} X^{\mathrm{B}} \equiv -\eta_{\mu\nu} X^\mu X^\nu + W^2 = H^{-2}, \tag{11.6.20}$$

$$(\mathrm{d}s^2)_{d+1} = \eta_{\mathrm{AB}} \mathrm{d}X^{\mathrm{A}} \mathrm{d}X^{\mathrm{B}}; \tag{11.6.21}$$

where the capital letters A, B, ... run over $\{0, 1, \dots, d\}$; the Greek letters $\mu, \nu, \dots$ over $\{0, 1, \dots, d-1\}$; $W \equiv X^d$; and H is usually dubbed the Hubble constant. We see why this embedding perspective is useful for enumerating the number of symmetry directions – the hyperboloid in eq. (11.6.20) is *Lorentz invariant* in the higher $(d+1)-$Minkowski, with the $((d+1)^2 - (d+1))/2 = (d+1)d/2$ Killing vectors

$$J^{\mathrm{AB}} \equiv X^{[\mathrm{A}} \partial^{\mathrm{B}]}; \tag{11.6.22}$$

i.e., it is rotation and boost invariant.

Turning to the $d-$dimensional metric itself, we first solving for W,

$$W = \pm H^{-1} \sqrt{1 + H^2 X^2}, \tag{11.6.23}$$

which implies

$$(\mathrm{d}s^2)[d - \text{dim. max. symm.}] = \eta_{\mu\nu} \mathrm{d}X^\mu \mathrm{d}X^\nu - \frac{H^2 (X_\mu \mathrm{d}X^\mu)^2}{1 + H^2 X^2}, \tag{11.6.24}$$

$$g_{\mu\nu} = \eta_{\mu\nu} - \frac{H^2 X_\mu X_\nu}{1 + H^2 X^2}. \tag{11.6.25}$$

A direct computation would reveal

$$\Gamma^\mu{}_{\alpha\beta} = -H^2 X^\mu g_{\alpha\beta}. \tag{11.6.26}$$

If we re-scale the affine parameter λ of some geodesic $Z^\mu(\lambda)$ such that

$$g_{\mu\nu} \frac{\mathrm{d}Z^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}Z^\nu}{\mathrm{d}\lambda} = 1 \qquad \text{(Timelike)} \tag{11.6.27}$$

$$= -1 \qquad \text{(Spacelike)} \tag{11.6.28}$$

$$= 0 \qquad \text{(Null)}; \tag{11.6.29}$$

the corresponding geodesic equations become

$$\ddot{Z}^\mu = H^2 Z^\mu \qquad \text{(Timelike)} \tag{11.6.30}$$

$$= -H^2 Z^\mu \qquad \text{(Spacelike)} \tag{11.6.31}$$

$$= 0 \qquad \text{(Null)}. \tag{11.6.32}$$

If we solve the boundary value problem

$$Z^\mu(\lambda_0) = X_0 \qquad \text{and} \qquad Z^\mu(\lambda_1) = X_1; \tag{11.6.33}$$

the timelike solutions are

$$Z^\mu(\lambda) = X_0^\mu \cosh(H(\lambda - \lambda_0)) + \frac{X_1^\mu - X_0^\mu \cosh(H(\lambda_1 - \lambda_0))}{\sinh(H(\lambda_1 - \lambda_0))} \sinh(H(\lambda - \lambda_0)); \tag{11.6.34}$$

the spacelike solutions are

$$Z^\mu(\lambda) = X_0^\mu \cos(H(\lambda - \lambda_0)) + \frac{X_1^\mu - X_0^\mu \cos(H(\lambda_1 - \lambda_0))}{\sin(H(\lambda_1 - \lambda_0))} \sin(H(\lambda - \lambda_0)); \tag{11.6.35}$$

and the null solutions are

$$Z^\mu(\lambda) = X_0^\mu + \frac{\lambda - \lambda_0}{\lambda_1 - \lambda_0}(X_1^\mu - X_0^\mu). \tag{11.6.36}$$

Problem 11.46. Geodesic Intervals Use the constraint $g_{\mu\nu}\dot{Z}^\mu \dot{Z}^\nu = 0, \pm 1$, to solve for $\lambda_1 - \lambda_0$; and for the timelike and spacelike cases, the corresponding world functions σ. □

11.7 Additional Problems

Problem 11.47. Exterior Derivative, Divergence, and Poincaré Lemma In d spacetime dimensions, show that the curl of a vector contracted with the Levi-Civita tensor $\widetilde{\epsilon}$ (i.e., the volume form) yields its divergence multiplied by the same $\widetilde{\epsilon}$:

$$((d-1)!)^{-1} \partial_{[\alpha_1} \left(J^\mu \widetilde{\epsilon}_{\alpha_2 \dots \alpha_d]\mu} \right) = (-)^{d-1} (\nabla_\sigma J^\sigma) \widetilde{\epsilon}_{\alpha_1 \dots \alpha_d}. \tag{11.7.1}$$

(In differential form notation, one denotes $\widetilde{\epsilon}[J]_{\alpha_2 \dots \alpha_d} \equiv J^\sigma \widetilde{\epsilon}_{\sigma \alpha_2 \dots \alpha_d}$ and eq. (11.7.1) is short-handed as $\mathrm{d}\widetilde{\epsilon}[J] = (\nabla_\sigma J^\sigma)\widetilde{\epsilon}$.) Hint: Any rank d fully anti-symmetric tensor in $d-$dimensions must be proportional to $\widetilde{\epsilon}$; hence, the problem reduces to the determination of its coefficient.

If $\nabla_\sigma J^\sigma = 0$, what does the Poincaré lemma tell us about eq. (11.7.1)? Find the dual of your result and argue there must an antisymmetric tensor $\Sigma^{\mu\nu}$ such that

$$J^\mu = \nabla_\nu \Sigma^{\mu\nu}. \tag{11.7.2}$$

Noether Currents are Ambiguous This result teaches us, if we wish to construct an identically conserved current, it must be the divergence of an anti-symmetric rank-2 tensor. This in turn means, Noether currents – conserved currents arising from continuous symmetries – are inherently ambiguous because, if J^μ is conserved so is $J^\mu + \nabla_\nu K^{\mu\nu}$. In particular, since $\nabla_\mu \nabla_\nu K^{\mu\nu}$ is zero, and since Noether currents are derived by reading them off a divergence equation of the form $\nabla_\sigma J^\sigma = 0$, we cannot tell apart $\nabla_\mu J^\mu$ and $\nabla_\mu (J^\mu + \nabla_\nu K^{\mu\nu})$.

> If J^μ is a Noether current corresponding to some continuous symmetry, then so is $J^\mu + \nabla_\nu K^{\mu\nu}$ for arbitrary but anti-symmetric $K^{\mu\nu}$. □

Problem 11.48. Gauge-covariant derivative Let ψ be a vector under *group* transformations. By this we mean that, if $\psi^{\check{a}}$ corresponds to the ath component of ψ, then given some matrix $U^{\check{a}}{}_{\check{b}}$, ψ transforms as

$$\psi^{\check{a}'} = U^{\check{a}'}{}_{\check{b}}\psi^{\check{b}} \qquad \text{(or, } \psi' = U\psi\text{)}\,. \tag{11.7.3}$$

Compare eq. (11.7.3) to how a spacetime vector transforms under coordinate transformations:

$$V^{\mu'}(x') = \mathcal{J}^{\mu'}{}_{\sigma}V^{\sigma}(x), \qquad \mathcal{J}^{\mu}{}_{\sigma} \equiv \frac{\partial x'^{\mu}}{\partial x^{\sigma}}. \tag{11.7.4}$$

Now, let us consider taking the gauge-covariant derivative $\check{D}$ of ψ such that it still transforms 'covariantly' under group transformations, namely

$$\check{D}_{\alpha}\psi' = \check{D}_{\alpha}(U\psi) = U(\check{D}_{\alpha}\psi). \tag{11.7.5}$$

Crucially:

> We shall now demand that the gauge-covariant derivative transforms covariantly – i.e., eq. (11.7.5) holds – even when the group transformation $U(x)$ depends on spacetime coordinates.

First check that, the spacetime-covariant derivative cannot be equal to the gauge-covariant derivative in general, i.e.,

$$\nabla_{\alpha}\psi' \neq \check{D}_{\alpha}\psi', \tag{11.7.6}$$

by showing that eq. (11.7.5) is not satisfied.

Just as the spacetime-covariant derivative was built from the partial derivative by adding a Christoffel symbol, $\nabla = \partial + \Gamma$, we may build a gauge-covariant derivative by adding to the spacetime-covariant derivative a *gauge potential*:

$$(\check{D}_{\mu})^{\check{a}}{}_{\check{b}} \equiv \delta^{a}_{b}\nabla_{\mu} + (A_{\mu})^{\check{a}}{}_{\check{b}}. \tag{11.7.7}$$

Or, in gauge-index-free notation,

$$\check{D}_{\mu} \equiv \nabla_{\mu} + A_{\mu}. \tag{11.7.8}$$

With the definition in eq. (11.7.7), how must the gauge potential A_{μ} (or, equivalently, $(A_{\mu})^{\check{a}}{}_{\check{b}}$) transform so that eq. (11.7.5) is satisfied? Compare the answer to the transformation properties of the Christoffel symbol in eq. (11.3.6). (Since the answer can be found in most Quantum Field Theory textbooks, make sure you verify the covariance explicitly!)

Bonus: Here, we have treated ψ as a spacetime scalar and the gauge-covariant derivative $\check{D}_{\alpha}$ itself as a scalar under group transformations. Can you generalize the analysis here to the higher-rank tensor case? □

11.8 *Metric Perturbation Theory

Carrying out perturbation theory about some fixed 'background' geometry $\bar{g}_{\mu\nu}$ has important physical applications. As such, in this section, we will in fact proceed to set up a general and systematic perturbation theory involving the metric:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{11.8.1}$$

where $\bar{g}_{\mu\nu}$ is an arbitrary 'background' metric and $h_{\mu\nu}$ is a small deviation. I will also take the opportunity to discuss the transformation properties of $h_{\mu\nu}$ under infinitesimal coordinate transformations, i.e., the gauge transformations of gravitons.

Moving Tensor Indices To ensure covariance of tensors, its indices are moved with the full metric $g_{\mu\nu}$; namely, $V_\mu = g_{\mu\nu}V^\nu$ and $V^\mu = g^{\mu\nu}V_\nu$. But because the metric involves a perturbation $h_{\mu\nu}$, as we shall soon see, its inverse is an infinite series. This makes computations tricky – both technically and conceptually – especially so when the tensors involved themselves admit perturbative series. As such, we shall resort to moving the indices of individual terms with the background metric $\bar{g}_{\mu\nu}$ and its inverse $\bar{g}^{\mu\nu}$; while ensuring that the full tensor(s)' indices are still moved by $g_{\mu\nu}$ and $g^{\mu\nu}$.

Metric inverse Starting with $h_{\mu\nu}$, we shall move its with the $\bar{g}_{\alpha\beta}$ in the following manner.

$$h^\alpha{}_\beta \equiv \bar{g}^{\alpha\sigma} h_{\sigma\beta}, \qquad \text{and} \qquad h^{\alpha\beta} \equiv \bar{g}^{\alpha\sigma}\bar{g}^{\beta\rho} h_{\sigma\rho}. \tag{11.8.2}$$

With this convention in place, let us note that the inverse metric is a geometric series. Firstly,

$$g_{\mu\nu} = \bar{g}_{\mu\sigma}\left(\delta^\sigma_\nu + h^\sigma{}_\nu\right) \doteq \bar{g} \cdot (\mathbb{I} + \mathsf{h}). \tag{11.8.3}$$

(Here, h is a matrix, whose μth row and νth column is $h^\mu{}_\nu \equiv \bar{g}^{\mu\sigma} h_{\sigma\nu}$.) Remember that, for invertible matrices A and B, we have $(A \cdot B)^{-1} = B^{-1}A^{-1}$. Therefore

$$g^{-1} = (\mathbb{I} + \mathsf{h})^{-1} \cdot \bar{g}^{-1}. \tag{11.8.4}$$

If we were dealing with numbers instead of matrices, the geometric series $1/(1+z) = \sum_{\ell=0}^\infty (-)^\ell z^\ell$ may come to mind. You may directly verify that this prescription, in fact, still works.

$$g^{\mu\nu} = \left(\delta^\mu{}_\lambda + \sum_{\ell=1}^\infty (-)^\ell h^\mu{}_{\sigma_1} h^{\sigma_1}{}_{\sigma_2} \dots h^{\sigma_{\ell-2}}{}_{\sigma_{\ell-1}} h^{\sigma_{\ell-1}}{}_\lambda\right) \bar{g}^{\lambda\nu} \tag{11.8.5}$$

$$= \bar{g}^{\mu\nu} + \sum_{\ell=1}^\infty (-)^\ell h^\mu{}_{\sigma_1} h^{\sigma_1}{}_{\sigma_2} \dots h^{\sigma_{\ell-2}}{}_{\sigma_{\ell-1}} h^{\sigma_{\ell-1}\nu} \tag{11.8.6}$$

$$= \bar{g}^{\mu\nu} - h^{\mu\nu} + h^\mu{}_{\sigma_1} h^{\sigma_1\nu} - h^\mu{}_{\sigma_1} h^{\sigma_1}{}_{\sigma_2} h^{\sigma_2\nu} + \dots. \tag{11.8.7}$$

Lowering a tensor index involves two terms,

$$V_\mu = g_{\mu\nu}V^\nu = (\bar{g}_{\mu\nu} + h_{\mu\nu})V^\nu; \tag{11.8.8}$$

whereas raising it involves infinite ones,

$$V^\mu = g^{\mu\nu}V_\nu = (\bar{g}^{\mu\nu} - h^{\mu\nu} + h^{\mu\sigma} h_\sigma{}^\nu + \dots)V_\nu. \tag{11.8.9}$$

Metric Determinant The square root of the determinant of the metric can be computed order-by-order in perturbation theory via the following formula. For any matrix A,

$$\det A = \exp\left[\mathrm{Tr}\left[\ln A\right]\right], \tag{11.8.10}$$

where Tr is the matrix trace; for e.g., $\mathrm{Tr}\left[\mathsf{h}\right] = h^{\sigma}{}_{\sigma}$. Taking the determinant of both sides of eq. (11.8.3), and using the property $\det[A \cdot B] = \det A \cdot \det B$,

$$\det g_{\alpha\beta} = \det \bar{g}_{\alpha\beta} \cdot \det\left[\mathbb{I} + \mathsf{h}\right], \tag{11.8.11}$$

so that eq. (11.8.10) can be employed to state

$$\sqrt{|g|} = \sqrt{|\bar{g}|} \cdot \exp\left[\frac{1}{2}\mathrm{Tr}\left[\ln[\mathbb{I} + \mathsf{h}]\right]\right]. \tag{11.8.12}$$

The first few terms read

$$\begin{aligned} \sqrt{|g|} &= \sqrt{|\bar{g}|}\bigg(1 + \frac{1}{2}h + \frac{1}{8}h^2 - \frac{1}{4}h^{\sigma\rho}h_{\sigma\rho} \\ &\qquad + \frac{1}{48}h^3 - \frac{1}{8}h \cdot h^{\sigma\rho}h_{\sigma\rho} + \frac{1}{6}h^{\sigma\rho}h_{\rho\kappa}h^{\kappa}{}_{\sigma} + \mathcal{O}[h^4]\bigg) \qquad (11.8.13) \\ h &\equiv h^{\sigma}{}_{\sigma}. \qquad (11.8.14) \end{aligned}$$

Covariance, Covariant Derivatives, Geometric Tensors Under a coordinate transformation $x \equiv x(x')$, the full metric of course transforms as a tensor. The full metric $g_{\alpha'\beta'}$ in this new x' coordinate system reads

$$g_{\alpha'\beta'}(x') = \left(\bar{g}_{\mu\nu}(x(x')) + h_{\mu\nu}(x(x'))\right)\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}. \tag{11.8.15}$$

If we define the 'background metric' to transform covariantly; namely

$$\bar{g}_{\alpha'\beta'}(x') \equiv \bar{g}_{\mu\nu}(x(x'))\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}; \tag{11.8.16}$$

then, from eq. (11.8.15), the perturbation itself can be treated as a tensor

$$h_{\alpha'\beta'}(x') = h_{\mu\nu}(x(x'))\frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}. \tag{11.8.17}$$

These will now guide us to construct the geometric tensors – the full Riemann tensor, Ricci tensor and Ricci scalar – using the covariant derivative $\overline{\nabla}$ with respect to the 'background metric' $\bar{g}_{\mu\nu}$ and its associated geometric tensors. Let's begin by considering this background covariant derivative acting on the full metric in eq. (11.8.1):

$$\overline{\nabla}_\alpha g_{\mu\nu} = \overline{\nabla}_\alpha\left(\bar{g}_{\mu\nu} + h_{\mu\nu}\right) = \overline{\nabla}_\alpha h_{\mu\nu}. \tag{11.8.18}$$

On the other hand, the usual rules of covariant differentiation tell us

$$\overline{\nabla}_\alpha g_{\mu\nu} = \partial_\alpha g_{\mu\nu} - \overline{\Gamma}^{\sigma}{}_{\alpha\mu}g_{\sigma\nu} - \overline{\Gamma}^{\sigma}{}_{\alpha\nu}g_{\mu\sigma}; \tag{11.8.19}$$

where the Christoffel symbols here are built out of the 'background metric',

$$\overline{\Gamma}^{\sigma}{}_{\alpha\mu} = \frac{1}{2}\bar{g}^{\sigma\lambda}\left(\partial_\alpha\bar{g}_{\mu\lambda} + \partial_\mu\bar{g}_{\alpha\lambda} - \partial_\lambda\bar{g}_{\mu\alpha}\right). \tag{11.8.20}$$

Problem 11.49. Relation between 'background' and 'full' Christoffel Show that equations (11.8.18) and (11.8.19) can be used to deduce that the full Christoffel symbol

$$\Gamma^{\alpha}{}_{\mu\nu}[g] = \frac{1}{2} g^{\alpha\sigma} \left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \right) \tag{11.8.21}$$

can be related to that of its background counterpart through the relation

$$\Gamma^{\alpha}{}_{\mu\nu}[g] = \overline{\Gamma}^{\alpha}{}_{\mu\nu}[\bar{g}] + \delta\Gamma^{\alpha}{}_{\mu\nu}. \tag{11.8.22}$$

Here,

$$\delta\Gamma^{\alpha}{}_{\mu\nu} \equiv \frac{1}{2} g^{\alpha\sigma} H_{\sigma\mu\nu}, \tag{11.8.23}$$

$$H_{\sigma\mu\nu} \equiv \overline{\nabla}_\mu h_{\nu\sigma} + \overline{\nabla}_\nu h_{\mu\sigma} - \overline{\nabla}_\sigma h_{\mu\nu}. \tag{11.8.24}$$

Notice the difference between the 'full' and 'background' Christoffel symbols, namely $\Gamma^{\mu}{}_{\alpha\beta} - \overline{\Gamma}^{\mu}{}_{\alpha\beta}$, is a tensor. □

Problem 11.50. Geometric tensors With the result in eq. (11.8.22), show that for an arbitrary 1-form V_β,

$$\nabla_\alpha V_\beta = \overline{\nabla}_\alpha V_\beta - \delta\Gamma^{\sigma}{}_{\alpha\beta} V_\sigma. \tag{11.8.25}$$

Use this to compute $[\nabla_\alpha, \nabla_\beta] V_\lambda$ and proceed to show that the exact Riemann tensor is

$$R^{\alpha}{}_{\beta\mu\nu}[g] = \bar{R}^{\alpha}{}_{\beta\mu\nu}[\bar{g}] + \delta R^{\alpha}{}_{\beta\mu\nu}, \tag{11.8.26}$$

$$\delta R^{\alpha}{}_{\beta\mu\nu} \equiv \overline{\nabla}_{[\mu} \delta\Gamma^{\alpha}{}_{\nu]\beta} + \delta\Gamma^{\alpha}{}_{\sigma[\mu} \delta\Gamma^{\sigma}{}_{\nu]\beta} \tag{11.8.27}$$

$$= \frac{1}{2} \overline{\nabla}_\mu \left(g^{\alpha\lambda} H_{\lambda\nu\beta} \right) - \frac{1}{2} \overline{\nabla}_\nu \left(g^{\alpha\lambda} H_{\lambda\mu\beta} \right) + \frac{1}{4} g^{\alpha\lambda} g^{\sigma\rho} \left(H_{\lambda\mu\sigma} H_{\rho\beta\nu} - H_{\lambda\nu\sigma} H_{\rho\beta\mu} \right), \tag{11.8.28}$$

where $\bar{R}^{\alpha}{}_{\beta\mu\nu}[\bar{g}]$ is the Riemann tensor built entirely out of the background metric $\bar{g}_{\alpha\lambda}$. □

From eq. (11.8.26), the Ricci tensor and scalars can be written down:

$$R_{\mu\nu}[g] = R^{\sigma}{}_{\mu\sigma\nu} \qquad \text{and} \qquad \mathcal{R}[g] = g^{\mu\nu} R_{\mu\nu}. \tag{11.8.29}$$

From these formulas, perturbation theory can now be carried out. The primary reason why these geometric tensors admit an infinite series is because of the geometric series of the full inverse metric eq. (11.8.6). I find it helpful to remember, when one multiplies two infinite series which do not have negative powers of the expansion object $h_{\mu\nu}$, the terms that contain precisely n powers of $h_{\mu\nu}$ is a discrete convolution: for instance, such an nth order piece of the Ricci scalar is

$$\delta_n \mathcal{R} = \sum_{\ell=0}^{n} \delta_\ell g^{\mu\nu} \delta_{n-\ell} R_{\mu\nu}, \tag{11.8.30}$$

where $\delta_\ell g^{\mu\nu}$ is the piece of the full inverse metric containing exactly ℓ powers of $h_{\mu\nu}$ and $\delta_{n-\ell} R_{\mu\nu}$ is that containing precisely $n - \ell$ powers of the same.

Problem 11.51. Linearized geometric tensors The Riemann tensor that contains up to one power of $h_{\mu\nu}$ can be obtained readily from eq. (11.8.26). The H^2 terms begin at order h^2, so we may drop them; and since H is already linear in h, the g^{-1} contracted into it can be set to the background metric.

$$R^{\alpha}{}_{\beta\mu\nu}[g] = \bar{R}^{\alpha}{}_{\beta\mu\nu}[\bar{g}] + \frac{1}{2}\overline{\nabla}_{[\mu}\left(\overline{\nabla}_{\nu]}h_{\beta}{}^{\alpha} + \overline{\nabla}_{|\beta|}h_{\nu]}^{\alpha} - \overline{\nabla}^{\alpha}h_{\nu]\beta}\right) + \mathcal{O}(h^2) \tag{11.8.31}$$
$$= \bar{R}^{\alpha}{}_{\beta\mu\nu}[\bar{g}] + \frac{1}{2}\left([\overline{\nabla}_{\mu}, \overline{\nabla}_{\nu}]h_{\beta}{}^{\alpha} + \overline{\nabla}_{\mu}\overline{\nabla}_{\beta}h_{\nu}{}^{\alpha} - \overline{\nabla}_{\nu}\overline{\nabla}_{\beta}h_{\mu}{}^{\alpha} - \overline{\nabla}_{\mu}\overline{\nabla}^{\alpha}h_{\nu\beta} + \overline{\nabla}_{\nu}\overline{\nabla}^{\alpha}h_{\mu\beta}\right) + \mathcal{O}(h^2).$$

(The $|\beta|$ on the first line indicates the β is not to be antisymmetrized.) Starting from the linearized Riemann tensor in eq. (11.8.31), let us work out the linearized Ricci tensor, Ricci scalar, and Einstein tensor.

Specifically, show that one contraction of eq. (11.8.31) yields the linearized Ricci tensor:

$$R_{\beta\nu} = \overline{R}_{\beta\nu} + \delta_1 R_{\beta\nu} + \mathcal{O}(h^2), \tag{11.8.32}$$
$$\delta_1 R_{\beta\nu} \equiv \frac{1}{2}\left(\overline{\nabla}^{\mu}\overline{\nabla}_{\{\beta}h_{\nu\}\mu} - \overline{\nabla}_{\nu}\overline{\nabla}_{\beta}h - \overline{\nabla}^{\mu}\overline{\nabla}_{\mu}h_{\beta\nu}\right). \tag{11.8.33}$$

Contracting this Ricci tensor result with the full inverse metric, verify that the linearized Ricci scalar is

$$\mathcal{R} = \overline{\mathcal{R}} + \delta_1\mathcal{R} + \mathcal{O}(h^2), \tag{11.8.34}$$
$$\delta_1\mathcal{R} \equiv -h^{\beta\nu}\bar{R}_{\beta\nu} + \left(\overline{\nabla}^{\mu}\overline{\nabla}^{\nu} - \bar{g}^{\mu\nu}\overline{\nabla}^{\sigma}\overline{\nabla}_{\sigma}\right)h_{\mu\nu}. \tag{11.8.35}$$

Now, let us define the variable $\bar{h}_{\mu\nu}$ through the relation

$$h_{\mu\nu} \equiv \bar{h}_{\mu\nu} - \frac{\bar{g}_{\mu\nu}}{d-2}\bar{h}, \qquad \bar{h} \equiv \bar{h}^{\sigma}{}_{\sigma}. \tag{11.8.36}$$

First explain why this is equivalent to

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{\bar{g}_{\mu\nu}}{2}h. \tag{11.8.37}$$

(Hint: First calculate the trace of $\bar{h}$ in terms of h.) In (3+1)D this $\bar{h}_{\mu\nu}$ is often dubbed the "trace-reversed" perturbation – can you see why? Then show that the linearized Einstein tensor is

$$G_{\mu\nu} = \bar{G}_{\mu\nu}[\bar{g}] + \delta_1 G_{\mu\nu} + \mathcal{O}(\bar{h}^2), \tag{11.8.38}$$

where

$$\delta_1 G_{\mu\nu} \equiv -\frac{1}{2}\left(\overline{\Box}\bar{h}_{\mu\nu} + \bar{g}_{\mu\nu}\overline{\nabla}_{\sigma}\overline{\nabla}_{\rho}\bar{h}^{\sigma\rho} - \overline{\nabla}_{\{\mu}\overline{\nabla}^{\sigma}\bar{h}_{\nu\}\sigma}\right)$$
$$+ \frac{1}{2}\left(\bar{g}_{\mu\nu}\bar{h}^{\rho\sigma}\bar{R}_{\rho\sigma} + \bar{h}_{\{\mu}{}^{\sigma}\bar{R}_{\nu\}\sigma} - \bar{h}_{\mu\nu}\bar{\mathcal{R}} - 2\bar{h}^{\rho\sigma}\bar{R}_{\mu\rho\nu\sigma}\right). \tag{11.8.39}$$

Cosmology, Kerr/Schwarzschild black holes, and Minkowski spacetimes are three physically important geometries. This result may be used to study linear perturbations about them. □

Second order Ricci For later purposes, we collect the second order Ricci tensor – see, for e.g., equation 35.58b of [25]:[131]

$$\delta_2 R_{\mu\nu} = \frac{1}{2}\Bigg\{\frac{1}{2}\overline{\nabla}_\mu h_{\alpha\beta}\overline{\nabla}_\nu h^{\alpha\beta} + h^{\alpha\beta}\left(\overline{\nabla}_\nu\overline{\nabla}_\mu h_{\alpha\beta} + \overline{\nabla}_\beta\overline{\nabla}_\alpha h_{\mu\nu} - \overline{\nabla}_\beta\overline{\nabla}_\nu h_{\mu\alpha} - \overline{\nabla}_\beta\overline{\nabla}_\mu h_{\nu\alpha}\right) \tag{11.8.40}$$
$$+ \overline{\nabla}^\beta h^\alpha{}_\nu\left(\overline{\nabla}_\beta h_{\mu\alpha} - \overline{\nabla}_\alpha h_{\mu\beta}\right) - \overline{\nabla}_\beta\left(h^{\alpha\beta} - \frac{1}{2}\bar{g}^{\alpha\beta}h\right)\left(\overline{\nabla}_{\{\nu}h_{\mu\}\alpha} - \overline{\nabla}_\alpha h_{\mu\nu}\right)\Bigg\}.$$

Gauge transformations: Infinitesimal Coordinate Transformations In the above discussion, we regarded the 'background metric' as a tensor. As a consequence, the metric perturbation $h_{\mu\nu}$ was also a tensor. However, since it is the full metric that enters any generally covariant calculation, it really is the combination $\bar{g}_{\mu\nu} + h_{\mu\nu}$ that transforms as a tensor. As we will now explore, when the coordinate transformation

$$x^\mu = x'^\mu + \xi^\mu(x') \tag{11.8.41}$$

is infinitesimal, in that ξ^μ is small in the same sense that $h_{\mu\nu}$ is small, we may instead attribute all the ensuing coordinate transformations to a transformation of $h_{\mu\nu}$ alone. This will allow us to view 'small' coordinate transformations as gauge transformations, and will also be important for the discussion of the linearized Einstein's equations.

In what follows, we shall view the x and x' in eq. (11.8.41) as referring to the same spacetime point, but expressed within infinitesimally different coordinate systems. Now, transforming from x to x',

$$\begin{aligned} \mathrm{d}s^2 &= g_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu = \left(\bar{g}_{\mu\nu}(x) + h_{\mu\nu}(x)\right)\mathrm{d}x^\mu\mathrm{d}x^\nu \\ &= \left(\bar{g}_{\mu\nu}(x'+\xi) + h_{\mu\nu}(x'+\xi)\right)\left(\mathrm{d}x'^\mu + \partial_{\alpha'}\xi^\mu\mathrm{d}x'^\alpha\right)\left(\mathrm{d}x'^\nu + \partial_{\beta'}\xi^\nu\mathrm{d}x'^\beta\right) \\ &= \left(\bar{g}_{\mu\nu}(x') + \xi^\sigma\partial_{\sigma'}\bar{g}_{\mu\nu}(x') + h_{\mu\nu}(x') + \mathcal{O}\left(\xi^2, \xi\partial h\right)\right)\left(\mathrm{d}x'^\mu + \partial_{\alpha'}\xi^\mu\mathrm{d}x'^\alpha\right)\left(\mathrm{d}x'^\nu + \partial_{\beta'}\xi^\nu\mathrm{d}x'^\beta\right) \\ &= \left(\bar{g}_{\mu\nu}(x') + \xi^\sigma(x')\partial_{\sigma'}\bar{g}_{\mu\nu}(x') + \bar{g}_{\sigma\{\mu}(x')\partial_{\nu'\}}\xi^\sigma(x') + h_{\mu\nu}(x') + \mathcal{O}\left(\xi^2, \xi\partial h\right)\right)\mathrm{d}x'^\mu\mathrm{d}x'^\nu \\ &\equiv \left(\bar{g}_{\mu'\nu'}(x') + h_{\mu'\nu'}(x')\right)\mathrm{d}x'^\mu\mathrm{d}x'^\nu. \end{aligned} \tag{11.8.42}$$

This teaches us that, the infinitesimal coordinate transformation of eq. (11.8.41) amounts to keeping the background metric fixed,

$$\bar{g}_{\mu\nu}(x) \to \bar{g}_{\mu\nu}(x), \tag{11.8.43}$$

but shifting

$$h_{\mu\nu}(x) \to h_{\mu\nu}(x) + \xi^\sigma(x)\partial_\sigma\bar{g}_{\mu\nu}(x) + \bar{g}_{\sigma\{\mu}(x)\partial_{\nu\}}\xi^\sigma(x), \tag{11.8.44}$$

followed by replacing

$$x^\mu \to x'^\mu \qquad \text{and} \qquad \partial_\mu \equiv \frac{\partial}{\partial x^\mu} \to \frac{\partial}{\partial x'^\mu} \equiv \partial_{\mu'}. \tag{11.8.45}$$

[131] I have checked that eq. (11.8.40) is consistent with the output from xAct [47].

However, since x and x' refer to the same point in spacetime,[132] it is customary within the contemporary physics literature to drop the primes and simply phrase the coordinate transformation as replacement rules:

$$x^\mu \to x^\mu + \xi^\mu(x), \tag{11.8.46}$$

$$\bar{g}_{\mu\nu}(x) \to \bar{g}_{\mu\nu}(x), \tag{11.8.47}$$

$$h_{\mu\nu}(x) \to h_{\mu\nu}(x) + \overline{\nabla}_{\{\mu}\xi_{\nu\}}(x); \tag{11.8.48}$$

where we have recognized

$$\xi^\sigma \partial_\sigma \bar{g}_{\mu\nu} + \bar{g}_{\sigma\{\mu}\partial_{\nu\}}\xi^\sigma = \overline{\nabla}_{\{\mu}\xi_{\nu\}} \equiv (\pounds_\xi \bar{g})_{\mu\nu}(x). \tag{11.8.49}$$

Problem 11.52. Lie derivative of a tensor If x and x' are infinitesimally nearby coordinate systems related via eq. (11.8.41), show that $T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M}(x)$ (the components of a given tensor in the x^μ coordinate basis) and $T^{\mu'_1\dots\mu'_N}{}_{\nu'_1\dots\nu'_M}(x')$ (the components of the same tensor but in the x'^μ coordinate basis) are in turn related via

$$T^{\mu'_1\dots\mu'_N}{}_{\nu'_1\dots\nu'_M}(x') = T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M}(x \to x') + (\pounds_\xi T)^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M}(x \to x'); \tag{11.8.50}$$

where the Lie derivative of the tensor reads

$$\begin{aligned}(\pounds_\xi T)^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M} &= \xi^\sigma \partial_\sigma T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M} \\ &- T^{\sigma\mu_2\dots\mu_N}{}_{\nu_1\dots\nu_M}\partial_\sigma \xi^{\mu_1} - \dots - T^{\mu_1\dots\mu_{N-1}\sigma}{}_{\nu_1\dots\nu_M}\partial_\sigma \xi^{\mu_N} \\ &+ T^{\mu_1\dots\mu_N}{}_{\sigma\nu_2\dots\nu_M}\partial_{\nu_1}\xi^\sigma + \dots + T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_{M-1}\sigma}\partial_{\nu_M}\xi^\sigma.\end{aligned} \tag{11.8.51}$$

The $x \to x'$ on the right hand side of eq. (11.8.50) means, the tensor $T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M}$ and its Lie derivative are to be computed in the x^μ-coodinate basis – but x^μ is to be replaced with x'^μ afterwards.

Explain why the partial derivatives on the right hand side of eq. (11.8.51) may be replaced with covariant ones, namely

$$\begin{aligned}(\pounds_\xi T)^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M} &= \xi^\sigma \nabla_\sigma T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_M} \\ &- T^{\sigma\mu_2\dots\mu_N}{}_{\nu_1\dots\nu_M}\nabla_\sigma \xi^{\mu_1} - \dots - T^{\mu_1\dots\mu_{N-1}\sigma}{}_{\nu_1\dots\nu_M}\nabla_\sigma \xi^{\mu_N} \\ &+ T^{\mu_1\dots\mu_N}{}_{\sigma\nu_2\dots\nu_M}\nabla_{\nu_1}\xi^\sigma + \dots + T^{\mu_1\dots\mu_N}{}_{\nu_1\dots\nu_{M-1}\sigma}\nabla_{\nu_M}\xi^\sigma.\end{aligned} \tag{11.8.52}$$

(Hint: First explain why $\partial_\alpha \xi^\beta = \nabla_\alpha \xi^\beta - \Gamma^\beta{}_{\alpha\sigma}\xi^\sigma$.) That the Lie derivative of a tensor can be expressed in terms of covariant derivatives indicates the former is a tensor.

We defined the Lie derivative of the metric $\bar{g}_{\mu\nu}$ with respect to ξ^α in eq. (11.8.49). Is it consistent with equations (11.8.51) and (11.8.52)? □

[132]We had, earlier, encountered very similar mathematical manipulations while considering the geometric symmetries that left the metric in the same form upon an active coordinate transformation – an actual displacement from one point to another infinitesimally close by. Here, we are doing a passive coordinate transformation, where x and x' describe the same point in spacetime, but using infinitesimally different coordinate systems.

Problem 11.53. Lie derivative along Killing vector commutes with ∇ If ξ^μ is a Killing vector, show that the Lie derivative along it commutes with the covariant derivative – i.e., show that

$$[\nabla_\alpha, \pounds_\xi] T^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M} = 0. \tag{11.8.53}$$

Hint: You may need to use the Killing equation $\nabla_\alpha \xi_\beta = -\nabla_\beta \xi_\alpha$ and eq. (11.5.8). □

Lie Derivative of Vector Note that the Lie derivative of some vector field U^μ with respect to ξ^μ is, according to eq. (11.8.52),

$$\pounds_\xi U^\mu = \xi^\sigma \nabla_\sigma U^\mu - U^\sigma \nabla_\sigma \xi^\mu \tag{11.8.54}$$
$$= \xi^\sigma \partial_\sigma U^\mu - U^\sigma \partial_\sigma \xi^\mu = [\xi, U]^\mu. \tag{11.8.55}$$

We have already encountered the Lie bracket/commutator of vector fields, in eq. (11.1.27). There, we learned that $[\xi, U] = 0$ iff ξ and U may be integrated to form a 2D coordinate system (at least locally). On the other hand, we may view the Lie derivative with respect to ξ as an active coordinate transformation induced by the displacement $x \to x + \xi$. This in fact provides insight into the above mentioned theorem: if $\pounds_\xi U^\mu = 0$ that means U remains unaltered upon a coordinate transformation induced along the direction of ξ; that in turn indicates, it is possible to move along the integral curve of ξ, bringing us from one integral curve of U to the next – while consistently maintaining the same coordinate value along the latter. Similarly, since $[\xi, U] = -[U, \xi] = -\pounds_U \xi = 0$, the vanishing of the Lie bracket also informs us the coordinate value along the integral curve of ξ may be consistently held fixed while moving along the integral curve of U, since the former is invariant under the flow along U. Altogether, this is what makes a set good 2D coordinates; we may vary one while keeping the other fixed, and vice versa.

Problem 11.54. Gauge transformations of a tensor Consider perturbing a spacetime tensor

$$T^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M} \equiv \overline{T}^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M} + \delta T^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M}, \tag{11.8.56}$$

where $\delta T^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M}$ is small in the same sense that ξ^α and $h_{\mu\nu}$ are small. Perform the infinitesimal coordinate transformation in eq. (11.8.41) on the tensor in eq. (11.8.56) and attribute all the transformations to the $\delta T^{\mu_1 \dots \mu_N}{}_{\nu_1 \dots \nu_M}$. Write down the ensuing gauge transformation, in direct analogy to eq. (11.8.48). Then justify the statement:

> "If the background tensor is zero, the perturbed tensor is gauge-invariant at first order in infinitesimal coordinate transformations."

Hint: You may work this out from scratch, or you may employ the results from Problem (11.52). □

11.8.1 Perturbed Flat Spacetimes

In this subsection we shall study perturbations about flat spacetimes

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1. \tag{11.8.57}$$

In 4D, this is the context where gravitational waves are usually studied.

Under a Poincaré transformation in eq. (10.1.12), $x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu$, where $\Lambda^\mu{}_\nu$ satisfies (10.1.13), observe that the metric transforms as

$$g_{\alpha'\beta'}(x') = g_{\mu\nu}(x = \Lambda x' + a)\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \tag{11.8.58}$$

$$= (\eta_{\mu\nu} + h_{\mu\nu}(x = \Lambda x' + a))\, \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \equiv \eta_{\alpha\beta} + h_{\alpha'\beta'}(x'). \tag{11.8.59}$$

Hence, as far as Poincaré transformations are concerned, we may attribute all the transformations to those of the perturbations. In other words, $h_{\mu\nu}$ is a tensor under Poincaré transformations:

$$h_{\alpha'\beta'}(x') = h_{\mu\nu}(x(x'))\Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta, \tag{11.8.60}$$

$$x^\mu = \Lambda^\mu{}_\nu x'^\nu + a^\mu. \tag{11.8.61}$$

Since the Riemann tensor is zero when $h_{\mu\nu} = 0$, that means the linearized counterpart $\delta_1 R_{\mu\nu\alpha\beta}$ must be gauge-invariant. More specifically, what we have shown thus far is, under the infinitesimal coordinate transformation

$$x^\mu = x'^\mu + \xi^\mu(x'), \tag{11.8.62}$$

the linearized Riemann tensor written in the x versus x' systems are related as

$$\delta_1 R_{\mu\nu\alpha\beta}(x) = \delta_1 R_{\mu'\nu'\alpha'\beta'}(x') + \mathcal{O}(h^2, \xi \cdot h, \xi^2). \tag{11.8.63}$$

Here, the components $\delta_1 R_{\mu\nu\alpha\beta}$ are written in the x coordinate basis whereas $\delta_1 R_{\mu'\nu'\alpha'\beta'}$ are in the x' basis. But, since x and x' differ by an infinitesimal quantity ξ, we may in fact replace $x' \to x$ on the right hand side:

$$\delta_1 R_{\mu\nu\alpha\beta}(x) = \delta_1 R_{\mu'\nu'\alpha'\beta'}(x' \to x) + \mathcal{O}(h^2, \xi \cdot h, \xi^2). \tag{11.8.64}$$

To solve for the $h_{\mu\nu}$ in eq. (11.8.57), one typically has to choose a specific coordinate system. However, eq. (11.8.64) tells us, the tidal forces encoded within the linearized Riemann tensor yield the same expression *for all infinitesimally nearby coordinate systems*.

Two Common Gauges Two commonly used gauges are the synchronous and de Donder gauges. The former refers to the choice of coordinate system such that all perturbations are spatial:

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + h^{(s)}_{ij}\mathrm{d}x^i \mathrm{d}x^j \qquad \text{(Synchronous gauge)}. \tag{11.8.65}$$

The latter is defined by the Lorentz-covariant constraint

$$\partial^\mu h_{\mu\nu} = \frac{1}{2}\partial_\nu h, \qquad h \equiv \eta^{\alpha\beta}h_{\alpha\beta}, \qquad \text{(de Donder gauge)}. \tag{11.8.66}$$

The de Donder gauge is particularly useful for obtaining explicit perturbative solutions to Einstein's equations. Whereas, the synchronous gauge is useful for describing proper distances between co-moving free-falling test masses.

One may prove that both gauges always exist. According to eq. (11.8.48), the perturbation in a Minkowski background transforms as

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu, \tag{11.8.67}$$

$$h \to h + 2\partial_\sigma \xi^\sigma. \tag{11.8.68}$$

Hence, if h_{00} were not zero, we may render it so by choosing $\xi_0 = -(1/2)\int^t h_{00}\mathrm{d}t$; since

$$\begin{aligned} h_{00} &\to h_{00} + 2\partial_0 \xi_0 && (11.8.69) \\ &= h_{00} + 2\partial_0 \frac{-1}{2}\int^t h_{00}\mathrm{d}t = 0. && (11.8.70) \end{aligned}$$

Moreover, if h_{0i} were not zero, an infinitesimal coordinate transformation would yield

$$\begin{aligned} h_{0i} &\to h_{0i} + \partial_i \xi_0 + \partial_0 \xi_i && (11.8.71) \\ &= h_{0i} - \frac{1}{2}\int^t \partial_i h_{00}\mathrm{d}t + \partial_0 \xi_i. && (11.8.72) \end{aligned}$$

The right hand side is zero if we choose

$$\xi_i = -\int^t \left(h_{0i} - \frac{1}{2}\int^{t'} \partial_i h_{00}\mathrm{d}t'' \right)\mathrm{d}t'. \tag{11.8.73}$$

That is, by choosing ξ_μ appropriately, $h_{0\mu} = h_{\mu 0}$ may always be set to zero.

As for the de Donder gauge condition in eq. (11.8.66), we first re-write it using eq. (11.8.37)

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h. \tag{11.8.74}$$

Namely,

$$\partial^\mu \bar{h}_{\mu\nu} = \partial^\mu \left(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \right) = 0. \tag{11.8.75}$$

Utilizing eq. (11.8.67), we may deduce the gauge-transformation of $\bar{h}_{\mu\nu}$ is

$$\bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu}\partial \cdot \xi, \qquad \partial \cdot \xi \equiv \partial^\sigma \xi_\sigma. \tag{11.8.76}$$

Now, if eq. (11.8.75) were not obeyed, a gauge transformation would produce

$$\begin{aligned} \partial^\mu \bar{h}_{\mu\nu} &\to \partial^\mu \bar{h}_{\mu\nu} + \partial^\mu \left(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu \right) - \eta_{\mu\nu}\partial^\mu \partial \cdot \xi && (11.8.77) \\ &= \partial^\mu \bar{h}_{\mu\nu} + \partial^2 \xi_\nu. && (11.8.78) \end{aligned}$$

Therefore, by choosing ξ_ν to be the solution to $\partial^2 \xi_\nu = -\partial^\mu \bar{h}_{\mu\nu}$, we may always switch over to the de Donder gauge of eq. (11.8.75). We also note, suppose $\bar{h}_{\mu\nu}$ already obeys the de Donder gauge condition; then notice the transformed $\bar{h}_{\mu\nu}$ actually remains within the de Donder gauge whenever $\partial^2 \xi_\nu = 0$.

Problem 11.55. Are the synchronous and de Donder gauges "infinitesimally nearby" coordinate systems? □

Problem 11.56. Co-moving geodesics in synchronous gauge Prove that

$$Z^\mu(t) = (t, \vec{Z}_0), \tag{11.8.79}$$

where $\vec{Z}_0$ is time-independent, satisfies the geodesic equation in the spacetime

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \mathrm{d}t^2 + g_{ij}(t,\vec{x})\mathrm{d}x^i \mathrm{d}x^j. \tag{11.8.80}$$

This result translates to the following interpretation: each $\vec{x}$ in eq. (11.8.80) may be viewed as the location of a test mass free-falling in the given spacetime. This co-moving test mass remains still, for all time t, in such a synchronous gauge system. Of course, eq. (11.8.65) is a special case of eq. (11.8.80). □

Linearized Synge's World Function In the weak field metric of eq. (11.8.57), according to eq. (11.3.58), half the square of the geodesic distance between x and x' is

$$\bar{\sigma}(x,x') = \frac{1}{2}\int_0^1 \mathrm{d}\lambda \left(\eta_{\mu\nu} + h_{\mu\nu}(Z)\right) \frac{\mathrm{d}Z^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}Z^\nu}{\mathrm{d}\lambda}; \tag{11.8.81}$$

where the trajectories obey geodesic equation (11.3.60)

$$\frac{\mathrm{d}^2 Z^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}Z^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}Z^\beta}{\mathrm{d}\lambda} = 0 \tag{11.8.82}$$

subject to the boundary conditions

$$Z^\mu(\lambda=0) = x'^\mu \qquad \text{and} \qquad Z^\mu(\lambda=1) = x^\mu. \tag{11.8.83}$$

If the perturbations were not present, $h_{\mu\nu}=0$, the geodesic equation is

$$\frac{\mathrm{d}^2 \bar{Z}^\mu}{\mathrm{d}\lambda^2} = 0; \tag{11.8.84}$$

whose solution, in turn, is

$$\bar{Z}^\mu(\lambda) = x'^\mu + \lambda(x-x')^\mu, \tag{11.8.85}$$

$$\dot{\bar{Z}}^\mu(\lambda) = (x-x')^\mu. \tag{11.8.86}$$

When the perturbations are non-trivial, $h_{\mu\nu} \neq 0$, the full solution $Z^\mu = \bar{Z}^\mu + \delta Z^\mu$ should deviate from the zeroth order solution $\bar{Z}^\mu$ at linear order in the perturbations: $\delta Z^\mu \sim \mathcal{O}(h_{\mu\nu})$. One may see this from eq. (11.3.74). Hence, if we insert $Z^\mu = \bar{Z}^\mu + \delta Z^\mu$ into Synge's world function in eq. (11.8.81),

$$\begin{aligned}\sigma(x,x') = {} & \frac{1}{2}\int_0^1 \mathrm{d}\lambda \left(\eta_{\mu\nu} + h_{\mu\nu}(\bar{Z})\right)(x-x')^\mu (x-x')^\nu \\ & - \int_0^1 \delta Z^\mu(\lambda)\left(\eta_{\mu\nu} + h_{\mu\nu}(\bar{Z})\right)\frac{D^2 \bar{Z}^\nu}{\mathrm{d}\lambda^2}\mathrm{d}\lambda + \mathcal{O}\left((\delta Z)^2\right);\end{aligned} \tag{11.8.87}$$

because the zeroth order geodesic equation is satisfied, namely $\mathrm{d}^2\bar{Z}/\mathrm{d}\lambda^2 = 0$, $D^2\bar{Z}^\mu/\mathrm{d}\lambda^2 = \Gamma^\mu{}_{\alpha\beta}\dot{\bar{Z}}^\alpha\dot{\bar{Z}}^\beta \sim \mathcal{O}(h_{\mu\nu})$ and therefore the second line scales as $\mathcal{O}(h_{\mu\nu}^2)$ and higher. At linear order in perturbation theory, half the square of the geodesic distance between $Z(\lambda = 0) = x'$ and $Z(\lambda = 1) = x$ is therefore Synge's world function evaluated on the zeroth order geodesic solution – namely, the straight line in eq. (11.8.85).[133]

$$\sigma(x,x') = \frac{1}{2}(x-x')^2 + \frac{1}{2}(x-x')^\mu(x-x')^\nu \int_0^1 h_{\mu\nu}\left(\bar{Z}(\lambda)\right)\mathrm{d}\lambda + \mathcal{O}(h^2) \tag{11.8.88}$$

Proper Distance Between Free-Falling Masses: Synchronous Gauge Consider a pair of free-falling test masses at $(t,\vec{y})$ and $(t',\vec{y}')$. Within the synchronous gauge of eq. (11.8.65), where $h_{\mu 0} = h_{0\mu} = 0$, the square of their geodesic spatial separation at a fixed time $t = t'$ is gotten from eq. (11.8.88) through

$$\ell^2 = -2\sigma(t=t';\vec{y},\vec{y}') \tag{11.8.89}$$

$$= (\vec{y}-\vec{y}')^2 - (y-y')^i(y-y')^j \int_0^1 h_{ij}^{(s)}\left(t,\vec{y}'+\lambda(\vec{y}-\vec{y}')\right)\mathrm{d}\lambda + \mathcal{O}(h^2) \tag{11.8.90}$$

Taking the square root on both sides, and using the Taylor expansion result $(1+z)^{1/2} = 1 + z/2 + \mathcal{O}(z^2)$, we surmise that the synchronous gauge form of the metric in eq. (11.8.65) indeed allows us to readily calculate the proper spatial geodesic distance between pairs of free-falling test masses.

$$\ell(t;\vec{y}\leftrightarrow\vec{y}') = |\vec{y}-\vec{y}'|\left(1 - \frac{1}{2}\widehat{R}^i\widehat{R}^j\int_0^1 h_{ij}^{(s)}\left(t,\bar{Z}(\lambda)\right)\mathrm{d}\lambda + \mathcal{O}(h^2)\right), \tag{11.8.91}$$

$$\widehat{R} \equiv \frac{\vec{y}-\vec{y}'}{|\vec{y}-\vec{y}'|}. \tag{11.8.92}$$

(Remember $\bar{Z}$ in eq. (11.8.85).)

Gravitational Wave Polarization & Oscillation Patterns We may re-phrase eq. (11.8.91) as a fractional distortion of space $\delta\ell/\delta_0$ away from the flat space value of $\ell_0 \equiv |\vec{y}-\vec{y}'|$, due to the presence of the perturbation $h_{ij}^{(s)}$,

$$\left(\frac{\delta\ell}{\ell_0}\right)(t;\vec{y}\leftrightarrow\vec{y}') = -\frac{1}{2}\widehat{R}^i\widehat{R}^j\int_0^1 h_{ij}^{(s)}\left(t,\bar{Z}(\lambda)\right)\mathrm{d}\lambda + \mathcal{O}(h^2). \tag{11.8.93}$$

If we define gravitational waves to be simply the finite frequency portion of the tidal signal in eq. (11.8.93), then we see that the fractional distortion of space due to a passing gravitational wave could consist of up to a maximum of $D(D+1)/2$ distinct oscillatory patterns, in a $D+1$ dimensional weakly curved spacetime. In detail, if we decompose

$$h_{ij}^{(s)}\left(t,\bar{Z}(\lambda)\right) = \int_{\mathbb{R}} \widetilde{h}_{ij}^{(s)}\left(\omega,\bar{Z}(\lambda)\right)e^{-i\omega t}\frac{\mathrm{d}\omega}{2\pi}, \tag{11.8.94}$$

[133] This sort of "first-order-variation-vanishes" argument occurs frequently in field theory as well.

then eq. (11.8.93) reads

$$\widetilde{\left(\frac{\delta\ell}{\ell_0}\right)}(\omega; \vec{y} \leftrightarrow \vec{y}') = -\frac{1}{2}\widehat{R}^i \widehat{R}^j \int_0^1 \widetilde{h}_{ij}^{(s)}\left(\omega, \vec{y} + \lambda(\vec{y}' - \vec{y})\right)\mathrm{d}\lambda + \mathcal{O}(h^2). \tag{11.8.95}$$

Now, a direct calculation will reveal

$$\delta_1 R_{0i0j}(t, \vec{x}) = -\frac{1}{2}\partial_0^2 h_{ij}^{(s)}(t, \vec{x}), \qquad \text{(Synchronous gauge)}. \tag{11.8.96}$$

To translate this statement to frequency space, we replace $\partial_0 = \partial_t \to -i\omega$,

$$\delta_1 \widetilde{R}_{0i0j}(\omega, \vec{x}) = \frac{\omega^2}{2}\widetilde{h}_{ij}^{(s)}(\omega, \vec{x}), \qquad \text{(Synchronous gauge)}. \tag{11.8.97}$$

Gravitational waves are associated with time dependent radiative processes, capable of performing dissipative work through their oscillatory tidal forces. To this end, eq. (11.8.97) teaches us it is the finite frequency modes – i.e., the $\omega \neq 0$ portion – of the linearized Riemann tensor that is to be associated with such gravitational radiation. By inserting eq. (11.8.97) into eq. (11.8.95), we see that the finite frequency gravitational-wave-driven fractional distortion of space – namely,

$$\widetilde{\left(\frac{\delta\ell}{\ell_0}\right)}(\omega \neq 0; \vec{y} \leftrightarrow \vec{y}') = \frac{\widehat{R}^i \widehat{R}^j}{\omega^2} \int_0^1 \delta_1 \widetilde{R}_{0i0j}\left(\omega, \vec{y} + \lambda(\vec{y}' - \vec{y})\right)\mathrm{d}\lambda + \mathcal{O}(h^2) \tag{11.8.98}$$

– is not only gauge-invariant (since the linearized Riemann components are); it has $(D^2 - D)/2 + D = D(D+1)/2$ algebraically independent components, since $\delta_1 \widetilde{R}_{0i0j}$ is a symmetric rank-2 spatial tensor in the ij indices.

Problem 11.57. Verify eq. (11.8.96). □

Problem 11.58. 4D Gravitational Wave Polarizations In 3+1 dimensional spacetime, choose the unit vector along the $3-$axis $\widehat{e}_3$ to be the direction of propagation of the finite frequency $\widetilde{h}_{ij}^{(s)}$ in eq. (11.8.95). Then proceed to build upon Problems (5.98) and (5.100) to decompose the fractional distortion of space in eq. (11.8.95) into its irreducible constituents – i.e., the spin-0, spin-1 and spin-2 finite-frequency waves. You should find that up to six different polarizations are allowed.

In 4D linearized de Donder gauge General Relativity, only null traveling waves are admitted in vacuum. As we will see in the next problem, this implies only the helicity-2 waves are predicted to exist. However, it is conceivable that alternate theories of gravity could allow for the other irreducible modes to carry gravitational radiation. □

Problem 11.59. Synchronous-de Donder Gauge & Null Traveling 'TT' Waves In this problem we shall see how the gauge-invariant linearized Riemann tensor may be used to relate the synchronous gauge metric perturbation to its de Donder counterpart – at least for source-free traveling waves.

Let us begin by performing a Fourier transform in spacetime,

$$h_{ij}^{(s)}(t, \vec{x}) = \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \widetilde{h}_{ij}^{(s)}(\omega, \vec{k}) e^{-i\omega t} e^{+i\vec{k}\cdot\vec{x}}; \tag{11.8.99}$$

so that $\partial_\mu \leftrightarrow -i(\omega, k_i)_\mu$. The associated synchronous gauge Riemann tensor components then read

$$\delta_1 \widetilde{R}_{0i0j}(\omega, \vec{k}) = +\frac{\omega^2}{2}\widetilde{h}^{(s)}_{ij}(\omega, \vec{k}), \qquad \text{(Synchronous gauge)}. \tag{11.8.100}$$

Up to this point, we have not assumed a dispersion relation between ω and $\vec{k}$. Suppose we impose the null condition

$$\omega^2 = \vec{k}^2 \tag{11.8.101}$$

on both the synchronous and de Donder gauge perturbations, so they are both superpositions of traveling waves propagating at unit speed –

$$h^{(s)}_{ij}(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{1}{2}\left\{\widetilde{h}^{(s)}_{ij}(k)e^{-i|\vec{k}|t} + \widetilde{h}^{(s)}_{ij}(k)^* e^{+i|\vec{k}|t}\right\} e^{i\vec{k}\cdot\vec{x}} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}, \qquad k^\mu \equiv (|\vec{k}|, \vec{k}) \tag{11.8.102}$$

– now, verify directly that the corresponding Riemann components are

$$\delta_1 \widetilde{R}_{0i0j}(\omega, \vec{k}) = \frac{\omega^2}{2}\left(\widetilde{h}_{ij} + \widehat{k}_{\{i}\widetilde{h}_{j\}l}\widehat{k}^l + \widehat{k}_i\widehat{k}_j\widetilde{h}_{mn}\widehat{k}^m\widehat{k}^n\right), \qquad \text{(de Donder)}; \tag{11.8.103}$$

$$\widehat{k}^i \equiv k^i/|\vec{k}|, \qquad \omega^2 = \vec{k}^2. \tag{11.8.104}$$

Next, verify $\delta_1 \widetilde{R}_{0i0j}$ in eq. (11.8.103) is transverse and traceless:

$$\delta^{ij}\delta_1 \widetilde{R}_{0i0j} = 0 = \widehat{k}^i \delta_1 \widetilde{R}_{0i0j}. \tag{11.8.105}$$

Finally, demonstrate that such a traveling-wave $\delta_1 \widetilde{R}_{0i0j}$ in de Donder gauge is simply the transverse-traceless (TT) portion of the metric perturbation itself:

$$\delta_1 \widetilde{R}_{0i0j}(\omega, \vec{k}) = \frac{\omega^2}{2}\widetilde{P}_{ijab}\widetilde{h}_{ab}(\omega, \vec{k}), \tag{11.8.106}$$

where the TT projector is

$$\widetilde{P}_{ijab} = \frac{1}{2}\widetilde{P}_{i\{a}\widetilde{P}_{b\}j} - \frac{1}{D-1}\widetilde{P}_{ij}\widetilde{P}_{ab}, \tag{11.8.107}$$

$$\widetilde{P}_{ij} = \delta_{ij} - \widehat{k}_i\widehat{k}_j. \tag{11.8.108}$$

It enjoys the following properties:

$$\widetilde{P}_{ijab} = \widetilde{P}_{abij}, \qquad \widetilde{P}_{jiab} = \widetilde{P}_{ijab}, \qquad \delta^{ij}\widetilde{P}_{ijab} = 0 = \widehat{k}^i\widetilde{P}_{ijab}. \tag{11.8.109}$$

Helicity−2 modes Finally, by choosing $\widehat{k} \equiv \widehat{e}_3$, the unit vector along the 3−axis, verify the claim in the previous problem, that the null traveling waves described by these linearized $\delta_1 \widetilde{R}_{0i0j}$ are purely helicity−2 modes only.

Hint: Throughout these calculations, you would need to repeatedly employ the de Donder gauge condition (eq. (11.8.66)) in Fourier spacetime: $k^\mu \widetilde{h}_{\mu\nu} = (1/2)k_\nu \widetilde{h}$, with $k^\mu \equiv (\omega, \vec{k})$. □

From our previous discussion, since the linearized Riemann tensor is gauge-invariant, we may immediately equate the $0i0j$ components in the synchronous (eq. (11.8.100)) and de Donder (eq. (11.8.103)) gauges to deduce: for finite frequencies $|\omega| = |\vec{k}| \neq 0$, the synchronous gauge metric perturbation is the TT part of the de Donder gauge one.

$$\widetilde{h}^{(s)}_{ij}[\text{Synchronous}] = \widetilde{P}_{ijab}\widetilde{h}_{ab}[\text{de Donder}] \tag{11.8.110}$$

That this holds only for finite frequencies – the formulas in equations (11.8.100) and (11.8.103) do not contain $\delta(\omega)$ or $\delta'(\omega)$ terms – because $\omega^2\delta(\omega) = 0 = \omega^2\delta'(\omega)$. More specifically, since eq. (11.8.96) involved a second time derivative on $h^{(s)}_{ij}$, by equating it to the (position-spacetime version of) eq. (11.8.103), we may solve the synchronous gauge metric perturbation only up to its initial value and time derivative:

$$\begin{aligned} h^{(s)}_{ij}(t,\vec{x}) = -2\int_{t_0}^{t}\int_{t_0}^{\tau_2} \delta_1 R_{0i0j}(\tau_1,\vec{x})\mathrm{d}\tau_1\mathrm{d}\tau_2 \\ + (t-t_0)\dot{h}^{(s)}_{ij}(t_0,\vec{x}) + h^{(s)}_{ij}(t_0,\vec{x}). \end{aligned} \tag{11.8.111}$$

Note that the initial velocity term $(t-t_0)\dot{h}^{(s)}_{ij}(t_0,\vec{x})$ is proportional to $\delta'(\omega)$ in frequency space; whereas the initial value $h^{(s)}_{ij}(t_0,\vec{x})$ is proportional to $\delta(\omega)$.

Unlike eq. (11.8.110), eq. (11.8.111) does not depend on specializing to traveling waves obeying the null dispersion relation $k^2 \equiv k_\mu k^\mu = 0$.[134] Moreover, eq. (11.8.111) suggests, up to the two initial condiitions, $h^{(s)}_{ij}$ itself is almost gauge-invariant – afterall it measures something geometrical, eq. (11.8.91), the proper distances between free-falling test masses – and we may attempt to further understand this through the following considerations. Since the synchronous gauge perturbation allows us to easily compute proper distances between co-moving test masses, let us ask how much coordinate freedom is available while still remaining with the synchronous gauge itself. For the 00 component to remain 0, we have from eq. (11.8.67)

$$0 = h^{(s)}_{00} \to 2\partial_0\xi_0 = 0. \tag{11.8.112}$$

That is, ξ_0 needs to be time-independent. For the $0i$ component to remain zero,

$$0 = h^{(s)}_{0i} \to \partial_0\xi_i + \partial_i\xi_0 = 0. \tag{11.8.113}$$

This allows us to assert

$$\xi_i(t,\vec{x}) = -(t-t_0)\partial_i\xi_0(\vec{x}) + \xi_i(t_0,\vec{x}). \tag{11.8.114}$$

Under such a coordinate transformation, $x \to x + \xi$,

$$\begin{aligned} h^{(s)}_{ij} &\to h^{(s)}_{ij} + \partial_i\xi_j + \partial_j\xi_i & (11.8.115)\\ &= h^{(s)}_{ij}(t,\vec{x}) - 2(t-t_0)\partial_i\partial_j\xi_0(\vec{x}) + \partial_{\{i}\xi_{j\}}(t_0,\vec{x}). & (11.8.116) \end{aligned}$$

Comparison with eq. (11.8.111) tells us $\partial_i\partial_j\xi_0$ may be identified with the freedom to redefine the initial velocity of $h^{(s)}_{ij}$; and $\partial_{\{i}\xi_{j\}}(t_0,\vec{x})$ its initial value.

[134] More specifically, eq. (11.8.110) holds whenever the linearized vacuum Einstein's equations hold; whereas eq. (11.8.111) is true regardless of the underlying dynamics of the metric perturbations.

11.9 ⋆Conformal/Weyl Transformations

In this section, we collect the conformal transformation properties of various geometric objects. We shall define a conformal transformation on a metric to be a change of the geometry by an overall spacetime dependent scale. That is,

$$g_{\mu\nu}(x) \equiv \Omega^2(x)\bar{g}_{\mu\nu}(x). \tag{11.9.1}$$

The inverse metric is

$$g^{\mu\nu}(x) = \Omega(x)^{-2}\bar{g}^{\mu\nu}(x), \qquad\qquad \bar{g}^{\mu\sigma}\bar{g}_{\sigma\nu} \equiv \delta^\mu_\nu. \tag{11.9.2}$$

We shall now enumerate how the geometric objects/operations built out of $g_{\mu\nu}$ is related to that built out of $\bar{g}_{\mu\nu}$. In what follows, all indices on barred tensors are raised and lowered with $\bar{g}^{\mu\nu}$ and $\bar{g}_{\mu\nu}$ while all indices on un-barred tensors are raised/lowered with $g^{\mu\nu}$ and $g_{\mu\nu}$; the covariant derivative ∇ is with respect to $g_{\mu\nu}$ while the $\overline{\nabla}$ is with respect to $\bar{g}_{\mu\nu}$.

Metric Determinant Since

$$\det g_{\mu\nu} = \det\left(\Omega^2\bar{g}_{\mu\nu}\right) = \Omega^{2d}\det\bar{g}_{\mu\nu}, \tag{11.9.3}$$

we must also have

$$|g|^{1/2} = \Omega^d|\bar{g}|^{1/2}. \tag{11.9.4}$$

Scalar Gradients The scalar gradient with a lower index is just a partial derivative. Therefore

$$\nabla_\mu\varphi = \overline{\nabla}_\mu\varphi = \partial_\mu\varphi. \tag{11.9.5}$$

while $\nabla^\mu\varphi = g^{\mu\nu}\nabla_\nu\varphi = \Omega^{-2}\bar{g}^{\mu\nu}\overline{\nabla}_\nu\varphi$, so

$$\nabla^\mu\varphi = \Omega^{-2}\overline{\nabla}^\mu\varphi. \tag{11.9.6}$$

Scalar Wave Operator The wave operator $\Box$ in the geometry $g_{\mu\nu}$ is defined as

$$\Box \equiv g^{\mu\nu}\nabla_\mu\nabla_\nu = \nabla_\mu\nabla^\mu. \tag{11.9.7}$$

By a direct calculation, the wave operator $\Box$ with respect to $g_{\mu\nu}$ acting on a scalar ψ is

$$\Box\varphi = \frac{1}{\Omega^2}\left(\frac{d-2}{\Omega}\overline{\nabla}_\mu\Omega\cdot\overline{\nabla}^\mu\varphi + \overline{\Box}\varphi\right), \tag{11.9.8}$$

where $\overline{\Box}$ is the wave operator with respect to $\bar{g}_{\mu\nu}$. We also have

$$\begin{aligned}\Box\left(\Omega^s\psi\right) = \frac{1}{\Omega^2}\Big\{ &\left(s\Omega^{s-1}\overline{\Box}\Omega + s\left(d+s-3\right)\Omega^{s-2}\overline{\nabla}_\mu\Omega\overline{\nabla}^\mu\Omega\right)\psi \\ &+ \left(2s+d-2\right)\Omega^{s-1}\overline{\nabla}_\mu\Omega\overline{\nabla}^\mu\psi + \Omega^s\overline{\Box}\psi\Big\}.\end{aligned} \tag{11.9.9}$$

Christoffel Symbols A direct calculation shows:

$$\Gamma^{\mu}_{\alpha\beta}[g] = \overline{\Gamma}^{\mu}_{\alpha\beta}[\bar{g}] + \left(\partial_{\{\alpha} \ln \Omega\right) \delta^{\mu}_{\beta\}} - \bar{g}_{\alpha\beta}\bar{g}^{\mu\nu} \left(\partial_{\nu} \ln \Omega\right) \tag{11.9.10}$$

$$= \overline{\Gamma}^{\mu}_{\alpha\beta}[\bar{g}] + \left(\overline{\nabla}_{\{\alpha} \ln \Omega\right) \delta^{\mu}_{\beta\}} - \bar{g}_{\alpha\beta}\overline{\nabla}^{\mu} \ln \Omega. \tag{11.9.11}$$

Riemann Tensor By viewing the difference between $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$ as a 'perturbation',

$$g_{\mu\nu} - \bar{g}_{\mu\nu} = \left(\Omega^2 - 1\right) \bar{g}_{\mu\nu} \equiv h_{\mu\nu}, \tag{11.9.12}$$

we may employ the results in §(11.8). In particular, eq. (11.8.26) may be used to infer that the Riemann tensor is

$$\begin{aligned} R^{\alpha}{}_{\beta\mu\nu}[g] = \bar{R}^{\alpha}{}_{\beta\mu\nu}[\bar{g}] &+ \overline{\nabla}_{\beta}\overline{\nabla}_{[\mu} \ln \Omega \delta^{\alpha}_{\nu]} - \bar{g}_{\beta[\nu}\overline{\nabla}_{\mu]}\overline{\nabla}^{\alpha} \ln \Omega \\ &+ \delta^{\alpha}_{[\mu}\overline{\nabla}_{\nu]} \ln \Omega \overline{\nabla}_{\beta} \ln \Omega + \overline{\nabla}^{\alpha} \ln \Omega \overline{\nabla}_{[\mu} \ln \Omega \bar{g}_{\nu]\beta} + \left(\overline{\nabla} \ln \Omega\right)^2 \bar{g}_{\beta[\mu}\delta^{\alpha}_{\nu]}. \end{aligned} \tag{11.9.13}$$

Ricci Tensor In turn, the Ricci tensor is

$$R_{\beta\nu}[g] = \bar{R}_{\beta\nu}[\bar{g}] - \bar{g}_{\beta\nu}\overline{\Box} \ln \Omega - (d-2)\overline{\nabla}_{\beta}\overline{\nabla}_{\nu} \ln \Omega \tag{11.9.14}$$

$$+ (d-2)\left(\overline{\nabla}_{\beta} \ln \Omega \overline{\nabla}_{\nu} \ln \Omega - \bar{g}_{\beta\nu} \left(\overline{\nabla} \ln \Omega\right)^2\right). \tag{11.9.15}$$

Ricci Scalar Contracting the Ricci tensor with $g^{\beta\nu} = \Omega^{-2}\bar{g}^{\beta\nu}$, we conclude

$$\mathcal{R}[g] = \Omega^{-2} \left(\overline{\mathcal{R}}[\bar{g}] - 2(d-1)\overline{\Box} \ln \Omega - (d-2)(d-1) \left(\overline{\nabla} \ln \Omega\right)^2\right). \tag{11.9.16}$$

Weyl Tensor The Weyl tensor $C_{\mu\nu\alpha\beta}$ is well defined for spacetime dimensions greater than two ($d > 2$): it is the completely trace-free portion of the Riemann tensor. For $d = 3$, Weyl is identically zero, $C_{\mu\nu\alpha\beta} = 0$. For $d \geq 4$, it can be expressed as "the Riemann tensor minus its trace parts", where the "trace parts" are the Ricci tensor and scalar terms:

$$C_{\mu\nu\alpha\beta} \equiv R_{\mu\nu\alpha\beta} - \frac{1}{d-2} \left(R_{\alpha[\mu}g_{\nu]\beta} - R_{\beta[\mu}g_{\nu]\alpha}\right) + \frac{g_{\mu[\alpha}g_{\beta]\nu}}{(d-2)(d-1)}\mathcal{R}[g]. \tag{11.9.17}$$

By a direct calculation, one may verify $C_{\mu\nu\alpha\beta}$ has the same index-symmetries as $R_{\mu\nu\alpha\beta}$, namely

$$C_{\mu\nu\alpha\beta} = C_{\alpha\beta\mu\nu} \tag{11.9.18}$$

$$C_{\mu\nu\alpha\beta} = -C_{\nu\mu\alpha\beta}, \tag{11.9.19}$$

and is indeed completely traceless: $g^{\mu\alpha}C_{\mu\nu\alpha\beta} = 0$. It also obeys the Bianchi identity

$$C_{\mu[\nu\alpha\beta]} = 0. \tag{11.9.20}$$

Using equations (11.9.1), (11.9.13), (11.9.14), and (11.9.16), one may then deduce the Weyl tensor with one upper index is *invariant* under conformal transformations:

$$C^{\mu}{}_{\nu\alpha\beta}[g] = C^{\mu}{}_{\nu\alpha\beta}[\bar{g}]. \tag{11.9.21}$$

If we lower the index μ on both sides,

$$C_{\mu\nu\alpha\beta}[g] = \Omega^2 C_{\mu\nu\alpha\beta}[\bar{g}]. \tag{11.9.22}$$

Let us also record that:

In spacetime dimensions greater than 3, i.e., $d \geq 4$, a metric $g_{\mu\nu}$ is locally conformally flat – i.e., it can be put into the form $g_{\mu\nu} = \Omega^2 \eta_{\mu\nu}$ – iff its Weyl tensor is zero.[135]

Problem 11.60. Weyl Tensor: Construction Given that the Ricci tensor is the only non-trivial single-contraction of Riemann, and the Ricci scalar is the only non-trivial twice-contracted form of Riemann, argue from the (anti-)symmetries of its indices that the Weyl has to take the form

$$C_{\mu\nu\alpha\beta} \equiv R_{\mu\nu\alpha\beta} + C_1 \left(R_{\alpha[\mu} g_{\nu]\beta} - R_{\beta[\mu} g_{\nu]\alpha} \right) + C_2 \; g_{\mu[\alpha} g_{\beta]\nu} \mathcal{R}. \tag{11.9.23}$$

By requiring the Weyl to be traceless, solve for $C_{1,2}$ and obtain eq. (11.9.17). □

Problem 11.61. Divergence of Weyl Prove the identity

$$\nabla^\mu C_{\mu\nu\alpha\beta} = \frac{d-3}{d-2} \nabla_{[\alpha} R_{\beta]\nu} + \frac{d-3}{2(d-1)(d-2)} g_{\nu[\alpha} \nabla_{\beta]} \mathcal{R}. \tag{11.9.24}$$

Hint: Use eq. (11.3.53). □

Problem 11.62. Number of Algebraically Independent Components Prove that the Weyl tensor $C^\mu{}_{\nu\alpha\beta}$ in d dimensions has

$$\frac{d(d+1)(d+2)(d-3)}{12} \tag{11.9.25}$$

algebraically independent components. This tells us Weyl is zero in $d = 3$ dimensions – a fact that can also be verified by brute force using eq. (11.9.17).

Hints: Weyl has the same algebraic index symmetries as Riemann. Furthermore, it has to obey the traceless condition $C^\mu{}_{\nu\mu\beta} = 0$. How many constraints are there in this traceless condition? □

Problem 11.63. Weyl and Maximal Symmetry Prove that the Weyl tensor is zero in a maximally symmetric space(time). □

Problem 11.64. Cosmological Perturbation Theory & Gauge-Invariance Consider a perturbed metric of the form

$$g_{\mu\nu} = \Omega^2 \left(\eta_{\mu\nu} + \chi_{\mu\nu} \right), \qquad |\chi_{\mu\nu}| \ll 1. \tag{11.9.26}$$

(Cosmological perturbation theory is a special case; where Ω describes the relative size of the universe.) Explain why the linearized Weyl tensor $\delta_1 C_{\mu\nu\alpha\beta}$ – i.e., the part of $C_{\mu\nu\alpha\beta}[g]$ linear in $\chi_{\mu\nu}$ – is gauge-invariant. Hint: See Problem (11.54). □

Einstein Tensor From equations (11.9.1), (11.9.14) and (11.9.16), we may also compute the transformation of the Einstein tensor $G_{\beta\nu} \equiv R_{\beta\nu} - (g_{\beta\nu}/2)\mathcal{R}$.

$$G_{\beta\nu}[g] = \overline{G}_{\beta\nu}[\bar{g}] - (d-2) \left(\overline{\nabla}_\beta \overline{\nabla}_\nu \ln \Omega - \bar{g}_{\beta\nu} \overline{\Box} \ln \Omega \right)$$

[135]In $d = 3$ dimensions, a spacetime is locally conformally flat iff its Cotton tensor vanishes.

$$+ (d-2)\left(\overline{\nabla}_\beta \ln \Omega \overline{\nabla}_\nu \ln \Omega + \frac{d-3}{2}\bar{g}_{\beta\nu}\left(\overline{\nabla}\ln\Omega\right)^2\right) \tag{11.9.27}$$

The first consequence of eq. (11.9.27) is that the Einstein tensor is invariant under constant conformal transformations in any dimension:

$$G_{\beta\nu}[g] = \overline{G}_{\beta\nu}[\bar{g}] \qquad \text{whenever } \partial_\mu \Omega = 0. \tag{11.9.28}$$

The second consequence of eq. (11.9.27) is that, since 2D space(time)s are always locally conformally flat, i.e., $g_{\mu\nu}[2\text{D}] = \Omega^2 \eta_{\mu\nu}$, and since $G_{\mu\nu}[\eta] = 0$, we see that the Einstein tensor must be identically zero when $d = 2$.

Problem 11.65. 2D Geometric Tensors YZ: Riemann is already traceless in 2D and 3D. That relates Riemann to Ricci tensor & scalar. But Ricci tensor is related to Ricci scalar by Einstein is zero. Hence, Riemann is related to Ricci scalar.

Explain why the Einstein tensor is zero in 2D. This implies the 2D Ricci tensor is proportional to the Ricci scalar:

$$R_{\alpha\beta} = \frac{1}{2} g_{\alpha\beta}\mathcal{R}. \tag{11.9.29}$$

Hint: Refer to eq. (11.9.27). □

Scalar Field Action In d dimensional spacetime, the following action involving the scalar φ and Ricci scalar $\mathcal{R}[g]$,

$$S[\varphi] \equiv \int \mathrm{d}^d x \sqrt{|g|}\frac{1}{2}\left(g^{\alpha\beta}\nabla_\alpha\varphi\nabla_\beta\varphi + \frac{d-2}{4(d-1)}\mathcal{R}\varphi^2\right), \tag{11.9.30}$$

is invariant – up to surface terms – under the simultaneous replacements

$$g_{\alpha\beta} \to \Omega^2 g_{\alpha\beta}, \qquad g^{\alpha\beta} \to \Omega^{-2} g^{\alpha\beta}, \qquad \sqrt{|g|} \to \Omega^d \sqrt{|g|}, \tag{11.9.31}$$

$$\varphi \to \Omega^{1-\frac{d}{2}}\varphi. \tag{11.9.32}$$

The jargon here is that φ transforms covariantly under conformal transformations, with weight $s = 1 - (d/2)$. We see in two dimensions, $d = 2$, a minimally coupled massless scalar theory automatically enjoys conformal/Weyl symmetry.

To reiterate: on the right-hand-sides of these expressions for the Riemann tensor, Ricci tensor and scalar, all indices are raised and lowered with $\bar{g}$; for example, $(\overline{\nabla}A)^2 \equiv \bar{g}^{\sigma\tau}\overline{\nabla}_\sigma A \overline{\nabla}_\tau A$ and $\overline{\nabla}^\alpha A \equiv \bar{g}^{\alpha\lambda}\overline{\nabla}_\lambda A$. The $R^\alpha{}_{\beta\mu\nu}[g]$ is built out of the metric $g_{\alpha\beta}$ but the $\bar{R}^\alpha{}_{\beta\mu\nu}[\bar{g}]$ is built entirely out of $\bar{g}_{\mu\nu}$, etc.

Problem 11.66. Weyl Gravity Suppose

$$C^2[g] \equiv C_{\mu\nu\alpha\beta}[g] C^{\mu\nu\alpha\beta}[g] \tag{11.9.33}$$

denotes the square of the Weyl tensor, where $C_{\mu\nu\alpha\beta}$ is built out of $g_{\mu\nu} = \Omega^2 \bar{g}_{\mu\nu}$, and all indices are moved with the same metric $g_{\mu\nu}$. Show that

$$\int \mathrm{d}^d x \sqrt{|g|} C^2[g] = \int \mathrm{d}^d x \sqrt{|\bar{g}|}\Omega^{d-4} C^2[\bar{g}]; \tag{11.9.34}$$

where $C^2[\bar{g}]$ is now built entirely out of $\bar{g}_{\mu\nu}$ and not $g_{\mu\nu}$ – i.e., it does not depend on Ω. Notice: this integral object is conformally invariant in $d = 4$ spacetime dimensions. □

Problem 11.67. de Sitter as a Maximally Symmetric Spacetime Verify that de Sitter spacetime, with coordinates $x^\mu \equiv (\eta, x^i)$,

$$g_{\mu\nu}(\eta, \vec{x}) = \Omega(\eta)^2 \eta_{\mu\nu}, \qquad \Omega(\eta) \equiv -\frac{1}{H\eta} \tag{11.9.35}$$

has the following Riemann tensor:

$$R_{\mu\nu\alpha\beta} = \frac{\mathcal{R}}{d(d-1)} g_{\mu[\alpha} g_{\beta]\nu}. \tag{11.9.36}$$

Also verify that the Ricci tensor and scalar are

$$R_{\mu\nu} = \frac{\mathcal{R}}{d} g_{\mu\nu} \qquad \text{and} \qquad \mathcal{R} = -\frac{2d\Lambda}{d-2}. \tag{11.9.37}$$

de Sitter spacetime is a maximally symmetric spacetime, with $d(d+1)/2$ Killing vectors.[136] Verify that the following are Killing vector of eq. (11.9.35):

$$T^\mu \partial_\mu \equiv -H x^\mu \partial_\mu \tag{11.9.38}$$

and

$$K^\mu_{(i)} \partial_\mu \equiv x^i T^\mu \partial_\mu - H \bar{\sigma} \partial_{x^i}, \tag{11.9.39}$$

$$\bar{\sigma} \equiv \frac{1}{2}\left(\eta^2 - \vec{x}^2\right) = \frac{1}{2}\eta_{\mu\nu} x^\mu x^\nu. \tag{11.9.40}$$

(Hint: It is easier to use the right hand side of eq. (11.5.4) in eq. (11.5.5).) Can you write down the remaining Killing vectors? (Hint: Think about the symmetries on a constant-η surface.) Using (some of) these $d(d+1)/2$ Killing vectors and eq. (11.5.31), explain why the Ricci scalar of the de Sitter geometry is a spacetime constant.

Observer time de Sitter spacetime may also be written as

$$\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht} \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}. \tag{11.9.41}$$

Can you describe the relation between η and t? Why is t dubbed the observer time? (Hint: What is the unit timelike geodesic vector?) Now, explain why the Killing vector in eq. (11.9.38) may also be expressed as

$$T^\mu \partial_\mu = \frac{1}{\Omega(\eta)} \partial_\eta - H x^i \partial_i = \partial_t - H x^i \partial_i. \tag{11.9.42}$$

This means we may take the flat spacetime limit by setting $H \to 0$, and hence identify $T^\mu \partial_\mu$ as the de Sitter analog of the generator of time translation symmetry in Minkowski spacetime. □

[136] As Weinberg explains in [24], maximally symmetric spacetimes are essentially unique, in that they are characterized by a single dimension-ful scale. We see that this scale is nothing but the cosmological constant Λ.

11.10 ⋆2D Spacetimes

In this section we shall lay out the properties of two dimensional space(time)s.[137] The differential geometry of 2D spacetimes is relevant for the study of cosmic strings and (super)string theory.

Conformal Flatness In Problem (9.67), we have already seen a proof that all 2D metrics are locally conformally flat. That is, any 2D metric

$$\mathrm{d}s^2 = g_{00}(\vec{x})(\mathrm{d}x^0)^2 + 2g_{01}(\vec{x})\mathrm{d}x^0\mathrm{d}x^1 + g_{11}(\vec{x})(\mathrm{d}x^1)^2 \tag{11.10.1}$$

may *always* be re-cast into

$$\mathrm{d}s^2 = f(u,v)\mathrm{d}u\mathrm{d}v = f(t,x)\left(\mathrm{d}t^2 - \mathrm{d}x^2\right), \tag{11.10.2}$$

$$t - x = u \qquad \text{and} \qquad t + x = v. \tag{11.10.3}$$

Riemann and Ricci Because of the (anti)symmetry properties of Riemann's indices, its only algebraically independent component is $R_{0101} = -R_{1001} = -R_{0110}$. This allows us to assert, the Riemann tensor is determined completely by the Ricci scalar $\mathcal{R}$:

$$R_{\mu\nu\alpha\beta} = \frac{\mathcal{R}}{2}\left(g_{\mu\alpha}g_{\beta\nu} - g_{\mu\beta}g_{\alpha\nu}\right) = \frac{\mathcal{R}}{2}g_{\mu[\alpha}g_{\beta]\nu}. \tag{11.10.4}$$

This in turn tells us the Ricci tensor is proportional to both the Ricci scalar and the metric:

$$R_{\alpha\beta} = \frac{\mathcal{R}}{2}g_{\alpha\beta}. \tag{11.10.5}$$

Notice the right hand side of eq. (11.10.4) is antisymmetric under $(\mu \leftrightarrow \nu)$ and $(\alpha \leftrightarrow \beta)$; and symmetric under $(\mu\nu) \leftrightarrow (\alpha\beta)$ – exactly the same (anti)symmetries of Riemann. Let us verify eq. (11.10.4) in a 'locally flat' coordinate system; see, for e.g., eq. (11.4.8)–(11.4.10). As already alluded to, the only independent component is 0101. On the left hand side, we have R_{0101}. Whereas on the right hand side,

$$\frac{\mathcal{R}}{2}\left(\eta_{00}\eta_{11} - \eta_{01}\eta_{01}\right) = -\frac{1}{2}\eta^{\alpha\mu}\eta^{\beta\nu}R_{\alpha\beta\mu\nu} = -\frac{1}{2}\eta^{\beta\nu}\left(R_{0\beta0\nu} - R_{1\beta1\nu}\right) \tag{11.10.6}$$

$$= -\frac{1}{2}\left(-R_{0101} - R_{1010}\right) = R_{0101}. \tag{11.10.7}$$

This completes the proof, since eq. (11.10.4) is a tensor equation and we are therefore allowed to verify it in any coordinate system.

Weyl In 4 dimensions and higher, the 'trace-less' part of the Riemann tensor is non-trivial, and is dubbed the Weyl tensor $C_{\mu\nu\alpha\beta}$. Because the 'trace' parts of Riemann consists of Ricci, we see from eq. (11.10.4) that Riemann is pure trace in 2D and therefore the corresponding Weyl tensor is identically zero.

$$C_{\mu\nu\alpha\beta} = 0 \tag{11.10.8}$$

[137]Note that 1D spaces $\mathrm{d}\ell^2 = g(x)\mathrm{d}x^2$ are always (locally) flat, because we may also find y up to an additive constant as $y = \int\sqrt{g(x)}\mathrm{d}x$ and thereby set $g(x)\mathrm{d}x^2 = \mathrm{d}y^2$. To this end, we also note that the only component of the Riemann tensor is $R_{1111} = -R_{1111} = 0$, since it is antisymmetric in the first two and last two indices.

Also note that, while our discussion here pertains to 2D spacetimes, much of it holds for 2D spaces as well.

In other words, if we try to construct the Weyl tensor by subtracting out the trace parts of Riemann – finding the right coefficients $A_{1,2}$ in

$$C_{\mu\nu\alpha\beta} = R_{\mu\nu\alpha\beta} - A_1 R_{\mu[\alpha} g_{\beta]\nu} - A_2 \mathcal{R} \cdot g_{\mu[\alpha} g_{\beta]\nu} \tag{11.10.9}$$

– such that $g^{\mu\alpha} C_{\mu\nu\alpha\beta} = 0$, we will find a trivial result $C_{\mu\nu\alpha\beta} = 0$ because Riemann is already proportional to the Ricci scalar.

Cotton Tensor

Strings in space

11.11 *Metric Sign Convention Dependence

12 Linear Partial Differential Equations (PDEs)

A partial differential equation (PDE) is a differential equation involving more than one variable. Much of fundamental physics – electromagnetism, quantum mechanics, gravitation and more – involves PDEs. We will first examine Poisson's equation, and introduce the concept of the Green's function, in order to solve it. Because the Laplacian $\vec{\nabla}^2$ will feature a central role in our study of PDEs, we will study its eigenfunctions/values in various contexts. Then we will use their spectra to tackle the heat/diffusion equation via an initial value formulation. In the final sections we will study the wave equation in flat spacetime, and study various routes to obtain its solutions, both in position/real spacetime and in Fourier space.

12.1 Laplacians and Poisson's Equation

12.1.1 Poisson's equation, uniqueness of solutions

Poisson's equation in D-space is defined to be

$$-\vec{\nabla}^2\psi(\vec{x}) = J(\vec{x}), \tag{12.1.1}$$

where J is to be interpreted as some given mass/charge density that sources the Newtonian/electric potential ψ. The most physically relevant case is in 3D; if we use Cartesian coordinates, Poisson's equation reads

$$-\vec{\nabla}^2\psi(\vec{x}) = -\left(\frac{\partial^2\psi}{\partial(x^1)^2} + \frac{\partial^2\psi}{\partial(x^2)^2} + \frac{\partial^2\psi}{\partial(x^3)^2}\right) = J(\vec{x}). \tag{12.1.2}$$

We will soon see how to solve eq. (12.1.1) by first solving for the inverse of the negative Laplacian ($\equiv$ Green's function).

Uniqueness of solution We begin by showing that the solution of Poisson's equation (eq. (12.1.1)) in some domain $\mathfrak{D}$ is unique once ψ is specified on the boundary of the domain $\partial\mathfrak{D}$. As we shall see, this theorem holds even in curved spaces. If it is the normal derivative $n^i\nabla_i\psi$ that is specified on the boundary $\partial\mathfrak{D}$, then ψ is unique up to an additive constant.

The proof goes by contradiction. Suppose there were two distinct solutions, ψ_1 and ψ_2. Let us define their difference as

$$\Psi \equiv \psi_1 - \psi_2 \tag{12.1.3}$$

and start with the integral

$$I \equiv \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}\nabla_i\Psi^\dagger\nabla^i\Psi \geq 0. \tag{12.1.4}$$

That this is greater or equal to zero, even in curved spaces, can be seen by writing the gradients in an orthonormal frame (cf. eq. (9.4.11)), where $g^{ij} = \varepsilon_{\widehat{a}}{}^i\varepsilon_{\widehat{b}}{}^j\delta^{ab}$.[138] The $\sqrt{|g|}$ is always positive, since it describes volume, whereas $\nabla_i\Psi\nabla^i\Psi$ is really a sum of squares.

$$\sqrt{|g|}\delta^{ab}\nabla_{\widehat{a}}\Psi^\dagger\nabla_{\widehat{b}}\Psi = \sqrt{|g|}\sum_a \left|\nabla_{\widehat{a}}\Psi\right|^2 \geq 0. \tag{12.1.5}$$

[138] Expressing the gradients in an orthonormal frame is, in fact, the primary additional ingredient to this proof, when compared to the flat space case. Moreover, notice this proof relies on the Euclidean (positive definite) nature of the metric.

We may now integrate-by-parts eq. (12.1.4) and use the curved space Gauss' theorem in eq. (9.7.84).

$$I = \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i \cdot \Psi^\dagger \nabla^i \Psi - \int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g|} \cdot \Psi^\dagger \nabla_i \nabla^i \Psi. \tag{12.1.6}$$

Remember from eq. (9.7.76) that $\mathrm{d}^{D-1}\Sigma_i \nabla^i \Psi = \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|} n^i \nabla_i \Psi$, where $\vec{x}(\vec{\xi})$ parametrizes the boundary $\partial\mathfrak{D}$; $H(\vec{\xi})$ is the determinant of the induced metric on $\partial\mathfrak{D}$ so that $\mathrm{d}^{D-1}\vec{\xi}\sqrt{|H|}$ is its infinitesimal area element and $n^i(\partial\mathfrak{D})$ its unit outward normal. If either $\psi(\partial\mathfrak{D})$ or $n^i\partial_i\psi(\partial\mathfrak{D})$ is specified, therefore, the first term on the right hand side of eq. (12.1.6) is zero – since $\Psi(\partial\mathfrak{D}) = \psi_1(\partial\mathfrak{D}) - \psi_2(\partial\mathfrak{D})$ and $n^i\partial_i\Psi(\partial\mathfrak{D}) = n^i\partial_i\psi_1(\partial\mathfrak{D}) - n^i\partial_i\psi_2(\partial\mathfrak{D})$. The seccond term is zero too, since

$$-\nabla_i\nabla^i\Psi = -\nabla_i\nabla^i(\psi_1 - \psi_2) = J - J = 0. \tag{12.1.7}$$

But we have just witnessed how I is itself the integral, over the domain, of the sum of squares of $|\nabla_{\widehat{a}}\Psi|$. The only way summing squares of something is zero is that something is identically zero.

$$\nabla_{\widehat{a}}\Psi = \varepsilon_{\widehat{a}}{}^i\partial_i\Psi = 0, \qquad \text{(everywhere in } \mathfrak{D}). \tag{12.1.8}$$

Viewing the $\varepsilon_{\widehat{a}}{}^i$ as a vector field, so $\nabla_{\widehat{a}}\Psi$ is the derivative of Ψ in the ath direction, this translates to the conclusion that $\Psi = \psi_1 - \psi_2$ is constant in every direction, all the way up to the boundary; i.e., ψ_1 and ψ_2 can at most differ by an additive constant. If the normal derivative $n^i\nabla_i\psi(\partial\mathfrak{D})$ were specified, so that $n^i\nabla_i\Psi = 0$ there, then $\psi_1(\vec{x}) - \psi_2(\vec{x}) =$ non-zero constant can still yield the same normal derivative. However, if instead $\psi(\partial\mathfrak{D})$ were specified on the boundary, $\Psi(\partial\mathfrak{D}) = 0$ there, and must therefore be zero everywhere in $\mathfrak{D}$. In other words $\psi_1 = \psi_2$, and there cannot be more than 1 distinct solution. This completes the proof.

12.1.2 (Negative) Laplacian as a Hermitian operator

We will now demonstrate that the negative Laplacian in some domain $\mathfrak{D}$ can be viewed as a Hermitian operator, if its eigenfunctions obey

$$\{\psi_\lambda(\partial\mathfrak{D}) = 0\} \qquad \text{(Dirichlet)} \tag{12.1.9}$$

or

$$\{n^i\nabla_i\psi_\lambda(\partial\mathfrak{D}) = 0\} \qquad \text{(Neumann)}, \tag{12.1.10}$$

or if there are *no boundaries*.[139] The steps we will take here are very similar to those in the uniqueness proof above. Firstly, by Hermitian we mean the negative Laplacian enjoys the property that

$$\left\langle \psi_1 \left| \vec{P}^2 \right| \psi_2 \right\rangle \equiv \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\psi_1^\dagger(\vec{x}) \left(-\vec{\nabla}_{\vec{x}}^2\psi_2(\vec{x})\right) \tag{12.1.11}$$

[139]In this chapter on PDEs we will focus mainly on Dirichlet (and occasionally, Neumann) boundary conditions. There are plenty of other possible boundary conditions, of course.

$$= \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\left(-\vec{\nabla}^2_{\vec{x}}\psi_1^\dagger(\vec{x})\right)\psi_2(\vec{x}) \equiv \left(\vec{P}^2\left|\psi_1\right\rangle\right)^\dagger\left|\psi_2\right\rangle, \tag{12.1.12}$$

for any functions $\psi_{1,2}(\vec{x})$ spanned by the eigenfunctions of $-\vec{\nabla}^2$, and therefore satisfy the same boundary conditions. We begin on the left hand side and again employ the curved space Gauss' theorem in eq. (9.7.84).

$$\begin{aligned}\left\langle\psi_1\left|\vec{P}^2\right|\psi_2\right\rangle &= \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i\psi_1^\dagger\left(-\nabla^i\psi_2\right) + \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}\nabla_i\psi_1^\dagger\nabla^i\psi_2, \\ &= \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i\left\{\psi_1^\dagger\left(-\nabla^i\psi_2\right) + \left(\nabla^i\psi_1^\dagger\right)\psi_2\right\} + \left(\vec{P}^2\left|\psi_1\right\rangle\right)^\dagger\left|\psi_2\right\rangle. \end{aligned} \tag{12.1.13}$$

We see that, if either $\psi_{1,2}(\partial\mathfrak{D}) = 0$, or $n^i\nabla_i\psi_{1,2}(\partial\mathfrak{D}) = 0$, the surface integrals vanish, and the Hermitian nature of the Laplacian is established. If there were no boundaries to begin with, note that all the surface terms would not be present – i.e., $\vec{P}^2 = -\vec{\nabla}^2$ would then be automatically Hermitian.

Non-negative eigenvalues Let us understand the bounds on the spectrum of the negative Laplacian subject to the Dirichlet (eq. (12.1.9)) or Neumann boundary (eq. (12.1.10)) conditions, or when there are no boundaries. Let ψ_λ be an eigenfunction obeying

$$-\vec{\nabla}^2\psi_\lambda = \lambda\psi_\lambda. \tag{12.1.14}$$

We have previously argued that

$$I' = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}\nabla_i\psi_\lambda^\dagger\nabla^i\psi_\lambda \tag{12.1.15}$$

is strictly non-negative. If we integrate-by-parts,

$$I' = \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_i\psi_\lambda^\dagger\nabla^i\psi_\lambda + \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}\psi_\lambda^\dagger\left(-\nabla_i\nabla^i\psi_\lambda\right) \geq 0. \tag{12.1.16}$$

If there are no boundaries – for example, if $\mathfrak{D}$ is a $(n \geq 2)$-sphere (usually denoted as $\mathbb{S}^n$) – there will be no surface terms; if there are boundaries but the eigenfunctions obey either Dirichlet conditions in eq. (12.1.9) or Neumann conditions in eq. (12.1.10), the surface terms will vanish. In all three cases, we see that the corresponding eigenvalues $\{\lambda\}$ are strictly non-negative, since $\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}|\psi_\lambda|^2 \geq 0$:

$$I' = \lambda\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}|\psi_\lambda|^2 \geq 0. \tag{12.1.17}$$

Problem 12.1. Robin Boundary Conditions Instead of Dirichlet or Neumann boundary conditions, let us allow for mixed (aka Robin) boundary conditions, namely

$$\alpha\cdot\psi + \beta\cdot n^i\nabla_i\psi = 0 \tag{12.1.18}$$

on the boundary $\partial\mathfrak{D}$. Show that the negative Laplacian is Hermitian if we impose $\alpha/\alpha^* = \beta/\beta^*$. In particular, eq. (12.1.18) automatically yields a Hermitian Laplacian whenever α and β are both real. □

12.1.3 Inverse of the negative Laplacian: Green's function and reciprocity

Given the Dirichlet boundary condition in eq. (12.1.9), i.e., $\{\psi_\lambda(\partial\mathfrak{D}) = 0\}$, we will now understand how to solve Poisson's equation, through the inverse of the negative Laplacian. Roughly speaking,

$$-\vec{\nabla}^2\psi = J \qquad \Rightarrow \qquad \psi = \left(-\vec{\nabla}^2\right)^{-1} J. \tag{12.1.19}$$

(The actual formula, in a finite domain, will be a tad more complicated, but here we are merely motivating the reason for defining G.) Since, given any Hermitian operator

$$H = \sum_\lambda \lambda \left|\lambda\right\rangle \left\langle\lambda\right|, \qquad \{\lambda \in \mathbb{R}\}, \tag{12.1.20}$$

its inverse is

$$H^{-1} = \sum_\lambda \frac{\left|\lambda\right\rangle \left\langle\lambda\right|}{\lambda}, \qquad \{\lambda \in \mathbb{R}\}; \tag{12.1.21}$$

we see that the inverse of the negative Laplacian in the position space representation is the following mode expansion involving its eigenfunctions $\{\psi_\lambda\}$.

$$G(\vec{x}, \vec{x}') = \left\langle \vec{x} \left| \frac{1}{-\vec{\nabla}^2} \right| \vec{x}' \right\rangle = \sum_\lambda \frac{\psi_\lambda(\vec{x})\psi_\lambda(\vec{x}')^\dagger}{\lambda}, \tag{12.1.22}$$

$$-\vec{\nabla}^2\psi_\lambda = \lambda\psi_\lambda, \qquad \psi_\lambda(\vec{x}) \equiv \left\langle \vec{x} \right| \lambda\rangle . \tag{12.1.23}$$

(The summation sign is schematic; it can involve either (or both) a discrete sum or/and an integral over a continuum.) Since the mode functions are subject to $\{\psi_\lambda(\partial\mathfrak{D}) = 0\}$, the Green's function itself also obeys Dirichlet boundary conditions:

$$G(\vec{x} \in \mathfrak{D}, \vec{x}') = G(\vec{x}, \vec{x}' \in \mathfrak{D}) = 0. \tag{12.1.24}$$

The Green's function G satisfies the PDE

$$-\vec{\nabla}^2_{\vec{x}} G(\vec{x}, \vec{x}') = -\vec{\nabla}^2_{\vec{x}'} G(\vec{x}, \vec{x}') = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt[4]{|g(\vec{x})g(\vec{x}'))|}}, \tag{12.1.25}$$

because the negative Laplacian is Hermitian and thus its eigenfunctions obey the following completeness relation (cf. (4.3.23))

$$\begin{aligned} \sum_\lambda \psi_\lambda(\vec{x}')\psi_\lambda(\vec{x})^\dagger &= \left\langle \vec{x}' \right| \left(\sum_\lambda \left|\lambda\right\rangle \left\langle\lambda\right| \right) \left|\vec{x}\right\rangle \\ &= \left\langle \vec{x}' \right| \vec{x}\rangle = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt[4]{|g(\vec{x})g(\vec{x}'))|}}. \end{aligned} \tag{12.1.26}$$

Eq. (12.1.25) follows from $-\vec{\nabla}^2\psi_\lambda = \lambda\psi_\lambda$ and

$$-\vec{\nabla}^2_{\vec{x}} G(\vec{x}, \vec{x}') = \sum_\lambda \frac{-\vec{\nabla}^2_{\vec{x}}\psi_\lambda(\vec{x})\psi_\lambda(\vec{x}')^\dagger}{\lambda} = \sum_\lambda \psi_\lambda(\vec{x})\psi_\lambda(\vec{x}')^\dagger, \tag{12.1.27}$$

$$-\vec{\nabla}^2_{\vec{x}'} G(\vec{x}, \vec{x}') = \sum_\lambda \frac{\psi_\lambda(\vec{x})(-\vec{\nabla}^2_{\vec{x}'}\psi_\lambda(\vec{x}')^\dagger)}{\lambda} = \sum_\lambda \psi_\lambda(\vec{x})\psi_\lambda(\vec{x}')^\dagger. \tag{12.1.28}$$

Because the $\delta^{(D)}$-functions on the right hand side of eq. (12.1.25) is the (position representation) of the identity operator, the Green's function itself is really the inverse of the negative Laplacian.

Field at $\vec{x}$ due to point source at $\vec{x}'$ Physically speaking, by comparing Poisson's equation with the corresponding Green's function equation in eq. (12.1.25), the δ-functions on the right hand side of the latter admit the interpretation that the Green's function is the field at $\vec{x}$ produced by a point source at $\vec{x}'$. Therefore, the Green's function of the negative Laplacian *is* the gravitational/electric potential produced by a unit strength point charge/mass.

Flat $\mathbb{R}^D$ *The* example illustrating the above discussion is provided by the eigenfunctions of the negative Laplacian in infinite D-space.

$$\psi_{\vec{k}}(\vec{x}) = \frac{e^{i\vec{k}\cdot\vec{x}}}{(2\pi)^{D/2}}, \qquad -\vec{\nabla}^2_{\vec{x}}\psi_{\vec{k}}(\vec{x}) = \vec{k}^2\psi_{\vec{k}}(\vec{x}). \tag{12.1.29}$$

Because we know the integral representation of the δ-function, eq. (12.1.26) now reads

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} = \delta^{(D)}(\vec{x}-\vec{x}'). \tag{12.1.30}$$

Through eq. (12.1.22), we may write down the integral representation of the inverse of the negative Laplacian in Euclidean D-space.

$$G(\vec{x}, \vec{x}') = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \frac{e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}}{\vec{k}^2} = \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}|\vec{x}-\vec{x}'|^{D-2}}. \tag{12.1.31}$$

As alluded to, this may be interpreted as the field at $\vec{x}$ due to the unit 'point charge' at $\vec{x}'$. In 3D, this result simplifies to the (hopefully familiar) result from both Newtonian gravity and Coulomb's law:

$$G_3(\vec{x}, \vec{x}') = \frac{1}{4\pi|\vec{x}-\vec{x}'|}. \tag{12.1.32}$$

Boundaries and Method of Images Suppose we now wish to solve the Green's function $G_D(\mathfrak{D})$ of the negative Laplacian in a finite domain of flat space, $\mathfrak{D} \subset \mathbb{R}^D$. One may view $G_D(\mathfrak{D})$ as the sum of its counterpart in infinite $\mathbb{R}^D$ plus a term that is a homogeneous solution $H_D(\mathfrak{D})$ in the finite domain $\mathfrak{D}$, such that the desired boundary conditions are achieved on $\partial\mathfrak{D}$. Namely,

$$\begin{aligned}
G_D(\vec{x}, \vec{x}'; \mathfrak{D}) &= \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}|\vec{x}-\vec{x}'|^{D-2}} + H(\vec{x}, \vec{x}'; \mathfrak{D}), \\
-\vec{\nabla}^2_{\vec{x}} G_D(\vec{x}, \vec{x}'; \mathfrak{D}) = -\vec{\nabla}^2_{\vec{x}'} G_D(\vec{x}, \vec{x}'; \mathfrak{D}) &= \delta^{(D)}\left(\vec{x}-\vec{x}'\right), \qquad \text{(Cartesian coordinates)} \\
-\vec{\nabla}^2_{\vec{x}} H_D(\vec{x}, \vec{x}'; \mathfrak{D}) = -\vec{\nabla}^2_{\vec{x}'} H_D(\vec{x}, \vec{x}'; \mathfrak{D}) &= 0, \qquad \vec{x}, \vec{x}' \in \mathfrak{D}.
\end{aligned} \tag{12.1.33}$$

If Dirichlet boundary conditions are desired, we would demand

$$\frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}|\vec{x}-\vec{x}'|^{D-2}} + H(\vec{x}, \vec{x}'; \mathfrak{D}) = 0 \tag{12.1.34}$$

whenever $\vec{x} \in \partial\mathfrak{D}$ or $\vec{x}' \in \partial\mathfrak{D}$.

The *method of images*, which you will likely learn about in an electromagnetism course, is a special case of such a strategy of solving the Green's function. We will illustrate it through the following example. Suppose we wish to solve the Green's function in a half-infinite space, i.e., for $x^D \geq 0$ only, but let the rest of the $\{x^1, \dots, x^{D-1}\}$ run over the real line. We further want the boundary condition

$$G_D(x^D = 0) = G_D(x'^D = 0) = 0. \tag{12.1.35}$$

The strategy is to notice that the infinite plane that is equidistant between one positive and one negative point mass/charge has zero potential, so if we wish to solve the Green's function (the potential of the positive unit mass) on the half plane, we place a negative unit mass on the opposite side of the boundary at $x^D = 0$. Since the solution to Poisson's equation is unique, the solution for $x^D \geq 0$ is therefore

$$\begin{aligned} &G_D(\vec{x}, \vec{x}'; \mathfrak{D}) \\ &= \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}} \left(\frac{1}{|\vec{x}_\perp - \vec{x}'_\perp + \widehat{e}_D(x^D - x'^D)|^{D-2}} - \frac{1}{|\vec{x}_\perp - \vec{x}'_\perp + \widehat{e}_D(x^D + x'^D)|^{D-2}} \right), \end{aligned} \tag{12.1.36}$$

where $\vec{x}_\perp$ and $\vec{x}'_\perp$ are the coordinate vector perpendicular to the x^D-direction, while $\widehat{e}_D$ is the unit vector pointing in the positive x^D-direction. Mathematically speaking, when the negative Laplacian is applied to the second term in eq. (12.1.36), it yields $\prod_{j=1}^{D-1} \delta(x^j - x'^j)\delta(x^D + x'^D)$, but since $x^D, x'^D \geq 0$, the very last δ-function can be set to zero. Hence, the second term is a homogeneous solution when attention is restricted to $x^D \geq 0$.

Problem 12.2. Dirichlet Green's Function Outside Sphere Consider a spherical surface of radius R in $(D \geq 3)-$flat space. Solve the Dirichlet Green's function G_D of the negative Laplacian *outside* this surface. In particular, ensure that G vanishes *on* this surface by showing that – if the origin of the coordinate system is placed at the center of the sphere –

$$G_D\left(\vec{x}, \vec{x}'; |\vec{x}|, |\vec{x}'| \geq R\right) = \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}} \left(\frac{1}{|\vec{x} - \vec{x}'|^{D-2}} - \frac{Q^{D-2}}{|\vec{x} - Q \cdot R \cdot \widehat{x}'|^{D-2}} \right), \tag{12.1.37}$$

$$Q \equiv R/|\vec{x}'|, \qquad \widehat{x}' \equiv \vec{x}'/|\vec{x}'|. \tag{12.1.38}$$

Explain how this may be viewed as an 'image charge' problem.

Hint: First show that, with $\vec{0}$ located at the center of the sphere, $(\vec{x}-\vec{x}')^2 = (\vec{x}-Q\cdot R\cdot\widehat{x}')^2/Q^2$ whenever $|\vec{x}| = R$ and $Q = R/|\vec{x}'|$. $\square$

Reciprocity We will also now show that the Green's function itself is a Hermitian object, in that

$$G(\vec{x}, \vec{x}')^\dagger = G(\vec{x}', \vec{x}) = G(\vec{x}, \vec{x}'). \tag{12.1.39}$$

The first equality follows from the real positive nature of the eigenvalues, as well as the mode expansion in eq. (12.1.22)

$$G(\vec{x}, \vec{x}')^* = \sum_\lambda \frac{\psi_\lambda(\vec{x}')\psi_\lambda(\vec{x})^\dagger}{\lambda} = G(\vec{x}', \vec{x}). \tag{12.1.40}$$

The second requires considering the sort of integrals we have been examining in this section.

$$I(x,x') \equiv \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}'' \sqrt{|g(\vec{x}'')|} \Big\{ G\left(\vec{x},\vec{x}''\right)(-\vec{\nabla}^2_{\vec{x}''})G\left(\vec{x}',\vec{x}''\right) \\ - G\left(\vec{x}',\vec{x}''\right)(-\vec{\nabla}^2_{\vec{x}''})G\left(\vec{x},\vec{x}''\right) \Big\}. \tag{12.1.41}$$

Using the PDE obeyed by G,

$$I(x,x') = G(\vec{x},\vec{x}') - G(\vec{x}',\vec{x}). \tag{12.1.42}$$

We may integrate-by-parts too.

$$I(x,x') = \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_{i''} \left\{ G(\vec{x},\vec{x}'')(-\nabla^{i''})G(\vec{x}',\vec{x}'') - G(\vec{x}',\vec{x}'')(-\nabla^{i''})G(\vec{x},\vec{x}'') \right\} \\ + \int \mathrm{d}^D\vec{x}'' \sqrt{|g(\vec{x}'')|} \left\{ \nabla_{i''}G(\vec{x},\vec{x}'')\nabla^{i''}G(\vec{x}',\vec{x}'') - \nabla_{i''}G(\vec{x}',\vec{x}'')\nabla^{i''}G(\vec{x},\vec{x}'') \right\}. \tag{12.1.43}$$

The terms in the last line cancel. Moreover, for precisely the same boundary conditions that make the negative Laplacian Hermitian, we see the surface terms have to vanish too. Therefore $I(x,x') = 0 = G(\vec{x},\vec{x}') - G(\vec{x}',\vec{x})$, and we have established the reciprocity of the Green's function.

Problem 12.3. Verify that the Green's function solutions in equations (12.1.36) and (12.1.37) obeys reciprocity. □

Non-invertible Laplacian We see from the mode sum in eq. (12.1.22) that a Hermitian Laplacian has no inverse – its Green's function does not exist – when it has an isolated zero eigenvalue; i.e., when there are no eigenvalues continuously connected to $\lambda = 0$. There are at least two cases where this occurs.

Neumann Boundary Conditions Within a finite domain $\mathfrak{D}$, we see that the Neumann boundary conditions $\{n^i\nabla_i\psi_\lambda(\partial\mathfrak{D}) = 0\}$ imply there must be a zero eigenvalue; for, the $\psi_0 =$ constant is the corresponding eigenvector, whose normal derivative on the boundary is zero:

$$-\vec{\nabla}^2\psi_0 = -\frac{\partial_i\left(\sqrt{|g|}g^{ij}\partial_j\psi_0\right)}{\sqrt{|g|}} = 0\cdot\psi_0. \tag{12.1.44}$$

As long as this is an isolated zero, this mode will contribute a discrete term in the mode sum of eq. (12.1.22) that yields a $1/0$ infinity. That is, the inverse of the Laplacian does not make sense if there is an isolated zero mode.[140]

Another way to see understand why G does not exist is by integrating $-\vec{\nabla}^2 G = \delta^{(D)}$ on both sides over the finite volume of relevance. This would yield

$$-\int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_{i'}\nabla^{i'}G(\vec{x},\vec{x}' \in \partial\mathfrak{D}) = 1. \tag{12.1.45}$$

[140] In the infinite flat $\mathbb{R}^D$ case above, we have seen the $\{\exp(i\vec{k}\cdot\vec{x})\}$ are the eigenfunctions and hence there is also a zero mode, gotten by setting $\vec{k} \to \vec{0}$. However the inverse does exist because the mode sum of eq. (12.1.22) is really an integral, and the integration measure $\mathrm{d}^D\vec{k}$ ensures convergence of the integral.

But, since G obeys Neumann boundary conditions, the left hand side – involving the normal derivative of G with respect to $\vec{x}'$ evaluated on the boundary – is zero. We then obtain a $0 = 1$ contradiction.

Domain without boundary If the domain under study has no boundary – for example, the 2D *closed* surface of a sphere or soap bubble – then the Green's function eq. (12.1.25) cannot be satisfied. For, we may integrate both sides over the domain. The left hand side of eq. (12.1.25), being the divergence of a gradient, would integrate to zero by Gauss' theorem because there will be no surface terms; whereas the right hand side would integrate to unity.

Problem 12.4. Total charge is zero in a closed space Define the total charge to be

$$Q \equiv \int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g(\vec{x})|} J(\vec{x}). \tag{12.1.46}$$

Show that, if the Poisson equation in eq. (12.1.1) holds in a space with no boundaries, the total charge must be zero. The Green's function may be viewed as the field ψ sourced by a unit strength point charge located at $\vec{x}'$; that is why its solution does not exist in a closed space. □

Problem 12.5. Pseudo-Inverse of Laplacian on Unit 2−sphere The 2−sphere described by $\mathrm{d}\ell^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2$ is an example of a geometry without an inverse of its associated Laplacian. Using the spherical harmonics $\{Y_\ell^m(\theta, \phi) | \ell = 0, 1, 2, \dots, -\ell \leq m \leq +\ell\}$, which obey $-\vec{\nabla}^2 Y_\ell^m = \ell(\ell+1) Y_\ell^m$, we may attempt to write the inverse $\sum_\lambda \lambda^{-1} |\lambda\rangle \langle\lambda|$ via the expression

$$\left(-\vec{\nabla}^{-2}\right)(\Omega, \Omega') \stackrel{?}{=} \sum_{\ell=0} \sum_{-\ell \leq m \leq +\ell} \frac{Y_\ell^m(\Omega) \overline{Y_\ell^m}(\Omega')}{\ell(\ell+1)}, \tag{12.1.47}$$

$$\Omega \equiv (\theta, \phi), \qquad \Omega' \equiv (\theta', \phi'); \tag{12.1.48}$$

only to realize the $\ell = 0$ is ill defined because we are dividing by 0. This (discrete) zero eigenvalue is intimately related to the fact that the 2−sphere has no boundaries.

We may, however, proceed to define the pseudo-inverse by simply excluding the offending $\ell = 0$ term:

$$\left(-\vec{\nabla}^{-2}\right)'(\Omega, \Omega') \equiv \sum_{\ell=1} \sum_{-\ell \leq m \leq +\ell} \frac{Y_\ell^m(\Omega) \overline{Y_\ell^m}(\Omega')}{\ell(\ell+1)}. \tag{12.1.49}$$

(Note the prime on the left hand side.) Start by verifying that

$$-\vec{\nabla}^2 \left(-\vec{\nabla}^{-2}\right)'(\Omega, \Omega') = \frac{\delta^{(2)}(\Omega - \Omega')}{\sqrt[4]{g(\Omega) g(\Omega')}} - \frac{1}{4\pi}. \tag{12.1.50}$$

Next, demonstrate that

$$\left(-\vec{\nabla}^{-2}\right)'(\Omega, \Omega') = \frac{\ln[2] - 1 - \ln\left(1 - \widehat{r} \cdot \widehat{r}'\right)}{4\pi}, \tag{12.1.51}$$

where $\widehat{r}(\Omega) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ is the unit vector and $\widehat{r}'$ is its Ω'−counterpart. (Hints: By using Rodrigues' formula or otherwise, first compute the overlap between the ℓ−th Legendre polynomial $P_\ell(\widehat{r} \cdot \widehat{r}')$ with $\ln(1 - \widehat{r} \cdot \widehat{r}')$; then use the addition formula relating $P_\ell(\widehat{r} \cdot \widehat{r}')$ and $Y_\ell^m(\widehat{r})\overline{Y_\ell^m}(\widehat{r}')$.) Finally, explain why

$$G(\Omega, \Omega_\pm) \equiv \frac{1}{4\pi} \ln\left(\frac{1 - \widehat{r} \cdot \widehat{r}_-}{1 - \widehat{r} \cdot \widehat{r}_+}\right) \tag{12.1.52}$$

– where the three unit vectors $\widehat{r}$ and $\widehat{r}_\pm$ are, respectively, functions of the angular coordinates $\Omega \equiv (\theta, \phi)$ and $\Omega_\pm \equiv (\theta_\pm, \phi_\pm)$ – obeys the Poisson equation

$$-\vec{\nabla}_\Omega^2 G(\Omega, \Omega_\pm) = \frac{\delta^{(2)}(\Omega - \Omega_+)}{\sqrt[4]{g(\Omega)g(\Omega_+)}} - \frac{\delta^{(2)}(\Omega - \Omega_-)}{\sqrt[4]{g(\Omega)g(\Omega_-)}}. \tag{12.1.53}$$

The ± 1 coefficients of the $\delta-$functions on the right hand side indicate, even though the inverse Laplacian – the field due to a *single* unit point charge – does not exist, the field due to the unit positive charge located at $\widehat{r}_+$ and a unit negative one located at $\widehat{r}_-$ does in fact exist, since the constraint from Problem (12.4) that total charge be zero is fulfilled. □

Discontinuous first derivatives Because it may not be apparent from the mode expansion in eq. (12.1.22), we close this section by highlighting that the Green's function must contain discontinuous first derivatives as $\vec{x} \to \vec{x}'$ in order to yield, from a second order Laplacian, δ-functions on the right hand side of eq. (12.1.25). For Green's functions in a finite domain $\mathfrak{D}$, there are potentially additional discontinuities when both $\vec{x}$ and $\vec{x}'$ are near the boundary of the domain $\partial\mathfrak{D}$.

12.1.4 Kirchhoff integral theorem and Dirichlet boundary conditions

Within a finite domain $\mathfrak{D}$ we will now understand why the choice of boundary conditions that makes the negative Laplacian a Hermitian operator, is intimately tied to the type of boundary conditions imposed in solving Poisson's equation eq. (12.1.1).

Suppose we have specified the field on the boundary $\psi(\partial\mathfrak{D})$. To solve Poisson's equation $-\vec{\nabla}^2\psi = J$, we will start by imposing Dirichlet boundary conditions on the eigenfunctions of the Laplacian, i.e., $\{\psi_\lambda(\partial\mathfrak{D}) = 0\}$, so that the resulting Green's function obey eq. (12.1.24). The solution to Poisson's equation within the domain $\mathfrak{D}$ can now be solved in terms of G, the source J, and its boundary values $\psi(\partial\mathfrak{D})$ through the following Kirchhoff integral representation:

$$\psi(\vec{x}) = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}') - \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_{i'} \nabla^{i'} G(\vec{x}, \vec{x}') \psi(\vec{x}'). \tag{12.1.54}$$

If there are no boundaries, then the boundary integral terms in eq. (12.1.54) are zero. Similarly, if the boundaries are infinitely far away, the same boundary terms can usually be assumed to vanish, provided the fields involved decay sufficiently quickly at large distances. Physically, the first term can be interpreted to be the ψ directly due to J the source (i.e., the particular solution); whereas the surface integral terms are independent of J and thus correspond to the homogeneous solutions.

Derivation of eq. (12.1.54) Let us now consider the following integral

$$I(\vec{x} \in \mathfrak{D}) \equiv \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}' \sqrt{|g(\vec{x}')|} \left\{ G(\vec{x}, \vec{x}') \left(-\vec{\nabla}_{\vec{x}'}^2 \psi(\vec{x}')\right) - \left(-\vec{\nabla}_{\vec{x}'}^2 G(\vec{x}, \vec{x}')\right) \psi(\vec{x}') \right\} \tag{12.1.55}$$

If we use the equations (12.1.1) and (12.1.25) obeyed by ψ and G respectively, we obtain immediately

$$I(\vec{x}) = \int_{\mathfrak{D}} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}') - \psi(\vec{x}). \tag{12.1.56}$$

On the other hand, we may integrate-by-parts,

$$\begin{aligned} I(\vec{x}) = &\int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_{i'} \left\{ G(\vec{x}, \vec{x}') \left(-\nabla^{i'} \psi(\vec{x}') \right) - \left(-\nabla^{i'} G(\vec{x}, \vec{x}') \right) \psi(\vec{x}') \right\} \\ &+ \int_{\mathfrak{D}} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \left\{ \nabla_{i'} G(\vec{x}, \vec{x}') \nabla^{i'} \psi(\vec{x}') - \nabla^{i'} G(\vec{x}, \vec{x}') \nabla_{i'} \psi(\vec{x}') \right\}. \end{aligned} \tag{12.1.57}$$

The second line cancels. Combining equations (12.1.56) and (12.1.57) then hands us the following Kirchhoff representation:

$$\begin{aligned} \psi(\vec{x} \in \mathfrak{D}) = &\int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\Sigma_{i'} \left\{ G(\vec{x}, \vec{x}') \left(\nabla^{i'} \psi(\vec{x}') \right) - \left(\nabla^{i'} G(\vec{x}, \vec{x}') \right) \psi(\vec{x}') \right\} \\ &+ \int_{\mathfrak{D}} \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}'). \end{aligned} \tag{12.1.58}$$

(The prime on the index in $\nabla^{i'}$ indicates the covariant derivative is with respect to $\vec{x}'$.) If we recall the Dirichlet boundary conditions obeyed by the Green's function $G(\vec{x}, \vec{x}')$ (eq. (12.1.24)), the first term on the right hand side of the first line drops out and we obtain eq. (12.1.54).

Problem 12.6. Dirichlet B.C. Variational Principle In a finite domain where

$$\int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g|} < \infty, \tag{12.1.59}$$

let all fields vanish on the boundary $\partial\mathfrak{D}$ and denote the smallest non-zero eigenvalue of the negative Laplacian $-\vec{\nabla}^2$ as λ_0. Let ψ be an arbitrary function obeying the same boundary conditions as the eigenfunctions of $-\vec{\nabla}^2$. For this problem, assume that the spectrum of the negative Laplacian is discrete. Prove that

$$\frac{\int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g|} \nabla_i \psi^\dagger \nabla^i \psi}{\int_{\mathfrak{D}} \mathrm{d}^D \vec{x} \sqrt{|g|} |\psi|^2} \geq \lambda_0. \tag{12.1.60}$$

Just like in quantum mechanics, we have a variational principle for the spectrum of the negative Laplacian in a finite volume curved space: you can exploit any trial complex function ψ that vanishes on $\mathfrak{D}$ to derive an upper bound for the lowest eigenvalue of the negative Laplacian.

Hint: Expand ψ as a superposition of the eigenfunctions of $-\vec{\nabla}^2$. Then integrate-by-parts one of the ∇^i in the integrand. □

Example Suppose, within a finite 1D box, $x \in [0, L]$ we are provided a real field ψ obeying

$$\psi(x = 0) = \alpha, \qquad \psi(x = L) = \beta \tag{12.1.61}$$

without any external sources. You can probably solve this 1D Poisson's equation ($-\partial_x^2\psi = 0$) right away; it is a straight line:

$$\psi(0 \leq x \leq L) = \alpha + \frac{\beta-\alpha}{L}x. \tag{12.1.62}$$

But let us try to solve it using the methods developed here. First, we recall the orthonormal eigenfunctions of the negative Laplacian with Dirichlet boundary conditions,

$$\begin{aligned}
\langle x|\, n\rangle &= \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi}{L}x\right), \qquad n \in \{1,2,3,\dots\}, \qquad \sum_{n=1}^{\infty}\langle x|\, n\rangle \langle n|\, x'\rangle = \delta(x-x'), \\
-\partial_x^2 \langle x|\, n\rangle &= \left(\frac{n\pi}{L}\right)^2 \langle x|\, n\rangle .
\end{aligned} \tag{12.1.63}$$

The mode sum expansion of the Green's function in eq. (12.1.22) is

$$G(x,x') = \frac{2}{L}\sum_{n=1}^{\infty}\left(\frac{n\pi}{L}\right)^{-2}\sin\left(\frac{n\pi}{L}x\right)\sin\left(\frac{n\pi}{L}x'\right). \tag{12.1.64}$$

The J term in eq. (12.1.54) is zero, while the surface integrals really only involve evaluation at $x = 0, L$. Do be careful that the normal derivative refers to the outward normal.

$$\begin{aligned}
\psi(0 < x < L) &= \partial_{x'}G(x,x'=0)\psi(x'=0) - \partial_{x'}G(x,x'=L)\psi(x'=L) \\
&= -\frac{2}{L}\sum_{n=1}^{\infty}\frac{L}{n\pi}\sin\left(\frac{n\pi}{L}x\right)\left[\cos\left(\frac{n\pi}{L}x'\right)\psi(x')\right]_{x'=0}^{x'=L} \\
&= \sum_{n=1}^{\infty}\frac{2}{n\pi}\sin\left(\frac{n\pi}{L}x\right)(\alpha - (-)^n\cdot\beta)
\end{aligned} \tag{12.1.65}$$

We may check this answer in the following way. Because the solution in eq. (12.1.65) is odd under $x \to -x$, let us we extend the solution in the following way:

$$\begin{aligned}
\psi_\infty(-L \leq x \leq L) &= \alpha + \frac{\beta-\alpha}{L}x, \qquad 0 \leq x \leq L, \\
&= -\left(\alpha + \frac{\beta-\alpha}{L}x\right), \qquad -L \leq x < 0.
\end{aligned} \tag{12.1.66}$$

We will then extend the definition of ψ_∞ by imposing periodic boundary conditions, $\psi_\infty(x+2L) = \psi_\infty(x)$. This yields the Fourier series

$$\psi_\infty(x) = \sum_{\ell=-\infty}^{+\infty} C_\ell e^{i\frac{2\pi\ell}{2L}x}. \tag{12.1.67}$$

Multiplying both sides by $\exp(-i(\pi n/L)x)$ and integrating over $x \in [-L, L]$.

$$C_n = \int_{-L}^{L}\psi_\infty(x)e^{-i\frac{\pi n}{L}x}\frac{\mathrm{d}x}{2L} = \int_{-L}^{L}\psi_\infty(x)\left(\cos\left(\frac{\pi n}{L}x\right) - i\sin\left(\frac{\pi n}{L}x\right)\right)\frac{\mathrm{d}x}{2L}$$

$$= -i \int_0^L \left(\alpha + \frac{\beta - \alpha}{L} x \right) \sin\left(\frac{\pi n}{L} x \right) \frac{\mathrm{d}x}{L}$$
$$= \frac{i}{\pi n} \left((-)^n \beta - \alpha \right). \tag{12.1.68}$$

Putting this back to into the Fourier series,

$$\psi_\infty(x) = i \sum_{n=1}^{+\infty} \frac{1}{\pi n} \left\{ \left((-)^n \beta - \alpha \right) e^{i \frac{\pi n}{L} x} - \left((-)^{-n} \beta - \alpha \right) e^{-i \frac{\pi n}{L} x} \right\}$$
$$= - \sum_{n=1}^{+\infty} \frac{2}{\pi n} \left((-)^n \beta - \alpha \right) \sin\left(\frac{\pi n}{L} x \right). \tag{12.1.69}$$

Is it not silly to obtain a complicated infinite sum for a solution, when it is really a straight line? The answer is that, while the Green's function/mode sum method here does appear unnecessarily complicated, this mode expansion method is very general and is oftentimes the only known means of solving the problem analytically.

Problem 12.7. 2D Case Solve the 2D flat space Poisson equation $-(\partial_x^2 + \partial_y^2)\psi(0 \leq x \leq L_1, 0 \leq y \leq L_2) = 0$, up to quadrature, with the following boundary conditions

$$\psi(0, y) = \varphi_1(y), \qquad \psi(L_1, y) = \varphi_2(y), \qquad \psi(x, 0) = \rho_1(x), \qquad \psi(x, L_2) = \rho_2(x). \tag{12.1.70}$$

Write the solution as a mode sum, using the eigenfunctions

$$\psi_{m,n}(x, y) \equiv \langle x, y | \, m, n \rangle = \frac{2}{\sqrt{L_1 L_2}} \sin\left(\frac{\pi m}{L_1} x \right) \sin\left(\frac{\pi n}{L_2} y \right). \tag{12.1.71}$$

Hint: your answer will involve 1D integrals on the 4 boundaries of the rectangle. $\square$

12.2 Laplacians and their spectra

Let us recall our discussions from both linear algebra and differential geometry. Given a (Euclidean signature) metric

$$\mathrm{d}\ell^2 = g_{ij}(\vec{x}) \mathrm{d}x^i \mathrm{d}x^j, \tag{12.2.1}$$

the Laplacian acting on a scalar ψ can be written as

$$\vec{\nabla}^2 \psi \equiv \nabla_i \nabla^i \psi = \frac{\partial_i \left(\sqrt{|g|} g^{ij} \partial_j \psi \right)}{\sqrt{|g|}}, \tag{12.2.2}$$

where $\sqrt{|g|}$ is the square root of the determinant of the metric.
Spectra Now we turn to the primary goal of this section, to study the eigenvector/value problem

$$-\vec{\nabla}^2 \psi_\lambda(\vec{x}) = -\vec{\nabla}^2 \langle \vec{x} | \, \lambda \rangle = \lambda \langle \vec{x} | \, \lambda \rangle. \tag{12.2.3}$$

If these eigenfunctions are normalized to unit length, namely

$$\int_{\mathfrak{D}} \mathrm{d}^D\vec{x} \, \langle\lambda|\,\vec{x}\rangle \, \langle\vec{x}|\,\lambda'\rangle = \delta^{\lambda}_{\lambda'} \tag{12.2.4}$$

– where the $\delta^{\lambda}_{\lambda'}$ on the right hand side can either be the kronecker delta (for discrete spectra) or the Dirac delta (for continuous ones) – then we have the completeness relation

$$\sum_{\lambda} \langle\vec{x}|\,\lambda\rangle \, \langle\lambda|\,\vec{x}'\rangle = \frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}. \tag{12.2.5}$$

The summation on the left hand side will become an integral for continuous spectra; and the Dirac delta functions on the right hand side should be viewed as the identity operator in the position representation.

12.2.1 Infinite $\mathbb{R}^D$ in Cartesian coordinates

In infinite flat Euclidean D-space $\mathbb{R}^D$, we have already seen that the plane waves $\{\exp(i\vec{k}\cdot\vec{x})\}$ are the eigenvectors of $-\vec{\nabla}^2$ with eigenvalues $\{k^2| -\infty < k < \infty\}$. This is a coordinate invariant statement, since the ψ and Laplacian in eq. (12.2.3) are coordinate scalars. Also notice that the eigenvalue/vector equation (12.2.3) is a "local" PDE in that it is possible to solve it only in the finite neighborhood of $\vec{x}$; it therefore requires appropriate boundary conditions to pin down the correct eigen-solutions.

In Cartesian coordinates, moreover,

$$\psi_{\vec{k}}(\vec{x}) = \langle\vec{x}|\vec{k}\rangle = e^{i\vec{k}\cdot\vec{x}} = \prod_{j=1}^{D} e^{ik_j x^j}, \qquad \vec{k}^2 = \delta^{ij}k_i k_j = \sum_{i=1}^{D}(k_i)^2 \equiv \vec{k}^2, \tag{12.2.6}$$

with completeness relations (cf. eq. (12.1.26)) given by

$$\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x} \left\langle\vec{k}\middle|\,\vec{x}\right\rangle \left\langle\vec{x}\,\middle|\vec{k}'\right\rangle = (2\pi)^D\delta^{(D)}\left(\vec{k}-\vec{k}'\right), \tag{12.2.7}$$

$$\int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left\langle\vec{x}\,\middle|\vec{k}\right\rangle \left\langle\vec{k}\middle|\,\vec{x}'\right\rangle = \delta^{(D)}\left(\vec{x}-\vec{x}'\right). \tag{12.2.8}$$

Translation symmetry and degeneracy For a fixed $1 \leq j \leq D$, notice the translation operator in the jth Cartesian direction, namely $-i\partial_j \equiv -i\partial/\partial x^j$ commutes with $-\vec{\nabla}^2$. The translation operators commute amongst themselves too. This is why one can simultaneously diagonalize the Laplacian, and all the D translation operators.

$$-i\partial_j \langle\vec{x}|\,k^2\rangle = k_j \langle\vec{x}|\,k^2\rangle \tag{12.2.9}$$

In fact, we see that the eigenvector of the Laplacian $|k^2\rangle$ can be viewed as a tensor product of the eigenstates of P_j.

$$\left|k^2 = \vec{k}^2\right\rangle = |k_1\rangle \otimes |k_2\rangle \otimes \cdots \otimes |k_D\rangle \tag{12.2.10}$$

$$\begin{aligned}\langle\vec{x}|\,k^2\rangle &= \left(\langle x^1| \otimes \dots \otimes \langle x^D|\right)\left(|k_1\rangle \otimes \dots \otimes |k_D\rangle\right) \\ &= \langle x^1|\,k_1\rangle\,\langle x^2|\,k_2\rangle \dots \langle x^D|\,k_D\rangle = \prod_{j=1}^{D} e^{ik_j x^j}. \end{aligned} \tag{12.2.11}$$

As we have already highlighted in the linear algebra of continuous spaces section, the spectrum of the negative Laplacian admits an infinite fold degeneracy here. Physically speaking we may associate it with the translation symmetry of $\mathbb{R}^D$.

12.2.2 1 Dimension

Infinite Flat Space In one dimension, the metric[141] is

$$\mathrm{d}\ell^2 = \mathrm{d}z^2, \tag{12.2.12}$$

for $z \in \mathbb{R}$, and eq. (12.2.6) reduces to

$$-\vec{\nabla}_1^2 \psi_k(z) = -\partial_z^2 \psi_k(z) = k^2 \psi_k(z), \qquad \langle z|\,k\rangle \equiv \psi_k(z) = e^{ikz}; \tag{12.2.13}$$

and their completeness relation (cf. eq. (12.1.26)) is

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} \langle z|\,k\rangle\,\langle k|\,z'\rangle = \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} e^{ik(z-z')} = \delta(z-z'). \tag{12.2.14}$$

Periodic infinite space If the 1D space obeys periodic boundary conditions, with period L, we have instead

$$\begin{gathered} -\vec{\nabla}_1^2 \psi_m(z) = -\partial_z^2 \psi_m(z) = \left(\frac{2\pi m}{L}\right)^2 \psi_m(z), \\ \langle z|\,m\rangle \equiv \psi_m(z) = L^{-1/2} e^{i\frac{2\pi m}{L}z}, \qquad m = 0, \pm 1, \pm 2, \dots . \end{gathered} \tag{12.2.15}$$

The orthonormal eigenvectors obey

$$\int_0^L \mathrm{d}z\,\langle m|\,z\rangle\,\langle z|\,m'\rangle = \delta^m_{m'}, \qquad \langle z|\,m\rangle = L^{-1/2} e^{i\frac{2\pi m}{L}z}; \tag{12.2.16}$$

while their completeness relation (eq. (12.1.26)) reads, for $0 \leq z, z' \leq L$,

$$\sum_{m=-\infty}^{\infty} \langle z|\,m\rangle\,\langle m|\,z'\rangle = \frac{1}{L} \sum_{m=-\infty}^{\infty} e^{\frac{2\pi m}{L} i(z-z')} = \delta(z-z'). \tag{12.2.17}$$

Unit Circle A periodic infinite space can be thought of as a circle, and vice versa. Simply identify $L \equiv 2\pi r$, where r is the radius of the circle as embedded in 2D space. For concreteness we will consider a circle of radius 1. Then we may write the metric as

$$\mathrm{d}\ell^2 = (\mathrm{d}\phi)^2, \qquad \phi \in [0, 2\pi). \tag{12.2.18}$$

[141] One dimensional space(time)s are always flat – the Riemann tensor is identically zero.

We may then bring over the results from the previous discussion.

$$-\vec{\nabla}^2_{\mathbb{S}^1}\psi_m(\phi) = -\partial_\phi^2\psi_m(\phi) = m^2\psi_m(\phi),$$
$$\langle\phi|\,m\rangle \equiv \psi_m(\phi) = (2\pi)^{-1/2}e^{im\phi}, \qquad m = 0, \pm 1, \pm 2, \dots. \tag{12.2.19}$$

The orthonormal eigenvectors obey

$$\int_0^{2\pi} \mathrm{d}\phi\, \langle m|\,\phi\rangle\, \langle\phi|\,m'\rangle = \delta^m_{m'}, \qquad \langle\phi|\,m\rangle = (2\pi)^{-1/2}e^{im\phi}. \tag{12.2.20}$$

while their completeness relation reads, for $0 \le z, z' \le L$,

$$\sum_{m=-\infty}^{\infty} \langle\phi|\,m\rangle\, \langle m|\,\phi'\rangle = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im(\phi-\phi')} = \delta(\phi-\phi'). \tag{12.2.21}$$

Fourier series re-visited. Note that $-i\partial_\phi$ can be thought of as the "momentum operator" on the unit circle (in the position representation) with eigenvalues $\{m\}$ and corresponding eigenvectors $\{\langle\phi|\,m\rangle\}$. Namely, if we define

$$\langle\phi\,|P_\phi|\,\psi\rangle = -i\partial_\phi\, \langle\phi|\,\psi\rangle \tag{12.2.22}$$

for any state $|\psi\rangle$, we shall see it is Hermitian with discrete spectra:

$$P_\phi\,|m\rangle = m\,|m\rangle\,, \qquad m = 0, \pm 1, \pm 2, \pm 3, \dots, \tag{12.2.23}$$
$$\langle\phi|\,m\rangle = e^{im\phi}/\sqrt{2\pi}. \tag{12.2.24}$$

Given arbitrary states $|\psi_{1,2}\rangle$,

$$\begin{aligned}\langle\psi_1\,|P_\phi|\,\psi_2\rangle &= \int_0^{2\pi} \mathrm{d}\phi\, \langle\psi_1|\,\phi\rangle\, (-i\partial_\phi\, \langle\phi|\,\psi_2\rangle) \qquad (12.2.25)\\ &= \left[-i\, \langle\psi_1|\,\phi\rangle\, \langle\phi|\,\psi_2\rangle\right]_{\phi=0}^{\phi=2\pi} + \int_0^{2\pi} \mathrm{d}\phi\, (i\partial_\phi\, \langle\psi_1|\,\phi\rangle)\, \langle\phi|\,\psi_2\rangle\,.\end{aligned}$$

As long as we are dealing with the space of *continuous* functions $\psi_{1,2}(\phi)$ on a circle, the boundary terms must vanish because $\phi = 0$ and $\phi = 2\pi$ really refer to the same point. Therefore,

$$\begin{aligned}\langle\psi_1\,|P_\phi|\,\psi_2\rangle &= \int_0^{2\pi} \mathrm{d}\phi\, (-i\partial_\phi\, \langle\phi|\,\psi_1\rangle)^*\, \langle\phi|\,\psi_2\rangle = \int_0^{2\pi} \mathrm{d}\phi\overline{\langle\phi\,|P_\phi|\,\psi_1\rangle}\, \langle\phi|\,\psi_2\rangle\\ &= \int_0^{2\pi} \mathrm{d}\phi\, \left\langle\psi_1\,\middle|P_\phi^\dagger\middle|\,\phi\right\rangle\, \langle\phi|\,\psi_2\rangle = \left\langle\psi_1\,\middle|P_\phi^\dagger\middle|\,\psi_2\right\rangle. \qquad (12.2.26)\end{aligned}$$

We must therefore have

$$\left\langle\phi\,\middle|e^{-i\theta P_\phi}\middle|\,\psi\right\rangle = e^{-i\theta(-i\partial_\phi)}\, \langle\phi|\,\psi\rangle = e^{-\theta\partial_\phi}\, \langle\phi|\,\psi\rangle = \langle\phi-\theta|\,\psi\rangle\,. \tag{12.2.27}$$

Any function on a circle can be expanded in the eigenstates of P_ϕ, which in turn can be expressed through its position representation.

$$\begin{aligned}|\psi\rangle &= \sum_{m=-\infty}^{+\infty} |m\rangle\, \langle m|\,\psi\rangle = \sum_{m=-\infty}^{+\infty}\int_0^{2\pi} \mathrm{d}\phi\,|\phi\rangle\, \langle\phi|\,m\rangle\, \langle m|\,\psi\rangle = \sum_{m=-\infty}^{+\infty}\int_0^{2\pi} \frac{\mathrm{d}\phi}{\sqrt{2\pi}}\,|\phi\rangle\, \langle m|\,\psi\rangle\, e^{im\phi},\\ \langle m|\,\psi\rangle &= \int_0^{2\pi} \mathrm{d}\phi'\, \langle m|\,\phi'\rangle\, \langle\phi'|\,\psi\rangle = \int_0^{2\pi} \frac{\mathrm{d}\phi'}{\sqrt{2\pi}} e^{-im\phi'}\psi(\phi'). \qquad (12.2.28)\end{aligned}$$

This is nothing but the Fourier series expansion of $\psi(\phi)$.

12.2.3 2 Dimensions ◦ Separation-of-Variables for PDEs

Flat Space, Cylindrical Coordinates The 2D flat metric in cylindrical coordinates reads

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\phi^2, \qquad r \geq 0, \qquad \phi \in [0, 2\pi), \qquad \sqrt{|g|} = r. \tag{12.2.29}$$

The negative Laplacian is therefore

$$-\vec{\nabla}_2^2 \varphi_k(r,\phi) = -\frac{1}{r}\left(\partial_r\left(r\partial_r\varphi_k\right) + \frac{1}{r}\partial_\phi^2\varphi_k\right) \tag{12.2.30}$$

$$= -\left\{\frac{1}{r}\partial_r\left(r\partial_r\varphi_k\right) + \frac{1}{r^2}\partial_\phi^2\varphi_k\right\}. \tag{12.2.31}$$

Our goal here is to diagonalize the negative Laplacian in cylindrical coordinates, and re-write the plane wave using its eigenstates. In this case we will in fact tackle the latter and use the results to do the former. To begin, note that the plane wave in 2D cylindrical coordinates is

$$\langle\vec{x}|\vec{k}\rangle = \exp(i\vec{k}\cdot\vec{x}) = \exp(ikr\cos(\phi - \phi_k)), \qquad k \equiv |\vec{k}|,\ r \equiv |\vec{x}|; \tag{12.2.32}$$

because the Cartesian components of $\vec{k}$ and $\vec{x}$ are

$$k_i = k\left(\cos\phi_k, \sin\phi_k\right) \qquad x^i = r\left(\cos\phi, \sin\phi\right). \tag{12.2.33}$$

We observe that this is a periodic function of the angle $\Delta\phi \equiv \phi - \phi_k$ with period $L = 2\pi$, which means it must admit a Fourier series expansion. Referring to equations (5.3.31) and (5.3.32),

$$\langle\vec{x}|\vec{k}\rangle = \sum_{m=-\infty}^{+\infty} \chi_m(kr)\frac{e^{im(\phi-\phi_k)}}{\sqrt{2\pi}}. \tag{12.2.34}$$

Setting $\phi - \phi_k \to \phi''$, multiplying both sides with $\exp(-im\phi'')/\sqrt{2\pi}$, followed by integrating ϕ'' over the unit circle,

$$\chi_m(kr) = \int_0^{2\pi}\frac{\mathrm{d}\phi''}{\sqrt{2\pi}}e^{ikr\cos\phi''}e^{-im\phi''} \tag{12.2.35}$$

$$= \sqrt{2\pi}\int_{\phi''=0}^{\phi''=2\pi}\frac{\mathrm{d}(\phi''+\pi/2)}{2\pi}e^{ikr\cos(\phi''+\pi/2-\pi/2)}e^{-im(\phi''+\pi/2-\pi/2)}$$

$$= \sqrt{2\pi}\int_{\pi/2}^{5\pi/2}\frac{\mathrm{d}\phi'}{2\pi}e^{ikr\sin\phi'}e^{-im\phi'}i^m = i^m\sqrt{2\pi}\int_{-\pi}^{+\pi}\frac{\mathrm{d}\phi'}{2\pi}e^{ikr\sin\phi'}e^{-im\phi'}. \tag{12.2.36}$$

(In the last line, we have used the fact that the integrand is itself a periodic function of ϕ' with period 2π to change the limits of integration.) As it turns out, the Bessel function J_m admits an integral representation (cf. eq. (10.9.2) of the NIST page here.)

$$J_m(z) = \int_{-\pi}^{\pi}\frac{\mathrm{d}\phi'}{2\pi}e^{iz\sin\phi' - im\phi'}, \qquad m \in \{0, \pm1, \pm2, \dots\}, \tag{12.2.37}$$

$$J_{-m}(z) = (-)^m J_m(z). \tag{12.2.38}$$

As an aside, let us record that $J_\nu(z)$ also has a series representation

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-)^k (z/2)^{2k}}{k!\Gamma(\nu+k+1)}; \tag{12.2.39}$$

and the large argument asymptotic expansion

$$J_{\pm\nu}(z \gg \nu) \sim \sqrt{\frac{2}{\pi z}} \cos\left(z \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right). \tag{12.2.40}$$

Utilizing eq. (5.4.141) in eq. (12.2.36), we see the plane wave in eq. (12.2.34) admits the cylindrical coordinate expansion:

$$\begin{aligned} \langle \vec{x}|\vec{k}\rangle &= \exp(i\vec{k}\cdot\vec{x}) = \exp(ikr\cos(\phi-\phi_k)), \qquad k \equiv |\vec{k}|, \; r \equiv |\vec{x}| \\ &= \sum_{m=-\infty}^{\infty} i^m J_m(kr) e^{im(\phi-\phi_k)}. \end{aligned} \tag{12.2.41}$$

Because the $\{e^{im\phi}\}$ are basis vectors on the circle of fixed radius r, every term in the infinite sum is a linearly independent eigenvector of $-\vec{\nabla}_2^2$. That is, we can now read off the basis eigenvectors of the negative Laplacian in 2D cylindrical coordinates. To obtain orthonormal ones, however, let us calculate their normalization using the following orthogonality relation, written in cylindrical coordinates,

$$\begin{aligned} (2\pi)^2 \frac{\delta(k-k')\delta(\phi_k-\phi_{k'})}{\sqrt{kk'}} &= \int_{\mathbb{R}^2} \mathrm{d}^2x \exp(i(\vec{k}-\vec{k}')\cdot\vec{x}) \\ &= \sum_{m,m'=-\infty}^{+\infty} \int_0^\infty \mathrm{d}r \cdot r \int_0^{2\pi} \mathrm{d}\phi \cdot i^m (-i)^{m'} J_m(kr) J_{m'}(k'r) e^{im(\phi-\phi_k)} e^{-im'(\phi-\phi_{k'})} \\ &= (2\pi) \sum_{m=-\infty}^{+\infty} \int_0^\infty \mathrm{d}r \cdot r J_m(kr) J_m(k'r) e^{im(\phi_{k'}-\phi_k)}. \end{aligned} \tag{12.2.42}$$

Problem 12.8. The left hand side of eq. (12.2.42) is $(2\pi)^2\delta^{(2)}(\vec{k}-\vec{k}')$ if we used Cartesian coordinates in $\vec{k}-$space – see eq. (12.2.7). Can you explain why it takes the form it does? Hint: Use cylindrical coordinates in $k-$space and refer to eq. (12.1.26). □

We now replace the $\delta(\phi-\phi_k)$ on the left hand side of eq. (12.2.42) with the completeness relation in eq. (12.2.17), where now $z = \phi_k$, $z' = \phi_{k'}$ and the period is $L = 2\pi$. Equating the result to the last line then brings us to

$$\sum_{m=-\infty}^{+\infty} \frac{\delta(k-k')}{\sqrt{kk'}} e^{im(\phi_k-\phi_{k'})} = \sum_{m=-\infty}^{+\infty} \int_0^\infty \mathrm{d}r \cdot r J_m(kr) J_m(k'r) e^{im(\phi_{k'}-\phi_k)}. \tag{12.2.43}$$

The coefficients of each (linearly independent) vector $e^{im(\phi_k-\phi_{k'})}$ on both sides should be the same. This yields the completeness relation of the radial mode functions:

$$\int_0^\infty \mathrm{d}r \cdot r J_m(kr) J_m(k'r) = \frac{\delta(k-k')}{\sqrt{kk'}}, \tag{12.2.44}$$

$$\int_0^\infty \mathrm{d}k \cdot k J_m(kr) J_m(kr') = \frac{\delta(r-r')}{\sqrt{rr'}}. \tag{12.2.45}$$

To summarize, we have found, in 2D infinite flat space, that the eigenvectors/values of the negative Laplacian in cylindrical coordinates ($r \geq 0, 0 \leq \phi < 2\pi$) are

$$-\vec{\nabla}_2^2 \langle r, \phi | k, m \rangle = k^2 \langle r, \phi | k, m \rangle, \qquad \langle r, \phi | k, m \rangle \equiv J_m(kr) \frac{\exp(im\phi)}{\sqrt{2\pi}},$$
$$m = 0, \pm 1, \pm 2, \pm 3, \dots. \tag{12.2.46}$$

The eigenvectors are normalized as

$$\int_0^\infty \mathrm{d}r \cdot r \int_0^{2\pi} \mathrm{d}\phi \langle k, m | r, \phi \rangle \langle r, \phi | k', m' \rangle = \delta^m_{m'} \frac{\delta(k-k')}{\sqrt{kk'}}. \tag{12.2.47}$$

Rotational symmetry and degeneracy Note that $-i\partial_\phi$ is the translation operator in the azimuthal direction ($\equiv$ rotation operator), with eigenvalue m. The spectrum here is discretely and infinitely degenerate, which can be physically interpreted to be due to the presence of rotational symmetry.

Bessel's equation As a check of our analysis here, we may now directly evaluate the 2D negative Laplacian acting on the its eigenvector $\langle r, \phi | k, m \rangle$, and see that we are lead to Bessel's equation. Starting from the eigenvector/value equation in (12.2.46), followed by using the explicit expression in eq. (12.2.30) and the angular eigenvalue/vector equation $\partial_\phi^2 \exp(im\phi) = -m^2 \exp(im\phi)$, this hands us

$$k^2 J_m(kr) = -\left\{ \frac{1}{r} \partial_r \left(r \partial_r J_m(kr) \right) - \frac{m^2}{r^2} J_m(kr) \right\}. \tag{12.2.48}$$

Let us then re-scale $\rho \equiv kr$, where $k \equiv |\vec{k}|$, so that $\partial_r = k\partial_\rho$.

$$\rho^2 \cdot J''(\rho) + \rho \cdot J'(\rho) + (\rho^2 - m^2) J(\rho) = 0 \tag{12.2.49}$$

Equation 10.2.1 of the NIST page here tells us we have indeed arrived at Bessel's equation. Two linearly independent solutions are $J_m(kr)$ and $Y_m(kr)$. However, eq. (10.2.2) of the NIST page here and eq. (10.8.1) of the NIST page here tell us, for small argument, $Y_m(z \to 0)$ has at least a log singularity of the form $\ln(z/2)$ and for $m \neq 0$ has also a power law singularity that goes as $1/z^{|m|}$. Whereas, $J_m(z)$ is $(z/2)^{|m|}$ times a power series in the variable $(z/2)^2$, and is not only smooth for small z, the power series in fact has an infinite radius of convergence. It makes sense that our plane wave expansion only contains J_m and not Y_m because it is smooth for all r.

Problem 12.9. 2D wedge Explain how you would modify the analysis here, if we were not dealing with an infinite 2D space, but only a wedge of 2D space – namely, $r \geq 0$ but $0 \leq \phi \leq \phi_0 < 2\pi$. How would you modify the analysis here, if $\phi \in [0, 2\pi)$, but now $0 \leq r \leq r_0 < \infty$? You do not need to carry out the calculations in full, but try to be as detailed as you can. Assume Dirichlet boundary conditions. □

2-sphere $\mathbb{S}^2$, Separation-Of-Variables, and the Spherical Harmonics[142] The 2-sphere of radius R can be viewed as a curved surface embedded in 3D flat space parametrized as

$$\vec{x}(\vec{\xi} = (\theta, \phi)) = R\left(\sin\theta\ \cos\phi, \sin\theta\ \sin\phi, \cos\theta\right), \qquad \vec{x}^2 = R^2. \tag{12.2.50}$$

For concreteness we will consider the case where $R = 1$. Its metric is therefore given by

$$H_{\mathrm{IJ}} \mathrm{d}\xi^{\mathrm{I}} \mathrm{d}\xi^{\mathrm{J}} = \delta_{ij} \mathrm{d}x^i \mathrm{d}x^j\big|_{R=1} = \delta_{ij} \partial_{\mathrm{I}} x^i \partial_{\mathrm{J}} x^j \mathrm{d}\xi^{\mathrm{I}} \mathrm{d}\xi^{\mathrm{J}}, \tag{12.2.51}$$

$$= \mathrm{d}\theta^2 + (\sin\theta)^2 \mathrm{d}\phi^2, \qquad \sqrt{|H|} = \sin\theta. \tag{12.2.52}$$

(Or, simply take the 3D flat space metric in spherical coordinates, and set $\mathrm{d}r \to 0$ and $r \to 1$.)

We wish to diagonalize the negative Laplacian on this unit radius 2−sphere. The relevant eigenvector/value equation is

$$-\vec{\nabla}^2_{\mathbb{S}^2} Y(\theta, \phi) = \nu(\nu + 1) Y(\theta, \phi), \tag{12.2.53}$$

where for now ν is some arbitrary real number greater or equal to 0 so that $\nu(\nu+1)$ itself can be equal to any non-negative number. We have chosen the form $\nu(\nu + 1)$ for technical convenience – as we shall see, ν is actually 0 or a positive integer, with its discrete nature due to the finite area of the 2−sphere.

To do so, we now turn to the *separation of variables* technique, which is a method to reduce a PDE into a bunch of ODEs – and hence more manageable. The main idea is, for highly symmetric problems such as the Laplacian in flat space(time)s or on the D-sphere, one postulates that a multi-variable eigenfunction factorizes into a product of functions, each depending only on one variable. The crux of the method then involves re-arranging the ensuing eigenvector equation into sums of terms, $\sum_i \tau_i = 0$, such that each τ_i depends solely on the ith variable of the system. Once this has been done – and since no other term now depends on the ith coordinate so we may vary it without varying others – we may then conclude that each τ_i has to be a constant because upon varying this ith term the entire sum must still remain zero. This in turn leads us to one ODE for every τ_i. If solutions can be found, we are assured that such an ansatz works.

For the unit radius 2−sphere, we postulate

$$Y(\theta, \phi) = \Lambda(\theta)\Phi(\phi). \tag{12.2.54}$$

First work out the Laplacian explicitly, with $s \equiv \sin\theta$,

$$-\left\{\frac{1}{s}\partial_\theta\left(s\partial_\theta Y\right) + \frac{1}{s^2}\partial_\phi^2 Y\right\} = -\left\{\frac{1}{s}\partial_\theta\left(s\partial_\theta Y\right) + \frac{1}{s^2}\vec{\nabla}^2_{\mathbb{S}^1} Y\right\} = \nu(\nu+1)Y(\theta, \phi). \tag{12.2.55}$$

We have identified $\vec{\nabla}^2_{\mathbb{S}^1} = \partial_\phi^2$ to be the Laplacian on the circle, from eq. (12.2.19). To reiterate, the key step in the separation-of-variables technique is to arrange the eigenvalue equation into sums of individual terms that depend on only one variable at a time. In the case at hand, let us

[142] In these notes we focus solely on the spherical harmonics on $\mathbb{S}^2$; for spherical harmonics in arbitrary dimensions, see arXiv:1205.3548.

multiply the above equation throughout by s^2, use the ansatz in eq. (12.2.54), and re-arrange it into:

$$\left\{s\partial_\theta\left(s\partial_\theta\Lambda\cdot\Phi\right)+s^2\nu(\nu+1)\Lambda\cdot\Phi\right\}+\partial_\phi^2(\Lambda\cdot\Phi)=0, \tag{12.2.56}$$

$$\frac{1}{\Lambda}\left\{s\partial_\theta\left(s\partial_\theta\Lambda\right)+s^2\nu(\nu+1)\Lambda\right\}+\frac{\partial_\phi^2\Phi}{\Phi}=0. \tag{12.2.57}$$

Notice the first term involving the $\{\dots\}$ depends only on θ and not on ϕ. Whereas the second term $(\partial^2\Phi)/\Phi$ only depends on ϕ and not on θ. This immediately implies both terms must be a constant. For, we may first differentiate both sides with respect to θ,

$$\partial_\theta\left\{\frac{1}{\Lambda}\Big(s\partial_\theta\left(s\partial_\theta\Lambda\right)+s^2\nu(\nu+1)\Lambda\Big)\right\}=0 \tag{12.2.58}$$

and conclude the terms in the curly brackets must be independent of θ. And since they are already independent of ϕ by assumption, these terms must be a constant. Similarly, differentiating eq. (12.2.57) with respect to ϕ,

$$\partial_\phi\left\{\frac{\partial_\phi^2\Phi}{\Phi}\right\}=0. \tag{12.2.59}$$

At this point, we deduce

$$\frac{1}{\Lambda}\left\{s\partial_\theta\left(s\partial_\theta\Lambda\right)+s^2\nu(\nu+1)\Lambda\right\}=m^2, \tag{12.2.60}$$

$$\frac{\partial_\phi^2\Phi}{\Phi}=-m^2. \tag{12.2.61}$$

Note the relative $-$ sign on the right hand sides of equations (12.2.60) and (12.2.61): this ensures their sum in eq. (12.2.57) is zero. At this point, m^2 is an arbitrary constant, but we may see that eq. (12.2.61) is nothing but the simple harmonic oscillator equation: $\partial_\phi^2\Phi+m^2\Phi=0$, whose solutions are $\Phi\propto\exp(im\phi)$. Demanding that $\Phi(\phi+2\pi)=\Phi(\phi)$ we obtain

$$\Phi(\phi)\propto\exp(im\phi),\qquad m=0,\pm1,\pm2,\dots \tag{12.2.62}$$

Notice this amounts to setting Φ to be the eigenvector of $\vec{\nabla}^2_{\mathbb{S}^1}$, which we could have guessed from the outset, since the only occurrence of ∂_ϕ in the 2-sphere Laplacian is in the $\partial_\phi^2\Phi$ term.

Moreover, it will turn out to be very useful to change variables to $c\equiv\cos\theta$, which runs from -1 to $+1$ over the range $0\leq\theta\leq\pi$. Since $s\equiv\sin\theta$ is strictly positive there, we have the positive root $s_\theta=(1-c^2)^{1/2}$ and $\partial_\theta=(\partial c/\partial\theta)\partial_c=-\sin\theta\partial_c=-(1-c^2)^{1/2}\partial_c$. Eq. (12.2.60) then reads

$$\partial_c\left((1-c^2)\partial_c\Lambda\right)+\left(\nu(\nu+1)-\frac{m^2}{1-c^2}\right)\Lambda=0. \tag{12.2.63}$$

This is solved – see eq. 14.2.2 of the NIST page here – by the two associated Legendre functions $P_\nu^m(c)$ and $Q_\nu^m(c)$. It turns out, to obtain a solution that does not blow up over the entire range $-1\leq c\leq+1$, we need to choose $P_\nu^m(c)$, set $\nu\equiv\ell$ to be 0 or a positive integer, and have m run from $-\ell$ to ℓ.

$$\Lambda\propto P_\ell^m(\cos\theta),\qquad \ell\in\{0,1,2,3,\dots\},\ m\in\{-\ell,-\ell+1,\dots.\ell-1,\ell\}. \tag{12.2.64}$$

Note that

$$P_\ell^0(x) = P_\ell(x), \tag{12.2.65}$$

where $P_\ell(x)$ is the ℓth Legendre polynomial. A common phase convention that yields an orthonormal basis set of functions on the 2−sphere is the following definition for the *spherical harmonics*

$$\begin{aligned} -\vec{\nabla}^2_{\mathbb{S}^2} Y_\ell^m(\theta,\phi) &= \ell(\ell+1) Y_\ell^m(\theta,\phi), \\ \langle \theta,\phi | \, \ell, m\rangle = Y_\ell^m(\theta,\phi) &= \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) e^{im\phi}, \\ \ell \in \{0,1,2,3,\dots\}, \ m &\in \{-\ell, -\ell+1, \dots . \ell-1, \ell\}. \end{aligned} \tag{12.2.66}$$

Spherical harmonics should be viewed as "waves" on the 2−sphere, with larger ℓ modes describing the higher frequency/shorter wavelength/finer features of the state/function on the sphere. Let us examine the spherical harmonics from $\ell = 0, 1, 2, 3$. The $\ell = 0$ spherical harmonic is a constant.

$$Y_0^0 = \frac{1}{\sqrt{4\pi}} \tag{12.2.67}$$

The $\ell = 1$ spherical harmonics are:

$$Y_1^{-1} = \frac{1}{2}\sqrt{\frac{3}{2\pi}} e^{-i\phi}\sin(\theta), \qquad Y_1^0 = \frac{1}{2}\sqrt{\frac{3}{\pi}}\cos(\theta), \qquad Y_1^1 = -\frac{1}{2}\sqrt{\frac{3}{2\pi}} e^{i\phi}\sin(\theta). \tag{12.2.68}$$

The $\ell = 2$ spherical harmonics are:

$$\begin{aligned} Y_2^{-2} &= \frac{1}{4}\sqrt{\frac{15}{2\pi}} e^{-2i\phi}\sin^2(\theta), \qquad Y_2^{-1} = \frac{1}{2}\sqrt{\frac{15}{2\pi}} e^{-i\phi}\sin(\theta)\cos(\theta), \qquad Y_2^0 = \frac{1}{4}\sqrt{\frac{5}{\pi}}\left(3\cos^2(\theta) - 1\right), \\ Y_2^1 &= -\frac{1}{2}\sqrt{\frac{15}{2\pi}} e^{i\phi}\sin(\theta)\cos(\theta), \qquad Y_2^2 = \frac{1}{4}\sqrt{\frac{15}{2\pi}} e^{2i\phi}\sin^2(\theta). \end{aligned} \tag{12.2.69}$$

The $\ell = 3$ spherical harmonics are:

$$\begin{aligned} Y_3^{-3} &= \frac{1}{8}\sqrt{\frac{35}{\pi}} e^{-3i\phi}\sin^3(\theta), \qquad Y_3^{-2} = \frac{1}{4}\sqrt{\frac{105}{2\pi}} e^{-2i\phi}\sin^2(\theta)\cos(\theta), \\ Y_3^{-1} &= \frac{1}{8}\sqrt{\frac{21}{\pi}} e^{-i\phi}\sin(\theta)\left(5\cos^2(\theta) - 1\right), \qquad Y_3^0 = \frac{1}{4}\sqrt{\frac{7}{\pi}}\left(5\cos^3(\theta) - 3\cos(\theta)\right), \\ Y_3^1 &= -\frac{1}{8}\sqrt{\frac{21}{\pi}} e^{i\phi}\sin(\theta)\left(5\cos^2(\theta) - 1\right), \qquad Y_3^2 = \frac{1}{4}\sqrt{\frac{105}{2\pi}} e^{2i\phi}\sin^2(\theta)\cos(\theta), \\ Y_3^3 &= -\frac{1}{8}\sqrt{\frac{35}{\pi}} e^{3i\phi}\sin^3(\theta). \end{aligned} \tag{12.2.70}$$

For later purposes, note that the $m = 0$ case removes any dependence on the azimuthal angle ϕ, and in fact returns the Legendre polynomial.

$$\langle \theta,\phi | \, \ell, m = 0\rangle = Y_\ell^0(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}} P_\ell(\cos\theta). \tag{12.2.71}$$

Orthonormality and completeness of the spherical harmonics read, respectively,

$$\begin{aligned}\langle \ell', m' | \, \ell, m \rangle &= \int_{\mathbb{S}^2} \mathrm{d}^2\vec{\xi} \sqrt{|H|} \; \overline{Y_{\ell'}^{m'}(\theta,\phi)} Y_\ell^m(\theta,\phi) \\ &= \int_{-1}^{+1} \mathrm{d}(\cos\theta) \int_0^{2\pi} \mathrm{d}\phi \overline{Y_{\ell'}^{m'}}(\theta,\phi) Y_\ell^m(\theta,\phi) = \delta_\ell^{\ell'} \delta_m^{m'}, \end{aligned} \tag{12.2.72}$$

and

$$\begin{aligned}\langle \theta', \phi' | \, \theta, \phi \rangle &= \frac{\delta(\theta'-\theta)\delta(\phi-\phi')}{\sqrt{\sin(\theta)\sin(\theta')}} = \delta\left(\cos(\theta') - \cos(\theta)\right)\delta(\phi - \phi') \\ &= \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \overline{Y_\ell^m(\theta',\phi')} Y_\ell^m(\theta,\phi). \end{aligned} \tag{12.2.73}$$

In 3D flat space, let us write the Cartesian components of the momentum vector $\vec{k}$ and the position vector $\vec{x}$ in spherical coordinates.

$$k_i = k\left(\sin\theta_k \cdot \cos\phi_k, \sin\theta_k \cdot \sin\phi_k, \cos\theta_k\right) \equiv k\widehat{k} \tag{12.2.74}$$

$$x^i = r\left(\sin\theta \cdot \cos\phi, \sin\theta \cdot \sin\phi, \cos\theta\right) \equiv r\widehat{x} \tag{12.2.75}$$

Addition formula In terms of these variables we may write down a useful identity involving the spherical harmonics and the Legendre polynomial, usually known as the addition formula.

$$P_\ell\left(\widehat{k} \cdot \widehat{x}\right) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{+\ell} \overline{Y_\ell^m(\theta,\phi)} Y_\ell^m(\theta_k,\phi_k) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta,\phi) \overline{Y_\ell^m(\theta_k,\phi_k)}, \tag{12.2.76}$$

where $\widehat{k} \equiv \vec{k}/k$ and $\widehat{x} \equiv \vec{x}/r$. The second equality follows from the first because the Legendre polynomial is real.

For a fixed direction $\widehat{k}$, note that $P_\ell(\widehat{k} \cdot \widehat{x})$ in eq. (12.2.76) is an eigenvector of the negative Laplacian on the 2−sphere. For, as we have already noted, the eigenvalue equation $-\vec{\nabla}^2\psi = \lambda\psi$ is a coordinate scalar. In particular, we may choose coordinates such that $\widehat{k}$ is pointing 'North', so that $\widehat{k} \cdot \widehat{x} = \cos\theta$, where θ is the usual altitude angle. By recalling eq. (12.2.71), we see therefore,

$$-\vec{\nabla}^2_{\vec{x},\mathbb{S}^2} P_\ell\left(\widehat{k} \cdot \widehat{x}\right) = \ell(\ell+1) P_\ell\left(\widehat{k} \cdot \widehat{x}\right). \tag{12.2.77}$$

Since $P_\ell(\widehat{k} \cdot \widehat{x})$ is symmetric under the swap $k \leftrightarrow x$, it must also be an eigenvector of the Laplacian with respect to $\vec{k}$,

$$-\vec{\nabla}^2_{\vec{k},\mathbb{S}^2} P_\ell\left(\widehat{k} \cdot \widehat{x}\right) = \ell(\ell+1) P_\ell\left(\widehat{k} \cdot \widehat{x}\right). \tag{12.2.78}$$

Complex conjugation Under complex conjugation, the spherical harmonics obey

$$\overline{Y_\ell^m(\theta,\phi)} = (-)^m Y_\ell^{-m}(\theta,\phi). \tag{12.2.79}$$

Parity Under a parity flip, meaning if you compare Y_ℓ^m evaluated at the point (θ, ϕ) to the point on the opposite side of the sphere $(\pi - \theta, \phi + \pi)$, we have the relation

$$Y_\ell^m(\pi - \theta, \phi + \pi) = (-)^\ell Y_\ell^m(\theta, \phi). \tag{12.2.80}$$

The odd ℓ spherical harmonics are thus odd under parity; whereas the even ℓ ones are invariant (i.e., even) under parity. That the Laplacian on the sphere $\vec{\nabla}^2_{\mathbb{S}^2}$ and the parity operator P share a common set of eigenvectors is because they commute: $[P, \vec{\nabla}^2_{\mathbb{S}^2}] = 0$.

Poisson Equation on the 2-sphere Having acquired some familiarity of the spherical harmonics, we can now tackle Poisson's equation

$$-\vec{\nabla}^2_{\mathbb{S}^2}\psi(\theta, \phi) = J(\theta, \phi) \tag{12.2.81}$$

on the 2−sphere. Because the spherical harmonics are complete on the sphere, we may expand both ψ and J in terms of them.

$$\psi = \sum_{\ell,m} A_\ell^m Y_\ell^m, \qquad J = \sum_{\ell,m} B_\ell^m Y_\ell^m. \tag{12.2.82}$$

(This means, if J is a given function, then we may calculate $B_\ell^m = \int_{\mathbb{S}^2} \mathrm{d}^2\Omega \overline{Y_\ell^m(\theta, \phi)} J(\theta, \phi)$.) Inserting these expansions into eq. (12.2.81), and recalling the eigenvalue equation $-\vec{\nabla}^2_{\mathbb{S}^2} Y_\ell^m = \ell(\ell+1)Y_\ell^m$,

$$\sum_{\ell \neq 0, m} \ell(\ell+1) A_\ell^m Y_\ell^m = \sum_{\ell,m} B_\ell^m Y_\ell^m. \tag{12.2.83}$$

On the left hand side, because the eigenvalue of Y_0^0 is zero, there is no longer any $\ell = 0$ term. Therefore, we see that for there to be a consistent solution, J itself cannot contain a $\ell = 0$ term. (This is intimately related to the fact that the sphere has no boundaries.[143]) At this point, we may then equate the $\ell > 0$ coefficients of the spherical harmonics on both sides, and deduce

$$A_\ell^m = \frac{B_\ell^m}{\ell(\ell+1)}, \qquad \ell > 0. \tag{12.2.84}$$

To summarize, given a $J(\theta, \phi)$ that has no "zero mode," such that it can be decomposed as

$$J(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} B_\ell^m Y_\ell^m(\theta, \phi) \quad \Leftrightarrow \quad B_\ell^m = \int_{-1}^{+1} \mathrm{d}(\cos\theta) \int_0^{2\pi} \mathrm{d}\phi \overline{Y_\ell^m(\theta, \phi)} J(\theta, \phi), \tag{12.2.85}$$

the solution to (12.2.81) is

$$\psi(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{+\ell} \frac{B_\ell^m}{\ell(\ell+1)} Y_\ell^m(\theta, \phi). \tag{12.2.86}$$

Problem 12.10. Eigensystem of 2D Laplacian Diagonalize the Laplacian in 2D flat space in cylindrical coodrdinates – i.e., obtain the results in eq. (12.2.46) – using the separation-of-variables technique. Hints: What is the boundary condition in the ϕ direction? For the radial function, consider the appropriate boundary conditions at $r = 0$; you may need to refer to here, here, and here. □

[143] For, suppose there is a solution to $-\vec{\nabla}^2\psi = \chi/(4\pi)$, where χ is a constant. Let us now integrate both sides over the sphere's surface, and apply the Gauss/Stokes' theorem. On the left hand side we get zero because the sphere has no boundaries. On the right hand side we have χ. This inconsistency means no such solution exist.

12.2.4 3 Dimensions

Infinite Flat Space, Cylindrical Coordinates We now turn to 3D flat space, written in cylindrical coordinates,

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2 + \mathrm{d}z^2, \qquad r \geq 0,\ \phi \in [0, 2\pi),\ z \in \mathbb{R}, \qquad \sqrt{|g|} = r. \tag{12.2.87}$$

Because the negative Laplacian on a scalar is the sum of the 1D and the 2D cylindrical case,

$$-\vec{\nabla}_3^2\psi = -\vec{\nabla}_2^2\psi - \partial_z^2\psi, \tag{12.2.88}$$

we may try the separation-of-variables ansatz involving the product of the eigenvectors of the respective Laplacians.

$$\psi(r, \phi, z) = \psi_2(r, \phi)\psi_1(z), \qquad \psi_2(r, \phi) \equiv J_m(kr)\frac{e^{im\phi}}{\sqrt{2\pi}}, \qquad \psi_1(z) \equiv e^{ik_z z}. \tag{12.2.89}$$

This yields

$$-\vec{\nabla}^2\psi = -\psi_1\vec{\nabla}_2^2\psi_2 - \psi_2\partial_z^2\psi_1 = (k^2 + (k_z)^2)\psi, \tag{12.2.90}$$

To sum, the orthonormal eigenfunctions are

$$\langle r, \phi, z |\, k, m, k_z\rangle = J_m(kr)\frac{e^{im\phi}}{\sqrt{2\pi}}e^{ik_z z} \tag{12.2.91}$$

$$\int_0^{2\pi}\mathrm{d}\phi\int_0^{\infty}\mathrm{d}rr\int_{-\infty}^{+\infty}\mathrm{d}z\,\langle k', m', k_z' |\, r, \phi, z\rangle\,\langle r, \phi, z |\, k, m, k_z\rangle = \delta_m^{m'}\frac{\delta(k-k')}{\sqrt{kk'}}\cdot(2\pi)\delta(k_z'-k_z). \tag{12.2.92}$$

Since we already figured out the 2D plane wave expansion in cylindrical coordinates in eq. (12.2.41), and since the 3D plane wave is simply the 2D one multiplied by the plane wave in the z direction, i.e., $\exp(i\vec{k}\cdot\vec{x}) = \exp(ikr\cos(\phi-\phi_k))\exp(ik_z z)$, we may write down the 3D expansion immediately

$$\langle\vec{x}|\vec{k}\rangle = \exp(i\vec{k}\cdot\vec{x}) = \sum_{\ell=-\infty}^{\infty} i^\ell J_\ell(kr)e^{i\ell(\phi-\phi_k)}e^{ik_z z}, \tag{12.2.93}$$

where

$$k_i = (k\cos\phi_k, k\sin\phi_k, k_z)\,, \qquad x^i = (r\cos\phi, r\sin\phi, z)\,. \tag{12.2.94}$$

Infinite Flat Space, Spherical Coordinates We now turn to 3D flat space written in spherical coordinates,

$$\begin{aligned}\mathrm{d}\ell^2 &= \mathrm{d}r^2 + r^2\mathrm{d}\Omega_{\mathbb{S}^2}^2, \qquad \mathrm{d}\Omega_{\mathbb{S}^2}^2 \equiv \mathrm{d}\theta^2 + (\sin\theta)^2\mathrm{d}\phi^2, \\ r &\geq 0,\ \phi \in [0, 2\pi),\ \theta \in [0, \pi], \qquad \sqrt{|g|} = r^2\sin\theta.\end{aligned} \tag{12.2.95}$$

The Laplacian on a scalar is

$$\vec{\nabla}^2\psi = \frac{1}{r^2}\partial_r\left(r^2\partial_r\psi\right) + \frac{1}{r^2}\vec{\nabla}^2_{\mathbb{S}^2}\psi. \tag{12.2.96}$$

where $\vec{\nabla}^2_{\mathbb{S}^2}$ is the Laplacian on a 2−sphere.
Plane wave With

$$k_i = k\left(\sin(\theta_k)\cos(\phi_k), \sin(\theta_k)\sin(\phi_k), \cos(\theta_k)\right) \equiv k\widehat{k}, \tag{12.2.97}$$

$$x^i = r\left(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)\right) \equiv r\widehat{x}, \tag{12.2.98}$$

we have

$$\langle\vec{x}|\vec{k}\rangle = \exp(i\vec{k}\cdot\vec{x}) = \exp\left(ikr\widehat{k}\cdot\widehat{x}\right). \tag{12.2.99}$$

If we view $\widehat{k}$ as the 3−direction, this means the plane wave has no dependence on the azimuthal angle describing rotation about the 3−direction. This in turn indicates we should be able to expand $\langle\vec{x}|\vec{k}\rangle$ using $P_\ell(\widehat{k}\cdot\vec{x})$.

$$\exp\left(ikr\widehat{k}\cdot\widehat{x}\right) = \sum_{\ell=0}^{\infty}\chi_\ell(kr)\sqrt{\frac{2\ell+1}{4\pi}}P_\ell\left(\widehat{k}\cdot\widehat{x}\right). \tag{12.2.100}$$

For convenience we have used the Y_ℓ^0 in eq. (12.2.71)) as our basis. Exploiting the orthonormality of the spherical harmonics to solve for the expansion coefficients:

$$\chi_\ell(kr) = 2\pi\int_{-1}^{+1}\mathrm{d}c e^{ikrc}\overline{Y_\ell^0(\theta,\phi)} = \sqrt{(4\pi)(2\ell+1)}\frac{1}{2}\int_{-1}^{+1}\mathrm{d}c e^{ikrc}P_\ell(c). \tag{12.2.101}$$

(Even though the integral is over the entire solid angle, the azimuthal integral is trivial and yields 2π immediately.) At this point we may refer to eq. (10.54.2) of the NIST page here for the following integral representation of the spherical Bessel function of integer order,

$$i^\ell j_\ell(z) = \frac{1}{2}\int_{-1}^{+1}\mathrm{d}c e^{izc}P_\ell(c), \qquad \ell = 0, 1, 2, \dots. \tag{12.2.102}$$

(The spherical Bessel function $j_\ell(z)$ is real when z is positive.) We have arrived at

$$\langle\vec{x}|\vec{k}\rangle = \exp(i\vec{k}\cdot\vec{x}) = \sum_{\ell=0}^{\infty}(2\ell+1)i^\ell j_\ell(kr)P_\ell\left(\widehat{k}\cdot\widehat{x}\right), \qquad k \equiv |\vec{k}| \tag{12.2.103}$$

$$= 4\pi\sum_{\ell=0}^{\infty}i^\ell j_\ell(kr)\sum_{m=-\ell}^{+\ell}Y_\ell^m(\theta,\phi)\overline{Y_\ell^m(\theta_k,\phi_k)}, \tag{12.2.104}$$

where, for the second equality, we have employed the additional formula in eq. (12.2.76).
Spectrum Just as we did for the 2D plane wave, we may now read off the eigenfunctions of the 3D flat Laplacian in spherical coordinates. First we compute the normalization.

$$\int_{\mathbb{R}^3}\mathrm{d}^3\vec{x}\exp(i(\vec{k}-\vec{k}')\cdot\vec{x}) = (2\pi)^3\frac{\delta(k-k')}{kk'}\delta\left(\cos(\theta_{k'})-\cos(\theta_k)\right)\delta\left(\phi_k-\phi_{k'}\right) \tag{12.2.105}$$

Switching to spherical coordinates within the integral on the left-hand-side, namely $\mathrm{d}^3\vec{x} = \mathrm{d}(\cos\theta)\mathrm{d}\phi\mathrm{d}r r^2 \equiv \mathrm{d}\Omega\mathrm{d}r r^2$; re-expressing $\exp(i\vec{k}\cdot\vec{x})$ and $\exp(-i\vec{k}'\cdot\vec{x})$ using eq. (12.2.103) and its complex conjugate; followed by using eq. (12.2.72) to integrate over the solid angle,

$$\begin{aligned}(4\pi)^2 \int_{\mathbb{S}^2} \mathrm{d}^2\Omega \int_0^\infty \mathrm{d}r r^2 \sum_{\ell,\ell'=0}^{\infty} i^\ell (-i)^{\ell'} j_\ell(kr) j_{\ell'}(k'r) \\ \times \sum_{m=-\ell}^{+\ell} \sum_{m'=-\ell'}^{+\ell'} Y_\ell^m(\theta,\phi)\overline{Y_\ell^m(\theta_k,\phi_k)} Y_{\ell'}^{m'}(\theta_k,\phi_k)\overline{Y_{\ell'}^{m'}(\theta,\phi)} \\ = (4\pi)^2 \int_0^\infty \mathrm{d}r r^2 \sum_{\ell=0}^{\infty} j_\ell(kr) j_\ell(k'r) \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k,\phi_k)\overline{Y_\ell^m(\theta_k,\phi_k)}. \end{aligned} \tag{12.2.106}$$

Let us compare the right hand sides of the two preceding equations, and utilize the completeness relation obeyed by the spherical harmonics (cf. eq. (12.2.73)):

$$\begin{aligned}4(2\pi)^2 \int_0^\infty \mathrm{d}r r^2 \sum_{\ell=0}^{\infty} j_\ell(kr) j_\ell(k'r) \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k,\phi_k)\overline{Y_\ell^m(\theta_k,\phi_k)} \\ = (2\pi)^3 \frac{\delta(k-k')}{kk'} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k,\phi_k)\overline{Y_\ell^m(\theta_k,\phi_k)}. \end{aligned} \tag{12.2.107}$$

Therefore it must be that

$$\int_0^\infty \mathrm{d}r r^2 j_\ell(kr) j_\ell(k'r) = \frac{\pi}{2}\frac{\delta(k-k')}{kk'}. \tag{12.2.108}$$

Referring to eq. (10.47.3) of the NIST page here,

$$j_\ell(z) = \sqrt{\frac{\pi}{2z}} J_{\ell+\frac{1}{2}}(z) \tag{12.2.109}$$

we see this is in fact the same result as in eq. (12.2.44).

To sum, we have diagonalized the 3D flat space negative Laplacian in spherical coordinates as follows.

$$\begin{aligned} -\vec{\nabla}^2 \langle r,\theta,\phi | k,\ell,m\rangle &= k^2 \langle r,\theta,\phi | k,\ell,m\rangle , \\ \langle r,\theta,\phi | k,\ell,m\rangle &= \sqrt{\frac{2}{\pi}} j_\ell(kr) Y_\ell^m(\theta,\phi), \\ \langle k',\ell',m' | k,\ell,m\rangle &= \int_{\mathbb{S}^2} \mathrm{d}^2\Omega \int_0^\infty \mathrm{d}r r^2 \langle k',\ell',m' | r,\theta,\phi\rangle \langle r,\theta,\phi | k,\ell,m\rangle , \\ &= \frac{\delta(k-k')}{kk'} \delta_\ell^{\ell'} \delta_m^{m'}. \end{aligned} \tag{12.2.110}$$

Problem 12.11. Prolate Ellipsoidal Coordinates in 3D Flat Space 3D Euclidean space can be foliated by prolate ellipsoids in the following way. Let $\vec{x} \equiv (x^1, x^2, x^3)$ be Cartesian coordinates; s be the size of a given prolate ellipsoid; and the angular coordinates $(0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$ specify a point on its 2D surface. Then,

$$\vec{x} = \frac{1}{2}\left(\sqrt{s^2 - R^2}\sin\theta\cos\phi, \sqrt{s^2-R^2}\sin\theta\sin\phi, \rho\cos\theta\right); \tag{12.2.111}$$

$$s \geq R, \qquad (\theta, \phi) \in \mathbb{S}^2. \tag{12.2.112}$$

Explain the geometric meaning of the constant R. First show that the 3D flat metric $\mathrm{d}\vec{x}^2$ in prolate ellipsoidal coordinates (s, θ, ϕ) is

$$\mathrm{d}\ell^2 = \frac{s^2 - R^2\cos^2\theta}{4(s^2-R^2)}\mathrm{d}s^2 + \frac{1}{4}\left(s^2 - R^2\cos^2\theta\right)\mathrm{d}\theta^2 + \frac{1}{4}(s^2-R^2)\sin^2\theta\mathrm{d}\phi^2. \tag{12.2.113}$$

Then proceed to employ separation-of-variables to derive the differential equations for the $\sigma-$, $\chi-$ and $\phi-$dependent eigenfunctions of the scalar Laplacian $\vec{\nabla}^2 \equiv g^{ij}\nabla_i\nabla_j$. If we define

$$\psi(s, \theta, \phi) \equiv \rho(\sigma)P(\chi)\Phi(\phi), \tag{12.2.114}$$

$$\sigma \equiv s/R, \qquad \chi \equiv \cos\theta; \tag{12.2.115}$$

you should discover

$$\partial_\sigma\left((1-\sigma^2)\rho'(\sigma)\right) + \left(\frac{1}{4}k^2R^2\left(1-\sigma^2\right) - \lambda - \frac{m^2}{1-\sigma^2}\right)\rho(\sigma) = 0, \tag{12.2.116}$$

$$\partial_\chi\left((1-\chi^2)P'(\chi)\right) + \left(\frac{1}{4}k^2R^2\left(1-\chi^2\right) - \lambda - \frac{m^2}{1-\chi^2}\right)P(\chi) = 0, \tag{12.2.117}$$

$$\Phi''(\phi) + m^2\Phi(\phi) = 0; \tag{12.2.118}$$

where $-k^2$ is the eigenvalue of $\vec{\nabla}^2$; and λ and m^2 are separation constants. Notice, in the spherical coordinates limit, we would put $R \to 0$ and $\lambda = -\ell(\ell+1)$. For more information on the solutions to these equations, see DLMF.

Hint: Work out the appropriate eigenvector equation and first multiply throughout by $(1/4)(s^2-R^2)(1-\chi^2)$ to isolate the Φ terms. Next, multiply throughout by $4(s^2-R^2\chi^2)(s^2-R^2)^{-1}(1-\chi^2)^{-1}$. You should find the σ-dependent portions separating from the $\chi-$dependent ones after re-writing $s^2 - R^2\chi^2 = (s^2-R^2) + R^2(1-\chi^2)$. □

12.3 Heat/Diffusion Equation

12.3.1 Definition, uniqueness of solutions

We will define the heat or diffusion equation to be the PDE

$$\partial_t\psi\left(t, \vec{x}\right) = \sigma\vec{\nabla}^2_{\vec{x}}\psi\left(t, \vec{x}\right) = \frac{\sigma}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j\psi\right), \qquad \sigma > 0, \tag{12.3.1}$$

where $\vec{\nabla}^2_{\vec{x}}$ is the Laplacian with respect to some metric $g_{ij}(\vec{x})$, which we will assume *does not* depend on the time t. As the heat equation, it describes the temperature distribution as a function of space and time. As the diffusion equation in flat space, it describes the probability density of finding a point particle undergoing (random) Brownian motion.

We shall assume the $\psi(t,\vec{x})$ is specified on the boundary of the domain described by $g_{ij}(\vec{x})$, i.e., it obeys Dirichlet boundary conditions. The diffusion constant σ has dimensions of length if $\vec{\nabla}^2$ is of dimensions $1/[\text{Length}^2]$. We may set $\sigma = 1$ and thereby describe all other lengths in the problem in units of σ. As we shall witness, the solution of eq. (12.3.1) is aided by the knowledge of the eigenfunctions/values of the Laplacian in question.

Uniqueness of solution Suppose the following initial conditions are given

$$\psi(t = t_0, \vec{x}) = \varphi_0(\vec{x}), \tag{12.3.2}$$

and suppose the field ψ or its normal derivative is specified on the boundaries $\partial\mathfrak{D}$,

$$\psi(t, \vec{x} \in \partial\mathfrak{D}) = \varphi_3(\partial\mathfrak{D}), \qquad \text{(Dirichlet)}, \tag{12.3.3}$$

$$\text{or } n^i\nabla_i\psi(t, \vec{x} \in \partial\mathfrak{D}) = \varphi_4(\partial\mathfrak{D}), \qquad \text{(Neumann)}, \tag{12.3.4}$$

where $n^i(\partial\mathfrak{D})$ is the unit outward normal vector. Then, the solution to the heat/diffusion equation in eq. (12.3.1) is unique.

Proof Without loss of generality, since our heat/diffusion equation is linear, we may assume the field is real. We then suppose there are two such solutions ψ_1 and ψ_2; the proof is established if we can show, in fact, that ψ_1 has to be equal to ψ_2. Note that the difference $\Psi \equiv \psi_1 - \psi_2$ is subject to the initial conditions

$$\Psi(t = t_0, \vec{x}) = 0, \tag{12.3.5}$$

and the spatial boundary conditions

$$\Psi(t, \vec{x} \in \partial\mathfrak{D}) = 0 \qquad \text{or} \qquad n^i\nabla_i\Psi(t, \vec{x} \in \partial\mathfrak{D}) = 0. \tag{12.3.6}$$

Let us then consider the following (non-negative) integral

$$\rho(t) \equiv \frac{1}{2}\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\Psi(t,\vec{x})^2 \geq 0, \tag{12.3.7}$$

as well as its time derivative

$$\partial_t\rho(t) = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\Psi\dot{\Psi}. \tag{12.3.8}$$

We may use the heat/diffusion equation on the $\dot{\Psi}$ term, and integrate-by-parts one of the gradients on the second term,

$$\begin{aligned}\partial_t\rho(t) &= \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\Psi\vec{\nabla}^2\Psi \\ &= \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}\Psi n^i\nabla_i\Psi - \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\nabla_i\Psi\nabla^i\Psi. \end{aligned}\tag{12.3.9}$$

By assumption either Ψ or $n^i\nabla_i\Psi$ is zero on the spatial boundary; therefore the first term on the second line is zero. We have previously argued that the integrand in the second term on the second line is strictly non-negative

$$\nabla_i\Psi\nabla^i\Psi = \sum_i (\nabla_{\widehat{i}}\Psi)^2 \geq 0. \tag{12.3.10}$$

This implies

$$\partial_t\rho(t) = -\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\nabla_i\Psi\nabla^i\Psi \leq 0. \tag{12.3.11}$$

However, the initial conditions $\Psi(t = t_0, \vec{x}) = 0$ indicate $\rho(t = t_0) = 0$ (cf. eq. (12.3.7)). Moreover, since $\rho(t \geq t_0)$ has to be non-negative from its very definition and since we have just shown its time derivative is non-positive, $\rho(t \geq t_0)$ therefore has to remain zero for all subsequent time $t \geq t_0$; i.e., it cannot decrease below zero. And because $\rho(t)$ is the integral of the square of Ψ, the only way it can be zero is $\Psi = 0 \Rightarrow \psi_1 = \psi_2$. This establishes the theorem. $\square$

12.3.2 Heat Kernel, Solutions with $\psi(\partial\mathfrak{D}) = 0$

In this section we introduce the propagator, otherwise known as the *heat kernel*, which will prove to be key to solving the heat/diffusion equation. It is the matrix element

$$K(\vec{x}, \vec{x}'; s \geq 0) \equiv \left\langle \vec{x} \middle| e^{s\vec{\nabla}^2} \middle| \vec{x}' \right\rangle. \tag{12.3.12}$$

It obeys the heat/diffusion equation

$$\begin{aligned}\partial_s K(\vec{x}, \vec{x}'; s) &= \left\langle \vec{x} \middle| \vec{\nabla}^2 e^{s\vec{\nabla}^2} \middle| \vec{x}' \right\rangle = \left\langle \vec{x} \middle| e^{s\vec{\nabla}^2}\vec{\nabla}^2 \middle| \vec{x}' \right\rangle \\ &= \vec{\nabla}^2_{\vec{x}} K(\vec{x}, \vec{x}'; s) = \vec{\nabla}^2_{\vec{x}'} K(\vec{x}, \vec{x}'; s),\end{aligned} \tag{12.3.13}$$

where we have assumed $\vec{\nabla}^2$ is Hermitian. K also obeys the initial condition

$$K(\vec{x}, \vec{x}'; s = 0) = \langle \vec{x} | \vec{x}' \rangle = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}. \tag{12.3.14}$$

If we demand the eigenfunctions of $\vec{\nabla}^2$ obey Dirichlet boundary conditions,

$$\left\{ \psi_\lambda(\partial\mathfrak{D}) = 0 \middle| -\vec{\nabla}^2\psi_\lambda = \lambda\psi_\lambda \right\}, \tag{12.3.15}$$

then the heat kernel obeys the same boundary conditions.

$$K(\vec{x} \in \partial\mathfrak{D}, \vec{x}'; s) = K(\vec{x}, \vec{x}' \in \partial\mathfrak{D}; s) = 0. \tag{12.3.16}$$

To see this we need to perform a *mode expansion*. By inserting in eq. (12.3.14) a complete set of the eigenstates of $\vec{\nabla}^2$, the heat kernel has an explicit solution

$$K(\vec{x}, \vec{x}'; s \geq 0) = \left\langle \vec{x} \middle| e^{s\vec{\nabla}^2} \middle| \vec{x}' \right\rangle = \sum_\lambda e^{-s\lambda} \langle \vec{x} | \lambda \rangle \langle \lambda | \vec{x}' \rangle, \tag{12.3.17}$$

where the sum is schematic: depending on the setup at hand, it can consist of either a sum over discrete eigenvalues and/or an integral over a continuum. In this form, it is manifest the heat kernel vanishes when either $\vec{x}$ or $\vec{x}'$ lies on the boundary $\partial\mathfrak{D}$.

Initial value problem In this section we will focus on solving the initial value problem when the field itself is zero on the boundary $\partial\mathfrak{D}$ for all relevant times. This will in fact be the case for infinite domains; for example, flat $\mathbb{R}^D$, whose heat kernel we will work out explicitly below. The setup is thus as follows:

$$\psi(t=t',\vec{x}) \equiv \langle \vec{x}|\,\psi(t')\rangle \ \text{(given)}, \qquad \psi(t\geq t',\vec{x}\in\mathfrak{D})=0. \tag{12.3.18}$$

Then $\psi(t,\vec{x})$ at any later time $t>t'$ is given by

$$\begin{aligned}\psi(t\geq t',\vec{x}) = \left\langle \vec{x}\left|e^{(t-t')\vec{\nabla}^2}\right|\psi(t')\right\rangle &= \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\left\langle \vec{x}\left|e^{(t-t')\vec{\nabla}^2}\right|\vec{x}'\right\rangle \langle \vec{x}'|\,\psi(t')\rangle \\ &= \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}K(\vec{x},\vec{x}';t-t')\psi(t',\vec{x}').\end{aligned} \tag{12.3.19}$$

That this is the correct solution is because the right hand side obeys the heat/diffusion equation through eq. (12.3.13). As $t\to t'$, we also see from eq. (12.3.14) that the initial condition is recovered.

$$\psi(t=t',\vec{x}) = \langle \vec{x}|\,\psi(t')\rangle = \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{|g(\vec{x}')g(\vec{x})|}}\psi(t',\vec{x}') = \psi(t',\vec{x}). \tag{12.3.20}$$

Moreover, since the heat kernel obeys eq. (12.3.16), the solution automatically maintains the $\psi(t\geq t',\vec{x}\in\mathfrak{D})=0$ boundary condition.

Decay times, Asymptotics Suppose we begin with some temperature distribution $T(t',\vec{x})$. By expanding it in the eigenfunctions of the Laplacian, let us observe that it is the component along the eigenfunction with the small eigenvalue that dominates the late time temperature distribution. From eq. (12.3.19) and (12.3.17),

$$\begin{aligned}T(t\geq t',\vec{x}) &= \sum_\lambda \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\left\langle \vec{x}\left|e^{(t-t')\vec{\nabla}^2}\right|\lambda\right\rangle \langle\lambda|\,\vec{x}'\rangle\langle \vec{x}'|\,T(t')\rangle \\ &= \sum_\lambda e^{-(t-t')\lambda}\langle \vec{x}|\,\lambda\rangle \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\langle\lambda|\,\vec{x}'\rangle\langle \vec{x}'|\,T(t')\rangle \\ &= \sum_\lambda e^{-(t-t')\lambda}\langle \vec{x}|\,\lambda\rangle\langle\lambda|\,T(t')\rangle\,.\end{aligned} \tag{12.3.21}$$

Remember we have proven that the eigenvalues of the Laplacian are strictly non-positive. That means, as $(t-t')\to\infty$, the dominant temperature distribution is

$$T(t-t'\to\infty,\vec{x}) \approx e^{-(t-t')\lambda_{\min}}\langle \vec{x}|\,\lambda_{\min}\rangle \int \mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\langle\lambda_{\min}|\,\vec{x}'\rangle\langle \vec{x}'|\,T(t')\rangle\,, \tag{12.3.22}$$

because all the $\lambda>\lambda_{\min}$ become exponentially suppressed (relative to the $\lambda_{\min}$ state) due to the presence of $e^{-(t-t')\lambda}$. As long as the minimum eigenvalue $\lambda_{\min}$ is strictly positive, we see the final temperature is zero.

$$T(t-t'\to\infty,\vec{x})=0, \qquad \text{if } \lambda_{\min}>0. \tag{12.3.23}$$

When the minimum eigenvalue is zero, $\lambda_{\min} = 0$, we have instead

$$T(t - t' \to \infty, \vec{x}) \to \langle \vec{x} | \, \lambda = 0 \rangle \int \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \, \langle \lambda = 0 | \, \vec{x}' \rangle \, \langle \vec{x}' | \, T(t') \rangle \, . \tag{12.3.24}$$

The exception to the dominant behavior in eq. (12.3.22) is when there is zero overlap between the initial distribution and that eigenfunction with the smallest eigenvalue, i.e., if

$$\int \mathrm{d}^D \vec{x}' \sqrt{|g(\vec{x}')|} \, \langle \lambda_{\min} | \, \vec{x}' \rangle \, \langle \vec{x}' | \, T(t') \rangle = 0. \tag{12.3.25}$$

Generically, we may say that, with the passage of time, the component of the initial distribution along the eigenfunction corresponding to the eigenvalue λ decays as $1/\lambda$; i.e., when $t - t' = 1/\lambda$, its amplitude falls by $1/e$.

Static limit: Laplace from Heat/Diffusion Another way of phrasing the $(t - t') \to \infty$ behavior is that – since every term in the sum-over-eigenvalues that depends on time decays exponentially, it must be that the late time asymptotic limit is simply the static limit, when the time derivative on the left hand side of eq. (12.3.1) is zero and we obtain Laplace's equation

$$0 = \vec{\nabla}^2 \psi(t \to \infty, \vec{x}). \tag{12.3.26}$$

In the late time limit, equilibrium $\partial_t \psi = 0$ *will* be achieved.

Probability interpretation in flat infinite space In the context of the diffusion equation in flat space, because of the δ-functions on the right hand side of eq. (12.3.14), the propagator $K(\vec{x}, \vec{x}'; t - t')$ itself can be viewed as the probability density ($\equiv$ probability per volume) of finding the Brownian particle – which was infinitely localized at $\vec{x}'$ at the initial time t' – at a given location $\vec{x}$ some later time $t > t'$. To support this probability interpretation it has to be that

$$\int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} K(\vec{x}, \vec{x}'; t - t') = 1. \tag{12.3.27}$$

The integral on the left hand side corresponds to summing the probability of finding the Brownian particle over all space – that has to be unity, since the particle has to be *somewhere*. We can verify this directly, by inserting a complete set of states.

$$\begin{aligned} \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} K(\vec{x}, \vec{x}'; t - t') &= \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} \left\langle \vec{x} \left| e^{(t-t')\vec{\nabla}^2} \right| \vec{x}' \right\rangle \qquad (12.3.28) \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} \left\langle \vec{x} \left| e^{(t-t')\vec{\nabla}^2} \right| \vec{k} \right\rangle \left\langle \vec{k} \right| \left. \vec{x}' \right\rangle \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} e^{-(t-t')\vec{k}^2} \langle \vec{x} | \vec{k} \rangle \left\langle \vec{k} \right| \left. \vec{x}' \right\rangle \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x} e^{-(t-t')\vec{k}^2} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} e^{-(t-t')\vec{k}^2} e^{-i\vec{k}\cdot\vec{x}'} \delta^{(D)}(\vec{k}) = 1. \qquad (12.3.29) \end{aligned}$$

Problem 12.12. Probability interpretation in closed spaces In a space without boundaries, i.e., closed, use the mode expansion of eq. (12.3.17) to argue that probability is conserved:

$$\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}K(\vec{x},\vec{x}';t-t') = 1. \tag{12.3.30}$$

Hint: In a closed space, recall that the Laplacian must have a zero eigenvalue – what is the corresponding $\langle\vec{x}|\,\lambda=0\rangle$? □

Heat Kernel in Infinite Flat Space We now turn to evaluating the heat kernel in flat $\mathbb{R}^D$. By inserting a compete set of momentum states,

$$\overline{K}(\vec{x},\vec{x}';t-t') \equiv \left\langle \vec{x}\left| e^{(t-t')\vec{\nabla}^2}\right|\vec{x}'\right\rangle \tag{12.3.31}$$

$$= \int_{\mathbb{R}^D} \mathrm{d}^D\vec{k}\left\langle \vec{x}\left| e^{(t-t')\vec{\nabla}^2}\right|\vec{k}\right\rangle\left\langle\vec{k}\right|\vec{x}'\right\rangle \tag{12.3.32}$$

$$= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{-(t-t')\vec{k}^2} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} = \prod_{j=1}^{D}\int_{-\infty}^{+\infty}\frac{\mathrm{d}k_j}{2\pi}e^{-(t-t')(k_j)^2}e^{ik_j(x^j-x'^j)}.$$

We may "complete the square" in the exponent by considering

$$-(t-t')\left(k_j - i\frac{x^j-x'^j}{2(t-t')}\right)^2 = -(t-t')\left((k_j)^2 - ik_j\frac{x^j-x'^j}{t-t'} - \left(\frac{x^j-x'^j}{2(t-t')}\right)^2\right). \tag{12.3.33}$$

The heat kernel in flat $\mathbb{R}^D$ is therefore

$$\overline{K}(\vec{x},\vec{x}';t-t') \equiv \left\langle \vec{x}\left| e^{(t-t')\sigma\vec{\nabla}^2}\right|\vec{x}'\right\rangle \tag{12.3.34}$$

$$= (4\pi\sigma(t-t'))^{-D/2}\exp\left(-\frac{(\vec{x}-\vec{x}')^2}{4\sigma(t-t')}\right), \qquad t>t', \tag{12.3.35}$$

where we have put back the diffusion constant σ. If you have taken quantum mechanics, you may recognize this result to be very similar to the path integral ${}_{\mathrm{H}}\langle\vec{x},t|\,\vec{x}',t'\rangle_{\mathrm{H}}$ of a free particle – compare eq. (12.3.34) with eq. (5.2.85) and notice the former may be obtained from the latter via the simultaneous replacements $m\to 1/(2\sigma)$ and $t-t'\to i(t-t')$. In fact, the heat equation itself becomes the Schrödinger equation upon similar replacements.

Heat Kernel on Circle $\mathbb{S}^1$ A circle of radius R (embedded in 2D flat space) is simply infinite 1D flat space but divided into periodic domains of length $2\pi R$. The metric is $\mathrm{d}x^2 \equiv R^2\mathrm{d}\phi^2$, with the polar angle running over the interval $[0,2\pi)$. Hence, the heat kernel on such a circle should be that of the 1D flat space case, summed over the infinite number of image charges at $x'+2\pi Rn$, for $n=0,\pm1,\pm2,\dots$. That is, if ϕ and ϕ' denote the polar angles on the circle, the heat kernel on the circle is

$$K_{\mathbb{S}^1}(\phi-\phi',t-t') = \sum_{n=-\infty}^{+\infty}\overline{K}\left(t-t',R(\phi-\phi'+2\pi n)\right) \tag{12.3.36}$$

$$= \sum_{n=-\infty}^{+\infty} \frac{\exp\left(-\frac{R^2(\phi-\phi'+2\pi n)^2}{4\sigma(t-t')}\right)}{\sqrt{4\pi\sigma(t-t')}}, \tag{12.3.37}$$

where $\overline{K}$ is the 1D version of eq. (12.3.31) and (12.3.34). The Laplacian on the circle is $R^{-2}\partial_\phi^2$, hence we may readily see that

$$\left(\partial_t - \sigma\partial_{(R\phi)}^2\right) \overline{K}\left(t-t', R(\phi-\phi'+2\pi n)\right) = \left(\partial_t - \sigma\partial_x^2\right) \overline{K}\left(t-t', x-x'\right) = 0, \tag{12.3.38}$$

where $x \equiv R\phi$ and $x' \equiv R(\phi'-2\pi n)$, by virtue of $\overline{K}$ being the flat space heat kernel. Taking the $t \to t'$ to recover the initial conditions,

$$\lim_{t\to t'} K_{\mathbb{S}^1}\left(\phi-\phi', t-t'\right) = \frac{1}{R}\sum_{n=-\infty}^{+\infty} \delta\left(\phi-\phi'+2\pi n\right), \tag{12.3.39}$$

where the flat space version of eq. (12.3.14) has been employed. If ϕ and ϕ' are polar angles on the circle, they are restricted to lie in the interval $[0, 2\pi)$, and therefore cannot have a difference equal to or greater than 2π. Hence, only the $n = 0$ term in the above summation is of relevance.

$$\lim_{t\to t'} K_{\mathbb{S}^1}\left(\phi-\phi', t-t'\right) = \frac{1}{R}\delta\left(\phi-\phi'\right). \tag{12.3.40}$$

Problem 12.13. Mode Expansion Since the circle has no boundaries, its Laplacian is automatically hermitian. Exploit the result in eq. (12.3.17) to deduce

$$K_{\mathbb{S}^1}(\phi-\phi', t-t') = \sum_{n=-\infty}^{+\infty} \exp\left(-(t-t')\sigma\cdot\left(\frac{n}{R}\right)^2\right)\frac{\exp(in(\phi-\phi'))}{2\pi}. \tag{12.3.41}$$

Show that eq. (12.3.41) is in fact equivalent to eq. (12.3.37). Hint: The Poisson summation formula in eq. (7.8.202) may be useful. □

12.3.3 Green's functions and initial value formulation in a finite domain

Green's function from Heat Kernel Given the heat kernel defined with Dirichlet boundary conditions, the associated Green's function is defined as

$$G_K(t-t'; \vec{x}, \vec{x}') \equiv \Theta(t-t')K(\vec{x}, \vec{x}'; t-t'), \tag{12.3.42}$$

where we define $\Theta(s) = 1$ for $s \geq 0$ and $\Theta(s) = 0$ for $s < 0$. This Green's function G_K obeys

$$\left(\partial_t - \vec{\nabla}_{\vec{x}}^2\right) G_K(t-t'; \vec{x}, \vec{x}') = \left(\partial_t - \vec{\nabla}_{\vec{x}'}^2\right) G_K(t-t'; \vec{x}, \vec{x}') = \delta(t-t')\frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}, \tag{12.3.43}$$

with the Dirichlet boundary condition

$$G_K(\tau; \vec{x} \in \partial\mathfrak{D}, \vec{x}') = G_K(\tau; \vec{x}, \vec{x}' \in \partial\mathfrak{D}) = 0, \tag{12.3.44}$$

as well as the causality condition

$$G_K(\tau;\vec{x},\vec{x}')=0 \qquad \text{when} \qquad \tau<0. \tag{12.3.45}$$

The boundary condition in eq. (12.3.44) follows directly from eq. (12.3.16); whereas eq. (12.3.43) follows from equations (12.3.13) and (12.3.14):

$$\begin{aligned}\left(\partial_t-\vec{\nabla}^2\right)G_K(t-t';\vec{x},\vec{x}') &= \delta(t-t')K(\vec{x},\vec{x}';t-t')+\Theta(t-t')\left(\partial_t-\vec{\nabla}^2\right)K(\vec{x},\vec{x}';t-t') \\ &= \delta(t-t')\frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}. \end{aligned} \tag{12.3.46}$$

Initial value problem Within a spatial domain $\mathfrak{D}$, suppose the initial field configuration $\psi(t',\vec{x}\in\mathfrak{D})$ is given and suppose its value on the spatial boundary $\partial\mathfrak{D}$ is also provided (i.e., $\psi(t\geq t',\vec{x}\in\partial\mathfrak{D})$ are specified). The unique solution $\psi(t\geq t',\vec{x}\in\mathfrak{D})$ to the heat/diffusion equation (12.3.1) is

$$\begin{aligned}\psi(t\geq t',\vec{x}) &= \int_{\mathfrak{D}}\mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}K(t-t';\vec{x},\vec{x}')\psi(t',\vec{x}') \\ &\quad -\int_{t'}^{t}\mathrm{d}t''\int_{\partial\mathfrak{D}}\mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}n^{i'}\nabla_{i'}K\left(t-t'';\vec{x},\vec{x}'(\vec{\xi})\right)\psi\left(t'',\vec{x}'(\vec{\xi})\right),\end{aligned} \tag{12.3.47}$$

where K is the heat kernel in eq. (12.3.12), which in turn obeys the Dirichlet boundary conditions in eq. (12.3.16).

Proof We shall now employ the Green's function G_K to derive eq. (12.3.47). Let's begin by multiplying both sides of eq. (12.3.43) by $\psi(t'',\vec{x}')$ and integrating over both space and time (from t' to infinity).

$$\begin{aligned}\psi(t\geq t',\vec{x}) &= \int_{t'-}^{\infty}\mathrm{d}t''\int_{\mathfrak{D}}\mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\left(\partial_t-\vec{\nabla}^2_{\vec{x}'}\right)G_K(t-t'';\vec{x},\vec{x}')\psi(t'',\vec{x}') \\ &= \int_{t'}^{\infty}\mathrm{d}t''\int_{\mathfrak{D}}\mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\left(-\partial_{t''}G_K\psi+\nabla_{i'}G_K\nabla^{i'}\psi\right) \\ &\qquad -\int_{t'}^{\infty}\mathrm{d}t''\int_{\partial\mathfrak{D}}\mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}n^{i'}\nabla_{i'}G_K\psi \\ &= \int_{\mathfrak{D}}\mathrm{d}^D\vec{x}'\sqrt{|g(\vec{x}')|}\left\{[-G_K\psi]^{t''=\infty}_{t''=t'}+\int_{t'}^{\infty}\mathrm{d}t''G_K\left(\partial_{t''}-\vec{\nabla}^2_{\vec{x}''}\right)\psi\right\} \\ &\qquad +\int_{t'}^{\infty}\mathrm{d}t''\int_{\partial\mathfrak{D}}\mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}\left(G_K\cdot n^{i'}\nabla_{i'}\psi-n^{i'}\nabla_{i'}G_K\cdot\psi\right).\end{aligned} \tag{12.3.48}$$

If we impose the boundary condition in eq. (12.3.45), we see that $[-G_K\psi]^{t''=\infty}_{t''=t'}=G_K(t-t')\psi(t')$ because the upper limit contains $G_K(t-\infty)\equiv\lim_{t'\to-\infty}\Theta(t-t')K(\vec{x},\vec{x}';t-t')=0$. The heat/diffusion eq. (12.3.1) removes the time-integral term on the first line of the last equality. If Dirichlet boundary conditions were chosen, we may choose $G_K(t-t'';\vec{x},\vec{x}'\in\partial\mathfrak{D})=0$ (i.e., eq. (12.3.44)) and obtain eq. (12.3.47). Note that the upper limit of integration in the last line is really t, because eq. (12.3.45) tells us the Green's function vanishes for $t''>t$. As long as $t>t'$, the $G_K(t-t';\vec{x},\vec{x}')$ may be replaced with $K(\vec{x},\vec{x}';t-t')$. This completes the derivation of eq. (12.3.47).

Finally, recall we have already in §(12.3.1) proven the uniqueness of the solution to the heat equation obeying Dirichlet or Neumann boundary conditions.

12.3.4 Problems

Problem 12.14. In infinite flat $\mathbb{R}^D$, suppose we have some initial probability distribution of finding a Brownian particle, expressed in Cartesian coordinates as

$$\psi(t = t_0, \vec{x}) = \left(\frac{\omega}{\pi}\right)^{D/2} \exp\left(-\omega(\vec{x} - \vec{x}_0)^2\right), \qquad \omega > 0. \tag{12.3.49}$$

Solve the diffusion equation for $t \geq t_0$.

Problem 12.15. Suppose we have some initial temperature distribution $T(t = t_0, \theta, \phi) \equiv T_0(\theta, \phi)$ on a thin spherical shell. This distribution admits some multipole expansion:

$$T_0(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_\ell^m Y_\ell^m(\theta, \phi), \qquad a_\ell^m \in \mathbb{C}. \tag{12.3.50}$$

The temperature as a function of time obeys the heat/diffusion equation

$$\partial_t T(t, \theta, \phi) = \sigma \vec{\nabla}^2 T(t, \theta, \phi), \qquad \sigma > 0, \tag{12.3.51}$$

where $\vec{\nabla}^2$ is now the Laplacian on the 2−sphere. Since $\vec{\nabla}^2$ is dimensionless here, σ has units of 1/[Time].

1. Solve the propagator K for the heat/diffusion equation on the 2−sphere, in terms of a spherical harmonic $\{Y_\ell^m(\theta, \phi)\}$ expansion.
2. Find the solution for $T(t > t_0, \theta, \phi)$.
3. What is the decay rate of the ℓth multipole, i.e., how much time does the ℓth term in the multipole sum take to decay in amplitude by $1/e$? Does it depend on both ℓ and m? And, what is the final equilibrium temperature distribution?

Problem 12.16. Inverse of Laplacian from Heat Kernel In this problem we want to point out how the Green's function of the Laplacian is related to the heat/diffusion equation. To re-cap, the Green's function itself obeys the D-dimensional PDE:

$$-\vec{\nabla}^2 G(\vec{x}, \vec{x}') = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt[4]{g(\vec{x}) g(\vec{x}')}}. \tag{12.3.52}$$

As already suggested by our previous discussions, the Green's function $G(\vec{x}, \vec{x}')$ can be viewed as the matrix element of the operator $G \equiv 1/(-\vec{\nabla}^2)$, namely[144]

$$G(\vec{x}, \vec{x}') = \langle \vec{x} | G | \vec{x}' \rangle \equiv \left\langle \vec{x} \left| \frac{1}{-\vec{\nabla}^2} \right| \vec{x}' \right\rangle. \tag{12.3.53}$$

[144] The perspective that the Green's function be viewed as an operator acting on some Hilbert space was advocated by theoretical physicist Julian Schwinger.

Now use eq. (7.2.1) to justify

$$G(\vec{x},\vec{x}') = \int_0^\infty \mathrm{d}t K\left(\vec{x},\vec{x}';t\right), \tag{12.3.54}$$

where K is the propagator (eq. (12.3.12)) of the heat/diffusion equation.

Flat Space We will borrow from our previous linear algebra discussion that $-\vec{\nabla}^2 = \vec{P}^2$, as can be seen from its position space representation. Now proceed to re-write this integral by inserting to both the left and to the right of the operator $e^{t\vec{\nabla}^2}$ the completeness relation in momentum space. Use the fact that $\vec{P}^2 = -\vec{\nabla}^2$ and eq. (7.2.1) to deduce

$$G(\vec{x},\vec{x}') = \int_0^\infty \mathrm{d}t \int \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{-t\vec{k}^2} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{12.3.55}$$

(Going to momentum space allows you to also justify in what sense the restriction $\mathrm{Re}(b) > 0$ of the formula in eq. (7.2.1) was satisfied.) By appropriately "completing the square" in the exponent, followed by an application of eq. (7.2.1), evaluate this integral to arrive at the Green's function of the Laplacian in D spatial dimensions:

$$G(\vec{x},\vec{x}') = \left\langle \vec{x} \left| \frac{1}{-\vec{\nabla}^2} \right| \vec{x}' \right\rangle = \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}|\vec{x}-\vec{x}'|^{D-2}}, \tag{12.3.56}$$

where $|\vec{x}-\vec{x}'|$ is the Euclidean distance between $\vec{x}$ and $\vec{x}'$.

Next, can you use eq. 18.12.4 of the NIST page here to perform an expansion of the Green's function of the negative Laplacian in terms of Gegenbauer polynomials $C_\ell^{(n)}$, $r_> \equiv \max(r,r')$, $r_< \equiv \min(r,r')$ and $\widehat{n}\cdot\widehat{n}'$, where $r \equiv |\vec{x}|$, $r' \equiv |\vec{x}'|$, $\widehat{n} \equiv \vec{x}/r$, and $\widehat{n}' \equiv \vec{x}'/r'$? The arbitrary D result is

$$\frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2}|\vec{x}-\vec{x}'|^{D-2}} = \frac{\Gamma\left(\frac{D}{2}-1\right)}{4\pi^{D/2} r_>^{D-2}} \sum_{\ell=0}^{\infty} C_\ell^{\left(\frac{D-2}{2}\right)}\left(\widehat{n}\cdot\widehat{n}'\right)\left(\frac{r_<}{r_>}\right)^\ell. \tag{12.3.57}$$

Be sure to explain how the $D=3$ case reduces to

$$\frac{1}{4\pi|\vec{x}-\vec{x}'|} = (4\pi r_>)^{-1} \sum_{\ell=0}^{\infty} P_\ell\left(\widehat{n}\cdot\widehat{n}'\right)\left(\frac{r_<}{r_>}\right)^\ell \tag{12.3.58}$$

$$= \frac{1}{r_>} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{\overline{Y_\ell^m(\widehat{n})} Y_\ell^m(\widehat{n}')}{2\ell+1} \left(\frac{r_<}{r_>}\right)^\ell, \tag{12.3.59}$$

where the $P_\ell = C_\ell^{\left(\frac{1}{2}\right)}$ are Legendre polynomials and in the second line the addition formula of eq. (12.2.76) was invoked.

Note that while it is not easy to verify by direct differentiation that eq. (12.3.56) is indeed the Green's function $1/(-\vec{\nabla}^2)$; one can do so *after* first performing the integral over t in eq. (12.3.55) to obtain

$$G(\vec{x},\vec{x}') = \int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}}{\vec{k}^2}. \tag{12.3.60}$$

We have already seen this in eq. (12.1.31).

Laplace from Heat/Diffusion Finally, can you use the relationship between the heat kernel and the Green's function of the Laplacian in eq. (12.3.54), to show how in a finite domain, eq. (12.3.47) leads to eq. (12.1.54) by taking the late time $t \to \infty$ limit? (You may assume the smallest eigenvalue of the negative Laplacian is strictly positive; recall eq. (12.1.60).) $\square$

Problem 12.17. Laplacian Inverse on $\mathbb{S}^2$? Is it possible to solve for the Green's function of the Laplacian on the 2-sphere; i.e., the surface defined by $\vec{x}^2 = R^2$ in flat 3−space? Use the methods of the last two problems, or simply try to write down the mode sum expansion in eq. (12.1.22), to show that you would obtain a 1/0 infinity. What is the reason for this apparent pathology? Suppose we could solve

$$-\vec{\nabla}^2 G(\vec{x}, \vec{x}') = \frac{\delta^{(2)}(\vec{x} - \vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}. \tag{12.3.61}$$

Perform a volume integral of both sides over the 2−sphere – explain the contradiction you get. (Recall the discussion in the differential geometry section.) Can you generalize the argument to $\mathbb{S}^D$, the surface of the round sphere in D−space? Can you generalize the argument to *any* surface that can be obtained by smoothly deforming $\mathbb{S}^D$?

Hint: For the $\mathbb{S}^2$ case, apply the curved space Gauss' law in eq. (9.7.84) and remember the 2-sphere is a closed surface.

12.4 Massless Scalar Wave Equation (Mostly) In Flat Spacetime $\mathbb{R}^{D,1}$

12.4.1 Spacetime Metric, Uniqueness of Minkowski Wave Solutions

Spacetime Metric In Cartesian coordinates $(t, \vec{x})$, it is possible associate a metric to flat spacetime as follows

$$\mathrm{d}s^2 = c^2 \mathrm{d}t^2 - \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x} \equiv \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu, \qquad x^\mu \equiv (ct, x^i), \tag{12.4.1}$$

where c is the speed of light in vacuum; $\mu \in \{0, 1, 2, \dots, D\}$; and D is still the dimension of space.[145] We also have defined the flat (Minkowski) spacetime metric

$$\eta_{\mu\nu} \equiv \mathrm{diag}\,(1, -1, -1, \dots, -1)\,. \tag{12.4.2}$$

The generalization of eq. (12.4.1) to curved spacetime is

$$\mathrm{d}s^2 = g_{\mu\nu}(t, \vec{x}) \mathrm{d}x^\mu \mathrm{d}x^\nu, \qquad x^\mu = (ct, x^i). \tag{12.4.3}$$

It is common to use the symbol $\square$, especially in curved spacetime, to denote the spacetime-'Laplacian':

$$\square\psi \equiv \nabla_\mu \nabla^\mu \psi = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|} g^{\mu\nu} \partial_\nu \psi \right), \tag{12.4.4}$$

[145] In this section it is important to distinguish Greek $\{\mu, \nu, \dots\}$ and Latin/English alphabets $\{a, b, i, j, \dots\}$. The former run over 0 through D, where the 0th index refers to time and the 1st through Dth to space. The latter run from 1 through D, and are thus strictly "spatial" indices. Also, be aware that the opposite sign convention, $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}\vec{x} \cdot \mathrm{d}\vec{x}$, is commonly used too. For most physical applications both sign conventions are valid; see, however, [36].

where $\sqrt{|g|}$ is now the square root of the absolute value of the determinant of the metric $g_{\mu\nu}$. In Minkowski spacetime of eq. (12.4.1), we have $\sqrt{|g|} = 1$, $\eta^{\mu\nu} = \eta_{\mu\nu}$, and

$$\Box\psi = \eta^{\mu\nu}\partial_\mu\partial_\nu\psi \equiv \partial^2\psi = \left(c^{-2}\partial_t^2 - \delta^{ij}\partial_i\partial_j\right)\psi; \tag{12.4.5}$$

where $\delta^{ij}\partial_i\partial_j = \vec{\nabla}^2$ is the spatial Laplacian in flat Euclidean space. The Minkowski "dot product" between vectors u and v in Cartesian coordinates is now

$$u \cdot v \equiv \eta_{\mu\nu}u^\mu v^\nu = u^0 v^0 - \vec{u}\cdot\vec{v}, \qquad u^2 \equiv (u^0)^2 - \vec{u}^2, \quad \text{etc.} \tag{12.4.6}$$

From here on, x, x' and k, etc. – without an arrow over them – denotes collectively the $D+1$ coordinates of spacetime. Indices of spacetime tensors are moved with $g^{\mu\nu}$ and $g_{\mu\nu}$. For instance,

$$u^\mu = g^{\mu\nu}u_\nu, \qquad\qquad u_\mu = g_{\mu\nu}u^\nu. \tag{12.4.7}$$

In the flat spacetime geometry of eq. (12.4.1), written in Cartesian coordinates,

$$u^0 = u_0, \qquad\qquad u^i = -u_i. \tag{12.4.8}$$

Indefinite signature The subtlety with the metric of spacetime, as opposed to that of space only, is that the "time" part of the distance in eq. (12.4.1) comes with a different sign from the "space" part of the metric. In curved or flat space, if $\vec{x}$ and $\vec{x}'$ have zero geodesic distance between them, they are really the same point. In curved or flat spacetime, however, x and x' may have zero geodesic distance between them, but they could either refer to the same spacetime point (aka "event") – or they could simply be lying on each other's light cone:

$$0 = (x-x')^2 = \eta_{\mu\nu}(x^\mu - x'^\mu)(x^\nu - x'^\nu) \qquad \Leftrightarrow \qquad (t-t')^2 = (\vec{x}-\vec{x}')^2. \tag{12.4.9}$$

To understand this statement more systematically, let us work out the geodesic distance between any pair of spacetime points in flat spacetime.

Problem 12.18. World Function In Minkowski spacetime expressed in Cartesian coordinates, the Christoffel symbols are zero. Therefore the geodesic equation in (9.5.44) returns the following "acceleration-is-zero" ODE:

$$0 = \frac{\mathrm{d}^2 Z^\mu(\lambda)}{\mathrm{d}\lambda^2}. \tag{12.4.10}$$

Show that the geodesic joining the initial spacetime point $Z^\mu(\lambda = 0) = x'^\mu$ to the final location $Z^\mu(\lambda = 1) = x^\mu$ is the straight line

$$Z^\mu(0 \leq \lambda \leq 1) = x'^\mu + \lambda\,(x^\mu - x'^\mu). \tag{12.4.11}$$

Use eq. (11.3.58) to show that half the *square* of the geodesic distance between x' and x is

$$\bar{\sigma}(x,x') = \frac{1}{2}(x-x')^2. \tag{12.4.12}$$

$\bar{\sigma}$ is commonly called Synge's world function in the gravitation literature. $\Box$

Some jargon needs to be introduced here. (Drawing a spacetime diagram would help.)

- When $\bar{\sigma} > 0$, we say x and x' are timelike separated. If you sit at rest in some inertial frame, then the tangent vector to your world line is $u^\mu = (1, \vec{0})$, and $u = \partial_t$ is a measure of how fast the time on your watch is running. Or, simply think about setting $\mathrm{d}\vec{x} = 0$ in the Minkowski metric: $\mathrm{d}s^2 \to \mathrm{d}t^2 > 0$.

- When $\bar{\sigma} < 0$, we say x and x' are spacelike separated. If you and your friend sit at rest in the same inertial frame, then at a fixed time $\mathrm{d}t = 0$, the (square of the) spatial distance between the both of you is now given by integrating $\mathrm{d}s^2 \to -\mathrm{d}\vec{x}^2 < 0$ between your two locations.

- When $\bar{\sigma} = 0$, we say x and x' are null (or light-like) separated. As already alluded to, in 4 dimensional flat spacetime, light travels strictly on null geodesics $\mathrm{d}s^2 = 0$. Consider a coordinate system for spacetime centered at x'; then we would say x lies on the light cone of x' (and vice versa).

As we will soon discover, the indefinite metric of spacetimes – as opposed to the positive definite one of space itself – is what allows for wave solutions, for packets of energy/momentum to travel over space and time. In Minkowski spacetime, we will show below, by solving explicitly the Green's function G_{D+1} of the wave operator, that these waves ψ, subject to eq. (12.4.16), will obey causality: they travel strictly on and/or within the light cone, independent of what the source J is.

Poincaré symmetry Analogous to how rotations $\{R^i{}_a | \delta_{ij} R^i{}_a R^j{}_b = \delta_{ab}\}$ and spatial translations $\{a^i\}$ leave the flat Euclidean metric δ_{ij} invariant,

$$x^i \to R^i{}_j x^j + a^i \qquad \Rightarrow \qquad \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \to \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{12.4.13}$$

(The $R^i{}_j$ and a^i are constants.) Lorentz transformations $\{\Lambda^\alpha{}_\mu | \eta_{\alpha\beta}\Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu = \eta_{\mu\nu}\}$ and spacetime translations $\{a^\mu\}$ are ones that leave the flat Minkowski metric $\eta_{\mu\nu}$ invariant.

$$x^\alpha \to \Lambda^\alpha{}_\mu x^\mu + a^\alpha \qquad \Rightarrow \qquad \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \to \eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu. \tag{12.4.14}$$

(The $\Lambda^\alpha{}_\mu$ and a^α are constants.) This in turn leaves the light cone condition $\mathrm{d}s^2 = 0$ invariant – the speed of light is unity, $|\mathrm{d}\vec{x}|/\mathrm{d}t = 1$, in all inertial frames related via eq. (12.4.14).

Wave Equation In Curved Spacetime The wave equation (for a minimally coupled massless scalar) in some spacetime geometry $g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$ is a 2nd order in time PDE that takes the following form:

$$\nabla_\mu \nabla^\mu \psi = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}g^{\mu\nu}\partial_\nu\psi\right) = J(x), \tag{12.4.15}$$

where J is some specified external source of ψ.

Minkowski We will mainly deal with the case of infinite flat (aka "Minkowski") spacetime in eq. (12.4.1), where in Cartesian coordinates $x^\mu = (ct, \vec{x})$. This leads us to the wave equation

$$\left(\partial_t^2 - c^2\vec{\nabla}_{\vec{x}}^2\right)\psi(t,\vec{x}) = c^2 J(t,\vec{x}), \qquad \vec{\nabla}_{\vec{x}}^2 \equiv \delta^{ij}\partial_i\partial_j. \tag{12.4.16}$$

Here, c will turn out to be the speed of propagation of the waves themselves. Because it will be the most important speed in this chapter, I will set it to unity, $c = 1$.[146] We will work mainly in flat infinite spacetime, which means the $\vec{\nabla}^2$ is the Laplacian in flat space. This equation describes a diverse range of phenomenon, from the vibrations of strings to that of spacetime itself.

2D Minkowski We begin the study of the homogeneous wave equation in 2 dimensions. In Cartesian coordinates (t, z),

$$\left(\partial_t^2 - \partial_z^2\right)\psi(t, z) = 0. \tag{12.4.17}$$

We see that the solutions are a superposition of either left-moving $\psi(z+t)$ or right-moving waves $\psi(z-t)$, where ψ can be any arbitrary function,

$$\left(\partial_t^2 - \partial_z^2\right)\psi(z \pm t) = (\pm)^2\psi''(z \pm t) - \psi''(z \pm t) = 0. \tag{12.4.18}$$

Remark It is worth highlighting the difference between the nature of the general solutions to 2nd order linear homogeneous ODEs versus those of PDEs such as the wave equation here. In the former, they span a 2 dimensional vector space, whereas the wave equation admits arbitrary functions as general solutions. This is why the study of PDEs involve infinite dimensional (oftentimes continuous) Hilbert spaces.

Let us put back the speed c – by dimensional analysis we know $[c]$=[Length/Time], so $x \pm ct$ would yield the correct dimensions.

$$\psi(t, x) = \psi_{\mathrm{L}}(x + ct) + \psi_{\mathrm{R}}(x - ct). \tag{12.4.19}$$

These waves move strictly at speed c.

Problem 12.19. Homogeneous Solutions by Direct Integration Let us define light cone coordinates as $x^\pm \equiv t \pm z$. Write down the Minkowski metric in eq. (12.4.1)

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}z^2 \tag{12.4.20}$$

in terms of $x^\pm$ and show that the wave equation in eq. (12.4.17) is converted to

$$\partial^2\psi = 4\partial_+\partial_-\psi = 0. \tag{12.4.21}$$

By direct integration of eq. (12.4.21), argue that the most general homogeneous wave solution in 2D is the superposition of left- and right-moving (otherwise arbitrary) profiles in eq. (12.4.19). □

Problem 12.20. Inhomogeneous Solutions by Direct Integration Let us examine the 2D wave equation expressed in null coordinates of eq. (12.4.21), but now with an arbitrary source:

$$4\partial_+\partial_-\psi = J. \tag{12.4.22}$$

[146] This is always a good labor-saving strategy when you solve problems. Understand all the distinct dimensionful quantities in your setup – pick the most relevant/important length, time, and mass, etc. Then set them to one, so you don't have to carry their symbols around in your calculations. Every other length, time, mass, etc. will now be respectively, expressed as multiples of them. For instance, now that $c = 1$, the speed(s) $\{v_i\}$ of the various constituents of the source J measured in some center of mass frame, would be measured in multiples of c – for instance, "$v^2 = 0.76$" really means $(v/c)^2 = 0.76$.

Suppose ψ is known on the forward light cone of $(x'^{\pm})$; i.e., $\psi(x^+, x'^-)$ and $\psi(x'^+, x^-)$ are specified for $(x^{\pm} \geq x'^{\pm})$. Show by direct integration that the solution to eq. (12.4.22) is given entirely by these 'light cone boundary conditions':

$$\psi(x^{\pm} > x'^{\pm}) = \psi(x^+, x'^-) + \psi(x'^+, x^-) - \psi(x'^+, x'^-) \\ + \int_{x'^+}^{x^+} \mathrm{d}x''^+ \int_{x'^-}^{x^-} \mathrm{d}x''^- J(x''^+, x''^-). \tag{12.4.23}$$

The first two terms on the right hand side may be viewed as the right- and left-movers; whereas the third term ensures the consistency of the $\psi(x^{\pm} \to x'^{\pm})$ limit. In Problem (12.55) below, you will witness the 'initial value' formulation of this system. □

Problem 12.21. Sound waves on a drum and eigensystem of 2D Laplacian The acoustic (i.e., sound) waves on a drum's surface obeys the 2+1 dimensional PDE

$$\left(\partial_t^2 - c_s^2 \vec{\nabla}_{2\mathrm{D}}^2\right) \psi(t, \vec{x}) = 0, \tag{12.4.24}$$

where c_s is the speed of the sound waves.

We may view ψ as the perpendicular displacement of the drum's 2D surface from its equilibrium position, at a particular location $\vec{x}$. The drum's membrane is usually pinned down at the edges, so we require Dirichlet boundary conditions

$$\psi(t, \vec{x} \in \partial\mathfrak{D}) = 0. \tag{12.4.25}$$

Let us study the normal modes of the drum by focusing on a particular angular frequency ω,

$$\psi(t, \vec{x}) = \mathrm{Re}\left(e^{-i\omega t}\widetilde{\psi}(\omega, \vec{x})\right). \tag{12.4.26}$$

Solve for the set of oscillation frequencies $\{\omega\}$ for a (A) circular, (B) rectangular, (C) triangular, (D) elliptical drum. Can you come up with other solvable shapes? □

Uniqueness of Minkowski solutions Suppose the following initial conditions are given

$$\psi(t = t_0, \vec{x}) = \varphi_0(\vec{x}), \qquad \partial_t\psi(t = t_0, \vec{x}) = \varphi_1(\vec{x}); \tag{12.4.27}$$

and suppose the scalar field ψ or its normal derivative is specified on the spatial boundaries $\partial\mathfrak{D}$,

$$\psi(t, \vec{x} \in \partial\mathfrak{D}) = \varphi_3(\partial\mathfrak{D}), \qquad \text{(Dirichlet)}, \tag{12.4.28}$$
$$\text{or } n^\alpha\nabla_\alpha\psi(t, \vec{x} \in \partial\mathfrak{D}) = \varphi_4(\partial\mathfrak{D}), \qquad \text{(Neumann)}, \tag{12.4.29}$$

where $n^\alpha(\partial\mathfrak{D})$ is the unit outward normal vector. Then, the solution to the wave equation in eq. (12.4.16) – namely, $\partial^2\psi = J$ – is unique.

Proof Without loss of generality, since our wave equation is linear, we may assume the scalar field is real. We then suppose there are two such solutions ψ_1 and ψ_2 obeying the same initial and boundary conditions. The proof is established if we can show, in fact, that ψ_1 has to be equal to ψ_2. Note that the difference $\Psi \equiv \psi_1 - \psi_2$ is subject to the homogeneous wave equation

$$\partial^2\Psi = \ddot{\Psi} - \vec{\nabla}^2\Psi = 0 \tag{12.4.30}$$

since the J cancels out when we subtract the wave equations of $\psi_{1,2}$. For similar reasons the Ψ obeys the initial conditions

$$\Psi(t = t_0, \vec{x}) = 0 \qquad \text{and} \qquad \partial_t\Psi(t = t_0, \vec{x}) = 0, \tag{12.4.31}$$

and the spatial boundary conditions

$$\Psi(t, \vec{x} \in \partial\mathfrak{D}) = 0 \qquad \text{or} \qquad n^\alpha\nabla_\alpha\Psi(t, \vec{x} \in \partial\mathfrak{D}) = 0. \tag{12.4.32}$$

Let us then consider the following integral

$$E(t) \equiv \frac{1}{2}\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\left(\dot{\Psi}^2(t, \vec{x}) + \vec{\nabla}\Psi(t, \vec{x}) \cdot \vec{\nabla}\Psi(t, \vec{x})\right) \tag{12.4.33}$$

[147]as well as its time derivative

$$\partial_t E(t) = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\left(\dot{\Psi}\ddot{\Psi} + \vec{\nabla}\dot{\Psi} \cdot \vec{\nabla}\Psi\right). \tag{12.4.34}$$

We may integrate-by-parts the gradients on the $\dot{\Psi}$,

$$\partial_t E(t) = \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}\dot{\Psi}n^\alpha\nabla_\alpha\Psi + \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\dot{\Psi}\left(\ddot{\Psi} - \vec{\nabla}^2\Psi\right). \tag{12.4.35}$$

By assumption either Ψ or $n^\alpha\nabla_\alpha\Psi$ is zero on the spatial boundary. Either way, the surface integral is zero. In addition, Ψ obeys the homogeneous wave equation. Therefore the right hand side vanishes and we conclude that E is actually a constant in time. Together with the initial conditions $\dot{\Psi}(t = t_0, \vec{x})^2 = 0$ and $\Psi(t = t_0, \vec{x}) = 0$ (which implies $(\vec{\nabla}\Psi(t = t_0, \vec{x}))^2 = 0$), we see that $E(t = t_0) = 0$, and therefore has to remain zero for all subsequent times $t \geq t_0$. Moreover, since $E(t \geq t_0) = 0$ is the integral of the sum of $(D + 1)$ positive terms $\{\dot{\Psi}^2, (\partial_i\Psi)^2\}$, each term must individually vanish, which in turn implies Ψ must be a constant in both space and time. But, since it is zero at the initial time $t = t_0$, it must be in fact zero for $t \geq t_0$. That means $\psi_1 = \psi_2$. □

Remark Armed with the knowledge that the "initial value problem" for the Minkowski spacetime wave equation has a unique solution, we will see how to actually solve it first in Fourier space and then with the retarded Green's function.

[147]The integrand, for Ψ obyeing the homogeneous wave equation, is in fact its energy density. Therefore $E(t)$ is the total energy stored in Ψ at a given time t.

12.4.2 Waves, Initial Value Problem via Fourier, Green's Functions

Dispersion relations, Homogeneous solutions You may guess that any function $f(t, \vec{x})$ in flat (Minkowski) spacetime can be Fourier transformed:

$$f(t, \vec{x}) = \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \widetilde{f}(\omega, \vec{k}) e^{-i\omega t} e^{i\vec{k}\cdot\vec{x}} \qquad \text{(not quite \dots)}, \tag{12.4.36}$$

where

$$k^\mu \equiv (\omega, k^i). \tag{12.4.37}$$

Remember the first component is now the 0th one; so

$$\exp(-ik_\mu x^\mu) = \exp(-i\eta_{\mu\nu}k^\mu x^\nu) = \exp(-i\omega t)\exp(i\vec{k}\cdot\vec{x}). \tag{12.4.38}$$

Furthermore, these plane waves in eq. (12.4.38) obey

$$\partial^2 \exp(-ik_\mu x^\mu) = -k^2 \exp(-ik_\mu x^\mu), \qquad k^2 \equiv k_\mu k^\mu. \tag{12.4.39}$$

This comes from a direct calculation; note that $\partial_\mu(ik_\alpha x^\alpha) = ik_\alpha \delta^\alpha_\mu = ik_\mu$ and similarly $\partial^\mu(ik_\alpha x^\alpha) = ik^\mu$.

$$\partial^2 \exp(-ik_\mu x^\mu) = \partial_\mu \partial^\mu \exp(-ik_\mu x^\mu) = (ik_\mu)(ik^\mu)\exp(-ik_\mu x^\mu). \tag{12.4.40}$$

Therefore, a particular mode $\widetilde{\psi} e^{-ik_\alpha x^\alpha}$ satisfies the homogeneous scalar wave equation in eq. (12.4.16) with $J = 0$ – provided that

$$0 = \partial^2\left(\widetilde{\psi} e^{-ik_\alpha x^\alpha}\right) = -k^2 \widetilde{\psi} e^{-ik_\alpha x^\alpha} \qquad \Rightarrow \qquad k^2 = 0 \qquad \Rightarrow \qquad \omega^2 = \vec{k}^2. \tag{12.4.41}$$

In other words, the two solutions are

$$\widetilde{\psi}(\vec{k}) \exp\left(\pm i|\vec{k}|\left\{t \pm \widehat{k}\cdot\vec{x}\right\}\right), \qquad \widehat{k} \equiv \frac{\vec{k}}{|\vec{k}|}. \tag{12.4.42}$$

The $e^{+i\omega t}$ waves propagate along the $\widehat{k}$ direction; while the $e^{-i\omega t}$ ones along $-\widehat{k}$.

This relationship between the zeroth component of the momentum and its spatial ones, is often known as the *dispersion relation*. Moreover, the positive root

$$\omega = |\vec{k}| \tag{12.4.43}$$

can be interpreted as saying the energy ω of the photon – or, the massless particle associated with ψ obeying eq. (12.4.16) – is equal to the magnitude of its momentum $\vec{k}$.

Therefore, if ψ satisfies the homogeneous wave equation, the Fourier expansion is actually D-dimensional not $(D+1)$ dimensional:

$$\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left(\widetilde{A}(\vec{k}) e^{-i|\vec{k}|t} + \widetilde{B}(\vec{k}) e^{i|\vec{k}|t}\right) e^{i\vec{k}\cdot\vec{x}}. \tag{12.4.44}$$

There are two terms in the parenthesis, one for the positive solution $\omega = +|\vec{k}|$ and one for the negative $\omega = -|\vec{k}|$. For a real scalar field ψ, the $\widetilde{A}$ and $\widetilde{B}$ are related.

$$\begin{aligned}\psi(t, \vec{x})^* = \psi(t, \vec{x}) &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left(\widetilde{A}(\vec{k})^* e^{i|\vec{k}|t} + \widetilde{B}(\vec{k})^* e^{-i|\vec{k}|t} \right) e^{-i\vec{k}\cdot\vec{x}} \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left(\widetilde{B}(-\vec{k})^* e^{-i|\vec{k}|t} + \widetilde{A}(-\vec{k})^* e^{i|\vec{k}|t} \right) e^{i\vec{k}\cdot\vec{x}}. \end{aligned} \tag{12.4.45}$$

Comparing equations (12.4.44) and (12.4.45) indicate $\widetilde{A}(-\vec{k})^* = \widetilde{B}(\vec{k}) \Leftrightarrow \widetilde{A}(\vec{k}) = \widetilde{B}(-\vec{k})^*$. Therefore,

$$\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \left(\widetilde{A}(\vec{k}) e^{-i|\vec{k}|t} + \widetilde{A}(-\vec{k})^* e^{i|\vec{k}|t} \right) e^{i\vec{k}\cdot\vec{x}}. \tag{12.4.46}$$

Note that $\widetilde{A}(\vec{k})$ itself, for a fixed $\vec{k}$, has two independent parts – its real and imaginary portions.[148]

Contrast this homogeneous wave solution against the infinite Euclidean (flat) space case, where $-\vec{\nabla}^2\psi = 0$ does not admit any solutions that are regular everywhere ($\equiv$ does not blow up anywhere), except the $\psi =$ constant solution.

Initial value formulation through mode expansion Unlike the heat/diffusion equation, the wave equation is second order in time. We therefore expect that, to obtain a unique solution to the latter, we have to supply both the initial field configuration and its first time derivative (conjugate momentum). It is possible to see it explicitly through the mode expansion in eq. (12.4.46) – the need for two independent coefficients $\widetilde{A}$ and $\widetilde{A}^*$ to describe the homogeneous solution is intimately tied to the need for two independent initial conditions.

Suppose

$$\psi(t = 0, \vec{x}) = \psi_0(\vec{x}) \qquad \text{and} \qquad \partial_t\psi(t = 0, \vec{x}) = \dot{\psi}_0(\vec{x}), \tag{12.4.47}$$

where the right hand sides are given functions of space. Then, from eq. (12.4.46),

$$\begin{aligned} \psi_0(\vec{x}) &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \widetilde{\psi}_0(\vec{k}) e^{i\vec{k}\cdot\vec{x}} = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \left(\widetilde{A}(\vec{k}) + \widetilde{A}(-\vec{k})^* \right) e^{i\vec{k}\cdot\vec{x}} \\ \dot{\psi}_0(\vec{x}) &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} \widetilde{\dot{\psi}}_0(\vec{k}) e^{i\vec{k}\cdot\vec{x}} = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D k}{(2\pi)^D} (-i|\vec{k}|) \left(\widetilde{A}(\vec{k}) - \widetilde{A}(-\vec{k})^* \right) e^{i\vec{k}\cdot\vec{x}}. \end{aligned} \tag{12.4.48}$$

We have also assumed that the initial field and its time derivative admits a Fourier expansion. By equating the coefficients of the plane waves,

$$\begin{aligned} \widetilde{\psi}_0(\vec{k}) &= \widetilde{A}(\vec{k}) + \widetilde{A}(-\vec{k})^*, \\ \frac{i}{|\vec{k}|} \widetilde{\dot{\psi}}_0(\vec{k}) &= \widetilde{A}(\vec{k}) - \widetilde{A}(-\vec{k})^*. \end{aligned} \tag{12.4.49}$$

[148] In quantum field theory, the coefficients $\widetilde{A}(\vec{k})$ and $\widetilde{A}(\vec{k})^*$ of the Fourier expansion in (12.4.46) will become operators obeying appropriate commutation relations.

Inverting this relationship tells us the $\widetilde{A}(\vec{k})$ and $\widetilde{A}(\vec{k})^*$ are indeed determined by (the Fourier transforms) of the initial conditions:

$$\begin{aligned}\widetilde{A}(\vec{k}) &= \frac{1}{2}\left(\widetilde{\psi}_0(\vec{k}) + \frac{i}{|\vec{k}|}\widetilde{\dot{\psi}}_0(\vec{k})\right) \\ \widetilde{A}(-\vec{k})^* &= \frac{1}{2}\left(\widetilde{\psi}_0(\vec{k}) - \frac{i}{|\vec{k}|}\widetilde{\dot{\psi}}_0(\vec{k})\right) \end{aligned} \tag{12.4.50}$$

In other words, given the initial conditions $\psi(t=0,\vec{x}) = \psi_0(\vec{x})$ and $\partial_t\psi(t=0,\vec{x}) = \dot{\psi}_0(\vec{x})$, we can evolve the homogeneous wave solution forward/backward in time through their Fourier transforms:

$$\begin{aligned}\psi(t,\vec{x}) &= \frac{1}{2}\int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left\{\left(\widetilde{\psi}_0(\vec{k}) + \frac{i}{|\vec{k}|}\widetilde{\dot{\psi}}_0(\vec{k})\right)e^{-i|\vec{k}|t} + \left(\widetilde{\psi}_0(\vec{k}) - \frac{i}{|\vec{k}|}\widetilde{\dot{\psi}}_0(\vec{k})\right)e^{i|\vec{k}|t}\right\}e^{i\vec{k}\cdot\vec{x}} \\ &= \int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left(\widetilde{\psi}_0(\vec{k})\cos(|\vec{k}|t) + \widetilde{\dot{\psi}}_0(\vec{k})\frac{\sin(|\vec{k}|t)}{|\vec{k}|}\right)e^{i\vec{k}\cdot\vec{x}}. \end{aligned} \tag{12.4.51}$$

We see that the initial profile contributes to the part of the field even under time reversal $t \to -t$; whereas its initial time derivative contributes to the portion odd under time reversal.

Suppose the initial field configuration and its time derivative were specified at some other time t_0 (instead of 0),

$$\psi(t=t_0,\vec{x}) = \psi_0(\vec{x}), \qquad \partial_t\psi(t=t_0,\vec{x}) = \dot{\psi}_0(\vec{x}). \tag{12.4.52}$$

Because of time-translation symmetry, eq. (12.4.51) becomes

$$\psi(t,\vec{x}) = \int_{\mathbb{R}^D}\frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\left(\widetilde{\psi}_0(\vec{k})\cos\left(|\vec{k}|(t-t_0)\right) + \widetilde{\dot{\psi}}_0(\vec{k})\frac{\sin\left(|\vec{k}|(t-t_0)\right)}{|\vec{k}|}\right)e^{i\vec{k}\cdot\vec{x}}. \tag{12.4.53}$$

Thus far, we have viewed t_0 as some 'initial time'. But since the solution in eq. (12.4.53) holds for all $t \in \mathbb{R}$, we should really phrase this as an intermediate value problem – namely, given ψ and its velocity at some 'intermediate time' t_0, eq. (12.4.53) then provides the corresponding homogeneous solution $\psi(t,\vec{x})$ throughout the rest of Minkowski spacetime.

Problem 12.22. Let's consider an initial Gaussian wave profile with zero time derivative,

$$\psi(t=0,\vec{x}) = \exp(-(\vec{x}/\sigma)^2), \qquad \partial_t\psi(t=0,\vec{x}) = 0. \tag{12.4.54}$$

If ψ satisfies the homogeneous wave equation, what is $\psi(t>0,\vec{x})$? Express the answer as a Fourier integral; the integral itself may be very difficult to evaluate. □

Inhomogeneous solution in Fourier space If there is a non-zero source J, we could try the strategy we employed with the 1D damped driven simple harmonic oscillator: first go to Fourier space and then inverse-transform it back to position spacetime. That is, starting with,

$$\partial_x^2\psi(x) = J(x), \tag{12.4.55}$$

$$\partial_x^2 \int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \widetilde{\psi}(k) e^{-ik_\mu x^\mu} = \int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \widetilde{J}(k) e^{-ik_\mu x^\mu} \tag{12.4.56}$$

$$\int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} (-k^2) \widetilde{\psi}(k) e^{-ik_\mu x^\mu} = \int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \widetilde{J}(k) e^{-ik_\mu x^\mu}, \qquad k^2 \equiv k_\mu k^\mu. \tag{12.4.57}$$

Because the plane waves $\{\exp(-ik_\mu x^\mu)\}$ are basis vectors, their coefficients on both sides of the equation must be equal.

$$\widetilde{\psi}(k) = -\frac{\widetilde{J}(k)}{k^2}. \tag{12.4.58}$$

The advantage of solving the wave equation in Fourier space is, we see that this is the particular solution for ψ – the portion that is sourced by J. Turn off J and you'd turn off (the inhomogeneous part of) ψ.

Inhomogeneous solution via Green's function We next proceed to transform eq. (12.4.58) back to spacetime.

$$\begin{aligned}\psi(x) &= -\int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \frac{\widetilde{J}(k)}{k^2} e^{-ik\cdot x} = -\int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \int_{\mathbb{R}^{D,1}} \mathrm{d}^{D+1}x'' \frac{J(x'')e^{ik\cdot x''}}{k^2} e^{-ik\cdot x} \\ &= \int_{\mathbb{R}^{D,1}} \mathrm{d}^{D+1}x'' \left(\int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \frac{e^{-ik\cdot(x-x'')}}{-k^2} \right) J(x'')\end{aligned} \tag{12.4.59}$$

That is, if we define the Green's function of the wave operator as

$$\begin{aligned}G_{D+1}(x-x') &= \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \frac{e^{-ik_\mu(x-x')^\mu}}{-k^2} \\ &= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} G_{\text{SHO}}(k; t-t') e^{i\vec{k}\cdot(\vec{x}-\vec{x}')},\end{aligned} \tag{12.4.60}$$

where we have identified – recall the friction-less $\gamma = 0$ limit of eq. (7.8.34) –

$$\widetilde{G}_{\text{SHO}}(k; T \equiv t-t') \equiv \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \frac{e^{-i\omega T}}{-\omega^2 + \vec{k}^2} = \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \frac{e^{-i\omega T}}{2|\vec{k}|} \left(\frac{1}{\omega + |\vec{k}|} - \frac{1}{\omega - |\vec{k}|} \right); \tag{12.4.61}$$

the eq. (12.4.59) translates to

$$\psi(x) = \int_{\mathbb{R}^{D+1}} \mathrm{d}^{D+1}x'' G_{D+1}(x-x'') J(x''). \tag{12.4.62}$$

The Green's function $G_{D+1}(x, x')$ itself satisfies the following PDE:

$$\partial_x^2 G_{D+1}(x,x') = \partial_{x'}^2 G_{D+1}(x,x') = \delta^{(D+1)}(x-x') = \delta(t-t')\delta^{(D)}\left(\vec{x}-\vec{x}'\right). \tag{12.4.63}$$

This is why we call it the Green's function. Like its counterpart for the Poisson equation, we can view G_{D+1} as the inverse of the wave operator. A short calculation using the Fourier representation in eq. (12.4.60) will verify eq. (12.4.63). If ∂^2 denotes the wave operator with respect to either x or x', and if we recall the eigenvalue equation (12.4.39) as well as the integral representation of the δ-function,

$$\partial^2 G_{D+1}(x-x') = \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \frac{\partial^2 e^{-ik_\mu (x-x')^\mu}}{-k^2}$$
$$= \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}^{D+1}k}{(2\pi)^{D+1}} \frac{-k^2 e^{-ik_\mu (x-x')^\mu}}{-k^2} = \delta^{(D+1)}(x-x'). \tag{12.4.64}$$

Relation to the Driven Simple Harmonic Oscillator If we had performed a Fourier transform only in space, notice that eq. (12.4.55) would read

$$\ddot{\widetilde{\psi}}(t,\vec{k}) + \vec{k}^2 \cdot \widetilde{\psi}(t,\vec{k}) = \widetilde{J}(t,\vec{k}). \tag{12.4.65}$$

Comparing this to the driven simple harmonic oscillator equation $\ddot{x} + \Omega^2 x = f$, we may thus identify $\vec{k}^2$ as the frequency-squared Ω^2, and the source $\widetilde{J}$ as the external force; even though the wave equation is relativistic while the SHO is non-relativistic.

Problem 12.23. Each Fourier Mode as a SHO Employ eq. (5.11.194) to explain why, for each $\vec{k}$ mode, the *causal* solution is

$$\widetilde{\psi}(t,\vec{k}) = \int_{-\infty}^{t} \mathrm{d}t' \frac{\sin\left(k(t-t')\right)}{k} \widetilde{J}(t',\vec{k}), \qquad k \equiv |\vec{k}|. \tag{12.4.66}$$

More generally, we see that

$$\widetilde{G}^{\pm}_{\text{SHO}}(t-t') \equiv \pm\Theta\left(\pm(t-t')\right) \frac{\sin\left(k(t-t')\right)}{k} \tag{12.4.67}$$

must correspond to a single Fourier mode of the retarded (+) and advanced (-) Green's functions in eq. (12.4.60). In particular, by performing an inverse Fourier transform, further explain why

$$G^{\pm}_{d}(t-t',\vec{x}-\vec{x}') = \pm\Theta\left(\pm(t-t')\right) \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \frac{\sin\left(k(t-t')\right)}{k} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{12.4.68}$$

Below, we will recover these results by a direct evaluation of the Fourier integrals. □

Initial/Final Conditions for $G^{\pm}_{D+1}$ Let us define $\Delta(x-x')$ to be the difference between the retarded and advanced Green's functions. From eq. (12.4.68),

$$\Delta_{d=D+1}(x-x') \equiv G^{+}_{d}(x-x') - G^{-}_{d}(x-x') \tag{12.4.69}$$
$$= \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \frac{\sin\left(k(t-t')\right)}{k} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}, \tag{12.4.70}$$

where we have used the identity $\Theta(x) + \Theta(-x) = 1$. This Δ obeys the *homogeneous* wave equation, because $\partial^2 G^{\pm} = \delta^{(d)}$.

$$\partial^2 \Delta_d = \partial^2 G^{+}(x-x') - \partial^2 G^{-}(x-x') \tag{12.4.71}$$
$$= \delta^{(d)}(x-x') - \delta^{(d)}(x-x') \tag{12.4.72}$$

$$= 0. \tag{12.4.73}$$

At this point, we may gather

$$\Delta_{D+1}(t = t', \vec{x} - \vec{x}') = 0 \qquad \text{and} \qquad \partial_t \Delta_{D+1}(t = t', \vec{x} - \vec{x}') = \delta^{(D)}(\vec{x} - \vec{x}'). \tag{12.4.74}$$

We may turn this result around, and use eq. (12.4.74) and $\partial^2 \Delta = 0$ to *define* the retarded or advanced Green's function via the relations

$$G^{\pm}(x - x') \equiv \pm\Theta\left(\pm(t - t')\right)\Delta(x - x'). \tag{12.4.75}$$

In this sense, eq. (12.4.74) may be viewed as the initial conditions at $t = t'$ for the retarded Green's function G^+, since $\Theta(t - t') = 1$ for $t > t'$; or, as final conditions for the advanced Green's function G^-, since $\Theta(t' - t) = 1$ for $t' > t$. Defined in this manner, the uniqueness of the initial/final value problem tells us the retarded Green's function itself is thus unique. In Problem (12.52) below, we will re-visit this initial value formulation for retarded Green's functions, but in curved spacetimes.

We may further define the symmetric Green's function $G^S = G^+ + G^-$ as the sum of the retarded and advanced Green's functions.

$$G^S(x - x') = \left(\Theta(t - t') - \Theta(t' - t)\right) \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} \frac{\sin\left(k(t - t')\right)}{k} e^{i\vec{k}\cdot(\vec{x} - \vec{x}')}. \tag{12.4.76}$$

The $t = t'$ limit sets the sine to zero.

$$G^S(t = t', \vec{x} - \vec{x}') = 0 \tag{12.4.77}$$

The first derivative would in time at $t = t'$ would yield $\delta(t - t')$ terms multiplied by $\sin(k(t - t'))$ as well as $(\Theta(0) - \Theta(0))\cos(k \cdot 0)$.

$$G^S(t = t', \vec{x} - \vec{x}') = 0 \tag{12.4.78}$$

Observer and Source, G_{D+1} as a field by a point source If we compare $\delta^{(D+1)}(x - x')$ in the wave equation obeyed by the Green's function itself (eq. (12.4.63)) with that of an external source J in the wave equation for ψ (eq. (12.4.55)), we see $G_{D+1}(x, x')$ itself admits the interpretation that it is the field observed at the spacetime location x produced by a spacetime point source at $x' \equiv (t', \vec{x}')$. Note that this interpretation is also consistent with the initial conditions in eq. (12.4.74).

Furthermore, according to eq. (12.4.62), the $\psi(t, \vec{x})$ is then the superposition of the fields due to all such spacetime points, weighted by the physical source J. (For a localized J, it sweeps out a world tube in spacetime – try drawing a spacetime diagram to show how its segments contribute to the signal at a given x.)

Contour prescriptions and causality From your experience with the mode sum expansion you may already have guessed that the Green's function for the wave operator ∂^2, obeying eq. (12.4.63), admits the mode sum expansion in eq. (12.4.60). However, you will soon run into a stumbling block if you begin with the $k^0 = \omega$ integral, because the denominator of the integrand of eq. (12.4.61) gives rise to two singularities on the real line at $\omega = \pm|\vec{k}|$. To ensure the mode expansion in eq. (12.4.60) is well defined, we would need to append to it an appropriate contour prescription for the ω-integral. It will turn out that, each distinct contour prescription will give rise to a Green's function with distinct causal properties.

On the complex ω-plane, 3 of the most commonly used contour prescriptions we can choose to avoid the singularities at $\omega = \pm|\vec{k}|$ are:

1. Making a tiny semi-circular clockwise contour around each of them; ensuring all the singularities of the integrand lie *below* the contour. For the sign convention $e^{-i\omega(t-t')}$, this will yield the *retarded Green's function* G^+_{D+1}, where signals from the source propagate forward in time; observers will see signals only from the past. For, when $t < t'$, we need to close the contour in eq. (12.4.61) in the upper half imaginary $\omega-$plane (recall $e^{-i(i\omega)(t-t')} = 0$) but because the integrand is analytic there, the result is zero. When $t > t'$, on the other hand, we need to close the contour in eq. (12.4.61) in the lower half imaginary $\omega-$plane (recall $e^{-i(-i\omega)(t-t')} = 0$). We thus obtain
$$\widetilde{G}^+_{\text{SHO}}(T) = -\frac{i\Theta(T)}{2|\vec{k}|}\left(e^{i|\vec{k}|T} - e^{-i|\vec{k}|T}\right) = \Theta(T)\frac{\sin\left(|\vec{k}|T\right)}{|\vec{k}|}. \tag{12.4.79}$$
Altogether, this recovers the + result in eq. (12.4.67).

2. Making a tiny semi-circular counterclockwise contour around each of them; ensuring all the singularities of the integrand lie *above* the contour. For the sign convention $e^{-i\omega(t-t')}$, this will yield the *advanced Green's function* G^-_{D+1}, where signals from the source propagate backward in time; observers will see signals only from the future. For, when $t > t'$, we need to close the contour in eq. (12.4.61) in the lower half imaginary $\omega-$plane (recall $e^{-i(-i\omega)(t-t')} = 0$) but because the integrand is analytic there, the result is zero. When $t < t'$, on the other hand, we need to close the contour in eq. (12.4.61) in the upper half imaginary $\omega-$plane (recall $e^{-i(i\omega)(t-t')} = 0$). We thus obtain
$$\widetilde{G}^-_{\text{SHO}}(T) = +\frac{i\Theta(-T)}{2|\vec{k}|}\left(e^{i|\vec{k}|T} - e^{-i|\vec{k}|T}\right) = -\Theta(-T)\frac{\sin\left(|\vec{k}|T\right)}{|\vec{k}|}. \tag{12.4.80}$$
Altogether, this recovers the − result in eq. (12.4.67).

3. Making a tiny semi-circular counterclockwise contour around $\omega = -|\vec{k}|$ and a clockwise one at $\omega = +|\vec{k}|$. When $t > t'$ we need to close the contour in eq. (12.4.61) in the lower half imaginary $\omega-$plane and obtain
$$\widetilde{G}^{\text{F}}_{\text{SHO}}(T > 0) = \frac{i \cdot e^{-i|\vec{k}|T}}{2|\vec{k}|} = \frac{1}{2|\vec{k}|}\left(\sin\left(|\vec{k}|T\right) + i\cos\left(|\vec{k}|T\right)\right); \tag{12.4.81}$$
whereas when $t < t'$ we need to close the contour in eq. (12.4.61) in the upper half imaginary $\omega-$plane and obtain
$$\widetilde{G}^{\text{F}}_{\text{SHO}}(T < 0) = \frac{i \cdot e^{i|\vec{k}|T}}{2|\vec{k}|} = \frac{1}{2|\vec{k}|}\left(-\sin\left(|\vec{k}|T\right) + i\cos\left(|\vec{k}|T\right)\right); \tag{12.4.82}$$

so altogether (cf. eq. (12.4.67))

$$\widetilde{G}^{\mathrm{F}}_{\mathrm{SHO}}(T) = \frac{1}{2}\widetilde{G}^{+}_{\mathrm{SHO}}(T) + \frac{1}{2}\widetilde{G}^{-}_{\mathrm{SHO}}(T) + \frac{i}{2|\vec{k}|}\cos\left(|\vec{k}|T\right). \tag{12.4.83}$$

Upon integrating over the Fourier $\vec{k}$ modes, namely

$$G^{\mathrm{F}}_{D+1}(x-x') = \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\widetilde{G}^{\mathrm{F}}_{\mathrm{SHO}}(T)e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}, \tag{12.4.84}$$

this $G^{\mathrm{F}}(x-x')$ will yield the Feynman(-like) Green's function, named after the theoretical physicist Richard P. Feynman. The Feynman Green's function is used heavily in Minkowski spacetime perturbative Quantum Field Theory, though in QFT it is $-i\hbar$ times the object here (i.e., $-i\hbar G^{\mathrm{F}}(x-x')$) that is usually called the Feynman Green's function. Unlike its retarded and advanced cousins – which are purely real – the Feynman Green's function is complex. If the symmetric Green's function G^S is defined as the sum of the retarded and the advanced Green's functions, we see from eq. (12.4.83) that

$$\mathrm{Re}\ G^{\mathrm{F}}_{D+1}(x-x') = \frac{1}{2}G^{+}_{D+1}(x-x') + \frac{1}{2}G^{-}_{D+1}(x-x') \tag{12.4.85}$$

$$= \frac{1}{2}G^{S}_{D+1}(x-x'). \tag{12.4.86}$$

The imaginary part of G^{F}_{D+1}, in turn, is given by

$$\mathrm{Im}\ G^{\mathrm{F}}_{D+1}(x-x') = \frac{1}{2}\int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D}\frac{\cos\left(|\vec{k}|T\right)}{|\vec{k}|}e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{12.4.87}$$

You may readily verify, that the right hand side is in fact real is the result of its invariance under the parity transformation $\vec{x}-\vec{x}' \leftrightarrow \vec{x}'-\vec{x}$. (Notice this expression does not converge when $D=1$ due to the $\vec{k}\to\vec{0}$ singularity of its integrand.) In the quantum field theory context, $\mathrm{Im}\ G^{\mathrm{F}}_{D+1}(x-x')$ describes particle creation by an external source.

Below, we will also examine the heat kernel representation of the Green's function of the Minkowski spacetime wave operator, i.e., the generalization of eq. (12.3.54) to "spacetime Laplacians". The subtlety here is that the eigenvalues of ∂^2, the $\{-k^2\}$, are not positive definite; to ensure convergence of the proper time t-integral in eq. (12.3.54) one would in fact be lead to the Feynman Green's function.

For classical physics, we will focus mainly on the retarded Green's function G^{+}_{D+1} because it obeys causality – the cause (the source J) precedes the effect (the field it generates). We will see this explicitly once we work out the G^{+}_{D+1} below, for all $D \geq 1$.

1+1 Dimensions: Retarded Case Let us tackle the 2 dimensional case. The retarded Green's function is

$$G^{+}_{2}\left(x-x' \equiv (t,z)\right) = \Theta(t)\int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\frac{\sin(kt)}{k}e^{ikz}. \tag{12.4.88}$$

Let's now observe that this integral is invariant under the replacement $z \to -z$. In fact,

$$G_2^+(t,-z) = \Theta(t)\int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\frac{e^{-ikz}}{k}\sin(kt) = G_2^+(t,z)^* \tag{12.4.89}$$

$$= \Theta(t)\int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\frac{e^{ikz}}{-k}\sin(-kt) = G_2^+(t,z)\,. \tag{12.4.90}$$

Therefore not only is $G_2^+(t,z)$ real, we can also put an absolute value around the z – the answer for G_2^+ has to be the same whether z is positive or negative anyway.

$$G_2^+(t,z) = \Theta(t)\int_{\mathbb{R}} \frac{\mathrm{d}k}{2\pi}\frac{e^{ik|z|}}{k}\sin(kt) \tag{12.4.91}$$

Note that the integrand $\exp(ik|z|)\sin(kt)/k$ is smooth on the entire real $k-$line. Therefore, if we view this integral as one on the complex $k-$plane, we may displace the contour slightly 'upwards' towards the positive imaginary axis:

$$G_2^+(t,z) = \frac{\Theta(t)}{2}\int_{-\infty+i0^+}^{+\infty+i0^+} \frac{\mathrm{d}k}{2\pi i}\frac{e^{ik|z|}}{k}\left(e^{ikt}-e^{-ikt}\right) \tag{12.4.92}$$

$$= \frac{\Theta(t)}{2}(-)\left(\Theta(-t-|z|)-\Theta(t-|z|)\right) \tag{12.4.93}$$

$$= \frac{1}{2}\Theta\left(t-|z|\right). \tag{12.4.94}$$

Problem 12.24. Can you explain why

$$\Theta(t)\Theta(t^2-z^2) = \Theta(t-|z|)? \tag{12.4.95}$$

Re-write $\Theta(-t)\Theta(t^2-z^2)$ as a single step function. □

We have arrived at the solution

$$G_2^+(x-x') = \frac{1}{2}\Theta(t-t')\Theta(\bar{\sigma}) = \frac{1}{2}\Theta\left(t-t'-|z-z'|\right), \tag{12.4.96}$$

$$\bar{\sigma} \equiv \frac{(t-t')^2-(z-z')^2}{2} = \frac{1}{2}(x-x')^2. \tag{12.4.97}$$

While the $\Theta(\bar{\sigma})$ allows the signal due to the spacetime point source at x' to propagate both forward and backward in time – actually, throughout the interior of the light cone of x' – the $\Theta(t-t')$ implements retarded boundary conditions: the observer time t always comes after the emission time t'. If you carry out a similar analysis for G_2 but for the advanced contour, you would find

$$G_2^-(x-x') = \frac{1}{2}\Theta(t'-t)\Theta(\bar{\sigma}). \tag{12.4.98}$$

Problem 12.25. (2+1)D From its Fourier representation , calculate $G_3^\pm(x-x')$, the retarded and advanced Green's function of the wave operator in 3 dimensional Minkowski spacetime. You should find

$$G_3^\pm(x-x') = \frac{\Theta(\pm(t-t'))}{\sqrt{2}(2\pi)}\frac{\Theta(\bar{\sigma})}{\sqrt{\bar{\sigma}}}. \tag{12.4.99}$$

Bonus problem: Can you perform the Fourier integral in eq. (12.4.60) for all G_{D+1}?

Hints: For the (2+1)D case, denoting $T \equiv t - t'$ and $R \equiv |\vec{x} - \vec{x}'|$, first show that

$$G_3^+(x - x') = \Theta(T) \int_0^{+\infty} \frac{\mathrm{d}k}{2\pi^2} \sin[kT] \int_{-1}^{+1} \mathrm{d}\chi \frac{\cos[kR\chi]}{\sqrt{1-\chi^2}}. \tag{12.4.100}$$

Now, view R on the right-hand-sides of equations (12.4.99) and (12.4.100) as a real quantity. Show that their Fourier transforms are the same; namely, demonstrate that

$$\int_{\mathbb{R}} \mathrm{d}R e^{i\omega R} \Theta(T) \int_0^{+\infty} \frac{\mathrm{d}k}{2\pi^2} \sin[kT] \int_{-1}^{+1} \mathrm{d}\chi \frac{\cos[kR\chi]}{\sqrt{1-\chi^2}} = \int_{\mathbb{R}} \mathrm{d}R e^{i\omega R} \frac{\Theta[T]\Theta[T^2 - R^2]}{2\pi\sqrt{T^2 - R^2}} \tag{12.4.101}$$

$$= \frac{\Theta[T]}{2} J_0[\omega T]; \tag{12.4.102}$$

where J_0 is the Bessel function. You may find this page useful. Since the Fourier transform is invertible, explain why this proves eq. (12.4.99). □

Problem 12.26. Initial/Final Conditions for G^+_{D+1}/G^-_{D+1} from Fourier Space Starting from

$$\Delta(x - x') = -\int_C \frac{\mathrm{d}\omega}{2\pi} \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} \frac{e^{-i\omega(t-t')} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}}{\omega^2 - \vec{k}^2}; \tag{12.4.103}$$

derive the correct contour prescription for the $\omega-$integral that would yield the result in eq. (12.4.70). This is the analogy of eq. (12.4.60) for the retarded/advanced Green's functions.

Then, compute $\Delta(t = t')$ and $\partial_t \Delta(t - t')$ – set $t = t'$ *before* evaluating the ω integral – and verify they yield (12.4.74). □

Green's Functions From Recursion Relations With the 2 and 3 dimensional Green's function under our belt, I will now show how we can generate the Green's function of the Minkowski wave operator in all dimensions, just by differentiating $G_{2,3}$. The primary observation that allow us to do so, is that a line source in $(D+2)$ spacetime is a point source in $(D+1)$ dimensions; and a plane source in $(D+2)$ spacetime is a point source in D dimensions.[149]

For this purpose let's set the notation. In $(D+1)$ dimensional flat spacetime, let the spatial coordinates be denoted as $x^i = (\vec{x}_\perp, w^1, w^2)$; and in $(D-1)$ dimensions let the spatial coordinates be the $\vec{x}_\perp$. Then $|\vec{x} - \vec{x}'|$ is a D dimensional Euclidean distance between the observer and source in the former, whereas $|\vec{x}_\perp - \vec{x}'_\perp|$ is the $D - 1$ counterpart in the latter.

Starting from the integral representation for G_{D+1} in eq. (12.4.60), we may integrate with respect to the Dth spatial coordinate w^2:

$$\int_{-\infty}^{+\infty} \mathrm{d}w'^2 G_{D+1}(t - t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}')$$
$$= \int_{-\infty}^{+\infty} \mathrm{d}w'^2 \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}\omega \mathrm{d}^{D-2}k_\perp \mathrm{d}^2 k_\parallel}{(2\pi)^{D+1}} \frac{e^{-i\omega(t-t')} e^{i\vec{k}_\perp\cdot(\vec{x}_\perp - \vec{x}'_\perp)} e^{ik_\parallel\cdot(\vec{w}-\vec{w}')}}{-\omega^2 + \vec{k}_\perp^2 + \vec{k}_\parallel^2}$$

[149] I will make this statement precise very soon, but you are encouraged to read Soodak and Tiersten [44] for a pedagogical treatment.

$$
\begin{aligned}
&= \int_{\mathbb{R}^{D+1}} \frac{\mathrm{d}\omega \mathrm{d}^{D-2}k_\perp \mathrm{d}^2 k_\parallel}{(2\pi)^{D+1}} (2\pi)\delta(k_\parallel^2) \frac{e^{-i\omega(t-t')}e^{i\vec{k}_\perp\cdot(\vec{x}_\perp - \vec{x}'_\perp)}e^{ik_\parallel^1(w^1-w'^1)}e^{ik_\parallel^2 w^2}}{-\omega^2 + \vec{k}_\perp^2 + \vec{k}_\parallel^2} \\
&= \int_{\mathbb{R}^{D}} \frac{\mathrm{d}\omega \mathrm{d}^{D-2}k_\perp \mathrm{d}k_\parallel^1}{(2\pi)^{D}} \frac{e^{-i\omega(t-t')}e^{i\vec{k}_\perp\cdot(\vec{x}_\perp - \vec{x}'_\perp)}e^{ik_\parallel^1(w^1-w'^1)}}{-\omega^2 + \vec{k}_\perp^2 + (k_\parallel^1)^2} \\
&= G_D(t-t', \vec{x}_\perp - \vec{x}'_\perp, w^1 - w'^1). \qquad (12.4.104)
\end{aligned}
$$

The notation is cumbersome, but the math can be summarized as follows. Integrating G_{D+1} over the Dth spatial coordinate amounts to discarding the momentum integral with respect to its D component and setting its value to zero everywhere in the integrand. But that is nothing but the integral representation of G_D. Moreover, because of translational invariance, we could have integrated with respect to either w'^2 or w^2. If we compare our integral here with eq. (12.4.62), we may identify $J(x'') = \delta(t''-t')\delta^{(D-2)}(\vec{x}'_\perp - \vec{x}''_\perp)\delta(w^1 - w''^1)$, an instantaneous line source of unit strength lying parallel to the Dth axis, piercing the $(D-1)$ space at $(\vec{x}'_\perp, w'^1)$.

We may iterate this integral recursion relation once more,

$$
\int_{\mathbb{R}^2} \mathrm{d}^2 w G_{D+1}\left(t-t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}'\right) = G_{D-1}\left(t-t', \vec{x}_\perp - \vec{x}'_\perp\right). \qquad (12.4.105)
$$

This is saying G_{D-1} is sourced by a 2D plane of unit strength, lying in $(D+1)$ spacetime. On the left hand side, we may employ cylindrical coordinates to perform the integral

$$
2\pi \int_0^\infty \mathrm{d}\rho\rho G_{D+1}\left(t-t', \sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + \rho^2}\right) = G_{D-1}\left(t-t', |\vec{x}_\perp - \vec{x}'_\perp|\right), \qquad (12.4.106)
$$

where we are now highlighting the fact that, the Green's function really has only two arguments: one, the time elapsed $t-t'$ between observation t and emission t'; and two, the Euclidean distance between observer and source. (We will see this explicitly very shortly.) For G_{D+1} the relevant Euclidean distance is

$$
|\vec{x} - \vec{x}'| = \sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + (\vec{w} - \vec{w}')^2}. \qquad (12.4.107)
$$

A further change of variables

$$
R' \equiv \sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + \rho^2} \qquad \Rightarrow \qquad \mathrm{d}R' = \frac{\rho \mathrm{d}\rho}{R'}. \qquad (12.4.108)
$$

This brings us to

$$
2\pi \int_R^\infty \mathrm{d}R' R' G_{D+1}(t-t', R') = G_{D-1}(t-t', R), \qquad R \equiv |\vec{x}_\perp - \vec{x}'_\perp|. \qquad (12.4.109)
$$

At this point we may differentiate both sides with respect to R (see Leibniz's rule for differentiation), to obtain the Green's function in $(D+1)$ dimensions from its counterpart in $(D-1)$ dimensions.

$$
G_{D+1}(t-t', R) = -\frac{1}{2\pi R}\frac{\partial}{\partial R} G_{D-1}(t-t', R). \qquad (12.4.110)
$$

The meaning of R on the left hand side is the D-space length $|\vec{x} - \vec{x}'|$; on the right hand side it is the $(D-2)$-space length $|\vec{x}_\perp - \vec{x}'_\perp|$.

Problem 12.27. Massive Scalar Green's Functions Explain why the solution to

$$\left(\partial^2 + m^2\right) G_{D+1}[x - x'; m] = \delta^{(D+1)}[x - x'], \tag{12.4.111}$$

for $m > 0$, is given by the Fourier transform

$$G_{D+1}[x - x'; m] = \int_{\mathbb{R}^{D,1}} \frac{\mathrm{d}\omega \mathrm{d}^D \vec{k}}{(2\pi)^{D+1}} \frac{e^{-i\omega \cdot T} e^{i\vec{k}\cdot\vec{R}}}{-\omega^2 + \vec{k}^2 + m^2}, \tag{12.4.112}$$

$$T \equiv t - t', \qquad \vec{R} \equiv \vec{x} - \vec{x}'. \tag{12.4.113}$$

What are the appropriate contours for the retarded and advanced Green's functions? (Hint: It is the same as the massless case – why?) In particular, demonstrate that the generalization of equations (12.4.79), (12.4.80), and (12.4.83) are, respectively

$$\widetilde{G}^+_{\text{SHO}}(T) = \Theta(T) \frac{\sin\left(\sqrt{\vec{k}^2 + m^2} T\right)}{\sqrt{\vec{k}^2 + m^2}}, \tag{12.4.114}$$

$$\widetilde{G}^-_{\text{SHO}}(T) = -\Theta(T) \frac{\sin\left(\sqrt{\vec{k}^2 + m^2} T\right)}{\sqrt{\vec{k}^2 + m^2}}, \tag{12.4.115}$$

and

$$\widetilde{G}^{\text{F}}_{\text{SHO}}(T) = \frac{1}{2}\widetilde{G}^+_{\text{SHO}}(T) + \frac{1}{2}\widetilde{G}^-_{\text{SHO}}(T) + \frac{i}{2\sqrt{\vec{k}^2 + m^2}} \cos\left(\sqrt{\vec{k}^2 + m^2} T\right). \tag{12.4.116}$$

Explain why eq. (12.4.110) also applies to the massive scalar Green's functions. Proceed to evaluate eq. (12.4.112) for $D = 1, 2$. Hint: The answers can be found in equations (12.4.139) and (12.4.140) below. $\square$

Problem 12.28. 'Dimension-Raising' Operator Since the Green's functions in Minkowski are spacetime translationally invariant, we may set $(t', \vec{x}') = (0, \vec{0})$ and view their wave equations as

$$\mathcal{W}_D G_D(t, r) \equiv \partial_t^2 G_D - \frac{\partial_r (r^{D-1} \partial_r G_D(t, r))}{r^{D-1}} = \frac{\delta(r - 0^+)}{r^{D-1} \Omega_D}, \tag{12.4.117}$$

$$\Omega_D = \frac{2\pi^{D/2}}{\Gamma(D/2)}. \tag{12.4.118}$$

The Ω_D is the solid angle subtended by a $D - 1$ sphere; see Problem (7.4). By applying the operator $-(2\pi r)^{-1}\partial_r$ on both sides of eq. (12.4.117), show that

$$\mathcal{W}_{D+2} \left(-\frac{1}{2\pi r} \partial_r G_D(t, r)\right) = \frac{\delta(r - 0^+)}{r^{D+1} \Omega_{D+2}}. \tag{12.4.119}$$

That is, $-(2\pi r)^{-1}\partial_r G_D$ is the solution to the Green's function equation in two higher dimensions – and, hence, may be regarded (in this specific sense) as a 'dimension-raising' operator. This provides an additional confirmation of the relation in eq. (12.4.110). $\square$

Green's Functions From Extra Dimensional Line Source There is an alternate means of obtaining the integral relation in eq. (12.4.104), which was key to deriving eq. (12.4.110). In particular, it does not require explicit use of the Fourier integral representation. Let us postulate that G_D is sourced by a "line charge" $J(w^2)$ extending in the extra spatial dimension of $\mathbb{R}^{D,1}$.

$$G_D(t-t', \vec{x}_\perp - \vec{x}'_\perp, w^1 - w'^1) \stackrel{?}{=} \int_{-\infty}^{+\infty} \mathrm{d}w'^2 G_{D+1}(t-t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}') J(w'^2) \qquad (12.4.120)$$

Applying the wave operator in the $((D-1)+1)$-space on the right hand side, and suppressing arguments of the Green's function whenever convenient,

$$\begin{aligned}
&\partial_D^2 \int_{-\infty}^{+\infty} \mathrm{d}w'^2 G_{D+1} \cdot J \qquad \qquad \left(\text{where } \partial_D^2 \equiv \partial_{t'}^2 - \sum_{i=1}^{D-1} \partial_{i'}^2\right) \\
&\quad = \int_{-\infty}^{+\infty} \mathrm{d}w'^2 J(w'^2) \left(\partial_D^2 - \left(\frac{\partial}{\partial w'^2}\right)^2 + \left(\frac{\partial}{\partial w'^2}\right)^2\right) G_{D+1}(w^2 - w'^2) \\
&\quad = \int_{-\infty}^{+\infty} \mathrm{d}w'^2 J(w'^2) \left(\partial_{D+1}^2 + \left(\frac{\partial}{\partial w'^2}\right)^2\right) G_{D+1}(w^2 - w'^2) \\
&\quad = \int_{-\infty}^{+\infty} \mathrm{d}w'^2 J(w'^2) \left(\delta(t-t')\delta^{(D-2)}(\vec{x}_\perp - \vec{x}'_\perp)\delta^{(2)}(\vec{w} - \vec{w}') + \left(\frac{\partial}{\partial w'^2}\right)^2 G_{D+1}(w^2 - w'^2)\right) \\
&\quad = \delta^{(D-1)}(x-x')\delta(w^1 - w'^1) J(w^2) \\
&\qquad\qquad + \left[J(w'^2) \frac{\partial G_{D+1}(w^2 - w'^2)}{\partial w'^2}\right]_{w'^2=-\infty}^{w'^2=+\infty} - \left[\frac{\partial J(w'^2)}{\partial w'^2} G_{D+1}(w^2 - w'^2)\right]_{w'^2=-\infty}^{w'^2=+\infty} \\
&\qquad\qquad + \int_{-\infty}^{+\infty} \mathrm{d}w'^2 J''(w'^2) G_{D+1}(w^2 - w'^2).
\end{aligned} \qquad (12.4.121)$$

That is, we would have verified the $((D-1)+1)$ flat space wave equation is satisfied if only the first term in the final equality survives. Moreover, that it needs to yield the proper δ-function measure, namely $\delta^{(D-1)}(x-x')\delta(w^1 - w'^1)$, translates to the boundary condition on J:

$$J(w^2) = 1. \qquad (12.4.122)$$

That the second and third terms of the final equality of eq. (12.4.121) are zero, requires knowing causal properties of the Green's function: in particular, because the $w'^2 = \pm\infty$ limits correspond to sources infinitely far away from the observer at $(\vec{x}_\perp, w^1, w^2)$, they must lie outside the observer's light cone, where the Green's function is identically zero. The final term of eq. (12.4.121) is zero if the source obeys the ODE

$$0 = J''(w'^2). \qquad (12.4.123)$$

The solution to eq. (12.4.123), subject to eq. (12.4.122), is

$$J(w'^2) = \cos^2\vartheta + \frac{w'^2}{w^2}\sin^2\vartheta. \qquad (12.4.124)$$

Choosing $\vartheta = 0$ and $\vartheta = \pi/2$ would return, respectively,

$$J(w'^2) = 1 \qquad \text{and} \qquad J(w'^2) = \frac{w'^2}{w^2}. \tag{12.4.125}$$

To sum, we have deduced the Green's function in D spacetime dimensions G_D may be sourced by a line source of a one-parameter family of charge densities extending in the extra spatial dimension of $\mathbb{R}^{D+1,1}$:

$$G_D(t-t', \vec{x}_\perp - \vec{x}'_\perp, w^1 - w'^1) = \int_{-\infty}^{+\infty} \mathrm{d}w'^2 \left(\cos^2\vartheta + \frac{w'^2}{w^2}\sin^2\vartheta\right) \tag{12.4.126}$$

$$\times G_{D+1}(t-t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}'). \tag{12.4.127}$$

Using the simpler expressions in eq. (12.4.125), we obtain

$$G_D(t-t', \vec{x}_\perp - \vec{x}'_\perp, w^1 - w'^1) = \int_{-\infty}^{+\infty} \mathrm{d}w'^2 G_{D+1}(t-t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}') \tag{12.4.128}$$

$$= \int_{-\infty}^{+\infty} \mathrm{d}w'^2 \frac{w'^2}{w^2} G_{D+1}(t-t', \vec{x}_\perp - \vec{x}'_\perp, \vec{w} - \vec{w}') \tag{12.4.129}$$

As a reminder, $\vec{x}_\perp$ and $\vec{x}'_\perp$ are $D-1$ dimensional spatial coordinates; whereas $\vec{w}$ and $\vec{w}'$ are two dimensional ones.

Problem 12.29. 2D from 3D Green's Functions As a consistency check, show that

$$\int_{\mathbb{R}} \mathrm{d}x'^2 G_3^{\pm}(x-x') = G_2^{\pm}(x-x'). \tag{12.4.130}$$

The (1+1)D and (2+1)D Green's functions can be found in equations (12.4.96), (12.4.98), and (12.4.99). □

$G^{\pm}_{D+1}$ in all dimensions, Causal structure of physical signals At this point we may gather $G^{\pm}_{2,3}$ in equations (12.4.96), (12.4.98), and (12.4.99) and apply to them the recursion relation in eq. (12.4.110) to record the explicit expressions of the retarded G^{+}_{D+1} and advanced G^{-}_{D+1} Green's functions in all ($D \geq 2$) dimensions.[150]

- In even dimensional spacetimes, $D+1 = 2+2n$ and $n = 0, 1, 2, 3, 4, \dots$,

$$G^{\pm}_{2+2n}(x-x') = \Theta\left(\pm(t-t')\right)\left(\frac{1}{2\pi}\frac{\partial}{\partial\bar{\sigma}}\right)^n \frac{\Theta(\bar{\sigma})}{2}. \tag{12.4.131}$$

Equivalently,

$$G^{\pm}_{2+2n}(T \equiv t-t', R \equiv |\vec{x}-\vec{x}'|) = \left(-\frac{1}{2\pi}\frac{\partial}{\partial R}\right)^n \frac{\Theta(\pm T - R)}{2}. \tag{12.4.132}$$

[150]When eq. (12.4.110) applied to $G^{\pm}_{2,3}$ in equations (12.4.96), (12.4.98), and (12.4.99), note that the $(2\pi R)^{-1}\partial_R$ passes through the $\Theta(\pm(t-t'))$ and because the rest of the $G^{\pm}_{2,3}$ depends solely on $\bar{\sigma}$, it becomes $-(2\pi R)^{-1}\partial_R = (2\pi)^{-1}\partial_{\bar{\sigma}}$.

- In odd dimensional spacetime, $D+1=3+2n$ and $n=0,1,2,3,4,\dots$,

$$G^{\pm}_{3+2n}(x-x') = \Theta\left(\pm(t-t')\right)\left(\frac{1}{2\pi}\frac{\partial}{\partial\bar{\sigma}}\right)^{n}\left(\frac{\Theta(\bar{\sigma})}{2\pi\sqrt{2\bar{\sigma}}}\right). \tag{12.4.133}$$

Equivalently,

$$G^{\pm}_{3+2n}(T \equiv t-t', R \equiv |\vec{x}-\vec{x}'|) = \left(-\frac{1}{2\pi}\frac{\partial}{\partial R}\right)^{n}\left(\frac{\Theta(\pm T - R)}{2\pi\sqrt{T^2-R^2}}\right). \tag{12.4.134}$$

Recall that $\bar{\sigma}(x,x')$ is half the square of the geodesic distance between the observer at x and point source at x',

$$\bar{\sigma} \equiv \frac{1}{2}(x-x')^2 = \frac{1}{2}(t-t')^2 - \frac{1}{2}(\vec{x}-\vec{x}')^2. \tag{12.4.135}$$

Hence, $\Theta(\bar{\sigma}) = \Theta(T^2-R^2)$ is unity inside the light cone $|T|>R$ and zero outside $|T|<R$; whereas $\delta(\bar{\sigma}) = 2\delta(T^2-R^2)$ and its derivatives are non-zero strictly on the light cone $|T|=R$. Note that the inside-the-light-cone portion of a signal – i.e., the $\Theta(\bar{\sigma})$ term of the Green's function – is known as the tail. Notice too, the $\Theta(\pm(t-t'))$ multiplies an expression that is symmetric under interchange of observer and source ($x \leftrightarrow x'$); namely, $G^{\pm}(x-x') = \Theta(\pm(t-t'))\mathcal{G}(x-x')$. For a fixed source at x', we may interpret these coefficients of $\Theta(\pm(t-t'))$ as the symmetric Green's function $\mathcal{G}(x-x')$: the field due to the source at x' travels both backwards and forward in time. The retarded $\Theta(t-t')$ (observer time is later than emission time) selects the future light cone $T \geq R$ portion of this symmetric signal; while the advanced $\Theta(-(t-t'))$ (observer time earlier than emission time) selects the backward light cone $T \leq -R$ part of it. The last two statements also explain why we may replace the $\Theta(\pm(t-t'))\dots\Theta(\bar{\sigma})$ in equations (12.4.131) and (12.4.133) with $\Theta(\pm T - R)$ in equations (12.4.132) and (12.4.134).

As already advertised earlier, because the Green's function of the scalar wave operator in Minkowski is the field generated by a unit strength point source in spacetime – the field ψ generated by an arbitrary source $J(t,\vec{x})$ obeys causality. By choosing the *retarded* Green's function, the field generated by the source propagates on and possibly within the forward light cone of J. Specifically, ψ travels strictly on the light cone for even dimensions greater or equal to 4, because $G_{D+1=2n}$ involves only $\delta(\bar{\sigma})$ and its derivatives. In 2 dimensions, the Green's function is pure tail, and is in fact a constant $1/2$ inside the light cone. In 3 dimensions, the Green's function is also pure tail, going as $\bar{\sigma}^{-1/2}$ inside the light cone. For odd dimensions greater than 3, the Green's function has non-zero contributions from both on and inside the light cone. However, the $\partial_{\bar{\sigma}}$s occurring within eq. (12.4.133) can be converted into $\partial_{t'}$s and – at least for material/timelike J – integrated-by-parts within the integral in eq. (12.4.62) to act on the J. The result is that, in all odd dimensional Minkowski spacetimes ($d \geq 3$), physical signals propagate strictly inside the null cone, despite the massless nature of the associated particles.[151]

Problem 12.30. Massive Scalar Green's Functions Employ the ansatz

$$G_2(x-x';m) = G_2(x-x';m=0)F_2\left(m\sqrt{(x-x')^2}\right), \tag{12.4.136}$$

[151] Explicit formulas for the electromagnetic and linear gravitational case can be found in appendices A and B of arXiv: 1611.00018 [43].

$$G_3(x-x';m) = G_3(x-x';m=0)F_3\left(m\sqrt{(x-x')^2}\right); \tag{12.4.137}$$

and show that the retarded solutions of the massive scalar Green's functions equations

$$(\partial^2 + m^2)G = \delta^{(d)}[x-x']. \tag{12.4.138}$$

are, for $\bar{\sigma} \equiv (1/2)(x-x')^2$ and in $d = D+1$ spacetime dimensions,

$$G^+_{d=2+2n}(x-x';m) = \frac{\Theta(t-t')}{2(2\pi)^n}\frac{\partial^n}{\partial\bar{\sigma}^n}\left\{\Theta(\bar{\sigma})J_0\left(m\sqrt{2\bar{\sigma}}\right)\right\}, \tag{12.4.139}$$

$$G^+_{d=3+2n}(x-x';m) = \frac{\Theta(t-t')}{(2\pi)^n}\frac{\partial^n}{\partial\bar{\sigma}^n}\left\{\frac{\Theta(\bar{\sigma})}{\sqrt{2\bar{\sigma}}}\cos\left(m\sqrt{2\bar{\sigma}}\right)\right\}. \tag{12.4.140}$$

Hints: You should find that, with $\xi \equiv m\sqrt{2\bar{\sigma}}$, and if G_2 and G_3 are the symmetric (retarded plus advanced) Green's functions in 2D and 3D respectively,

$$\Box G_2 = 2\delta^{(2)}[x-x']\cdot F_2(\xi) + \frac{m^2}{2}\Theta(\bar{\sigma})\left(F_2''(\xi) + \frac{F_2'(\xi)}{\xi} + F_2(\xi)\right), \tag{12.4.141}$$

$$\Box G_3 = 2\delta^{(3)}[x-x']\cdot F_3(\xi) + \frac{m}{\pi}F_3'(\xi)\cdot\delta\left(\bar{\sigma}\right) + \frac{m^2\cdot\Theta(\bar{\sigma})}{2\pi\sqrt{2\bar{\sigma}}}\left(F_3''(\xi) + F_3(\xi)\right). \tag{12.4.142}$$

These results should provide ordinary differential equations and boundary conditions for F_2 and F_3. Finally, recall eq. (12.4.110). $\Box$

Comparison to Laplace Equation The sign difference between the 'time component' versus the 'space components' of the flat spacetime metric is responsible for the sign difference between the time derivatives and spatial derivatives in the wave operator: $\partial^2\psi = (\partial_t^2 - \vec{\nabla}_{\vec{x}}^2)\psi = 0$. This can be contrasted against Laplace's equation $\vec{\nabla}^2\psi = \partial_i\partial_i\psi = 0$, where there are no sign differences because the Euclidean metric is diag$(+1,\cdots+1)$. In turn, let us recognize, this is why non-trivial smooth solutions exist in vacuum for the former and not for the latter, at least in infinite space(time). Physically, we may interpret this as telling us that the wave equation allows for radiation – i.e., waves that propagate through spacetime, capable of carrying energy-momentum to infinity – while the Laplace equation does not. To this end, let us go to Fourier space(time).

$$\partial^2\left(\widetilde{\psi}(k)e^{-ik_\mu x^\mu}\right) = 0 = -\left(k_0^2 - \vec{k}\cdot\vec{k}\right)\widetilde{\psi}e^{-ik_\mu x^\mu}, \qquad \text{(Wave Equation)} \tag{12.4.143}$$

$$\vec{\nabla}^2\left(\widetilde{\psi}(\vec{k})e^{i\vec{k}\cdot\vec{x}}\right) = 0 = -\left(\vec{k}\cdot\vec{k}\right)\widetilde{\psi}e^{i\vec{k}\cdot\vec{x}} \qquad \text{(Laplace Equation)} \tag{12.4.144}$$

We see that, for the wave equation, either $\widetilde{\psi} = 0$ or $k^2 = k_0^2 - \vec{k}^2 = 0$. But $\widetilde{\psi} = 0$ would render the whole solution trivial. Hence, for non-singular $\widetilde{\psi}$ this means $k_0 = \pm|\vec{k}|$ and we have

$$\psi = \widetilde{\psi}(k)\exp\left(i|\vec{k}|(\widehat{k}\cdot\vec{x}\mp t)\right), \qquad \widehat{k} \equiv \vec{k}/|\vec{k}|. \tag{12.4.145}$$

(We have already encountered this result in eq. (12.4.44).) Whereas, for the Laplace equation either $\widetilde{\psi} = 0$ or $\vec{k}^2 = 0$. Again, the former would render the whole solution trivial, which tells us we must have $\vec{k}^2 = 0$. However, since $\vec{k}^2 \geq 0$ – this positive definite nature of $\vec{k}^2$ is a consequence of the analogous one of the Euclidean metric – we conclude there are simply no non-trivial regular solutions in Fourier space.[152] For the wave equation, the non-trivial solutions $k_0 = \pm|\vec{k}|$ are a direct consequence of the Lorentzian nature of Minkowski spacetime.

Comparison to Heat Equation The causal structure of the solutions to the wave equation here can be contrasted against those of the infinite flat space heat equation. Referring to the heat kernel in eq. (12.3.34), we witness how at initial time t', the field K is infinitely sharply localized at $\vec{x} = \vec{x}'$. However, immediately afterwards, it becomes spread out over all space, with a Gaussian profile peaked at $\vec{x} = \vec{x}'$ – thereby violating causality. In other words, the "waves" in the heat/diffusion equation of eq. (12.3.1) propagates with infinite speed. Physically speaking, we may attribute this property to the fact that time and space are treated asymmetrically both in the heat/diffusion eq. (12.3.1) itself – one time derivative versus two derivatives per spatial coordinate – as well as in the heat kernel solution of eq. (12.3.34). On the other hand, the symmetric portion of the spacetime Green's functions in equations (12.4.131) and (12.4.133) depend on spacetime solely through $2\bar{\sigma} \equiv (t-t')^2 - (\vec{x}-\vec{x}')^2$, which is invariant under global Poincaré transformations (cf. eq. (12.4.14)).

Heat Kernel, Feynman vs. Symmetric Green's functions Despite the apparently non-relativistic character of the heat equation, let us now turn to study the relationship between the heat kernel, Feynman Green's function; as well as the latter's connection to the symmetric Green's function, defined as the sum of retarded and advanced Green's function:

$$G^S \equiv G^+ + G^-. \tag{12.4.146}$$

This in turn means, we may extract the retarded $(+)$ or advanced $(-)$ Green's functions via

$$G^\pm = \Theta(\pm(t-t'))G^S. \tag{12.4.147}$$

Notice the Feynman prescription on the complex $\omega-$plane for the Green's function can be summarized by the algebraic equation

$$\omega^2 - \vec{k}^2 + i0^+ = 0 \tag{12.4.148}$$

because the roots are $\omega_\pm = \pm|\vec{k}|(1 - i0^+)$; which is equivalent to skirting the $-|\vec{k}|$ pole from below because it now lies slightly above the Re axis; as well as skirting the $+|\vec{k}|$ pole from above, since it now lies slightly below the Re axis. We may now write

$$-\frac{1}{\omega^2 - \vec{k}^2 + i0^+} = i\int_0^\infty \mathrm{d}s e^{is(\omega^2 - \vec{k}^2 + i0^+)}, \tag{12.4.149}$$

where the $i0^+$ factor arising from the Feynman contour prescription is crucial for the upper limit of the integral to converge. This brings us to

$$G^{\mathrm{F}}(T, \vec{x}-\vec{x}') = i\int_0^\infty \mathrm{d}s \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \int_{\mathbb{R}^D} \frac{\mathrm{d}^D\vec{k}}{(2\pi)^D} e^{-i\omega T} e^{is(\omega^2 - \vec{k}^2 + i0^+)} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{12.4.150}$$

[152] One could allow for singular solutions proportional to the $\vec{k}-$space $\delta^{(d-1)}$-function and its derivatives, such as $\widetilde{\psi}_0 = \delta^{(3)}(\vec{k})\exp(i\vec{k}\cdot\vec{x})$ and $\widetilde{\psi}_1 = \partial_{k^i}\delta^{(3)}(\vec{k})\exp(i\vec{k}\cdot\vec{x})$ (for fixed i), so that $\vec{\nabla}^2\psi = 0$ because $\vec{k}^2\delta^{(3)}(\vec{k}) = 0 = \vec{k}^2\partial_{k^i}\delta^{(3)}(\vec{k})$. However, the ψ_0 in position space is simply a spatial constant; while the ψ_1 is proportional to x^i, which blows up as $x^i \to \pm\infty$. In fact, there are an infinite number of linearly independent homogeneous solutions to the Laplace equation, namely $\{r^\ell Y_\ell^m(\theta,\phi) | \ell = 0,1,2,3,\dots; m = -\ell, -\ell+1, \dots, +\ell-1, +\ell\}$, but for $\ell > 0$ they all blow up at spatial infinity.

Defining $\omega' \equiv \omega\sqrt{s} - T/(2\sqrt{s})$, and upon 'completing the square' in the exponent,

$$G^{\mathrm{F}}(T, \vec{x} - \vec{x}') = i \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-i\frac{T^2}{4s}} \left(\int_{\mathbb{R}} \frac{\mathrm{d}\omega'}{2\pi} e^{i\omega'^2} \right) \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} e^{-is(\vec{k}^2 - i0^+)} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}. \tag{12.4.151}$$

The ω' integral may be evaluated by a counter-clockwise rotation of its contour by $\pi/4$; i.e., $\omega' \to e^{i\pi/4}\omega'$, so that exponential itself becomes a real Gaussian. Since the $\omega-$integrals have been evaluated, we shall assume the remaining integrals are well-defined and proceed to drop the 0^+ dependence. We then arrive at the following relation between the Feynman Green's function and 'Wick rotated' heat kernel

$$G^{\mathrm{F}}_{D+1}(T, \vec{x} - \vec{x}') = \frac{1}{2\sqrt{\pi}} \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-i\left(\frac{T^2}{4s} - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; is\right), \tag{12.4.152}$$

$$K_D(\vec{x}, \vec{x}'; s) \equiv \int_{\mathbb{R}^D} \frac{\mathrm{d}^D \vec{k}}{(2\pi)^D} e^{-s\cdot\vec{k}^2} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}; \tag{12.4.153}$$

where K itself can be found in eq. (12.3.31). Employing the result in eq. (12.3.34) and recalling eq. (12.4.85), we may further infer

$$2 \text{ Re } G^{\mathrm{F}}_{D+1}(T, \vec{x} - \vec{x}') = G^{S}_{D+1}(T, \vec{x} - \vec{x}') \tag{12.4.154}$$

$$= \frac{1}{\sqrt{4\pi}} \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} \left(e^{-i\left(\frac{T^2}{4s} - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; is\right) + e^{i\left(\frac{T^2}{4s} - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; -is\right) \right) \tag{12.4.155}$$

$$= \frac{1}{\sqrt{4\pi}} \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} \left(\frac{e^{\frac{\bar{\sigma}}{2is} + i\frac{3\pi}{4}}}{(4\pi is)^{D/2}} + \frac{e^{-\frac{\bar{\sigma}}{2is} - i\frac{3\pi}{4}}}{(-4\pi is)^{D/2}} \right), \tag{12.4.156}$$

where $\bar{\sigma} \equiv (1/2)(T^2 - (\vec{x} - \vec{x}')^2)$ is Synge's world function in flat spacetime. We see that this method of computing Green's functions leads directly to the symmetric Green's function, whose integral representation is manifestly Poincaré invariant – because it depends on spacetime solely through $\bar{\sigma}$ – unlike its retarded/advanced counterparts.

Problem 12.31. Recursion Relation Use eq. (12.4.156) to show that

$$G^{S}_{D+1+2n}(\bar{\sigma}) = \left(\frac{1}{2\pi} \frac{\partial}{\partial \bar{\sigma}} \right)^n G^{S}_{D+1}(\bar{\sigma}), \tag{12.4.157}$$

for $n = 0, 1, 2, 3, \dots$. If we multiply this result by $\Theta(\pm(t - t'))$, we will arrive at equations (12.4.131) and (12.4.133). □

Let us compute the $D = 2$ case of eq. (12.4.156). To begin, put $s' \equiv 1/s$.

$$G^{S}_{3}[x, x'] = \frac{1}{(4\pi)^{3/2}} \int_0^\infty \frac{\mathrm{d}s'}{\sqrt{s'}} \left(e^{-i\frac{s'}{2}\cdot\bar{\sigma}} e^{i\frac{\pi}{4}} + e^{i\frac{s'}{2}\cdot\bar{\sigma}} e^{-i\frac{\pi}{4}} \right) \tag{12.4.158}$$

Consider, for real α,

$$\int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-is\alpha} = \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-is(\mathrm{sgn}[\alpha])|\alpha|}. \tag{12.4.159}$$

Let us rotate the contour $s \to (-i\mathrm{sgn}[\alpha])s$, where $s \in [0, \infty)$.

$$\int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-is\alpha} = (-i\mathrm{sgn}[\alpha])^{1/2} \int_0^\infty \frac{\mathrm{d}s}{\sqrt{s}} e^{-s|\alpha|} \tag{12.4.160}$$

$$= \frac{e^{-i\frac{\pi}{4}\mathrm{sgn}[\alpha]}}{\sqrt{|\alpha|}} \int_0^\infty \frac{\mathrm{d}(s|\alpha|)}{\sqrt{s|\alpha|}} e^{-(s|\alpha|)} = \frac{e^{-i\frac{\pi}{4}\mathrm{sgn}[\alpha]}}{\sqrt{|\alpha|}} \Gamma[1/2] \tag{12.4.161}$$

$$= \sqrt{\frac{\pi}{|\alpha|}} e^{-i\frac{\pi}{4}\mathrm{sgn}[\alpha]} \tag{12.4.162}$$

Applying this result to the symmetric Green's function,

$$G_3^S[x, x'] = \frac{1}{(4\pi)^{3/2}} \sqrt{\frac{2\pi}{|\bar{\sigma}|}} \left(e^{-i\frac{\pi}{4}\mathrm{sgn}[\bar{\sigma}]} e^{i\frac{\pi}{4}} + e^{i\frac{\pi}{4}\mathrm{sgn}[\bar{\sigma}]} e^{-i\frac{\pi}{4}} \right) \tag{12.4.163}$$

When $\bar{\sigma} > 0$, the exponentials become unity. When $\bar{\sigma} < 0$, we have

$$e^{-i\frac{\pi}{4}\mathrm{sgn}[\bar{\sigma}]} e^{i\frac{\pi}{4}} + e^{i\frac{\pi}{4}\mathrm{sgn}[\bar{\sigma}]} e^{-i\frac{\pi}{4}} = e^{i\frac{\pi}{2}} + e^{-i\frac{\pi}{2}} = 0. \tag{12.4.164}$$

Hence, the symmetric Green's function in 3D Minkowski is

$$G_3^S[x, x'] = \frac{\Theta[\bar{\sigma}]}{2\pi\sqrt{2\bar{\sigma}}}. \tag{12.4.165}$$

Next, let us turn to evaluating the $D = 3$ case of eq. (12.4.156). Recognizing $i^{3/2} = e^{i3\pi/4}$ and switching to $s' \equiv 1/s$, we arrive at

$$G_4^S(T, \vec{x} - \vec{x}') = \frac{1}{(4\pi)^2} \int_0^\infty \mathrm{d}s' \left(e^{-is'\frac{\bar{\sigma}}{2}} + e^{is'\frac{\bar{\sigma}}{2}} \right) \tag{12.4.166}$$

$$= \frac{1}{(4\pi)^2} \int_{\mathbb{R}} \mathrm{d}s' e^{is'\frac{\bar{\sigma}}{2}} = \frac{\delta(\bar{\sigma})}{4\pi}. \tag{12.4.167}$$

Now that we have the 3D and 4D symmetric Green's functions, we may employ eq. (12.4.157) to derive the G_{D+1}^S for all $D \geq 2$.

Problem 12.32. (1+1)D Work out the $D = 1$ case of eq. (12.4.156). You should find that it *does not* recover the symmetric massless scalar Green's function in $1 + 1$ dimensional flat spacetime – can you explain why? □

Problem 12.33. Hermitian Spatial Laplacians in Curved Spacetime Let us consider spacetimes of the form

$$\mathrm{d}s^2 = \mathrm{d}t^2 - H_{ij}(\vec{x})\mathrm{d}x^i \mathrm{d}x^j, \tag{12.4.168}$$

where H_{ij} is a time-independent spatial metric. Let D_i denote the covariant derivative compatible with H_{ij}. Verify that the spacetime wave operator is

$$\Box\psi = \left(\partial_t^2 - \vec{\nabla}^2 \right) \psi, \tag{12.4.169}$$

$$\vec{\nabla}^2 \equiv D_i D^i = H^{ij} D_i D_j, \qquad H^{il} H_{lj} \equiv \delta^i_j. \tag{12.4.170}$$

Eigenfunction expansion Let us impose Dirichlet boundary conditions in the space described by H_{ij}, so that the Laplacian $\vec{\nabla}^2$ becomes a Hermitian operator. If the space itself has no boundaries, then $\vec{\nabla}^2$ is automatically Hermitian. Denote the orthonormal eigenfunctions of $D_i D^i$ by $\{\Psi_L\}$, namely

$$-\vec{\nabla}^2 \Psi_L = \lambda_L \Psi_L; \tag{12.4.171}$$

where L is a schematic label, which can be either discrete or continuous. For example, for infinite flat space $H_{ij} = \delta_{ij}$ the eigenfunctions are $\Psi_L \leftrightarrow e^{i\vec{k}\cdot\vec{x}}$, $L \leftrightarrow \vec{k}$, and $\lambda_L \leftrightarrow \vec{k}^2$; whereas for the 2−sphere $H_{ij} = \text{diag}[1, \sin^2\theta]$, $\Psi_L \leftrightarrow Y_\ell^m(\theta,\phi)$, $L \leftrightarrow (\ell, m)$, and $\lambda_L \leftrightarrow \ell(\ell+1)$. In any case, that means we may decompose the wave ψ and its source J into modes:

$$\psi(t,\vec{x}) = \sum_L \widetilde{\psi}_L(t)\Psi_L(\vec{x}), \tag{12.4.172}$$

$$J(t,\vec{x}) = \sum_L \widetilde{J}_L(t)\Psi_L(\vec{x}). \tag{12.4.173}$$

If the wave equation

$$\Box\psi = J \tag{12.4.174}$$

holds, show that the individual modes are driven simple harmonic oscillators (SHOs); namely,

$$\partial_t^2 \widetilde{\psi}_L + \lambda_L \widetilde{\psi}_L = \widetilde{J}_L. \tag{12.4.175}$$

If we recall the Greem's functions of the SHO, explain why the retarded (+) and advanced (-) Green's functions of our wave operator must therefore be

$$G^{\pm}_{D+1}(T \equiv t - t', \vec{x}, \vec{x}') = \pm\Theta(\pm T) \sum_L \frac{\sin\left(\sqrt{\lambda_L} T\right)}{\sqrt{\lambda_L}} \Psi_L(\vec{x}) \overline{\Psi_L(\vec{x}')}. \tag{12.4.176}$$

Heat Kernel Next, argue that the associated heat kernel satisfying the Schrödinger equation – without $\hbar$s –

$$i\partial_s K_D(\vec{x},\vec{x}'; is) = -\vec{\nabla}^2 K_D(\vec{x},\vec{x}'; is), \tag{12.4.177}$$

$$K_D(\vec{x},\vec{x}'; is = 0) = \frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}} \tag{12.4.178}$$

is

$$K_D(\vec{x},\vec{x}'; s \geq 0) = \sum_L e^{-s\cdot\lambda_L} \Psi_L(\vec{x}) \overline{\Psi_L(\vec{x}')}. \tag{12.4.179}$$

Explain why the symmetric Green's function admits the following relation to the heat kernel –

$$G^S_{D+1}(T = t - t', \vec{x}, \vec{x}') = \int_0^\infty \frac{\mathrm{d}s}{\sqrt{4\pi s}} \Big(e^{-i\left(\frac{T^2}{4s} - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; is\right)$$

$$+ e^{i\left(\frac{T^2}{4s}-\frac{3\pi}{4}\right)} K_D\left(\vec{x},\vec{x}';-is\right)\Big) \tag{12.4.180}$$

– i.e., it takes the same form as eq. (12.4.155).

Frequency Decomposition Alternatively, we may exploit the time translation invariance of the geometry to decompose the Green's function in frequency space. Argue that

$$G^{\pm}_{D+1}(T \equiv t-t', \vec{x}, \vec{x}') = \pm\Theta(\pm T)\int_C \frac{\mathrm{d}\omega}{2\pi}\frac{\sin(\omega T)}{\omega}\widetilde{G}(\omega;\vec{x},\vec{x}'); \tag{12.4.181}$$

where $\widetilde{G}$ satisfies the homogeneous Helmholtz equation

$$\left(-\omega^2 - \vec{\nabla}^2_{\vec{x}}\right)\widetilde{G} = \left(-\omega^2 - \vec{\nabla}^2_{\vec{x}'}\right)\widetilde{G} = 0 \tag{12.4.182}$$

and the 'completeness relation'

$$\int_C \frac{\mathrm{d}\omega}{2\pi}\widetilde{G}(\omega;\vec{x},\vec{x}') = \frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{H(\vec{x})H(\vec{x}')}} \tag{12.4.183}$$

for some appropriate contour C on the complex $\omega-$plane. $\square$

Problem 12.34. Green's Functions on 2-Sphere Let D_i be the covariant derivative with respect to the 2-sphere geometry

$$\mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2; \tag{12.4.184}$$

and let $\widetilde{\epsilon}_{ab}$ be its associated Levi-Civita tensor. Define the retarded (+) and advanced (−) scalar and vector Green's function on the 2-sphere to be

$$G^{\pm}(T,\Omega,\Omega') = \pm\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}\Theta(\pm T)\frac{\sin(\sqrt{\ell(\ell+1)}T)}{\sqrt{\ell(\ell+1)}}Y_\ell^m(\Omega)\overline{Y_\ell^m}(\Omega'), \tag{12.4.185}$$

$$\begin{aligned} G^{\pm}{}_i{}^{j'}(T,\Omega,\Omega') = \pm\sum_{\ell\geq 1}\sum_{m=-\ell}^{\ell}\Theta(\pm T)\frac{\sin(\sqrt{\ell(\ell+1)}T)}{\sqrt{\ell(\ell+1)}} \\ \times\frac{D_i Y_\ell^m(\Omega)\overline{D^{j'}Y_\ell^m}(\Omega') + \widetilde{\epsilon}_{ia}D^a Y_\ell^m(\Omega)\overline{\widetilde{\epsilon}^{j'b'}D_{b'}Y_\ell^m}(\Omega')}{\ell(\ell+1)}; \end{aligned} \tag{12.4.186}$$

where $T \equiv t-t'$, $\{Y_\ell^m\}$ are the usual spherical harmonics and Ω and Ω' denote the angular coordinates of the observer and source points. If the spacetime is

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}\Omega^2, \tag{12.4.187}$$

show that these Green's functions obey

$$\Box G^{\pm}(t-t',\vec{x},\vec{x}') = \frac{\delta^{(2)}(\Omega-\Omega')}{\sqrt[4]{g(\Omega)g(\Omega')}}, \tag{12.4.188}$$

$$(\Box+1)G^{\pm}{}_i{}^{j'}(t-t',\vec{x},\vec{x}') = \delta_i^{\ j}\frac{\delta^{(2)}(\Omega-\Omega')}{\sqrt[4]{g(\Omega)g(\Omega')}}. \tag{12.4.189}$$

Hint: Problem (12.33), as well as the completeness and eigen-equations of both the scalar and vector spherical harmonics are important. For the vector harmonics case, see Problems (9.97) and (9.98). $\square$

Problem 12.35. Massive Green's function in Minkowski In this problem, we shall extend the analysis in Problem (12.33) to solve for the Green's function associated with $\Box + m^2$, where m is some positive mass. Namely, we seek the symmetric solution to

$$\left(\Box + m^2\right) G^S_{D+1}(x, x') = \frac{\delta^{(D+1)}[x - x']}{\sqrt[4]{g[x]g[x']}}, \tag{12.4.190}$$

for the metric in eq. (12.4.168). Argue that eq. (12.4.180) becomes

$$\begin{aligned} G^S_{D+1}(T, \vec{x}, \vec{x}') = \int_0^\infty \frac{\mathrm{d}s}{\sqrt{4\pi s}} \Big(& e^{-i\left(\frac{T^2}{4s} + s \cdot m^2 - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; is\right) \\ & + e^{i\left(\frac{T^2}{4s} + s \cdot m^2 - \frac{3\pi}{4}\right)} K_D\left(\vec{x}, \vec{x}'; -is\right) \Big). \end{aligned} \tag{12.4.191}$$

Use this formula to recover equations (12.4.139) and (12.4.140).

Hints: For the 2D Minkowski case, consider re-scaling $s \to s/m^2$ followed by the change-of-variables $s \to (\zeta/2)e^s$ where $\zeta \equiv m\sqrt{2|\bar{\sigma}|}$. For 3D Minkowski, after re-scaling $s \to s/m^2$, consider instead the change-of-variables $s \to (\zeta^2/2)s^{-2}$. You may find it helpful to consider the $\bar{\sigma} < 0$ and $\bar{\sigma} > 0$ cases separately. □

12.4.3 Massless Fields in (3+1)D: Position Spacetime

4 Dimensions: Massless Scalar Field We highlight the 4 dimensional retarded case, because it is most relevant to the real world. Using equations (12.4.131) and (12.4.132), the retarded massless scalar Green's function reads

$$G^+_4(x - x') = \frac{\Theta(t - t')\delta(\bar{\sigma})}{4\pi} = \frac{\delta\left(t - t' - |\vec{x} - \vec{x}'|\right)}{4\pi|\vec{x} - \vec{x}'|}. \tag{12.4.192}$$

The G_4 says the point source at $(t', \vec{x}')$ produces a spherical wave that propagates strictly on the light cone $t - t' = |\vec{x} - \vec{x}'|$, with amplitude that falls off as 1/(observer-source spatial distance) $= 1/|\vec{x} - \vec{x}'|$.[153]

Problem 12.36. 3D Green's function from 4D Can you use eq. (12.4.128) to compute the (2+1)D massless scalar retarded Green's function in eq. (12.4.99) from its (3+1)D counterpart in eq. (12.4.192)? □

The solution to ψ from eq. (12.4.62) is now

$$\begin{aligned} \psi(t, \vec{x}) &= \int_{-\infty}^{+\infty} \mathrm{d}t' \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' G^+_4(t - t', \vec{x} - \vec{x}')J(t', \vec{x}') \\ &= \int_{-\infty}^{+\infty} \mathrm{d}t' \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' \frac{\delta\left(t - t' - |\vec{x} - \vec{x}'|\right) J(t', \vec{x}')}{4\pi|\vec{x} - \vec{x}'|} && (12.4.193) \\ &= \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' \frac{J(t_r, \vec{x}')}{4\pi|\vec{x} - \vec{x}'|}, \qquad t_r \equiv t - |\vec{x} - \vec{x}'|. && (12.4.194) \end{aligned}$$

[153] In the first equality of eq. (12.4.192), one may verify: $\delta(\bar{\sigma})/(4\pi) = (4\pi|\vec{x} - \vec{x}'|)^{-1}(\delta(t - t' - |\vec{x} - \vec{x}'|) + \delta(t - t' + |\vec{x} - \vec{x}'|))$ The $\delta(t - t' + |\vec{x} - \vec{x}'|)/(4\pi|\vec{x} - \vec{x}'|)$ would be the advanced Green's function, where the elapsed time $t - t' = -|\vec{x} - \vec{x}'| < 0$, and is eliminated by the retarded condition encoded within $\Theta(t - t')$.

The t_r is called retarded time. With $c = 1$, the time it takes for a signal traveling at unit speed to travel from $\vec{x}'$ to $\vec{x}$ is $|\vec{x} - \vec{x}'|$, and so at time t, what the observer detects at $(t, \vec{x})$ is what the source produced at time $t - |\vec{x} - \vec{x}'|$. Drawing a spacetime diagram here would be useful.

4D Far Zone Let us center the coordinate system so that $\vec{x} = \vec{x}' = \vec{0}$ lies within the body of the source J itself. When the observer is located at very large distances from the source compared to the latter's characteristic size, we may approximate

$$\begin{aligned} |\vec{x} - \vec{x}'| &= \exp\left(-x'^j \partial_j\right) r, \qquad\qquad r \equiv |\vec{x}| \\ &= r - \vec{x}' \cdot \widehat{r} + r\mathcal{O}\left(\left(\frac{r'}{r}\right)^2\right), \qquad \widehat{r} \equiv \frac{x^i}{r}, \qquad r' \equiv |\vec{x}'| \qquad (12.4.195) \\ &= r - \vec{x}' \cdot \widehat{r} + r'\mathcal{O}\left(\frac{r'}{r}\right) = r\left(1 - \frac{\vec{x}'}{r} \cdot \widehat{r} + \mathcal{O}\left((r'/r)^2\right)\right). \qquad (12.4.196) \end{aligned}$$

(By dimensional analysis, you should be able to deduce this is, schematically, a power series in r'/r.) This leads us from eq. (12.4.194) to the following far zone scalar solution

$$\begin{aligned} \psi(t, \vec{x}) = \frac{1}{4\pi r} \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' \left\{1 + \frac{\vec{x}'}{r} \cdot \widehat{x} + \mathcal{O}\left((r'/r)^2\right)\right\} \qquad (12.4.197) \\ \times J\left(t - r + \vec{x}' \cdot \widehat{r} + r'\mathcal{O}(r'/r), \vec{x}'\right). \end{aligned}$$

The term in curly brackets arises from the $1/|\vec{x} - \vec{x}'|$ portion of the 4D Green's function in eq. (12.4.192). In turn, this far zone leading order $1/r$ behavior teaches us, the field due to some field always fall off as 1/(observer-source distance).[154] On the other hand, by recognizing $\vec{x}' \cdot \widehat{r} + r'\mathcal{O}(r'/r) = \widehat{x}' \cdot \widehat{r}(1 + \mathcal{O}(r'/r))$, followed by Taylor expanding the time argument of the source,

$$J\left(t - |\vec{x} - \vec{x}'|, \vec{x}'\right) = J(t - r, \vec{x}') + \sum_{\ell=1}^{+\infty} \frac{\left(\vec{x}' \cdot \widehat{r}\right)^\ell}{\ell!} \left(1 + \mathcal{O}(r'/r)\right) \partial_t^\ell J(t - r, \vec{x}'). \qquad (12.4.198)$$

If we associate each time derivative acting on J to scale as

$$\partial_t^\ell J \sim \frac{J}{(\text{characteristic timescale of source})^\ell}, \qquad (12.4.199)$$

then the Taylor expansion in eq. (12.4.198) becomes one in powers of the ratio $r'/\tau_J \equiv$ (characteristic size of the source)/(characteristic timescale of source). (Recall from eq. (12.4.193) the $\vec{x}'$ always lies within the source.) In the $c = 1$ units we are employing here, this corresponds to a non-relativistic expansion, since the characteristic size of the source is the time it takes for light to traverse it. Furthermore, at each order in this non-relativistic expansion, there is a further 'finite size' correction that begins at order $r'/r \sim$ (characteristic size of source)/(observer-source distance).

Relativistic Far Zone To sum, if we take the far zone limit – i.e., neglect all (characteristic size of source)/(observer-source distance) $\ll 1$ corrections – but allow for a fully relativistic source, eq. (12.4.197) now reads

$$\psi(t, \vec{x}) \approx \frac{1}{4\pi r} \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' J\left(t - r + \vec{x}' \cdot \widehat{r}, \vec{x}'\right), \qquad r \equiv |\vec{x}|. \qquad (12.4.200)$$

[154] In Quantum Field Theory, this $1/r$ is attributed to the massless-ness of the $\psi-$particles.

Non-relativistic Far Zone If we further assume the source is non-relativistic, namely (characteristic size of source)/(timescale of source) $\ll 1$,

$$\psi(t,\vec{x}) \approx \frac{\mathcal{A}(t-r)}{4\pi r}, \tag{12.4.201}$$

$$\mathcal{A}(t-r) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' J(t-r,\vec{x}'). \tag{12.4.202}$$

In the far zone the amplitude of the wave falls off with increasing distance as 1/(observer-source spatial distance); and the time-dependent portion of the wave $\mathcal{A}(t-r)$ is consistent with that of an outgoing wave, one emanating from the source J.[155]

Problem 12.37. Spherical $s-$Waves The $\mathcal{A}(t-r)/r$ in eq. (12.4.201) turns out to be an *exact* solution, despite our arrival at it via a non-relativistic and far zone approximation. Referring to eq. (12.4.193), identify the form of $J(t',\vec{x}')$ that would yield the following exact solution to $\partial^2\psi = J$:

$$\psi(t,\vec{x}) = \frac{\mathcal{A}(t-r)}{4\pi r}, \qquad r \equiv |\vec{x}|. \tag{12.4.203}$$

Hint: J describes a point charge sitting at the spatial origin, but with a time dependent strength. $\square$

Problem 12.38. Spherical $s-$Waves vs Plane Waves In this problem, we will compare the homogeneous plane wave solutions in eq. (12.4.42) with the spherical wave in eq. (12.4.203). We will assume the amplitude $\mathcal{A}$ in eq. (12.4.203) admits a Fourier transform:

$$\mathcal{A}(\xi) = \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega\xi} \widetilde{A}(\omega). \tag{12.4.204}$$

Then each frequency mode must itself be a solution to the wave equation:

$$\psi = \mathrm{Re}\left(\widetilde{A}(\omega)\frac{e^{-i\omega(t-r)}}{4\pi r}\right). \tag{12.4.205}$$

Throughout this analysis, we shall assume the high frequency and far zone limit to hold:

$$\omega r \gg 1. \tag{12.4.206}$$

First show that the Minkowski metric in spherical coordinates is

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right). \tag{12.4.207}$$

Then verify that

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\left(\frac{e^{-i\omega(t-r)}}{4\pi r}\right) = 0, \tag{12.4.208}$$

[155] Even for the relativistic case in eq. (12.4.200), we see from eq. (12.4.198) that it consists of an infinite series of various rank amplitudes that are functions of retarded time $t-r$: $\psi(t,\vec{x}) = (4\pi r)^{-1}\sum_{\ell=0}^{+\infty} \widehat{r}^{i_1}\dots\widehat{r}^{i_\ell}\mathcal{A}^{i_1\dots i_\ell}(t-r)$.

as long as $r \neq 0$; as well as the null character of the constant-phase surfaces, in that

$$g^{\mu\nu}\nabla_\mu \left(\omega(t-r)\right) \nabla_\nu \left(\omega(t-r)\right) = 0. \tag{12.4.209}$$

This latter condition is consistent with the property that the spherical wave is traveling radially outwards from the source at the speed of light. Now, since $\exp(-i\omega(t-r))$ is the 'fast' part of the spherical wave (at least for $\omega r \gg 1$) whereas $1/r \ll \omega$ is the 'slow' part, we see that $\exp(-i\omega(t-r))$ in eq. (12.4.205) may be identified with $\exp(-ik(t-\widehat{k}\cdot\vec{x}))$ in eq. (12.4.42) if we identify the propagation direction $\widehat{r}$ in the former with the the propagation direction $\widehat{k}$ in the latter:

$$\widehat{r} \leftrightarrow \widehat{k} \qquad \text{and} \qquad \omega \leftrightarrow k. \tag{12.4.210}$$

Afterall, as the radius of curvature grows ($r \to \infty$), we expect the constant phase surfaces of the spherical wave to appear locally flatter – and hence, to a good approximation, behaving more like plane waves, at least within a region whose extent is much smaller than r itself.

To further support this identification, we recognize that, each t or r derivative on $e^{-i\omega(t-r)}$ yields a factor of $\omega \sim 1/(\text{period of wave})$. So one might have expected that $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$ applied to the same should scale as $\Box \sim \omega^2$. However, show that – due to the null condition in eq. (12.4.209),

$$\Box e^{-i\omega(t-r)} = -2i\frac{\omega}{r}e^{-i\omega(t-r)} \tag{12.4.211}$$

instead. Thus, relative to the expectation that $\Box \sim \omega^2$, the actual result scales as $1/(\omega r)$ relative to it.

$$\frac{\Box[\text{Actual}]}{\Box[\text{Expectation based on first derivatives}]} \sim \frac{1}{\omega r} \ll 1. \tag{12.4.212}$$

To sum:

> In the high-frequency and far zone limit, namely $\omega r \gg 1$, a single frequency mode of the spherical wave approximates that of a plane wave, as $r \to \infty$, in a given region whose size is much smaller than r itself. The slowly varying amplitude of the spherical wave scales as $1/r$.

We will see below, the spherical wave $\exp(-i\omega(t-r))/r$ can also be viewed as a special case of the JWKB solution of wave equations. $\Box$

4D photons In 4 dimensional flat spacetime, the vector potential of electromagnetism, in the Lorenz gauge

$$\partial_\mu A^\mu = 0 \qquad \text{(Cartesian coordinates)}, \tag{12.4.213}$$

obeys the wave equation

$$\partial^2 A^\mu = J^\mu. \tag{12.4.214}$$

Here, ∂^2 is the scalar wave operator, and J^μ is a conserved electromagnetic current describing the motion of some charge density

$$\partial_\mu J^\mu = \partial_t J^t + \partial_i J^i = 0. \tag{12.4.215}$$

The electromagnetic fields are the "curl" of the vector potential

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{12.4.216}$$

In particular, for a given inertial frame, the electric E and magnetic B fields are, with $i, j, k \in \{1, 2, 3\}$,

$$E^i = F^{i0} = \partial^i A^0 - \partial^0 A^i = -\partial_i A_0 + \partial_0 A_i = -F_{i0}, \tag{12.4.217}$$

$$B^k = -\epsilon^{ijk}\partial_i A_j = -\frac{1}{2}\epsilon^{ijk} F_{ij}, \qquad \epsilon^{123} \equiv 1. \tag{12.4.218}$$

Problem 12.39. Lorenz Gauge, Relativity & Current Conservation Comparison of equations (12.4.200) and (12.4.214) indicates, in the far zone,

$$A^\mu(t, \vec{x}) \approx \frac{1}{4\pi r}\int \mathrm{d}^3\vec{x}' J^\mu\left(t - r + \vec{x}' \cdot \widehat{r}, \vec{x}'\right). \tag{12.4.219}$$

If one takes the non-relativistic limit too (cf. eq. (12.4.201)),

$$A^\mu(t, \vec{x}) \approx \frac{1}{4\pi r}\int \mathrm{d}^3\vec{x}' J^\mu\left(t - r, \vec{x}'\right). \tag{12.4.220}$$

Compute $\partial_\mu A^\mu$ using equations (12.4.219) and (12.4.220) to leading order in $1/r$. Hint: a key step is to recognize, for a conserved current obeying eq. (12.4.215),

$$\partial_t J^0(\tau, \vec{x}') = -\left(\partial_{i'} J^i(\tau, \vec{x}')\right)_t + \widehat{r}^i \partial_t J^i(\tau, \vec{x}'); \tag{12.4.221}$$

$$\tau \equiv t - r + \vec{x}' \cdot \widehat{r}, \qquad \partial_{i'} \equiv \frac{\partial}{\partial x'^i}; \tag{12.4.222}$$

where the subscript t on the first term on the right-hand-side of eq. (12.4.221) means the spatial derivatives are carried out with the observation time t held fixed – which is to be distinguished from doing so but with τ held fixed.

You should find that the Lorenz gauge in eq. (12.4.213) is respected only by the relativistic solution in eq. (12.4.219), and not by the non-relativistic one in eq. (12.4.220). This is an important point because, even though the Lorenz gauge in eq. (12.4.213) was a mathematical choice, once we have chosen it to solve Maxwell's equations, violating it may lead to a violation of current conservation: to see this, simply take the 4−divergence of eq. (12.4.214) to obtain $\partial^2(\partial_\mu A^\mu) = \partial_\mu J^\mu$. $\square$

Problem 12.40. Electromagnetic Radiation Refer to eq. (12.4.219), the solution of A^μ in terms of J^μ in the far zone. Like the scalar case, take the far zone limit. In this problem we wish to study some basic properties of A^μ in this limit. Throughout this analysis, assume that J^i is sufficiently localized that it vanishes at spatial infinity; and assume J^i is a non-relativistic source.

1. Using $\partial_t J^t = -\partial_i J^i$, the conservation of the current, show that A^0 is independent of time in the far zone and non-relativistic limit. Is there a difference between taking the time derivative of the non-relativistic limit of the far zone A_0 and taking the non-relativistic limit of its time derivative?

2. Now define the dipole moment as

$$I^i(s) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' x'^i J^0(s, \vec{x}'). \tag{12.4.223}$$

Can you show its first time derivative is

$$\dot{I}^i(s) \equiv \frac{\mathrm{d}I^i(t)}{\mathrm{d}s} = \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' J^i(s, \vec{x}')? \tag{12.4.224}$$

Hint: Use current conservation and integration-by-parts.

Compute the spatial derivative of the far zone A^0. Compare the result of acting ∂^i on the non-relativistic A^0 versus taking the non-relativistic limit of $\partial^i A^0$. You should find that the latter yields the correct answer: **YZ: Jia-En says – first line should not have a − sign.**

$$\begin{aligned}\partial^i A^0(t, \vec{x}) &\overset{\text{Far zone}}{\to} -\frac{\widehat{r}^i\widehat{r}^j}{4\pi r}\ddot{I}^j[t-r] \\ &= \widehat{r}^i\widehat{r}^j \left(\partial_t A^j(t, \vec{x})\right)_{\text{Non-relativistic, far zone}}. \end{aligned}\tag{12.4.225}$$

3. From the above results, we shall infer it is the 'transverse' part of the 'velocity' $\partial_t A^i$ that contains radiative effects. First show that in the far zone, i.e., to leading order in $1/r$, **YZ: Jia-En says – second line should have a − sign; i.e., $-P^{ij}$ ....**

$$\begin{aligned}E^i(t, \vec{x}) &\to -\frac{1}{4\pi r}\frac{\mathrm{d}^2 I^i_{(t)}(t-r, \widehat{r})}{\mathrm{d}t^2} \\ &\equiv P^{ij}\left(\partial_t A^j(t, \vec{x})\right)_{\text{Non-relativistic, far zone}}\end{aligned}\tag{12.4.226}$$

$$B^i(t, \vec{x}) \to -\frac{1}{4\pi r}\epsilon^{ijk}\widehat{r}^j\frac{\mathrm{d}^2 I^k_{(t)}(t-r, \widehat{r})}{\mathrm{d}t^2} = \epsilon^{ijk}\widehat{r}^j E^k, \tag{12.4.227}$$

$$I^i_{(t)}(s, \widehat{r}) \equiv P^{ij}(\widehat{r}) I^j(s), \qquad P^{ij} \equiv \delta^{ij} - \widehat{r}^i\widehat{r}^j, \qquad \widehat{r}^i \equiv \frac{x^i}{|\vec{x}|}. \tag{12.4.228}$$

The subscript '(t)' stands for 'transverse'; the projector, which obeys $\widehat{r}^i P^{ij} = 0$, ensures the $I^i_{(t)}$ now consists only of the 'transverse' portion of the dipole moment: $\widehat{r}^i I^i_{(t)} = 0$. Moreover, this result indicates, not only the electric and magnetic fields are mutually perpendicular, they are also orthogonal to the radial direction and hence transverse to the propagation direction of (far zone) electromagnetic radiation. Finally, from eq. (12.4.225), notice we would have incorrectly concluded that the electric field in eq. (12.4.226) is not built solely from the transverse acceleration of the dipole, if the non-relativistic limit were taken too early.

Origin-independence Can you explain whether the results in equations (12.4.226) and (12.4.227) would change if we had shifted by a constant vector $\vec{b}$ the origin of integration in the definition of the dipole moment in eq. (12.4.223)? That is, what becomes of $\vec{E}$ and $\vec{B}$ if instead of eq. (12.4.223), we defined

$$I^i(\tau) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}'(x'^i - b^i)J^0(\tau, \vec{x}')? \tag{12.4.229}$$

4. Use the above results in equations (12.4.226) and (12.4.227) to compute the far zone Poynting vector $\vec{S} \equiv \vec{E} \times \vec{B}$, which describes the direction and rate of flow of momentum carried by electromagnetic waves. (The energy density $\mathcal{E}$ is the average $(\vec{E}^2 + \vec{B}^2)/2$.) Also verify the following projector property of P^{ij} in eq. (12.4.228):

$$P^{ia}P^{ib} = P^{ab} \tag{12.4.230}$$

and show that the dot product of the Poynting vector with the unit radial vector is

$$\frac{1}{r^2}\frac{\mathrm{d}^3 E}{\mathrm{d}t\mathrm{d}\Omega} \equiv \vec{S} \cdot \widehat{r} = \frac{1}{(4\pi)^2 r^2}\left(\ddot{\vec{I}}^2 - \left(\widehat{r} \cdot \ddot{\vec{I}}\right)^2\right) \tag{12.4.231}$$

$$\equiv \frac{\sin^2(\theta)}{(4\pi)^2 r^2}\left|\frac{\mathrm{d}^2\vec{I}(t-r)}{\mathrm{d}t^2}\right|^2. \tag{12.4.232}$$

For an arbitrary unit vector $\widehat{n}$, the dot product $\vec{S}(t,\vec{x}) \cdot \widehat{n}(t,\vec{x})$ is the energy per unit time per unit area passing through the infinitesimal plane orthogonal to the vector $\widehat{n}$ based at $(t,\vec{x})$. The quantity in eq. (12.4.231), if integrated over the 2-sphere, therefore describes the rate of loss of total energy to infinity as $r \to \infty$. □

4D gravitational waves In a 4D weakly curved spacetime, the metric can be written as one deviating slightly from Minkowski,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \qquad \text{(Cartesian coordinates)}, \tag{12.4.233}$$

where the dimensionless components of $h_{\mu\nu}$ are assumed to be much smaller than unity.

The (trace-reversed) graviton

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\eta^{\alpha\beta}h_{\alpha\beta}, \tag{12.4.234}$$

in the de Donder gauge

$$\partial^\mu \bar{h}_{\mu\nu} = \partial_t \bar{h}_{t\nu} - \delta^{ij}\partial_i \bar{h}_{j\nu} = 0, \tag{12.4.235}$$

obeys the wave equation[156]

$$\partial^2 \bar{h}_{\mu\nu} = -16\pi G_{\mathrm{N}} T_{\mu\nu} \qquad \text{(Cartesian coordinates)}. \tag{12.4.236}$$

[156] The following equation is only approximate; it comes from linearizing Einstein's equations about a flat spacetime background, i.e., where all terms quadratic and higher in $h_{\mu\nu}$ are discarded.

(The $G_{\rm N}$ is the same Newton's constant you see in Newtonian gravity $\sim G_{\rm N} M_1 M_2/r^2$; both $\bar{h}_{\mu\nu}$ and $T_{\mu\nu}$ are symmetric.) The $T_{\mu\nu}$ is a 4×4 matrix describing the energy-momentum-shear-stress density of matter, and has zero divergence – i.e., it is conserved – whenever the matter is held together primarily by non-gravitational forces:[157]

$$\partial_\mu T^{\mu\nu} = \partial_t T^{t\nu} + \partial_i T^{i\nu} = 0. \tag{12.4.237}$$

Problem 12.41. de Donder Gauge, Relativity & Energy-Momentum Conservation Comparison of equations (12.4.200) and (12.4.214) indicates, in the far zone,

$$\bar{h}^{\mu\nu}(t,\vec{x}) \approx -\frac{4G_{\rm N}}{r}\int \mathrm{d}^3\vec{x}' T^{\mu\nu}\left(t - r + \vec{x}'\cdot\widehat{r}, \vec{x}'\right). \tag{12.4.238}$$

If one takes the non-relativistic limit too (cf. eq. (12.4.201)),

$$\bar{h}^{\mu\nu}(t,\vec{x}) \approx -\frac{4G_{\rm N}}{r}\int \mathrm{d}^3\vec{x}' T^{\mu\nu}\left(t - r, \vec{x}'\right). \tag{12.4.239}$$

Compute $\partial_\mu \bar{h}^{\mu\nu}$ using equations (12.4.238) and (12.4.239) to leading order in $1/r$. Hint: a key step is to recognize, for a conserved energy-momentum-stress tensor obeying eq. (12.4.237),

$$\partial_t T^{0\mu}(\tau,\vec{x}') = -\left(\partial_{i'} T^{i\mu}(\tau,\vec{x}')\right)_t + \widehat{r}^i \partial_t T^{i\mu}(\tau,\vec{x}'); \tag{12.4.240}$$

$$\tau \equiv t - r + \vec{x}'\cdot\widehat{r}, \qquad \partial_{i'} \equiv \frac{\partial}{\partial x'^i}; \tag{12.4.241}$$

where the subscript t on the first term on the right-hand-side of eq. (12.4.240) means the spatial derivatives are carried out with the observation time t held fixed – which is to be distinguished from doing so but with τ held fixed.

You should find that the de Donder gauge in eq. (12.4.235) is respected only by the relativistic solution in eq. (12.4.238), and not by the non-relativistic one in eq. (12.4.239). This is an important point because, even though the de Donder gauge in eq. (12.4.235) was a mathematical choice, once we have chosen it to solve the linearized Einstein's equations, violating it may lead to a violation of stress-energy-momentum conservation: to see this, simply take the 4−divergence of eq. (12.4.236) to obtain $\partial^2(\partial^\mu \bar{h}_{\mu\nu}) = -16\pi G_{\rm N}\partial^\mu T_{\mu\nu}$. □

Problem 12.42. Gravitational Radiation YZ: This problem needs to be updated / revised. Can you carry out a similar analysis in Problem (12.40), but for gravitational radiation? Using G_4^+ in eq. (12.4.192), write down the solution of $\bar{h}^{\mu\nu}$ in terms of $T^{\mu\nu}$. Then take the far zone limit. Throughout this analysis, assume that $T^{\mu\nu}$ is sufficiently localized that it vanishes at spatial infinity; and assume $T^{\mu\nu}$ is a non-relativistic source.

1. Using $\partial_t T^{t\nu} = -\partial_i T^{i\nu}$, the conservation of the stress-tensor, show that $\bar{h}^{\nu 0} = \bar{h}^{0\nu}$ is independent of time in the far zone limit.

[157]For systems held together primarily by gravity, such as the Solar System or compact binary black hole(s)/neutron star(s) emitting gravitational radiation, the corresponding matter stress tensor will *not* be divergence-less.

2. Now define the quadrupole moment as

$$I^{ij}(\tau) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' x'^i x'^j T^{00}(\tau, \vec{x}'). \tag{12.4.242}$$

Can you show its second time derivative is

$$\ddot{I}^{ij}(\tau) \equiv \frac{\mathrm{d}^2 I^{ij}(\tau)}{\mathrm{d}\tau^2} = 2\int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' T^{ij}(\tau, \vec{x}')? \tag{12.4.243}$$

and from it infer that the (trace-reversed) gravitational wave form in the far zone is proportional to the acceleration of the quadrupole moment evaluated at retarded time:

$$\bar{h}^{ij}(t, \vec{x}) \to -\frac{2G_{\mathrm{N}}}{r}\frac{\mathrm{d}^2 I^{ij}(t-r)}{\mathrm{d}t^2}, \qquad r \equiv |\vec{x}|. \tag{12.4.244}$$

Origin-Independence Can you explain what would become of this result if, instead of the quadrupole moment defined in eq. (12.4.242), we had shifted its integration origin by a constant vector $\vec{b}$, namely

$$I^{ij}(t) \equiv \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' (x'^i - b^i)(x'^j - b^j) T^{00}(t, \vec{x}')? \tag{12.4.245}$$

3. Note that the transverse-traceless portion of this (trace-reversed) gravitational wave $\bar{h}_{ij}(t, \vec{x})$ can be detected by how it squeezes and stretches arms of a laser interferometer such as aLIGO and VIRGO.

$$h^{\mathrm{tt}}_{ij} = P_{ijab}\bar{h}_{ab}, \tag{12.4.246}$$

$$P_{ijab} \equiv \frac{1}{2}\left(P_{ia}P_{jb} + P_{ib}P_{ja} - P_{ij}P_{ab}\right), \qquad \text{(cf. eq. (12.4.228))}, \tag{12.4.247}$$

$$\widehat{r}^i h^{\mathrm{tt}}_{ij} = 0 \quad \text{(Transverse)} \qquad \delta^{ij} h^{\mathrm{tt}}_{ij} = 0 \quad \text{(Traceless)}. \tag{12.4.248}$$

Averaged over multiple wavelengths, the energy-momentum-stress tensor of gravitational waves takes the form

$$\langle t_{\mu\nu}[h^{\mathrm{tt}}]\rangle = \frac{1}{32\pi G_{\mathrm{N}}}\langle \partial_\mu h^{\mathrm{tt}}_{ij} \partial_\nu h^{\mathrm{tt}}_{ij}\rangle. \tag{12.4.249}$$

Can you work out the energy density $\mathcal{E} \equiv \langle t_{00}\rangle$ and the momentum flux $\mathcal{P}^i \equiv \langle t^{0i}\rangle = -\langle t_{0i}\rangle$ (the gravitational analog to the electromagnetic Poynting vector) in terms of the quadrupole moment? □

Problem 12.43. Waves Around Schwarzschild Black Hole. The geometry of a non-rotating black hole is described by

$$\mathrm{d}s^2 = \left(1 - \frac{r_s}{r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{r_s}{r}} - r^2\left(\mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\phi^2\right), \tag{12.4.250}$$

where $x^\mu = (t \in \mathbb{R}, r \geq 0, 0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$, and r_s (proportional to the mass of the black hole itself) is known as the Schwarzschild radius – nothing can fall inside the black hole ($r < r_s$) and still get out.

Consider the (massless scalar) homogeneous wave equation in this black hole spacetime, namely

$$\Box\psi(t, r, \theta, \phi) = \nabla_\mu \nabla^\mu \psi = 0. \tag{12.4.251}$$

Consider the following separation-of-variables ansatz

$$\psi(t, r, \theta, \phi) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega t} \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{+\ell} \frac{R_\ell(\omega r_*)}{r} Y_\ell^m(\theta, \phi), \tag{12.4.252}$$

where $\{Y_\ell^m\}$ are the spherical harmonics on the 2-sphere and the "tortoise coordinate" is

$$r_* \equiv r + r_s \ln\left(\frac{r}{r_s} - 1\right). \tag{12.4.253}$$

Show that the wave equation is reduced to an ordinary differential equation for the ℓth radial mode function

$$R_\ell''(\xi_*) + \left(\frac{\xi_s^2}{\xi^4} + \frac{(\ell(\ell+1) - 1)\,\xi_s}{\xi^3} - \frac{\ell(\ell+1)}{\xi^2} + 1\right) R_\ell(\xi_*) = 0, \tag{12.4.254}$$

where $\xi \equiv \omega r$, $\xi_s \equiv \omega r_s$ and $\xi_* \equiv \omega r_*$.

An alternative route is to first perform the change-of-variables

$$x \equiv 1 - \frac{\xi}{\xi_s}, \tag{12.4.255}$$

and the change of radial mode function

$$\frac{R_\ell(\xi_*)}{r} \equiv \frac{Z_\ell(x)}{\sqrt{x(1-x)}}. \tag{12.4.256}$$

Show that this returns the ODE

$$Z_\ell''(x) + \left(\frac{1}{4(x-1)^2} + \frac{1 + 4\xi_s^2}{4x^2} + \xi_s^2 + \frac{2\ell(\ell+1) + 1 - 4\xi_s^2}{2x} - \frac{2\ell(\ell+1)+1}{2(x-1)}\right) Z_\ell(x) = 0. \tag{12.4.257}$$

You may use Mathematica or similar software to help you with the tedious algebra/differentiation; but make sure you explain the intermediate steps clearly.

The solutions to eq. (12.4.257) are related to the confluent Heun function. For a recent discussion, see for e.g., §I of arXiv: 1510.06655. The properties of Heun functions are not as well studied as, say, the Bessel functions you have encountered earlier. This is why it is still a subject of active research – see, for instance, the Heun Project. □

12.4.4 Massless Fields in (3+1)D: Frequency Space

Wave Equation in Frequency Space We begin with eq. (12.4.59), and translate it to frequency space.

$$\begin{aligned}\psi(t,\vec{x}) &= \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\widetilde{\psi}(\omega,\vec{x})e^{-i\omega t} \\ &= \int_{-\infty}^{+\infty}\mathrm{d}t''\int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}''G_{D+1}(t-t'',\vec{x}-\vec{x}'')\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\widetilde{J}(\omega,\vec{x}'')e^{-i\omega t''} \\ &= \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\int_{-\infty}^{+\infty}\mathrm{d}(t-t'')e^{i\omega(t-t'')}e^{-i\omega t}\int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}''G_{D+1}(t-t'',\vec{x}-\vec{x}'')\widetilde{J}(\omega,\vec{x}'') \\ &= \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega t}\int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}''\widetilde{G}^+_{D+1}(\omega,\vec{x}-\vec{x}'')\widetilde{J}(\omega,\vec{x}''). \end{aligned} \tag{12.4.258}$$

Equating the coefficients of $e^{-i\omega t}$ on both sides,

$$\widetilde{\psi}(\omega,\vec{x}) = \int_{\mathbb{R}^D}\mathrm{d}^D\vec{x}''\widetilde{G}^+_{D+1}(\omega,\vec{x}-\vec{x}'')\widetilde{J}(\omega,\vec{x}''); \tag{12.4.259}$$

$$\widetilde{G}^+_{D+1}(\omega,\vec{x}-\vec{x}'') \equiv \int_{-\infty}^{+\infty}\mathrm{d}\tau e^{i\omega\tau}G_{D+1}(\tau,\vec{x}-\vec{x}''). \tag{12.4.260}$$

Equation (12.4.259) tells us that the ω-mode of the source is directly responsible for that of the field $\widetilde{\psi}(\omega,\vec{x})$. This is reminiscent of the driven harmonic oscillator system, except now we have one oscillator per point in space $\vec{x}'$ – hence the integral over all of them.

4D Retarded Green's Function in Frequency Space Next, we focus on the $(D+1)=(3+1)$ case, and re-visit the 4D retarded Green's function result in eq. (12.4.192), but replace the δ-function with its integral representation. This leads us to $\widetilde{G}^+_4(\omega,\vec{x}-\vec{x}')$, the frequency space representation of the retarded Green's function of the wave operator.

$$\begin{aligned}G^+_4(x-x') &= \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{\exp\left(-i\omega(t-t'-|\vec{x}-\vec{x}'|)\right)}{4\pi|\vec{x}-\vec{x}'|} \\ &\equiv \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega(t-t')}\widetilde{G}^+_4(\omega,\vec{x}-\vec{x}'),\end{aligned} \tag{12.4.261}$$

where

$$\widetilde{G}^+_4(\omega,\vec{x}-\vec{x}') \equiv \frac{\exp\left(i\omega|\vec{x}-\vec{x}'|\right)}{4\pi|\vec{x}-\vec{x}'|}. \tag{12.4.262}$$

As we will see, ω can be interpreted as the frequency of the source of the waves. In this section we will develop a multipole expansion of the field in frequency space by performing one for the source as well. This will allow us to readily take the non-relativistic/static limit, where the motion of the sources (in some center of mass frame) is much slower than 1.

Because the $(3+1)$-dimensional case of eq. (12.4.63) in frequency space reads

$$\left(\partial_0^2-\vec{\nabla}^2\right)\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{\exp\left(-i\omega(t-t'-|\vec{x}-\vec{x}'|)\right)}{4\pi|\vec{x}-\vec{x}'|} = \delta(t-t')\delta^{(3)}\left(\vec{x}-\vec{x}'\right), \tag{12.4.263}$$

$$\int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega(t-t')} \left(-\omega^2 - \vec{\nabla}^2\right) \frac{\exp\left(i\omega|\vec{x}-\vec{x}'|\right)}{4\pi|\vec{x}-\vec{x}'|} = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega(t-t')} \delta^{(3)}\left(\vec{x}-\vec{x}'\right), \quad (12.4.264)$$

– where ∂_0^2 can be either ∂_t^2 or $\partial_{t'}^2$; $\vec{\nabla}^2$ can be either $\vec{\nabla}_{\vec{x}}$ or $\vec{\nabla}_{\vec{x}'}$; and we have replaced $\delta(t-t')$ with its integral representation – we can equate the coefficients of the (linearly independent) functions $\{\exp(-i\omega(t-t'))\}$ on both sides to conclude, for fixed ω, the frequency space Green's function of eq. (12.4.262) obeys the PDE

$$\left(-\omega^2 - \vec{\nabla}^2\right) \widetilde{G}_4^+(\omega, \vec{x}-\vec{x}') = \delta^{(3)}\left(\vec{x}-\vec{x}'\right). \quad (12.4.265)$$

Problem 12.44. Far Zone In Frequency Space Show that the frequency transform of the far zone wave in eq. (12.4.200) is

$$\widetilde{\psi}(\omega, \vec{x}) \approx \frac{e^{i\omega r}}{4\pi r} \widetilde{J}\left(\omega, \omega\widehat{r}\right), \qquad r \equiv |\vec{x}|, \quad (12.4.266)$$

where

$$\widetilde{J}(\omega, \vec{k}) \equiv \int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x} e^{+i\omega t} e^{-i\vec{k}\cdot\vec{x}} J\left(t, \vec{x}\right). \quad (12.4.267)$$

We will re-derive this result below, but as a multi-pole expansion. □

Static Limit Equals Zero Frequency Limit In any (curved) spacetime that enjoys time translation symmetry – which, in particular, means there is some coordinate system where the metric $g_{\mu\nu}(\vec{x})$ depends only on space $\vec{x}$ and not on time t – we expect the Green's function of the wave operator to reflect the symmetry and take the form $G^+(t-t'; \vec{x}, \vec{x}')$. Furthermore, the wave operator only involves time through derivatives, i.e., eq. (12.4.15) now reads

$$\begin{aligned} \nabla_\mu \nabla^\mu G &= g^{tt}\partial_t\partial_t G + g^{ti}\partial_t\partial_i G + \frac{\partial_i\left(\sqrt{|g|}g^{ti}\partial_t G\right)}{\sqrt{|g|}} + \frac{1}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j G\right) \\ &= \frac{\delta(t-t')\delta^{(D)}\left(\vec{x}-\vec{x}'\right)}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}; \end{aligned} \quad (12.4.268)$$

since $\sqrt{|g|}$ and $g^{\mu\nu}$ are time-independent. In such a time-translation-symmetric situation, we may perform a frequency transform

$$\widetilde{G}^+(\omega; \vec{x}, \vec{x}') = \int_{-\infty}^{+\infty} \mathrm{d}\tau e^{i\omega\tau} G^+\left(\tau; \vec{x}, \vec{x}'\right), \quad (12.4.269)$$

and note that solving the *static* equation

$$\begin{aligned} \nabla_\mu \nabla^\mu G^{(\text{static})}\left(\vec{x}, \vec{x}'\right) &= \frac{\partial_i\left(\sqrt{|g(\vec{x})|}g^{ij}(\vec{x})\partial_j G^{(\text{static})}\left(\vec{x}, \vec{x}'\right)\right)}{\sqrt{|g(\vec{x})|}} \\ &= \frac{\partial_{i'}\left(\sqrt{|g(\vec{x}')|}g^{ij}(\vec{x}')\partial_{j'} G^{(\text{static})}\left(\vec{x}, \vec{x}'\right)\right)}{\sqrt{|g(\vec{x}')|}} = \frac{\delta^{(D)}(\vec{x}-\vec{x}')}{\sqrt[4]{g(\vec{x})g(\vec{x}')}}, \end{aligned} \quad (12.4.270)$$

amounts to taking the zero frequency limit of the frequency space retarded Green's function. Note that the static equation still depends on the full $(D+1)$ dimensional metric, but the δ-functions on the right hand side is D-dimensional.

The reason is the frequency transform of eq. (12.4.268) replaces $\partial_t \to -i\omega$ and the $\delta(t-t')$ on the right hand side with unity.

$$\begin{aligned} g^{tt}(-i\omega)^2\widetilde{G} + g^{ti}(-i\omega)\partial_i\widetilde{G} + \frac{\partial_i\left(\sqrt{|g|}g^{ti}(-i\omega)G\right)}{\sqrt{|g|}} \\ +\frac{1}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j\widetilde{G}\right) = \frac{\delta^{(D)}\left(\vec{x}-\vec{x}'\right)}{\sqrt[4]{g(\vec{x})g(\vec{x}')}} \end{aligned} \tag{12.4.271}$$

In the zero frequency limit $(\omega \to 0)$ we obtain eq. (12.4.270). And since the static limit is the zero frequency limit,

$$\begin{aligned} G^{(\text{static})}(\vec{x},\vec{x}') &= \lim_{\omega\to 0}\int_{-\infty}^{+\infty}\mathrm{d}\tau e^{i\omega\tau}G^{+}\left(\tau;\vec{x},\vec{x}'\right), \qquad (12.4.272) \\ &= \int_{-\infty}^{+\infty}\mathrm{d}\tau G^{+}\left(\tau;\vec{x},\vec{x}'\right) = \int_{-\infty}^{+\infty}\mathrm{d}\tau\int \mathrm{d}^D\vec{x}''\sqrt{|g(\vec{x}'')|}G^{+}\left(\tau;\vec{x},\vec{x}''\right)\frac{\delta^{(D)}(\vec{x}'-\vec{x}'')}{\sqrt{|g(\vec{x}')g(\vec{x}'')|}}. \end{aligned}$$

This second line has the following interpretation: not only is the static Green's function the zero frequency limit of its frequency space retarded counterpart, it can also be viewed as the field generated by a point "charge/mass" held still at $\vec{x}'$ from past infinity to future infinity.[158]

4D Minkowski Example We may illustrate our discussion here by examining the 4D Minkowski case. The field generated by a charge/mass held still at $\vec{x}'$ is nothing but the Coulomb/Newtonian potential $1/(4\pi|\vec{x}-\vec{x}'|)$. Since we also know the 4D Minkowski retarded Green's function in eq. (12.4.192), we may apply the infinite time integral in eq. (12.4.272).

$$G^{(\text{static})}(\vec{x},\vec{x}') = \int_{-\infty}^{+\infty}\mathrm{d}\tau\frac{\delta(\tau-|\vec{x}-\vec{x}'|)}{4\pi|\vec{x}-\vec{x}'|} = \frac{1}{4\pi|\vec{x}-\vec{x}'|}, \tag{12.4.273}$$

$$-\delta^{ij}\partial_i\partial_j G^{(\text{static})}(\vec{x},\vec{x}') = -\vec{\nabla}^2 G^{(\text{static})}(\vec{x},\vec{x}') = \delta^{(3)}(\vec{x}-\vec{x}'). \tag{12.4.274}$$

On the other hand, we may also take the zero frequency limit of eq. (12.4.262) to arrive at the same answer.

$$\lim_{\omega\to 0}\frac{\exp\left(i\omega|\vec{x}-\vec{x}'|\right)}{4\pi|\vec{x}-\vec{x}'|} = \frac{1}{4\pi|\vec{x}-\vec{x}'|}. \tag{12.4.275}$$

Problem 12.45. Outgoing Waves: Frequency Space For a fixed frequency,

$$\psi(t,\vec{x}) \equiv e^{-i\omega\cdot t}\widetilde{\psi}(\vec{x}), \tag{12.4.276}$$

show that the *outgoing* waves take the form – expressed in spherical coordinates (r,θ,ϕ) –

$$\widetilde{\psi}(\vec{x}) = \sum_{\ell=0}^{\infty}h_\ell^{(1)}\left(\omega r\right)\sum_{m=-\ell}^{+\ell}\chi_\ell^m(\omega)Y_\ell^m\left(\widehat{r}\right). \tag{12.4.277}$$

[158]Note, however, that in curved spacetimes, holding still a charge/mass – ensuring it stays put at $\vec{x}'$ – requires external forces. For example, holding a mass still in a spherically symmetric gravitational field of a star requires an outward external force, for otherwise the mass will move towards the center of the star.

Next, show that the far zone $\omega r \gg 1$ limit of a general outgoing wave must take the form

$$\widetilde{\psi}(\vec{x}) \to \frac{e^{i\omega \cdot r}}{r} f(\widehat{r}), \tag{12.4.278}$$

where, with the unit radial vector denoted as $\widehat{r}(\theta, \phi)$, $f(\widehat{r})$ is an arbitrary function of the angles on the 2−sphere.

Hint: You may find the following large argument limit useful:

$$h_\ell^{(1)}(x \gg 1) \to (-i)^{\ell+1} \frac{e^{ix}}{x}. \tag{12.4.279}$$

Problem 12.46. Mode Decomposition of $\widetilde{G}_4^+(\omega, \vec{x} - \vec{x}')$: Spherical Coordinates In this problem, let us perform in spherical coordinates a mode decomposition of the frequency space retarded Green's function in eq. (12.4.262), expressing it as a sum over mode functions taking a separated form; whereby, the dependence on each coordinate (1 radial and 2 angular ones) occurs in a separate function. On a more technical level, we will also understand the discontinuity in the radial portion of the Green's function.

We begin by switching from Cartesian to spherical coordinates,

$$\vec{x} = r\left(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta\right), \tag{12.4.280}$$
$$\vec{x}' = r'\left(\sin\theta'\cos\phi', \sin\theta'\sin\phi', \cos\theta'\right); \tag{12.4.281}$$

and utilizing the following ansatz

$$\frac{\exp\left(i\omega|\vec{x} - \vec{x}'|\right)}{4\pi|\vec{x} - \vec{x}'|} = \sum_{\ell=0}^{\infty} \widetilde{g}_\ell(r, r') \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta, \phi)\overline{Y_\ell^m(\theta', \phi')}, \tag{12.4.282}$$

Show that this leads to the following ODE(s) for the ℓth radial Green's function $\widetilde{g}_\ell$:

$$\frac{1}{r^2}\partial_r\left(r^2\partial_r\widetilde{g}_\ell\right) + \left(\omega^2 - \frac{\ell(\ell+1)}{r^2}\right)\widetilde{g}_\ell = -\frac{\delta(r - r')}{rr'}, \tag{12.4.283}$$
$$\frac{1}{r'^2}\partial_{r'}\left(r'^2\partial_{r'}\widetilde{g}_\ell\right) + \left(\omega^2 - \frac{\ell(\ell+1)}{r'^2}\right)\widetilde{g}_\ell = -\frac{\delta(r - r')}{rr'}. \tag{12.4.284}$$

Because $\widetilde{G}_4^+(\omega, \vec{x} - \vec{x}') = \widetilde{G}_4^+(\omega, \vec{x}' - \vec{x})$, i.e., it is symmetric under the exchange of the spatial coordinates of source and observer, it is reasonable to expect that the radial Green's function is symmetric too: $\widetilde{g}(r, r') = \widetilde{g}(r', r)$. That means the results in §(5.11) may be applied here. Show that

$$\widetilde{g}_\ell(r, r') = i\omega j_\ell(\omega r_<)h_\ell^{(1)}(\omega r_>), \tag{12.4.285}$$

where $j_\ell(z)$ is the spherical Bessel function and $h_\ell^{(1)}(z)$ is the Hankel function of the first kind. Then check that the static limit in eq. (12.7.47) is recovered, by taking the limits $\omega r, \omega r' \to 0$.

Some useful formulas include

$$j_\ell(x) = (-x)^\ell \left(\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^\ell \frac{\sin x}{x}, \qquad h_\ell^{(1)}(x) = -i(-x)^\ell \left(\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^\ell \frac{\exp(ix)}{x}, \tag{12.4.286}$$

their small argument limits

$$j_\ell(x \ll 1) \to \frac{x^\ell}{(2\ell+1)!!}\left(1 + \mathcal{O}(x^2)\right), \qquad h_\ell^{(1)}(x \ll 1) \to -\frac{i(2\ell-1)!!}{x^{\ell+1}}\left(1 + \mathcal{O}(x)\right), \tag{12.4.287}$$

as well as their large argument limits

$$j_\ell(x \gg 1) \to \frac{1}{x}\sin\left(x - \frac{\pi\ell}{2}\right), \qquad h_\ell^{(1)}(x \gg 1) \to (-i)^{\ell+1}\frac{e^{ix}}{x}. \tag{12.4.288}$$

Their Wronskian is

$$\mathrm{Wr}_z\left(j_\ell(z), h_\ell^{(1)}(z)\right) = \frac{i}{z^2}. \tag{12.4.289}$$

Hints: First explain why

$$\widetilde{g}_\ell(r, r') = A_\ell^1 j_\ell(\omega r) j_\ell(\omega r') + A_\ell^2 h_\ell^{(1)}(\omega r) h_\ell^{(1)}(\omega r') + \mathcal{G}_\ell(r, r'), \tag{12.4.290}$$

$$\mathcal{G}_\ell(r, r') \equiv F\left\{(\chi_\ell - 1) j_\ell(\omega r_>) h_\ell^{(1)}(\omega r_<) + \chi_\ell \cdot j_\ell(\omega r_<) h_\ell^{(1)}(\omega r_>)\right\}, \tag{12.4.291}$$

where $A_\ell^{1,2}$, F and χ_ℓ are constants. Fix F by ensuring the "jump" in the first r-derivative at $r = r'$ yields the correct δ-function measure. Then consider the limits $r \to 0$ and $r \gg r'$. For the latter, note that

$$|\vec{x} - \vec{x}'| = e^{-\vec{x}'\cdot\vec{\nabla}_{\vec{x}}}|\vec{x}| = |\vec{x}|\left(1 - (r'/r)\widehat{r}\cdot\widehat{r}' + \mathcal{O}((r'/r)^2)\right), \tag{12.4.292}$$

where $\widehat{r} \equiv \vec{x}/r$ and $\widehat{r}' \equiv \vec{x}'/r'$. $\square$

We will now proceed to understand the utility of obtaining such a mode expansion of the frequency space Green's function.

Localized source(s): Static Multipole Expansion In infinite flat $\mathbb{R}^3$, Poisson's equation

$$-\vec{\nabla}^2\psi(\vec{x}) = J(\vec{x}) \tag{12.4.293}$$

is solved via the static limit of the 4D retarded Green's function we have been discussing. This static limit is given in eq. (12.7.47) in spherical coordinates, which we will now exploit to display its usefulness. In particular, assuming the source J is localized in space, we may now ask:

> What is the field generated by J and how does it depend on the details of its interior?

Let the origin of our coordinate system lie at the center of mass of the source J, and let R be its maximum radius, i.e., $J(r > R) = 0$. Therefore we may replace $r_< \to r'$ and $r_> \to r$ in eq. (12.7.47), and the exact solution to ψ now reads

$$\psi(\vec{x}; r > R) = \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}' G(\vec{x} - \vec{x}') J(\vec{x}') = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{+\ell} \frac{\rho_\ell^m}{2\ell+1}\frac{Y_\ell^m(\theta,\phi)}{r^{\ell+1}}, \tag{12.4.294}$$

where the multipole moments $\{\rho_\ell^m\}$ are defined

$$\rho_\ell^m \equiv \int_{\mathbb{S}^2} \mathrm{d}(\cos\theta')\mathrm{d}\phi' \int_0^\infty \mathrm{d}r' r'^{\ell+2}\, \overline{Y_\ell^m(\theta',\phi')} J(r',\theta',\phi'). \tag{12.4.295}$$

It is worthwhile to highlight the following.

- The spherical harmonics can be roughly thought of as waves on the 2−sphere. Therefore, the multipole moments ρ_ℓ^m in eq. (12.4.295) with larger ℓ and m values, describe the shorter wavelength/finer features of the interior structure of J. (Recall the analogous discussion for Fourier transforms.)

- Moreover, since there is a $Y_\ell^m(\theta,\phi)/r^{\ell+1}$ multiplying the (ℓ,m)-moment of J, we see that the finer features of the field detected by the observer at $\vec{x}$ is not only directly sourced by finer features of J, it falls off more rapidly with increasing distance from J. As the observer moves towards infinity, the dominant part of the field ψ is the monopole which goes as $1/r$ times the total mass/charge of J.

- We see why separation-of-variables is not only a useful mathematical technique to reduce the solution of Green's functions from a PDE to a bunch of ODE's, it was the form of eq. (12.7.47) that allowed us to cleanly separate the contribution from the source (the multipoles $\{\rho_\ell^m\}$) from the form of the field they would generate, at least on a mode-by-mode basis.

Localized source(s): General Multipole Expansions, Far Zone Let us generalize the static case to the fully time dependent one, but in frequency space and in the far zone. By the far zone, we mean the observer is located very far away from the source J, at distances (from the center of mass) much further than the typical inverse frequency of $\widetilde{J}$, i.e., mathematically, $\omega r \gg 1$. We begin with eq. (12.4.285) inserted into eq. (12.4.282).

$$\widetilde{G}_4^+(\omega,\vec{x}-\vec{x}') = \frac{\exp\left(i\omega|\vec{x}-\vec{x}'|\right)}{4\pi|\vec{x}-\vec{x}'|} \tag{12.4.296}$$

$$= i\omega \sum_{\ell=0}^{\infty} j_\ell(\omega r_<) h_\ell^{(1)}(\omega r_>) \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta,\phi) Y_\ell^m(\theta',\phi')^* \tag{12.4.297}$$

Our far zone assumptions means we may replace the Hankel function in eq. (12.4.285) with its large argument limit in eq. (12.4.288).

$$\widetilde{G}_4^+(\omega r \gg 1) = \frac{e^{i\omega r}}{r}\left(1+\mathcal{O}\left((\omega r)^{-1}\right)\right)\sum_{\ell=0}^{\infty}(-i)^\ell j_\ell(\omega r')\sum_{m=-\ell}^{\ell} Y_\ell^m(\theta,\phi)Y_\ell^m(\theta',\phi')^*. \tag{12.4.298}$$

Applying this limit to the general wave solution in eq. (12.4.259),

$$\widetilde{\psi}(\omega,\vec{x}) = \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}'' \widetilde{G}_4^+(\omega,\vec{x}-\vec{x}'')\widetilde{J}(\omega,\vec{x}''), \tag{12.4.299}$$

$$\widetilde{\psi}(\omega r \gg 1) \approx \frac{e^{i\omega r}}{r}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{Y_\ell^m(\theta,\phi)}{2\ell+1}\widetilde{\Omega}_\ell^m(\omega), \tag{12.4.300}$$

where now the frequency dependent multipole moments are defined as

$$\widetilde{\Omega}_\ell^m(\omega) \equiv (2\ell+1)(-i)^\ell \int_{\mathbb{S}^2}\mathrm{d}(\cos\theta')\mathrm{d}\phi' \int_0^\infty \mathrm{d}r' r'^2 j_\ell(\omega r')\overline{Y_\ell^m(\theta',\phi')}\widetilde{J}(\omega,r',\theta',\phi'). \tag{12.4.301}$$

Problem 12.47. Far zone in position/real space Use the plane wave expansion in eq. (12.2.104) to show that eq. (12.4.300) is equivalent to eq. (12.4.266). □

Low frequency limit equals slow motion limit How are the multipole moments $\{\rho_\ell^m\}$ in eq. (12.4.295) (which are pure numbers) related to the frequency dependent ones $\{\widetilde{\Omega}_\ell^m(\omega)\}$ in eq. (12.4.301)? The answer is that the low frequency limit is the slow-motion/non-relativistic limit. To see this in more detail, we take the $\omega r' \ll 1$ limit, which amounts to the physical assumption that the object described by J is localized so that its maximum radius R (from its center of mass) is much smaller than the inverse frequency. In other words, in units where the speed of light is unity, the characteristic size R of the source J is much smaller than the time scale of its typical time variation. Mathematically, this $\omega r' \ll 1$ limit is achieved by replacing $j_\ell(\omega r')$ with its small argument limit in eq. (12.4.287).

$$\widetilde{\Omega}_\ell^m(\omega R \ll 1) \approx \frac{(-i\omega)^\ell}{(2\ell-1)!!}\left(1+\mathcal{O}(\omega^2)\right)\int_{\mathbb{S}^2}\mathrm{d}(\cos\theta')\mathrm{d}\phi'\int_0^\infty \mathrm{d}r' r'^{2+\ell}\overline{Y_\ell^m(\theta',\phi')}\widetilde{J}(\omega,r',\theta',\phi') \tag{12.4.302}$$

Another way to see this "small ω equals slow motion limit" is to ask: what is the real time representation of these $\{\widetilde{\Omega}_\ell^m(\omega R \ll 1)\}$? By recognizing every $-i\omega$ as a t-derivative,

$$\begin{aligned}\Omega_\ell^m(t) &\equiv \int_{\mathbb{R}}\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega t}\widetilde{\Omega}_\ell^m(\omega) \\ &\approx \frac{\partial_t^\ell}{(2\ell-1)!!}\int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega t}\int_{\mathbb{S}^2}\mathrm{d}(\cos\theta')\mathrm{d}\phi'\int_0^\infty \mathrm{d}r' r'^{2+\ell}\overline{Y_\ell^m(\theta',\phi')}\widetilde{J}(\omega,r',\theta',\phi'), \\ &\equiv \frac{\partial_t^\ell \rho_\ell^m(t)}{(2\ell-1)!!}.\end{aligned} \tag{12.4.303}$$

We see that the $\omega R \ll 1$ is the slow motion/non-relativistic limit because it is in this limit that time derivatives vanish. This is also why the only $1/r$ piece of the static field in eq. (12.4.294) comes from the monopole.

Spherical waves in small ω limit In this same limit, we may re-construct the real time scalar field, and witness how it is a superposition of spherical waves $\exp(i\omega(r-t))/r$. The observer detects a field that depends on the time derivatives of the multipole moments evaluated at retarded time $t-r$.

$$\begin{aligned}\psi(t,\vec{x}) &= \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega t}\widetilde{\psi}(\omega,\vec{x}) \\ &\approx \int_{-\infty}^{+\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{e^{i\omega(r-t)}}{r}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{Y_\ell^m(\theta,\phi)}{2\ell+1}\widetilde{\Omega}_\ell^m(\omega), \qquad \text{(Far zone spherical wave expansion)} \\ &\approx \frac{1}{r}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{Y_\ell^m(\theta,\phi)}{(2\ell+1)!!}\frac{\mathrm{d}^\ell\rho_\ell^m(t-r)}{\mathrm{d}t^\ell}, \qquad \text{(Slow motion limit).}\end{aligned} \tag{12.4.304}$$

Problem 12.48. Mode Decomposition of $\widetilde{G}_4^+(\omega,\vec{x}-\vec{x}')$: Cylindrical Coordinates In this problem, let us perform in cylindrical coordinates a mode decomposition of the frequency space retarded Green's function in eq. (12.4.262), expressing it as a sum over mode functions taking a separated form; whereby, the dependence on each of the 3 spatial coordinates occurs in a separate function. By switching from Cartesian to spherical coordinates,

$$\vec{x} = (\rho\cos\phi, \rho\sin\phi, z)\,, \tag{12.4.305}$$

$$\vec{x}' = (\rho'\cos\phi', \rho'\sin\phi', z')\,; \tag{12.4.306}$$

show that the appropriate decomposition of eq. (12.4.262) reads

$$\frac{\exp\left(i\omega|\vec{x}-\vec{x}'|\right)}{4\pi|\vec{x}-\vec{x}'|} = \frac{i}{4}\sum_{m\in\mathbb{Z}}\int_{\mathbb{R}}\frac{\mathrm{d}k}{2\pi}J_m\left(\sqrt{\omega^2-k^2}\rho_<\right)H_m^{(1)}\left(\sqrt{\omega^2-k^2}\rho_>\right)$$
$$\times\exp(im(\phi-\phi'))\exp(ik(z-z')), \tag{12.4.307}$$

where $\rho_<$ and $\rho_>$ are, respectively, the smaller and larger of the pair (ρ, ρ'). □

Problem 12.49. Static Limit: 3D Take the $\omega \to 0$ limit of equations (12.4.297) and (12.4.307) to obtain the variables-separated mode expansion in the static limit – i.e., that of

$$\widetilde{G}(\omega = 0, \vec{x}, \vec{x}') = (-\vec{\nabla}^2)^{-1}(\vec{x}-\vec{x}'). \tag{12.4.308}$$

You should find

$$\widetilde{G}(\omega = 0, \vec{x}, \vec{x}') = \sum_{\ell=0}^{\infty}\frac{1}{r_>}\left(\frac{r_<}{r_>}\right)^{\ell}\frac{Y_\ell^m(\widehat{r})\overline{Y_\ell^m(\widehat{r}')}}{2\ell+1} \qquad \text{(Spherical coords.)} \tag{12.4.309}$$

$$= \sum_{m\in\mathbb{Z}}\int_0^{\infty}\frac{\mathrm{d}k}{2\pi^2}I_m\left(k\rho_<\right)K_m\left(k\rho_>\right) \qquad \text{(Cylindrical coords.)}$$
$$\times\, e^{im(\phi-\phi')}\cos(k(z-z')); \tag{12.4.310}$$

where I_m and K_m are modified Bessel functions. □

Problem 12.50. Static Limit: 2D from 3D Employ the result in eq. (12.4.310) to deduce, the inverse of the negative Laplacian in 2D is

$$\widetilde{G}_{\mathrm{2D}}(\omega = 0, \vec{x}, \vec{x}') = -\frac{1}{2\pi}\left\{\ln\left(\mu\cdot\rho_>\right) - \sum_{\ell=1}^{\infty}\frac{1}{\ell}\left(\frac{\rho_<}{\rho_>}\right)^{\ell}\cos\Big(\ell(\phi-\phi')\Big)\right\}, \tag{12.4.311}$$

where μ is an arbitrary inverse length scale to render the argument of the natural logarithm dimensionless; and

$$\vec{x} = \rho(\cos\phi, \sin\phi), \tag{12.4.312}$$

$$\vec{x}' = \rho'(\cos\phi', \sin\phi'), \tag{12.4.313}$$

are the Cartesian coordinate vectors. Upon comparison with eq. (2.0.54), we see that

$$(-\vec{\nabla}^2)^{-1}(\vec{x}-\vec{x}') = -\frac{\ln(\mu|\vec{x}-\vec{x}'|)}{2\pi}. \tag{12.4.314}$$

Hint: Use the insight that the Green's function in $D-$dimensional flat space can be sourced by a line source extending into one higher dimension. For this purpose it may be useful to first extend the $k-$integration domain in eq. (12.4.310) from $\mathbb{R}^+$ to $\mathbb{R}$ after replacing the ks in the Bessel functions with $|k|$ and dividing the integrand by 2. □

12.4.5 Frequency Space Green's Functions of Wave Equations in Flat $D+1 \geq 3$

In the previous section, §(12.4.4), we have witnessed how the 4D frequency space Green's function of the wave operator was very useful in

12.4.6 Initial/final value problem via Kirchhoff representation

Massless scalar fields Previously we showed how, if we specified the initial conditions for the scalar field ψ – then via their Fourier transforms – eq. (12.4.51) tells us how they will evolve forward in time. Now we will derive an analogous expression that is valid in curved spacetime, first using the retarded Green's function G^+_{D+1}. To begin, the appropriate generalization of equations (12.4.16) and (12.4.63) are

$$\begin{aligned}\Box_x \psi(x) &= J(x), \\ \Box_x G^+_{D+1}(x,x') = \Box_{x'} G^+_{D+1}(x,x') &= \frac{\delta^{(D+1)}(x-x')}{\sqrt[4]{|g(x)g(x')|}}.\end{aligned} \tag{12.4.315}$$

Let us consider some 'cylindrical' domain of spacetime $\mathfrak{D}^+$ with spatial boundaries $\partial\mathfrak{D}^+_s$, initial time t_0 and final time $t_>$ hypersurfaces $\partial\mathfrak{D}(t_0)$ and $\partial\mathfrak{D}^+(t_>)$ respectively. (These constant time hypersurfaces need not correspond to the same time coordinate used in the integration.) We will consider an observer residing at x within this domain $\mathfrak{D}^+$ and the following integral.

$$I(x \in \mathfrak{D}^+) \equiv \int_{\mathfrak{D}^+} \mathrm{d}^{D+1}x' \sqrt{|g(x')|} \left\{ G^+_{D+1}(x,x')\Box_{x'}\psi(x') - \Box_{x'}G^+_{D+1}(x,x') \cdot \psi(x') \right\} \tag{12.4.316}$$

If we employ the equations for the scalar $\Box\psi = J$ and that for the Green's function $\Box G = \delta/\sqrt{|g|}$,

$$I(x \in \mathfrak{D}^+) = \int_{\mathfrak{D}^+} \sqrt{|g(x')|} G^+_{D+1}(x,x') J(x') \mathrm{d}^{D+1}x' - \psi(x). \tag{12.4.317}$$

On the other hand, if we integrate-by-parts,

$$\begin{aligned}I(x \in \mathfrak{D}^+) &= \int_{\partial\mathfrak{D}^+} \mathrm{d}^D\Sigma_{\alpha'} \left\{ G^+_{D+1}(x,x')\nabla^{\alpha'}\psi(x') - \nabla^{\alpha'}G^+_{D+1}(x,x') \cdot \psi(x') \right\} \\ &\quad - \int_{\mathfrak{D}^+} \mathrm{d}^{D+1}x' \sqrt{|g(x')|} \left\{ \nabla_{\alpha'}G^+_{D+1}(x,x')\nabla^{\alpha'}\psi(x') - \nabla_{\alpha'}G^+_{D+1}(x,x')\nabla^{\alpha'}\psi(x') \right\}.\end{aligned} \tag{12.4.318}$$

The terms in the very last line cancel. What remains in the second equality is the surface integrals over the spatial boundaries $\partial\mathfrak{D}^+_s$, and constant time hypersurfaces $\partial\mathfrak{D}^+(t_>)$ and $\partial\mathfrak{D}(t_0)$ – where we have used the Gauss' theorem in eq. (9.7.84). Here is where there is a significant difference between the curved space setup and the curved spacetime one at hand. By causality, since we have G^+_{D+1} in the integrand, the constant time hypersurface $\partial\mathfrak{D}^+(t_>)$ cannot contribute to the integral because it lies to the future of x. (Drawing a spacetime diagram here helps.) Finally, note that if ψ on spatial boundaries are known, we shall proceed to impose Dirichlet boundary conditions on the Green's function,

$$G^\pm_{D+1}\left(x \in \partial\mathfrak{D}^+_s, x'\right) = 0 = G^\pm_{D+1}\left(x, x' \in \partial\mathfrak{D}^+_s\right). \tag{12.4.319}$$

Equating equations (12.4.317) and (12.4.318), we arrive at the relationship between the field $\psi(x)$ at some arbitrary observer location $x \in \mathfrak{D}^+$ and its initial profile $\psi(x'(\vec{\xi}))$, initial velocity $u^{\alpha'}\nabla_{\alpha'}\psi(x'(\vec{\xi}))$, and its spatial boundary profile $\psi(x'(y))$:

$$\begin{aligned}
&\psi(x; x^0 > t_0) \qquad (12.4.320)\\
&= \int_{\mathfrak{D}^+} \mathrm{d}^{D+1}x' \sqrt{|g(x')|} G^+_{D+1}(x,x') J(x') + \int_{\partial\mathfrak{D}_s^+} \mathrm{d}^D y \sqrt{|\mathcal{H}(y)|} n^{\alpha'} \nabla_{\alpha'} G^+_{D+1}\left(x, x'(y)\right) \psi\left(x'(y)\right)\\
&+ \int_{\partial\mathfrak{D}(t_0)} \mathrm{d}^D\vec{\xi} \sqrt{|H(\vec{\xi})|} \left\{ G^+_{D+1}\left(x, x'(\vec{\xi})\right) u^{\alpha'}\nabla_{\alpha'}\psi\left(x'(\vec{\xi})\right) - u^{\alpha'}\nabla_{\alpha'} G^+_{D+1}\left(x, x'(\vec{\xi})\right)\psi\left(x'(\vec{\xi})\right) \right\}.
\end{aligned}$$

Here, we have assumed there are D coordinates $\vec{\xi}$ such that $x'^{\mu}(\vec{\xi})$ parametrizes our initial time hypersurface $\partial\mathfrak{D}(t_0)$. The $\sqrt{|H|}$ is the square root of the determinant of its induced metric. More specifically,

$$H_{ij}(\vec{\xi})\mathrm{d}\xi^i \mathrm{d}\xi^j = \left(g_{\mu\nu}(x(\vec{\xi})) \frac{\partial x^\mu}{\partial \xi^i} \frac{\partial x^\nu}{\partial \xi^j} \right) \mathrm{d}\xi^i \mathrm{d}\xi^j. \qquad (12.4.321)$$

We have defined the unit timelike vector u^μ, orthogonal to the surface $\partial\mathfrak{D}(t_0)$, to be pointing *towards* the future. Since the unit normal vector dotted into the gradient $\nabla_{\alpha'}$ occurring in Gauss' theorem (eq. (9.7.84)) points *outwards* – see equations (9.7.76) and (9.7.77) – it translates in our case at hand to pointing *backward* in time; this is our $-u^{\alpha'}$, namely,

$$\mathrm{d}^D\Sigma_{\alpha'} = \mathrm{d}^D\vec{\xi}\sqrt{|H(\vec{\xi})|}\left(-u_{\alpha'}(\vec{\xi}')\right). \qquad (12.4.322)$$

For the spatial boundaries $\partial\mathcal{D}_s^+$, the unit *outward*-pointing orthogonal vector is the $n^{\alpha'}$ in eq. (12.4.320); whereas the $\{y^{\mathrm{I}}\}$ are the spacetime coordinates parametrizing $\partial\mathcal{D}_s$ itself, so that the induced metric is

$$\mathcal{H}_{\mathrm{IJ}}(y)\mathrm{d}y^{\mathrm{I}}\mathrm{d}y^{\mathrm{J}} = \left(g_{\mu\nu}(x(y)) \frac{\partial x^\mu}{\partial y^{\mathrm{I}}} \frac{\partial x^\nu}{\partial y^{\mathrm{J}}} \right) \mathrm{d}y^{\mathrm{I}}\mathrm{d}y^{\mathrm{J}}. \qquad (12.4.323)$$

Finally, we note that the spatial boundary $\partial\mathfrak{D}_s^+$ terms in eq. (12.4.320) are zero if the spatial boundary is infinitely far away or if there are no spatial boundaries.

Physically speaking the initial conditions in eq. (12.4.320) are usually due to ψ engendered by the same source J but before $t < t_0$. In other words, if ψ were produced entirely by J, then the solution ψ ought to be obtained by setting t_0 to the infinite past, or the beginning of time in the spacetime at hand, while discarding the initial condition hypersurface integral:

$$\begin{aligned}
\psi(x) = &\int_{t' \leq t} \mathrm{d}^{D+1}x' \sqrt{|g(x')|} G^+_{D+1}(x,x') J(x')\\
&+ \int_{\substack{\partial\mathfrak{D}_s^+ \\ t' \leq t}} \mathrm{d}^D y \sqrt{|\mathcal{H}(y)|} n^{\alpha'} \nabla_{\alpha'} G^+_{D+1}\left(x, x'(y)\right)\psi\left(x'(y)\right). \qquad (12.4.324)
\end{aligned}$$

In electromagnetism, even the initial conditions at the beginning of the universe, say, need to obey Gauss' law constraints; and therefore cannot be zero. To sum: for a given observer location x, it is only when $J = 0$ on and within its entire past cone that we then regard the initial conditions at arbitrary $t_0 \leq t$ as freely specifiable.

Problem 12.51. Final Value Problem Let us now instead consider some 'cylindrical' domain of spacetime $\mathfrak{D}^-$ with spatial boundaries $\partial\mathfrak{D}_s^-$ and initial time $t_<$ and final time t_0 hypersurfaces $\partial\mathfrak{D}^-(t_<)$ and $\partial\mathfrak{D}(t_0)$ respectively. (This latter t_0 is the same one from above, but is now the final rather than the initial time.) We will consider an observer residing at x within this domain $\mathfrak{D}^-$. Show using the advanced Green's function G^-_{D+1} that

$$\begin{aligned}
&\psi(x; x^0 < t_0) \qquad (12.4.325)\\
&= \int_{\mathfrak{D}^-} \mathrm{d}^{D+1}x' \sqrt{|g(x')|} G^-_{D+1}(x,x') J(x') + \int_{\partial\mathfrak{D}_s^-} \mathrm{d}^D y \sqrt{|\mathcal{H}(y)|} n^{\alpha'} \nabla_{\alpha'} G^-_{D+1}(x, x'(y)) \psi(x'(y)) \\
&- \int_{\partial\mathfrak{D}(t_0)} \mathrm{d}^D\vec{\xi} \sqrt{|H(\vec{\xi})|} \left\{ G^-_{D+1}\left(x, x'(\vec{\xi})\right) u^{\alpha'} \nabla_{\alpha'} \psi\left(x'(\vec{\xi})\right) - u^{\alpha'} \nabla_{\alpha'} G^-_{D+1}\left(x, x'(\vec{\xi})\right) \psi\left(x'(\vec{\xi})\right) \right\}
\end{aligned}$$

where $u^{\alpha'}$ is the same future-pointing unit timelike vector from before. □

Intermediate Value Problem and Homogeneous Solutions We may now add the results from equations (12.4.320) and (12.4.325) but, for simplicity, focus on the source-free case (defined by $J = 0$) and further assume the spatial boundaries are either infinitely far away or are not present at all. These assumptions mean, only the integral on the 'intermediate-time' hypersurface $\partial\mathfrak{D}(t_0)$ remains. Then, for an observer lying between the initial-time hypersurface $\partial\mathfrak{D}^-(t_<)$ and the final-time hypersurface $\partial\mathfrak{D}^+(t_>)$, the homogeneous solution obeying $\Box\psi = 0$, for a given pair of intermediate time profile $\psi(t_0, \vec{x}(\vec{\xi}))$ and velocity $u \cdot \nabla\psi(t_0, \vec{x}(\vec{\xi}))$, is expressed by the constant$-t_0$ hypersurface integral

$$\begin{aligned}
&\psi(x; t_< < x^0 < t_>) \qquad (12.4.326)\\
&= \int_{\partial\mathfrak{D}(t_0)} \mathrm{d}^D\vec{\xi} \sqrt{|H(\vec{\xi})|} \left\{ \Delta_{D+1}\left(x, x'(\vec{\xi})\right) u^{\alpha'} \nabla_{\alpha'} \psi\left(x'(\vec{\xi})\right) - u^{\alpha'} \nabla_{\alpha'} \Delta_{D+1}\left(x, x'(\vec{\xi})\right) \psi\left(x'(\vec{\xi})\right) \right\},
\end{aligned}$$

where $\Delta \equiv G^+ - G^-$ is the difference between the retarded and advanced Green's functions.

Problem 12.52. Initial/Final Value Formulation of Δ For a given 'initial/final time' t_0 hypersurface $x(\vec{\xi})$, argue for the validity of the initial/final conditions

$$\Delta\left(x(\vec{\xi}), x'(\vec{\xi}')\right) = 0 \qquad (12.4.327)$$

$$\text{and} \qquad u^{\alpha'}(x') \nabla_{x'^\alpha} \Delta\left(x(\vec{\xi}), x'(\vec{\xi}')\right) = -u^\alpha(x) \nabla_{x^\alpha} \Delta\left(x(\vec{\xi}), x'(\vec{\xi}')\right) = -\frac{\delta^{(D)}\left(\vec{\xi} - \vec{\xi}'\right)}{\sqrt[4]{H(\vec{\xi}) H(\vec{\xi}')}}.$$

Hint: Start by explaining why $G^\pm(x,x') = \pm\Theta(t-t')\Delta(x,x')$. Then examine the equation $\Box G^\pm = \delta^{(d)}[x-x']/\sqrt{|g|}$ but in terms of Δ.

This result tells us Δ itself may be viewed as a field that vanishes on the t_0 surface but grows away from it, emanating away from the x' location. In QFT these would be analogous to the $[x,x] = 0$ and $[x,p] = i\hbar$ 'equal time' commutation relations. □

Flat Spacetime In Minkowski spacetime, we may choose t_0 to be the constant t surface of $\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}\vec{x}^2$. Then, expressed in these Cartesian coordinates, equations (12.4.320) and

(12.4.325) respectively reduces to

$$\psi(t > t_0, \vec{x}) = \int_{t' \geq t_0} \mathrm{d}t' \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' G^+_{D+1}\left(t - t', \vec{x} - \vec{x}'\right) J(t', \vec{x}') \tag{12.4.328}$$
$$+ \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \left\{ G^+_{D+1}(t - t_0, \vec{x} - \vec{x}') \partial_{t_0} \psi(t_0, \vec{x}') - \partial_{t_0} G^+_{D+1}(t - t_0, \vec{x} - \vec{x}') \cdot \psi(t_0, \vec{x}') \right\},$$

and

$$\psi(t < t_0, \vec{x}) = \int_{t' \leq t_0} \mathrm{d}t' \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' G^-_{D+1}\left(t - t', \vec{x} - \vec{x}'\right) J(t', \vec{x}') \tag{12.4.329}$$
$$- \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \left\{ G^-_{D+1}(t - t_0, \vec{x} - \vec{x}') \partial_{t_0} \psi(t_0, \vec{x}') - \partial_{t_0} G^-_{D+1}(t - t_0, \vec{x} - \vec{x}') \cdot \psi(t_0, \vec{x}') \right\}.$$

We see in both equations (12.4.320) and (12.4.328), that the time evolution of the field $\psi(x)$ can be solved once the retarded Green's function G^+_{D+1}, as well as ψ's initial profile and first time derivative is known at t_0. Generically, the field at the observer location x is the integral of the contribution from its initial profile and first time derivative on the $t = t_0$ surface from both on and within the past light cone of x. (Even in flat spacetime, while in 4 and higher even dimensional flat spacetime, the field propagates only on the light cone – in 2 and all odd dimensions, we have seen that scalar waves develop tails.)

Let us also observe that the wave solution in eq. (12.4.62) is in fact a special case of eq. (12.4.328): the initial time surface is the infinite past $t_0 \to -\infty$, upon which it is further assumed the initial field and its time derivatives are trivial – the signal detected at x can therefore be entirely attributed to J.

Problem 12.53. 4D Plane Waves In 4 dimensional infinite flat spacetime, let the initial conditions for the source-free ($J = 0$) massless scalar field be given by

$$\psi(t = 0, \vec{x}) = e^{i\vec{k}\cdot\vec{x}}, \qquad \partial_t \psi(t = 0, \vec{x}) = \pm i |\vec{k}| e^{i\vec{k}\cdot\vec{x}}. \tag{12.4.330}$$

Use the rightmost expression of eq. (12.4.192) in the Kirchhoff representation of eq. (12.4.328) to find $\psi(t > 0, \vec{x})$. You can probably guess the final answer, but this is a simple example to show you the Kirchhoff representation really works. □

Problem 12.54. Connection to Fourier Space Starting from the Minkowski version of the source-free ($J = 0$) eq. (12.4.326), namely

$$\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \mathrm{d}^D \vec{x}' \Big\{ \Delta_{D+1}(t - t_0, \vec{x} - \vec{x}') \partial_{t_0} \psi(t_0, \vec{x}')$$
$$- \partial_{t_0} \Delta_{D+1}(t - t_0, \vec{x} - \vec{x}') \cdot \psi(t_0, \vec{x}') \Big\}, \tag{12.4.331}$$

derive eq. (12.4.53). Hint: Employ the representation in eq. (12.4.70). □

Problem 12.55. Two Dimensions In 1+1 dimensional flat spacetime, suppose $\partial^2 \psi = (\partial_t^2 - \partial_x^2)\psi = 0$ and

$$\psi(t = 0, x) = Q(x), \qquad \partial_t \psi(t = 0, x) = P(x). \tag{12.4.332}$$

Explain why

$$\psi(t>0,x) = \frac{1}{2}Q(x+t) + \frac{1}{2}Q(x-t) + \frac{1}{2}\int_{x-t}^{x+t} P(x')\mathrm{d}x'. \tag{12.4.333}$$

Hint: Remember eq. (12.4.96). Note that, if $t > 0$, the $\delta(t-|z|)$ implies $z = t$ and $z = -t$. Also contrast this 'initial value problem' to the 'null cone' one in Problem (12.20). □

12.4.7 Helmholtz Wave Equation in Isotropic Non-conducting Media

The scalar analog of the electromagnetic wave equation in a frequency-dependent medium is given by the following Helmholtz wave equation in (3+1)-dimensions:

$$\left(\vec{\nabla}^2 + \Omega^2(\omega)\right)\widetilde{\psi}(\omega,\vec{x}) = -\widetilde{J}(\omega,\vec{x}), \tag{12.4.334}$$

$$\Omega(\omega) \equiv \omega\cdot\widetilde{n}(\omega), \tag{12.4.335}$$

where

$$\widetilde{\psi}(\omega,\vec{x}) \equiv \int_{\mathbb{R}} \widetilde{\psi}(\omega,\vec{x})e^{i\omega t}\mathrm{d}t. \tag{12.4.336}$$

We see that the corresponding retarded Green's function – recall equations (12.4.262) and (12.4.265) – is

$$\widetilde{G}^+(\Omega R) = \frac{\exp(i\Omega R)}{4\pi R}, \tag{12.4.337}$$

$$\left(\vec{\nabla}^2 + \Omega^2\right)\widetilde{G}^+(\Omega R) = -\delta^{(3)}(\vec{x}-\vec{x}'). \tag{12.4.338}$$

The solution to the wave equation in frequency space is

$$\widetilde{\psi}(\omega,\vec{x}) = \int_{\mathbb{R}^3} \mathrm{d}^3\vec{x}'\widetilde{G}^+(\Omega R)\,\widetilde{J}(\omega,\vec{x}'), \tag{12.4.339}$$

$$R \equiv |\vec{x}-\vec{x}'|. \tag{12.4.340}$$

Multiplying both sides by $e^{-i\omega t}$ and integrating over frequency, we then obtain the position spacetime wave solution as

$$\psi(t,\vec{x}) = \int_{\mathbb{R}^3}\mathrm{d}^3\vec{x}'\int_{\mathbb{R}}\frac{\mathrm{d}\omega}{2\pi}\frac{\exp(-i\omega(t-\widetilde{n}(\omega)R))}{4\pi R}\widetilde{J}(\omega,\vec{x}'). \tag{12.4.341}$$

[159]The physical interpretation of eq. (12.4.341) is: there are spherical waves emitted from every point $\vec{x}'$ within the source; and at each driving frequency ω, these waves propagate at speed $1/\widetilde{n}(\omega)$ since

$$\exp(-i\omega(t-\widetilde{n}(\omega)R)) = \exp\left(+i\Omega\left(R - \frac{t}{\widetilde{n}(\omega)}\right)\right). \tag{12.4.342}$$

[159]Note: The wave equation for ψ is *not* $\partial^2\psi = J$; its form will not concern us here.

In other words, $\widetilde{n}(\omega)$ may be viewed as the refractive index of the medium at frequency ω.

Analyticity Implies Causality Let us study the causal properties of the ψ signal sourced by $\widetilde{J}$. To this end, let us suppose $J(t,\vec{x})$ is not only spatially localized, but also of finite duration:

$$J(t,\vec{x}) \neq 0 \qquad t_1 \leq t \leq t_2 \tag{12.4.343}$$

$$= 0 \qquad \text{Otherwise.} \tag{12.4.344}$$

Because of its finite duration, it is reasonable to suppose

$$\int_{-\infty}^{+\infty} |J(t,\vec{x}')|\mathrm{d}t = \int_{t_1}^{t_2} |J(t,\vec{x}')|\mathrm{d}t < \infty. \tag{12.4.345}$$

By factoring out $e^{i\omega t_2}$ from its frequency $\omega-$transform

$$\widetilde{J}(\omega,\vec{x}) = \int_{-\infty}^{+\infty} J(t,\vec{x}')e^{+i\omega t}\mathrm{d}t \tag{12.4.346}$$

$$= e^{+i\omega t_2}\int_{t_1-t_2}^{0} J(t+t_2,\vec{x}')e^{+i\omega t}\mathrm{d}t \equiv e^{+i\omega t_2}\widetilde{J}'(\omega,\vec{x};t_1,t_2). \tag{12.4.347}$$

the remaining factor $\widehat{J}'$ is necessarily *analytic* in the $\omega \equiv \omega_{\mathrm{R}} + i\omega_{\mathrm{I}}$ for $\omega_{\mathrm{I}} < 0$ because

$$\left|\partial_\omega \int_{t_1-t_2}^{0} J(t+t_2,\vec{x}')e^{+i\omega t}\mathrm{d}t\right| \leq \int_{t_1-t_2}^{0} |t\cdot J(t+t_2,\vec{x}')|e^{-\omega_{\mathrm{I}} t}\mathrm{d}t < \infty \qquad \forall \omega_{\mathrm{I}} < 0. \tag{12.4.348}$$

Let us now assume $\widetilde{n}$ is not only analytic everywhere on the negative imaginary portion of the complex $\omega-$plane, but also tends to some constant

$$\widetilde{n}(|\omega| \to +\infty) = \widetilde{n}_\infty \tag{12.4.349}$$

for $\omega_{\mathrm{I}} < 0$. (Remember, it is not possible for a non-constant function to be entire and not blow up somewhere at infinity.) Since both $\widetilde{n}$ and $\widetilde{J}$ are analytic for $\omega_{\mathrm{I}} < 0$, we may deform the contour in eq. (12.4.339) from running along the real line to the infinite-radius semi-circle that tends to $|\omega| \to \infty$ in the lower half $\omega-$plane. Using polar coordinates $\omega = \rho e^{i\theta}$ for $\rho \geq 0$ and $\pi \leq \theta \leq 2\pi$,

$$\psi(t,\vec{x}) = \int_{\mathbb{R}^3} \frac{\mathrm{d}^3\vec{x}'}{4\pi R} \lim_{\rho\to\infty} \int_{\pi\leq\theta\leq 2\pi} \frac{i\rho e^{i\theta}\mathrm{d}\theta}{2\pi} \widetilde{J}'\left(\omega = \rho e^{i\theta}, \vec{x}'; t_1, t_2\right)$$
$$\times \exp\left[\rho(\sin(\theta) - i\cos(\theta))(t - t_2 - \widetilde{n}_\infty R)\right]. \tag{12.4.350}$$

Therefore, for a $\psi-$detector located at the event $(t,\vec{x})$ and time elapsed satisfying

$$\frac{t-t_2}{\widetilde{n}_\infty} > |\vec{x} - \vec{x}'|, \tag{12.4.351}$$

the ψ is strictly zero because the magnitude of the integrand goes to zero

$$\lim_{\rho\to\infty} \frac{1}{2\pi} |\widetilde{J}'\left(\omega = \rho e^{i\theta}, \vec{x}'; t_1, t_2\right)| \cdot \rho \exp\left[\rho\sin(\theta)(t - t_2 - \widetilde{n}_\infty R)\right] = 0. \tag{12.4.352}$$

Eq. (12.4.351) determines the wave front emanating from the source $\widetilde{J}(\omega,\vec{x}')$ at $\vec{x}'$ for arbitrary sources, and also allows us to identify $1/\widetilde{n}_\infty$ as the wavefront speed. To sum:

The reciprocal of the refractive index $1/\widetilde{n}(\omega)$ evaluated at frequencies with infinite magnitudes – assuming $\widetilde{n}(\omega)$ is analytic and $\widetilde{n}(|\omega| \to \infty, \omega_{\rm I} < 0) = \widetilde{n}_\infty$ is a constant – yields the wavefront speed of waves obeying the Helmholtz wave equation in eq. (12.4.334).

Problem 12.56. Since the source $J(t,\vec{x})$ was only switched on at $t = t_1$, it is intuitively clear there should be no ψ waves at all for $t < t_1$. On the other hand, can you show directly from eq. (12.4.341) that $\psi(t < t_1, \vec{x})$ is indeed zero for all $\vec{x}$? Hints: You may find it useful to define a different shifted $\widetilde{J}'$ through the relation

$$\widetilde{J}(\omega, \vec{x}) = e^{+i\omega t_1} \int_0^{t_2 - t_1} J(t + t_1, \vec{x}') e^{+i\omega t} \mathrm{d}t \equiv e^{+i\omega t_1} \widetilde{J}'(\omega, \vec{x}; t_1, t_2). \tag{12.4.353}$$

Then perform the appropriate contour deformation of eq. (12.4.341) that parallels eq. (12.4.350) – should you push the contour 'downwards' or 'upwards' on the $\omega-$plane? □

Problem 12.57. Vibrations of a Drum's Surface Frequencies. Normal modes. Eigenfunctions.

Diffraction from Kirchhoff Representation We now turn to the study of diffraction of waves with a fixed frequency through an aperture. A telescope with a circular aperture is a good example of such a physical problem. Away from obstacles, we shall assume the field $\widetilde{\psi}$ obeys the homogeneous equation

$$\left(\vec{\nabla}^2 + \Omega(\omega)^2\right) \widetilde{\psi} = 0, \tag{12.4.354}$$

$$\Omega \equiv \omega \cdot \widetilde{n}(\omega). \tag{12.4.355}$$

We shall first consider a (closed) domain $\mathfrak{D}$ within infinite (flat) space and exploit the Helmholtz Green's function obeying

$$\left(\vec{\nabla}^2 + \Omega(\omega)^2\right) \widetilde{G}(\omega; \vec{x}, \vec{x}') = -\delta^{(3)}(\vec{x} - \vec{x}'), \tag{12.4.356}$$

but *not* necessarily the retarded boundary condition obeyed by the $\widetilde{G}^+$ in eq. (12.4.337). We shall then use $\widetilde{G}$ to solve $\widetilde{\psi}(\vec{x} \in \mathfrak{D})$ using $\widetilde{\psi}(\vec{x} \in \partial\mathfrak{D})$; namely, the field within the domain from its value on the boundary.

The starting point is

$$\int_{\mathfrak{D}} \mathrm{d}^3\vec{x}' \left(\widetilde{G}(\vec{x}, \vec{x}') \cdot (\vec{\nabla}^2_{\vec{x}'} + \Omega^2) \widetilde{\psi}(\vec{x}') - \widetilde{\psi}(\vec{x}') \cdot (\vec{\nabla}^2_{\vec{x}'} + \Omega^2) \widetilde{G}(\vec{x}, \vec{x}') \right) = \widetilde{\psi}(\vec{x} \in \mathfrak{D}). \tag{12.4.357}$$

Let us notice the Ω^2 terms cancel out. Integration-by-parts then yields a surface term and a volume term, but the latter cancels out to zero.

$$\widetilde{\psi}(\vec{x} \in \mathfrak{D}) = \int_{\partial\mathfrak{D}} \mathrm{d}^2\vec{\Sigma}' \cdot \left(\widetilde{G}(\vec{x}, \vec{x}') \cdot \vec{\nabla}_{\vec{x}'} \widetilde{\psi}(\vec{x}') - \widetilde{\psi}(\vec{x}') \cdot \vec{\nabla}_{\vec{x}'} \widetilde{G}(\vec{x}, \vec{x}') \right). \tag{12.4.358}$$

We see that, $\widetilde{\psi}$ may indeed be related solely to its value on the boundary, and not to its normal derivative there, if we subject $\widetilde{G}$ to Dirichlet boundary conditions:

$$\widetilde{G}(\vec{x} \in \partial\mathfrak{D}, \vec{x}') = 0 = \widetilde{G}(\vec{x}, \vec{x}' \in \partial\mathfrak{D}). \tag{12.4.359}$$

This hands us the desired formula:

$$\widetilde{\psi}(\vec{x} \in \mathfrak{D}) = -\int_{\partial\mathfrak{D}} \mathrm{d}^2\vec{\Sigma}' \cdot \vec{\nabla}_{\vec{x}'}\widetilde{G}(\vec{x}, \vec{x}') \cdot \widetilde{\psi}(\vec{x}') \tag{12.4.360}$$

$$= -\int_{\partial\mathfrak{D}} \mathrm{d}^2\Sigma_{i'} \nabla^i_{\vec{x}'}\widetilde{G}(\vec{x}, \vec{x}') \cdot \widetilde{\psi}(\vec{x}'). \tag{12.4.361}$$

Let us now apply this to the setup with an infinite 2D opaque flat screen in 3D space, with an opening O with an arbitrary shape. We place some source of light on one side of the screen, and we wish to solve for the signal detected by some observer on the other side. For the setup at hand, $\mathfrak{D}$ is simply the observer side of the screen; and its boundary $\partial\mathfrak{D}$ is the infinite 2D plane of the screen itself plus the infinite hemisphere at infinity. Let $\vec{x}_\perp$ and $\vec{x}'_\perp$ describe respectively the observer's and unit-point-source's Cartesian coordinates on the plane parallel to the 2D screen, and z and z' be the ones orthogonal to it, with positive $z-$values parametrizing the observer side of the screen. Then, the Dirichlet Green's function is

$$\widetilde{G}(\vec{x}, \vec{x}') = \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z - z')|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z - z')|} - \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z + z')|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z + z')|}. \tag{12.4.362}$$

YZ: Normal D is zero G. We have defined $\widehat{n}$ to be the unit outward normal, pointing away from the observer side of the screen; whereas the screen itself is at $z = 0 = z'$. That this $\widetilde{G}$ obeys Dirichlet boundary conditions can be understood by first noting that the length

$$|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z \pm z')| = \sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + (z \pm z')^2} \tag{12.4.363}$$

itself is invariant under the replacement $\widehat{n} \to -\widehat{n}$. Thus, when either $z = 0$ or $z' = 0$, we have

$$\widetilde{G}(z = 0) = \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp + \widehat{n} \cdot z'|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp + \widehat{n} \cdot z'|} - \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z'|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z'|} = 0 \tag{12.4.364}$$

or

$$\widetilde{G}(z' = 0) = \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z|} - \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n} \cdot z|} = 0. \tag{12.4.365}$$

In addition, if we apply the Helmholtz wave operator with respect to $\vec{x} \equiv (\vec{x}_\perp, z)$ or $\vec{x}' \equiv (\vec{x}'_\perp, z')$, and utilize the retarded Green's function eq. (12.4.338),

$$-\left(\vec{\nabla}^2 + \Omega^2\right)\widetilde{G} = \delta^{(2)}(\vec{x} - \vec{x}')\left(\delta(z - z') - \delta(z + z')\right). \tag{12.4.366}$$

However, since we are focusing on the observer side of the screen – i.e., $\mathfrak{D}$ is the region where $z, z' \geq 0$ – the $z-$coordinates can never be negative of each other and therefore $\delta(z + z')$ may be dropped. This recovers $-(\vec{\nabla}^2 + \Omega^2)\widetilde{G} = \delta^{(3)}(\vec{x} - \vec{x}')$ within $\mathfrak{D}$. For the observer at $\vec{x} = (\vec{x}_\perp, z)$, we may view the left term on the right hand side of eq. (12.4.362) as the spherical wave engendered by the point-source at $(\vec{x}'_\perp, z')$ whereas the rightmost term is that by the 'image charge' at $(\vec{x}'_\perp, -z')$ on the source side of the screen. These two waves then act to cancel each other out on the screen at $z = 0$.

Because the

$$R_\pm \equiv |\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z \pm z')| \tag{12.4.367}$$

occurring within the Dirichlet Helmholtz Green's function (cf. eq. (12.4.362)) tend to infinity on the infinite hemisphere portion of the surface integral in eq. (12.4.360), thereby setting the integrand to zero,**YZ Need to verify this explicitly** the remaining contribution comes only from the aperture O, since we have also assumed the rest of the screen is opaque. Our final formula for this diffraction-through-aperture problem is therefore

$$\widetilde{\psi}(\vec{x} \in \mathfrak{D}) = -i\Omega \cdot z \cdot \int_O \frac{\mathrm{d}^2\vec{x}'_\perp}{2\pi}\widetilde{\psi}(\vec{x}'_\perp)\frac{e^{i\Omega R_0}}{R_0^2}\left(1 - \frac{1}{i\Omega R_0}\right), \tag{12.4.368}$$

$$R_0 \equiv \sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + z^2}. \tag{12.4.369}$$

Problem 12.58. Insert eq. (12.4.362) into eq. (12.4.360) and carry out the necessary analysis to verify eq. (12.4.368). □

Problem 12.59. Neumann Green's Function Instead of making $\widetilde{G}$ vanish on the $(1,2)-$plane, let us make its normal derivative vanish instead; i.e., require $\partial_z\widetilde{G}(z=0) = 0 = \partial_{z'}\widetilde{G}(z'=0)$. Show that the answer is given by

$$\widetilde{G}(\vec{x}, \vec{x}') = \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z - z')|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z - z')|} + \frac{\exp\left(i\Omega|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z + z')|\right)}{4\pi|\vec{x}_\perp - \vec{x}'_\perp - \widehat{n}(z + z')|}. \tag{12.4.370}$$

□

Apertures and Obstacles The primary issue with eq. (12.4.368) is that, while it provides a mathematically consistent formula, it does not provide us with a prescription for the correct boundary value of $\widetilde{\psi}$ that corresponds to the diffraction problem of a given setup.

Far Zone For concreteness, let us place the origin of the coordinate system parallel to the screen somewhere within the central region of the aperture O. This allows us to define the radii

$$r' \equiv |\vec{x}'_\perp| \qquad \text{and} \qquad r \equiv |\vec{x}_\perp|. \tag{12.4.371}$$

Furthermore – for simplicity – let us make a far zone (FZ) approximation, defined by the conditions

$$\Omega \cdot r \gg 1 \qquad \text{and} \qquad \frac{r'}{r} \ll \sqrt{\frac{2\pi}{\omega \cdot r}}. \tag{12.4.372}$$

The first condition asserts the signal is measured at distances much larger than the inverse frequency of the input plane wave; whereas the second condition is the assumption that the signal is measured at distances much larger than the size of the aperture O itself. These conditions render the $1/(\Omega \cdot R)$ term much smaller than the one to its immediate left; and, moreover, allow us to approximate $z/R \approx z/r \equiv \cos\theta$. We also expand

$$\sqrt{(\vec{x}_\perp - \vec{x}'_\perp)^2 + z^2} = e^{-\vec{x}'_\perp \cdot \vec{\nabla}_{\vec{x}}} |\vec{x}| \tag{12.4.373}$$

$$= r\left(1 - \frac{\vec{x}'_\perp}{r} \cdot \widehat{r} + \mathcal{O}\left((r'/r)^2\right)\right). \tag{12.4.374}$$

Then, eq. (12.4.368) takes the far zone form

$$\widetilde{\psi}(\vec{x} \in \text{FZ}) = -i\Omega \cdot \cos\theta \cdot \frac{e^{i\Omega r}}{2\pi r} \cdot \int_O \text{d}^2\vec{x}'_\perp \widetilde{\psi}(\vec{x}'_\perp) e^{-i(\Omega\widehat{r})\cdot\vec{x}'_\perp(1+\mathcal{O}(r'/r))} \tag{12.4.375}$$

$$\times \left(1 + \mathcal{O}\left((\Omega \cdot r)^{-1}, r'/r\right)\right). \tag{12.4.376}$$

The remaining exponential within the integrand goes as $\exp(-i\Omega\widehat{r} \cdot \vec{x}' + i\mathcal{O}[(\Omega r)(r'/r)^2] + \dots)$; this explains the inequality on the right in eq. (12.4.372), as the correction to the phase must be much smaller than 2π. **YZ: Do we recover $\widetilde{\psi}$ on O itself? Solve the complement problem; Babinet. Shadow. Cylindrical coordinates?**

Point Source Input We already know from the Green's function solution that

$$\widetilde{\psi}_0(\vec{x}') \equiv \frac{e^{i\Omega|\vec{x}'-\vec{x}_0|}}{4\pi|\vec{x}' - \vec{x}_0|} \tag{12.4.377}$$

is an outgoing spherical wave emanating from the unit 'point charge' at $\vec{x}_0$. Let us, therefore, suppose we have such a point source of waves on the source side ($z' < 0$) of the plane. If we assume that the source is far enough from the aperture O such that $\Omega|\vec{x}'_\perp - \vec{x}_0| \gg 1$ and $r_0 \equiv |\vec{x}_0| \gg |\vec{x}'_\perp|$ are satisfied, we may therefore take the 'input wave' arising from our point source at $\vec{x}_0$ is therefore

$$\widetilde{\psi}(\vec{x}'_\perp) \approx \frac{e^{i\Omega r_0}}{4\pi r_0} e^{-i\Omega\widehat{r}_0\cdot\vec{x}'_\perp}, \tag{12.4.378}$$

where $\widehat{r}_0 \equiv \vec{x}_0/|\vec{x}_0|$ is a fixed unit radial vector pointing from the origin within O to the point source. Defining A_O to be the area of the opening, eq. (12.4.375) now becomes

$$\widetilde{\psi}(\vec{x} \in \text{FZ}) \approx -i\Omega A_O \cos\theta \frac{e^{i\Omega(r_0+r)}}{8\pi^2 r_0 r} \mathcal{M}\left(\widehat{r}, \widehat{r}_0; \Omega\right) \tag{12.4.379}$$

where the screen-dependent modulation factor is

$$\mathcal{M}\left(\widehat{r}, \widehat{r}_0; \Omega\right) \equiv \int_O \frac{\text{d}^2\vec{x}'_\perp}{A_O} \exp\left(-i\Omega(\widehat{r}_0 + \widehat{r}) \cdot \vec{x}'_\perp\right). \tag{12.4.380}$$

YZ: Be careful about using eq. (12.4.337) for diffraction from a small aperture. It does not obey Dirichlet BC on the rest of the opaque screen; though, asymptotic analysis may show that $\mathcal{M}$ is small on the 'lens screen' regardless of the shape of aperture and regardless of the G used (Dirichlet vs retarded). To leading order in the far zone approximation, we see the signal is simply a spherical wave emanating from O, modulated by the $\cos\theta$ angular dependence: $\cos\theta \cdot e^{i\Omega r}/(4\pi r)$. We kept the first non-trivial dependence of the phase $\exp(i\Omega R)$ on $\vec{x}'_\perp$ in order to capture the impact on the far zone signal due to the details of O – its shape, size, etc. In particular, it yields a Fourier transform of O as a function of 'wave vectors' $\Omega\widehat{r}_0$ and $\Omega\widehat{r}$. For very low frequencies, where $\Omega r' \ll 1$, the exponential is approximately unity and the $\mathcal{M} \approx 1$. A high enough frequency, $\Omega r' \gtrsim 1$, is needed to obtain a non-trivial modulation factor $\mathcal{M}$.

Filter or Lens at O Up to this point we have assumed the medium is homogeneous and isotropic. We may consider placing a filter or lens at O to modify the trajectory of the incoming plane wave. For instance, a telescope to leading order is simply such a device, focusing the rays instead of allowing them to spread out as $\cos\theta$:

$$\widetilde{\psi}_{\text{telescope}}(\vec{x}) \approx -i\Omega A_O \cos\theta \frac{e^{i\Omega(r_0+r)}}{8\pi^2 r_0 r} f(\theta,\phi)\mathcal{M}\left(\widehat{r},\widehat{r}_0;\Omega\right); \tag{12.4.381}$$

where $f(\theta,\phi)$ is non-zero only for very small deflection angles $\theta \ll 1$.

Circular Aperture For a circular aperture of radius R_O, if $\widehat{n} = (0,0,-1)$, we may then parametrize

$$\vec{x}'_\perp = r'(\cos\phi', \sin\phi', 0), \tag{12.4.382}$$
$$\widehat{r}_0 = -\left(\sin\theta_0\cos\phi_0, \sin\theta_0\sin\phi_0, \cos\theta_0\right), \tag{12.4.383}$$
$$\widehat{r} = \left(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta\right), \tag{12.4.384}$$

– the θs are the angles from the $z-$axis and the ϕs are the azimuthal angules on the $\vec{x}_\perp-$plane – so that

$$\vec{x}'_\perp \cdot (\widehat{r}_0 + \widehat{r}) = r'(\sin\theta - \sin\theta_0)\cos(\phi - \phi'). \tag{12.4.385}$$

Thus the screen modulation factor itself is

$$\begin{aligned}
\mathcal{M}\left(\widehat{r},\widehat{r};\Omega\right) &= \int_0^{R_O} \frac{\mathrm{d}r' \cdot r'}{\pi R_O^2} \int_0^{2\pi} \mathrm{d}\phi \exp\left[-i\Omega r'(\sin\theta - \sin\theta_0)\cos\phi\right] \\
&= 2\pi \int_0^{R_O} \frac{\mathrm{d}r' \cdot r'}{\pi R_O^2} J_0\left(\Omega(\sin\theta - \sin\theta_0) r'\right) \\
&= \frac{2}{(\Omega(\sin\theta - \sin\theta_0)R_O)^2} \int_0^{\Omega(\sin\theta - \sin\theta_0)R_O} \mathrm{d}z \cdot z J_0(z) \\
&= 2\frac{J_1\left(\Omega(\sin\theta - \sin\theta_0)R_O\right)}{\Omega(\sin\theta - \sin\theta_0)R_O}.
\end{aligned} \tag{12.4.386}$$

We have employed the integral representation of $2\pi \cdot J_0(z) = \int_0^{2\pi} \exp(-iz\cos\phi)\mathrm{d}\phi$ in eq. (5.4.141); and the $m = n = 1$ case of eq. (5.4.200) – namely, $(z \cdot J_1(z))' = z \cdot J_0(z)$.

Problem 12.60. On-Axis Telescope Resolving Power Consider the light from a star placed at the center of the field-of-view of a telescope by first assuming $\widehat{r}_0$ is anti-parallel to the $z-$axis. Its light can be approximated as a superposition of spherical waves $\exp(-i\Omega(t - |\vec{x} - \vec{x}_0|))/(4\pi|\vec{x} - \vec{x}_0|)$, involving different frequencies $\{\Omega\}$, impinging upon the circular aperture of a telescope at $z = 0$. The ideal telescope would then focus the star light into as small a spot as allowed by the modulation factor $\mathcal{M}$, onto a distant screen (or, your retina). The Bessel $J_1(\dots)$ in eq. (12.4.386) then tells us – making a plot here will help – most of the signal is concentrated within the first peak. The smaller the angular width of this first peak, the *greater* the ability to resolve fine details, since they will be better separated.

For a fixed wavelength λ, we may identify

$$\Omega = \frac{2\pi}{\lambda}. \tag{12.4.387}$$

Given that the first zero of Bessel $J_1(z)$ lies at $z \approx 3.83171$, show that the width of the FZ signal's first peak also corresponds to an angular width of

$$\theta \approx 1.21967\frac{\lambda}{D_O}, \tag{12.4.388}$$

where $D_O \equiv 2R_O$ is the diameter of the telescope (aka aperture). This is known as the Rayleigh criterion.

f/ratio Consider placing a CMOS or CCD detector, whose pixels are squares with length ℓ, at the focus of a telescope. Suppose we wish to ensure that each pixel is able to sample $1/N$ of the telescope's angular resolution θ, where N is typically $\mathcal{O}$(few). Argue that, for a fixed ℓ, this N is proportional to the wavelength of light λ times the f/ratio of the telescope itself, where f/ratio $\equiv$ (focal Length)/(aperture). □

Problem 12.61. Central Obstruction Many telescope designs involve a central obstruction – usually a secondary mirror. Suppose the central obstruction is circular and completely opaque; and suppose it is centered at the center of the telescope's aperture and its diameter is $0 \leq \xi < 1$ that of the telescope's. Compute the analog to eq. (12.4.386) and, by plotting $\mathcal{M}$ for different ξs, discuss its effect on resolving power.

Hint: You should find that larger central obstructions yield lower resolutions because the signal gets distributed over a larger angular width. □

Problem 12.62. Gaussian Apodization Up till now, we have assumed the transparency of the telescope's aperture at O is uniform across its entire diameter. Let us now apply a Gaussian modulation, i.e., assume that, upon the spherical waves hitting the aperture, its amplitude is multiplied by a factor that falls off with radius from the center as a Gaussian. This modifies eq. (12.4.380) into the expression

$$\mathcal{M}\left(\Omega\widehat{r}\right) \equiv F_0 \int_O \frac{\mathrm{d}^2\vec{x}'_\perp}{A_O} \exp\left[-\frac{|\vec{x}'_\perp|^2}{2\sigma^2}\right] \exp\left[-i\Omega\vec{x}'_\perp \cdot \widehat{r}(\theta,\phi)\right], \tag{12.4.389}$$

where F_0 is a dimensionless constant and we have, for simplicity, assumed $\widehat{r}_0 = \widehat{n}$. Show that this $\mathcal{M}\left(\Omega\widehat{r}\right)$ is approximately

$$\mathcal{M}\left(\Omega\widehat{r}\right) \approx 2F_0 \left(\frac{\sigma}{R_O}\right)^2 \exp\left[-\frac{1}{2}\Omega^2\sigma^2\sin^2(\theta)\right], \tag{12.4.390}$$

by assuming $R_O/\sigma \gg 1$.

Notice the diffraction patterns due to the Bessel function J_1 in eq. (12.4.386) are now gone due to this Gaussian 'filter'. By adjusting σ appropriately, one may attempt to increase the resolving power of the telescope – focus the star's light into a smaller spot – albeit at the cost of losing signal strength. □

Problem 12.63. Young's Double Slit Model Young's double slit experiment as follows. For $0 < \varepsilon \ll a \ll L$, the two slits – each of width ε and height L – are parallel to the $y-$axis and separated by distance a along the $x-$axis.

$$\vec{x}'_\perp(\text{slit } 1) = \left(-\frac{\varepsilon}{2} \leq x'^1 - \frac{a}{2} \leq \frac{\varepsilon}{2}, -\frac{L}{2} \leq x'^2 \leq \frac{L}{2}\right), \tag{12.4.391}$$

$$\vec{x}'_\perp(\text{slit } 2) = \left(-\frac{\varepsilon}{2} \leq x'^1 + \frac{a}{2} \leq \frac{\varepsilon}{2}, -\frac{L}{2} \leq x'^2 \leq \frac{L}{2}\right). \tag{12.4.392}$$

Compute the approximate interference pattern near the $z-$axis but in the far zone. □

Problem 12.64. Retarded Green's Function (12.4.337)

12.5 Linear Wave Equations in Curved Spacetimes

12.5.1 JWKB (Short Wavelength) Approximation and Null Geodesics

YZ: For now, copied+pasted from Physics in Curved Spacetime notes. Plan: modify it to discuss scalars. In this section we will apply the JWKB (more commonly dubbed WKB) approximation to study the vacuum (i.e., $J_\mu = 0$ limit of) Maxwell's equations in eq. (**??**). At leading orders in perturbation theory, we will argue – in the limit where the wavelength of the photons are much shorter than that of the background geometric curvature – that photons propagate on the light cone and their polarization tensors are largely parallel transported along their null geodesics. We will also see that the photon's phase S would allow us to define its frequency as the number density of constant-S surfaces piercing the timelike worldline of the observer. This also leads us to recognize that, not only is $k^\mu \equiv \nabla^\mu S$ null it obeys the geodesic equation $k^\sigma \nabla_\sigma k^\mu = 0$.

Eikonal/Geometric Optics/JWKB Ansatz We will begin by postulating that the vector potential can be modeled as the (real part of) a slowly varying amplitude a_μ multiplied by a rapidly oscillating phase $\exp(iS)$:

$$A_\mu = \text{Re}\left\{a_\mu \exp(iS/\epsilon)\right\}. \tag{12.5.1}$$

[160]The $\{a_\mu\}$ can be complex but S is real. We shall also allow the amplitude itself to be a power series in ϵ:

$$a_\mu = \sum_{\ell=0}^{\infty} \epsilon^\ell \, {}_\ell a_\mu. \tag{12.5.2}$$

The $0 < \epsilon \ll 1$ is a fictitious parameter that reminds us of the hierarchy of length scales in the problem – specifically, ϵ should be viewed as the ratio between the short wavelength of the

[160]Recall that this ansatz becomes an exact solution in Minkowski spacetime, where $S = \pm k_\mu x^\mu$ and both k_μ and a_μ are constant.

photon to the long wavelength of the background geometric curvature. To this end, we shall re-write the vacuum version of the Lorenz-gauge Maxwell's equation (**??**) with ϵ^2 multiplying the wave operator $\Box$:

$$\Box A_\mu - \epsilon^2 R_\mu{}^\sigma A_\sigma = 0. \tag{12.5.3}$$

In a locally freely-falling frame (i.e., flat coordinate system), this equation takes the schematic form

$$\partial^2 A - \epsilon^2 (\partial^2 g) A = 0. \tag{12.5.4}$$

The first term from the left goes as $A/(\text{wavelength of } A)^2$ while the second as $A/(\text{wavelength of } g)^2$, and as thus already advertised ϵ^2 is a power counting parameter reminding us of the relative strength of the two terms.

Wave Equation Plugging the ansatz of eq. (12.5.1) into eq. (12.5.3):

$$\begin{aligned} 0 &= (\Box a_\mu - \epsilon^2 R_\mu{}^\sigma a_\sigma) e^{iS/\epsilon} + 2\nabla_\sigma a_\mu \frac{i}{\epsilon} \nabla^\sigma S \cdot e^{iS/\epsilon} + a_\mu \nabla_\sigma \left(i(\nabla^\sigma S/\epsilon) e^{iS/\epsilon} \right) \\ &= (\Box a_\mu - \epsilon^2 R_\mu{}^\sigma a_\sigma) e^{iS/\epsilon} + 2\nabla_\sigma a_\mu \frac{i}{\epsilon} (\nabla^\sigma S) \cdot e^{iS/\epsilon} + a_\mu \left(i(\Box S/\epsilon) e^{iS/\epsilon} + (i\nabla S/\epsilon)^2 e^{iS/\epsilon} \right). \end{aligned} \tag{12.5.5}$$

Employing the power series of eq. (12.5.2),

$$\begin{aligned} 0 &= \Box\, {}_0a_\mu + \epsilon \Box\, {}_1a_\mu + \epsilon^2 \Box\, {}_2a_\mu + \dots \\ &\quad - R_\mu{}^\sigma \left(\epsilon^2\, {}_0a_\sigma + \epsilon^3\, {}_1a_\sigma + \dots \right) \\ &\quad + 2i\epsilon^{-1} (\nabla\, {}_0a_\mu \cdot \nabla S) + 2i\epsilon^0 (\nabla\, {}_1a_\mu \cdot \nabla S) + 2i\epsilon (\nabla\, {}_2a_\mu \cdot \nabla S) + \dots \\ &\quad + i\epsilon^{-1}\, {}_0a_\mu \Box S + i\epsilon^0\, {}_1a_\mu \Box S + i\epsilon\, {}_2a_\mu \Box S + \dots \\ &\quad - (\nabla S)^2 \left(\epsilon^{-2}\, {}_0a_\mu + \epsilon^{-1}\, {}_1a_\mu + \epsilon^0\, {}_2a_\mu + \epsilon\, {}_3a_\mu + \dots \right). \end{aligned} \tag{12.5.6}$$

Negative Two Setting the coefficient of ϵ^{-2} to zero

$$k_\mu k^\mu = 0, \qquad k_\mu \equiv \nabla_\mu S. \tag{12.5.7}$$

Because S is a scalar, $\nabla_\nu k_\mu = \nabla_\nu \nabla_\mu S = \nabla_\mu \nabla_\nu S = \nabla_\mu k_\nu$ and hence

$$0 = \nabla_\nu (k^2) = 2k^\mu \nabla_\nu k_\mu = 2k^\mu \nabla_\mu k_\nu. \tag{12.5.8}$$

That is, the gradient of the phase S sweeps out null geodesics in spacetime:

$$(k \cdot \nabla) k^\mu = 0. \tag{12.5.9}$$

Negative One Setting the coefficient of ϵ^{-1} to zero:

$$0 = {}_0a_\mu \Box S + 2\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu, \tag{12.5.10}$$

$$0 = \overline{{}_0a_\mu} \Box S + 2\nabla^\sigma S \nabla_\sigma \overline{{}_0a_\mu}, \tag{12.5.11}$$

where the second line is simply the complex conjugate of the first. Note that $\nabla|a|^2 = (\nabla a)\bar{a} + a(\nabla \bar{a})$. Guided by this, we may multiply the first equation by $\overline{{}_0a^\mu}$ and the second equation by ${}_0a^\mu$, followed by adding them.

$$0 = |\,{}_0a|^2 \Box S + 2\overline{{}_0a^\mu} \nabla^\sigma S \nabla_\sigma\, {}_0a_\mu \tag{12.5.12}$$

$$0 = |{}_0a|^2 \Box S + 2\,{}_0a^\mu \nabla^\sigma S \nabla_\sigma \overline{{}_0a_\mu} \tag{12.5.13}$$

$$0 = 2|{}_0a|^2 \Box S + 2\nabla^\sigma S \nabla_\sigma |{}_0a|^2, \qquad |{}_0a|^2 \equiv {}_0a_\mu \overline{{}_0a^\mu}. \tag{12.5.14}$$

The right hand side of the final equation can be expressed as a divergence.

$$0 = \nabla_\sigma \left(|{}_0a|^2 \nabla^\sigma S\right) = \nabla_\sigma \left(|{}_0a|^2 k^\sigma\right) \tag{12.5.15}$$

Up to an overall normalization constant, we may interpret $n^\sigma \equiv |{}_0a|^2 k^\sigma$ as a photon number current, and this equation as its conservation law.

We turn to examining the derivative along $k \equiv \nabla S$ the normalized leading order photon amplitude ${}_0a_\mu/\sqrt{|{}_0a|^2}$:

$$\nabla^\sigma S \nabla_\sigma \left(\frac{{}_0a_\mu}{\sqrt{|{}_0a|^2}}\right) = \frac{\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu}{\sqrt{|{}_0a|^2}} - \frac{{}_0a_\mu}{2(|{}_0a|^2)^{3/2}} \nabla^\sigma S \nabla_\sigma |{}_0a|^2. \tag{12.5.16}$$

Eq. (12.5.15) says $\nabla S \cdot \nabla |{}_0a|^2 = -|{}_0a|^2 \Box S$, while eq. (12.5.10), in turn, states ${}_0a_\mu \Box S = -2\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu$.

$$\begin{aligned}
\nabla^\sigma S \nabla_\sigma \left(\frac{{}_0a_\mu}{\sqrt{|{}_0a|^2}}\right) &= \frac{\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu}{\sqrt{|{}_0a|^2}} + \frac{|{}_0a|^2}{2(|{}_0a|^2)^{3/2}}\, {}_0a_\mu \Box S \\
&= \frac{\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu}{\sqrt{|{}_0a|^2}} - \frac{\nabla^\sigma S \nabla_\sigma\, {}_0a_\mu}{\sqrt{|{}_0a|^2}} = 0.
\end{aligned} \tag{12.5.17}$$

To summarize, we have worked out the first two orders of the Lorenz gauge vacuum Maxwell's equations in the JWKB/eikonal/geometric optics limit. Up to this level of accuracy, perturbation theory teaches us:

- The gradient of the phase of the photon field $k^\mu \equiv \nabla^\mu S$ – which we may interpret as its dominant direction of propagation – follows null geodesics in the curved spacetime.
- The photon number current is covariantly conserved.
- The normalized polarization vector is parallel transported along k^μ.
- This same wave vector is orthogonal to the polarization of the photon at leading order; and the first deviation to non-orthogonality occurring at the next order is proportional to the divergence of the polarization vector itself.

12.6 Variational Principle in Field Theory

You may be familiar with the variational principle – or, the principle of stationary action – from classical mechanics. Here, we will write down one for the classical field theories leading to the Poisson and wave equations.

Poisson equation Consider the following *action* for the real field ψ sourced by some externally prescribed $J(\vec{x})$.

$$S_{\text{Poisson}}[\psi] \equiv \int_{\mathfrak{D}} \mathrm{d}^D\vec{x} \sqrt{|g(\vec{x})|} \left(\frac{1}{2}\nabla_i \psi(\vec{x}) \nabla^i \psi(\vec{x}) - \psi(\vec{x}) J(\vec{x})\right) \tag{12.6.1}$$

We claim that the action S_{Poisson} is extremized iff ψ is a solution to Poisson's equation (eq. (12.1.1)), provided the field at the boundary $\partial\mathfrak{D}$ of the domain is specified and fixed.

Given a some field $\bar{\psi}$, not necessarily a solution, let us consider some deviation from it; namely,

$$\psi(\vec{x}) = \bar{\psi}(\vec{x}) + \delta\psi(\vec{x}). \tag{12.6.2}$$

($\delta\psi$ is one field; the δ is pre-pended as a reminder this is a deviation from $\bar{\psi}$.) A direct calculation yields

$$\begin{aligned} S_{\text{Poisson}}[\bar{\psi} + \delta\psi] &= \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\left(\frac{1}{2}\nabla_i\bar{\psi}\nabla^i\bar{\psi} - \bar{\psi}J\right) \\ &+ \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\left(\nabla_i\bar{\psi}\nabla^i\delta\psi - J\delta\psi\right) \\ &+ \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\left(\frac{1}{2}\nabla_i\delta\psi\nabla^i\delta\psi\right). \end{aligned} \tag{12.6.3}$$

We may integrate-by-parts, in the second line, the gradient acting on $\delta\psi$.

$$\begin{aligned} S_{\text{Poisson}}[\bar{\psi} + \delta\psi] = \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\left(\frac{1}{2}\nabla_i\bar{\psi}\nabla^i\bar{\psi} - \bar{\psi}J + \frac{1}{2}\nabla_i\delta\psi\nabla^i\delta\psi + \delta\psi\left\{-\vec{\nabla}^2\bar{\psi} - J\right\}\right) \\ + \int_{\partial\mathfrak{D}} \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}\delta\psi n^i\nabla_i\bar{\psi} \end{aligned} \tag{12.6.4}$$

Provided Dirichlet boundary conditions are specified and not varied, i.e., $\psi(\partial\mathfrak{D})$ is given, then by definition $\delta\psi(\partial\mathfrak{D}) = 0$ and the surface term on the second line is zero. Now, suppose Poisson's equation is satisfied by $\bar{\psi}$, then $-\vec{\nabla}^2\bar{\psi} - J = 0$ and because the remaining quadratic-in-$\delta\psi$ is strictly positive (as argued earlier) we see that any deviation *increases* the value of S_{Poisson} and therefore the solution $\bar{\psi}$ yields a minimal action.

Conversely, just as we say a (real) function $f(x)$ is extremized at $x = x_0$ when $f'(x_0) = 0$, we would say S_{Poisson} is extremized by $\bar{\psi}$ if the first-order-in-$\delta\psi$ term

$$\int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g(\vec{x})|}\delta\psi\left\{-\vec{\nabla}^2\bar{\psi} - J\right\} \tag{12.6.5}$$

vanishes for *any* deviation $\delta\psi$. But if this were to vanish for any deviation $\delta\psi(\vec{x})$, the terms in the curly brackets must be zero, and Poisson's equation is satisfied.

Wave equation in infinite space Assuming the fields fall off sufficiently quickly at spatial infinity and suppose the initial $\psi(t_{\mathrm{i}}, \vec{x})$ and final $\psi(t_{\mathrm{f}}, \vec{x})$ configurations are specified and fixed, we now discuss why the action

$$S_{\text{Wave}} \equiv \int_{t_{\mathrm{i}}}^{t_{\mathrm{f}}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left\{\frac{1}{2}\nabla_\mu\psi(t'', \vec{x})\nabla^\mu\psi(t'', \vec{x}) + J(t'', \vec{x})\psi(t'', \vec{x})\right\} \tag{12.6.6}$$

(where $x \equiv (t'', \vec{x})$) is extremized iff the wave equation in eq. (12.4.15) is satisfied.

Just as we did for S_{Poisson}, let us consider adding to some given field $\bar{\psi}$, a deviation $\delta\psi$. That is, we will consider

$$\psi(x) = \bar{\psi}(x) + \delta\psi(x), \tag{12.6.7}$$

without first assuming $\bar{\psi}$ solves the wave equation. A direct calculation yields

$$\begin{aligned}
S_{\text{Wave}}[\bar{\psi}+\delta\psi] = &\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\frac{1}{2}\nabla_\mu\bar{\psi}\nabla^\mu\bar{\psi}+\bar{\psi}J\right) \\
&+\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\nabla_\mu\bar{\psi}\nabla^\mu\delta\psi+J\delta\psi\right) \\
&+\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\frac{1}{2}\nabla_\mu\delta\psi\nabla^\mu\delta\psi\right).
\end{aligned} \tag{12.6.8}$$

We may integrate-by-parts, in the second line, the gradient acting on $\delta\psi$. By assuming that the fields fall off sufficiency quickly at spatial infinity, the remaining surface terms involve the fields at the initial and final time hypersurfaces.

$$\begin{aligned}
S_{\text{Wave}}[\bar{\psi}+\delta\psi] &= \int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\frac{1}{2}\nabla_\mu\bar{\psi}\nabla^\mu\bar{\psi}+\bar{\psi}J+\frac{1}{2}\nabla_\mu\delta\psi\nabla^\mu\delta\psi+\delta\psi\left\{-\nabla_\mu\nabla^\mu\bar{\psi}+J\right\}\right) \\
&\quad+\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\delta\psi(t=t_\text{f},\vec{x})g^{0\mu}\partial_\mu\bar{\psi}(t=t_\text{f},\vec{x}) \\
&\quad-\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\delta\psi(t=t_\text{i},\vec{x})g^{0\mu}\partial_\mu\bar{\psi}(t=t_\text{i},\vec{x}) \\
&\quad+\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{S}^{D-1}} \mathrm{d}^{D-1}\vec{\xi}\sqrt{|H(\vec{\xi})|}\delta\psi n^\mu\nabla_\mu\bar{\psi} \\
&= \int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\frac{1}{2}\nabla_\mu\bar{\psi}\nabla^\mu\bar{\psi}+\bar{\psi}J+\frac{1}{2}\nabla_\mu\delta\psi\nabla^\mu\delta\psi+\delta\psi\left\{-\nabla_\mu\nabla^\mu\bar{\psi}+J\right\}\right) \\
&\quad+\left[\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\delta\psi(t,\vec{x})g^{0\mu}\partial_\mu\bar{\psi}(t,\vec{x})\right]_{t=t_\text{i}}^{t=\text{f}}.
\end{aligned} \tag{12.6.9}$$

The second and third lines of the first equality (and the second line of the second equality) come from the time derivative part of

$$\begin{aligned}
\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{g(x)}\nabla_\mu\left(\delta\psi\nabla^\mu\bar{\psi}\right) &= \int_{t'}^{t} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\partial_\mu\left(\sqrt{g(x)}\delta\psi g^{\mu\nu}\nabla_\nu\bar{\psi}\right) \\
&= \left[\int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{g(x)}\delta\psi g^{0\nu}\partial_\nu\bar{\psi}\right]_{t''=t_\text{i}}^{t''=t_\text{f}}+\dots
\end{aligned} \tag{12.6.10}$$

Provided the initial and final field values are specified and not varied, then $\delta\psi(t''=t_{\text{i,f}})=0$ and the surface terms are zero. In eq. (12.6.9), we see that the action is extremized, i.e., when the term

$$\int_{t_\text{i}}^{t_\text{f}} \mathrm{d}t'' \int_{\mathbb{R}^D} \mathrm{d}^D\vec{x}\sqrt{|g(x)|}\left(\delta\psi\left\{-\nabla_\mu\nabla^\mu\bar{\psi}+J\right\}\right) \tag{12.6.11}$$

is zero for all deviations $\delta\psi$, iff the terms in the curly brackets vanish, and the wave equation eq. (12.4.15) is satisfied. Note that, unlike the case for S_{Poisson}, because $\nabla_\mu\psi\nabla^\mu\psi$ may not be positive definite, it is not possible to conclude from this analysis whether all solutions minimize, maximize, or merely extremizes the action S_{Wave}.

Why? Why bother coming up with an action to describe dynamics, especially if we already have the PDEs governing the fields themselves? Apart from the intellectual interest/curiosity in formulating the same physics in different ways, having an action to describe dynamics usually allows the symmetries of the system to be made more transparent. For instance, all of the currently known fundamental forces and fields in Nature – the Standard Model (SM) of particle physics and gravitation – can be phrased as an action principle, and the mathematical symmetries they exhibit played key roles in humanity's attempts to understand them. Furthermore, having an action for a given theory allows it to be quantized readily, through the path integral formulation of quantum field theory due to Richard P. Feynman. In fact, our discussion of the heat kernel in, for e.g. eq. (12.3.17), is in fact an example of Norbert Wiener's version of the path integral, which was the precursor of Feynman's.

Problem 12.65. Euler-Lagrange Equations Let us consider a more general action built out of some field $\psi(x)$ and its first derivatives $\nabla_\mu\psi(x)$, for $x^\mu \equiv (t'', \vec{x})$.

$$S[\psi] \equiv \int_{t_i}^{t_f} \mathrm{d}t'' \int_{\mathfrak{D}} \mathrm{d}^D\vec{x}\sqrt{|g|}\mathcal{L}\left(\psi, \nabla\psi\right) \tag{12.6.12}$$

Show that, demanding the action to be extremized leads to the Euler-Lagrange equations

$$\frac{\partial\mathcal{L}}{\partial\psi} = \nabla_\mu \frac{\partial\mathcal{L}}{\partial\nabla_\mu\psi}. \tag{12.6.13}$$

What sort of boundary conditions are sufficient to make the variational principle well-defined? What happens when $\mathfrak{D}$ no longer has an infinite spatial extent (as we have assumed in the preceding above)? Additionally, make sure you check that the Poisson and wave equations are recovered by applying the appropriate Euler-Lagrange equations. □

12.7 Appendix to linear PDEs discourse: Symmetric Green's Function of a real 2nd Order ODE

Setup In this section we wish to write down the symmetric Green's function of the most general 2nd order real linear ordinary differential operator D, in terms of its homogeneous solutions. We define such as differential operator as

$$D_z f(z) \equiv p_2(z)\frac{\mathrm{d}^2 f(z)}{\mathrm{d}z^2} + p_1(z)\frac{\mathrm{d}f(z)}{\mathrm{d}z} + p_0(z)f(z), \qquad a \leq z \leq b, \tag{12.7.1}$$

where $p_{0,1,2}$ are assumed to be smooth real functions and we are assuming the setup at hand is defined within the domain $z \in [a, b]$. By homogeneous solutions $f_{1,2}(z)$, we mean they both obey

$$D_z f_{1,2}(z) = 0. \tag{12.7.2}$$

Because this is a 2nd order ODE, we expect two linearly independent solutions $f_{1,2}(z)$. What we wish to solve here is the symmetric Green's function $G(z, z') = G(z', z)$ equation

$$D_z G(z, z') = \lambda(z)\delta(z - z'), \qquad \text{and} \qquad D_{z'} G(z, z') = \lambda(z')\delta(z - z'), \tag{12.7.3}$$

where $\delta(z-z')$ is the Dirac δ-function and λ is a function to be determined. With the Green's function $G(z,z')$ at hand we may proceed to solve the particular solution $f_p(z)$ to the inhomogeneous equation, with some prescribed external source J,

$$D_z f_p(z) = J(z) \qquad \Rightarrow \qquad f_p(z) = \int_a^b \frac{\mathrm{d}z'}{\lambda(z')} G(z,z') J(z'). \tag{12.7.4}$$

Of course, for a given problem, one needs to further impose appropriate boundary conditions to obtain a unique solution. Here, we will simply ask: what's the most general ansatz that would solve eq. (12.7.3) in terms of $f_{1,2}$?

Discontinuous first derivative at $z=z'$ The key observation to solving the symmetric Green's function is that, as long as $z \neq z'$ then the $\delta(z-z')=0$ in eq. (12.7.3). Therefore $G(z,z')$ has to obey the homogeneous equation

$$D_z G(z,z') = D_{z'} G(z,z') = 0, \qquad z \neq z'. \tag{12.7.5}$$

For $z > z'$, if we solve $D_z G = 0$ first,

$$G(z,z') = \alpha^{\mathrm{I}}(z') f_{\mathrm{I}}(z), \tag{12.7.6}$$

i.e., it must be a superposition of the linearly independent solutions $\{f_{\mathrm{I}}(z)\}$ (in the variable z). Because $G(z,z')$ is a function of both z and z', the coefficients of the superposition must depend on z'. If we then solve

$$D_{z'} G(z,z') = D_{z'} \alpha^{\mathrm{I}}(z') f_{\mathrm{I}}(z) = 0, \tag{12.7.7}$$

(for $z \neq z'$), we see that the $\{\alpha^{\mathrm{I}}(z')\}$ must in turn each be a superposition of the linearly independent solutions in the variable z'.

$$\alpha^{\mathrm{I}}(z') = A_>^{\mathrm{IJ}} f_{\mathrm{J}}(z'). \tag{12.7.8}$$

(The $\{A_>^{\mathrm{IJ}}\}$ are now constants, because $\alpha^{\mathrm{I}}(z')$ has to depend only on z' and not on z.) What we have deduced is that $G(z>z')$ is a sum of 4 independent terms:

$$G(z>z') = A_>^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z'), \qquad A_>^{\mathrm{IJ}} = \text{constant}. \tag{12.7.9}$$

Similar arguments will tell us,

$$G(z<z') = A_<^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z'), \qquad A_<^{\mathrm{IJ}} = \text{constant}. \tag{12.7.10}$$

This may be summarized as

$$G(z,z') = \Theta(z-z') A_>^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z') + \Theta(z'-z) A_<^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z'). \tag{12.7.11}$$

Now we examine the behavior of $G(z,z')$ near $z=z'$. Suppose $G(z,z')$ is discontinuous at $z=z'$. Then its first derivative there will contain $\delta(z-z')$ and its second derivative will contain $\delta'(z-z')$, and G itself will thus not satisfy the right hand side of eq. (12.7.3). Therefore we may impose the continuity conditions

$$A_<^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z) = A_>^{\mathrm{IJ}} f_{\mathrm{I}}(z) f_{\mathrm{J}}(z), \tag{12.7.12}$$

$$A^{11}_< f_1(z)^2 + A^{22}_< f_2(z)^2 + (A^{12}_< + A^{21}_<) f_1(z) f_2(z) = A^{11}_> f_1(z)^2 + A^{22}_> f_2(z)^2 + (A^{12}_> + A^{21}_>) f_1(z) f_2(z).$$

Since this must hold for all $a \leq z \leq b$, the coefficients of $f_1(z)^2$, $f_2(z)^2$ and $f_1(z)f_2(z)$ on both sides must be equal,

$$A^{11}_< = A^{11}_> \equiv A^1, \qquad A^{22}_< = A^{22}_> \equiv A^2, \qquad A^{12}_< + A^{21}_< = A^{12}_> + A^{21}_>. \tag{12.7.13}$$

Now let us integrate $D_z G(z,z') = \lambda(z)\delta(z-z')$ around the neighborhood of $z \approx z'$; i.e., for $0 < \epsilon \ll 1$, and a prime denoting ∂_z,

$$\begin{aligned}
\int_{z'-\epsilon}^{z'+\epsilon} \mathrm{d}z \lambda(z)\delta(z-z') &= \int_{z'-\epsilon}^{z'+\epsilon} \mathrm{d}z \left\{ p_2 G'' + p_1 G' + p_0 G \right\} \\
\lambda(z') &= \left[p_2 G' + p_1 G \right]_{z'-\epsilon}^{z'+\epsilon} + \int_{z'-\epsilon}^{z'+\epsilon} \mathrm{d}z \left\{ -p_2' G' - p_1' G + p_0 G \right\} \\
&= \left[(p_1(z) - \partial_z p_2(z)) G(z,z') + p_2(z) \partial_z G(z,z') \right]_{z=z'-\epsilon}^{z=z'+\epsilon} \\
&\quad + \int_{z'-\epsilon}^{z'+\epsilon} \mathrm{d}z \left\{ p_2''(z) G(z,z') - p_1'(z) G(z,z') + p_0(z) G(z,z') \right\}.
\end{aligned} \tag{12.7.14}$$

Because $p_{0,1,2}(z)$ are smooth and because G is continuous at $z = z'$, as we set $\epsilon \to 0$, only the G' remains on the right hand side.

$$\lim_{\epsilon \to 0} \left\{ p_2(z'+\epsilon) \frac{\partial G(z = z'+\epsilon, z')}{\partial z} - p_2(z'-\epsilon) \frac{\partial G(z = z'-\epsilon, z')}{\partial z} \right\} = \lambda(z') \tag{12.7.15}$$

We can set $z' \pm \epsilon \to z'$ in the p_2 because it is smooth; the error incurred would go as $\mathcal{O}(\epsilon)$. We have thus arrived at the following "jump" condition: the first derivative of the Green's function on either side of $z = z'$ *has to be* discontinuous and their difference multiplied by $p_2(z')$ is equal to the function $\lambda(z')$, the measure multiplying the $\delta(z-z')$ in eq. (12.7.3).

$$p_2(z') \left\{ \frac{\partial G(z = z'^+, z')}{\partial z} - \frac{\partial G(z = z'^-, z')}{\partial z} \right\} = \lambda(z') \tag{12.7.16}$$

This translates to

$$p_2(z') \left(A^{\mathrm{IJ}}_> f_{\mathrm{I}}'(z') f_{\mathrm{J}}(z') - A^{\mathrm{IJ}}_< f_{\mathrm{I}}'(z') f_{\mathrm{J}}(z') \right) = \lambda(z'). \tag{12.7.17}$$

By taking into account eq. (12.7.13),

$$p_2(z') \left((A^{12}_> - A^{12}_<) f_1'(z') f_2(z') + (A^{21}_> - A^{21}_<) f_2'(z') f_1(z') \right) = \lambda(z'), \tag{12.7.18}$$

Since $A^{12}_< + A^{21}_< = A^{12}_> + A^{21}_> \Leftrightarrow A^{12}_> - A^{12}_< = -(A^{21}_> - A^{21}_<)$,

$$\begin{aligned}
p_2(z')(A^{21}_> - A^{21}_<) \mathrm{Wr}_{z'}(f_1, f_2) &= \lambda(z'), \\
p_2(z')(A^{21}_> - A^{21}_<) W_0 \exp\left(-\int_b^{z'} \frac{p_1(z'')}{p_2(z'')} \mathrm{d}z'' \right) &= \lambda(z'),
\end{aligned} \tag{12.7.19}$$

where eq. (7.7.60) was employed in the second line. We see that, given a differential operator D of the form in eq. (12.7.1), this amounts to solving for the measure $\lambda(z')$: it is fixed, up to an overall multiplicative constant $(A^{21}_{>} - A^{21}_{<})W_0$, by the $p_{1,2}$. (Remember the Wronskian itself is fixed up to an overall constant by $p_{1,2}$; cf. eq. (7.7.60).) Furthermore, note that $A^{21}_{>} - A^{21}_{<}$ can be absorbed into the functions $f_{1,2}$, since the latter's normalization has remained arbitrary till now. Thus, we may choose $A^{21}_{>} - A^{21}_{<} = 1 = -(A^{12}_{>} - A^{12}_{<})$. At this point,

$$\begin{aligned} G(z,z') = A^1 f_1(z) f_1(z') + A^2 f_2(z) f_2(z') & \\ + \Theta(z-z')((A^{12}_{<} - 1) f_1(z) f_2(z') + A^{21}_{>} f_2(z) f_1(z')) & \\ + \Theta(z'-z)(A^{12}_{<} f_1(z) f_2(z') + (A^{21}_{>} - 1) f_2(z) f_1(z')). & \end{aligned} \tag{12.7.20}$$

Because we are seeking a symmetric Green's function, let us also consider

$$\begin{aligned} G(z',z) = A^1 f_1(z') f_1(z) + A^2 f_2(z') f_2(z) & \\ + \Theta(z'-z)((A^{12}_{<} - 1) f_1(z') f_2(z) + A^{21}_{>} f_2(z') f_1(z)) & \\ + \Theta(z-z')(A^{12}_{<} f_1(z') f_2(z) + (A^{21}_{>} - 1) f_2(z') f_1(z)). & \end{aligned} \tag{12.7.21}$$

Comparing the first lines of equations (12.7.20) and (12.7.21) tells us the $A^{1,2}$ terms are automatically symmetric; whereas the second line of eq. (12.7.20) versus the third line of eq. (12.7.21), together with the third line of eq. (12.7.20) versus second line of eq. (12.7.21), says the terms involving $A^{12}_{\lessgtr}$ are symmetric iff $A^{12}_{<} = A^{21}_{>} \equiv \chi$. We gather, therefore,

$$G(z,z') = A^1 f_1(z) f_1(z') + A^2 f_2(z) f_2(z') + \mathcal{G}(z,z';\chi), \tag{12.7.22}$$

$$\begin{aligned} \mathcal{G}(z,z';\chi) \equiv (\chi - 1) \left\{ \Theta(z-z') f_1(z) f_2(z') + \Theta(z'-z) f_1(z') f_2(z) \right\} & \\ + \chi \left\{ \Theta(z-z') f_2(z) f_1(z') + \Theta(z'-z) f_2(z') f_1(z) \right\}. & \end{aligned} \tag{12.7.23}$$

The terms in the curly brackets can be written as $(\chi - 1) f_1(z_>) f_2(z_<) + \chi \cdot f_1(z_<) f_2(z_>)$, where $z_>$ is the larger and $z_<$ the smaller of the pair (z, z'). Moreover, we see it is these terms that contributes to the 'jump' in the first derivative across $z = z'$. The terms involving A^1 and A^2 are smooth across $z = z'$ provided, of course, the functions $f_{1,2}$ themselves are smooth; they are also homogeneous solutions with respect to both z and z'.

Summary Given any pair of linearly independent solutions to

$$D_z f_{1,2}(z) \equiv p_2(z) \frac{\mathrm{d}^2 f_{1,2}(z)}{\mathrm{d}z^2} + p_1(z) \frac{\mathrm{d} f_{1,2}(z)}{\mathrm{d}z} + p_0(z) f_{1,2}(z) = 0, \qquad a \leq z \leq b, \tag{12.7.24}$$

we may solve the symmetric Green's function equation(s)

$$D_z G(z,z') = p_2(z) W_0 \exp\left(-\int_b^z \frac{p_1(z'')}{p_2(z'')} \mathrm{d}z'' \right) \delta(z-z'), \tag{12.7.25}$$

$$D_{z'} G(z,z') = p_2(z') W_0 \exp\left(-\int_b^{z'} \frac{p_1(z'')}{p_2(z'')} \mathrm{d}z'' \right) \delta(z-z'), \tag{12.7.26}$$

$$G(z,z') = G(z',z), \tag{12.7.27}$$

by using the general ansatz

$$G(z,z') = G(z',z) = A^1 f_1(z) f_1(z') + A^2 f_2(z) f_2(z') + \mathcal{G}(z,z';\chi), \tag{12.7.28}$$

$$\mathcal{G}(z, z'; \chi) \equiv (\chi - 1) f_1(z_>) f_2(z_<) + \chi\ f_2(z_>) f_1(z_<), \tag{12.7.29}$$

$$z_> \equiv \max(z, z'), \quad z_< \equiv \min(z, z'). \tag{12.7.30}$$

Here W_0, $A^{1,2}$, and χ are arbitrary constants. However, once W_0 is chosen, the $f_{1,2}$ needs to be normalized properly to ensure the constant W_0 is recovered. Specifically,

$$\begin{aligned} \mathrm{Wr}_z(f_1, f_2)(z) = f_1(z) f_2'(z) - f_1'(z) f_2(z) &= \left(\frac{\partial \mathcal{G}(z = z'^+, z')}{\partial z} - \frac{\partial \mathcal{G}(z = z'^-, z')}{\partial z} \right) \bigg|_{z' \to z} \\ &= W_0 \exp\left(- \int_b^z \frac{p_1(z'')}{p_2(z'')} \mathrm{d}z'' \right). \end{aligned} \tag{12.7.31}$$

We also reiterate, up to the overall multiplicative constant W_0, the right hand side of eq. (12.7.25) is fixed once the differential operator D (in eq. (12.7.24)) is specified; in particular, one may not always be able to set the right hand side of eq. (12.7.25) to $\delta(z - z')$.

Problem 12.66. Hermitian Case

3D Green's Function of Laplacian As an example of the methods described here, let us work out the radial Green's function of the Laplacian in 3D Euclidean space. That is, we shall employ spherical coordinates

$$x^i = r(s_\theta c_\phi, s_\theta s_\phi, c_\theta), \tag{12.7.32}$$

$$x'^i = r'(s_{\theta'} c_{\phi'}, s_{\theta'} s_{\phi'}, c_{\theta'}); \tag{12.7.33}$$

and try to solve

$$-\vec{\nabla}_{\vec{x}}^2 G(\vec{x} - \vec{x}') = -\vec{\nabla}_{\vec{x}'}^2 G(\vec{x} - \vec{x}') = \frac{\delta(r - r')}{r r'} \delta(c_\theta - c_{\theta'}) \delta(\phi - \phi'). \tag{12.7.34}$$

Because of the rotation symmetry of the problem – we know, in fact,

$$G(\vec{x} - \vec{x}') = \frac{1}{4\pi |\vec{x} - \vec{x}'|} = (4\pi)^{-1} \left(r^2 + r'^2 - 2 r r' \cos\gamma \right)^{-1/2} \tag{12.7.35}$$

depends on the angular coordinates through the dot product $\cos\gamma \equiv \vec{x} \cdot \vec{x}'/(r r') = \widehat{x} \cdot \widehat{x}'$. This allows us to postulate the ansatz

$$G(\vec{x} - \vec{x}') = \sum_{\ell=0}^{\infty} \frac{\widetilde{g}_\ell(r, r')}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta, \phi) \overline{Y_\ell^m(\theta', \phi')}. \tag{12.7.36}$$

By applying the Laplacian in spherical coordinates (cf. eq. (12.2.96)) and using the completeness relation for spherical harmonics in eq. (12.2.73), eq. (12.7.34) becomes

$$\begin{aligned} \sum_{\ell=0}^{\infty} \frac{\widetilde{g}_\ell'' + (2/r)\widetilde{g}_\ell' - \ell(\ell+1) r^{-2} \widetilde{g}_\ell}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta, \phi) \overline{Y_\ell^m(\theta', \phi')} \\ = -\frac{\delta(r - r')}{r r'} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta, \phi) \overline{Y_\ell^m(\theta', \phi')}, \end{aligned} \tag{12.7.37}$$

with each prime representing ∂_r. Equating the (ℓ, m) term on each side,

$$D_r \widetilde{g}_\ell \equiv \widetilde{g}_\ell'' + \frac{2}{r}\widetilde{g}_\ell' - \frac{\ell(\ell+1)}{r^2}\widetilde{g}_\ell = -(2\ell+1)\frac{\delta(r-r')}{rr'}. \tag{12.7.38}$$

We already have the δ-function measure – it is $-(2\ell+1)/r^2$ – but it is instructive to check its consistency with the right hand side of (12.7.25); here, $p_1(r) = 2/r$ and $p_2(r) = 1$, and

$$W_0 \exp\left(-2\int^r \mathrm{d}r''/r''\right) = W_0 e^{-2\ln r} = W_0 r^{-2}. \tag{12.7.39}$$

Now, the two linearly independent solutions to $D_r f_{1,2}(r) = 0$ are

$$f_1(r) = \frac{F_1}{r^{\ell+1}}, \qquad f_2(r) = F_2 r^\ell, \qquad F_{1,2} = \text{constant}. \tag{12.7.40}$$

The radial Green's function must, according to eq. (12.7.28), take the form

$$\widetilde{g}_\ell(r,r') = \frac{A_\ell^1}{(rr')^{\ell+1}} + A_\ell^2 (rr')^\ell + \mathcal{G}_\ell(r,r'), \tag{12.7.41}$$

$$\mathcal{G}_\ell(r,r') \equiv F\left\{\frac{\chi_\ell - 1}{r_>}\left(\frac{r_<}{r_>}\right)^\ell + \frac{\chi_\ell}{r_<}\left(\frac{r_>}{r_<}\right)^\ell\right\}, \tag{12.7.42}$$

$$r_> \equiv \max(r,r'),\ r_< \equiv \min(r,r'), \tag{12.7.43}$$

where $A_\ell^{1,2}$, F, and χ_ℓ are constants. (What happened to $F_{1,2}$? Strictly speaking $F_1 F_2$ should multiply $A_\ell^{1,2}$ but since the latter is arbitrary their product(s) may be assimilated into one constant(s); similarly, in $\mathcal{G}_\ell(r,r')$, $F = F_1 F_2$ but since $F_{1,2}$ occurs as a product, we may as well call it a single constant.) To fix F, we employ eq. (12.7.31).

$$-\frac{2\ell+1}{r^2} = F\ \mathrm{Wr}_r\left(r^{-\ell-1}, r^\ell\right) = \frac{\partial \mathcal{G}(r=r'^+)}{\partial r} - \frac{\partial \mathcal{G}(r=r'^-)}{\partial r}. \tag{12.7.44}$$

Carrying out the derivatives explicitly,

$$\begin{aligned} -\frac{2\ell+1}{r^2} &= F\left\{\frac{\partial}{\partial r}\left(\frac{1}{r'}\left(\frac{r}{r'}\right)^\ell\right)_{r=r'^-} - \frac{\partial}{\partial r}\left(\frac{1}{r}\left(\frac{r'}{r}\right)^\ell\right)_{r=r'^+}\right\} \\ &= F\left\{\frac{\ell \cdot r^{\ell-1}}{r^{\ell+1}} + \frac{(\ell+1)r^\ell}{r^{\ell+2}}\right\} = F\frac{2\ell+1}{r^2}. \end{aligned} \tag{12.7.45}$$

Thus, $F = -1$. We may take the limit $r \to 0$ or $r' \to 0$ and see that the terms involving A_ℓ^1 and $(\chi_\ell/r_<)(r_>/r_<)^\ell$ in eq. (12.7.41) will blow up for any ℓ; while $1/(4\pi|\vec{x}-\vec{x}'|) \to 1/(4\pi r')$ or $\to 1/(4\pi r)$ does not. This implies $A_\ell^1 = 0$ and $\chi_\ell = 0$. Next, by considering the limits $r \to \infty$ or $r' \to \infty$, we see that the A_ℓ^2 term will blow up for $\ell > 0$, whereas, in fact, $1/(4\pi|\vec{x}-\vec{x}'|) \to 0$. Hence $A_{\ell>0}^2 = 0$. Moreover, the $\ell = 0$ term involving A_0^2 is a constant in space because $Y_{\ell=0}^m = 1/\sqrt{4\pi}$ and does not decay to zero for $r, r' \to \infty$; therefore, $A_0^2 = 0$ too. Equation (12.7.41) now stands as

$$\widetilde{g}_\ell(r,r') = \frac{1}{r_>}\left(\frac{r_<}{r_>}\right)^\ell, \tag{12.7.46}$$

which in turn means eq. (12.7.36) is

$$G(\vec{x}-\vec{x}') = \frac{1}{4\pi|\vec{x}-\vec{x}'|} = \frac{1}{r_>}\sum_{\ell=0}^{\infty}\frac{1}{2\ell+1}\left(\frac{r_<}{r_>}\right)^{\ell}\sum_{m=-\ell}^{\ell}Y_\ell^m(\theta,\phi)\overline{Y_\ell^m(\theta',\phi')}. \tag{12.7.47}$$

If we use the addition formula in eq. (12.2.76), we then recover eq. (12.3.58).

Problem 12.67. Can you perform a similar "jump condition" analysis for the 2D Green's function of the negative Laplacian? Your answer should be proportional to eq. (2.0.54). Hint: Start by justifying the ansatz

$$G_2(\vec{x}-\vec{x}') = \sum_{\ell=-\infty}^{+\infty}\widetilde{g}_\ell(r,r')e^{i\ell(\phi-\phi')}, \tag{12.7.48}$$

where $\vec{x} \equiv r(\cos\phi, \sin\phi)$ and $\vec{x}' \equiv r'(\cos\phi', \sin\phi')$. Carry out the jump condition analysis, assuming the radial Green's function $\widetilde{g}_\ell$ is a symmetric one. You should be able to appeal to the solution in eq. (9.7.101) to argue there are no homogeneous contributions to this 2D Green's function; i.e., the $A^1 = A^2 = 0$ in eq. (12.7.28) are zero in this case. □

12.8 ⋆Covariant Helmholtz Decomposition of 3-Vectors

Consider an infinite curved 3-dimensional space

$$\mathrm{d}\ell^2 = g_{ij}(\vec{x})\mathrm{d}x^i\mathrm{d}x^j \tag{12.8.1}$$

such that $\vec{\nabla}^2\psi = \delta^D(\vec{x}-\vec{x}')/\sqrt[4]{|g(\vec{x})g(\vec{x}')|}$ admits a well-defined solution. (A closed space such a sphere would not admit a solution, because the volume integral of $\vec{\nabla}^2\psi$ on the left hand side is always zero; while that of the right hand side would have to yield 1.) Then the Helmholtz decomposition of a vector states that any arbitrary V^i may always be written as the sum of a gradient and a curl,

$$V^i = \nabla^i\psi - \widetilde{\epsilon}^{ijk}\nabla_j U_k, \tag{12.8.2}$$

where

$$\psi(\vec{x}) = \int \mathrm{d}^3\vec{x}'\sqrt{|g(\vec{x}')|}G(\vec{x},\vec{x}')\nabla_{i'}V^{i'}(\vec{x}'), \tag{12.8.3}$$

$$U_i(\vec{x}) = \sigma_g\int \mathrm{d}^3\vec{x}'\sqrt{|g(\vec{x}')|}G_{ij'}(\vec{x},\vec{x}')\widetilde{\epsilon}^{j'a'b'}\partial_{a'}V_{b'}(\vec{x}'), \tag{12.8.4}$$

$$\sigma_g \equiv \text{sign}\ \det g_{ab}. \tag{12.8.5}$$

The vector is divergence-free, $\nabla_i U^i = 0$, whereas the Green's functions obey

$$\vec{\nabla}^2_{\vec{x}}G(\vec{x},\vec{x}') = \vec{\nabla}^2_{\vec{x}'}G(\vec{x},\vec{x}') = \frac{\delta^{(3)}(\vec{x}-\vec{x}')}{\sqrt[4]{|g(\vec{x})g(\vec{x}')|}} \tag{12.8.6}$$

and

$$\vec{\nabla}^2_{\vec{x}} G_{ij'} - R_i^{\ l}(\vec{x}) G_{lj'} = \vec{\nabla}^2_{\vec{x}'} G_{ij'} - R_{j'}^{\ l'}(\vec{x}') G_{il'} = g_{ij'}(\vec{x}, \vec{x}') \frac{\delta^{(3)}(\vec{x} - \vec{x}')}{\sqrt[4]{|g(\vec{x}) g(\vec{x}')|}}. \tag{12.8.7}$$

The $\vec{\nabla}^2 = \nabla_i \nabla^i$ is the Laplacian; $R_i^{\ l}$ the Ricci tensor; and $g_{ij'}(\vec{x}, \vec{x}')$ the parallel propagator, whose coincidence limit returns the metric, $g_{ij'}(\vec{x} \to \vec{x}') = g_{i'j'}(\vec{x}')$.

The divergence of the vector Green's function $G_{ij'}$ with respect to $\vec{x}$ is the negative gradient of the scalar one with respect to $\vec{x}'$,

$$\nabla^i G_{ij'}(\vec{x}, \vec{x}') = -\nabla_{j'} G(\vec{x}, \vec{x}'). \tag{12.8.8}$$

Curl and Divergence To understand eq. (12.8.2), we start by checking its consistency with the curl and divergence of V^i. Via a direct calculation, the curl of V^i yields (cf. eq. (9.6.81))

$$\begin{aligned} \widetilde{\epsilon}^{ijk} \partial_j V_k &= \sigma_g \left(\vec{\nabla}^2 U^i - R^i_{\ j} U^j \right) && (12.8.9) \\ &= \sigma_g^2 \int \mathrm{d}^3\vec{x}' \sqrt{|g(\vec{x}')|} \left(\vec{\nabla}^2_{\vec{x}} G^i_{\ j'} - R^{il}(\vec{x}) G_{lj'} \right) \widetilde{\epsilon}^{j'a'b'} \partial_{a'} V_{b'} && (12.8.10) \\ &= \widetilde{\epsilon}^{iab} \partial_a V_b(\vec{x}). && (12.8.11) \end{aligned}$$

Since the divergence of the second term on the right hand side of eq. (12.8.2) is identically zero, the divergence of V^i is simply the Laplacian of ψ.

$$\begin{aligned} \nabla_i V^i &= \int \mathrm{d}^3\vec{x}' \sqrt{|g(\vec{x}')|} \vec{\nabla}^2_{\vec{x}} G(\vec{x}, \vec{x}') \nabla_{a'} V^{a'} && (12.8.12) \\ &= \nabla_a V^a(\vec{x}). && (12.8.13) \end{aligned}$$

Let us put

$$V^i = \nabla^i \psi - \widetilde{\epsilon}^{ijk} \nabla_j U_k + W^i, \tag{12.8.14}$$

where W^i is arbitrary for now. Taking the divergence and curl on both sides would teach us, W^i itself must be curl and divergence free:

$$\widetilde{\epsilon}^{ijk} \nabla_j W_k = 0 = \nabla^i W_i. \tag{12.8.15}$$

Curl free implies $W^i = \nabla^i \varphi$ for some scalar φ and divergence free, in turn, tell us $\vec{\nabla}^2 \varphi = 0$. Provided φ itself does not blow up at infinity, the only solution is $\varphi = \varphi_0 =$ constant. That in turn says $W^i = \nabla^i \varphi_0 = 0$.

Scalar and Vector Green's Functions

A Copyleft

You should feel free to re-distribute these notes, as long as they remain freely available. *Please do not* post on-line solutions to the problems I have written here!

B Group Theory

What is a group? A group is a collection of objects $\{a, b, c, \dots\}$ with a well defined multiplication $\cdot$ rule, with the following axioms.

- *Closure* If a and b are group elements, so is $a \cdot b$.
- *Associativity* The multiplication is always associative: $a \cdot b \cdot c = (a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- *Identity* There is an identity e, which obeys $a \cdot e = a$ for any group element a.
- *Inverse* For any group element b, there is always an inverse b^{-1} which obeys $b \cdot b^{-1} = e$.

Basic facts The left and right inverse of a group element is the same $b^{-1} \cdot b = b \cdot b^{-1} = e$. The identity is its own inverse $e^{-1} = e$; and the left identity is the same as that of the right, namely $e \cdot a = a \cdot e = a$ for all a.

Problem B.1. Group elements and linear operators Prove that *invertible* linear operators acting on a given vector space themselves form a vector space. (Hint: In §(4) we have already seen that the space of all linear operators form a vector space; so you merely need to refer to the discussion at the end of §(4.1).)

Suppose $\{X_i | i = 1, \dots, N\}$ is a collection of such invertible linear operators that are closed under multiplication, namely

$$X_i X_j = c_{ijk} X_k \tag{B.0.1}$$

for any $i, j, k \in \{1, 2, \dots, N\}$; where c_{ijk} are complex numbers. Prove that these $\{X_i\}$ form a group. This result is the basis of group representation theory – turning the study of groups to that of linear operators. □

Group representations A representation of a group is a map from its elements $\{g_1, g_2, \dots\}$ to a set of invertible linear operators $\{D(g_1), D(g_2), \dots\}$ which are closed under multiplication, in such a way that preserves the group multiplication rules. In other words, the linear operators are functions of the group elements $D(g)$, such that

$$D(g_1)D(g_2) = D(g_1 g_2). \tag{B.0.2}$$

The identity maps to the identity

$$D(e) = \mathbb{I}. \tag{B.0.3}$$

because $D(e)D(g) = D(e \cdot g) = D(g) = \mathbb{I} \cdot D(g)$ for all g. Also,

$$D(g^{-1}) = D(g)^{-1} \tag{B.0.4}$$

because $D(g^{-1})D(g) = D(g^{-1}g) = \mathbb{I} = D(g)^{-1}D(g)$.

Examples & Motivation Examples of groups representations can be found in §(5.4). Quantum mechanical motivation for group representations can be found in §(5.5).

C Conventions

Function argument There is a notational ambiguity whenever we write "f is a function of the variable x" as $f(x)$. If you did not know f were meant to be a function, what is $f(x+\sin(\theta))$? Is it some number f times $x + \sin\theta$? For this reason, in my personal notes and research papers I reserve square brackets exclusively to denote the argument of functions – I would always write $f[x + \sin[\theta]]$, for instance. (This is a notation I borrowed from the software Mathematica.) However, in these lecture notes I will stick to the usual convention of using parenthesis; but I wish to raise awareness of this imprecision in our mathematical notation.

Einstein summation and index notation Repeated indices are always summed over, unless otherwise stated:

$$\xi^i p_i \equiv \sum_i \xi^i p_i. \tag{C.0.1}$$

Often I will remain agnostic about the range of summation, unless absolutely necessary.

In such contexts when the Einstein summation is in force – unless otherwise stated – both the superscript and subscript are enumeration labels. ξ^i is the ith component of $(\xi^1, \xi^2, \xi^3, \dots)$, not some variable ξ raised to the ith power. The position of the index, whether it is super- or sub-script, usually represents how it transforms under the change of basis or coordinate system used. For instance, instead of calling the 3D Cartesian coordinates (x, y, z), we may now denote them collectively as x^i, where $i = 1, 2, 3$. When you rotate your coordinate system $x^i \to R^i{}_j y^j$, the derivative transforms as $\partial_i \equiv \partial/\partial x^i \to (R^{-1})^j{}_i \partial_j$.

Dimensions Unless stated explicitly, the number of space dimensions is D; it is an arbitrary positive integer greater or equal to one. Unless stated explicitly, the number of spacetime dimensions is $d = D + 1$; it is an arbitrary positive integer greater or equal to 2.

Spatial vs. spacetime indices I will employ the common notation that spatial indices are denoted with Latin/English alphabets whereas spacetime ones with Greek letters. Spacetime indices begin with 0; the 0th index is in fact time. Spatial indices start at 1. I will also use the "mostly minus" convention for the metric; for e.g., the flat spacetime geometry in Cartesian coordinates reads

$$\eta_{\mu\nu} = \mathrm{diag}\left[1, -1, \dots, -1\right], \tag{C.0.2}$$

where "diag$[a_1, \dots, a_N]$" refers to the diagonal matrix, whose diagonal elements (from the top left to the bottom right) are respectively a_1, a_2, $\dots$, a_N. Spatial derivatives are $\partial_i \equiv \partial/\partial x^i$; and spacetime ones are $\partial_\mu \equiv \partial/\partial x^\mu$. The scalar wave operator in flat spacetime, in Cartesian coordinates, read

$$\partial^2 = \Box = \eta^{\mu\nu}\partial_\mu\partial_\nu. \tag{C.0.3}$$

The Laplacian in flat space, in Cartesian coordinates, read instead

$$\vec{\nabla}^2 = \delta^{ij}\partial_i\partial_i, \tag{C.0.4}$$

where δ_{ij} is the Kronecker delta, the unit $D \times D$ matrix $\mathbb{I}$:

$$\delta_{ij} = 1, \qquad i = j$$

$$= 0, \qquad i \neq j. \tag{C.0.5}$$

Index (anti-)symmetrization The symbols $[\dots]$ and $\{\dots\}$ denote anti-symmetrization and symmetrization respectively. In particular,

$$T_{[i_1 \dots i_N]} = \sum_{\text{even permutations }\Pi} T_{\Pi[i_1,\dots i_N]} - \sum_{\text{odd permutations }\Pi} T_{\Pi[i_1,\dots i_N]}, \tag{C.0.6}$$

$$T_{\{i_1 \dots i_N\}} = \sum_{\text{all permutations }\Pi} T_{\Pi[i_1,\dots i_N]}. \tag{C.0.7}$$

For example,

$$T_{[ijk]} = T_{ijk} - T_{ikj} - T_{jik} + T_{jki} - T_{kji} + T_{kij} \tag{C.0.8}$$

$$T_{\{ijk\}} = T_{ijk} + T_{ikj} + T_{jik} + T_{jki} + T_{kji} + T_{kij}. \tag{C.0.9}$$

Caution Beware that many relativity texts define their (anti-)symmetrization with a division by a factorial; namely,

$$T_{[i_1 \dots i_N]} = \frac{1}{N!}\left(\sum_{\text{even permutations }\Pi} T_{\Pi[i_1,\dots i_N]} - \sum_{\text{odd permutations }\Pi} T_{\Pi[i_1,\dots i_N]}\right), \tag{C.0.10}$$

$$T_{\{i_1 \dots i_N\}} = \frac{1}{N!}\sum_{\text{all permutations }\Pi} T_{\Pi[i_1,\dots i_N]}. \tag{C.0.11}$$

I prefer *not* to do so, because of the additional baggage incurred by these numerical factors when performing concrete computations.

D Physical Constants and Dimensional Analysis

In much of these notes we will set Planck's reduced constant and the speed of light to unity: $\hbar = c = 1$. (In the General Relativity literature, Newton's gravitational constant $G_{\rm N}$ is also often set to one.) What this means is, we are using $\hbar$ as our base unit for angular momentum; and c for speed.

Since $[c]$ is Length/Time, setting it to unity means

$$[\text{Length}] = [\text{Time}]\ .$$

Since in SI units $c = 299,792,458$ meters/second, we have

$$1 \text{ second} = 299,792,458 \text{ meters}, \qquad (c = 1). \tag{D.0.1}$$

Einstein's $E = mc^2$, once $c = 1$, becomes the statement that

$$[\text{Energy}] = [\text{Mass}]\ .$$

Because $[\hbar]$ is Energy $\times$ Time, setting it to unity means

$$[\text{Energy}] = [1/\text{Time}]\ .$$

In SI units, $\hbar \approx 1.0545718 \times 10^{-34}$ Joules second – hence,

$$1 \text{ second} \approx 1/(1.0545718 \times 10^{-34} \text{ Joules}) \qquad (\hbar = 1). \tag{D.0.2}$$

Altogether, with $\hbar = c = 1$, we may state

$$[\text{Mass}] = [\text{Energy}] = [1/\text{Time}] = [1/\text{Length}] \,.$$

Physically speaking, the energy-mass and time-length equivalence can be attributed to relativity (c); whereas the (energy/mass)-(time/length)$^{-1}$ equivalence can be attributed to quantum mechanics ($\hbar$).

High energy physicists prefer to work with eV (or its multiples, such as MeV or GeV); and so it is useful to know the relation

$$\hbar c \approx 197.326, 98 \text{MeV fm} \tag{D.0.3}$$

where fm = femtometer = 10^{-15} meters. Hence,

$$10^{-15} \text{ meters} \approx 1/(197.326, 98 \text{ MeV}), \qquad (\hbar c = 1). \tag{D.0.4}$$

Using these 'natural units' $\hbar = c = 1$ is a very common practice throughout the physics literature.

One key motivation behind setting to unity physical constants occurring frequently in your physics analysis, is that it allows you to focus on the quantities that are more specific (and hence more important) to the problem at hand. Carrying these physical constants around clutter your calculation, and increases the risk of mistakes due to this additional burden. For instance, in the Bose-Einstein or Fermi-Dirac statistical distribution $1/(\exp(E/(k_B T)) \pm 1)$ – where E is the energy of the particle(s), k_B the Boltzmann constant, and T the temperature of the system – what's physically important is the ratio of the energy scales, E versus $k_B T$. The Boltzmann constant k_B is really a distraction, and ought to be set to one, so that temperature is now measured in units of energy: the cleaner expression now reads $1/(\exp(E/T) \pm 1)$.

Another reason why one may want to set a physical constant to unity is because, it could be such an important benchmark in the problem at hand that it should be employed as a base unit. For instance, most down-to-Earth engineering problems may not benefit from using the speed of light c as their basic unit for speed. In non-relativistic astrophysical systems bound by their mutual gravity, however, it turns out that General Relativistic corrections to the Newtonian law of gravity will be akin to a series in v/c, where v is the typical speed of the bodies that comprise the system. The expansion parameter then becomes $0 \leq v < 1$ if we set $c = 1$ – i.e., if we measure all speeds relative to c – which in turn means this 'post-Newtonian' expansion is a series in the gravitational potential $G_{\mathrm{N}} M/r$ through the virial theorem (kinetic energy $\sim$ potential energy) $v \sim \sqrt{G_{\mathrm{N}} M/r}$.

Newton's gravitational constant takes the form

$$G_{\mathrm{N}} = 6.67430(15) \times 10^{-11} m^3 \cdot \text{kg}^{-1} \cdot s^{-2} \tag{D.0.5}$$

$$\approx 6.7086 \times 10^{-39} \hbar c (\text{GeV}/c^2)^{-2}. \tag{D.0.6}$$

Just from this dimensional analysis alone, when $\hbar = c = 1$, one may form a mass-energy scale ('Planck mass')

$$M_{\mathrm{pl}} \equiv \frac{1}{\sqrt{32\pi G_{\mathrm{N}}}} \approx 10^{18} \text{GeV}. \tag{D.0.7}$$

(The 32π is for technical convenience.) This suggests – since $M_{\rm pl}$ appears to involve relativity (c), quantum mechanics ($\hbar$) and gravitation ($G_{\rm N}$) – that the energy scale required to probe quantum aspects of gravity is roughly $M_{\rm pl}$. Therefore, it may be useful to set $M_{\rm pl} = 1$ in quantum gravity calculations, so that all other energy scales in a given problem, say the quantum amplitude of scattering gravitons, are now measured relative to it.

I recommend the following resource for physical and astrophysical constants, particle physics data, etc.:

Particle Data Group: http://pdg.lbl.gov .

Problem D.1. Let $\hbar = c = 1$.

- If angular momentum is 3.34 in natural units, convert it to SI units.
- What is the mass of the Sun in MeV? What is its mass in parsec?
- If Pluto is orbiting roughly 40 astronomical units from the Sun, how many seconds is this orbital distance? How many GeV is it?
- Work out the Planck mass in eq. (D.0.7) in seconds, meters, and GeV.

□

Problem D.2. Schwarzschild radius of Sun Explain why, when $c = 1$ and $M_\odot$ is the mass of the Sun, the latter's Schwarzschild radius is

$$r_s = 2G_{\rm N}M_\odot \approx 2.953 \text{ km}. \tag{D.0.8}$$

Can you put back the factors of c in r_s, assuming it is of physical dimensions length? (There are no $\hbar$s in this classical problem.) By replacing $M_\odot$ with $M_{\rm Earth}$, compute the Schwarzschild radius of the Earth. □

Problem D.3. Yukawa Potential In $(3+1)-$dimensional Quantum Field Theory, an exchange of a massless (integer spin) boson between two objects results in a $1/r$ Coulomb-like potential, where r is the distance between them. (For example, the Coulomb potential between two point charges in fact arises from an exchange of a virtual photon.) When a boson of mass $m > 0$ is exchanged, a short range Yukawa potential $V(r) = e^{-mr}/(4\pi r)$ is produced instead. Restore the appropriate factors of $\hbar$ and c in the exponential $\exp(-mr)$. Hint: I find it convenient to remember the dimensions of $\hbar c$; see eq. (D.0.3).

The $W^\pm$ and Z^0 gauge bosons have masses $M_W = 80$ GeV$/c^2$ and $M_Z = 91$ GeV$/c^2$ respectively; they mediate the weak interactions. Compute the length scales $1/M_W$ and $1/M_Z$ in Fermi fm $= 10^{-15}m$. □

Problem D.4. Consider the following wave operator for a particle of mass $m > 0$,

$$\mathcal{W} \equiv \partial_\mu \partial^\mu + m^2, \qquad x^\mu \equiv (t, \vec{x}). \tag{D.0.9}$$

- In $\mathcal{W}$, put back the $\hbar$s only.
- In $\mathcal{W}$, put back the cs only.

- In $\mathcal{W}$, put back both the $\hbar$s and cs.

Assume that $\mathcal{W}$ has dimensions of $1/[\text{Length}^2]$. $\square$

Problem D.5. Dirac Operator At the core of the Dirac equation is the differential operator $\not{D} \equiv i\gamma^\mu \partial_\mu - m\mathbb{I}_{4\times 4}$, where $c = \hbar = 1$, $\{\gamma^\mu\}$ are dimensionless 4×4 matrices, $\mathbb{I}_{4\times 4}$ is the identity, and m is a (real) mass parameter. Assuming $[\not{D}]$ is inverse length, put back the $\hbar$s while maintaining $c = 1$. $\square$

E Acknowledgments

I wish to thank the following people for pointing out errors/typos, and/or for discussions that lead to new or better presentations, etc.: Jia-En Chen, Wei Hao Chen, Li-Ying Chou, Chun-Chih Huang, Afif Ismail, Jake Leistico, Kuan-Nan Lin, Alec Lovlein, Kiersten Meigs, Chern Xi Pang, Hadi Papei, Jason Payne, Evan Severson, Leon Tsai, Chun-Yu Wei (Viona), Cheng Han Wu, Yi-Chen Wu, Yu-Hung Yeh, Jia-You Yuan, and Afidah Zuroida.

I thank Tolga Birkandan for pointing me to Sages Math (linked below, [46]); and Leo Stein for clarifying how xAct [47] carries out perturbation theory.

I would also like to thank Tanmay Vachaspati for encouraging me to post these lecture notes on the arXiv (@ 1701.00776).

F Last update: July 1, 2025

References

[1] C. R. Frye and C. J. Efthimiou, "Spherical Harmonics in p Dimensions," arXiv:1205.3548 [math.CA].

[2] Digital Library of Mathematical Functions (DLMF)

[3] Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables (AMS55)

[4] D. Zwillinger, I.S. Gradshteyn, I.M. Ryzhik, "Table of Integrals, Series, and Products," Academic Press (2014)

[5] N. N. Lebedev, "Special Functions & Their Applications," Dover Publications; Revised ed. edition (June 1, 1972)

[6] Harry Hochstadt, "The Functions of Mathematical Physics," Dover Publications; Revised edition (February 29, 2012)

[7] A.D. Polyanin "Handbook of Linear Partial Differential Equations for Engineers and Scientists," Chapman and Hall/CRC (2001)

[8] Bertrand, J., Bertrand, P., Ovarlez, J. "The Mellin Transform." The Transforms and Applications Handbook: Second Edition. Ed. Alexander D. Poularikas Boca Raton: CRC Press LLC, 2000

[9] V.F. Zaitsev and A.D. Polyanin, "Handbook of Exact Solutions for Ordinary Differential Equations," Chapman and Hall/CRC (2001)

[10] A.D. Polyanin, V.F. Zaitsev, "Handbook of Nonlinear Partial Differential Equations," Chapman and Hall/CRC (2001)

[11] J. J. Sakurai, Jim Napolitano, "Modern Quantum Mechanics," Cambridge University Press; 2nd edition (September 22, 2017)

[12] Steven Weinberg, "Lectures on Quantum Mechanics," Cambridge University Press (2015), 2nd Edition.

[13] Philip M. Morse, Herman Feshbach, "Methods of Theoretical Physics," Volumes 1 and 2.

[14] Byron and Fuller, "Mathematics of Classical and Quantum Physics," Dover Publications; Revised ed. edition (August 20, 1992)

[15] J. Mathews, R.L. Walker, "Mathematical Methods of Physics," New York : W. A. Benjamin 1964

[16] C.M. Bender, S.A. Orszag, "Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory," Springer (2010)

[17] Frank Olver, "Asymptotics and Special Functions" A K Peters/CRC Press; 2nd edition (January 24, 1997)

[18] George B. Arfken, Hans J. Weber, Frank E. Harris "Mathematical Methods for Physicists," 7th Edition, Academic Press (2013)

[19] J. Brown, R. Churchill "Complex Variables and Applications" McGraw-Hill (2013)

[20] R. P. Woodard, "Avoiding dark energy with 1/r modifications of gravity," Lect. Notes Phys. **720**, 403-433 (2007) doi:10.1007/978-3-540-71013-4_14 [arXiv:astro-ph/0601672 [astro-ph]].

[21] R. P. Woodard, "Ostrogradsky's theorem on Hamiltonian instability," Scholarpedia **10**, no.8, 32243 (2015) doi:10.4249/scholarpedia.32243 [arXiv:1506.02210 [hep-th]].

[22] E. Poisson, "A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics," Cambridge University Press (2007). (Online draft here.)

[23] B. Schutz, "Geometrical Methods of Mathematical Physics," Cambridge University Press; (January 28, 1980)

[24] S. Weinberg, "Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity,"

[25] Misner, Thorne, Wheeler, "Gravitation," Princeton University Press (2017)

[26] M. P. Hobson, G. P. Efstathiou and A. N. Lasenby, "General relativity: An introduction for physicists," Cambridge, UK: Univ. Pr. (2006) 572 p

[27] S. M. Carroll, "Lecture Notes on General Relativity," Webpage here.

S. M. Carroll, "Spacetime and Geometry," San Francisco, USA: Addison-Wesley (2004) 513p

[28] Gerad 't Hooft, "Introduction to General Relativity," Available here.

[29] Matthias Blau, "Lecture Notes on General Relativity," Available here.

Matthias Blau, "Plane Waves and Penrose Limits," Available here.

[30] J. C. Baez and E. F. Bunn, "The Meaning of Einstein's equation," Am. J. Phys. **73**, 644 (2005) doi:10.1119/1.1852541 [gr-qc/0103044].

[31] P. K. Townsend, "Black holes: Lecture notes," gr-qc/9707012.

[32] Wu-Ki Tung, "Group Theory in Physics," World Scientific (September 1, 1985)

[33] J. Schwichtenberg, "Physics from Symmetry," Springer International Publishing, 2018, ISBN 978-3-319-66630-3, 978-3-319-66631-0 doi:10.1007/978-3-319-66631-0

[34] George F. Simmons, "Differential Equations with applications and historical notes," CRC Press, Taylor and Francis, Third Edition

[35] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory," Addison-Wesley, 1995, ISBN 978-0-201-50397-5

[36] S. Carlip and C. DeWitt-Morette, "Where the Sign of the Metric Makes a Difference," Phys. Rev. Lett. **60**, 1599 (1988). doi:10.1103/PhysRevLett.60.1599

[37] G. W. Gibbons, H. Lu, D. N. Page and C. N. Pope, "The General Kerr-de Sitter metrics in all dimensions," J. Geom. Phys. **53**, 49 (2005) doi:10.1016/j.geomphys.2004.05.001 [hep-th/0404008].

[38] N. Ashby, "Relativity in the Global Positioning System," Living Rev. Rel. **6**, 1 (2003) [100 Years Of Relativity : space-time structure: Einstein and beyond, 257 (2005)]. doi:10.12942/lrr-2003-1, 10.1142/9789812700988_0010

[39] C. M. Will, "The Confrontation between General Relativity and Experiment," Living Rev. Rel. **17**, 4 (2014) doi:10.12942/lrr-2014-4 [arXiv:1403.7377 [gr-qc]].

[40] I. I. Shapiro, "Fourth Test of General Relativity," Phys. Rev. Lett. **13**, 789 (1964). doi:10.1103/PhysRevLett.13.789

[41] I. I. Shapiro, G. H. Pettengill, M. E. Ash, M. L. Stone, W. B. Smith, R. P. Ingalls and R. A. Brockelman, "Fourth Test of General Relativity: Preliminary Results," Phys. Rev. Lett. **20**, 1265 (1968). doi:10.1103/PhysRevLett.20.1265

[42] I. I. Shapiro, M. E. Ash, R. P. Ingalls, W. B. Smith, D. B. Campbell, R. B. Dyce, R. F. Jurgens and G. H. Pettengill, "Fourth test of general relativity - new radar result," Phys. Rev. Lett. **26**, 1132 (1971). doi:10.1103/PhysRevLett.26.1132

[43] Y. Z. Chu, "More On Cosmological Gravitational Waves And Their Memories," arXiv:1611.00018 [gr-qc].

[44] H. Soodak and M. S. Tiersten, "Wakes and waves in N dimensions," Am. J. Phys. **61** (395), May 1993

[45] Numerical Receipes (Older versions available online.)

[46] Sages Math ("*. . . free open-source mathematics software system. . .*")

[47] http://www.xact.es