Analytical Methods in Physics

Yi-Zen Chu

Contents

1 Preface 4

2 Complex Numbers and Functions 7

3 Matrix Algebra: A Review 12

3.1 Basics, Matrix Operations, and Special types of matrices 12
3.2 Determinants, Linear (In)dependence, Inverses and Eigensystems 15
3.3 Special Topic 1: 2D real orthogonal matrices 23

4 Linear Algebra 26

4.1 Definition 26
4.2 Inner Products 30
4.3 Linear Operators 36

4.3.1 Definitions and Fundamental Concepts 36
4.3.2 Hermitian Operators 44
4.3.3 Unitary Operation as Change of Orthonormal Basis 55
4.3.4 Additional Problems 62
4.4 Tensor Products of Vector Spaces 65
4.5 Continuous Spaces and Infinite $D$–Space 67

4.5.1 Preliminaries: Dirac’s $\delta$-“function”, eigenket integrals, and continuous (Lie group) operators 68
4.5.2 Spatial translations and the Fourier transform 75
4.5.3 Boundary Conditions, Finite Box, Periodic functions and the Fourier Series 86
4.5.4 Rotations in $D = 2$ Spatial Dimensions 92
4.5.5 Rotations in $D \geq 3$ Spatial Dimensions; Angular Momentum & Spin 98
4.5.6 Rotations in 3 Spatial Dimensions: Integer Spin & Spherical Harmonics 116
4.5.7 Rotations in 3 Spatial Dimensions: Half Integer Spin, SU$_2$ 121
4.5.8 Rotations in 3 Spatial Dimensions: ‘Adding’ Angular Momentum, Tensor Operators, Wigner-Eckart Theorem 122

4.6 Special Topic: Clebsch-Gordan Coefficients 141
4.7 Special Topic: Approximation Methods for Eigensystems 143

4.7.1 Rayleigh-Schrödinger Perturbation Theory 143
4.7.2 Variational Method 154
5 Calculus on the Complex Plane

5.1 Differentiation ............................................. 157
5.2 Cauchy’s integral theorems, Laurent Series, Analytic Continuation ........ 161
5.3 Poles and Residues ........................................ 169
5.4 Branch Points, Branch Cuts .................................. 177
5.5 Fourier Transforms ........................................ 182
5.5.1 Application: Damped Driven Simple Harmonic Oscillator ................. 184
5.6 Fourier Series ............................................. 188

6 Advanced Calculus: Special/Approximation Techniques, Asymptotic Expan-
sions and Variational Methods ................................ 190

6.1 Gaussian integrals ........................................... 190
6.2 Complexification ............................................. 192
6.3 Differentiation under the integral sign (Leibniz’s theorem) ......................... 192
6.4 Symmetry ..................................................... 194
6.5 Asymptotic expansion of integrals ................................ 197
6.5.1 Integration-by-parts (IBP) .................................. 197
6.5.2 Laplace’s Method, Method of Stationary Phase, Steepest Descent ........ 199
6.6 JWKB solution to $-\epsilon^2 \psi''(x) + U(x)\psi(x) = 0$, for $0 < \epsilon \ll 1$ .................. 205
6.7 Calculus of Variation ....................................... 207

7 Differential Geometry of Curved Spaces .................................. 208

7.1 Preliminaries, Tangent Vectors, Metric, and Curvature ......................... 208
7.2 Locally Flat Coordinates & Symmetries, Infinitesimal Volumes, General Tensors, Orthonormal Basis ..................... 214
7.3 Covariant derivatives, Parallel Transport, Geodesics, Levi-Civita, Hodge Dual .......... 228
7.4 Hypersurfaces ............................................... 253
7.4.1 Induced Metrics ............................................ 253
7.4.2 Fluxes, Gauss-Stokes’ theorems, Poincaré lemma ............... 256

8 Differential Geometry In Curved Spacetimes ................................ 266

8.1 Constancy of $c$, Poincaré and Lorentz symmetry ................................ 266
8.2 Lorentz Transformations in 4 Dimensions .................................. 282
8.2.1 $SL_2(C)$ Spinors and Spin-Half Vectors; Gravitational Time Dilation .......... 283
8.3 Curved Metrics, Orthonormal Frames & Volume; Timelike, Spacelike vs. Null Geodesics .................................................. 302
8.4 Connections, Curvature, Geodesics .................................. 310
8.5 Equivalence Principles, Geometry-Induced Tidal Forces, Isometries & Geometric Tensors ........................................... 323
8.6 Special Topic: Metric Perturbation Theory ................................ 338
8.6.1 Perturbed Flat Spacetimes & Gravitational Waves .................. 344
8.7 Special Topic: Conformal/Weyl Transformations .......................... 352
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. 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 §9, 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 \([X^i, P_j] = i\delta^j_i\) is not necessarily a quantum mechanical statement. This emphasis is based on the belief that the power of a given

\[\text{That the textbook originally assigned for this course relegated the axioms of Linear Algebra towards the very end of the discussion was one major reason why I decided to write these notes. This same book also cost nearly two hundred (US) dollars – a fine example of exorbitant textbook prices these days – so I am glad I saved my students quite a bit of their educational expenses that semester.}\]
mathematical tool is very much tied to its versatility – this issue arises again in the JWKB discussion within §(6), 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 §(5) 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 §(7) 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, §(8), 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 §(7). It is my hope that §(7) and §(8) can be used to build the differential geometric tools one could then employ to understand General Relativity, Einstein’s field equations for gravitation.

In §(9) 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

The motivational introduction to complex numbers, in particular the number $i$, is the solution to the equation

$$i^2 = -1. \quad (2.0.1)$$

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

$$z = x + iy \quad (2.0.2)$$

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

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 $\phi$ is the angle between the vector joining $(0, 0)$ to $(x, y)$ and the positive horizontal axis – then

$$(x, y) = (r \cos \phi, r \sin \phi). \quad (2.0.3)$$

Therefore a complex number must be expressible as

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

This actually takes a compact form using the exponential:

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

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.,

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

where the final equality follows from Pythagoras’ Theorem. The angle $\phi$ is denoted as

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

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

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

**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

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

\[2\text{Some of the material in this section is based on James Nearing’s } Mathematical Tools for Physics\]

\[3\text{Engineers use } j.\]
Arithmetic

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

\[ z_1 = x_1 + iy_1 \quad \text{and} \quad z_2 = x_2 + iy_2, \quad (2.0.10) \]

then

\[ z_1 \pm z_2 = (x_1 \pm x_2) + i(y_1 \pm y_2). \quad (2.0.11) \]

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

\[ z_1z_2 = r_1r_2 e^{i(\phi_1 + \phi_2)}. \quad (2.0.12) \]

To summarize:

Complex numbers \( \{z = x + iy = re^{i\phi} \mid x, y \in \mathbb{R}; r \geq 0, \phi \in \mathbb{R}\} \) 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 \( i \equiv -1 \) 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., \( z_1z_2z_3 = z_1(z_2z_3) = (z_1z_2)z_3 \).

**Complex conjugation**

Taking the complex conjugate of \( z = x + iy \) means we flip the sign of its imaginary part, i.e.,

\[ z^* = x - iy; \quad (2.0.13) \]

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

\[ e^{-i\phi} = \cos(-\phi) + i \sin(-\phi) = \cos \phi - i \sin \phi. \quad (2.0.14) \]

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

\[ z^*z = zz^* = (x + iy)(x - iy) = x^2 + y^2 = |z|^2. \quad (2.0.15) \]

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,

\[ \frac{x + iy}{a + ib} = \frac{(a - ib)(x + iy)}{a^2 + b^2} = \frac{(ax + by) + i(ay - bx)}{a^2 + b^2}. \quad (2.0.16) \]
Problem 2.6. Is \((z_1z_2)^* = z_1^*z_2^*\), 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 \(|z_1z_2| = |z_1||z_2|\)? What about \(|z_1/z_2| = |z_1|/|z_2|\)? Also show that \(\text{arg}(z_1 \cdot z_2) = \text{arg}(z_1) + \text{arg}(z_2)\). Strictly speaking, \(\text{arg}(z)\) is well defined only up to an additive multiple of \(2\pi\). Can you explain why? Hint: polar coordinates are very useful in this problem. \(\square\)

Problem 2.7. 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 \(z + z^* = 2\text{Re}(z)\) and \(z - z^* = 2i\text{Im}(z)\). Hint: use Cartesian coordinates. \(\square\)

Problem 2.8. Prove that the roots of a polynomial with real coefficients

\[ P_N(z) \equiv c_0 + c_1z + c_2z^2 + \cdots + c_Nz^N, \quad \{c_i \in \mathbb{R}\}, \quad (2.0.17) \]

come in complex conjugate pairs; i.e., if \(z\) is a root then so is \(z^*\). \(\square\)

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

\[ e^{\pm i \phi} = \cos \phi \pm i \sin \phi. \quad (2.0.18) \]

These two equations can be added and subtracted to yield

\[ \cos(z) = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin(z) = \frac{e^{iz} - e^{-iz}}{2i}, \quad \tan(z) = \frac{\sin(z)}{\cos(z)}. \quad (2.0.19) \]

We have made the replacement \(\phi \rightarrow z\). This change is cosmetic if \(0 \leq z < 2\pi\), but we can in fact now use eq. (2.0.19) to define the trigonometric functions in terms of the exp function for any complex \(z\).

Trigonometric identities can be readily obtained from their exponential definitions. For example, the addition formulas would now begin from

\[ e^{i(\theta_1 + \theta_2)} = e^{i\theta_1}e^{i\theta_2}. \quad (2.0.20) \]

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

\[ \cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2) = (\cos \theta_1 + i \sin \theta_1)(\cos \theta_2 + i \sin \theta_2) \]
\[ = (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\sin \theta_1 \cos \theta_2 + \sin \theta_2 \cos \theta_1). \quad (2.0.21) \]

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

\[ \cos(\theta_1 + \theta_2) = \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2, \quad (2.0.22) \]
\[ \sin(\theta_1 + \theta_2) = \sin \theta_1 \cos \theta_2 + \sin \theta_2 \cos \theta_1. \quad (2.0.23) \]

Problem 2.9. You are probably familiar with the hyperbolic functions, now defined as

\[ \cosh(z) = \frac{e^z + e^{-z}}{2}, \quad \sinh(z) = \frac{e^z - e^{-z}}{2}, \quad \tanh(z) = \frac{\sinh(z)}{\cosh(z)}, \quad (2.0.24) \]

for any complex \(z\). Show that

\[ \cosh(iz) = \cos(z), \quad \sinh(iz) = i\sin(z), \cos(iz) = \cosh(z), \quad \sin(iz) = i\sinh(z). \quad (2.0.25) \]
Problem 2.10. Calculate, for real $\theta$ and positive integer $N$:

$$\cos(\theta) + \cos(2\theta) + \cos(3\theta) + \cdots + \cos(N\theta) = ?$$ \hspace{1cm} (2.0.26)

$$\sin(\theta) + \sin(2\theta) + \sin(3\theta) + \cdots + \sin(N\theta) = ?$$ \hspace{1cm} (2.0.27)

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

Problem 2.11. 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.

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

$$z = re^{i\theta + i2\pi n}, \hspace{1cm} n = 0, \pm 1, \pm 2, \pm 3, \ldots \hspace{1cm} (2.0.28)$$

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

$$1^{1/m} = e^{i2\pi n/m}, \hspace{1cm} n = 0, 1, \ldots, m - 1. \hspace{1cm} (2.0.29)$$

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

$$1^{a/b} = e^{i2\pi n(a/b)} \hspace{1cm} \text{for} \hspace{1cm} n = 0, 1, \ldots, b - 1, \hspace{1cm} (2.0.30)$$

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

$$1^{1/\pi} = e^{i2\pi n/\pi} = e^{i2n}, \hspace{1cm} (2.0.31)$$

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 $m$th 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 $n$,

$$(e^{i(\pi + 2\pi n)})^{1/2} = e^{i\pi/2 + i\pi n} = \pm i. \hspace{1cm} n = 0, 1. \hspace{1cm} (2.0.32)$$

Problem 2.12. 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

$$e^w = e^x e^{i(y+2\pi n)}, \hspace{1cm} n = 0, \pm 1, \pm 2, \pm 3, \ldots \hspace{1cm} (2.0.33)$$
We define $\ln$ to be such that

$$\ln e^w = x + i(y + 2\pi n). \quad (2.0.34)$$

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

$$\ln(z) = \ln |z| + i(\arg(z) + 2\pi n). \quad (2.0.35)$$

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 $\ln$:

$$z^w = e^{w \ln z} = e^{(x+iy)(\ln(r)+i(\theta+2\pi n))} = e^{x \ln r - y \theta - y(\theta+2\pi n)} e^{i(y \ln(r)+x(\theta+2\pi n))}. \quad (2.0.36)$$

This is, of course, a multi-valued function. We will have more to say about such multi-valued functions when discussing their calculus in §5.

**Problem 2.13.** Find the inverse hyperbolic functions of eq. (2.0.24) 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.

**Problem 2.14.** Let $\vec{\xi}$ and $\vec{\xi}'$ be vectors in a 2D Euclidean space, i.e., you may assume their Cartesian components are

$$\vec{\xi} = (x, y) = r(\cos \phi, \sin \phi), \quad \vec{\xi}' = (x', y') = r'(\cos \phi', \sin \phi'). \quad (2.0.37)$$

Use complex numbers, and assume that the following complex Taylor expansion of $\ln$ holds

$$\ln(1 - z) = -\sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell}, \quad |z| < 1, \quad (2.0.38)$$

to show that

$$\ln |\vec{\xi} - \vec{\xi}'| = \ln r_> - \sum_{\ell=1}^{\infty} \frac{1}{\ell} \left( \frac{r_<}{r_>} \right)^\ell \cos \left( \ell(\phi - \phi') \right), \quad (2.0.39)$$

where $r_>$ is the larger and $r_<$ is the smaller of the $(r, r')$, 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 $z = re^{i\phi}$ and $z' = r'e^{i\phi'}$; then consider $\ln(z - z')$ – how do you extract $\ln |\vec{\xi} - \vec{\xi}'|$ from it?
3 Matrix Algebra: A Review

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 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 $i$th row and $j$th column of $M$ and that of $N$ can be written as

$$M_{ij}^i \quad \text{and} \quad N_{ij}^j.$$  \hfill (3.1.1)

As an example, if $M$ is a $2 \times 2$ matrix, we have

$$M = \begin{bmatrix} M_{11}^1 & M_{12}^1 \\ M_{21}^2 & M_{22}^2 \end{bmatrix}. \hfill (3.1.2)$$

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 differently. However, it is common to see the notation $M_{ij}$ and $M^{ij}$, etc., too.

A vector $v$ can be written as

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

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

$$(M \cdot N)^i_j = \sum_{k=1}^{D} M_{ik}^i N_{kj}^k \equiv M_{ik}^i N_{kj}^k. \hfill (3.1.4)$$

In words: the $ij$ component of the product $MN$, for a fixed $i$ and fixed $j$, means we are taking the $i$th row of $M$ and “dotting” it into the $j$th 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

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad N = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \hfill (3.1.5)$$

then

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

---

\footnote{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 $\lambda$ goes respectively as

\[(M + N)^i_j = M^i_j + N^i_j \quad (3.1.7)\]

and

\[(\lambda M)^i_j = \lambda M^i_j. \quad (3.1.8)\]

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

\[A^i_k B^k_l C^l_j = (AB)^i_l C^l_j = A^i_k (BC)^l_j = (ABC)^i_j. \quad (3.1.9)\]

**Tr**  The trace of a square matrix $A$, denoted by $\text{Tr}(A)$, is defined as $\sum_i A^i_i$. The index notation makes it clear that $\text{Tr}(AB)$ is that of $BA$ because

\[\text{Tr} [A \cdot B] = A^i_k B^k_l = B^k_l A^i_k = \text{Tr} [B \cdot A]. \quad (3.1.10)\]

This immediately implies the trace of the product of any number of matrices is cyclic, in the sense that

\[\text{Tr} [X_1 \cdot X_2 \cdot \cdots \cdot X_N] = \text{Tr} [X_N \cdot X_1 \cdot X_2 \cdot \cdots \cdot X_{N-1}] = \text{Tr} [X_2 \cdot X_3 \cdot \cdots \cdot X_N \cdot X_1]. \quad (3.1.11)\]

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

\[\text{Tr} [X + Y] = \text{Tr} [X] + \text{Tr} [Y], \quad \text{Tr} [\lambda X] = \lambda \text{Tr} [X]. \quad (3.1.12)\]

Comment on whether it makes sense to define $\text{Tr}(A) = A^i_i$, if $A$ is not a square matrix.

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

\[I^i_j = \delta^i_j = 1, \quad i = j \]

\[= 0, \quad i \neq j \quad (3.1.13)\]

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

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 $A^i_j = a_i \delta^i_j = a_j \delta^j_i$ is the diagonal matrix with $a_1, a_2, \ldots, a_D$ filling its diagonal components, from the upper left to the lower right. Diagonal matrices are also often denoted, for instance, as

\[A = \text{diag}[a_1, \ldots, a_D]. \quad (3.1.14)\]

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

\[(AB)^i_j = \sum_i a_i \delta^i_i b_j \delta^j_i. \quad (3.1.15)\]
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
\[
(AB)^{ij}_j = a_ib_j. \tag{3.1.16}
\]

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

\[
(AB)^{ij}_j = a_ib_j\delta^i_j \quad \text{(No sum over } i, j). \tag{3.1.17}
\]

**Transpose**

The transpose \( T \) of any matrix \( A \) is
\[
(A^T)^{ij}_j = A^j_i. \tag{3.1.18}
\]

In words: the \( i \) row of \( A^T \) is the \( i \)th column of \( A \); the \( j \)th column of \( A^T \) is the \( j \)th 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 \((A \cdot B)^T = B^T A^T\).

**Adjoint**

The adjoint \( \dagger \) of any matrix is given by
\[
(A^\dagger)^{ij}_j = (A^j_i)^* = (A^*)^j_i. \tag{3.1.19}
\]

In other words, \( A^\dagger = (A^T)^* \); to get \( A^\dagger \), you start with \( A \), take its transpose, then take its complex conjugate. An example is,
\[
A = \begin{bmatrix} 1 + i & e^{i\theta} \\ x + iy & \sqrt{10} \end{bmatrix}, \quad 0 \leq \theta < 2\pi, \ x, y \in \mathbb{R} \tag{3.1.20}
\]
\[
A^T = \begin{bmatrix} 1 + i & x + iy \\ e^{i\theta} & \sqrt{10} \end{bmatrix}, \quad A^\dagger = \begin{bmatrix} 1 - i & x - iy \\ e^{-i\theta} & \sqrt{10} \end{bmatrix}. \tag{3.1.21}
\]

**Orthogonal, Unitary, Symmetric, and Hermitian**

A \( D \times D \) matrix \( A \) is

1. **Orthogonal** if \( A^T A = AA^T = I \). The set of real orthogonal matrices implement rotations in a \( D \)-dimensional real (vector) space.

2. **Unitary** if \( A^\dagger A = AA^\dagger = I \). 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 \( A^T = A \); **anti-symmetric** if \( A^T = -A \).

4. **Hermitian** if \( A^\dagger = A \); **anti-hermitian** if \( A^\dagger = -A \).

**Problem 3.3.** Explain why, if \( A \) is an orthogonal matrix, it obeys the equation
\[
A^i_k A^j_l \delta_{ij} = \delta_{kl}. \tag{3.1.22}
\]

Now explain why, if \( A \) is a unitary matrix, it obeys the equation
\[
(A^i_k)^* A^j_l \delta_{ij} = \delta_{kl}. \tag{3.1.23}
\]
Problem 3.4. Prove that \((AB)^T = B^TA^T\) 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):

\[
\begin{bmatrix}
e^{i\theta} & 0 \\
0 & e^{i\delta}
\end{bmatrix}, \quad \begin{bmatrix}
e^{i\theta} & X \\
X & e^{i\delta}
\end{bmatrix}, \quad \begin{bmatrix}
\sqrt{109} & 1 - i \\
1 + i & \theta^\delta
\end{bmatrix}, \quad \theta, \delta \in \mathbb{R}.
\]

(3.1.24)

3.2 Determinants, Linear (In)dependence, Inverses and 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 \(\epsilon_{i_1i_2...i_{D-1}i_D}\):

\[
\det A \equiv \epsilon_{i_1i_2...i_{D-1}i_D}A_{i_1}^{i_{i_2}}A_{i_2}^{i_{i_3}}...A_{i_{D-1}}^{i_{i_D}}.
\]

(3.2.1)

Every index on the Levi-Civita runs from 1 through \(D\). 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

\[
\epsilon_{i_1i_2...i_{a-1}i_ai_{a+1}...i_{b-1}i_bi_{b+1}...i_{D-1}i_D} = -\epsilon_{i_1i_2...i_{a-1}i_{b}i_{a+1}...i_{b-1}i_ai_{b+1}...i_{D-1}i_D},
\]

(3.2.2)

- In matrix algebra and flat Euclidean space, \(\epsilon_{123...D} = \epsilon^{123...D} \equiv 1^5\)

These are sufficient to define every component of the Levi-Civita symbol. Because \(\epsilon\) 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 \(i_1 \ldots i_D\) are distinct indices, \(\epsilon_{i_1i_2...i_{D-1}i_D}\) is really the sign of the permutation (\(\equiv (-)^{\text{number of swaps of index pairs}}\)) that brings \(\{1, 2, \ldots, D-1, D\}\) to \(\{i_1, i_2, \ldots, i_{D-1}, i_D\}\).

Hence, \(\epsilon_{i_1i_2...i_{D-1}i_D}\) is +1 when it takes zero/even number of swaps, and −1 when it takes odd.

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

\[
1 = \epsilon_{12} = -\epsilon_{21}.
\]

(3.2.3)

In the 3 dimensional case,

\[
1 = \epsilon_{123} = -\epsilon_{213} = -\epsilon_{321} = -\epsilon_{132} = \epsilon_{231} = \epsilon_{312}.
\]

(3.2.4)

Properties of the determinant include

\[
\det A^T = \det A, \quad \det(A \cdot B) = \det A \cdot \det B, \quad \det A^{-1} = \frac{1}{\det A}.
\]

(3.2.5)

\(^5\)In Lorentzian flat spacetimes, the Levi-Civita tensor with upper indices will need to be carefully distinguished from its counterpart with lower indices.
for square matrices $A$ and $B$. As a simple example, let us use eq. (3.2.1) to calculate the determinant of

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

Remember the only non-zero components of $\epsilon_{i_1i_2}$ are $\epsilon_{12} = 1$ and $\epsilon_{21} = -1$.

$$ \det A = \epsilon_{12} A^1_1 A^2_2 + \epsilon_{21} A^2_1 A^1_2 = A^1_1 A^2_2 - A^2_1 A^1_2 = ad - bc. $$

(3.2.7)

**Problem 3.5. Inverse of $2 \times 2$ matrix**

By viewing $\epsilon$ as a $2 \times 2$ matrix, prove that, whenever the inverse of a matrix $M$ exist, it can be written as

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

(3.2.8)

Hint: Can you explain why eq. (3.2.1) implies $\epsilon_{AB} M^A_1 M^B_j = \epsilon_{1J} \det M$? (3.2.9)

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

**Problem 3.6.** Explain why eq. (3.2.1) implies

$$ \epsilon_{i_1i_2\ldots i_{D-1}i_D} A^i_{i_1} A^{i_2 j_1} A^{i_3 j_2} \ldots A^{i_{D-1} j_{D-1}} A^{i_D j_D} = \epsilon_{j_1j_2\ldots j_{D-1}j_D} \det A. $$

(3.2.10)

Hint: What happens when you swap $A^{i_m}_{m}$ and $A^{i_n}_{n}$ in eq. (3.2.1)?

**Linear (in)dependence**

Given a set of $D$ vectors $\{v_1, \ldots, v_D\}$, 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,

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

(3.2.11)

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

**Det 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. (3.2.15) 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 (≡ 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 3.7. 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 3.8. Show that, for a $D \times D$ matrix $A$ and some complex number $\lambda$,

$$\det(\lambda A) = \lambda^D \det A.$$  \hspace{1cm} (3.2.12)

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

Problem 3.9. Relation to cofactor expansion

The cofactor expansion definition of the determinant is

$$\det A = \sum_{i=1}^{D} A^i_k C^i_k,$$  \hspace{1cm} (3.2.13)

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 $i$th row and $k$th 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 $3 \times 3$ example, we have

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & l \end{bmatrix} = b(-)^{1+2} \det \begin{bmatrix} d & f \\ g & l \end{bmatrix} + e(-)^{2+2} \det \begin{bmatrix} a & c \\ g & l \end{bmatrix} + h(-)^{3+2} \det \begin{bmatrix} a & c \\ d & f \end{bmatrix}.$$  \hspace{1cm} (3.2.14)

Cramer’s rule

Can you show the equivalence of equations (3.2.1) and (3.2.13)? Can you also show that

$$\delta_{kl} \det A = \sum_{i=1}^{D} A^i_k C^i_l.$$  \hspace{1cm} (3.2.15)

That is, show that when $k \neq l$, the sum on the right hand side is zero. What does eq. (3.2.15) tell us about $(A^{-1})^k_i$?

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

$$\det A = \epsilon_{j_1 \ldots j_D} A^{j_1} \ldots A^{j_D}$$

$$= A^i_k \left( \epsilon_{j_1 \ldots j_{k-1} j_k+1 \ldots j_D} A^{j_1} \ldots A^{j_{k-1}} A^{j_k+1} \ldots A^{j_D} \right),$$

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

Pauli Matrices

The $2 \times 2$ identity together with the Pauli matrices are Hermitian matrices.

$$\sigma^0 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma^1 \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$  \hspace{1cm} (3.2.17)
Problem 3.10. 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,

\[
\det p_\mu \sigma^\mu = \sum_{0 \leq \mu, \nu \leq 3} \eta^{\mu\nu} p_\mu p_\nu \equiv p^2, \tag{3.2.18}
\]

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

\[
\eta^{\mu\nu} \equiv \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{bmatrix}. \tag{3.2.19}
\]

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

\[
\det \sigma^0 = 1, \quad \det \sigma^i = -1, \quad \text{Tr} [\sigma^0] = 2, \quad \text{Tr} [\sigma^i] = 0 \quad \text{Tr} [\sigma^i] = 0 \tag{3.2.20}
\]

\[
\sigma^i \sigma^j = \delta^{ij} \mathbb{I} + i \sum_{1 \leq k \leq 3} \epsilon^{ijk} \sigma^k, \quad \epsilon \sigma^i \epsilon = (\sigma^i)^*. \tag{3.2.21}
\]

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

\[
\theta_i \theta_j \epsilon^{ijk} = 0. \tag{3.2.22}
\]

Use these facts to derive the result:

\[
U(\bar{\theta}) \equiv \exp \left[ -\frac{i}{2} \sum_{j=1}^{3} \theta_j \sigma^j \right] \equiv e^{-(i/2)\bar{\theta} \cdot \sigma} \\
= \cos \left( \frac{1}{2} |\bar{\theta}| \right) - i \frac{\bar{\theta} \cdot \sigma}{|\bar{\theta}|} \sin \left( \frac{1}{2} |\bar{\theta}| \right), \quad |\bar{\theta}| = \sqrt{\theta \cdot \bar{\theta}} \equiv \sqrt{\bar{\theta} \cdot \sigma}, \tag{3.2.23}
\]

which is valid for complex \( \{\theta_i\} \). (Hint: Taylor expand \( \exp X = \sum_{\ell=0}^{\infty} X^\ell / \ell! \), followed by applying the first relation in eq. (3.2.21).)

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

\[
p_\mu \sigma^\mu = A. \tag{3.2.24}
\]

Inverse

The inverse of the \( D \times D \) matrix \( A \) is defined to be

\[
A^{-1} A = A A^{-1} = \mathbb{I}. \tag{3.2.25}
\]

The inverse \( A^{-1} \) of a finite dimensional matrix \( A \) is unique; moreover, the left \( A^{-1} A = \mathbb{I} \) and right inverses \( A A^{-1} = \mathbb{I} \) are the same object. The inverse exists if and only if (\( \equiv \text{iff} \)) \( \det A \neq 0 \).
Problem 3.11. How does eq. (3.2.15) allow us to write down the inverse matrix \((A^{-1})^i_k\)?

Problem 3.12. Why are the left and right inverses of (an invertible) matrix \(A\) the same? Hint: Consider \(LA = I\) and \(AR = I\); for the first, multiply \(R\) on both sides from the right.

Problem 3.13. Prove that \((A^{-1})^T = (A^T)^{-1}\) and \((A^{-1})^\dagger = (A^\dagger)^{-1}\).

Eigenvectors and Eigenvalues If \(A\) is a \(D \times D\) matrix, \(v\) is its \((D\text{-component})\) eigenvector with eigenvalue \(\lambda\) if it obeys

\[ A^i_j v^j = \lambda v^i. \]  (3.2.26)

This means

\[ (A^i_j - \lambda \delta^i_j) v^j = 0 \]  (3.2.27)

has non-trivial solutions iff

\[ P_D(\lambda) \equiv \det (A - \lambda I) = 0. \]  (3.2.28)

Equation (3.2.28) is known as the characteristic equation. For a \(D \times D\) matrix, it gives us a \(D\)th degree polynomial \(P_D(\lambda)\) for \(\lambda\), whose roots are the eigenvalues of the matrix \(\lambda\) – the set of all eigenvalues of a matrix is called its spectrum. For each solution for \(\lambda\), we then proceed to solve for the \(v^i\) in eq. (3.2.27). 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 \(A - \lambda I\) are necessarily linearly dependent. As already discussed above, this amounts to the statement that there is some sum of multiples of these columns (≡ “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 \(A - \lambda I\),

\[ A - \lambda I \equiv [w_1 w_2 \ldots w_D] \Rightarrow (A - \lambda I)v = \sum_j w_j v^j = 0. \]  (3.2.29)

(Note that, if \(\sum_j w_j v^j = 0\) then \(\sum_j w_j (Kv^j) = 0\) 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 \(D\)th order polynomial equation (3.2.28); 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. (3.2.28) as \(\sum_{i=0}^D q_i \lambda^i = 0\) (for appropriate complex constants \(\{q_i\}\)), then replace \(\lambda^i \rightarrow A^i\) (namely, the \(i\)th power of \(\lambda\) with the \(i\)th power of \(A\)), we would find

\[ P_D(A) = 0. \]  (3.2.30)

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:

\[ U^\dagger AU = \text{diag}(\lambda_1, \ldots, \lambda_D) + N, \]  (3.2.31)
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.

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

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

(3.2.32)

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 $v^\dagger v = (v^\dagger v)^\dagger \delta_{ij} = 1$.) In index notation,

$$A^i_j U^j_k = \lambda_k U^i_k = U^i_l \delta^j_k \lambda_k, \quad (\text{No sum over } k).$$

(3.2.33)

In matrix notation,

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

(3.2.34)

Here, $U^i_k$ for fixed $k$, is the $k$th eigenvector, and $\lambda_k$ is the corresponding eigenvalue. By multiplying both sides with $U^\dagger$, we have

$$U^\dagger AU = D, \quad D^i_l \equiv \lambda_l \delta^j_l \quad (\text{No sum over } l).$$

(3.2.35)

Equivalently,

$$A = UDU^\dagger.$$  

(3.2.36)

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 $(A - \lambda I) \cdot v$, we are really solving for the null space of the matrix $M \equiv A - \lambda I$, because for a fixed eigenvalue $\lambda$, there could be more than one solution – that’s what we mean by degeneracy.

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

$$UAU^\dagger = \text{diag}(\lambda_1, \ldots, \lambda_D) + N,$$ 

(3.2.37)

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,

$$(UAU^\dagger)^\dagger = UAU^\dagger \quad \Rightarrow \quad \text{diag}(\lambda_1^*, \ldots, \lambda_D^*) + N^\dagger = \text{diag}(\lambda_1, \ldots, \lambda_D) + N.$$  

(3.2.38)

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
\[
UAU^\dagger = M, \tag{3.2.39}
\]
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
\[
(UAU^\dagger)^\dagger (UAU^\dagger) = U^\dagger A^\dagger UAU^\dagger = UAU^\dagger = UU^\dagger = I. \tag{3.2.40}
\]
That means
\[
M^\dagger M = I \Rightarrow (M^\dagger M)^k_i = (M^\dagger)^k_s M^*_i = \sum_s M^*_i M^k_s = \delta_{ij} M^* M^k_l = \delta_{kl}, \tag{3.2.41}
\]
where we have recalled eq. (3.1.23) in the last equality. If \( w_i \) denotes the \( i \)th 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., \( w_1^\dagger w_1 = 1 \Rightarrow |\lambda_1|^2 = 1 \). That
\[
M^\dagger w = \lambda w \Rightarrow \lambda \neq 0 \Rightarrow M^\dagger w = 0 \Rightarrow M^\dagger w = 0 \Rightarrow M^\dagger = 0.
\]

Diagonalization example

As an example, let’s diagonalize \( \sigma^2 \) from eq. (3.2.17).
\[
P_2(\lambda) = \det \begin{bmatrix} -\lambda & -i \\ i & -\lambda \end{bmatrix} = \lambda^2 - 1 = 0 \tag{3.2.42}
\]
(We can even check Caley-Hamilton here: \( P_2(\sigma^2) = (\sigma^2)^2 - I = I - I = 0 \); see eq. (3.2.21).) The solutions are \( \lambda = \pm 1 \) and
\[
\begin{bmatrix} \mp 1 & -i \\ i & \mp 1 \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow v^1 = \mp iv^2. \tag{3.2.43}
\]
The subscripts on \( v \) refer to their eigenvalues, namely
\[
\sigma^2 v^\pm = \pm v^\pm. \tag{3.2.44}
\]
By choosing \( v^2 = 1/\sqrt{2} \), we can check \((v^\pm)^* v^\pm \delta_{ij} = 1\) and therefore the normalized eigenvectors are
\[
v^\pm = \frac{1}{\sqrt{2}} \begin{bmatrix} \mp i \\ 1 \end{bmatrix}. \tag{3.2.45}
\]
Furthermore you can check directly that eq. (3.2.44) is satisfied. We therefore have
\[
\left( \frac{1}{\sqrt{2}} \begin{bmatrix} i & 1 \\ -i & 1 \end{bmatrix} \right) \sigma^2 \left( \frac{1}{\sqrt{2}} \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.2.46}
\]

An example of a matrix that cannot be diagonalized is
\[
A \equiv \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}. \tag{3.2.47}
\]

The characteristic equation is \( \lambda^2 = 0 \), so both eigenvalues are zero. Therefore \( A - \lambda I = A \), and
\[
\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow v^1 = 0, \ v^2 \text{ arbitrary.} \tag{3.2.48}
\]

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 \(2 \times 2\) matrix \(U\) whose columns are distinct unit length eigenvectors of \(\sigma^2\).

**Problem 3.14.** Show how to go from eq. (3.2.33) to eq. (3.2.35) using index notation. \( \square \)

**Problem 3.15.** Use the Schur decomposition to explain why, for any matrix \(A\), \(\text{Tr} [A]\) is equal to the sum of its eigenvalues and \(\det A\) is equal to their product:
\[
\text{Tr} [A] = \sum_{l=1}^{D} \lambda_l, \quad \det A = \prod_{l=1}^{D} \lambda_l. \tag{3.2.49}
\]

Hint: for \(\det A\), the key question is how to take the determinant of an upper triangular matrix. \( \square \)

**Problem 3.16.** For a strictly upper triangular matrix \(N\), prove that \(N\) multiplied to itself any number of times still returns a strictly upper triangular matrix. Can a strictly upper triangular matrix be diagonalized? (Explain.) \( \square \)

**Problem 3.17.** 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 \(f(A) = U f(X) U^\dagger\). Hint: Can you explain why \((UBU^\dagger)^\ell = UB^\ell U^\dagger\), for \(B\) some arbitrary matrix, \(U\) unitary, and \(\ell = 1, 2, 3, \ldots\)? \( \square \)

**Problem 3.18.** 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 3.19.** *Simplified example of neutrino oscillations.* We begin with the observation that the solution to the first order equation
\[
i\partial_t \psi(t) = E \psi(t), \tag{3.2.50}
\]
for \(E\) some real constant, is
\[
\psi(t) = e^{-iEt} \psi_0. \tag{3.2.51}
\]
The $\psi_0$ is some arbitrary (possibly complex) constant, corresponding to the initial condition $\psi(t = 0)$. Now solve the matrix differential equation

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

with the initial condition – describing the production of $\nu_1$-type of neutrino, say –

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

where the Hamiltonian $H$ is

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

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

The $p$ is the magnitude of the momentum, $m_{1,2}$ are masses, and $\theta$ is the “mixing angle”. Then calculate

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

Express $P_{1\rightarrow 1}$ and $P_{1\rightarrow 2}$ 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 $\nu_1$ and $\nu_2$ at time $t > 0$, given $\nu_1$ was produced at $t = 0$.) Hint: Start by diagonalizing $M = U^T A U$ where

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

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

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

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.

### 3.3 Special Topic 1: 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 $2 \times 2$ real orthogonal matrix. We will begin by writing its components as follows

$$A \equiv \begin{bmatrix} v^1 & v^2 \\ w^1 & w^2 \end{bmatrix}. $$
(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 $AA^T = I$.

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

$$v^i = r (\cos \phi, \sin \phi), \quad r \geq 0, \phi \in [0, 2\pi). \quad (3.3.3)$$

But $v^2 = 1$ means $r = 1$. Similarly,

$$w^i = (\cos \phi', \sin \phi'), \quad \phi' \in [0, 2\pi). \quad (3.3.4)$$

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

$$\mathbf{v} \cdot \mathbf{w} = \cos \phi \cdot \cos \phi' + \sin \phi \cdot \sin \phi' = \cos (\phi - \phi') = 0. \quad (3.3.5)$$

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

$$w^i = (\cos (\phi \pm \pi/2), \sin (\phi \pm \pi/2)) = (\mp \sin (\phi), \pm \cos (\phi)). \quad (3.3.6)$$

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 $\phi$ there are two distinct solutions.

$$R_1(\phi) = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}, \quad R_2(\phi) = \begin{bmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{bmatrix}. \quad (3.3.7)$$

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

$$R_2(\phi) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot R_1(\phi). \quad (3.3.8)$$

In other words, the $R_2(\phi = 0)$ in eq. (3.3.7) corresponds to a “parity flip” where the vector is reflected about the x-axis.

**Problem 3.20.** What about the matrix that reflects 2D vectors about the $y$-axis? What value of $\theta$ 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 $\theta_0$ such that $R_1(\theta_0) = R_2(\theta_0)$. Also verify that

$$R_1(\phi)R_1(\phi') = R_1(\phi + \phi'). \quad (3.3.9)$$

This makes geometric sense: rotating a vector clockwise by $\phi$ then by $\phi'$ should be the same as rotation by $\phi + \phi'$. Mathematically speaking, this composition law in eq. (3.3.9) tells us rotations form the $SO_2$ group. The set of $D \times D$ real orthogonal matrices obeying $R^T R = I$, 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).
Problem 3.21. $2 \times 2$ Unitary Matrices. Can you construct the most general $2 \times 2$ unitary matrix? First argue that the most general complex 2D vector $\vec{v}$ that satisfies $\vec{v}^\dagger \vec{v} = 1$ is

$$v^i = e^{i\phi_1} (\cos \theta, e^{i\phi_2} \sin \theta), \quad \phi_{1,2}, \theta \in [0, 2\pi). \quad (3.3.10)$$

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

$$w^i = e^{i\phi_1'} (\cos \theta', e^{i\phi_2'} \sin \theta'), \quad \phi_{1,2}', \theta' \in [0, 2\pi). \quad (3.3.11)$$

You should arrive at

$$\sin(\theta) \sin(\theta') e^{i(\phi_2' - \phi_2)} + \cos(\theta) \cos(\theta') = 0. \quad (3.3.12)$$

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

$$\phi_2' = \phi_2, \quad \theta = \theta' \pm \frac{\pi}{2}. \quad (3.3.13)$$

or

$$\phi_2' = \phi_2 + \pi, \quad \theta = -\theta' \pm \frac{\pi}{2}. \quad (3.3.14)$$

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

$$U = \begin{bmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{bmatrix} \cdot O(\theta) \cdot \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi_3} \end{bmatrix}. \quad (3.3.15)$$

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 $2 \times 2$ matrix – and therefore we should have 4 free parameters left.

**Bonus problem:** By imposing $\det U = 1$, can you connect eq. (3.3.15) to eq. (3.2.23)?
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 $D \in \{1, 2, 3, \ldots \}$ 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 $\alpha$ is just some arbitrary label, but later on it will often correspond to the eigenvalue of some linear operator. We may also use $\alpha$ as an enumeration label, where $|\alpha\rangle$ is the $\alpha$th 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 $\{|\alpha\rangle, |\beta\rangle, |\gamma\rangle, \ldots \}$ 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 $(-1)\,|\alpha\rangle$.

\[ |\alpha\rangle + (-|\alpha\rangle) = |\text{zero}\rangle. \quad (4.1.6) \]

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

\[ c\,|\alpha\rangle = |\gamma\rangle. \quad (4.1.7) \]

(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,

\[ a(|\alpha\rangle + |\beta\rangle) = a\,|\alpha\rangle + a\,|\beta\rangle \]
\[ (a + b)\,|\alpha\rangle = a\,|\alpha\rangle + b\,|\alpha\rangle. \quad (4.1.8) \]

**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

\[ |\vec{a}\rangle \equiv (a^1, a^2, \ldots, a^D), \quad (4.1.10) \]

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

\[ |\vec{v}\rangle = (v^1, v^2, \ldots, v^D), \]
\[ |\vec{w}\rangle = (w^1, w^2, \ldots, w^D) \quad \text{and} \]
\[ |\vec{x}\rangle = (x^1, x^2, \ldots, x^D) \quad (4.1.11) \]

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

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

\[ |\vec{v}\rangle + |\vec{w}\rangle = (v^1 + w^1, \ldots, v^D + w^D) \equiv |\vec{v} + \vec{w}\rangle. \quad (4.1.13) \]

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

\[ |\vec{v}\rangle + |\vec{w}\rangle = |\vec{v} + \vec{w}\rangle = (v^1 + w^1, \ldots, v^D + w^D) = (w^1 + v^1, \ldots, w^D + v^D) = |\vec{w}\rangle + |\vec{v}\rangle = |\vec{w} + \vec{v}\rangle, \quad (4.1.14) \]
\[ |\vec{v}\rangle + (|\vec{w}\rangle + |\vec{x}\rangle) = (v^1 + (w^1 + x^1), \ldots, v^D + (w^D + x^D)) = ((v^1 + w^1) + x^1, \ldots, (v^D + w^D) + x^D) \]
\[ = |\vec{v}\rangle + (|\vec{w}\rangle + |\vec{x}\rangle) = (|\vec{v}\rangle + |\vec{w}\rangle) + |\vec{x}\rangle = |\vec{v} + \vec{w} + \vec{x}\rangle. \quad (4.1.15) \]
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.,
\[
0 |\vec{v}\rangle = 0(v^1, \ldots, v^D) = (0, \ldots, 0) = |\text{zero}\rangle
\] (4.1.16)
– that acts as an additive identity. Namely, adding \(|\text{zero}\rangle\) to any vector returns the vector itself:
\[
|\text{zero}\rangle + |\vec{w}\rangle = (0, \ldots, 0) + (w^1, \ldots, w^D) = |\vec{w}\rangle.
\] (4.1.17)
For any vector \(|\vec{x}\rangle\) there exists an additive inverse; in fact, the inverse of \(|\vec{x}\rangle\) is just \((-1)|\vec{x}\rangle = |\vec{-x}\rangle\).
\[
|\vec{x}\rangle + (-|\vec{x}\rangle) = (x^1, \ldots, x^D) - (x^1, \ldots, x^D) = |\text{zero}\rangle.
\] (4.1.18)

3. Multiplication by scalar Any ket can be multiplied by an arbitrary real number \(c\) to yield another vector
\[
c|\vec{v}\rangle = c(v^1, \ldots, v^D) = (cv^1, \ldots, cv^D) \equiv |c\vec{v}\rangle.
\] (4.1.19)
This multiplication is distributive with respect to both vector and scalar addition; if \(a\) and \(b\) are arbitrary real numbers,
\[
a(|\vec{v}\rangle + |\vec{w}\rangle) = (av^1 + aw^1, av^2 + aw^2, \ldots, av^D + aw^D)
= |a\vec{v}\rangle + |a\vec{w}\rangle = a |\vec{v}\rangle + a |\vec{w}\rangle,
\] (4.1.20)
\[
(a + b) |\vec{x}\rangle = (ax^1 + bx^1, \ldots, ax^D + bx^D)
= |a\vec{x}\rangle + |b\vec{x}\rangle = a |\vec{x}\rangle + b |\vec{x}\rangle.
\] (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 \(\int_{\mathbb{R}^D} \, d^D \vec{x} |f(\vec{x})|^2 < \infty\).

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,

\[
|N\rangle = \sum_{i=1}^{N-1} c^i |i\rangle ,
\]

(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:

\[
\sum_{i=1}^{N} c^i |i\rangle = |\text{zero}\rangle , \quad \exists \{\chi^i\}.
\]

(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 \( \{|1\rangle, |2\rangle, |3\rangle, \ldots, |D\rangle\} \) 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 linear combination (aka superposition) of these vectors

\[
|\alpha\rangle = \sum_{i=1}^{D} c^i |i\rangle , \quad \{\chi^i \in \mathbb{C}\}.
\]

(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 \( \{|i\rangle | i = 1, 2, \ldots, D\} \) is a complete set of basis vectors; and such a set of (basis) vectors is said to span the vector space. 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 \( |\vec{a}\rangle \equiv (a^1, \ldots, a^D) \) from the real vector space of \( \mathbb{R}^D \), we may choose

\[
|1\rangle = (1, 0, 0, \ldots), \quad |2\rangle = (0, 1, 0, 0, \ldots), \quad |3\rangle = (0, 0, 1, 0, 0, \ldots), \ldots \quad |D\rangle = (0, 0, \ldots, 0, 0, 1).
\]

(4.1.25)

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

\[
|\vec{a}\rangle = (a^1, \ldots, a^D) = \sum_{i=1}^{D} a^i |i\rangle .
\]

(4.1.26)

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

\[
|\vec{e}\rangle \equiv (1, 1, 0, \ldots, 0) ,
\]

(4.1.27)

\[
|\vec{w}\rangle \equiv (1, -1, 0, \ldots, 0) ,
\]

(4.1.28)

\[
|\vec{u}\rangle \equiv (1, 0, 0, \ldots, 0).
\]

(4.1.29)

\footnote{The span of vectors \( \{|1\rangle, \ldots, |D\rangle\} \) 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 \( P_n(x) = c_0 + c_1x + \cdots + c_nx^n \), where the \( \{c_i|i = 1, 2, \ldots, n\} \) 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 Axioms 1 through Axiom 3 above. But a brief glance tells us, the axioms in Axiom 1 (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 Axiom 2 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, \( a^3 = \text{same constant for all vectors} \). 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 \( \{ \hat{H}_1, \hat{H}_2, \ldots \} \) with real coefficients yield another Hermitian matrix
\[
\left( c_1 \hat{H}_1 + c_2 \hat{H}_2 \right)^\dagger = c_1 \hat{H}_1^\dagger + c_2 \hat{H}_2^\dagger, \quad c_{1,2} \in \mathbb{R}; \tag{4.1.31}
\]
whereas the superposition of (anti)symmetric ones with complex coefficients return another (anti)symmetric matrix:
\[
\left( c_1 \hat{H}_1 + c_2 \hat{H}_2 \right)^T = c_1 \hat{H}_1 + c_2 \hat{H}_2, \quad c_{1,2} \in \mathbb{C}, \hat{H}_1^T = \hat{H}_1, \quad \hat{H}_2^T = \hat{H}_2, \tag{4.1.32}
\]
\[
\left( c_1 \hat{H}_1 + c_2 \hat{H}_2 \right)^T = -(c_1 \hat{H}_1 + c_2 \hat{H}_2), \quad c_{1,2} \in \mathbb{C}, \hat{H}_1^T = -\hat{H}_1, \quad \hat{H}_2^T = -\hat{H}_2. \tag{4.1.33}
\]

**4.2 Inner Products**

In Euclidean \( D \)-space \( \mathbb{R}^D \) the ordinary dot product, between the real vectors \( |\vec{a}\rangle \equiv (a^1, \ldots, a^D) \) and \( |\vec{b}\rangle \equiv (b^1, \ldots, b^D) \), 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 $|\alpha\rangle^\dagger \equiv \langle \alpha |$. (4.2.2)

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

$$(c |\alpha\rangle)^\dagger \equiv c^* \langle \alpha |.$$ (4.2.3)

Moreover, for complex numbers $a$ and $b$,

$$(a |\alpha\rangle + b |\beta\rangle)^\dagger \equiv a^* \langle \alpha | + b^* \langle \beta |.$$ (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:

$$\langle \alpha | \beta \rangle \in \mathbb{C}.$$ (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 $c$,

$$\langle \alpha | (|\beta\rangle + |\gamma\rangle) = \langle \alpha | \beta \rangle + \langle \alpha | \gamma \rangle$$ (4.2.6)

$$\langle \alpha | (c |\beta\rangle) = c \langle \alpha | \beta \rangle$$ (4.2.7)

$$\langle \alpha | \beta \rangle^* = \langle \beta | \alpha \rangle = \langle \alpha | \beta \rangle$$ (4.2.8)

$$\langle \alpha | \alpha \rangle \geq 0$$ (4.2.9)

and

$$\langle \alpha | \alpha \rangle = 0$$ (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 $(\alpha, \beta)$, where $\alpha$ and $\beta$ are vectors. Then the defining properties of the inner product would read instead

$$(\alpha, b\beta + c\gamma) = b(\alpha, \beta) + c(\alpha, \gamma),$$ (4.2.11)

(for any constants $b$ and $c$),

$$(\alpha, \beta)^* = (\beta, \alpha),$$ (4.2.12)

$$(\alpha, \alpha) \geq 0;$$ (4.2.13)
and

\[(\alpha, \alpha) = 0 \quad (4.2.14)\]

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

\[(b\beta + c\gamma, \alpha) = b^*(\beta, \alpha) + c^*(\gamma, \alpha). \quad (4.2.15)\]

**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

\[
|\vec{a}\rangle = (a_1, a_2, \ldots, a_D), \quad (4.2.16)
\]

\[
|\vec{b}\rangle = (b_1, b_2, \ldots, b_D), \quad (4.2.17)
\]

\[
|\vec{c}\rangle = (c_1, c_2, \ldots, c_D); \quad (4.2.18)
\]

where all the components \(a_i, b_i, \ldots\) are now real. Next, define

\[
\langle \vec{a} | \vec{b} \rangle = \vec{a} \cdot \vec{b}. \quad (4.2.19)
\]

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, \(\langle \vec{a} | (c|\vec{b}\rangle) = \langle \vec{a} | c\vec{b} \rangle = c\langle \vec{a} | \vec{b} \rangle = c \langle \vec{a} | \vec{b} \rangle\). Third, \(\langle \vec{a} | \vec{b} \rangle = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{a} = (\vec{b} \cdot \vec{a})^* = \langle \vec{b} | \vec{a} \rangle\).

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 \(\vec{a} = 0\) 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 \(|\alpha\rangle \equiv (\alpha^1, \ldots, \alpha^D)\) and \(|\beta\rangle \equiv (\beta^1, \ldots, \beta^D);\) and define the inner product to be

\[
\langle \alpha | \beta \rangle \equiv \sum_{i=1}^D (\alpha^i)^* \beta^i = \delta_{ij} (\alpha^i)^* \beta^j = \alpha^\dagger \beta. \quad (4.2.20)
\]

- Consider the space of \(D \times D\) complex matrices. Consider two such matrices \(X\) and \(Y\) and define their inner product to be

\[
\langle X | Y \rangle \equiv \text{Tr} \left[ X^\dagger Y \right]. \quad (4.2.21)
\]

Here, \(\text{Tr}\) means the matrix trace and \(X^\dagger\) is the adjoint of the matrix \(X\).

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

\[
\langle f | g \rangle \equiv \int_{-1}^1 dx f(x)^* g(x) \quad (4.2.22)
\]

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:
\[
\langle \alpha | \alpha \rangle \langle \beta | \beta \rangle \geq |\langle \alpha | \beta \rangle|^2.
\] (4.2.23)
The analogy in Euclidean space is \( |\vec{x}|^2 |\vec{y}|^2 \geq |\vec{x} \cdot \vec{y}|^2 \). Hint: Start with
\[
(\langle \alpha | + c^* \langle \beta |) (|\alpha \rangle + c |\beta \rangle) \geq 0.
\] (4.2.24)
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:
\[
\langle \alpha | \beta \rangle = 0 = \langle \beta | \alpha \rangle.
\] (4.2.25)
We also call \( \sqrt{\langle \alpha | \alpha \rangle} \) 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,
\[
|\tilde{\alpha} \rangle \equiv \frac{|\alpha \rangle}{\sqrt{\langle \alpha | \alpha \rangle}} \Rightarrow \langle \tilde{\alpha} | \tilde{\alpha} \rangle = 1.
\] (4.2.26)
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,
\[
\langle i | j \rangle = \delta^i_j.
\] (4.2.27)
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.

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 Let \( \{ |1 \rangle, |2 \rangle, \ldots, |N \rangle \} \) 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 \( \{ |i \rangle \} \).
\[
|\tilde{\alpha} \rangle \equiv |\alpha \rangle - \sum_{j=1}^N |j \rangle \langle j | \alpha \rangle
\] (4.2.28)
This result is key to the following Gram-Schmidt process.
Gram-Schmidt Let \( \{ \alpha_1, \alpha_2, \ldots, \alpha_D \} \) 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. (4.2.28), 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:

\[
|\tilde{\alpha}_1\rangle = \frac{|\alpha_1\rangle}{\sqrt{\langle \alpha_1 | \alpha_1 \rangle}}. \tag{4.2.29}
\]

2. Take the second vector \( |\alpha_2\rangle \) and project out \( |\tilde{\alpha}_1\rangle \):

\[
|\alpha'_2\rangle \equiv |\alpha_2\rangle - |\tilde{\alpha}_1\rangle \langle \tilde{\alpha}_1 | \alpha_2 \rangle, \tag{4.2.30}
\]

and normalize it to unit length

\[
|\tilde{\alpha}_2\rangle \equiv \frac{|\alpha'_2\rangle}{\sqrt{\langle \alpha'_2 | \alpha'_2 \rangle}}. \tag{4.2.31}
\]

3. Take the third vector \( |\alpha_3\rangle \) and project out \( |\tilde{\alpha}_1\rangle \) and \( |\tilde{\alpha}_2\rangle \):

\[
|\alpha'_3\rangle \equiv |\alpha_3\rangle - |\tilde{\alpha}_1\rangle \langle \tilde{\alpha}_1 | \alpha_3 \rangle - |\tilde{\alpha}_2\rangle \langle \tilde{\alpha}_2 | \alpha_3 \rangle, \tag{4.2.32}
\]

then normalize it to unit length

\[
|\tilde{\alpha}_3\rangle \equiv \frac{|\alpha'_3\rangle}{\sqrt{\langle \alpha'_3 | \alpha'_3 \rangle}}. \tag{4.2.33}
\]

4. Repeat ... Take the \( i \)th vector \( |\alpha_i\rangle \) and project out \( |\tilde{\alpha}_1\rangle \) through \( |\tilde{\alpha}_{i-1}\rangle \):

\[
|\alpha'_i\rangle \equiv |\alpha_i\rangle - \sum_{j=1}^{i-1} |\tilde{\alpha}_j\rangle \langle \tilde{\alpha}_j | \alpha_i \rangle, \tag{4.2.34}
\]

then normalize it to unit length

\[
|\tilde{\alpha}_i\rangle \equiv \frac{|\alpha'_i\rangle}{\sqrt{\langle \alpha'_i | \alpha'_i \rangle}}. \tag{4.2.35}
\]

By construction, \( |\tilde{\alpha}_i\rangle \) will be orthogonal to \( |\tilde{\alpha}_1\rangle \) through \( |\tilde{\alpha}_{i-1}\rangle \). 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

\[
|\alpha_1\rangle \doteq (2, 0, 0), \quad |\alpha_2\rangle \doteq (1, 1, 1), \quad |\alpha_3\rangle \doteq (1, 0, 1). \tag{4.2.36}
\]
You can check that these vectors are linearly independent by taking the determinant of the $3 \times 3$ 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,

$$|\tilde{\alpha}_1\rangle = \frac{|\alpha_1\rangle}{\sqrt{\langle \alpha_1 | \alpha_1 \rangle}} = \frac{(2,0,0)}{2} = (1,0,0).$$  \hspace{1cm} (4.2.37)

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

$$|\alpha'_2\rangle = |\alpha_2\rangle - |\tilde{\alpha}_1\rangle \langle \tilde{\alpha}_1 | \alpha_2 \rangle = (1,1,1) - (1,0,0)(1 + 0 + 0) = (0,1,1).$$  \hspace{1cm} (4.2.38)

Then we normalize it to unit length.

$$|\tilde{\alpha}_2\rangle = \frac{|\alpha'_2\rangle}{\sqrt{\langle \alpha'_2 | \alpha'_2 \rangle}} = \frac{(0,1,1)}{\sqrt{2}}.$$  \hspace{1cm} (4.2.39)

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

$$|\alpha'_3\rangle = |\alpha_3\rangle - |\tilde{\alpha}_1\rangle \langle \tilde{\alpha}_1 | \alpha_3 \rangle - |\tilde{\alpha}_2\rangle \langle \tilde{\alpha}_2 | \alpha_3 \rangle$$

$$= (1,0,1) - (1,0,0)(1 + 0 + 0) - \frac{(0,1,1)0 + 0 + 1}{\sqrt{2}} \frac{1}{\sqrt{2}}$$

$$= (1,0,1) - (1,0,0) - \frac{(0,1,1)}{2} = \left(0, -\frac{1}{2}, \frac{1}{2}\right).$$  \hspace{1cm} (4.2.40)

Then we normalize it to unit length.

$$|\tilde{\alpha}_3\rangle = \frac{|\alpha'_3\rangle}{\sqrt{\langle \alpha'_3 | \alpha'_3 \rangle}} = \frac{(0,-1,1)}{\sqrt{2}}.$$  \hspace{1cm} (4.2.41)

You can check that

$$|\tilde{\alpha}_1\rangle = (1,0,0), \quad |\tilde{\alpha}_2\rangle = \frac{(0,1,1)}{\sqrt{2}}, \quad |\tilde{\alpha}_3\rangle = \frac{(0,-1,1)}{\sqrt{2}}.$$  \hspace{1cm} (4.2.42)

are mutually perpendicular and of unit length.

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

$$\langle f | g \rangle \equiv \int_{-1}^{+1} dx f(x)^* g(x).$$  \hspace{1cm} (4.2.43)

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

$$P_\ell(x) \equiv \frac{1}{2^\ell \ell!} \frac{d^\ell}{dx^\ell} (x^2 - 1)^\ell, \quad \ell = 0, 1, 2.$$  \hspace{1cm} (4.2.44)
Orthogonality and Linear independence. We close this subsection with an observation. If a set of non-zero kets \( \{|i\rangle | i = 1, 2, \ldots, N - 1, N \} \) 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
\[
\sum_{i=1}^{N} C^i |i\rangle = 0. \tag{4.2.45}
\]
If we now act \( \langle j | \) on this equation, for any \( j \in \{1, 2, 3, \ldots, N\} \),
\[
\sum_{i=1}^{N} C^i \langle j | i \rangle = \sum_{i=1}^{N} C^i \delta_{ij} \langle j | j \rangle = C^j \langle j | j \rangle = 0. \tag{4.2.46}
\]
That means all the \( \{C^j | j = 1, 2, \ldots, N\} \) 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. (3.2.17). The vector space of \( 2 \times 2 \) complex matrices is 4-dimensional, since there are 4 independent components. Moreover, we have already seen that the trace \( \text{Tr} [X^{\dagger}Y] \) is one way to define an inner product of matrices \( X \) and \( Y \). Since
\[
\frac{1}{2} \text{Tr} \left[ (\sigma^\mu)^\dagger \sigma^\nu \right] = \frac{1}{2} \text{Tr} [\sigma^\mu \sigma^\nu] = \delta^{\mu\nu}, \quad \mu, \nu \in \{0, 1, 2, 3\}, \tag{4.2.47}
\]
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 \( 2 \times 2 \) matrices. That means it must be possible to choose \( \{p_\mu\} \) such that the superposition \( p_\mu \sigma^\mu \) is equal to any given \( 2 \times 2 \) complex matrix \( A \). In fact,
\[
p_\mu \sigma^\mu = A \quad \Leftrightarrow \quad p_\mu = \frac{1}{2} \text{Tr} [\sigma^\mu A]. \tag{4.2.48}
\]
In quantum mechanics and quantum field theory, these \( \{\sigma^\mu\} \) are fundamental to describing spin\(-1/2\) systems.

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? Let \( A \) be one. Firstly, when it acts from the left on a vector, it returns another vector
\[
A |\alpha\rangle = |\alpha'\rangle. \tag{4.3.1}
\]
In other words, if you can tell me what you want the ‘output’ \( |\alpha'\rangle \) to be, 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 \( d \),
\[
(cA + dB) |\alpha\rangle = c A |\alpha\rangle + d B |\alpha\rangle \tag{4.3.2}
\]
\[ A(c|\alpha\rangle + d|\beta\rangle) = c \ A|\alpha\rangle + d \ A|\beta\rangle. \]

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

\[ XYZ = (XY)Z = X(YZ). \quad (4.3.3) \]

**Identity**  
The identity operator obeys

\[ \mathbb{I}|\gamma\rangle = |\gamma\rangle \quad \text{for all } |\gamma\rangle. \quad (4.3.4) \]

**Inverse**  
The inverse of the operator \( X \) is still defined as one that obeys

\[ X^{-1}X = XX^{-1} = \mathbb{I}. \quad (4.3.5) \]

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,

\[ \langle \alpha|A = \langle \alpha'|. \quad (4.3.6) \]

The reason is that a bra is something that acts linearly on an arbitrary vector and returns a complex number. Since that is what \( \langle \alpha|A \) 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 \( ^\dagger \) of the ket \( X^\dagger |\alpha\rangle \) in the following way:

\[ (X^\dagger |\alpha\rangle)^\dagger = \langle \alpha|X. \quad (4.3.7) \]

If \( |\alpha\rangle \) and \( |\beta\rangle \) are arbitrary states,

\[ \langle \beta|X|\alpha\rangle = \langle \beta|(X|\alpha\rangle) = (X^\dagger |\beta\rangle)^\dagger |\alpha\rangle. \quad (4.3.8) \]

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

\[ \langle \alpha|X|\beta\rangle = \langle \beta|X^\dagger|\alpha\rangle. \quad (4.3.9) \]

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

\[ (X^\dagger)^\dagger = X, \quad (4.3.10) \]

so we could also have began with the definition \( (X|\alpha\rangle)^\dagger = \langle \alpha|X^\dagger \).

In the math literature, where \( \alpha \) and \( \beta \) denote the states and \( X \) is still some linear operator, the latter’s adjoint is expressed through the inner product as

\[ \langle \beta,X\alpha \rangle = (X^\dagger \beta,\alpha). \quad (4.3.11) \]
**Problem 4.9.** Prove that

\[(XY)^\dagger = Y^\dagger X^\dagger. \tag{4.3.12}\]

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

**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 \(a\):

\[X |\alpha\rangle = a |\alpha\rangle. \tag{4.3.13}\]

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:

\[X (z |\alpha\rangle) = a (z |\alpha\rangle). \tag{4.3.14}\]

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.

\[X (e^{i\theta} |\alpha\rangle) = a (e^{i\theta} |\alpha\rangle), \quad \theta \in \mathbb{R}. \tag{4.3.15}\]

**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 \(|\beta\rangle\),

\[(|\alpha\rangle \langle \beta|) |\gamma\rangle = |\alpha\rangle (\langle \beta| \gamma\rangle) = (\langle \beta| \gamma\rangle) \cdot |\alpha\rangle, \tag{4.3.16}\]

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,

\[(|\alpha\rangle \langle \beta| + |\alpha'\rangle \langle \beta'|) |\gamma\rangle = |\alpha\rangle \langle \beta| \gamma\rangle + |\alpha'\rangle \langle \beta'| \gamma\rangle; \tag{4.3.17}\]

and for complex numbers \(c\) and \(d\),

\[|\alpha\rangle \langle \beta| (c |\gamma\rangle + d |\gamma'\rangle) = c |\alpha\rangle \langle \beta| \gamma\rangle + d |\alpha\rangle \langle \beta| \gamma'\rangle. \tag{4.3.18}\]

**Problem 4.10.** Show that

\[(|\alpha\rangle \langle \beta|)^\dagger = |\beta\rangle \langle \alpha| \tag{4.3.19}\]

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 \(|\alpha\rangle\).

**Superposition, the Identity operator as a completeness relation** 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,

\[\langle i| j \rangle = \delta^i_j. \tag{4.3.20}\]
Then we may consider acting an linear operator \( X \) on some arbitrary vector \( |\gamma\rangle \), which we will express as a linear combination of the \( \{|i\rangle\} \):

\[
|\gamma\rangle = \sum_i \hat{\gamma}^i |i\rangle, \quad \{\hat{\gamma}^i \in \mathbb{C}\}. \tag{4.3.21}
\]

By acting both sides with respect to \( \langle j| \), we have

\[
\langle j| \gamma\rangle = \hat{\gamma}^j. \tag{4.3.22}
\]

In other words,

\[
|\gamma\rangle = \sum_i |i\rangle \langle i| \gamma\rangle. \tag{4.3.23}
\]

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

\[
I = \sum_i |i\rangle \langle i|. \tag{4.3.24}
\]

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. \( I \) 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. (4.3.23), 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

\[
|\gamma\rangle \doteq \begin{bmatrix} \langle 1| \gamma \rangle \\ \langle 2| \gamma \rangle \\ \langle 3| \gamma \rangle \\ \vdots \\ \langle D| \gamma \rangle \end{bmatrix}. \tag{4.3.25}
\]

The \( \doteq \) is not quite an equality; rather it means “represented by,” in that this column vector contains as much information as eq. (4.3.23), provided the orthonormal basis vectors are known.

We may also express an arbitrary bra through a superposition of the basis bras \( \{|i\rangle\} \), using the adjoint of eq. (4.3.24).

\[
\langle \alpha| = \sum_i \langle \alpha| i \rangle \langle i|. \tag{4.3.26}
\]

(According to eq. (4.3.24), this is simply \( \langle \alpha| I \).) In this case, the coefficients \( \{\langle \alpha| i \rangle\} \) may be arranged as a row vector:

\[
\langle \alpha| \doteq \begin{bmatrix} \langle \alpha| 1 \rangle \\ \langle \alpha| 2 \rangle \\ \vdots \\ \langle \alpha| D \rangle \end{bmatrix}. \tag{4.3.27}
\]

**Inner products** Let us consider the inner product \( \langle \alpha| \gamma \rangle \). By inserting the completeness relation in eq. (4.3.24), we obtain

\[
\langle \alpha| \gamma \rangle = \sum_i \langle \alpha| i \rangle \langle i| \gamma \rangle = \delta_{ij} (\hat{\alpha})^* \hat{\gamma}^j = \hat{\alpha}^\dagger \hat{\gamma}, \tag{4.3.28}
\]

39
\[ \hat{\alpha}^i \equiv \langle i | \alpha \rangle, \quad \hat{\gamma}^j \equiv \langle j | \gamma \rangle. \]  

(4.3.29)

This is the reason for writing a ket \(| \gamma \rangle\) as a column whose components are its representation (eq. (4.3.25)) and a bra \(\langle \alpha |\) as a row whose components are the complex conjugate of its representation (eq. (4.3.27)) – their inner product is in fact the complex ‘dot product’

\[ \langle \alpha | \gamma \rangle = \begin{bmatrix} \langle 1 | \alpha \rangle \\ \langle 2 | \alpha \rangle \\ \vdots \\ \langle D | \alpha \rangle \end{bmatrix}^\dagger \begin{bmatrix} \langle 1 | \gamma \rangle \\ \langle 2 | \gamma \rangle \\ \vdots \\ \langle D | \gamma \rangle \end{bmatrix}, \]  

(4.3.30)

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

\[ 1 = \langle \gamma | \gamma \rangle = \sum_i \langle \gamma | i \rangle \langle i | \gamma \rangle = \sum_i |\langle i | \gamma \rangle|^2 = \delta_{ij} \langle \gamma^i | \gamma \rangle = \hat{\gamma}^\dagger \hat{\gamma}. \]  

(4.3.31)

**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 \(X\),

\[ X | \gamma \rangle = \sum_{i,j} |j \rangle \langle j | X | i \rangle \langle i | \gamma \rangle. \]  

(4.3.32)

We can also apply the \(l\)th basis bra \(\langle l |\) from the left on both sides and obtain

\[ \langle l | X | \gamma \rangle = \sum_i \langle l | X | i \rangle \langle i | \gamma \rangle. \]  

(4.3.33)

Just as we read off the components of the vector in eq. (4.3.23) as a column vector, we can do the same here. Again supposing a \(D\) dimensional vector space for notational convenience,

\[ X | \gamma \rangle = \begin{bmatrix} \langle 1 | X | 1 \rangle \\ \langle 2 | X | 1 \rangle \\ \vdots \\ \langle D | X | 1 \rangle \\ \langle 1 | X | 2 \rangle \\ \langle 2 | X | 2 \rangle \\ \vdots \\ \langle D | X | 2 \rangle \\ \vdots \\ \langle 1 | X | D \rangle \\ \langle 2 | X | D \rangle \\ \vdots \\ \langle D | X | D \rangle \end{bmatrix} \begin{bmatrix} \langle 1 | \gamma \rangle \\ \langle 2 | \gamma \rangle \\ \vdots \\ \langle D | \gamma \rangle \end{bmatrix}. \]  

(4.3.34)

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\(^9\)

\[ \hat{X}^i_j \equiv \langle i | X | j \rangle \quad \text{and} \quad \hat{\gamma}^j \equiv \langle j | \gamma \rangle, \]  

(4.3.35)

we have

\[ \langle i | X | \gamma \rangle = \hat{X}^i_j \hat{\gamma}^j. \]  

(4.3.36)

---

\(^9\)In this chapter on the abstract formulation of Linear Algebra, I use a \(\hat{\cdot}\) to denote a matrix (representation), in order to distinguish it from the linear operator itself.
Since $|\gamma\rangle$ in eq. (4.3.32) was arbitrary, we may record that any linear operator $X$ admits a ket-bra operator expansion:

$$X = \sum_{i,j} |j\rangle \langle X |i\rangle = \sum_{i,j} |j\rangle \hat{X}^j_i \langle i|.$$

We have already seen, this result follows from inserting the completeness relation in eq. (4.3.24) on the left and right of $X$. Importantly, notice that specifying the matrix $\hat{X}^j_i$ amounts to defining the linear operator $X$ itself, once a orthonormal basis has been picked.

**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. (4.3.37) – which holds for an arbitrary linear operator $X$ – teaches us the set of $D^2$ ket-bra operators $\{|j\rangle \langle i|, j, i = 1, 2, 3, \ldots, D\}$ (4.3.38) form the basis of the space of linear operators. The matrix elements $\langle j |X |i\rangle = \hat{X}^j_i$ are the expansion coefficients.

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

$$\langle i | (|\alpha\rangle \langle \beta|) |j\rangle = \langle i | \alpha \rangle \langle \beta | j \rangle \equiv \hat{\beta}^j_i \hat{\alpha}^i_j.$$

(4.3.39)

**Products of Linear Operators** We can consider $YX$, where $X$ and $Y$ are linear operators. By inserting the completeness relation in eq. (4.3.24),

$$YX |\gamma\rangle = \sum_{i,j,k} |k\rangle \langle k |Y |j\rangle \langle j |X |i\rangle \langle i| \gamma\rangle$$

$$= \sum_{k} |k\rangle \hat{Y}^k_j \hat{X}^j_k \hat{\gamma}^i.$$

(4.3.40)

The product $YX$ can therefore be represented as

$$YX = \begin{bmatrix}
\langle 1 |Y| 1 \rangle & \langle 1 |Y| 2 \rangle & \cdots & \langle 1 |Y| D \rangle \\
\langle 2 |Y| 1 \rangle & \langle 2 |Y| 2 \rangle & \cdots & \langle 2 |Y| D \rangle \\
\cdots & \cdots & \cdots & \cdots \\
\langle D |Y| 1 \rangle & \langle D |Y| 2 \rangle & \cdots & \langle D |Y| D \rangle \\
\end{bmatrix}
\begin{bmatrix}
\langle 1 |X| 1 \rangle & \langle 1 |X| 2 \rangle & \cdots & \langle 1 |X| D \rangle \\
\langle 2 |X| 1 \rangle & \langle 2 |X| 2 \rangle & \cdots & \langle 2 |X| D \rangle \\
\cdots & \cdots & \cdots & \cdots \\
\langle D |X| 1 \rangle & \langle D |X| 2 \rangle & \cdots & \langle D |X| D \rangle \\
\end{bmatrix}.$$

(4.3.41)

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

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

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

(4.3.42)
where the \{\vert \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
\[
\langle X \vert Y \rangle \equiv \text{Tr} \left[ X^\dagger Y \right].
\] (4.3.43)

This is in fact equivalent to the matrix trace inner product because, by inserting the completeness relation (4.3.24) between \(X\) and \(Y\) and employing eq. (4.3.42),
\[
\langle X \vert Y \rangle = \sum_{i,j} \langle i \vert X^\dagger \vert j \rangle \langle j \vert Y \vert i \rangle = (\hat{X}^\dagger)_j^i \hat{Y}^i_j = \text{Tr} \left[ \hat{X}^\dagger \hat{Y} \right].
\] (4.3.44)

With such a tool, it is now possible to sharpen the statement that the set of \(D^2\) ket-bra operators \(\{\vert i \rangle \langle j \vert \mid i,j \in 1,2,3,\ldots,D\} \) 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 \(\vert i \rangle \langle j \vert\) and \(\vert m \rangle \langle n \vert\). Utilizing the result that \((\vert i \rangle \langle j \vert)^\dagger = \vert j \rangle \langle i \vert\):
\[
\text{Tr} \left[ \left(\vert i \rangle \langle j \vert \right)^\dagger \left(\vert m \rangle \langle n \vert \right) \right] = \sum_{\ell} \langle \ell \vert j \rangle \langle i \vert m \rangle \langle n \vert \ell \rangle.
\] (4.3.45)

Now, by assumption, \(\langle \ell \vert j \rangle\) is non-zero only when \(\ell = j\). Similarly, \(\langle n \vert \ell \rangle\) is non-zero only when \(\ell = n\). Therefore when \(j \neq n\) the entire sum is zero because \(\ell\) cannot be simultaneously equal to both \(j\) and \(n\). But again by the orthonormal assumption, when \(\ell = j = n\), \(\langle \ell \vert j \rangle \langle n \vert \ell \rangle = 1\). In other words, the sum is proportional to \(\delta^j_n\), likewise \(\langle i \vert m \rangle = \delta^i_m\) too. At this point, we have arrived at the orthonormality condition:
\[
\text{Tr} \left[ \left(\vert i \rangle \langle j \vert \right)^\dagger \left(\vert m \rangle \langle n \vert \right) \right] = \delta^j_n \delta^i_m.
\] (4.3.46)

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

**Problem 4.11.** 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? □

**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 \(\{\vert i \rangle\}\). Firstly, from eq. (4.3.37),
\[
X^\dagger \vert i \rangle = \sum_{a,b} \vert a \rangle \left(\hat{X}^\dagger\right)^a_b \langle b \vert i \rangle = \sum_a \vert a \rangle \left(\hat{X}^\dagger\right)^a_i.
\] (4.3.47)

Taking the \(^\dagger\) using eq. (4.2.3), and then applying it to \(\vert j \rangle\),
\[
\left(X^\dagger \vert i \rangle\right)^\dagger \vert j \rangle = \sum_a \left(\hat{X}^\dagger\right)^a_i \langle a \vert j \rangle = \left(\hat{X}^\dagger\right)^j_i.
\] (4.3.48)
According to eq. (4.3.8), this must be equal to \( \langle i | X | j \rangle = \hat{X}_{ij}^* \). 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 \( X \):

\[
\langle j | X^\dagger | i \rangle = \langle i | X | j \rangle^* \quad \Leftrightarrow \quad \hat{X}^\dagger = (\hat{X}^T)^*.
\]  

(4.3.49)

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. (4.3.49) more directly by starting with the ket-bra expansion (cf. (4.3.37)) of \( X \) and then using equations (4.2.3) and (4.3.19) to directly implement \( \dagger \) on the right hand side:

\[
X^\dagger = \sum_{i,j} \left( \langle \hat{X}^i_j | i \rangle \right)^j
\]

(4.3.50)

\[= \sum_{i,j} |j\rangle \hat{X}^i_j \langle i \rangle \equiv \sum_{i,j} |j\rangle (\hat{X}^\dagger)^j_i \langle i \rangle.
\]

(4.3.51)

**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 \( \langle \alpha | \gamma \rangle = \langle \gamma | \alpha \rangle \):

\[
\langle \alpha | X | \beta \rangle = (X | \beta \rangle)^\dagger |\alpha\rangle = ((X^\dagger)^\dagger |\beta\rangle)^\dagger |\alpha\rangle = \langle \beta | X^\dagger | \alpha \rangle.
\]

(4.3.52)

**Mapping finite dimensional vector spaces to \( \mathbb{C}^D \)**

We 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 \( \{|1\rangle, \ldots, |D\rangle\} \). Then, every vector \( |\alpha\rangle \) can be represented as a column vector; the \( i \)th component is the result of projecting the abstract vector on the \( i \)th 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 \( \hat{O} \) can be represented as a matrix with the element on the \( i \)th row and \( j \)th column given by \( \langle i | \hat{O} | j \rangle \); and conversely, writing any square matrix \( \hat{O}^i_j \) can be interpreted to define a linear operator, on this vector space, with matrix elements \( \langle i | \hat{O} | j \rangle \). Product of linear operators becomes products of matrices, with the usual rules of matrix multiplication.

<table>
<thead>
<tr>
<th>Object</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector/Ket: (</td>
<td>\alpha\rangle = \sum_i \langle i</td>
</tr>
<tr>
<td>Dual Vector/Bra: ( \langle \alpha</td>
<td>= \sum_i \langle \alpha</td>
</tr>
<tr>
<td>Inner product: ( \langle \alpha</td>
<td>\beta \rangle = \sum_i \langle \alpha</td>
</tr>
<tr>
<td>Linear operator (LO): ( X = \sum_{i,j} \langle i</td>
<td>X</td>
</tr>
<tr>
<td>LO acting on ket: ( X</td>
<td>\gamma\rangle = \sum_{i,j} \langle i</td>
</tr>
<tr>
<td>Products of LOs: ( XY = \sum_{i,j,k} \langle i</td>
<td>X</td>
</tr>
<tr>
<td>Adjoint of LO: ( X^\dagger = \sum_{i,j} \langle j</td>
<td>X^\dagger</td>
</tr>
</tbody>
</table>
Next we highlight two special types of linear operators.

**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

\[
\partial_t |\psi(t)\rangle \equiv \lim_{\delta t \to 0} \frac{|\psi(t + \delta t)\rangle - |\psi(t)\rangle}{\delta t}.
\]

(4.3.53)

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

\[
\partial_t \langle \psi(t)| \equiv \lim_{\delta t \to 0} \frac{\langle \psi(t + \delta t)| - \langle \psi(t)|}{\delta t} = (\partial_t |\psi(t)\rangle)^\dagger.
\]

(4.3.54)

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

\[
\partial_t A(t) = \lim_{\delta t \to 0} \frac{A(t + \delta t) - A(t)}{\delta t}.
\]

(4.3.55)

**Problem 4.13. Product rule** Can you prove the product rule holds for the derivative of matrix elements; i.e.,

\[
\partial_t \langle \psi_1(t)|A(t)|\psi_2(t)\rangle = (\partial_t \langle \psi_1(t)|) A(t) |\psi_2(t)\rangle + \langle \psi_1(t)| (\partial_t A(t)) |\psi_2(t)\rangle + \langle \psi_1(t)| A(t) \partial_t |\psi_2(t)\rangle?
\]

(4.3.56)

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

\(\square\)

### 4.3.2 Hermitian Operators

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

\[
X^\dagger = X.
\]

(4.3.57)

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

\[
\langle \alpha| X |\beta\rangle = \langle \beta| X |\alpha\rangle^*
\]

(4.3.58)

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

\[
\langle j| X |i\rangle = \langle i| X |j\rangle^*.
\]

(4.3.59)

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.,

\[
X |a\rangle = a |a\rangle
\]

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

\[
\langle a'| X = a'^* \langle a'|.
\]

We can multiply \( \langle a'| \) from the left on both sides of eq. (4.3.60); and multiply \( |a\rangle \) from the right on both sides of eq. (4.3.61).

\[
\langle a'| X |a\rangle = a \langle a' | a \rangle , \quad \langle a'| X |a\rangle = a'^* \langle a'| a \rangle
\]

Subtracting these two equations,

\[
0 = (a - a'^*) \langle a'| a \rangle.
\]

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: The eigenkets \( \{ |\lambda_k\rangle | k = 1, 2, \ldots, D \} \) of a Hermitian operator span the vector space upon which it is acting. The full set of eigenvalues \( \{ \lambda_k | k = 1, 2, \ldots, D \} \) of some Hermitian operator is called its spectrum; and from eq. (4.3.24), completeness of its eigenvectors reads

\[
\mathbb{I} = \sum_{k=1}^{D} |\lambda_k\rangle \langle \lambda_k|.
\]

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 observables such as spin, position, momentum, etc., correspond to Hermitian operators; their eigenvalues are then the possible outcomes of the measurements of these observables. This is because their spectrum are real, which guarantees we get a real number from performing a measurement on the system at hand.

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.

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 \( \lambda \). 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 \( \ell \in \{0, 1, 2, 3, \ldots \} \). 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 \( \phi \) is the azimuthal angle). A closely related fact is that \([-\nabla_{S^2}^2, J(\phi)] = 0\), where \([X, Y] \equiv XY - YX\).

\[
- \nabla_{S^2}^2 |\ell, m\rangle = \ell(\ell + 1) |\ell, m\rangle, \quad \ell \in \{0, 1, 2, \ldots\}, \quad m \in \{-\ell, -\ell + 1, \ldots, -1, 0, 1, \ldots, \ell - 1, \ell\}.
\]

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. 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 4.14.** Let \( X \) be a linear operator with eigenvalues \( \{\lambda_i | i = 1, 2, 3, \ldots, D\} \) and orthonormal eigenvectors \( \{ |\lambda_i \rangle | i = 1, 2, 3, \ldots, D\} \) that span the given vector space. Show that \( X \) can be expressed as

\[
X = \sum_i \lambda_i |\lambda_i \rangle \langle \lambda_i |.
\] (4.3.66)

(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. (4.3.66), 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.

**How to diagonalize a Hermitian operator?** To *diagonalize* a linear operator \( X \) means to get it in the form in eq. (4.3.66), where it is expanded in terms of projectors built out of its eigen-kets \( \{ |\lambda_i \rangle \} \). For, the matrix representation in such a basis is purely diagonal \( \langle \lambda_i | X | \lambda_j \rangle = \lambda_i \delta_{ij} \).

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

\[
H = \sum_{i,j} |i \rangle \hat{H}^i_j \langle j |.
\] (4.3.67)

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. (3.2.36). By treating \( \hat{H}^i_j \) as a matrix, you can find its eigenvectors and eigenvalues \( \{\lambda_k\} \). Specifically, what you are solving for is the unitary matrix \( \hat{U}^j_k \), whose \( k \)th column is the \( k \)th unit length eigenvector of \( \hat{H}^i_j \), with eigenvalue \( \lambda_k \):

\[
\hat{H}^i_j \hat{U}^j_k = \lambda_k \hat{U}^j_k \iff \sum_j \langle i | H | j \rangle \langle j | \lambda_k \rangle = \lambda_k \langle i | \lambda_k \rangle,
\] (4.3.68)

---

\(^{10}\)See Wikipedia articles on the [Stark] and [Zeeman] effects for plots of the energy levels vs. electric/magnetic field strengths.
with
\[
\langle i | H | j \rangle \equiv \hat{H}^i_j \quad \text{and} \quad \langle j | \lambda_k \rangle \equiv \hat{U}^j_k.
\] (4.3.69)

In other words,
\[
\hat{H}^i_j = (\hat{U} \text{diag}[\lambda_1, \ldots, \lambda_D] \hat{U}^\dagger)^i_j = \sum_{m,n} \langle i | \lambda_m \rangle \delta_n^m \lambda_n \langle \lambda_n | j \rangle.
\] (4.3.70)

Once you have obtained the representation of the \( k \)th eigenket \( \hat{U}^i_k = (\langle 1 | \lambda_k \rangle, \langle 2 | \lambda_k \rangle, \ldots, \langle D | \lambda_k \rangle)^T \), you can then write the eigenket itself as
\[
| \lambda_k \rangle = \sum_i | i \rangle \langle i | \lambda_k \rangle = \sum_i | i \rangle \hat{U}^i_k.
\] (4.3.71)

The adjoint of the same eigenket is
\[
\langle \lambda_k | = \sum_i \langle \lambda_k | i \rangle = \sum_i \bar{\hat{U}}^i_k \langle i | = \sum_i (\hat{U}^{\dagger})^k_i \langle i |.
\] (4.3.72)

The operator \( H \) has now been diagonalized as
\[
H = \sum_k \lambda_k | \lambda_k \rangle \langle \lambda_k |
\] (4.3.73)
because according to eq. (4.3.70),
\[
H = \sum_{i,j} | i \rangle \hat{H}^i_j \langle j | = \sum_{i,j,a} | i \rangle \langle i | \lambda_a \rangle \lambda_a \langle \lambda_a | j \rangle \langle j |.
\] (4.3.74)

Using the completeness relation in eq. (4.3.24) then leads us to eq. (4.3.73).

In summary, with the relations in eq. (4.3.69),
\[
H = \sum_{i,j} | i \rangle \hat{H}^i_j \langle j | = \sum_k \lambda_k | \lambda_k \rangle \langle \lambda_k |
\] (4.3.75)
\[
= \sum_{i,j,m,n} | i \rangle \hat{U}^i_m (\text{diag} [\lambda_1, \ldots, \lambda_D])^m_n (\hat{U}^{\dagger})^n_j \langle j |.
\] (4.3.76)

**Problem 4.15.** Consider a 2 dimension vector space with the orthonormal basis \{\( | 1 \rangle \), \( | 2 \rangle \}\. The operator \( H \) is defined through its actions:
\[
H | 1 \rangle = a | 1 \rangle + ib | 2 \rangle,
\] (4.3.77)
\[
H | 2 \rangle = -ib | 1 \rangle + a | 2 \rangle;
\] (4.3.78)

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,

$$[A, B] \equiv AB - BA = 0. \quad (4.3.79)$$

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.

$$X = \sum_a a \langle a \rvert a \rangle \quad (4.3.80)$$

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

$$\langle a' \rvert [X, Y] \rvert a \rangle = \langle a' \rvert XY - YX \rvert a \rangle = (a' - a) \langle a' \rvert Y \rvert a \rangle = 0. \quad (4.3.81)$$

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

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

$$X = \sum_{a, b} a \rvert a, b \rangle \langle a, b \rvert \quad \text{and} \quad Y = \sum_{a, b} b \rvert a, b \rangle \langle a, b \rvert, \quad (4.3.82)$$

Let’s compute the commutator

$$[X, Y] = \sum_{a, b, a', b'} ab' (\rvert a, b \rangle \langle a, b \rvert a', b' \rangle \langle a', b' \rangle - \rvert a', b' \rangle \langle a, b \rvert) \quad (4.3.83)$$

Remember that eigenvectors corresponding to distinct eigenvalues are orthogonal, namely $\langle a, b \rvert a', b' \rangle$ 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.

$$[X, Y] = \sum_{a, b} ab (\rvert a, b \rangle \langle a, b \rvert - \rvert a, b \rangle \langle a, b \rvert a, b \rangle \langle a, b \rvert) = 0. \quad (4.3.84)$$

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

\[2\]Remember, to say $X$ or $Y$ is diagonalized means it has been put in the form in eq. (4.3.66). To say both of them have been simultaneously diagonalized therefore means they can be put in the form in eq. (4.3.66) using the same set of eigenkets.
Problem 4.16. 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

$$Y = \sum_a |a\rangle \langle a| Y |a\rangle \langle a| .$$

(4.3.85)

Read off the eigenvalues.

Problem 4.17. Properties of Commutators

Show that, for linear operators $A$, $B$, and $C$, the following relations hold.

$$[AB, C] = A[B, C] + [A, C]B,$$

(4.3.86)

$$[A, BC] = B[A, C] + [A, B]C,$$

(4.3.87)

$$[A, B] \dagger = -[A \dagger, B \dagger];$$

(4.3.88)

and


(4.3.89)

If we have a collection of linear operators $\{A_i|i = 1, 2, \ldots, M\}$ and $\{B_j|i = 1, 2, \ldots, N\}$, explain why the commutator in linear in both slots, in that

$$\sum_{i=1}^M A(i) \sum_{j=1}^N B(j) = \sum_{i=1}^M \sum_{j=1}^N [A(i), B(j)].$$

(4.3.90)

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

$$\langle \psi| \Delta X^2 |\psi\rangle \equiv \langle \psi|(X - \langle \psi|X|\psi\rangle)^2 |\psi\rangle;$$

(4.3.91)

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; $\Delta X$ (and, similarly, $\Delta Y$) is Hermitian.

From the Cauchy-Schwarz inequality of eq. (4.2.23), if we identify $|\alpha\rangle = \Delta X |\psi\rangle$ and $|\beta\rangle = \Delta Y |\psi\rangle$, then

$$\langle \psi| \Delta X^2 |\psi\rangle \langle \psi| \Delta Y^2 |\psi\rangle \geq |\langle \psi| \Delta X \Delta Y |\psi\rangle|^2.$$ 

(4.3.92)

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

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

(4.3.93)
where the anti-commutator itself is defined as

$$\{A, B\} \equiv AB + BA.$$  \hfill (4.3.94)

(If eq. \ref{eq:4.3.93} is not apparent, simply expand the right hand side.) Now, let us note that the commutator of two observables is anti-Hermitian, in that

$$[\Delta X, \Delta Y]^\dagger = (\Delta X \Delta Y)^\dagger - (\Delta Y \Delta X)^\dagger = \Delta Y \Delta X - \Delta X \Delta Y = -[\Delta X, \Delta Y].$$  \hfill (4.3.95)

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

$$\{\Delta X, \Delta Y\}^\dagger = (\Delta X \Delta Y)^\dagger + (\Delta Y \Delta X)^\dagger = \Delta Y \Delta X + \Delta X \Delta Y = \{\Delta X, \Delta Y\}.$$  \hfill (4.3.96)

Additionally, if $A^\dagger = \pm A$, then

$$\langle \psi | A | \psi \rangle^* = \langle \psi | A^\dagger | \psi \rangle = \pm \langle \psi | A | \psi \rangle.$$

The expectation value of a (anti-)Hermitian operator is purely (imaginary) real.

Altogether, we learn that the expectation value of eq. \ref{eq:4.3.93}, when $A = \Delta X$ and $B = \Delta Y$ – which reads

$$\langle \psi | \Delta X \Delta Y | \psi \rangle = \frac{1}{2} \langle \psi | [\Delta X, \Delta Y] | \psi \rangle + \frac{1}{2} \langle \psi | \{\Delta X, \Delta Y\} | \psi \rangle,$$  \hfill (4.3.99)

– consists of a purely imaginary portion (the commutator term) plus a purely real one (the anti-commutator term). But since the modulus square of a complex number is the sum of the square of its real and imaginary pieces,

$$| \langle \psi | \Delta X \Delta Y | \psi \rangle |^2 = \frac{1}{4} | \langle \psi | [\Delta X, \Delta Y] | \psi \rangle |^2 + \frac{1}{4} | \langle \psi | \{\Delta X, \Delta Y\} | \psi \rangle |^2.$$  \hfill (4.3.100)

Plugging this result back into eq. \ref{eq:4.3.92},

$$\langle \psi | \Delta X^2 | \psi \rangle \langle \psi | \Delta Y^2 | \psi \rangle \geq \frac{1}{4} | \langle \psi | [\Delta X, \Delta Y] | \psi \rangle |^2 + \frac{1}{4} | \langle \psi | \{\Delta X, \Delta Y\} | \psi \rangle |^2.$$  \hfill (4.3.101)

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. \ref{eq:4.3.101} must certainly larger or equal to the first commutator term, we arrive at the famous uncertainty relation

$$\langle \psi | \Delta X^2 | \psi \rangle \langle \psi | \Delta Y^2 | \psi \rangle \geq \frac{1}{4} | \langle \psi | [X, Y] | \psi \rangle |^2.$$  \hfill (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,

$$|\alpha\rangle = \sum_i |\lambda_i\rangle \langle \lambda_i | \alpha\rangle, \quad O |\lambda_i\rangle = \lambda_i |\lambda_i\rangle.$$  \hfill (4.3.103)
It is a postulate of quantum theory that the probability of obtaining a specific \( \lambda_j \) 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 \(| \lambda_i; j \rangle \) corresponding to \( \lambda_i \), then the probability will be

\[
P(\lambda_i) = \sum_j \langle \alpha | \lambda_i; j \rangle \langle \lambda_i; j | \alpha \rangle.
\] (4.3.104)

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

\[
\langle \alpha | O | \alpha \rangle.
\] (4.3.105)

If \( O \) is Hermitian, then the expectation value is real, since

\[
\langle \alpha | O | \alpha \rangle^* = \langle \alpha | O^\dagger | \alpha \rangle = \langle \alpha | O | \alpha \rangle.
\] (4.3.106)

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. This can be seen by expanding \(| \alpha \rangle \) in terms of the eigenstates of \( O \).

\[
\langle \alpha | O | \alpha \rangle = \sum_{i,j} \langle \alpha | \lambda_i \rangle \langle \lambda_i | O | \lambda_j \rangle \langle \lambda_j | \alpha \rangle
\]

\[
= \sum_{i,j} \langle \alpha | \lambda_i \rangle \lambda_i \langle \lambda_i | \lambda_j \rangle \langle \lambda_j | \alpha \rangle
\]

\[
= \sum_i |\langle \alpha | \lambda_i \rangle|^2 \lambda_i.
\] (4.3.107)

The probability of finding \( \lambda_i \) 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.)

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.

\[
\langle \alpha | O | \alpha \rangle^* = \langle \alpha | O^\dagger | \alpha \rangle = -\langle \alpha | O | \alpha \rangle.
\] (4.3.108)

**Hellmann-Feynman** Whenever the Hermitian operator \( A(\alpha_1, \alpha_2, \ldots) \equiv A(\vec{\alpha}) \) 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

\[
\lambda(\vec{\alpha}) = \langle \lambda(\vec{\alpha}) | A(\vec{\alpha}) | \lambda(\vec{\alpha}) \rangle.
\] (4.3.109)

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

\[
\frac{\partial \lambda(\vec{\alpha})}{\partial \alpha_i} = \left. \lambda(\vec{\alpha}) \right| \frac{\partial A(\vec{\alpha})}{\partial \alpha_i} \left. \lambda(\vec{\alpha}) \right|,
\] \( i = 1, 2, 3, \ldots \) (4.3.110)
A straightforward differentiation would confirm
\[ \partial_\alpha \lambda = (\partial_\alpha, \langle \lambda \rangle A | \lambda \rangle + \langle \lambda | A \partial_\alpha | \lambda \rangle + \langle \lambda | \partial_\alpha A | \lambda \rangle. \]  
(4.3.111)

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 \( | \lambda \rangle \).
\[
\partial_\alpha \lambda = \lambda \left\{ \left( \partial_\alpha, \langle \lambda \rangle | \lambda \rangle + \langle \lambda | \partial_\alpha A | \lambda \rangle \right) + \langle \lambda | \partial_\alpha A | \lambda \rangle \right\}.
\]  
(4.3.112)

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. \( (3.2.17) \). 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
  \[ \sigma^i \sigma^j = \delta^{ij} \mathbb{I} + i \sum_k \epsilon^{ijk} \sigma^k. \]  
(4.3.113)

That this is consistent with the Hermitian nature of the \( \{ \sigma^i \} \) can be checked by taking \( ^\dagger \) 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
\[ [\sigma^i, \sigma^j] = (\delta^{ij} - \delta^{ji}) \mathbb{I} + \sum_k i(\epsilon^{ijk} - \epsilon^{jik}) \sigma^k = 2i \sum_k \epsilon^{ijk} \sigma^k. \]  
(4.3.114)

We then define the operators
\[ \sigma^\pm \equiv \sigma^1 \pm i \sigma^2 \Rightarrow (\sigma^\pm)^\dagger = \sigma^\mp; \]  
(4.3.115)

and calculate\(^\dagger\)
\[ [\sigma^3, \sigma^\pm] = [\sigma^3, \sigma^1] \pm i[\sigma^3, \sigma^2] = 2i \epsilon^{312} \sigma^2 \pm 2i \epsilon^{321} \sigma^1 \]  
\[ = 2i \sigma^2 \pm 2 \sigma^1 = \pm 2(\sigma^1 \pm i \sigma^2), \]  
(4.3.116)

\[ [\sigma^3, \sigma^\pm] = \pm 2 \sigma^\pm. \]  
(4.3.117)

Also,
\[ \sigma^\mp \sigma^\pm = (\sigma^1 \mp i \sigma^2)(\sigma^1 \pm i \sigma^2) \]
\[ = (\sigma^1)^2 + (\mp i)(\pm i)(\sigma^2)^2 \mp i \sigma^2 \sigma^1 \pm i \sigma^1 \sigma^2 \]
\[ = 2 \mathbb{I} \pm i(\sigma^1 \sigma^2 - \sigma^2 \sigma^1) = 2 \mathbb{I} \pm i[\sigma^1, \sigma^2] = 2 \mathbb{I} \pm 2i^2 \epsilon^{123} \sigma^3 \]
\[ \quad \text{[X,Y] + [X,Z].} \]
\[ \Rightarrow \sigma^\pm \sigma^\pm = 2(\mathbb{I} \mp \sigma^3). \]  

(4.3.118)

\[ \sigma^3 \text{ and its Matrix representation.} \]

Suppose \( |\lambda\rangle \) is a unit norm eigenket of \( \sigma^3 \). Using \( \sigma^3 |\lambda\rangle = \lambda |\lambda\rangle \) and \((\sigma^3)^2 = \mathbb{I}\),

\[ 1 = \langle \lambda | \lambda \rangle = \langle \lambda | \sigma^3 \sigma^3 | \lambda \rangle = (\sigma^3 |\lambda\rangle)^\dagger (\sigma^3 |\lambda\rangle) = \lambda^2 \langle \lambda | \lambda \rangle = \lambda^2. \]

(4.3.119)

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

\[ \sigma^3 = |+\rangle \langle +| - |\rangle \langle -|, \]

(4.3.120)

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:

\[
\langle j | \sigma^3 | i \rangle = \begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix}. 
\]

(4.3.121)

Observe that we could have considered \( \langle \lambda | \sigma^i \sigma^4 | \lambda \rangle \) 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 \( \sigma^i \)s 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

\[
\sigma^3 \sigma^\pm |\lambda\rangle = (\sigma^3 \sigma^\pm - \sigma^\pm \sigma^3 + \sigma^\pm \sigma^3) |\lambda\rangle \\
= ([\sigma^3, \sigma^\pm] + \sigma^\pm \sigma^3) |\lambda\rangle = (\pm 2 \sigma^\pm + \lambda \sigma^\pm) |\lambda\rangle \\
= (\lambda \pm 2) \sigma^\pm |\lambda\rangle \quad \Rightarrow \quad \sigma^\pm |\lambda\rangle = K^\pm_\lambda |\lambda \pm 2\rangle, \quad K^\pm_\lambda \in \mathbb{C}. 
\]

(4.3.122)

This is why the \( \sigma^\pm \) are often called raising/lowering operators: when applied to the eigenket \( |\lambda\rangle \) of \( \sigma^3 \) it returns an eigenket with eigenvalue raised/lowered by 2 relative to \( \lambda \). 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\(^{14}\) and even the notion of particles in quantum field theory.

What is the norm of \( \sigma^\pm |\lambda\rangle \)?

\[
\langle \lambda | \sigma^\mp \sigma^\mp | \lambda \rangle = |K^\pm_\lambda|^2 \langle \lambda \pm 2 | \lambda \pm 2 \rangle \\
\langle \lambda | 2(\mathbb{I} \mp \sigma^3) | \lambda \rangle = |K^\pm_\lambda|^2 \\
2(1 \mp \lambda) = |K^\pm_\lambda|^2.
\]

(4.3.123)

This means we can solve \( K^\pm_\lambda \) up to a phase

\[
K^\pm_\lambda = e^{i\delta^\pm_\lambda} \sqrt{2(1 \mp \lambda), \quad \lambda \in \{-1, +1\}.}
\]

(4.3.124)

\(^{14}\)For the H atom, the algebraic derivation of its energy levels involve the quantum analog of the classical Laplace-Runge-Lenz vector.
Note that \( K^+ = e^{i\delta_+} \sqrt{2(1 - (+1))} = 0 \), and \( K^- = e^{i\delta_-} \sqrt{2(1 + (-1))} = 0 \), which means

\[
\sigma^+ |+\rangle = 0, \quad \sigma^- |−\rangle = 0. \tag{4.3.125}
\]

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 \( K^+ = e^{i\delta_-} \sqrt{2(1 - (-1))} = 2e^{i\delta_-} \).

\[
\sigma^+ |−\rangle = 2e^{i\delta_-} |+\rangle, \quad \sigma^- |+\rangle = 2e^{i\delta_+} |−\rangle. \tag{4.3.126}
\]

At this point, we have proved that the spectrum of \( \sigma^3 \) has to include both \(|±\rangle\), because we can get from one to the other by applying \( \sigma^± \) 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 \( \sigma^± \) 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:

\[
\sigma^+ |−\rangle = 2e^{i\delta_-} |+\rangle \langle j|, \quad \sigma^- |+\rangle = 2e^{i\delta_+} |−\rangle \langle i|. \tag{4.3.127}
\]

Since \((\sigma^+)\dagger = \sigma^-\), we must have \(\delta_-^{(−)} = −\delta_+^{(+) \equiv \delta}\).

\[
\sigma^+ = 2e^{i\delta} |+\rangle \langle −|, \quad \sigma^- = 2e^{-i\delta} |−\rangle \langle +|. \tag{4.3.128}
\]

with the corresponding matrix representations, with \(\langle +|\sigma^±|+\rangle\) being the top left hand corner entry:

\[
\langle j|\sigma^+|i\rangle = \begin{bmatrix} 0 & 2e^{i\delta} \\ 0 & 2e^{-i\delta} \end{bmatrix}, \quad \langle j|\sigma^-|i\rangle = \begin{bmatrix} 0 & 0 \\ 2e^{-i\delta} & 0 \end{bmatrix}. \tag{4.3.129}
\]

Now, we have \(\sigma^± = \sigma^1 \pm i\sigma^2\), which means we can solve for

\[
2\sigma^1 = \sigma^+ + \sigma^−, \quad 2i\sigma^2 = \sigma^+ − \sigma^−. \tag{4.3.130}
\]

We have

\[
\sigma^1 = e^{i\delta} |+\rangle \langle −| + e^{-i\delta} |−\rangle \langle +|, \quad \sigma^2 = −ie^{i\delta} |+\rangle \langle −| + ie^{-i\delta} |−\rangle \langle +|, \quad \delta \in \mathbb{R}, \tag{4.3.131}
\]

with matrix representations

\[
\langle j|\sigma^1|i\rangle = \begin{bmatrix} 0 & e^{i\delta} \\ e^{-i\delta} & 0 \end{bmatrix}, \quad \langle j|\sigma^2|i\rangle = \begin{bmatrix} 0 & −ie^{i\delta} \\ ie^{-i\delta} & 0 \end{bmatrix}. \tag{4.3.132}
\]

You can check explicitly that the algebra in eq. (4.3.113) holds for any \(\delta\). However, we can also use the fact that unit normal eigenkets can be re-scaled by a phase and still remain unit norm eigenkets.

\[
\sigma^3 (e^{i\theta} |±\rangle) = \pm (e^{i\theta} |±\rangle), \quad (e^{i\theta} |±\rangle)\dagger (e^{i\theta} |±\rangle) = 1, \quad \theta \in \mathbb{R}. \tag{4.3.133}
\]
We re-group the phases occurring within our $\sigma^3$ and $\sigma^\pm$ as follows.

\[
\sigma^3 = (e^{i\delta/2}|+\rangle)(e^{i\delta/2}|+\rangle)^\dagger - (e^{-i\delta/2}|-\rangle)(e^{-i\delta/2}|-\rangle)^\dagger, \quad (4.3.135)
\]
\[
\sigma^+ = 2(e^{i\delta/2}|+\rangle)(e^{-i\delta/2}|-\rangle)^\dagger, \quad \sigma^- = 2(e^{-i\delta/2}|-\rangle)(e^{i\delta/2}|+\rangle)^\dagger. \quad (4.3.136)
\]

That is, if we re-define $|\pm\rangle' \equiv e^{\pm i\delta/2}|\pm\rangle$, followed by dropping the primes, we would have

\[
\sigma^3 = |+\rangle\langle +| - |-\rangle\langle -|, \quad (4.3.137)
\]
\[
\sigma^+ = 2|+\rangle\langle -|, \quad \sigma^- = 2|-\rangle\langle +|, \quad (4.3.138)
\]

and again using $\sigma^1 = (\sigma^1 + \sigma^2)/2$ and $\sigma^2 = -i(\sigma^1 - \sigma^2)/2$,

\[
\sigma^1 = |+\rangle\langle -| + |+\rangle\langle +|, \quad (4.3.139)
\]
\[
\sigma^2 = -i|+\rangle\langle -| + i|-\rangle\langle +|, \quad \delta \in \mathbb{R}. \quad (4.3.140)
\]

We see that the Pauli matrices in eq. (3.2.17) correspond to the matrix representations of $\sigma^i$ in the basis built out of the unit norm eigenkets of $\sigma^3$, with an appropriate choice of phase.

Note that there is nothing special about choosing our basis as the eigenkets of $\sigma^3$ – we could have chosen the eigenkets of $\sigma^1$ or $\sigma^2$ as well. The analogous raising and lower operators can then be constructed from the remaining $\sigma^i$'s.

Finally, for $\hat{U}$ unitary we have already noted that $\det(\hat{U}\hat{\sigma}^i\hat{U}^\dagger) = \det \hat{\sigma}^i$ and $\text{Tr} \left[ \hat{U}\hat{\sigma}^i\hat{U}^\dagger \right] = \text{Tr} \left[ \hat{\sigma}^i \right]$. Therefore, if we choose $\hat{U}$ such that $\hat{U}\hat{\sigma}^i\hat{U}^\dagger = \text{diag}(1,-1)$ – since we now know the eigenvalues of each $\hat{\sigma}^i$ are $\pm 1$ – we readily deduce that

\[
\det \hat{\sigma}^i = -1, \quad \text{Tr} \left[ \hat{\sigma}^i \right] = 0. \quad (4.3.141)
\]

(However, $\hat{\sigma}^2\hat{\sigma}^i\hat{\sigma}^2 = -\left(\hat{\sigma}^i\right)^*$ 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.,

\[
U^\dagger U =UU^\dagger = I. \quad (4.3.142)
\]

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 $\{|i\rangle | i = 1,2,\ldots,D\}$, 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 $\{\hat{x},\hat{y},\hat{z}\}$ to some other $\{\hat{x}',\hat{y}',\hat{z}'\}$
(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

$$\{\vert 1 \rangle, \ldots, \vert D \rangle \} \quad \text{and} \quad \{\vert 1' \rangle, \ldots, \vert D' \rangle \}.$$  

(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

$$U \vert i \rangle = \vert i' \rangle, \quad i \in \{1, 2, \ldots, D\}. \quad (4.3.144)$$

Its matrix representation in the unprimed basis \{\vert i \rangle\} is gotten by projecting both sides along $\vert j \rangle$.

$$\langle j \vert U \vert i \rangle = \langle j \vert i' \rangle, \quad i, j \in \{1, 2, \ldots, D\}. \quad (4.3.145)$$

Is $U$ really unitary? One way to verify this is through its matrix representation. We have

$$\langle j \vert U^\dagger \vert i \rangle = \langle i \vert U \vert j \rangle^* = \langle j' \vert i \rangle. \quad (4.3.146)$$

Whereas $U^\dagger U$ in matrix form is

$$\sum_k \langle j \vert U^\dagger \vert k \rangle \langle k \vert U \vert i \rangle = \sum_k \langle k \vert U \vert j \rangle^* \langle k \vert U \vert i \rangle = \sum_k \langle k \vert i' \rangle \langle k \vert j \rangle^* = \sum_k \langle j' \vert k \rangle \langle k \vert i \rangle. \quad (4.3.147)$$

Because both \{\vert k \rangle\} and \{\vert k' \rangle\} form an orthonormal basis, we may invoke the completeness relation eq. (4.3.24) to deduce

$$\sum_k \langle j \vert U^\dagger \vert k \rangle \langle k \vert U \vert i \rangle = \langle j' \vert i \rangle = \delta_{ij}. \quad (4.3.148)$$

That is, we recover the unit matrix when we multiply the matrix representation of $U^\dagger$ to that of $U$\textsuperscript{15} 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. (4.3.144) can be expressed as

$$U = \sum_j \vert j' \rangle \langle j \vert. \quad (4.3.149)$$

All we have to verify is $U \vert i \rangle = \vert i' \rangle$ for any $i \in \{1, 2, 3, \ldots, D\}$.

$$U \vert i \rangle = \sum_j \vert j' \rangle \langle j \vert \langle i \rangle = \sum_j \vert j' \rangle \delta_{ij} = \vert i' \rangle. \quad (4.3.150)$$

\textsuperscript{15}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.
The unitary nature of $U$ can also be checked explicitly. Remember $(|\alpha\rangle \langle \beta|)^\dagger = |\beta\rangle \langle \alpha|$. 

$$U^\dagger U = \sum_j |j\rangle \langle j'| \sum_k |k'\rangle \langle k| = \sum_j |j\rangle \langle j'| \sum_k |k'\rangle \langle k|$$

$$= \sum_{j,k} |j\rangle \delta^j_k \langle k| = \sum_j |j\rangle \langle j| = \mathbb{I}. \quad (4.3.151)$$

The very last equality is just the completeness relation in eq. (4.3.24).

Starting from $U$ defined in eq. (4.3.144) 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 $\{|i'\rangle\}$ by *defining*

$$|i'\rangle \equiv U |i\rangle. \quad (4.3.152)$$

All we have to show is that $\{|i'\rangle\}$ form an orthonormal set.

$$\langle i'|i'\rangle = (U |j\rangle)^\dagger (U |i\rangle) = \langle j |U^\dagger U |i\rangle = \langle j |i\rangle = \delta^j_i. \quad (4.3.153)$$

We may therefore pause to summarize our findings as follows.

**Change-of-basis of $\langle \alpha| i$** Given a bra $\langle \alpha|$, we may expand it either in the new $\{|i'\rangle\}$ or old $\{|i\rangle\}$ basis bras,

$$\langle \alpha | = \sum_i \langle \alpha | i \rangle \langle i | = \sum_i \langle \alpha | i' \rangle \langle i' |. \quad (4.3.154)$$

We can relate the components of expansions using $\langle i | U | k\rangle = \langle i | k'\rangle$ (cf. eq. (4.3.145)),

$$\sum_k \langle \alpha | k' \rangle \langle k'| = \sum_i \langle \alpha | i \rangle \langle i |$$

$$= \sum_{i,k} \langle \alpha | i \rangle \langle i | k' \rangle \langle k'| = \sum_k \left( \sum_i \langle \alpha | i \rangle \langle i | U | k\rangle \right) \langle k'|. \quad (4.3.155)$$

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 $\tilde{U}$,

$$\langle \alpha | k' \rangle = \sum_i \langle \alpha | i \rangle \langle i | U | k\rangle. \quad (4.3.156)$$

In words: the $\langle \alpha |$ row vector in the basis $\{|i'\rangle\}$ is equal to $U$, written in the basis $\{|j | U | i\rangle\}$, acting (from the right) on the $\langle \alpha | i \rangle$ row vector, the $\langle \alpha |$ in the basis $\{|i\rangle\}$. Moreover, in index notation,

$$\tilde{\alpha}_k' = \tilde{\alpha}_i \tilde{U}_k^i. \quad (4.3.157)$$
Problem 4.18. 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 $i$th 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

$$\langle i' | \alpha \rangle = \sum_k \langle i | U^\dagger k \rangle \langle k | \alpha \rangle, \quad U |i\rangle = |i'\rangle.$$  \hfill (4.3.158)

In words: the $|\alpha\rangle$ column vector in the basis $\{|i'\rangle\}$ is equal to $U^\dagger$ written in the basis $\{\langle j | U^\dagger | i \rangle\}$, acting (from the left) on the $\langle i | \alpha \rangle$ column vector, the $|\alpha\rangle$ in the basis $\{|i\rangle\}$.

Furthermore, in index notation,

$$\hat{\alpha}' = (\hat{U}^\dagger)^i_k \hat{\alpha}^k.$$ \hfill (4.3.159)

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 4.19. 2D rotation in 3D. Let’s rotate the basis vectors of the 2D plane, spanned by the $x$- and $z$-axis, by an angle $\theta$. 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, $U(\theta) |2\rangle = |2\rangle$. (Drawing a picture may help.) Can you then write down the matrix representation $\langle j | U(\theta) | i \rangle$?

Problem 4.20. Consider a 2 dimension vector space with the orthonormal basis $\{|1\rangle, |2\rangle\}$. The operator $U$ is defined through its actions:

$$U |1\rangle = \frac{1}{\sqrt{2}} |1\rangle + \frac{i}{\sqrt{2}} |2\rangle, \hfill (4.3.160)$$
$$U |2\rangle = \frac{i}{\sqrt{2}} |1\rangle + \frac{1}{\sqrt{2}} |2\rangle. \hfill (4.3.161)$$

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. (4.3.24) in the basis $\{|i\rangle\}$, on both the left and the right,

$$\langle i' | X | j' \rangle = \sum_{k,l} \langle i' | k \rangle \langle k | X | l \rangle \langle l | j' \rangle$$
$$= \sum_{k,l} \langle i | U^\dagger | k \rangle \langle k | X | l \rangle \langle l | U | j \rangle = \langle i | U^\dagger X U | j \rangle,$$ \hfill (4.3.162)

where we have recognized (from equations (4.3.145) and (4.3.146)) $\langle i' | k \rangle = \langle i | U^\dagger | k \rangle$ and $\langle l | j' \rangle = \langle l | U | j \rangle$. If we denote $\hat{X}'$ as the matrix representation of $X$ with respect to the primed
basis; and \( \hat{X} \) and \( \hat{U} \) as their corresponding operators with respect to the unprimed basis, we recover the similarity transformation

\[
\hat{X}' = \hat{U}^\dagger \hat{X} \hat{U}.
\]  

(4.3.163)

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 \( \{|i\rangle\} \),

\[
\hat{X}'_{i'j'} = (\hat{U}^\dagger)^i_k \hat{X}^k_l \hat{U}^j_l.
\]  

(4.3.164)

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, \( X'_{i'i'} = X_i^i \).

On the other hand, from the last equality of eq. (4.3.162), we may also view \( \hat{X}' \) as the matrix representation of the operator \( X' \equiv U^\dagger X U \)

(4.3.165)

written in the old basis \( \{|i\rangle\} \). To reiterate,

\[
\langle i' | X | j' \rangle = \langle i | U^\dagger X U | j \rangle.
\]  

(4.3.166)

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 \( \lambda \) be the corresponding eigenvalue of \( X \).

\[
X |\lambda\rangle = \lambda |\lambda\rangle
\]  

(4.3.167)

By inserting a \( I = U U^\dagger \) between \( X \) and \( |\lambda\rangle \); and multiplying both sides on the left by \( U^\dagger \),

\[
U^\dagger X U U^\dagger |\lambda\rangle = \lambda U^\dagger |\lambda\rangle,
\]  

(4.3.168)

\[
X'(U^\dagger |\lambda\rangle) = \lambda (U^\dagger |\lambda\rangle).
\]  

(4.3.169)

That is, given the eigenvector \( |\lambda\rangle \) of \( X \) with eigenvalue \( \lambda \), the corresponding eigenvector of \( X' \) is \( U^\dagger |\lambda\rangle \) with precisely the same eigenvalue \( \lambda \).

**Theorem.** Let \( X' \equiv U^\dagger X U \). Then \( X \) is Hermitian iff \( X' \) is Hermitian.

Moreover, \( X \) is unitary iff \( X' \) is unitary.

---

16This 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.
Proof If $X$ is Hermitian, we consider $X^\dagger$.

$$X^\dagger = (U^\dagger XU)^\dagger = U^\dagger X^\dagger (U^\dagger)^\dagger = U^\dagger X U = X'.$$  \hspace{1cm} (4.3.170)

If $X$ is unitary we consider $X^\dagger X'$.

$$X^\dagger X' = (U^\dagger XU)^\dagger (U^\dagger XU) = U^\dagger X^\dagger U U^\dagger X U = U^\dagger X X U = U^\dagger U = I.$$  \hspace{1cm} (4.3.171)

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

$$U^\dagger \sigma^i U = \sigma^j, \quad i \neq j.$$  \hspace{1cm} (4.3.172)

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

$$X = \sum_k \lambda_k |\lambda_k\rangle \langle \lambda_k|$$  \hspace{1cm} (4.3.173)

and defined the unitary operator

$$U |k\rangle = |\lambda_k\rangle \quad \Leftrightarrow \quad \langle i | U | k\rangle = \langle i | \lambda_k\rangle.$$  \hspace{1cm} (4.3.174)

Notice the $k$th column of $\hat{U}^i_k \equiv \langle i | U | k\rangle$ are the components of the $k$th unit norm eigenvector $|\lambda_k\rangle$ written in the $\{|i\rangle\}$ basis. This implies, via two insertions of the completeness relation in eq. \ref{eq:4.3.24},

$$X = \sum_{i,j,k} \lambda_k \langle i | \lambda_k \rangle \langle \lambda_k | j \rangle \langle j |.$$  \hspace{1cm} (4.3.175)

Taking matrix elements,

$$\langle i | X | j \rangle = \tilde{X}^i_j = \sum_{k,l} \langle i | \lambda_k \rangle \lambda_l \delta^k_l \langle \lambda_l | j \rangle = \sum_{k,l} \hat{U}^i_k \lambda_k \delta^k_l (\hat{U}^\dagger)^l_j.$$  \hspace{1cm} (4.3.176)

Multiplying both sides by $\hat{U}^\dagger$ on the left and $\hat{U}$ on the right, we have

$$\hat{U}^\dagger \hat{X} \hat{U} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_D).$$  \hspace{1cm} (4.3.177)

60
Schur decomposition. Not all linear operators are diagonalizable. However, we already know that any square matrix \( \hat{X} \) can be brought to an upper triangular form
\[
\hat{U}^\dagger \hat{X} \hat{U} = \hat{\Gamma} + \hat{N},
\]
where the \( \{ \lambda_i \} \) are the eigenvalues of \( X \) and \( \hat{N} \) is strictly upper triangular. We may now phrase the Schur decomposition as a change-of-basis from \( \hat{X} \) 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
\[
\text{Tr} [X] = \sum_i \langle i | X | i \rangle, \quad \langle i | j \rangle = \delta^i_j. \tag{4.3.179}
\]
The Trace yields a complex number\(^{17}\) 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
\[
\sum_i \langle i' | X | i' \rangle = \sum_{i,j,k} \langle i' | j \rangle \langle j | X | k \rangle \langle k | i' \rangle = \sum_{i,j,k} \langle k | i' \rangle \langle i' | j \rangle \langle j | X | k \rangle
= \sum_{j,k} \langle k | j \rangle \langle j | X | k \rangle = \sum_k \langle k | X | k \rangle. \tag{4.3.180}
\]
Because \( \text{Tr} \) 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 4.21. Prove the following statements. For linear operators \( X \) and \( Y \), and unitary operator \( U \),
\[
\text{Tr} [XY] = \text{Tr} [YX] \tag{4.3.181}
\]
\[
\text{Tr} [U^\dagger X U] = \text{Tr} [X] \tag{4.3.182}
\]
The second identity tells you \( \text{Tr} \) is a basis-independent operation.

Problem 4.22. 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 \( \{ A^i | i = 1, 2, \ldots, N \} \) obeys
\[
[A^i, A^j] = ij^{ijk} A^k \tag{4.3.183}
\]
\(^{17}\)Be aware that the trace may not make sense in an infinite dimensional continuous vector space.
for some constants \( \{ f^{ijk} \} \); then under

\[
A^i \equiv U^\dagger A^i U,
\]

one obtains

\[
[A^i, A^j] = i f^{ijk} A^k
\]

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. (4.3.185) still holds. \( \square \)

### 4.3.4 Additional Problems

**Problem 4.23.** If \( \{|i\rangle \mid i = 1, 2, 3 \ldots, D \} \) is a set of orthonormal basis vectors, what is \( \text{Tr} [|j\rangle \langle k|] \), where \( j, k \in \{1, 2, \ldots, D\} \)?

**Problem 4.24.** Verify the following Jacobi identity. For linear operators \( X, Y \) and \( Z \),

\[
[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.
\]

Furthermore, verify that

\[
[X, Y] = -[Y, X], \quad [X, Y + Z] = [X, Y] + [X, Z],
\]

\[
[X, Y Z] = [X, Y] Z + Y [X, Z].
\]

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 4.25.** 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

\[
X \equiv a (|1\rangle \langle 1| - |2\rangle \langle 2| + |1\rangle \langle 2| + |2\rangle \langle 1|).
\]

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 \( \{|i\rangle \mid i = 1, 2, 3\} \).

\[
Y \equiv a (|1\rangle \langle 1| - |2\rangle \langle 2| - |3\rangle \langle 3|),
\]

\[
Z \equiv b |1\rangle \langle 1| - ib |2\rangle \langle 3| + ib |3\rangle \langle 2|.
\]

(In equations (4.3.189) and (4.3.190), \( a \) 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 4.26.** Pauli matrices re-visited. Refer to the Pauli matrices \( \{ \sigma^\mu \} \) defined in eq. (3.2.17). Let \( p_\mu \) be a 4-component collection of real numbers. We may then view \( p_\mu \sigma^\mu \) (where \( \mu \) sums over 0 through 3) as a Hermitian operator acting on a 2 dimensional vector space.
1. Find the eigenvalues $\lambda_{\pm}$ and corresponding unit norm eigenvectors $\xi_{\pm}$ of $p_i \sigma^i$ (where $i$ sums over 1 through 3). These are called the helicity eigenstates. Are they also eigenstates of $p_\mu \sigma^\mu$? (Hint: consider $[p_i \sigma^i, p_\mu \sigma^\mu]$.)

2. Explain why

$$p_i \hat{\sigma}^i = \lambda_+ \xi^+ (\xi^+)\dagger + \lambda_- \xi^- (\xi^-)\dagger.$$  \hspace{1cm} (4.3.191)

Can you write down the analogous expansion for $p_\mu \hat{\sigma}^\mu$?

3. If we define the square root of an operator or matrix $\sqrt{A}$ as the solution to $\sqrt{A}^2 = A$, write down the expansion for $\sqrt{p_i \hat{\sigma}^i}$.

4. These 2 component spinors $\xi^{\pm}$ play a key role in the study of Lorentz symmetry in 4 space-time dimensions. Consider applying an invertible transformation $L_A^B$ on these spinors, i.e., replace

$$(\xi^\pm)_A \to L_A^B (\xi^\pm)_B.$$ \hspace{1cm} (4.3.192)

(The A and B indices run from 1 to 2, the components of $\xi^\pm$.) How does $p_\mu \hat{\sigma}^\mu$ change under such a transformation? And, how does its determinant change?

---

**Problem 4.27.** 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 $\{\hat{S}_j^i\}$ in

$$S |i\rangle = \sum_j |j\rangle \hat{S}_j^i \equiv |i'\rangle.$$ \hspace{1cm} (4.3.193)

Explain why $\hat{S}_j^i = \langle i | S | j \rangle$ is still $\langle i | j' \rangle$. (Compare with eq. (4.3.146).) On the other hand, since $S$ is no longer unitary, its matrix representation $\hat{S}$ is no longer a unitary matrix. Show, however, that the inverse transformation is directly related to the inverse matrix, which obeys $\hat{S}^{-1} \hat{S} = I$:

$$|i\rangle = \sum_j |j'\rangle (\hat{S}^{-1})^j_i.$$ \hspace{1cm} (4.3.194)

\[^{18}\text{As a concrete problem, let us perform the following change-of-basis, for } \theta \neq \phi:\]

$$S |1\rangle = \cos(\theta) |1\rangle + \sin(\theta) |2\rangle \equiv |1'\rangle,$$ \hspace{1cm} (4.3.195)

$$S |2\rangle = -\sin(\theta) |1\rangle + \cos(\theta) |2\rangle \equiv |2'\rangle.$$ \hspace{1cm} (4.3.196)

Solve $\hat{S}^{-1}$ and find $|1\rangle$ and $|2\rangle$ in terms of $\{|1'\rangle, |2'\rangle\}$. \hspace{1cm} \square

\[^{18}\text{Note that our discussion implicitly assumes } \hat{S}^{-1} \text{ exists, for otherwise we are not performing a faithful coordinate transformation but discarding information about the vector space. As a simple 2D example, we could define } S |1\rangle = |1\rangle \equiv |1'\rangle \text{ and } S |2\rangle = |1\rangle \equiv |2'\rangle \text{ but this basically collapses the 2-dimensional vector space to a 1-dimensional one – i.e., we ‘lose information’ and } \hat{S}^{-1} \text{ most certainly does not exist.}\]
Problem 4.28. Schrödinger’s equation & Dyson Series  

The primary equation in quantum mechanics (and quantum field theory), governing how states evolve in time, is

\[ i\hbar \partial_t |\psi(t)\rangle = H |\psi(t)\rangle, \quad (4.3.197) \]

where \( \hbar \approx 1.054572 \times 10^{-34} \text{ J s} \) is the reduced Planck’s constant, and \( H \) is the Hamiltonian (≡ 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 \( \{|E_i; n_j\}\} \) 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 \( (4.3.197) \) is

\[ |\psi(t)\rangle = \sum_{j,n_j} e^{-(i/\hbar)E_j t} |E_j; n_j\rangle \langle E_j; n_j |\psi(t = 0)\rangle, \quad (4.3.198) \]

where \( |\psi(t = 0)\rangle \) is the initial condition, i.e., the state \( |\psi(t)\rangle \) at \( t = 0 \). (Hint: Check that eq. \( (4.3.197) \) and the initial condition are satisfied.) Since the initial state was arbitrary, what you have verified is that the operator

\[ U(t,t') \equiv \sum_{j,n_j} e^{-(i/\hbar)E_j(t-t')} |E_j; n_j\rangle \langle E_j; n_j | \]

obeys Schrödinger’s equation,

\[ i\hbar \partial_t U(t,t') = HU(t,t'). \quad (4.3.200) \]

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 \( \langle \psi(t = 0) | H |\psi(t = 0)\rangle \).

What if the Hamiltonian in Schrödinger’s equation depends on time – what is the corresponding \( U \)? Consider the following (somewhat formal) solution for \( U \).

\[ U(t,t') \equiv \mathbb{I} - \frac{i}{\hbar} \int_{t'}^t d\tau_1 H(\tau_1) + \left( -\frac{i}{\hbar} \right)^2 \int_{t'}^t d\tau_2 \int_{t'}^{\tau_2} d\tau_1 H(\tau_2)H(\tau_1) + \ldots \quad (4.3.201) \]

\[ = \mathbb{I} + \sum_{\ell=1}^{\infty} \mathcal{I}_\ell(t,t'), \quad (4.3.202) \]

where the \( \ell \)-nested integral \( \mathcal{I}_\ell(t,t') \) is

\[ \mathcal{I}_\ell(t,t') \equiv \left( -\frac{i}{\hbar} \right)^\ell \int_{t'}^t d\tau_\ell \int_{t'}^{\tau_\ell} d\tau_{\ell-1} \ldots \int_{t'}^{\tau_3} d\tau_2 \int_{t'}^{\tau_2} d\tau_1 H(\tau_\ell)H(\tau_{\ell-1}) \ldots H(\tau_2)H(\tau_1). \quad (4.3.203) \]

(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 \( \tau_1 \neq \tau_2 \).) Verify that, for \( t > t' \),

\[ i\hbar \partial_t U(t,t') = H(t)U(t,t'). \quad (4.3.204) \]
What is $U(t = t')$? You should be able to conclude that $|\psi(t)\rangle = U(t, t')|\psi(t')\rangle$. Hint: Start with $i\hbar\partial_t I(t, t')$ and employ Leibniz’s rule:

$$\frac{d}{dt} \left( \int_{\alpha(t)}^{\beta(t)} F(t, z) dz \right) = \int_{\alpha(t)}^{\beta(t)} \frac{\partial F(t, z)}{\partial t} dz + F(t, \beta(t)) \beta'(t) - F(t, \alpha(t)) \alpha'(t).$$

(4.3.205)

**Bonus:** Can you prove Leibniz’s rule, by say, using the limit definition of the derivative?

**Problem 4.29.** If an operator $A$ is simultaneously unitary and Hermitian, what is $A$? Hint: Diagonalize it first.

### 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., $V_1 \otimes V_2 \otimes \cdots \otimes V_N$ to form a $V_L$. 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

$$|A; L\rangle \equiv |\alpha_1; 1\rangle \otimes |\alpha_2; 2\rangle \otimes \cdots \otimes |\alpha_N; N\rangle,$$

(4.4.1)

where it is understood the vector $|\alpha_i; i\rangle$ in the $i$th slot (from the left) is an element of the $i$th 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,

$$|\alpha\rangle \otimes (|\alpha'\rangle + |\beta'\rangle) \otimes |\alpha''\rangle = |\alpha\rangle \otimes |\alpha'\rangle \otimes |\alpha''\rangle + |\alpha\rangle \otimes |\beta'\rangle \otimes |\alpha''\rangle.$$

(4.4.2)

2. Scalar multiplication can be factored out. For example,

$$c (|\alpha\rangle \otimes |\alpha'\rangle) = (c |\alpha\rangle) \otimes |\alpha'\rangle = |\alpha\rangle \otimes (c |\alpha'\rangle).$$

(4.4.3)

Our larger vector space $V_L$ is spanned by all vectors of the form in eq. (4.4.1), meaning every vector in $V_L$ can be expressed as a linear combination:

$$|A'; L\rangle \equiv \sum_{\alpha_1, \ldots, \alpha_N} C^{\alpha_1, \ldots, \alpha_N} |\alpha_1; 1\rangle \otimes |\alpha_2; 2\rangle \otimes \cdots \otimes |\alpha_N; N\rangle \in V_L.$$  

(4.4.4)

(The $C^{\alpha_1, \ldots, \alpha_N}$ is just a collection complex numbers.) In fact, if we let $\{|i; j\rangle | i = 1, 2, \ldots, D_j\}$ be the basis vectors of the $j$th vector space $V_j$,

$$|A'; L\rangle = \sum_{\alpha_1, \ldots, \alpha_N} \sum_{i_1, \ldots, i_N} C^{\alpha_1, \ldots, \alpha_N} \langle i_1; 1 | \alpha_1 \rangle \langle i_2; 2 | \alpha_2 \rangle \cdots \langle i_N; N | \alpha_N \rangle$$
In other words, the basis vectors of this tensor product space $V_L$ are formed from products of the basis vectors from each and every vector space \{V_i\}.

**Dimension** If the $i$th vector space $V_i$ has dimension $D_i$, then the dimension of $V_L$ itself is $D_1D_2\ldots D_{N-1}D_N$. The reason is, for a given tensor product $|i_1; 1\rangle \otimes |i_2; 2\rangle \otimes \cdots \otimes |i_N; N\rangle$, 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

$$
|1; L\rangle = |+\rangle \otimes |+\rangle, \quad |2; L\rangle = |+\rangle \otimes |-\rangle, \quad (4.4.6)
$$

$$
|3; L\rangle = |-\rangle \otimes |+\rangle, \quad |4; L\rangle = |-\rangle \otimes |-\rangle. \quad (4.4.7)
$$

(Note that this ordering of the vectors is of course not unique.)

**Adjoint and Inner Product** Just as we can form tensor products of kets, we can do so for bras. We have

$$
(|\alpha_1\rangle \otimes |\alpha_2\rangle \otimes \cdots \otimes |\alpha_N\rangle)^\dagger = \langle \alpha_1 | \otimes \langle \alpha_2 | \otimes \cdots \otimes \langle \alpha_N |, \quad (4.4.8)
$$

where the $i$th slot from the left is a bra from the $i$th vector space $V_i$. We also have the inner product

$$
(\langle \alpha_1 | \otimes \langle \alpha_2 | \otimes \cdots \otimes \langle \alpha_N |) (c |\beta_1\rangle \otimes |\beta_2\rangle \otimes \cdots \otimes |\beta_N\rangle + d |\gamma_1\rangle \otimes |\gamma_2\rangle \otimes \cdots \otimes |\gamma_N\rangle)
= c \langle \alpha_1 | \beta_1 \rangle \langle \alpha_2 | \beta_2 \rangle \cdots \langle \alpha_N | \beta_N \rangle + d \langle \alpha_1 | \gamma_1 \rangle \langle \alpha_2 | \gamma_2 \rangle \cdots \langle \alpha_N | \gamma_N \rangle, \quad (4.4.9)
$$

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

$$
(\langle j_1; 1 | \otimes \cdots \otimes \langle j_N; N |) (|i_1; 1\rangle \otimes \cdots \otimes |i_N; N\rangle)
= \langle j_1; 1 | i_1; 1 \rangle \langle j_2; 2 | i_2; 2 \rangle \cdots \langle j_N; N | i_N; N \rangle
= \delta_{j_1 i_1} \cdots \delta_{j_N i_N}. \quad (4.4.10)
$$

**Linear Operators** If $X_i$ is a linear operator acting on the $i$th vector space $V_i$, we can form a tensor product of them. Their operation is defined as

$$
(X_1 \otimes X_2 \otimes \cdots \otimes X_N) (c |\beta_1\rangle \otimes |\beta_2\rangle \otimes \cdots \otimes |\beta_N\rangle + d |\gamma_1\rangle \otimes |\gamma_2\rangle \otimes \cdots \otimes |\gamma_N\rangle)
= c(X_1 |\beta_1\rangle) \otimes (X_2 |\beta_2\rangle) \otimes \cdots \otimes (X_N |\beta_N\rangle) + d(X_1 |\gamma_1\rangle) \otimes (X_2 |\gamma_2\rangle) \otimes \cdots \otimes (X_N |\gamma_N\rangle), \quad (4.4.11)
$$

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

$$
Y = \sum_{i_1, \ldots, i_N} \sum_{j_1, \ldots, j_N} (|i_1; 1\rangle \otimes \cdots \otimes |i_N; N\rangle) \hat{Y}_{i_1 \ldots i_N}^{j_1 \ldots j_N} (\langle j_1; 1 | \otimes \cdots \otimes \langle j_N; N |), \quad (4.4.12)
$$

$$
\hat{Y}_{i_1 \ldots i_N}^{j_1 \ldots j_N} \in \mathbb{C}. \quad (4.4.13)
$$
Problem 4.30. **Tensor transformations.** Consider the state

\[
|\Psi'; L\rangle = \sum_{1 \leq i_1 \leq D_1} \sum_{1 \leq i_2 \leq D_2} \cdots \sum_{1 \leq i_N \leq D_N} T^{i_1i_2\cdots i_N}_{1i_1i_2\cdots i_N} |i_1; 1\rangle \otimes |i_2; 2\rangle \otimes \cdots \otimes |i_N; N\rangle,
\]

where \(\{|i_j; j\rangle\}\) are the \(D_j\) orthonormal basis vectors spanning the \(j\)th vector space \(V_j\), and \(T^{i_1i_2\cdots i_N}_{1i_1i_2\cdots i_N}\) are complex numbers. Consider a change of basis for each vector space, i.e., \(|i; j\rangle \rightarrow |i'; j\rangle\). By defining the unitary operator that implements this change-of-basis

\[
U \equiv (1) U \otimes (2) U \otimes \cdots \otimes (N) U,
\]

\[
(i) U \equiv \sum_{1 \leq j \leq D_i} |j'; i\rangle \langle j; i|,
\]

expand \(|\Psi'; L\rangle\) in the new basis \(\{|j'_1; 1\rangle \otimes \cdots \otimes |j'_N; N\rangle\}\); this will necessarily involve the \(U^{\dagger i}\)'s. Define the coefficients of this new basis via

\[
|\Psi'; L\rangle = \sum_{1 \leq i'_1 \leq D_1} \sum_{1 \leq i'_2 \leq D_2} \cdots \sum_{1 \leq i'_N \leq D_N} T^{i_1'i_2'\cdots i_N'}_{1'i_1'i_2'\cdots i_N'} |i'_1; 1\rangle \otimes |i'_2; 2\rangle \otimes \cdots \otimes |i'_N; N\rangle.
\]

Now relate \(T^{i_1'i_2'\cdots i_N'}_{1'i_1'i_2'\cdots i_N'}\) to the coefficients in the old basis \(T^{i_1i_2\cdots i_N}_{1i_1i_2\cdots i_N}\) using the matrix elements

\[
\left((i) U^{\dagger}\right)^j_k \equiv \langle j; i| (i) U \otimes \cdots \otimes (N) U |k; i\rangle.
\]

Can you perform a similar change-of-basis for the following dual vector?

\[
\langle \Psi'; L| = \sum_{1 \leq i_1 \leq D_1} \sum_{1 \leq i_2 \leq D_2} \cdots \sum_{1 \leq i_N \leq D_N} T_{i_1i_2\cdots i_N 1} \langle i_1; 1| \otimes \langle i_2; 2| \otimes \cdots \otimes \langle i_N; N|
\]

In differential geometry, tensors will transform in analogous ways.

\[
\square
\]

Problem 4.31. **Product Rule** Suppose the collection of states \(\{|\psi_i(t)\rangle | i = 1, 2, \ldots, N\}\) depend on the real parameter \(t\). Explain why the product rule of differentiation holds for their tensor product.

\[
\partial_t \left( |\psi_1(t)\rangle \otimes |\psi_2(t)\rangle \otimes \cdots \otimes |\psi_N(t)\rangle \right)
\]

\[
= (\partial_t |\psi_1(t)\rangle) \otimes |\psi_2(t)\rangle \otimes \cdots \otimes |\psi_N(t)\rangle + |\psi_1(t)\rangle \otimes (\partial_t |\psi_2(t)\rangle) \otimes \cdots \otimes |\psi_N(t)\rangle
\]

\[
+ \cdots + |\psi_1(t)\rangle \otimes |\psi_2(t)\rangle \otimes \cdots \otimes (\partial_t |\psi_N(t)\rangle).
\]

\[
\square
\]

4.5 Continuous Spaces and Infinite \(D\)-Space

For the final section we will deal with vector spaces with continuous spectra, with infinite dimensionality. 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.
4.5.1 Preliminaries: Dirac’s \( \delta \)-“function”, eigenket integrals, and continuous (Lie group) operators

**Dirac’s \( \delta \)-“function”**  
We will see that transitioning from discrete, finite dimensional vector spaces to continuous ones means summations become integrals; while Kronecker-\( \delta \)s will be replaced with Dirac-\( \delta \) functions. In case the latter is not familiar, the Dirac-\( \delta \) function of one variable is to be viewed as an object that occurs within an integral, and is defined via

\[
\int_a^b f(x') \delta(x' - x) dx' = f(x),
\]

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-\( \delta \)-function\(^{19} \) 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

\[
\delta(x) = \lim_{\epsilon \to 0^+} \Theta\left(\frac{\epsilon}{2} - |x|\right) \frac{1}{\epsilon} \quad (4.5.2)
\]

\[
= \lim_{\epsilon \to 0^+} e^{-|x|/\epsilon} \quad (4.5.3)
\]

\[
= \lim_{\epsilon \to 0^+} \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2} \quad (4.5.4)
\]

For the first equality, \( \Theta(z) \) is the step function, defined to be

\[
\Theta(z) = 1, \quad \text{for } z > 0
\]

\[
= 0, \quad \text{for } z < 0. \quad (4.5.5)
\]

**Problem 4.32.** Justify these three definitions of \( \delta(x) \). What happens, for finite \( x \neq 0 \), when \( \epsilon \to 0^+ \)? Then, by holding \( \epsilon \) fixed, integrate them over the real line, before proceeding to set \( \epsilon \to 0^+ \).

For later use, we record the following integral representation of the Dirac \( \delta \)-function.

\[
\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega(z-z')} = \delta(z-z') \quad (4.5.6)
\]

Finally, for functions defined within the interval \( -1 < x < +1 \), the following is yet another representation of the Dirac delta function:

\[
\delta(x) = \lim_{n \to +\infty} \frac{(2n + 1)!}{2^{2n+1}(n!)^2} (1 - x^2)^n \quad (4.5.7)
\]

\(^{19}\)In the rigorous mathematical literature, Dirac’s \( \delta \) is not a function but a *distribution*, whose theory is due to Laurent Schwartz.
where \( n \) is to be viewed as an integer. We may understand eq. (4.5.7) 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. (4.5.7) ensure the total area of the right-hand-side is still unity. This representation in eq. (4.5.7) plays a central role in the Weierstrass approximation theorem which states that any continuous function within a finite interval (on the real line) may be approximated by a polynomial.

**Problem 4.33.** Can you justify the following?

\[
\Theta(z - z') = \int_{z_0}^z dz'' \delta(z'' - z'), \quad z' > z_0.
\] (4.5.8)

We may therefore assert the derivative of the step function is the \( \delta \)-function,

\[
\Theta'(z - z') = \delta(z - z').
\] (4.5.9)

Somewhat more rigorously, we may refer to the integral representation of the step function in eq. (5.3.41) below; and thereby justify its counterpart for the Dirac \( \delta \)-function in eq. (4.5.6). □

A few properties of the \( \delta \)-function are worth highlighting.

- From eq. (4.5.9) – that a \( \delta(z - z') \) follows from taking the derivative of a discontinuous function – in this case, \( \Theta(z - z') \) – will be important for the study of Green’s functions.

- If the argument of the \( \delta \)-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

\[
\delta(f(z)) = \sum_{z_i \equiv \text{th zero of } f(z)} \frac{\delta(z - z_i)}{|f'(z_i)|}.
\] (4.5.10)

To justify this we recall the fact that, the \( \delta \)-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 \( \delta \)-function near each zero. That is, what are the \( \varphi_i \)’s in

\[
\delta(f(z)) = \sum_{z_i \equiv \text{th zero of } f(z)} \frac{\delta(z - z_i)}{\varphi_i}.
\] (4.5.11)

We now use the fact that integrating a \( \delta \)-function around the small neighborhood of the \( i \)th 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 \( i \)th zero,

\[
f(z) = f'(z_i)(z - z_i) + \mathcal{O}((z - z_i)^2),
\] (4.5.12)

\[
\Rightarrow \quad df = f'(z_i)dz + \mathcal{O}((z - z_i)^1)dz.
\] (4.5.13)
The integration around the $i$th zero reads, for $0 < \epsilon \ll 1$,
\begin{equation}
1 = \int_{z=z_i-\epsilon}^{z=z_i+\epsilon} df \delta(f) = \int_{z=z_i-\epsilon}^{z=z_i+\epsilon} dz \left| \left( f'(z_i) + \mathcal{O}((z - z_i)1) \right) \right| \frac{\delta(z - z_i)}{\varphi_i} \tag{4.5.14}
\end{equation}
\begin{equation}
\epsilon \to 0 \left| f'(z_i) \right| \varphi_i \tag{4.5.15}
\end{equation}

(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. (4.5.10) is, for $a \in \mathbb{R}$,
\begin{equation}
\delta(az) = \frac{\delta(z)}{|a|}. \tag{4.5.16}
\end{equation}

- Since $\delta(z)$ is non-zero only when $z = 0$, it must be that $\delta(-z) = \delta(z)$ and more generally
\begin{equation}
\delta(z - z') = \delta(z' - z). \tag{4.5.17}
\end{equation}

- We may also take the derivative of a $\delta$-function. Under an integral sign, we may apply integration-by-parts as follows:
\begin{equation}
\int_a^b \delta'(x - x') f(x) dx = \delta(x - x') f(x) |_{x=a}^{x=b} - \int_a^b \delta(x - x') f'(x) dx = -f'(x') \tag{4.5.18}
\end{equation}

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 $\delta$-function? Turns out $\delta(\xi)$ has dimensions of $1/|\xi|$, i.e., the reciprocal of the dimension of its argument. The reason is
\begin{equation}
\int d\xi \delta(\xi) = 1 \quad \Rightarrow \quad [\xi][\delta(\xi)] = 1. \tag{4.5.19}
\end{equation}

**Problem 4.34.** We may generalize the identities in equations (4.5.10) and (4.5.16) 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:
\begin{equation}
\delta(g(z)f(z)) = \frac{\delta(f(z))}{g(z)} = \sum_{z_i \equiv \text{ith zero of } f(z)} \frac{\delta(z - z_i)}{g(z_i)|f'(z_i)|}. \tag{4.5.20}
\end{equation}

Hint: Simply apply eq. (4.5.10). □

**Continuous spectrum** Let $\Omega$ be a Hermitian operator whose spectrum is continuous; i.e., $\Omega |\omega\rangle = \omega |\omega\rangle$ with $\omega$ being a continuous parameter. If $|\omega\rangle$ and $|\omega'\rangle$ are both “unit norm” eigenvectors of different eigenvalues $\omega$ and $\omega'$, we have for example
\begin{equation}
\langle\omega|\omega'\rangle = \delta(\omega - \omega'). \tag{4.5.21}
\end{equation}
(This assumes a “translation symmetry” in this $\omega$-space; we will see later how to modify this inner product when the translation symmetry is lost.) The completeness relation in eq. (4.3.24) is given by

$$\int d\omega |\omega\rangle \langle \omega| = \mathbb{I}; \quad (4.5.22)$$

because for an arbitrary ket $|f\rangle$,

$$\langle \omega' | f \rangle = \langle \omega' | \mathbb{I} | f \rangle = \int d\omega \langle \omega' | \omega \rangle \langle \omega | f \rangle \quad (4.5.23)$$

$$= \int d\omega \delta(\omega' - \omega) \langle \omega | f \rangle. \quad (4.5.24)$$

An arbitrary vector $|\alpha\rangle$ can thus be expressed as

$$|\alpha\rangle = \int d\omega |\omega\rangle \langle \omega| \alpha\rangle. \quad (4.5.25)$$

When the state is normalized to unity, we say

$$\langle \alpha | \alpha \rangle = \int d\omega \langle \alpha | \omega \rangle \langle \omega| \alpha\rangle = \int d\omega |\langle \omega | \alpha\rangle|^2 = 1. \quad (4.5.26)$$

The inner product between arbitrary vectors $|\alpha\rangle$ and $|\beta\rangle$ now reads

$$\langle \alpha | \beta \rangle = \int d\omega \langle \alpha | \omega \rangle \langle \omega| \beta\rangle. \quad (4.5.27)$$

Since by assumption $\Omega$ is diagonal, i.e.,

$$\Omega = \int d\omega |\omega\rangle \langle \omega|, \quad (4.5.28)$$

the matrix elements of $\Omega$ are

$$\langle \omega | \Omega | \omega' \rangle = \omega \delta(\omega - \omega') = \omega' \delta(\omega - \omega'). \quad (4.5.29)$$

Because of the $\delta$-function, which enforces $\omega = \omega'$, it does not matter if we write $\omega$ or $\omega'$ 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,

$$A(\vec{\xi}) = e^{-iZ(\vec{\xi})}. \quad (4.5.30)$$
The exponential of an operator $Y$ is itself defined through the Taylor series
\[ e^Y \equiv I + Y + \frac{Y^2}{2!} + \frac{Y^3}{3!} + \cdots = \sum_{\ell=0}^{\infty} \frac{Y^\ell}{\ell!}. \tag{4.5.31} \]

For later use, note that
\[ (e^Y)^\dagger = \sum_{\ell=0}^{\infty} \frac{(Y^\dagger)^\ell}{\ell!} = e^{Y\dagger}. \tag{4.5.32} \]

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.
\[ Z = \vec{\xi} \cdot \vec{T} \equiv \xi^a T^a. \tag{4.5.33} \]

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 $\{A = \exp(-i\alpha^i T^i)\}$, 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
\[ \exp \left( -i\vec{a} \cdot \vec{T} \right) \exp \left( -i\vec{b} \cdot \vec{T} \right) = \exp \left( -i\vec{c} \cdot \vec{T} \right). \tag{4.5.34} \]

### Lie Group & Lie Algebra

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. (4.5.34) – that products of group elements yield another group element – is all we need to ensure these they indeed form a group.

The crucial property ensuring eq. (4.5.34) holds, is that the basis generators $\{T^a\}$ themselves obey a Lie algebra:
\[ [T^a,T^b] = i f^{abc} T^c \equiv i \sum_c f^{abc} T^c. \tag{4.5.35} \]

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 4.35.** Prove that the set of linear operators $\{T^a\}$ in eq. (4.5.35) 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? 

---

20If $A$ and $Y \equiv -iZ$ were complex numbers, then the $A = e^Y$ in eq. (4.5.30) 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 $Y \equiv \ln (I + (A - I)) \equiv -\sum_{\ell=1}^{\infty} (I - A)^\ell/\ell$. 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 $A - I$, is why there is a need to demand that $A$ is continuously connected to $I$ – i.e., $\ln A$ would cease to be valid if the operator norm $\|A - I\|$ is too large.

21Lie 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.
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., \([X,[X,Y]]\), \([Y,[X,Y]]\), \([X,[Y,[Y,X]]]\), etc. Because these are operators, note that \( e^X e^Y \neq e^X + e^Y \neq e^Y e^X \). In detail, the first few terms in the exponent read

\[
e^X e^Y = \exp \left( X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \ldots \right). \tag{4.5.36}
\]

(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 (4.5.33) and (4.5.34), if

\[
X = -ia_i T^i \quad \text{and} \quad Y = -ib_i T^i; \tag{4.5.37}
\]

then eq. (4.5.35) inserted into the right hand side of (4.5.36) reads

\[
-i Z = -i (a_i + b_i) T^i + \frac{1}{2} (-i)^2 a_i b_j [T^i, T^j] + \frac{1}{12} (-i)^3 a_i a_j b_k [T^i, [T^j, T^k]] - \frac{1}{12} (-i)^3 b_i a_j b_k [T^i, [T^j, T^k]] + \ldots
\]

\[
= -i \left\{ a_l + b_l + \frac{1}{2} a_i a_j f_{ijl} + \frac{1}{12} a_i a_j b_k f^{jks} f_{isl} - \frac{1}{12} b_l a_j b_k f^{jks} f_{isl} + \ldots \right\} T^l. \tag{4.5.38}
\]

From this, we may now read off – the exponent on the right hand side of eq. (4.5.34) is

\[
Z = \vec{c} \cdot \vec{T}, \tag{4.5.39}
\]

\[
c_l = a_l + b_l + \frac{1}{2} a_i a_j f_{ijl} + \frac{1}{12} a_i a_j b_k f^{jks} f_{isl} - \frac{1}{12} b_l a_j b_k f^{jks} f_{isl} + \ldots. \tag{4.5.40}
\]

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

\[
U^\dagger = (e^{-iZ})^\dagger = e^{+iZ^\dagger} = e^{iZ}. \tag{4.5.41}
\]
Therefore, since $iZ$ certainly commutes with $-iZ$,

$$U^\dagger U = e^{iZ}e^{-iZ} = e^{i(Z-Z)} = I. \quad (4.5.42)$$

On the other hand, if $U$ is unitary, we may introduce a fictitious real parameter $\epsilon$ and expand

$$\begin{aligned}
(e^{-i\epsilon Z})^\dagger e^{-i\epsilon Z} &= \mathbb{I}, \\
(\mathbb{I} + i\epsilon Z^\dagger + O(\epsilon^2)) (\mathbb{I} - i\epsilon Z + O(\epsilon^2)) &= \mathbb{I}, \\
\mathbb{I} + i\epsilon(Z^\dagger - Z) + O(\epsilon^2) &= \mathbb{I}. 
\end{aligned} \quad (4.5.43)$$

The presence of the parameter $\epsilon$ allows us to see that each order in $Z$ is independent, as we may view the product as a Taylor series in $\epsilon$. At first order, in particular, we have – as advertised –

$$Z^\dagger = Z. \quad (4.5.46)$$

At this juncture, we gather the following:

**Exp(anti-Hermitian operator) is unitary** 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.

**Example** 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

$$i\partial_t U = HU; \quad (4.5.47)$$

and the boundary condition

$$U(t = 0) = I. \quad (4.5.48)$$

We see the solution is provided by

$$U(t) = \exp(-itH). \quad (4.5.49)$$

**Symmetry & 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.

$$[A, U(\vec{\xi})] = 0. \quad (4.5.50)$$

But that implies, if $|\alpha\rangle$ is an eigenket of $A$ with eigenvalue $\alpha$, namely

$$A |\alpha\rangle = \alpha |\alpha\rangle, \quad (4.5.51)$$

74
so must \( U |\alpha\rangle \) be. For, \([A, U] = 0\) leads us to consider
\[
[A, U] |\alpha\rangle = 0,  \quad (4.5.52)
\]
\[
A(U |\alpha\rangle) = U A |\alpha\rangle = \alpha(U |\alpha\rangle).  \quad (4.5.53)
\]

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 \( U(\xi) |\alpha\rangle \) and \( |\alpha\rangle \) both correspond to eigenkets of \( A \) with the same eigenvalue \( \alpha \). To sum:

Symmetry implies degeneracy.

### 4.5.2 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
\[
\vec{X} |\vec{x}\rangle = \vec{x} |\vec{x}\rangle.  \quad (4.5.54)
\]
Or, in index notation,
\[
X^k |\vec{x}\rangle = x^k |\vec{x}\rangle, \quad k \in \{1, 2, \ldots, D\}.  \quad (4.5.55)
\]
The position eigenkets are normalized as, in Cartesian coordinates,
\[
\langle \vec{x} | \vec{x}\rangle = \delta^{(D)}(\vec{x} - \vec{x}) \equiv \prod_{i=1}^{D} \delta(x^i - x^i) = \delta(x^1 - x^1)\delta(x^2 - x^2)\ldots \delta(x^D - x^D).  \quad (4.5.56)
\]
As an important aside, the generalization of the 1D transformation law in eq. (4.5.10) involving the \( \delta \)-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
\[
\delta^{(D)}(\vec{x} - \vec{x}') = \frac{\delta^{(D)}(\vec{y} - \vec{y}')}{|\det \partial x^a(\vec{y})/\partial y^b|} = \frac{\delta^{(D)}(\vec{y} - \vec{y}')}{|\det \partial x^a(\vec{y})/\partial y^b|},  \quad (4.5.57)
\]
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 \( a \)th row and \( b \)th column is given by \( \partial x^a(\vec{y})/\partial y^b \). (The second and third equalities follow from each other because the \( \delta \)-functions allow us to assume \( \vec{y} = \vec{y}' \).) Equation (4.5.57) can be justified by demanding that its integral around the point \( \vec{x} = \vec{x}' \) gives one. For \( 0 < \epsilon \ll 1 \), and denoting \( \delta^{(D)}(\vec{x} - \vec{x}') = \delta^{(D)}(\vec{y} - \vec{y}')/\varphi(\vec{y}) \),
\[
1 = \int_{|\vec{x} - \vec{x}'| \leq \epsilon} d^D \vec{x} \delta^{(D)}(\vec{x} - \vec{x}') = \int_{|\vec{x} - \vec{x}'| \leq \epsilon} d^D \vec{y} \left| \partial x^a(\vec{y})/\partial y^b \right| \delta^{(D)}(\vec{y} - \vec{y}') = \left| \frac{\partial x^a(\vec{y})}{\partial y^b} \right| \varphi(\vec{y}).  \quad (4.5.58)
\]
Now, any vector $|\alpha\rangle$ in the Hilbert space can be expanded in terms of the position eigenkets.

$$|\alpha\rangle = \int_{\mathbb{R}^D} d^D \vec{x} \langle \vec{x} | \alpha \rangle. \quad (4.5.59)$$

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 $\int d^D \vec{x} |\langle \vec{x} | \alpha \rangle|^2 = 1$. 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 $\{|\vec{p}\rangle\}$) does not necessarily have to involve quantum theory. 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. (9.3.48), 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

$$\langle \alpha | Y | \beta \rangle = \int d^D \vec{x} \int d^D \vec{x}' \langle \alpha | \vec{x} \rangle \langle \vec{x} | Y | \vec{x}' \rangle \langle \vec{x}' | \beta \rangle = \int d^D \vec{x} \int d^D \vec{x}' \langle \vec{x} | Y | \vec{x}' \rangle \langle \vec{x}' | \beta \rangle. \quad (4.5.60)$$

If the operator $Y(\vec{X})$ were built solely from the position operator $\vec{X}$, then

$$\langle \vec{x}' | Y(\vec{X}) | \vec{x} \rangle = Y(\vec{x})\delta^{(D)}(\vec{x} - \vec{x}') = Y(\vec{x}')\delta^{(D)}(\vec{x} - \vec{x}'); \quad (4.5.61)$$

and the double integral collapses into one,

$$\langle \alpha | Y(\vec{X}) | \beta \rangle = \int d^D \vec{x} \langle \vec{x} | \alpha \rangle^* \langle \vec{x}' | \beta \rangle Y(\vec{x}). \quad (4.5.62)$$

**Problem 4.36.** Show that if $U$ is a unitary operator and $|\alpha\rangle$ is an arbitrary vector, then $|\alpha\rangle$, $U|\alpha\rangle$ and $U^+|\alpha\rangle$ have the same norm. \[\square\]

**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

$$\mathcal{T}(\vec{d}) |\vec{x}\rangle = |\vec{x} + \vec{d}\rangle. \quad (4.5.63)$$

\[\text{This is especially pertinent for those whose first contact with continuous Hilbert spaces was in the context of a quantum mechanics course.}\]
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. (4.5.63),
\[
\langle \vec{y} \rangle \mathcal{T}(\vec{d})^\dagger = \langle \vec{y} + \vec{d} \rangle .
\]
(4.5.64)
Therefore, recalling eq. (4.5.56),
\[
\langle \vec{y} \rangle \mathcal{T}(\vec{d})^\dagger \mathcal{T}(\vec{d}) |\vec{x}\rangle = \langle \vec{y} + \vec{d} \rangle \langle \vec{x} + \vec{d} \rangle = \delta^{(D)}(\vec{y} - \vec{x}) = \langle \vec{y} | \vec{x} \rangle ;
\]
(4.5.65)
and since this is true for arbitrary states $|\vec{x}\rangle$ and $|\vec{y}\rangle$,
\[
\mathcal{T}(\vec{d})^\dagger \mathcal{T}(\vec{d}) = \mathbb{I}.
\]
(4.5.66)
The inverse transformation of the translation operator is
\[
\mathcal{T}(\vec{d})^\dagger |\vec{x}\rangle = |\vec{x} - \vec{d}\rangle
\]
(4.5.67)
since
\[
\mathcal{T}(\vec{d})^\dagger \mathcal{T}(\vec{d}) |\vec{x}\rangle = \mathcal{T}(\vec{d})^\dagger |\vec{x} + \vec{d}\rangle = |\vec{x} + \vec{d} - \vec{d}\rangle = |\vec{x}\rangle .
\]
(4.5.68)
Of course we have the identity operator $\mathbb{I}$ when $\vec{d} = \vec{0}$,
\[
\mathcal{T}(\vec{0}) |\vec{x}\rangle = |\vec{x}\rangle \quad \Rightarrow \quad \mathcal{T}(\vec{0}) = \mathbb{I}.
\]
(4.5.69)
The following composition law has to hold
\[
\mathcal{T}(\vec{d}_1)\mathcal{T}(\vec{d}_2) = \mathcal{T}(\vec{d}_1 + \vec{d}_2),
\]
(4.5.70)
because translation is commutative
\[
\mathcal{T}(\vec{d}_1)\mathcal{T}(\vec{d}_2) |\vec{x}\rangle = \mathcal{T}(\vec{d}_1) |\vec{x} + \vec{d}_2\rangle = |\vec{x} + \vec{d}_2 + \vec{d}_1\rangle = |\vec{x} + \vec{d}_1 + \vec{d}_2\rangle = \mathcal{T}(\vec{d}_1 + \vec{d}_2) |\vec{x}\rangle .
\]
(4.5.71)
Problem 4.37. Translation operator is unitary. Show that
\[
\mathcal{T}(\vec{d}) = \int_{\mathbb{R}^D} d^D \vec{x}' |\vec{d} + \vec{x}'\rangle \langle \vec{x}'|
\]
(4.5.72)
satisfies eq. (4.5.63) 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. (4.5.72) with eq. (4.3.144).
Momentum operator

We now turn to demonstrate that eq. (4.5.30) can be expressed as

\[ T(\vec{d}) = \exp \left( -i \vec{d} \cdot \vec{P} \right) = \exp \left( -i d^k P_k \right). \] (4.5.73)

Since \( T(\vec{d}) \) is unitary in infinite space, the \( \vec{P} \) is Hermitian. We will call this Hermitian operator \( \vec{P} \) the momentum operator. For instance, in this exp form, eq. (4.5.70) reads

\[ \exp( -i \vec{d}_1 \cdot \vec{P} ) \exp( -i \vec{d}_2 \cdot \vec{P} ) = \exp( -i (\vec{d}_1 + \vec{d}_2) \cdot \vec{P} ) . \] (4.5.74)

To see that eq. (4.5.73) holds, we begin with an arbitrary state \( \lvert f \rangle \) and position eigenstate \( \lvert \vec{x} \rangle \).

\[ \langle \vec{x} \lvert T(\vec{d}) \lvert f \rangle = \langle \vec{x} \lvert f \rangle - i \langle \vec{x} \lvert \int d^D \vec{\xi} \frac{\partial}{\partial \vec{x}} \langle \vec{x} \lvert A \lvert \psi \rangle = \langle \vec{x} \lvert f \rangle - i \langle \vec{x} \lvert \int d^D \vec{\xi} \frac{\partial}{\partial \vec{x}} \langle \vec{x} \lvert f \rangle + O(\xi^2). \] (4.5.77)

Equating the first order in \( d \xi \) terms in equations (4.5.75) and (4.5.76), we arrive at the position space representation of the momentum operator:

\[ \langle \vec{x} \lvert P_j \lvert f \rangle = -i \partial_j \langle \vec{x} \lvert f \rangle. \] (4.5.77)

\[ \int_R d^n \vec{r} f(\vec{r} + \vec{\xi}) = \exp \left( \frac{\vec{\xi} \cdot \partial}{\partial \vec{x}} \right) \langle \vec{x} \lvert f \rangle . \] (4.5.79)

\[ \int_R d^n \vec{r} f(\vec{r} + \vec{\xi}) = \exp \left( \frac{\vec{\xi} \cdot \partial}{\partial \vec{x}} \right) \langle \vec{x} \lvert f \rangle . \] (4.5.79)

\[ \langle \vec{x} \lvert T(\vec{d}) \lvert f \rangle = \langle \vec{x} \lvert f \rangle - i \langle \vec{x} \lvert \int d^D \vec{\xi} \frac{\partial}{\partial \vec{x}} \langle \vec{x} \lvert f \rangle + O(\xi^2). \] (4.5.77)

This leads to the same conclusion in eq. (4.5.77).

---

\[ ^{23} \] Strictly speaking \( P_j \) here has dimensions of \( 1/[\text{length}] \), whereas the momentum you might be familiar with has units of \( [\text{mass} \times \text{length}/\text{time}^2] = [\text{angular momentum}]/[\text{length}] \). The reason for such nomenclature is because of its application in Quantum Mechanics.

\[ ^{24} \] If we had written \( T(\vec{d}) = \exp( -i Z(\vec{d}) ) \), i.e., without assuming \( Z = \vec{d} \cdot \vec{P} \), in eq. (4.5.76) we would arrive at \( \langle \vec{x} \lvert T(\vec{d}) \lvert f \rangle = \langle \vec{x} \lvert f \rangle - i \langle \vec{x} \lvert Z \lvert f \rangle + O(Z^2) \). But comparison with eq. (4.5.75) tells us \( Z \) must be linear in \( d \xi \). This leads to the same conclusion in eq. (4.5.77).
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:

$$\langle \vec{x} + \vec{d} | \vec{x}' + \vec{d} \rangle = \delta^{(D)} (\vec{x} - \vec{x}') = \langle \vec{x} | \vec{x}' \rangle.$$  (4.5.80)

This observation played a crucial role in the proof of the unitary character of $\mathcal{T}$ in eq. (4.5.65).

In turn, the unitary $\mathcal{T}(\vec{d}) = \exp(-i\vec{d} \cdot \vec{P})$ implies its generators $\{\vec{P}_j\}$ must be Hermitian. To reiterate:

**Symmetry, Unitarity & 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.

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

$$P_j |\vec{k}\rangle = k_j |\vec{k}\rangle.$$  (4.5.81)

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 $\mathcal{T}(d\vec{\xi})$.

$$\hat{X} \mathcal{T}(d\vec{\xi}) |\vec{x}\rangle = \hat{X} |\vec{x} + d\vec{\xi}\rangle = (\vec{x} + d\vec{\xi}) |\vec{x} + d\vec{\xi}\rangle,$$  (4.5.82)

and

$$\mathcal{T}(d\vec{\xi})\hat{X} |\vec{x}\rangle = |\vec{x} + d\vec{\xi}\rangle.$$  (4.5.83)

Since $|\vec{x}\rangle$ was an arbitrary vector, we may subtract the two equations

$$[\hat{X}, \mathcal{T}(d\vec{\xi})] |\vec{x}\rangle = d\vec{\xi} |\vec{x} + d\vec{\xi}\rangle = d\vec{\xi} |\vec{x}\rangle + \mathcal{O}(d^2).$$  (4.5.84)

At first order in $d\vec{\xi}$, we have the operator identity

$$[\hat{X}, \mathcal{T}(d\vec{\xi})] = d\vec{\xi}.$$  (4.5.85)

The left hand side involves operators, but the right hand side only real numbers. At this point we invoke eq. (4.5.73), and deduce, for infinitesimal displacements,

$$\mathcal{T}(d\vec{\xi}) = 1 - id\vec{\xi} \cdot \vec{P} + \mathcal{O}(d^2)$$  (4.5.86)
which in turn means eq. (4.5.85) now reads, as $d\vec{\xi} \to 0$,

$$[\vec{X}, -id\vec{\xi} \cdot \vec{P}] = d\vec{\xi}$$  

$$[X^i, P_j] \, d\xi^j = i\delta^i_j \, d\xi^j \quad \text{(the lth component)}$$ (4.5.87)

Since the $\{d\xi^j\}$ are independent, the coefficient of $d\xi^j$ on both sides must be equal. This leads us to the fundamental commutation relation between $k$th component of the position operator with the $j$ component of the momentum operator:

$$[X^k, P_j] = i\delta^k_j, \quad j, k \in \{1, 2, \ldots, D\}.$$ (4.5.88)

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.

$$|f\rangle = \int_{\mathbb{R}^D} d^D x' \, |x'\rangle \langle x' | f\rangle, \quad X^i \, |x'\rangle = x^i \, |x'\rangle \quad \text{ (4.5.90)}$$

$$|f\rangle = \int_{\mathbb{R}^D} d^D \vec{k}' \, |\vec{k}'\rangle \langle \vec{k}' | f\rangle, \quad P_j \, |\vec{k}'\rangle = k'_j \, |\vec{k}'\rangle \quad \text{ (4.5.91)}$$

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.

**Space-Translation of Momentum Eigenket**

Let $|\vec{k}\rangle$ be an eigenket of the momentum operator $\vec{P}$. Notice that

$$\mathcal{T}(d) \, |\vec{k}\rangle = \exp(-id \cdot \vec{P}) \, |\vec{k}\rangle = \exp(-id \cdot \vec{k}) \, |\vec{k}\rangle.$$ (4.5.92)

In words: the momentum eigenstate $|\vec{k}\rangle$ is an eigenvector of $\mathcal{T}(d)$ with eigenvalue $\exp(-id \cdot \vec{k})$.

Since this is merely a phase, in quantum mechanics, we would regard $\mathcal{T}(d) |\vec{k}\rangle$ and $|\vec{k}\rangle$ as the same physical ket: i.e., space-translation merely shifts the momentum eigenket by a phase.

**Problem 4.39. Planck’s constant: From inverse length to momentum**

Notice that eq. (4.5.77) tells us the “momentum operator” has dimension

$$[P_j] = 1/\text{Length}.$$ (4.5.93)

Explain why, to re-scale $P_j \to \kappa P_j$ to an object $\kappa P_j$ that truly has dimension of momentum, the $\kappa$ must have dimension of Planck’s (reduced) constant $\hbar$.

**Problem 4.40. 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:

$$[\mathcal{T}(\vec{d}_1), \mathcal{T}(\vec{d}_2)] = 0.$$ (4.5.94)
By considering infinitesimal displacements \( \vec{d}_1 = d\zeta_1 \) and \( \vec{d}_2 = d\zeta_2 \), show that eq. (4.5.73) leads to us to conclude that momentum operators commute among themselves,

\[
[P_i, P_j] = 0, \quad i, j \in \{1, 2, 3, \ldots, D\}. \tag{4.5.95}
\]

Comparing against eq. (4.5.35), we may conclude the structure constants occurring within the Lie algebra obeyed by the translation generators are all zero.

Problem 4.41. Check that the position representation of the momentum operator \( \vec{P} \) in eq. (4.5.77) is consistent with eq. (4.5.89) by considering

\[
\langle \vec{x} \mid [X^k, P_j] \mid \alpha \rangle = i\delta^k_j \langle \vec{x} \mid \alpha \rangle. \tag{4.5.96}
\]

Start by expanding the commutator on the left hand side, and show that you can recover eq. (4.5.77).

Problem 4.42. Express the following matrix element in the position space representation

\[
\langle \alpha \mid \vec{P} \mid \beta \rangle = \int d^D \vec{x} \left( \ ? \right). \tag{4.5.97}
\]

Problem 4.43. Show that the negative of the Laplacian, namely

\[
-\vec{\nabla}^2 \equiv -\sum_i \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^i} \quad \text{(in Cartesian coordinates \( \{x^i\} \)}, \tag{4.5.98}
\]

is the square of the momentum operator. That is, for an arbitrary state \( \mid \alpha \rangle \), show that

\[
\langle \vec{x} \mid \vec{P}^2 \mid \alpha \rangle = -\delta^{ij} \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} \langle \vec{x} \mid \alpha \rangle \equiv -\vec{\nabla}^2 \langle \vec{x} \mid \alpha \rangle. \tag{4.5.99}
\]

Problem 4.44. Prove the Campbell-Baker-Hausdorff lemma. For linear operators \( A \) and \( B \), and complex number \( \alpha \),

\[
e^{i\alpha A} B e^{-i\alpha A} = B + \sum_{\ell=1}^{\infty} \frac{(i\alpha)^\ell}{\ell!} [A, [A, \ldots [A, B]]]. \tag{4.5.100}
\]

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 \( \mid \psi \rangle \):

\[
\langle \psi \mid \vec{X} \mid \psi \rangle \quad \text{and} \quad \langle \psi \mid \vec{P} \mid \psi \rangle. \tag{4.5.101}
\]

What happens to these expectation values when we replace \( \mid \psi \rangle \to T(\vec{d}) \mid \psi \rangle \)?
Fourier analysis  We will now show how the concept of a Fourier transform readily arises from the formalism we have developed so far. To initiate the discussion we start with eq. (4.5.77), 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.

$$\langle \vec{x} | \vec{P} | \vec{k}\rangle = \vec{k} \langle \vec{x} | \vec{k}\rangle = -i \frac{\partial}{\partial \vec{x}} \langle \vec{x} | \vec{k}\rangle, \quad \Leftrightarrow \quad k_j \langle \vec{x} | \vec{k}\rangle = -i \frac{\partial}{\partial x^j} \langle \vec{x} | \vec{k}\rangle. \quad (4.5.102)$$

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. (4.5.102) is simply the plane wave

$$\langle \vec{x} | \vec{k}\rangle = \chi \exp \left( i \vec{k} \cdot \vec{x} \right). \quad (4.5.103)$$

where $\chi$ is complex constant to be fixed in the following way. We want

$$\int_{\mathbb{R}^D} d^Dk \langle \vec{x} | \vec{k}\rangle \langle \vec{k} | \vec{x}' \rangle = \langle \vec{x} | \vec{x}' \rangle = \delta^{(D)}(\vec{x} - \vec{x}'). \quad (4.5.104)$$

Using the plane wave solution,

$$(2\pi)^D|\chi|^2 \int \frac{d^Dk}{(2\pi)^D} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} = \delta^{(D)}(\vec{x} - \vec{x}'). \quad (4.5.105)$$

Now, recall the representation of the $D$-dimensional $\delta$-function

$$\int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} = \delta^{(D)}(\vec{x} - \vec{x}'). \quad (4.5.106)$$

Therefore, up to an overall multiplicative phase $e^{i\delta}$, which we will choose to be unity, $\chi = 1/(2\pi)^{D/2}$ and eq. (4.5.103) becomes

$$\langle \vec{x} | \vec{k}\rangle = (2\pi)^{-D/2} \exp \left( i \vec{k} \cdot \vec{x} \right). \quad (4.5.107)$$

By comparing eq. (4.5.107) with eq. (4.3.145), we see that the plane wave in eq. (4.5.107) 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.

$$\langle \vec{x} | f \rangle = \int_{\mathbb{R}^D} d^Dk \langle \vec{x} | \vec{k}\rangle \langle \vec{k} | f \rangle = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^{D/2}} e^{i\vec{k} \cdot \vec{x}} \langle \vec{k} | f \rangle \quad (4.5.108)$$

Similarly, we may expand the momentum representation of an arbitrary state $\langle \vec{k} | f \rangle$ in the position eigenbasis.

$$\langle \vec{k} | f \rangle = \int_{\mathbb{R}^D} d^D\vec{x} \langle \vec{k} | \vec{x}\rangle \langle \vec{x} | f \rangle = \int_{\mathbb{R}^D} \frac{d^D\vec{x}}{(2\pi)^{D/2}} e^{-i\vec{k} \cdot \vec{x}} \langle \vec{x} | f \rangle \quad (4.5.109)$$
Equations (4.5.108) and (4.5.109) are nothing but the Fourier expansion of some function $f(\vec{x})$ and its inverse transform. We may sum up the discussion here with the following expansions:

$$|\vec{x}\rangle = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^{D/2}} e^{-i\vec{k} \cdot \vec{x}} |\vec{k}\rangle,$$

$$|\vec{k}\rangle = \int_{\mathbb{R}^D} \frac{d^Dx}{(2\pi)^{D/2}} e^{i\vec{k} \cdot \vec{x}} |\vec{x}\rangle. \quad (4.5.110)$$

**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. (4.5.108) as the orthonormal basis vectors. The coefficients of the expansion are then the $\tilde{f}(\vec{k}) \equiv \langle \vec{k}|f \rangle$.

$$f(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^{D/2}} e^{i\vec{k} \cdot \vec{x}} \tilde{f}(\vec{k}). \quad (4.5.112)$$

By multiplying both sides by $\exp(-i\vec{k}' \cdot \vec{x})/(2\pi)^{D/2}$, integrating over all space, using the integral representation of the $\delta$-function in eq. (4.5.6), and finally replacing $\vec{k}' \rightarrow \vec{k}$,

$$\tilde{f}(\vec{k}) = \int_{\mathbb{R}^D} \frac{d^Dx}{(2\pi)^{D/2}} e^{-i\vec{k} \cdot \vec{x}} f(\vec{x}). \quad (4.5.113)$$

**Problem 4.45.** Prove that, for the eigenstate of momentum $|\vec{k}\rangle$, arbitrary states $|\alpha\rangle$ and $|\beta\rangle$,

$$\langle \vec{k} | \vec{X} | \alpha \rangle = i \frac{\partial}{\partial \vec{k}} \langle \vec{k} | \alpha \rangle \quad (4.5.114)$$

$$\langle \beta | \vec{X} | \alpha \rangle = \int d^Dk \langle \vec{k} | \beta \rangle^* i \frac{\partial}{\partial \vec{k}} \langle \vec{k} | \alpha \rangle. \quad (4.5.115)$$

The $\vec{X}$ is the position operator.

**Uncertainty relation** According to (4.3.102) and (4.5.89), if we work in 1 dimension for now,

$$\langle \psi | \Delta X^2 | \psi \rangle \langle \psi | \Delta P^2 | \psi \rangle \geq \frac{1}{4}. \quad (4.5.116)$$

In $D$—spatial dimensions, we may consider

$$\langle \psi | \Delta \vec{X}^2 | \psi \rangle \langle \psi | \Delta \vec{P}^2 | \psi \rangle = \sum_{1 \leq i,j \leq D} \langle \psi | (\Delta X^i)^2 | \psi \rangle \langle \psi | (\Delta P_j)^2 | \psi \rangle. \quad (4.5.117)$$

\(^{25}\) A warning on conventions: everywhere else in these notes, our Fourier transform conventions will be $\int d^Dk/(2\pi)^D$ for the momentum integrals and $\int d^Dx$ for the position space integrals. This is just a matter of where the $(2\pi)s$ are allocated, and no math/physics content is altered.
We may apply eq. (4.3.102) on each term in the sum – i.e., identify \( \Delta X^i \) (here) ↔ \( \Delta Y \) (4.3.102) and \( \Delta P_j \) here ↔ \( \Delta Z \) (4.3.102) – to deduce,

\[
\langle \psi | \Delta X^2 | \psi \rangle \langle \psi | \Delta P^2 | \psi \rangle \geq \frac{1}{4} \sum_{1 \leq i,j \leq D} | \langle \psi | [X^i, P^j] | \psi \rangle |^2 = \sum_{i,j} \frac{\delta^d_i}{4} = \frac{D}{4}. \tag{4.5.118}
\]

This is the generalization of eq. (4.5.116) to \( D \)-dimensions.

**Problem 4.46. Gaussian states & Uncertainty Relations**

Consider the function, with \( d > 0 \),

\[
\langle \vec{x} | \psi \rangle = (\sqrt{\pi d})^{-D/2} e^{i\vec{k} \cdot \vec{x}} \exp \left(-\frac{x^2}{2d^2}\right). \tag{4.5.119}
\]

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:

\[
\langle \psi | \vec{X} | \psi \rangle, \quad \langle \psi | \vec{X}^2 | \psi \rangle, \quad \langle \psi | \vec{P} | \psi \rangle, \quad \langle \psi | \vec{P}^2 | \psi \rangle. \tag{4.5.120}
\]

What is the value of

\[
\left( \langle \psi | \vec{X}^2 | \psi \rangle - \langle \psi | \vec{X} | \psi \rangle \right)^2 \left( \langle \psi | \vec{P}^2 | \psi \rangle - \langle \psi | \vec{P} | \psi \rangle \right)^2? \tag{4.5.121}
\]

Hint: In this problem you will need the following results

\[
\int_{-\infty}^{+\infty} dx e^{-ax^2} = \sqrt{\pi} a^{-1/2}, \quad a > 0, y \in \mathbb{R}. \tag{4.5.122}
\]

If you encounter an integral of the form

\[
\int_{\mathbb{R}^D} d^D \vec{x} e^{-a \vec{x}^2} e^{i\vec{q} \cdot \vec{x}}, \quad a > 0, \tag{4.5.123}
\]

you should try to combine the exponents and “complete the square”. Do you find that the uncertainty relation in eq. (4.5.118) to be saturated?

**Problem 4.47. Free Particle in Quantum Mechanics**

The free particle in Quantum Mechanics is described by the Hamiltonian

\[
H = \frac{\vec{P}^2}{2m}, \tag{4.5.124}
\]

where \( m > 0 \) is the particle’s mass. Remember the Schrödinger equation is \( i\hbar \partial_t |\psi\rangle = H |\psi\rangle \).

Show that the solution in momentum space is,

\[
\langle \vec{k} | \psi(t) \rangle = \exp \left(-\frac{i}{\hbar} \vec{k}^2 (t-t_0)\right), \quad \vec{k}^2 \equiv \vec{k} \cdot \vec{k}, \tag{4.5.125}
\]

84
if the initial condition is $\langle \vec{k} | \psi(t = t_0) \rangle = 1$. What is the corresponding position space solution $\langle \vec{x} | \psi(t) \rangle$? Hint: Start by applying $\langle \vec{k} |$ on the Schrödinger equation:

$$\langle \vec{k} | i\hbar \partial_t | \psi(t) \rangle = \langle \vec{k} | H | \psi(t) \rangle,$$

and note that, since $\langle \vec{k} |$ is time-independent, we have $\langle \vec{k} | i\hbar \partial_t | \psi(t) \rangle = i\hbar \partial_t \langle \vec{k} | \psi(t) \rangle$. □

**Translation in momentum space**

We have discussed how to implement translation in position space using the momentum operator $\vec{P}$, namely

$$T(\vec{d}) = \exp(-i\vec{d} \cdot \vec{P}).$$

What would be the corresponding translation operator in momentum space? That is, what is $\tilde{T}$ such that $\tilde{T}(\vec{d}) | \vec{k} \rangle = | \vec{k} + \vec{d} \rangle$, $P_j | \vec{k} \rangle = k_j | \vec{k} \rangle$?

Of course, one representation would be the analog of eq. (4.5.72):

$$\tilde{T}(\vec{d}) = \int_{\mathbb{R}^D} d^D \vec{k}' \langle \vec{k}' + \vec{d} | \vec{k} \rangle = \int_{\mathbb{R}^D} d^D \vec{x}' \langle \vec{x}' + \vec{d} | \vec{k} \rangle.$$  

(4.5.128)

But is there an exponential form, like there is one for the translation in position space (eq. (4.5.73))? We start with the observation that the momentum eigenstate $| \vec{k} \rangle$ can be written as a superposition of the position eigenkets using eq. (4.5.107),

$$| \vec{k} \rangle = \int_{\mathbb{R}^D} d^D \vec{x}' \langle \vec{x}' | \vec{k} \rangle = \int_{\mathbb{R}^D} \frac{d^D \vec{x}'}{(2\pi)^{D/2}} e^{ik \cdot \vec{x}'} | \vec{x}' \rangle.$$  

(4.5.129)

Now consider

$$\exp(+i\vec{d} \cdot \vec{X}) | \vec{k} \rangle = \int_{\mathbb{R}^D} \frac{d^D \vec{x}'}{(2\pi)^{D/2}} e^{ik \cdot \vec{x}'} e^{i\vec{d} \cdot \vec{x}'} | \vec{x}' \rangle = \int_{\mathbb{R}^D} \frac{d^D \vec{x}'}{(2\pi)^{D/2}} e^{i(\vec{k} + \vec{d}) \cdot \vec{x}'} | \vec{x}' \rangle = | \vec{k} + \vec{d} \rangle.$$  

(4.5.130)

That means

$$\tilde{T}(\vec{d}) = \exp \left( i\vec{d} \cdot \vec{X} \right).$$  

(4.5.131)

**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 $\{ \langle \vec{x} | \vec{k} \rangle \equiv \exp(i\vec{k} \cdot \vec{x})/(2\pi)^{D/2} \}$.

1. They can be viewed as the orthonormal basis vectors (in the $\delta$-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 $U | \vec{x} \rangle = | \vec{k} \rangle$.

---

26This question was suggested by Jake Leistico, who also correctly guessed the essential form of eq. (4.5.131).
3. They are simultaneous eigenstates of the momentum operators \{-i\partial_j \equiv -i\partial/\partial x^j | j = 1, 2, \ldots, D\} and the negative Laplacian \(-\vec{\nabla}^2\) in the position representation.

\[-\vec{\nabla}^2 (\vec{x}| \vec{k}) = \vec{k}^2 (\vec{x}| \vec{k}), \quad -i\partial_j (\vec{x}| \vec{k}) = k_j (\vec{x}| \vec{k}), \quad \vec{k}^2 \equiv \delta^{ij} k_i k_j.\] (4.5.132)

The eigenvector/value equation for the momentum operators had been solved previously in equations (4.5.102) and (4.5.103). For the negative Laplacian, we may check

\[-\vec{\nabla}^2 \vec{x} \langle \vec{x}| \vec{k} \rangle = \vec{k}^2 \langle \vec{x}| \vec{k} \rangle, \quad -i\partial_j \langle \vec{x}| \vec{k} \rangle = k_j \langle \vec{x}| \vec{k} \rangle, \quad \vec{k}^2 \equiv \delta^{ij} k_i k_j.\] (4.5.133)

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 \textit{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

\[\vec{k}^2 \equiv \sum_{j=1}^{D} (k_j)^2 = k^2, \quad \vec{P}^2 |k^2; k_1 \ldots k_D\rangle = k^2 |k^2; k_1 \ldots k_D\rangle.\] (4.5.134)

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. (4.5.92) 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 \([\mathcal{T}(\vec{d}), \vec{P}^2] = 0\) and is, in turn, a consequence of translation symmetry of the underlying flat space.

### 4.5.3 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,

\[\langle \psi_1 | \vec{P}^2 | \psi_2 \rangle = \langle \psi_1 | (\vec{P}^2)^\dagger | \psi_2 \rangle = \langle \psi_2 | \vec{P}^2 | \psi_1 \rangle^*,\] (4.5.135)

for any states \(|\psi_{1,2}\rangle\). Inserting a complete set of position eigenkets, and using

\[\langle \vec{x}| \vec{P}^2 | \psi_{1,2} \rangle = -\vec{\nabla}^2_{\vec{x}} \langle \vec{x}| \psi_{1,2} \rangle,\] (4.5.136)

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 \(\psi_1\) or \(\psi_2\).

\[\int_{\mathcal{D}} d^D x \langle \psi_1 | \vec{x} \rangle \langle \vec{x}| \vec{P}^2 | \psi_2 \rangle \equiv \int_{\mathcal{D}} d^D x \langle \psi_2 | \vec{x} \rangle^* \langle \vec{x}| \vec{P}^2 | \psi_1 \rangle^*,\]
\[
\int_{\mathcal{D}} d^Dx \psi_1(\vec{x})^* \left( -\nabla_x^2 \psi_2(\vec{x}) \right) \overset{?}{=} \int_{\mathcal{D}} d^Dx \left( -\nabla_x^2 \psi_1(\vec{x})^* \right) \psi_2(\vec{x}), \quad \psi_{1,2}(\vec{x}) \equiv \langle \vec{x} | \psi_{1,2} \rangle. \tag{4.5.137}
\]

Notice we have to specify a domain \(\mathcal{D}\) to perform the integral. If we now proceed to work from the left hand side, and use Gauss’ theorem from vector calculus,

\[
\begin{align*}
\int_{\mathcal{D}} d^Dx \psi_1(\vec{x})^* \left( -\nabla_x^2 \psi_2(\vec{x}) \right) &= \int_{\partial\mathcal{D}} d^{D-1}\Sigma \cdot \left( -\nabla \psi_1(\vec{x})^* \right) \psi_2(\vec{x}) + \int_{\mathcal{D}} d^Dx \nabla \psi_1(\vec{x})^* \cdot \nabla \psi_2(\vec{x}) \\
&= \int_{\partial\mathcal{D}} d^{D-1}\Sigma \cdot \left\{ \left( -\nabla \psi_1(\vec{x})^* \right) \psi_2(\vec{x}) + \psi_1(\vec{x})^* \nabla \psi_2(\vec{x}) \right\} \\
&\quad + \int_{\mathcal{D}} d^Dx \psi_1(\vec{x})^* \left( -\nabla_x^2 \psi_2(\vec{x}) \right) \tag{4.5.138}
\end{align*}
\]

Here, \(d^{D-1}\Sigma\) is the \((D-1)\)-dimensional analog of the 2D infinitesimal area element \(d\vec{A}\) in vector calculus, and is proportional to the unit (outward) normal \(\vec{n}\) to the boundary of the domain \(\partial\mathcal{D}\). We see that integrating-by-parts the \(\vec{P}^2\) from \(\psi_1\) onto \(\psi_2\) 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 \(\{\vec{n} \cdot \nabla \psi_\lambda\}\) to be zero:

\[
\psi_\lambda(\partial\mathcal{D}) = 0 \text{ (Dirichlet) } \quad \text{or} \quad \vec{n} \cdot \nabla \psi_\lambda(\partial\mathcal{D}) = 0 \text{ (Neumann).} \tag{4.5.139}
\]

**No boundaries**  The exception to the requirement for boundary conditions, is when the domain \(\mathcal{D}\) itself has no boundaries – there will then be no “surface terms” to speak of, and the 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 \(\mathcal{D}\) 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 \(\partial\mathcal{D}\).

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 \(i\)th Cartesian axis, the box is of length \(L^i\). If we demand that the eigenfunctions of \(-\nabla^2\) vanish at the boundary of the box, we find the eigensystem

\[
-\nabla^2 \langle \vec{x} | \vec{n} \rangle = \lambda(\vec{n}) \langle \vec{x} | \vec{n} \rangle, \quad \langle \vec{x}; x^i = 0 | \vec{n} \rangle = \langle \vec{x}; x^i = L^i | \vec{n} \rangle = 0, \tag{4.5.140}
\]

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 \(\vec{J} = \psi^* \nabla \psi - \psi \nabla \psi^*\) within the surface integral of eq. \((4.5.138)\) is proportional to the probability current. Choosing the right boundary conditions to set \(\vec{J} = 0\), so as to guarantee the hermiticity of \(\vec{P}^2\), then amounts to, in this limit, ensuring there is zero flow of probability outside the domain \(\mathcal{D}\) under consideration.
admits the solution
\[
\langle \vec{x} | \vec{n} \rangle \propto \prod_{i=1}^{D} \sin \left( \frac{\pi n_i}{L_i} x_i \right), \quad \lambda(\vec{n}) = \sum_{i=1}^{D} \left( \frac{\pi n_i}{L_i} \right)^2.
\]

These \( \{n_i\} \) runs over the positive integers only; because sine is an odd function, the negative integers do not yield new solutions.

**Problem 4.48.** Verify that the basis eigenkets in eq. (4.5.142) do solve eq. (4.5.140). What is the correct normalization for \( \langle \vec{x} | \vec{n} \rangle \)? Also verify that the basis plane waves in eq. (4.5.159) satisfy the normalization condition in eq. (4.5.158).

Use these \( \{|\vec{n}\rangle\} \) to solve the free particle Schrödinger equation:
\[
\langle \vec{n} | \psi(t) \rangle = \begin{pmatrix} \vec{P}^2 / (2m) \end{pmatrix} \langle \vec{n} | \psi(t) \rangle
\]

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 \( \mathcal{D} \); 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,
\[
f(\vec{x} + \vec{d}) = \exp \left( \vec{d} \cdot \vec{\partial} \right) \langle \vec{x} | f \rangle
\]

still holds, as long as both \( \vec{x} \) and \( \vec{x} + \vec{d} \) lie within \( \mathcal{D} \). This means equations (4.5.73), (4.5.77), and (4.5.79) 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. (4.5.72):
\[
\mathcal{T}(d > 0) \equiv \int_{0}^{L} dz' |z' + d\rangle \langle z'|
\]

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. (4.5.145) would not make sense. Likewise, for \( d < 0 \) and \( z' = 0 \), the \( |d\rangle \) in eq. (4.5.145) would, too, be non-existent. More generally, for \( d > 0 \), the bras \( \langle z' > L - d \rangle \) and ket \( \{ |z' > L - d \} \) when translated by \( d \),
\[
\langle z' | \to \langle z' + d | \quad \text{and} \quad |z' \rangle \to |z' + d \rangle
\]

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 $T(\xi) = \exp(-i\xi \cdot P)$ 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. (4.5.140) and (4.5.142), we recall the eigensystem relation

$$\langle z | P^2 | n \rangle = \left( \frac{\pi n}{L} \right)^2 \langle z | n \rangle ;$$

(4.5.147)

but on the other hand,

$$\langle z | P | n \rangle \propto \frac{\pi n}{L} \cdot \cos \left( \frac{\pi n}{L} z \right),$$

(4.5.148)

which is not proportional to $\langle z | n \rangle \propto \sin \left( \frac{\pi n}{L} x \right)$ – 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 $D$ 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. (4.5.73) 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. (4.5.145) – 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 $\varepsilon < \min(a, b)$, 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

$$T(\varepsilon | a, b) \equiv \int_a^b dz' | z' + \varepsilon \rangle \langle z' | .$$

(4.5.149)

For an arbitrary position eigenket $| z \rangle$, we may compute

$$T(\varepsilon | a, b) | z \rangle = \int_a^b dz' | z' + \varepsilon \rangle \delta(z' - z)$$

(4.5.150)

$$= \Theta(z - a)\Theta(b - z) | z + \varepsilon \rangle .$$

(4.5.151)

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 $\delta$-function within the integral of eq. (4.5.150)
is zero unless $z' = z$ and $a < z' < b$ simultaneously. Therefore, as expected, eq. (4.5.150) 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 $\varepsilon$.

**Problem 4.49. Local ‘Identity’ and ‘Unitary’-Translation Operators**

Explain why, for $0 < a < b < L$,

$$\mathbb{I}(a, b) \equiv \int_a^b dz' \ket{z'} \bra{z'}$$

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. (4.5.150) obeys

$$\mathcal{T}(\varepsilon|a, b)^\dagger \mathcal{T}(\varepsilon|a, b) = \mathbb{I}(a + \varepsilon, b + \varepsilon).$$

(4.5.153)

Since $\mathbb{I}(a + \varepsilon, b + \varepsilon)$ is the identity on the interval $(a + \varepsilon, b + \varepsilon)$, eq. (4.5.153) 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,

$$\langle \vec{x}; x^i \rightarrow x^i + L^i \mid f \rangle = \langle \vec{x}; x^i \mid f \rangle,$$

$$f(x^1, \ldots, x^i + L^i, \ldots, x^D) = f(x^1, \ldots, x^i, \ldots, x^D) = f(\vec{x}),$$

(4.5.154) (4.5.155)

this would mean, not all the basis plane waves from eq. (4.5.107) remains in the Hilbert space. Instead, periodicity means

$$\langle \vec{x}; x^j = x^j + L^j \mid \vec{k} \rangle = \langle \vec{x}; x^j = x^j \mid \vec{k} \rangle$$

$$e^{i k_j (x^j + L^j)} = e^{i k_j x^j}, \quad \text{(No sum over } j\text{.)}$$

(4.5.156)

(The rest of the plane waves, $e^{i k_l x^l}$ for $l \neq j$, cancel out of the equation.) This further implies the eigenvalue $k_j$ becomes discrete:

$$e^{i k_j L^j} = 1 \quad \text{(No sum over } j\text{.)} \quad \Rightarrow \quad k_j L^j = 2\pi n \quad \Rightarrow \quad k_j = \frac{2\pi n^j}{L^j};$$

$$n^j = 0, \pm 1, \pm 2, \pm 3, \ldots$$

(4.5.157)

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.

$$\int_{\{0 \leq x^i \leq L^i \mid i = 1, 2, \ldots, D\}} d^D \vec{x} \langle \vec{n}' \mid \vec{x} \rangle \langle \vec{x} \mid \vec{n} \rangle = \delta_{\vec{n}' \vec{n}} \equiv \prod_{i=1}^D \delta_{n_i'}.$$  

(4.5.158)

The set of orthonormal eigenvectors of the negative Laplacian may be taken as

$$\langle \vec{x} \mid \vec{n} \rangle \equiv \prod_{j=1}^D \frac{\exp \left( i \frac{2\pi n^j x^j}{L^j} \right)}{\sqrt{L^j}},$$

(4.5.159)
\[-\vec{\nabla}^2 \langle \vec{x} | \vec{n} \rangle = \lambda(\vec{n}) \langle \vec{x} | \vec{n} \rangle, \quad \lambda(\vec{n}) = \sum_i \left( \frac{2\pi n^i}{L^i} \right)^2. \] (4.5.160)

Notice the basis vectors in eq. (4.5.159) are momentum eigenkets too:

\[\langle \vec{x} | P_j | \vec{n} \rangle = -i \partial_j \langle \vec{x} | \vec{n} \rangle = k_i \langle \vec{x} | \vec{n} \rangle, \quad k_i(n^i) = \frac{2\pi n^i}{L^i}. \] (4.5.161)

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:

\[\langle \vec{x} | \vec{x}^\prime \rangle = D(D) - \iota \frac{2\pi n}{L^i} x^i\] (4.5.162)

We may also use the position eigenkets themselves to write \(I\), but instead of integrating over all space, we only integrate over one domain (since space is now periodic):

\[I = \int_{\{0 \leq x^i \leq L^i | i = 1, 2, ... D\}} d^D \vec{x} | \vec{x} \rangle \langle \vec{x} |. \] (4.5.164)

To summarize our discussion here: any periodic function \(f\), subject to eq. (4.5.155), can be expanded as a superposition of periodic plane waves in eq. (4.5.159),

\[f(\vec{x}) = \sum_{n^1 = -\infty}^{\infty} \cdots \sum_{n^D = -\infty}^{\infty} \tilde{f}(n^1, \ldots, n^D) \prod_{j=1}^{D} (L^j)^{-1/2} \exp \left( i \frac{2\pi n^j}{L^j} x^j \right). \] (4.5.165)

This is known as the Fourier series. By using the inner product in eq. (4.5.158), or equivalently, multiplying both sides of eq. (4.5.165) 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

\[\tilde{f}(n^1, n^2, \ldots, n^D) = \int_{0 \leq x^i \leq L^i} d^D \vec{x} \tilde{f}(\vec{x}) \prod_{j=1}^{D} (L^j)^{-1/2} \exp \left( -i \frac{2\pi n^j}{L^j} x^j \right). \] (4.5.166)

**Remark I** The \(\exp\) in eq. (4.5.159) 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 \(i\)th Cartesian coordinate from 0 to \(L^i\) in eq. (4.5.166), since the function is periodic, we really just need only to integrate over a complete period, from \(\kappa\) to \(\kappa + L^i\) (for \(\kappa\) real), to achieve the same result. For example, in 1D, and whenever \(f(x)\) is periodic (with a period of \(L\)),

\[\int_{0}^{L} dx f(x) = \int_{\kappa}^{\kappa + L} dx f(x). \] (4.5.167)

(Drawing a plot here may help to understand this statement.)
Problem 4.50. Translation operator in ket-bra form

Construct the translation operator in the ket-bra form, analogous to eq. \((4.5.72)\). Verify that the translation operator in a periodic space is unitary. Can you explain why it is so – in words?

From eq. \((4.5.163)\) we see that any state \(|f\rangle\) may be expanded as

\[
|f\rangle = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} |\vec{n}\rangle \langle \vec{n}| f\rangle. \tag{4.5.168}
\]

If we apply the translation operator directly to this, eq. \((4.5.161)\) tells us \(T(\vec{d}) |f\rangle\) is

\[
e^{-i\vec{d} \cdot \vec{P}} |f\rangle = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} e^{-i\vec{d} \cdot \vec{k}(\vec{n})} |\vec{n}\rangle \langle \vec{n}| f\rangle
= \left( \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \exp \left( -i\vec{d} \cdot \vec{k}(\vec{n}) \right) \right) |f\rangle. \tag{4.5.169}
\]

Since \(|f\rangle\) was arbitrary, we have managed to uncover the diagonal form of the translation operator:

\[
T(\vec{d}) = \sum_{n^1=-\infty}^{\infty} \cdots \sum_{n^D=-\infty}^{\infty} \exp \left( -i\vec{d} \cdot \vec{k}(\vec{n}) \right) |\vec{n}\rangle \langle \vec{n}|, \quad k_j = \frac{2\pi n^j}{L_j}. \tag{4.5.170}
\]

4.5.4 Rotations in \(D = 2\) Spatial Dimensions

In this section we will further develop linear operators in continuous vector spaces by extending our discussion in §4.5.2 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

\[
\vec{x}(r, \phi) \equiv r (\cos \phi, \sin \phi), \quad r \geq 0, \ 0 \leq \phi < 2\pi, \tag{4.5.172}
\]

\[
= r \cos \phi \hat{e}_1 + r \sin \phi \hat{e}_2; \tag{4.5.173}
\]

\[
\hat{e}_i = \delta_i^1, \quad i, I \in \{1, 2\}. \tag{4.5.174}
\]

We may also express this result through the following matrix that implements counter-clockwise rotation by \(\phi\):

\[
\hat{R}(\phi) \equiv \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}. \tag{4.5.175}
\]

Specifically, we see that (cf. \((4.5.172)\))

\[
\hat{R}(\phi) \hat{e}_1 = \vec{x}(r, \phi)/r, \tag{4.5.176}
\]

\[
\hat{R}(\phi) \hat{e}_2 = -\sin \phi \hat{e}_1 + \cos \phi \hat{e}_2. \tag{4.5.177}
\]
Problem 4.51. Verify through a direct calculation that

\[ \hat{R}(\phi)\hat{R}(\phi') = \hat{R}(\phi + \phi') \quad . \quad (4.5.178) \]

Draw a picture and explain what this result means.

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:

\[ D(\phi) | \psi \rangle = | \psi + \phi \rangle \quad ; \quad (4.5.179) \]

with the identity

\[ \mathbb{I} | \psi \rangle \equiv D(\phi = 0) | \psi \rangle = | \psi \rangle . \quad (4.5.180) \]

We will assume the periodic boundary condition

\[ | \psi + 2\pi \rangle = | \psi \rangle . \quad (4.5.181) \]

We will normalize these position eigenstates such that

\[ \langle \phi | \phi' \rangle = \delta(\phi - \phi') . \quad (4.5.182) \]

Starting from the definition in eq. (4.5.179), we may first check that it obeys the same product rule as the rotation matrix in eq. (4.5.178):

\[ D(\phi)D(\phi') | \psi \rangle = D(\phi) | \psi + \phi' \rangle = | \psi + \phi' + \phi \rangle \quad ; \quad (4.5.183) \]

\[ = D(\phi + \phi') | \psi \rangle . \quad (4.5.184) \]

Since addition is associative and commutative, i.e., \( | \psi + \phi \rangle = | \phi + \psi \rangle \) and \( | \psi + \phi + \phi' \rangle = | \psi + \phi' + \phi \rangle \), we see that rotation on 2D is associative and commutative – in accordance to our intuition. To sum:

\[ D(\phi)D(\phi') = D(\phi')D(\phi) = D(\phi + \phi') . \quad (4.5.185) \]

**Unitary** Acting on this abstract vector space of position eigenstates, the rotation operator in eq. (4.5.179) is unitary. To see this, we take the adjoint of \( D(\phi) | \psi' \rangle = | \psi' + \phi \rangle \).

\[ \langle \psi' | D(\phi) \rangle^\dagger = \langle \psi' + \phi | . \quad (4.5.186) \]

Combining equations (4.5.179) and (4.5.186) hands us

\[ \langle \psi' | D(\phi) \rangle^\dagger D(\phi) | \psi \rangle = \langle \psi' + \phi | \psi + \phi \rangle = \delta(\psi' - \psi) = \langle \psi' | \psi \rangle . \quad (4.5.187) \]

But since \( \langle \psi' | \) and \( | \psi \rangle \) are arbitrary, we must have

\[ D(\phi)^\dagger D(\phi) = \mathbb{I} . \quad (4.5.188) \]
Problem 4.52. Can you argue that
\[ D(\phi)^\dagger = D(\phi)^{-1} = D(-\phi)? \] (4.5.189)

Hint: We just proved the second equality. The third can be gotten by acting on an arbitrary state.

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
\[ D(\phi) = e^{-i\phi J} \] (4.5.190)
where the \( J \) is the generator of rotation (aka angular momentum operator). We may readily verify the multiplication rule in eq. (4.5.185),
\[ D(\phi)D(\phi') = e^{-i\phi J}e^{-i\phi' J} = e^{-i(\phi+\phi')J} = D(\phi+\phi') \] (4.5.191)
since \(-i\phi J \) and \(-i\phi' J \) commutes; i.e., \([\phi J,\phi' J]=0\) and refer to eq. (??). Since we just proved that \( D(\phi) \) is unitary (cf. eq. (4.5.188)) it must be that \( J \) is Hermitian. Firstly, note that
\[ (e^X)^\dagger = \sum_{\ell=0}^{+\infty} \frac{(X^\ell)^\dagger}{\ell!} = \sum_{\ell=0}^{+\infty} \frac{(X^\dagger)^\ell}{\ell!} = e^{X^\dagger}. \] (4.5.192)
Utilizing eq. (4.5.32), we see by Taylor expanding eq. (4.5.188) up to first order in \( \phi \) bring us
\[ (e^{-i\phi J})^\dagger e^{-i\phi J} = e^{i\phi J} e^{-i\phi J} = (I + i\phi J^\dagger + \ldots)(I - i\phi J + \ldots) \]
\[ = I + i\phi (J^\dagger - J) + \ldots = I. \] (4.5.194)
The coefficient of \( \phi \) must therefore vanish and we have
\[ J^\dagger = J. \] (4.5.195)

Problem 4.53. Argue that an alternate representation for the rotation operator is
\[ D(\phi) = \int_{0}^{2\pi} d\varphi \left|\varphi + \phi\right> \left<\varphi\right|. \] (4.5.196)
Show that it is unitary. (Hint: Recall the normalization in eq. (4.5.182).) Explain why the rotation will no longer unitary if the position eigenstates in eq. (4.5.196) are restricted to a wedge \( \{|\psi\> |0 \leq \psi \leq \phi_0 < 2\pi\} \) instead of a full circle.

Problem 4.54. To confirm eq. (4.5.190) is the right form of the rotation operator, argue, for an arbitrary state \(|f\rangle\) and \(|\phi\rangle\) a position eigenstate, that by denoting \( f(\phi) \equiv \langle \phi | f \rangle \),
\[ f(\phi - \phi') = e^{-i\phi \partial_\phi} f(\phi) = \left< \phi \left| e^{-i\phi' J} \right| f \right>. \] (4.5.197)
Can you also prove that
\[ \langle \phi | J | f \rangle = -i \partial_\phi \langle \phi | f \rangle? \] (4.5.198)
(Hint: Taylor expansion.) Compare these results to the one in eq. (4.5.77).
‘Orbital’ Angular Momentum

According to eq. (4.5.175), if we rotate the Cartesian coordinates \(\vec{x} \equiv (x^1, x^2)\) via \(\vec{x} \rightarrow \hat{R}\vec{x}\), and employ the Taylor expansions
\[
\cos \phi = 1 - \left(\frac{1}{2}\right) \phi^2 + \ldots
\]
and
\[
\sin \phi = \phi - \left(\frac{1}{3!}\right) \phi^3 + \ldots
\]
\[
\hat{R}(\phi)\vec{x} = \left(\begin{array}{cc}I_{2 \times 2} + \phi & -x^2 \partial_1 \\ x^1 \partial_2 & 1 - \left(\frac{1}{2}\right) \phi^2 \end{array}\right) + O(\phi^2). \tag{4.5.199}
\]
We may re-write this by first defining
\[
i \hat{J} = \begin{bmatrix}0 & 1 \\ -1 & 0\end{bmatrix}. \tag{4.5.201}
\]
Then eq. (4.5.199) reads
\[
\hat{R}(\phi)\vec{x} = \left(\begin{array}{cc}I - i \phi \hat{J} + O(\phi^2)\end{array}\right) \vec{x}. \tag{4.5.202}
\]
As we shall see below, that the \(-i \hat{J}\) is anti-symmetric – and therefore the ‘generator’ of rotations \(\hat{J}\) is Hermitian – is a feature that holds in general; not just in the 2D case here.

Moreover, we may consider the Taylor expansion resulting from the infinitesimal rotation carried out in eq. (4.5.200):
\[
\langle \vec{x} | f \rangle \rightarrow \langle x^1 - \phi x^2 + \ldots, x^2 + \phi x^1 + \ldots | f \rangle
\]
\[
= \langle \vec{x} | f \rangle - i(-\phi)(-i) \left(\partial_1 x^1 - \partial_2 x^2\right) \langle \vec{x} | f \rangle + O(\phi^2). \tag{4.5.203}
\]
On the other hand, we must have \(|x^1 - \phi x^2 + \ldots, x^2 + \phi x^1\rangle = \exp(-i\phi J) |x^1, x^2\rangle\).
\[
\langle x^1 - \phi x^2 + \ldots, x^2 + \phi x^1 + \ldots | f \rangle = (e^{-i\phi J} |\vec{x}\rangle)^\dagger |f\rangle = \langle \vec{x} | (1 + i\phi J + \ldots) | f \rangle
\]
\[
= \langle \vec{x} | f \rangle - i(-\phi) \langle \vec{x} | J | f \rangle + \ldots. \tag{4.5.204}
\]
Comparing equations (4.5.203) and (4.5.205) hands us the position representation of the generator of rotations – aka ‘orbital’ angular momentum – in 2D:
\[
\langle \vec{x} | J | f \rangle = -i \left(\partial_1 x^1 - \partial_2 x^2\right) \langle \vec{x} | f \rangle = \langle \vec{x} | X^1 P_2 - X^2 P_1 | f \rangle; \tag{4.5.206}
\]
where, in the final equality, we have recalled eq. (4.5.77). Below, we will generalize the identification
\[
J = X^1 P_2 - X^2 P_1 \tag{4.5.207}
\]
to higher dimensions.

**Eigenstates & Topology**

Since \(J\) is Hermitian, we are guaranteed its eigenstates form a complete basis \(\{|m\}\). Let us now witness how the choice of boundary conditions in eq. (4.5.181) will allow us to fix the eigenvalues. Consider, for \(|\psi\rangle\) some position eigenbra and \(|m\rangle\) an eigenstate of \(J\),
\[
\langle \psi | e^{-i(2\pi)J} | m \rangle = e^{-i(2\pi)m} \langle \psi | m \rangle \tag{4.5.208}
\]
Comparing the rightmost terms on the first and second line,
\[ e^{-i2\pi m} = 1 \iff m = \text{integer}. \] (4.5.210)

Choosing eq. (4.5.181) 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

\[ |\phi + 2\pi\rangle = -|\phi\rangle. \] (4.5.211)

Completeness We may construct the identity operator

\[ \mathbb{I} = \int_{0}^{2\pi} d\varphi |\varphi\rangle \langle \varphi|. \] (4.5.212)

As a check, we recall eq. (4.5.182) and calculate

\[ \langle \phi' | \mathbb{I} | \phi \rangle = \int_{0}^{2\pi} d\varphi \langle \phi' | \varphi \rangle \langle \varphi | \phi \rangle \]
\[ = \int_{0}^{2\pi} d\varphi \delta(\phi' - \varphi)\delta(\varphi - \phi) = \delta(\phi' - \phi). \] (4.5.214)

Problem 4.55. Change-of-basis Using eq. (4.5.198), can you show that
\[ \langle \phi | m \rangle \propto e^{im\phi}? \] (4.5.215)

If we agree to normalize the eigenstates as
\[ \langle m | n \rangle = \delta_{n}^{m} \] (4.5.216)

explain why, up to an overall phase factor,
\[ \langle \phi | m \rangle = \frac{\exp(im\phi)}{\sqrt{2\pi}}. \] (4.5.217)

If we assume the eigenstates of \( J \), namely \( \{|m\}\} \), are normalized to unity, we may also write

\[ \mathbb{I} = \sum_{m=-\infty}^{+\infty} |m\rangle \langle m|. \] (4.5.218)

For, we may check that we have the proper normalization; invoking eq. (4.5.217),

\[ \int_{0}^{2\pi} d\phi \langle \phi | \mathbb{I} | \phi' \rangle = 1 = \int_{0}^{2\pi} d\phi \sum_{m=-\infty}^{+\infty} \langle \phi | m \rangle \langle m | \phi' \rangle \] (4.5.219)
\[
\int_0^{2\pi} d\phi \sum_{m=-\infty}^{+\infty} e^{im(\phi-\phi')} \frac{1}{2\pi} = \sum_{m=-\infty}^{+\infty} \delta_m^0. \tag{4.5.220}
\]

Observe that eq. \((4.5.218)\) is essentially the 1D version of \((4.5.163)\). This is not a coincidence: the circle \(\phi \in [0, 2\pi)\) can, of course, be thought of as a periodic space with period \(2\pi\).

*Fourier Series* Any state \(|f\rangle\) may be decomposed into modes by inserting the completeness relation in eq. \((4.5.218)\):

\[
|f\rangle = \sum_{m=-\infty}^{+\infty} |m\rangle \langle m| f\rangle. \tag{4.5.221}
\]

Multiplying both sides by a position eigenbra \(\langle \phi|\) and using employing eq. \((4.5.217)\), we obtain the Fourier series expansion

\[
\langle \phi| f\rangle = \sum_{m=-\infty}^{+\infty} e^{im\phi} \frac{1}{\sqrt{2\pi}} \langle m| f\rangle. \tag{4.5.222}
\]

*Rotation Operator* At this point, recalling \((4.3.66)\) and the eigenvector equations

\[
J|m\rangle = m|m\rangle \quad \text{and} \quad e^{-i\phi J}|m\rangle = e^{-im\phi}|m\rangle \tag{4.5.223}
\]

hands us the following representation of the rotation operator:

\[
D(\phi) = \sum_{m=-\infty}^{+\infty} e^{-im\phi} |m\rangle \langle m|
\]

\[
= |0\rangle \langle 0| + \sum_{m=1}^{+\infty} \left( e^{-im\phi} |m\rangle \langle m| + e^{im\phi} |m\rangle \langle -m| \right). \tag{4.5.225}
\]

This result is the 1D analog of eq. \((4.5.171)\).

**Problem 4.56.** Recover eq. \((4.5.225)\) by employing eq. \((4.5.217)\) and inserting the completeness relation in eq. \((4.5.218)\) on the left and right of eq. \((4.5.196)\).

**Problem 4.57.** Our discussion thus far may seem a tad abstract. However, if we view the rotation matrix in eq. \((4.5.175)\) as the matrix element of some operator,

\[
|i|R|j\rangle = \hat{R}_{ij}, \quad i, j \in \{1, 2\}; \tag{4.5.226}
\]

show that this \(R\) is in fact related to the \(m = 1\) term in eq. \((4.5.225)\) 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 \(|\pm\rangle \equiv (\pm i/\sqrt{2})|\phi = 0\rangle + (1/\sqrt{2})|\phi = \pi/2\rangle\).

**Invariant subspaces** We close this section by making the observation that, due to the abelian (or, commutative) nature of 2D rotations in eq. \((4.5.185)\),

\[
D(\phi)\hat{D}(\phi)D(\phi) = D(\phi)^\dagger D(\phi)D(\phi) = D(\phi) \tag{4.5.227}
\]

since \(D(\phi)^\dagger D(\phi) = \mathbb{1}\) (cf. \((4.5.188)\)). 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. \((4.5.227)\) informs us that the 2D rotation operator in fact remains invariant under all change-of-basis transformations.
4.5.5 Rotations in $D \geq 3$ Spatial Dimensions; Angular Momentum & Spin

We now move on to study rotations in spatial dimensions $D$ higher than 2. For arbitrary $D \geq 2$, we may view rotations in Euclidean space as a continuous linear operation parametrized by $D(D-1)/2$ angles $\{\vec{\theta}\}$ that leaves the lengths of vectors invariant. More specifically, in $D$-space, the square of the distance between the point $\vec{x}$ and $\vec{x} + d\vec{x}$ is given by the Pythagorean theorem

$$d\ell^2 = d\vec{x} \cdot d\vec{x} = \delta_{ij} dx^i dx^j. \quad (4.5.228)$$

If we replace $dx^i$ with its linearly transformed counterpart via

$$dx^i \rightarrow \hat{R}^i_j dx^j, \quad (4.5.229)$$

the square of the distance is replaced with

$$\delta_{ij} dx^i dx^j \rightarrow \delta_{ij} (\hat{R}^a_i dx^a)(\hat{R}^b_j dx^b) = (\delta_{ab} \hat{R}^a_i \hat{R}^b_j) dx^i dx^j. \quad (4.5.230)$$

In matrix notation,

$$d\vec{x} \cdot d\vec{x} \rightarrow (\hat{R}d\vec{x}) \cdot (\hat{R}d\vec{x}) = d\vec{x}^T \hat{R}^T \hat{R} d\vec{x}. \quad (4.5.232)$$

Since we wish to preserve the length

$$\delta_{ij} dx^i dx^j \rightarrow \delta_{ij} dx^i dx^j \quad (4.5.233)$$

for arbitrary (infinitesimal) displacements $d\vec{x}$ – i.e., we are rotating the whole space centered at $\vec{x}$, not just a particular vector – we shall require that

$$\delta_{ij} = \delta_{ab} \hat{R}^a_i \hat{R}^b_j. \quad (4.5.234)$$

In matrix form,

$$I = \hat{R}(\vec{\theta})^T \hat{R}(\vec{\theta}). \quad (4.5.235)$$

which in turn is equivalent to

$$\hat{R}^T = \hat{R}^{-1}. \quad (4.5.236)$$

Notice the condition in eq. (4.5.235) not only preserves lengths, it also preserves angles between different vectors

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} \rightarrow (\hat{R} \vec{v})^T \hat{R} \vec{w} = \vec{v}^T \hat{R}^T \hat{R} \vec{w} = \vec{v} \cdot \vec{w}. \quad (4.5.237)$$

We will also assume we may set the rotation angles $\{\vec{\theta}\}$ (to, say, $\vec{0}$) such that the identity is recovered:

$$\hat{R}(\vec{\theta} = \vec{0}) = I_{D \times D}. \quad (4.5.238)$$

---

28Do 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 $D = 3$. 

98
Taking the determinant of eq. (4.5.235), and recalling \( \det AB = \det A \det B \) and \( \det \hat{R}^T = \det \hat{R} \), we have

\[
(\det \hat{R})^2 = 1 \quad \Rightarrow \quad \det \hat{R}(\vec{\theta}) = \pm 1.
\]

But since we have assumed we may continuously tune \( \hat{R} \) to the identity, which has determinant +1, it must be that

\[
\det \hat{R}(\vec{\theta}) = +1,
\]

because the determinant cannot abruptly jump from +1 to −1 as it must also depend on \( \{\vec{\theta}\} \) in a continuous manner.

Note that the product of two rotations \( \hat{R}_1 \) and \( \hat{R}_2 \) is another rotation because, as long as eq. (4.5.235) is obeyed – namely \( \hat{R}_1^T \hat{R}_1 = I \) and \( \hat{R}_2^T \hat{R}_2 = I \) – then we have

\[
\left( \hat{R}_1 \hat{R}_2 \right)^T \left( \hat{R}_1 \hat{R}_2 \right) = \hat{R}_2^T \hat{R}_1^T \hat{R}_1 \hat{R}_2
\]

\[
= \hat{R}_2^T \hat{R}_2 = I.
\]

That is, \( \hat{R}_3 \equiv \hat{R}_1 \hat{R}_2 \) satisfies \( \hat{R}_3^T \hat{R}_3 = I \). Moreover, as long as as eq. (4.5.240) is respected – namely, \( \det \hat{R}_1 = 1 = \det \hat{R}_2 \) – then \( \det \hat{R}_3 = 1 \) too, because

\[
\det \left( \hat{R}_1 \hat{R}_2 \right) = (\det \hat{R}_1)(\det \hat{R}_2).
\]

**Remark** Some jargon: the set of matrices \( \{\hat{R}\} \) obeying eq. (4.5.235) form the orthogonal group \( \text{O}_D(\equiv \text{O}(D)) \); if they are further restricted to be of unit determinant, i.e., obeying eq. (4.5.240) (aka ‘special’), they form the \( \text{SO}_D(\equiv \text{SO}(D)) \) group.

**Anti-symmetric Generators** We now turn to the construction of \( \hat{R} \). As we have argued previously, any operator continuously connected to the identity can be expressed as an exponential:

\[
\hat{R}(\vec{\theta}) = e^{\vec{\epsilon} \hat{\Omega}} = I + \vec{\epsilon} \cdot \hat{\Omega} + \mathcal{O}(\vec{\epsilon}^2),
\]

where the matrix \( \hat{\Omega} \) is known as the *generator* of rotations, which we will take to be real since \( \hat{R} \) is real. Furthermore, we have inserted a parameter \( \vec{\epsilon} \) so that eq. (4.5.235) may be now regarded as a Taylor series in \( \vec{\epsilon} \).

\[
\left( I + \vec{\epsilon} \cdot \hat{\Omega}^T + \mathcal{O}(\vec{\epsilon}^2) \right) \left( I + \vec{\epsilon} \cdot \hat{\Omega} + \mathcal{O}(\vec{\epsilon}^2) \right) = I,
\]

\[
I + \vec{\epsilon} \left( \hat{\Omega}^T + \hat{\Omega} \right) + \mathcal{O}(\vec{\epsilon}^2) = I
\]

The identity cancels out from both sides; leaving us with the conclusion that each order in \( \vec{\epsilon} \) must cancel. In particular, at first order,

\[
\hat{\Omega}^T = -\hat{\Omega}.
\]
Now, if eq. (4.5.247) were true then from eq. (4.5.244), we may verify eq. (4.5.235). By Taylor expanding the exponential, one may readily verify that $(\exp(\Omega))^T = \exp(\Omega^T) = \exp(-\Omega)$. Since $-\Omega$ and $\Omega$ commutes, we may combine the exponents in eq. (4.5.235)

$$\hat{R}^T \hat{R} = e^{-\hat{\Omega}} e^{\hat{\Omega}} = e^{\hat{\Omega} - \hat{\Omega}} = \mathbb{I}, \quad (4.5.248)$$

where we have now absorbed $\varepsilon$ into the generator $\hat{\Omega}$.

**Rotation angles** Moreover, note that antisymmetric matrices (with a total of $D^2$ entries) have zeros on the diagonal (since $\hat{\Omega}_{ii} = -\hat{\Omega}_{ii}$, 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 $\hat{\Omega}_{ij} = -\hat{\Omega}_{ji}$). 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 $\hat{J}^{ij}$ that span the space of $\{\hat{\Omega} = -i\omega_{ij}\hat{J}^{ij} | \hat{\Omega}^T = -\hat{\Omega}\}$ in fact generate rotations about these 2D planes, with rotation angle $\theta^i \leftrightarrow \omega_{ij}$. Hence, rotations in $D$ spatial dimensions are parametrized by a total of $D(D - 1)/2$ rotation angles $\{\theta^i | i = 1, 2, \ldots, D(D - 1)/2\}$.

**Basis generators** One such basis of anti-symmetric generators $\hat{\Omega}$ is as follows. First recall all diagonal components are zero. For the first generator basis matrix, set the $(1, 1)$ component to $-i\omega_{ij}\hat{J}^{ij}$, while $\hat{\Omega}$ is real and antisymmetric, $\hat{\Omega}^T = \hat{\Omega}^\dagger = -\hat{\Omega}$ (cf. eq. (4.5.247)),

$$\frac{1}{2}\omega_{ij}(-i\hat{J}^{ij})^\dagger = \frac{i}{2}\omega_{ij}\hat{J}^{ij} = -\frac{1}{2}\omega_{ij}(-i\hat{J}^{ij}). \quad (4.5.251)$$

But eq. (4.5.250) tells us $-i\hat{J}^{ij}$ is real and anti-symmetric, i.e., $(-i\hat{J}^{ij})^\dagger = +i\hat{J}^{ij}$; so not only is $\hat{J}^{ij}$ therefore Hermitian

$$+i\hat{J}^{ij} = (-i\hat{J}^{ij})^\dagger = i(\hat{J}^{ij})^\dagger \quad (4.5.252)$$

One may check that $\hat{J}^{ij} = -\hat{J}^{ji}$, and therefore the sum in eq. (4.5.249) 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, when $i = j$ we see eq. (4.5.250) 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 $j = a$. 

100
eq. \((4.5.251)\) informs us the parameters \(\{\omega_{ij}\}\) in eq. \((4.5.249)\) must therefore be real.

**Problem 4.58. 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 \(\{\hat{i}\}\), explain why

\[-iJ^{ij} |j\rangle = -|i\rangle \quad \text{and} \quad -iJ^{ij} |i\rangle = +|j\rangle; \quad (4.5.253)\]

while

\[-iJ^{ij} |k\rangle = 0, \quad \forall k \neq i,j. \quad (4.5.254)\]

In other words,

\[-iJ^{ij} = |j\rangle \langle i| - |i\rangle \langle j|. \quad (4.5.255)\]

Can you compute \((-iJ^{ij})^n\) for odd and even \(n\)? Then show that

\[e^{-i\theta J^{ij}} = \cos(\theta) |i\rangle \langle i| - \sin(\theta) |i\rangle \langle j| + \sin(\theta) |j\rangle \langle i| + \cos(\theta) |j\rangle \langle j| + \sum_{k \neq i,j} |k\rangle \langle k|. \quad (4.5.256)\]

In other words, the basis anti-symmetric rotation generators in eq. \((4.5.250)\) produce counter-clockwise rotations on the \((i,j)\) 2D plane, leaving the rest of the \(D\)-space untouched.

**Change-of-basis & rotating the \(\omega\)s** We may now show that

\[\hat{R} \exp\left(-\frac{i}{2} \omega_{ab} \hat{J}^{ab}\right) \hat{R}^T = \exp\left(-\frac{i}{2} \omega'_{ab} \hat{J}^{ab}\right), \quad (4.5.257)\]

with

\[\omega'_{ab} = \hat{R}_{am} \hat{R}_{bn} \omega_{mn}; \quad (4.5.258)\]

or, if we view \(\omega\) as a matrix,

\[\hat{\omega}' = \hat{R} \cdot \hat{\omega} \cdot \hat{R}^T. \quad (4.5.259)\]

To see this, we first note from Taylor expansion and \(\hat{R} \hat{R}^T = \mathbb{I}\) that

\[\hat{R} \exp(X) \hat{R}^T = \exp\left(\hat{R} \cdot X \cdot \hat{R}^T\right). \quad (4.5.260)\]

Therefore we may employ eq. \((4.5.250)\) to evaluate

\[-\frac{i}{2} \omega'_{ij} \left(\hat{J}^{ij}\right)_{ab} \equiv -\frac{i}{2} \omega_{ij} \left(\hat{R} \hat{J}^{ij} \hat{R}^T\right)_{ab} = -\frac{1}{2} \omega_{ij} \hat{R}_{am} \delta_{[a}^i \delta_{j]}^j \hat{R}_{mn} \quad (4.5.261)\]

\[= -\hat{R}_{am} \hat{R}_{bn} \omega_{mn} \quad (4.5.262)\]

On the other hand,

\[-\frac{i}{2} \omega'_{ij} \left(\hat{J}^{ij}\right)_{ab} = -\frac{1}{2} \omega'_{ij} \delta_{[a}^i \delta_{b]}^j \quad (4.5.263)\]

\[= -\omega'_{ab}. \quad (4.5.264)\]

Comparing equations \((4.5.262)\) and \((4.5.264)\), we arrive at eq. \((4.5.258)\).
To this end, we may discover from eq. (4.5.265) that
\[ D(\hat{R}) |\vec{x}\rangle \equiv \exp \left( -\frac{i}{2} \omega_{ab} J^{ab} \right) |\vec{x}\rangle = |\hat{R} \vec{x}\rangle, \]
where \( D(\hat{R}) \) is now the linear operator associated with the rotation matrix \( \hat{R} \) in eq. (4.5.249).
Now, we must have, for two rotation matrices \( \hat{R}_1 \) and \( \hat{R}_2 \),
\[ D(\hat{R}_1)D(\hat{R}_2) |\vec{x}\rangle = D(\hat{R}_1) |\hat{R}_2 \vec{x}\rangle = |\hat{R}_1 \hat{R}_2 \vec{x}\rangle = D(\hat{R}_1 \hat{R}_2) |\vec{x}\rangle. \]
Since \( |\vec{x}\rangle \) was arbitrary, we have the product rule
\[ D(\hat{R}_1)D(\hat{R}_2) = D(\hat{R}_1 \hat{R}_2). \] (4.5.267)

Now, according to the discourse enveloping equations (4.5.34) through (4.5.39), 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. (4.5.35)). On the other hand, eq. (4.5.267) tells us
\[ \exp \left( -\frac{i}{2} \omega_{ab} J^{ab} \right) \exp \left( -\frac{i}{2} \omega'_{ab} J^{ab} \right) = \exp \left( -\frac{i}{2} \omega''_{ab} J^{ab} \right); \]
where \( \omega_{ab} \) are the rotation angles describing \( \hat{R}_1(\omega), \) \( \omega'_a \) are those describing \( \hat{R}_2(\omega'), \) and \( \omega''_{ab} \) are those describing their product \( (\hat{R}_1 \hat{R}_2)(\omega''). \) One way to guarantee eq. (4.5.268) holds, is therefore to ensure the operators \( \{ J^{ab} \} \) obey the same Lie algebra as their matrix counterparts \( \{ \tilde{J}^{ab} \}. \)

**Problem 4.59. Lie Algebra of \( SO_D \)**

Use the choice of basis \( \{ \tilde{J}^{ab} \} \) in eq. (4.5.250) to argue there must a basis \( \{ J^{ab} \} \) such that
\[ [J^{kl}, J^{mn}] = -i \left( \delta^{[k,m} J^{n]l} - \delta^{[m,n} J^{k]l} \right). \] (4.5.269)

That the generators do not commute indicates rotations for \( D > 2 \) do not, in general, commute: \( \hat{R}_1 \hat{R}_2 \neq \hat{R}_2 \hat{R}_1. \) (The anti-symmetrization symbol means, for e.g., \( T^{[ij]} = T^{ij} - T^{ji}. \))

**Unitary \( D(\hat{R}) \) and Hermitian \( J^{ab} \)**

We will next see that these \( \{ J^{ab} \} \) are Hermitian because \( D(\hat{R}) = \exp(-i/2 \omega_{ab} J^{ab}) \) is unitary, since the \( \omega_{ab} \) are rotation angles and hence always real.

To this end, we may discover from eq. (4.5.265) that
\[ \langle \hat{R} \vec{x}' \rangle = \langle \vec{x}' | D(\hat{R})^\dagger. \] (4.5.270)
Together, we deduce
\[
\langle \vec{x}' | D(\hat{R})^{\dagger} D(\hat{R}) | \vec{x} \rangle = \langle \hat{R} \vec{x}' | \hat{R} \vec{x} \rangle = \delta^{(D)}(\hat{R}(\vec{x} - \vec{x}'))
\] (4.5.271)
\[
= \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{| \det \partial(\hat{R} x^i)/\partial x^a |^{\frac{1}{2}} | \det \partial(\hat{R} x^i)/\partial x^a |^{\frac{1}{2}}} = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{| \det \partial(\hat{R} x^i)/\partial x^a |^{\frac{1}{2}}} = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{| \det \partial(\hat{R} x^i)/\partial x^a |^{\frac{1}{2}}};
\]
where in the second line we have appealed to eq. (4.5.57). Moreover, we may compute
\[
\det \frac{\partial(\hat{R}^i x^j)}{\partial x^a} = \det \frac{\partial(\hat{R}^i x^j')}{\partial x'^a} = \det \hat{R}^i_j \delta^j_a = \det \hat{R}^i_a = 1,
\] (4.5.272)
if we recall eq. (4.5.240). At this point, we gather
\[
\langle \vec{x}' | D(\hat{R})^{\dagger} D(\hat{R}) | \vec{x} \rangle = \delta^{(D)}(\vec{x}' - \vec{x}) = \langle \vec{x}' | \vec{x} \rangle.
\] (4.5.273)
Since \( | \vec{x} \rangle \) and \( | \vec{x}' \rangle \) are arbitrary, we have proven
\[
D(\hat{R})^{\dagger} D(\hat{R}) = \mathbb{I}.
\] (4.5.274)

**Problem 4.60.** Can you argue that
\[
D(\hat{R})^{\dagger} | \vec{x} \rangle = | \hat{R}^T \vec{x} \rangle = D(\hat{R}^T) | \vec{x} \rangle?
\] (4.5.275)
This is the generalization of eq. (4.5.189) to \( D \geq 3 \). Hint: Apply both sides of eq. (4.5.273) on \( | \vec{x} \rangle \). \( \Box \)

**Problem 4.61.** ‘Orbital’ Angular Momentum: Position Representation

In this problem, we shall work out \( J^{ab} \) within the position representation.

According to eq. (4.5.244), an infinitesimal rotation may be implemented via the replacement
\[
x^i \rightarrow \hat{R}^i_j x^j = \left( \delta^i_j + \hat{\Omega}^i_j + \ldots \right) x^j,
\] (4.5.276)
where \( \Omega \) is anti-symmetric; cf. eq. (4.5.247). (Here, the placement of indices on \( \hat{\Omega} \), i.e., up versus down, is unimportant.) In eq. (4.5.265), take
\[
\hat{\Omega} = -i \theta \hat{J}^{ij},
\] (4.5.277)
where \( J^{ij} \) is one of the basis anti-symmetric matrices in eq. (4.5.250); and let \( f(\vec{x}) = \langle \vec{x} | f \rangle \) be an arbitrary function. Explain why the replacement in eq. (4.5.276) induces
\[
f(\vec{x}) \rightarrow f(\vec{x}) + \theta \cdot (x^i \partial_j - x^j \partial_i) f(\vec{x}).
\] (4.5.278)
Next show that, upon an infinitesimal rotation generated by \( J^{ij} \) – now acting on the \( \{ | \vec{x} \rangle \} \) –
\[
\langle \vec{x} | f \rangle \rightarrow \left( D(\hat{R}) | \vec{x} \rangle \right)^\dagger | f \rangle = \langle \vec{x} | f \rangle + i \theta \langle \vec{x} | J^{ij} | f \rangle + O(\theta^2).
\] (4.5.279)
We may therefore identify

$$\langle \vec{x} | J^{ij} | f \rangle = -i \left( x^i \partial_j - x^j \partial_i \right) \langle \vec{x} | f \rangle, \quad \partial_j \equiv \frac{\partial}{\partial x^j}. \quad (4.5.281)$$

That is, these $J^{ij}$ are the $D$–dimensional analogs of the ‘orbital angular-momentum’ operators in 3D space. Employing eq. (4.5.177), we may deduce from (4.5.281) that

$$J^{ij} = X^i P_j - X^j P_i, \quad (4.5.282)$$

where $\vec{X}$ and $\vec{P}$ are now, respectively, the position and momentum operators.

Can you verify that eq. (4.5.282) satisfies the Lie Algebra in eq. (4.5.269) 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 §7 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. (4.5.228) in the desired coordinate system. For our case, we will focus on the $D$–dimensional spherical coordinate system $(r, \vec{\theta})$, which yields

$$\delta_{ij} dx^i dx^j = dr^2 + r^2 H_{IJ} d\theta^I d\theta^J. \quad (4.5.283)$$

if we set the Cartesian coordinate vector to be equal to the radial distance $r$ times an appropriately defined unit radial vector parameterized by $\vec{\theta}$: $x^i = r \vec{\hat{r}}(\vec{\theta})$. In particular,

$$r^2 H_{IJ} = r^2 \delta_{ij} \frac{\partial \vec{\hat{r}}^i}{\partial \theta^I} \frac{\partial \vec{\hat{r}}^j}{\partial \theta^J}. \quad (4.5.284)$$

Generically, given a metric

$$d\ell^2 = g_{ij} dx^i dx^j \quad (4.5.285)$$

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

$$\nabla^2 \psi = \frac{1}{\sqrt{g}} \partial_i \left( \sqrt{g} g^{ij} \partial_j \psi \right). \quad (4.5.286)$$

For instance, the metric in Cartesian coordinates is simply $\delta_{ij}$, whose determinant is unity, inverse $\delta^{ij}$, and Laplacian

$$\nabla^2 \psi = \delta^{ij} \partial_i \partial_j \psi. \quad (4.5.287)$$

**Problem 4.62.** If $H$ denotes the determinant of $H_{IJ}$ in eq. (4.5.283), show that the $D$–space Laplacian is

$$\nabla^2 \psi = \frac{1}{r^{D-1}} \partial_r \left( r^{D-1} \partial_r \psi \right) + \frac{1}{r^2} \nabla^2_{S^{D-1}} \psi; \quad (4.5.288)$$
where $\vec{\nabla}^2_{S^{D-1}}$ is the Laplacian on the unit $(D - 1)$-sphere (i.e., $r = 1$), namely
\[
\vec{\nabla}^2_{S^{D-1}} \psi = \frac{1}{\sqrt{H}} \partial_I \left( \sqrt{H} H^{IJ} \partial_J \psi \right),
\]  
(4.5.289)
where the $I$ and $J$ indices run only over the angular coordinates $\{\theta^I\}$. This result will be used in the next problem.

**Problem 4.63.** $(2r^2)^{-1} J^{ab} J^{ab} = \vec{J}^2 / r^2$ and Laplacian on Sphere

Use eq. (4.5.281) to show that
\[
\frac{1}{2} \langle \vec{x} \mid J^{ab} J^{ab} \mid \psi \rangle = \left( (D - 1)x^a \partial_a + x^a x^b \partial_a \partial_b - \vec{x} \vec{\nabla}^2 \right) \langle \vec{x} \mid \psi \rangle.
\]  
(4.5.290)
Next, recall eq. (4.5.99) and show that
\[
- \left\langle \vec{x} \left| \vec{P}^2 - \frac{1}{2r^2} J^{ab} J^{ab} \right| \psi \right\rangle = \frac{1}{r^{D-1}} \partial_r \left( r^{D-1} \partial_r \langle \vec{x} \mid \psi \rangle \right).
\]  
(4.5.291)
From this, identify $(1/2) J^{ab} J^{ab}$ as the negative Laplacian on the $(D - 1)$-sphere:
\[
\left\langle \vec{x} \left| \vec{J}^2 \right| \psi \right\rangle \equiv \frac{1}{2} \left\langle \vec{x} \left| J^{ab} J^{ab} \right| \psi \right\rangle = - \vec{\nabla}^2_{S^{D-1}} \langle \vec{x} \mid \psi \rangle.
\]  
(4.5.292)
Hints: You may need the result from the previous problem. Recognize too, from $x^i = r \hat{r}^i(\vec{\theta})$,
\[
r \partial_r = r \frac{\partial x^i}{\partial r} \partial_i = r \hat{r}^i \partial_i = x^i \partial_i;
\]  
(4.5.293)
as well as (keeping in mind $\hat{r}^a \partial_a \hat{r}^b = 0$ – can you see why?)
\[
x^i x^j \partial_i \partial_j \psi = r^2 \partial^2 \psi.
\]  
(4.5.294)
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 $r$.

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
\[
x^i = r \hat{r}^i(\vec{\theta}), \quad \hat{r}^i \hat{r}^i = 1
\]  
(4.5.295)
these angles are orthogonal in the sense that
\[
\partial_i \hat{r} \cdot \partial_j \hat{r} = \delta_{i j} \partial_i \hat{r}^i \cdot \partial_j \hat{r}^j = H_{IJ} \equiv \text{diag} \left[ H_{22}, H_{33}, \ldots, H_{DD} \right].
\]  
(4.5.296)
In other words, we assume the angular metric in eq. (4.5.284) is diagonal.
Another consequence of eq. (4.5.295) follows from differentiating $\hat{r}^i \hat{r}^i = 1$ with respect to any of one of the angles is

$$\hat{r}^i \partial_{\hat{r}} \hat{r}^i = 0. \quad (4.5.297)$$

The Jacobian $\partial x^i / \partial (r, \theta)^a$ takes the form

$$\frac{\partial x^i}{\partial r} = \hat{r}^i \quad \text{and} \quad \frac{\partial x^i}{\partial \theta^j} = r \frac{\partial \hat{r}^i}{\partial \theta^j}. \quad (4.5.298)$$

Let us now observe, through the chain rule

$$\frac{\partial x^i}{\partial x^j} = \delta^i_j, \quad (4.5.299)$$

the matrix $\partial (r, \vec{\theta})^a / \partial x^j$ is simply the inverse of $\partial x^i / \partial (r, \vec{\theta})^a$; namely,

$$\frac{\partial (r, \vec{\theta})^a}{\partial x^j} = \left( \left( \frac{\partial (r, \vec{\theta})}{\partial \vec{x}^j} \right)^{-1} \right)^a_j. \quad (4.5.300)$$

It has components

$$\frac{\partial r}{\partial x^i} = \hat{r}^i \quad \text{and} \quad \frac{\partial \theta^l}{\partial x^i} = \frac{H^{1L}}{r} \frac{\partial \hat{r}^i}{\partial \theta^l}; \quad (4.5.301)$$

where the inverse angular metric is defined through the relation

$$H^{IK} H_{KJ} = \delta^I_J. \quad (4.5.302)$$

To see this, we simply check that our expressions for $\partial r / \partial x^i$ and $\partial \theta^l / \partial x^i$ do indeed yield the components of the inverse of $\partial x^i / \partial (r, \theta)^a$; namely,

$$\frac{\partial r}{\partial r} = \frac{\partial r}{\partial x^i} \frac{\partial x^i}{\partial r} = \hat{r}^i \hat{r}^i = 1; \quad (4.5.303)$$

$$\frac{\partial r}{\partial \theta^l} = \frac{\partial r}{\partial x^i} \frac{\partial x^i}{\partial \theta^l} = r \hat{r}^i \partial_{\theta^l} \hat{r}^i = 0; \quad (4.5.304)$$

$$\frac{\partial \theta^l}{\partial r} = \frac{\partial \theta^l}{\partial x^i} \frac{\partial x^i}{\partial r} = \frac{H^{1L}}{r} \frac{\partial \hat{r}^i}{\partial \theta^l} \hat{r}^i = 0; \quad (4.5.305)$$

and

$$\frac{\partial \theta^l}{\partial \theta^j} = \frac{\partial \theta^l}{\partial x^i} \frac{\partial x^i}{\partial \theta^j} = \frac{H^{IK}}{r} \partial_{\theta^j} \hat{r}^i \cdot r \partial_{\theta^l} \hat{r}^i = H^{IK} H_{KJ} = \delta^I_J. \quad (4.5.306)$$

Hence, from eq. (4.5.301),

$$\langle r, \vec{\theta} | J^{ab} | \psi \rangle = -i x^a \partial_b \langle \vec{x} | \psi \rangle = -i \left( x^a \frac{\partial r}{\partial x^b} \partial_r + x^a \frac{\partial \theta^l}{\partial x^b} \partial_{\theta^l} \right) \langle \vec{x} | \psi \rangle \quad (4.5.307)$$

106
Because \( r \hat{r} = \vec{x} \) (cf. eq. (4.5.295)), the left term in the last line is zero because \( r \hat{a} \hat{b} = 0 \). From eq. (4.5.301), we now arrive at the spherical coordinates analog of eq. (4.5.281):

\[
\langle r, \vec{\theta} \mid J_{ij} \mid \psi \rangle = -i H_{ij} \left( r^a \frac{\partial r^b}{\partial \theta^1} - r^b \frac{\partial r^a}{\partial \theta^1} \right) \frac{\partial}{\partial \theta^j} \langle r, \vec{\theta} \mid \psi \rangle.
\]  

(4.5.309)

**Problem 4.64.** Show that \( J_{ab} \) commutes with \((1/2) J^{ij} J^{ij}\), i.e.

\[
\left[ J_{ab}, \frac{1}{2} J^{ij} J^{ij} \right] = 0.
\]  

(4.5.310)

Hint: Remember eq. (4.3.87).

**Spherical Harmonics in \( D \)–dimensions**

The Poisson equation of Newtonian gravity or Coulomb’s law reads

\[
\vec{\nabla}^2 \psi = 4\pi \rho,
\]  

(4.5.311)

where \( \rho \) is either mass or charge density. Suppose we were solving \( \psi \) 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

\[
\psi(\vec{x}) = \sum_{\ell=0}^{+\infty} \frac{x^{i_1} \ldots x^{i_\ell}}{\ell!} \psi_{i_1 \ldots i_\ell},
\]  

(4.5.312)

\[
\psi_{i_1 \ldots i_\ell} = \partial_{i_1} \ldots \partial_{i_\ell} \psi(\vec{x} = \vec{0}).
\]  

(4.5.313)

Since \( \rho(\vec{x}) = 0 \) in this region, eq. (4.5.311) reduces to \( \vec{\nabla}^2 \psi = 0 \). Eq. (4.5.312) inserted into eq. (4.5.311) must yield the statement that, for a fixed \( \ell \) but with summation over the \( \ell \) indices \( \{i_1, \ldots, i_\ell\} \) still in force,

\[
\vec{\nabla}^2_{\vec{x}} \left( x^{i_1} x^{i_2} \ldots x^{i_{\ell-1}} x^{i_\ell} \right) \psi_{i_1 \ldots i_\ell} = 0.
\]  

(4.5.314)

Notice \( x^{i_1} \ldots x^{i_\ell} \psi_{i_1 \ldots i_\ell} \) is a homogeneous polynomial of degree \( \ell \). (Here, a homogeneous polynomial of degree \( \ell \), \( P_\ell \), is a polynomial built of out the Cartesian components of \( \vec{x} \) such that, under the re-scaling \( \vec{x} \rightarrow \lambda \vec{x} \), the polynomial scales as \( P_\ell \rightarrow \lambda^\ell P_\ell \).) Therefore, for each \( \ell \), the solution of the vacuum Poisson equation in \( D \)–dimensions involves eq. (4.5.314): homogeneous polynomials of degree \( \ell \) annihilated by the Laplacian – this is often the starting definition of the spherical harmonics.

**Problem 4.65.** Recall the space of polynomials of degree less than or equal to \( \ell \) 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. (4.5.314)? Hint: Remember the discussion at the end of §4.1.
If we employ spherical coordinates in \( D \)-dimensions,
\[
x^i = r \hat{r}^i(\vec{\theta}), \quad \vec{\theta} = (\theta^1, \ldots, \theta^{D-1});
\]
then eq. (4.5.314) takes the form
\[
\vec{\nabla}^2 \left( r^\ell Y(\vec{\theta}) \right) = 0;
\]
where the angular portion arises from
\[
Y(\vec{\theta}) = \hat{r}_1 \ldots \hat{r}_\ell \psi_{i_1 \ldots i_\ell}.
\]
\( \text{Problem 4.66. Eigenfunctions/values on the } (D - 1)\text{-sphere} \)

Show that eq. (4.5.316) leads to the eigenvector/value equation
\[
\vec{\nabla}^2_S D - 1 Y(\vec{\theta}) = -\ell(\ell + D - 2) Y(\vec{\theta}).
\]
These angular spherical harmonics, for \( D = 3 \), are usually denoted as \( Y_\ell^m(\theta, \phi) \), where \( \ell = 0, 1, 2, \ldots \) and \( -\ell \leq m \leq +\ell \). We shall examine them in the next section.

\( \text{Rotations in 3D} \)

We will now focus on the 3D case, 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
\[
J^i \equiv \frac{1}{2} \epsilon^{ijk} J^{jk},
\]
which, by multiplying both sides with \( \epsilon^{mni} \) and using the result
\[
\epsilon^{aij} \epsilon^{amn} = \delta^i_{[m} \delta^j_{n]} = \delta^i_m \delta^j_n - \delta^i_n \delta^j_m,
\]
is equivalent to
\[
J^{ij} = \epsilon^{ijm} J^m;
\]
so that
\[
-\frac{i}{2} \omega_{ij} J^{ij} = -i \theta^i J^i \quad \Leftrightarrow \quad \theta^i = \frac{1}{2} \epsilon_{abl} \epsilon^{abi}.
\]
For example,
\[
\theta^1 = \frac{1}{2} \epsilon^{123} \omega_{23} + \frac{1}{2} \epsilon^{132} \omega_{32} = \omega_{23}.
\]

\(^{30}\text{The proof of eq. (4.5.320) can be found in the discussion following eq. (7.3.139) below.}\)
Recall that \( J^{23} \) generates rotations of the \((2, 3)\) plane, and \( \omega_{23} \) is the corresponding angle (for e.g., eq. (4.5.256)); 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 (4.5.319) and (4.5.322), when \( D = 3 \), we thus specialize the \(D\)-dimensional result in eq. (4.5.249) as

\[
3D : \hat{R}(\theta) = \exp\left(-\frac{i}{2} \omega_{ab} J^{ab}\right) = \exp\left(-i \vec{\theta} \cdot \vec{J}\right).
\]

(4.5.324)

It is worth reiterating, that this rotation operator can be written either in terms of \( J^i \) or \( J^{ab} \) is unique to 3D.

Moreover, eq. (4.5.320) may be invoked to deduce

\[
\hat{J}^2 \equiv J^a J^a = \frac{1}{4} \epsilon^{amn} \epsilon^{aij} J^{mn} J^{ij}
\]

(4.5.325)

\[
= \frac{1}{4} (\delta^m_i \delta^n_j - \delta^m_j \delta^n_i) J^{mn} J^{ij} = \frac{1}{2} J^{mn} J^{mn}.
\]

(4.5.326)

In the previous section, we have already demonstrated that \( D(\hat{R}) \) 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 4.67. Lie Algebra of Rotation Generators in 3D**

Show that

\[
\left[ J^a, J^b \right] = \epsilon^{abc} J^c
\]

and

\[
\left[ J^a, J^2 \right] = 0, \quad J^2 \equiv J^i J^i.
\]

(4.5.327)

(4.5.328)

Hint: Recall equations (4.5.269) and (4.5.310). Eq. (4.5.327) may also be tackled by first utilizing eq. (4.5.250) to prove that the matrix generator is

\[
(\hat{J}^i)_{ab} = -i \delta^{i}{}_{ab}.
\]

(4.5.329)

**Eigenvalues of \( \hat{J}^2 \) and \( J^3 \) from Ladder Operators in 3D**

According to eq. (4.5.327), the \( \{J^a\} \) do not commute among themselves. However, eq. (4.5.328) tells us we may choose \( J^2 \) 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 \( \hat{J}^2 \) and \( J^3 \). Denote the simultaneous eigenket of \( \hat{J}^2 \) and \( J^3 \) as \( |\lambda J, m\rangle \).

\[
\hat{J}^2 |\lambda J, m\rangle = \lambda J |\lambda J, m\rangle \quad \text{and} \quad J^3 |\lambda J, m\rangle = m |\lambda J, m\rangle
\]

(4.5.330)

To this end, let us define the raising \( J^+ \) and lowering \( J^- \) operators

\[
J^\pm \equiv J^1 \pm i J^2;
\]

(4.5.331)
and compute, using the linearity of the commutator and the Lie Algebra of eq. (4.5.327),

\[
[J^3, J^\pm] = [J^3, J^1] \pm i [J^3, J^2] = -i \epsilon^{132} J^2 \mp i^2 \epsilon^{231} J^1 = \pm (J^1 \pm i J^2).
\] (4.5.333)

In other words,

\[
[J^3, J^\pm] = \pm J^\pm.
\] (4.5.334)

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.

\[
J^\pm |\lambda_J, m\rangle = c_{m\pm1} |\lambda_J, m \pm 1\rangle
\] (4.5.335)

To see this, we employ eq. (4.5.334),

\[
J^3 J^\pm |\lambda_J, m\rangle = (J^3 J^\pm - J^\pm J^3 + J^\pm J^3) |\lambda_J, m\rangle
\]

\[
= ([J^3, J^\pm] + J^\pm J^3) |\lambda_J, m\rangle
\]

\[
= (m \pm 1) J^\pm |\lambda_J, m\rangle.
\] (4.5.336)

**Problem 4.68.** Show that

\[
[J^+, J^-] = 2 J^3.
\] (4.5.337)

Next, let us prove that, for a fixed \(\lambda_J\), there is a maximum and minimum eigenvalue of \(J^3\). We shall use the non-negative character of the norm to do so. Specifically,

\[
(J^\pm |\lambda_J, m\rangle)^\dagger J^\pm |\lambda_J, m\rangle \geq 0.
\] (4.5.338)

Now,

\[
(J^\pm)^\dagger J^\pm = (J^1 \mp i J^2)(J^1 \pm i J^2)
\]

\[
= (J^1)^2 + (J^2)^2 \pm i (J^1 J^2 - J^2 J^1)
\]

\[
= (J^1)^2 + (J^2)^2 \pm \epsilon^2 J^3
\]

\[
= (J^1)^2 + (J^2)^2 + (J^3)^2 - (J^3)^2 \mp J^3 = J^2 - (J^3)^2 \mp J^3.
\] (4.5.342)

Therefore, their average is

\[
\frac{1}{2} (J^+)^\dagger J^+ + \frac{1}{2} (J^-)^\dagger J^- = J^2 - (J^3)^2;
\] (4.5.343)

and we have

\[
\frac{1}{2} (J^+ |\lambda_J, m\rangle)^\dagger J^+ |\lambda_J, m\rangle + \frac{1}{2} (J^- |\lambda_J, m\rangle)^\dagger J^- |\lambda_J, m\rangle \geq 0
\] (4.5.344)

\[
\langle \lambda_J, m | J^2 - (J^3)^2 | \lambda_J, m \rangle \geq 0
\] (4.5.345)
\[ \lambda_J \geq m^2. \quad (4.5.346) \]

If there were no \( m_{\text{max}} \), eq. \((4.5.336)\) 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 \( \lambda_J \) at some point, contradicting eq. \((4.5.346)\). By applying more and more powers of \( J^- \), we may similarly argue there has to be a \( m_{\text{min}} \), otherwise \( m^2 \) will eventually violate eq. \((4.5.346)\) again. These considerations also tell us,

\[ J^+ |\lambda_J, m_{\text{max}}\rangle = 0; \quad (4.5.347) \]

for otherwise eq. \((4.5.336)\) would imply there is no \( m_{\text{max}} \); likewise,

\[ J^- |\lambda_J, m_{\text{min}}\rangle = 0. \quad (4.5.348) \]

Let us in fact consider the former; this implies

\[
\langle \lambda_J, m_{\text{max}} | (J^+)\dagger J^+ |\lambda_J, m_{\text{max}}\rangle = 0 \quad (4.5.349) \\
\langle \lambda_J, m_{\text{max}} | \vec{J}^2 - (J^3)^2 |\lambda_J, m_{\text{max}}\rangle = 0 \quad (4.5.350) \\
\lambda_J = m_{\text{max}}(m_{\text{max}} + 1); \quad (4.5.351)
\]

where eq. \((4.5.342)\) was employed in the second line. If we instead considered \( J^- |\lambda_J, m_{\text{min}}\rangle = 0, \)

\[
\langle \lambda_J, m_{\text{min}} | (J^-)\dagger J^- |\lambda_J, m_{\text{min}}\rangle = 0 \quad (4.5.352) \\
\langle \lambda_J, m_{\text{min}} | \vec{J}^2 - (J^3)^2 + J^3 |\lambda_J, m_{\text{min}}\rangle = 0 \quad (4.5.353) \\
\lambda_J = m_{\text{min}}(m_{\text{min}} - 1); \quad (4.5.354)
\]

where once again eq. \((4.5.342)\) was employed in the second line. Equating the right hand sides of equations \((4.5.351)\) and \((4.5.354)\),

\[
m_{\text{max}} = \frac{-1 \pm \sqrt{1 - 4(1)(-1)m_{\text{min}}(m_{\text{min}} - 1)}}{2} \quad (4.5.355) \\
= -\frac{1}{2} \pm \left( m_{\text{min}} - \frac{1}{2} \right). \quad (4.5.356)
\]

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 \( (J^+)^n |\lambda_J, m_{\text{min}}\rangle \propto |\lambda_J, m_{\text{max}}\rangle \). At this point we gather

\[ m_{\text{min}} + n = -m_{\text{max}} + n = m_{\text{max}}; \quad (4.5.357) \]

which in turn implies

\[ m_{\text{max}} = \frac{n}{2}. \quad (4.5.358) \]

Since we have no further constraints on the integer \( n \), we now search the cases where \( m_{\text{max}} \) 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. \((4.5.351)\):
Spin & 3D Rotations  

Starting solely from the commutation relations between the angular momentum operators \( \{ J^a \} \) in eq. (4.5.327), we surmise: the simultaneous eigensystem of \( \vec{J}^2 \) and \( J^3 \) is encoded within

\[
\vec{J}^2 |\ell, m\rangle = \ell (\ell + 1) |\ell, m\rangle \quad \text{and} \quad J^3 |\ell, m\rangle = m |\ell, m\rangle . \tag{4.5.359}
\]

Here, the spin \( \ell \) can be a non-negative integer (\( \ell = 0, 1, 2, \ldots \)) or positive half-integer (\( \ell = 0.5, 1.5, 2.5, \ldots \)); whereas the azimuthal eigenvalue runs between \(-\ell\) to \(\ell\) in integer steps:

\[
m \in \{-\ell, -\ell + 1, -\ell + 2, \ldots, \ell - 2, \ell - 1, \ell\} . \tag{4.5.360}
\]

**Problem 4.69. Rotating the rotation axis**  
Show that the 3D version of eq. (4.5.258) is:

\[
\hat{R} \exp \left(-i\vec{\theta} \cdot \vec{J} \right) \hat{R}^T = \exp \left(-i\vec{\theta}' \cdot \vec{J} \right) , \quad \vec{\theta} \cdot \vec{J} \equiv \theta_a J^a , \tag{4.5.361}
\]

where

\[
\theta_a' \equiv \hat{R}_{ab} \theta_b . \tag{4.5.362}
\]

In the other words, a change-of-basis through a rotation \( \hat{R} \) amounts to rotating the angles \( \vec{\theta} \) .

We may also compute, up to an overall phase, the normalization constant in eq. (4.5.335). We have, from eq. (4.5.342),

\[
0 \leq |c_{m+1}^m|^2 \langle \ell, m+1 | \ell, m \rangle = \langle \ell, m | J^\pm | \ell, m \rangle
\]

\[
= \langle \ell, m | \vec{J}^2 - (J^3)^2 \mp J^3 | \ell, m \rangle
\]

\[
= \ell (\ell + 1) - m(m \pm 1) = (\ell \mp m)(\ell \pm m + 1) . \tag{4.5.365}
\]

Since eigenvectors are only defined up to a phase, we shall choose to simply take the positive square root on both sides.

\[
J^\pm |\ell, m\rangle = \sqrt{\ell (\ell + 1) - m(m \pm 1)} |\ell, m \pm 1\rangle . \tag{4.5.366}
\]

**Invariant Subspaces in 3D, Degeneracy & Symmetry**  
Because \([\vec{J}^2, J^a] = 0\), we must have, for \( D(\hat{R}) = \exp(-i\vec{\theta} \cdot \vec{J}) \),

\[
\left[ D(\hat{R}), \vec{J}^2 \right] = 0 , \tag{4.5.367}
\]

\[
\vec{J}^2 D(\hat{R}) |\ell, m\rangle = D(\hat{R}) \vec{J}^2 |\ell, m\rangle = \ell (\ell + 1) \cdot D(\hat{R}) |\ell, m\rangle . \tag{4.5.368}
\]

In words, \( D(\hat{R}) |\ell, m\rangle \) is an eigenvector of \( \vec{J}^2 \) with eigenvalue \( \ell (\ell + 1) \). Hence, we see that rotations do not ‘mix’ the eigenvectors \( \{|\ell, m\rangle\} \) of \( \vec{J}^2 \) with different \( \ell \). That is,

\[
D(\hat{R}) |\ell, m\rangle = \sum_{m' = -\ell}^{+\ell} |\ell, m'\rangle \hat{D}_{(\ell)}^{m'} m'(\hat{R}) ,
\]

112
\[
\hat{D}_{(\ell) m'}^m(\hat{\mathbf{R}}) \equiv \langle \ell, m' | D(\hat{\mathbf{R}}) | \ell, m \rangle.
\]

While the completeness relation should involve a sum over all \(\ell'\), namely
\[
\mathbb{1} = \sum_{\ell' = 0}^{\infty} \sum_{m' = -\ell'}^{\ell'} | \ell', m' \rangle \langle \ell', m' |
\]
only the \(\ell' = \ell\) terms will survive when employed in eq. (4.5.369) – due to the result in eq. (4.5.368).

Now, suppose a Hermitian operator \(A\) remains invariant under rotations; namely,
\[
D(\hat{\mathbf{R}})AD(\hat{\mathbf{R}})^\dagger = H.
\]

Then by Taylor expanding \(D(\hat{\mathbf{R}})\), we see \(A\) must commute with the generators \(\{J^i\}\).
\[
[J^i, A] = 0
\]

We must therefore be able to simultaneously diagonalize \(\{A, \hat{J}^2, J^3\}\). 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 §4.5.1. To see this, we first compute
\[
A \left( D(\hat{\mathbf{R}}) |\lambda; \ell, m\rangle \right) = D(\hat{\mathbf{R}})A |\lambda; \ell, m\rangle
\]
\[
= \lambda \left( D(\hat{\mathbf{R}}) |\lambda; \ell, m\rangle \right) .
\]

Inserting a complete set of eigenstates, and exploiting the fact that eigenstates with distinct eigenvalues are necessarily orthogonal,
\[
A \sum_{\lambda', \ell', m'} |\lambda'; \ell', m'\rangle \left\langle \lambda'; \ell', m' | D(\hat{\mathbf{R}}) |\lambda; \ell, m\rangle \right.
\]
\[
= A \sum_{m'} |\lambda; \ell, m'\rangle \left\langle \lambda; \ell, m' | D(\hat{\mathbf{R}}) |\lambda; \ell, m\rangle \right.
\]
\[
= \lambda \sum_{m'} |\lambda; \ell, m'\rangle \left\langle \lambda; \ell, m' | D(\hat{\mathbf{R}}) |\lambda; \ell, m\rangle \right.
\]

Since we have made no assumption of the rotation \(\hat{\mathbf{R}}\) 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 \(\lambda\). Note, however, this says nothing about states of distinct \(\hat{J}^2\) eigenvalues – i.e., eigenvectors with different \(\ell\)s can have either the same or different eigenvalues of \(A\), depending on what \(A\) itself actually is.

**Vector Operators** Suppose \(D(\hat{\mathbf{R}})\) is a rotation operator. Consider the following operation involving the position operator \(X^i\) and its eigenkets \(|\vec{x}\rangle\):
\[
D(\hat{\mathbf{R}})^\dagger X^i D(\hat{\mathbf{R}}) |\vec{x}\rangle = D(\hat{\mathbf{R}})^\dagger X^i \hat{\mathbf{R}} \vec{x}\rangle
\]
\[ = (\hat{R}\vec{x})^i D(\hat{R})^\dagger \left| \hat{R}\vec{x} \right\rangle = (\hat{R}\vec{x})^i \left| \hat{R}^T \hat{R}\vec{x} \right\rangle \quad (4.5.379) \]
\[ = \hat{R}^i_j \vec{p}^j \left| \vec{x} \right\rangle . \quad (4.5.380) \]

(We have employed eq. (4.5.275) in the third equality.) Since this holds for arbitrary position eigenkets, we must have the operator identity

\[ D(\hat{R})^\dagger X^i D(\hat{R}) = \hat{R}^i_j X^j. \quad (4.5.381) \]

**Problem 4.70.** Using eq. (4.5.129), first explain why the rotation operator applied to \( \left| k \right\rangle \), the eigenket of the momentum operator, behaves similarly as its position cousin:

\[ D(\hat{R}) \left| k \right\rangle = \left| \hat{R}k \right\rangle . \quad (4.5.382) \]

Then show that the analog to eq. (4.5.381) for the momentum operator holds; namely,

\[ D(\hat{R})^\dagger P_i D(\hat{R}) = \hat{R}^i_j P_j, \quad (4.5.383) \]

where we have defined \( \hat{R}^i_j \equiv \hat{R}_j^i \).

Equations (4.5.381) and (4.5.383) 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 \( D(\hat{R}) \):

\[ D(\hat{R})^\dagger V^i D(\hat{R}) = \hat{R}^i_j V^j. \quad (4.5.384) \]

Although we shall focus on the \( D = 3 \) case here, note that this definition holds in arbitrary dimensions \( D \geq 3 \).

**Problem 4.71. Vector Operator in 3D: Infinitesimal Version** In 3D, show that if \( V^i \) is a vector operator obeying eq. (4.5.384), then it also obeys

\[ [J^a, V^b] = i\epsilon^{abc} V^c. \quad (4.5.385) \]

Can you argue, if eq. (4.5.385) holds, then so does eq. (4.5.384) – i.e., they are equivalent definitions of a vector operator? Hint: Recall equations (4.5.100) and (4.5.329).

**Remark** Notice, from eq. (4.5.385), that the angular momentum generators \( \{ J^a \} \) are themselves vector operators.

**Problem 4.72. 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:

\[ [J^a, \vec{V} \cdot \vec{W}] = 0. \quad (4.5.386) \]

Through eq. (4.5.100), this means \( D(\hat{R})^\dagger (\vec{V} \cdot \vec{W}) D(\hat{R}) = \vec{V} \cdot \vec{W} \).
Rotations in 4D. We will spend the next few sections studying 3D rotations in detail. But before we do so, let us briefly examine the 4D case – 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

\[
J^I = \frac{1}{2} \epsilon^{JKL} J^{KL} \quad \Leftrightarrow \quad \epsilon^{JKL} J^I = J^{JK}.
\] (4.5.387)

The remaining 3 generators of SO\(_4\) are then \(\{J^{I4}\}\). Like the preceding 3D case, we need to compute the Lie Algebra of these angular momentum operators. We already know from eq. (4.5.327) that

\[
[J^A, J^B] = i\epsilon^{ABC} J^C.
\] (4.5.388)

We therefore only need to figure out the commutation relations among the \(\{J^{I4}\}\) and between them and the \(\{J^I\}\). From eq. (4.5.269), we have

\[
[J^{A4}, J^{B4}] = -i \left( \delta^{[A[B} J^{C]4} - \delta^{[B} J^{4]A} \right).
\] (4.5.389)

Keeping in mind \(A, B \neq 4\) and \(J^{I4} = 0\) because of anti-symmetry,

\[
[J^{A4}, J^{B4}] = -i J^{BA} \quad \text{and} \quad [J^{A4}, J^{B4}] = i\epsilon^{ABC} J^C.
\] (4.5.390) (4.5.391)

Next, we do

\[
[J^{A4}, J^{BC}] = -i \left( \delta^{A[B} J^{C]4} - \delta^{[B} J^{4]C} \right) \quad \text{and} \quad [J^{A4}, J^{K}] = -i \frac{1}{2} \epsilon^{KBC} \left( \delta^{A[B} J^{C]4} - \delta^{[B} J^{4]C} \right).
\] (4.5.392) (4.5.393)

This leads us to

\[
[J^{A4}, J^{B}] = i\epsilon^{ABC} J^{C4}.
\] (4.5.394)

SO\(_4\) Lie Algebra. If \(A, B, C\) runs from 1 through 3 only, and if we remember the definition in eq. (4.5.387), the angular momentum operators in 4D obey the Lie Algebra in equations (4.5.388), (4.5.390) and (4.5.394).

**Problem 4.73. Two copies of SO\(_3\) Lie Algebra** Define

\[
M^I_\pm \equiv \frac{J^1 \pm J^{14}}{2}.
\] (4.5.395)

Show that

\[
[M^I_+, M^I_-] = 0 \quad \text{and} \quad [M^I_+, M^I_+] = i\epsilon^{JKL} M^K_\pm.
\] (4.5.396)
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,

\[
\begin{align*}
\vec{M}_\pm^2 \frac{\ell_\pm, m_+}{\ell_\pm, m_-} &= \ell_\pm(\ell_\pm + 1) \frac{\ell_\pm, m_+}{\ell_\pm, m_-}, & \vec{M}_\pm^2 \equiv M_\pm^1 M_\pm^1 \\
M_\pm^3 \frac{\ell_\pm, m_+}{\ell_\pm, m_-} &= m_\pm \frac{\ell_\pm, m_+}{\ell_\pm, m_-}
\end{align*}
\]

and \(m_\pm \in \{-\ell_\pm, -\ell_\pm + 1, \ldots, \ell_\pm - 1, \ell_\pm\}\).

\[\square\]

### 4.5.6 Rotations in 3 Spatial Dimensions: Integer Spin & Spherical Harmonics

\[31\text{In this section, we shall witness how the angular spherical harmonics introduced in equations (4.5.316) and (4.5.318) are in fact the position representation of the integer spin case (} \ell = 0, 1, 2, 3, \ldots \text{) in eq. (4.5.359) for 3D rotations. Specifically, if we apply the position eigenket } \langle r, \theta, \phi | \text{ written in spherical coordinates (} x^1, x^2, x^3 = r(\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta)) = r\hat{r}(\theta, \phi) \rangle \text{ on both sides of eq. (4.5.359):}

\[
\langle r, \theta, \phi | \vec{J}^2 | \ell, m \rangle = \ell(\ell + 1) \langle r, \theta, \phi | \ell, m \rangle.
\]

Recalling the result in eq. (4.5.292),

\[
-\vec{\nabla}_g^2 \langle r, \theta, \phi | \ell, m \rangle = \ell(\ell + 1) \langle r, \theta, \phi | \ell, m \rangle.
\]

Notice, when \(D = 3\), eq. (4.5.318) reads

\[
-\vec{\nabla}_g^2 Y^m_\ell(\theta, \phi) = \ell(\ell + 1) Y^m_\ell(\theta, \phi).
\]

Below, we shall identify

\[
Y^m_\ell(\theta, \phi) = \langle \theta, \phi | \ell, m \rangle.
\]

Firstly, if we convert Cartesian to Spherical coordinates via eq. (4.5.399), the metric in 3D flat space becomes

\[
d\ell^2 = \delta_{ij} dx^i dx^j = \delta_{ij} \frac{\partial x^i}{\partial (r, \theta, \phi)} \frac{\partial x^j}{\partial (r, \theta, \phi)} (dr, d\theta, d\phi)^a (dr, d\theta, d\phi)^b (dr, d\theta, d\phi)^a (dr, d\theta, d\phi)^b \cdot (dr, d\theta, d\phi)^b \cdot (dr, d\theta, d\phi)^b \cdot (dr, d\theta, d\phi)^b \cdot (dr, d\theta, d\phi)^b \cdot (dr, d\theta, d\phi)^b
\]

\[= dr^2 + r^2 H_{ij} d\theta^i d\theta^j, \quad H_{ij} d\theta^i d\theta^j = d\theta^2 + \sin(\theta)^2 d\phi^2.
\]

The square root of the determinant is

\[
\sqrt{g} = r^2 \sqrt{H} = r^2 \sin \theta;
\]

\[31\text{This and the next 2 sections are under heavy construction.}\]

\[32\text{The above calculation is described in more detail in §(7.1).}\]
the non-zero components of the inverse metric are
\[ g_{rr} = 1, \quad g_{\theta\theta} = r^{-2} H^{\theta\theta} = r^{-2}, \quad g_{\phi\phi} = r^{-2} H^{\phi\phi} = (r \sin(\theta))^{-2}. \] (4.5.408)
Therefore, the Laplacian is
\[ \vec{\nabla}^2 \psi = \frac{1}{r^2 s_\theta} \left( \partial_r \left( r^2 s_\theta \partial_r \psi \right) + \partial_\theta \left( r^2 s_\theta r^{-2} \partial_\theta \psi \right) + \partial_\phi \left( r^2 s_\theta (r s_\theta)^{-2} \partial_\phi \psi \right) \right) \] (4.5.409)
\[ = \frac{1}{r^2} \partial_r \left( r^2 \partial_r \psi \right) + \frac{1}{r^2} \vec{\nabla}^2_\mathbb{S}^2 \psi, \] (4.5.410)
where \( \vec{\nabla}^2_\mathbb{S}^2 \) is the Laplacian on the 2-sphere of unit radius,
\[ - \langle \vec{x} | J^2 | \psi \rangle = \vec{\nabla}^2_\mathbb{S}^2 \psi = \frac{1}{\sin(\theta)} \left( \partial_\theta (\sin(\theta) \partial_\theta \psi) + \frac{1}{\sin(\theta)} \partial_\phi^2 \psi \right). \] (4.5.411)
We may directly infer from equations (4.5.281) and (4.5.319) that in 3D, the position representation of the generators of rotations (aka “angular momentum operators”) are
\[ \langle r, \theta, \phi | J^1 | \psi \rangle = i \sin(\phi) \partial_\theta + \cos(\phi) \cot(\theta) \partial_\phi \langle r, \theta, \phi | \psi \rangle, \] (4.5.412)
\[ \langle r, \theta, \phi | J^2 | \psi \rangle = i \left( - \cos(\phi) \partial_\theta + \sin(\phi) \cot(\theta) \partial_\phi \right) \langle r, \theta, \phi | \psi \rangle, \] (4.5.413)
\[ \langle r, \theta, \phi | J^3 | \psi \rangle = -i \partial_\phi \langle r, \theta, \phi | \psi \rangle. \] (4.5.414)

Problem 4.74. 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:
\[ (\vec{A} \times \vec{B})^i = \epsilon^{ijk} A^j B^k. \] (4.5.415)
For instance, \( (\vec{A} \times \vec{B})^1 = \epsilon^{1jk} A^j B^k = \epsilon^{123} A^2 B^3 + \epsilon^{132} A^3 B^2 = A^2 B^3 - A^3 B^2. \)

Problem 4.75. Orbital Angular Momentum Operators In the spherical coordinate system defined in eq. (4.5.399), show that the angular momentum operators, i.e., the generators of rotation in 3D, are
\[ \langle r, \theta, \phi | J^1 | \psi \rangle = i \left( \sin(\phi) \partial_\theta + \cos(\phi) \cot(\theta) \partial_\phi \right) \langle r, \theta, \phi | \psi \rangle, \] (4.5.416)
\[ \langle r, \theta, \phi | J^2 | \psi \rangle = i \left( - \cos(\phi) \partial_\theta + \sin(\phi) \cot(\theta) \partial_\phi \right) \langle r, \theta, \phi | \psi \rangle, \] (4.5.417)
\[ \langle r, \theta, \phi | J^3 | \psi \rangle = -i \partial_\phi \langle r, \theta, \phi | \psi \rangle. \] (4.5.418)
In turn, deduce that the position representations of the ladder operators in eq. (4.5.331) are
\[ \langle r, \theta, \phi | J^\pm | \psi \rangle = e^{\pm i \phi} \left( \pm \partial_\theta + i \cot(\theta) \partial_\phi \right) \langle r, \theta, \phi | \psi \rangle. \] (4.5.419)
Hint: Recall equations (4.5.309) and (4.5.319).
Spherical Harmonics in 3D

Let us now turn to solving the spherical harmonics in 3D, and the associated eigenfunctions of $J^2$ – recall equations (4.5.316) and (4.5.318). Remember, since $[J^3, J^2] = 0$, we must be able to simultaneously diagonalize $J^3$ and $J^2$. In fact, since $\langle r, \theta, \phi | J^3 | \psi \rangle = -i \partial_\phi \langle r, \theta, \phi | \psi \rangle$, we must have

$$\langle r, \theta, \phi | J^3 | \ell, m \rangle = m \langle r, \theta, \phi | \ell, m \rangle,$$  
(4.5.420)

$$-i \partial_\phi \langle r, \theta, \phi | \ell, m \rangle = m \langle r, \theta, \phi | \ell, m \rangle.$$  
(4.5.421)

The solution to the second line is the solution to

$$-i \partial_\phi f(\phi) = mf(\phi) \Rightarrow f(\phi) = f_0 \exp(im\phi),$$

except in our case $f_0$ can still depend on $\theta$ and other parameters in the problem. This implies the angular spherical harmonics takes the form

$$Y_{\ell}^m(\theta, \phi) = \langle r, \theta, \phi | \ell, m \rangle = \langle \theta | \ell, m \rangle \exp(im\phi).$$  
(4.5.422)

Next, we recall the discussions around equations (4.5.347) and (4.5.348), 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. (4.5.419) and (4.5.422), we may write the position representation of eq. (4.5.347) as

$$e^{i\phi} (\partial_\theta + i \cot(\theta) \partial_\phi) Y_{\ell}^\ell(\theta, \phi) = e^{i(\ell+1)\phi} (\partial_\theta - \ell \cdot \cot(\theta)) \langle \theta | \ell, \ell \rangle = 0.$$  
(4.5.423)

Using the results in eq. (4.5.419) and (4.5.422), we may write the position representation of eq. (4.5.348) as

$$e^{-i\phi} (-\partial_\theta + i \cot(\theta) \partial_\phi) Y_{\ell}^{-\ell}(\theta, \phi) = e^{-i(\ell+1)\phi} (-\partial_\theta + \ell \cdot \cot(\theta)) \langle \theta | \ell, -\ell \rangle = 0.$$  
(4.5.424)

**Problem 4.76.** Solve equations (4.5.423) and (4.5.424) to show that

$$Y_{\ell}^{\pm \ell}(\theta, \phi) = \chi_{\ell}^{\pm \ell} \sin^{\ell}(\theta) \exp(\pm i\ell\phi),$$  
(4.5.425)

where the $\chi_{\ell}^{\pm \ell}$ are $(\theta, \phi)$-independent.

For $m \geq 0$,

$$Y_{\ell}^m(\theta, \phi) = (-)^\ell \sqrt{\frac{2\ell + 1}{4\pi} \cdot \frac{(\ell + m)!}{(\ell - m)!} \sin^{m}(\theta)} \left(\frac{d}{d \cos(\theta)}\right)^{\ell-m} (1 - \cos^2(\theta))^{\ell}.$$  
(4.5.426)

Define negative $m$ via

$$Y_{\ell}^{-m} = (-)^m Y_{\ell}^m(\theta, \phi).$$  
(4.5.427)

If the function $\langle r, \phi | \psi \rangle$ is defined over the entire circle $\phi \in [0, 2\pi)$ for all $\theta$, and if we assume it is continuous, then

$$\langle \theta, \phi + 2\pi | \psi \rangle = \langle \theta, \phi | \psi \rangle$$  
(4.5.428)
which in turn means \( \exp(\pm im(\phi + 2\pi)) = \exp(\pm im\phi) \). Thus, \( m \) must be an integer. The spherical harmonics in 3D must therefore take the form
\[
r^\ell Y^m_\ell(\theta, \phi) = r^\ell P^m_\ell(\cos \theta)e^{im\phi}, \quad m = 0, \pm 1, \pm 2, \pm 3, \ldots
\]
Inserting this result into equations (4.5.318) and (4.5.411), we obtain the ordinary differential equation for the associated Legendre function:
\[
\partial_c ((1 - c^2)\partial_c P^m_\ell(c)) + \left( \ell(\ell + 1) - \frac{m^2}{1 - c^2} \right) P^m_\ell(c) = 0, \quad c \equiv \cos \theta.
\]

**Problem 4.77.** Verify eq. (4.5.430) from the \( D = 3 \) version of eq. (4.5.318).

We already know from the discussion in §(4.5.5) that \( r^\ell Y^m_\ell(\theta, \phi) \) is a homogeneous polynomial of degree \( \ell \geq 0 \). The \( \{P^m_\ell(c)\} \) may be constructed via the following means.

### Spherical Harmonics as Homogeneous Polynomials

We will turn to a different manner to obtain the spherical harmonics \( Y^m_\ell \), by viewing them as the angular portion of homogeneous polynomial of degree \( \ell \) which satisfies the (homogeneous) Laplace equation (4.5.314) (or, equivalently eq. (4.5.316)).

\[
\nabla^2 (r^\ell Y^m_\ell(\theta, \phi)) = 0
\]
To this end, let us define
\[
x^\pm = x^1 \pm ix^2 = r \sin(\theta)e^{\pm i\phi}.
\]
Let us consider homogeneous polynomials of degree \( \ell \) by superposing the products of positive powers of \( x^\pm \) and \( x^3 \), namely
\[
r^\ell Y^m_\ell = \psi_{a_+, a_-}(x^+)^{a_+}(x^-)^{a_-}(x^3)^b = \psi_{a_+, a_-} \cdot r^{a_+ + a_- + b}(\sin(\theta))^{a_+ + a_-} \cos^b(\theta) \exp(i(a_+ - a_-)\phi).
\]
In order to obtain a degree \( \ell \) polynomial, the sum of the powers must yield \( \ell \).
\[
a_+ + a_- + b = \ell
\]
To achieve this, for a fixed \( a_+ \), we may choose \( a_- \{0, 1, 2, \ldots, \ell - a_+\} \); followed by putting \( b = \ell - a_+ - a_- \). Therefore, the total number of independent terms in eq. (4.5.433) is
\[
N_\ell = \sum_{a_+=0}^{a_+} \sum_{a_-=0}^{\ell-a_+} 1 = \sum_{a_+=0}^{\ell} (\ell - a_+ + 1) = (\ell + 1)^2 - \frac{0 + \ell}{2}(\ell + 1) = \frac{(\ell + 1)(\ell + 2)}{2}.
\]

**Problem 4.78.** Show that the Laplacian acting on an arbitrary function \( \psi(x^+, x^-, x^3) \) is
\[
\delta^{ij}\partial_i \partial_j \psi = (4\partial_+ \partial_- + \partial_3^2)\psi,
\]
where \( \partial_\pm \) is the derivative with respect to \( x^\pm \equiv x^1 \pm ix^2 \).

\[\text{Part of the discussion here is modeled after the one in Weinberg [13].}\]
Inserting eq. (4.5.433) into eq. (4.5.314), one would find

\[
\tilde{\nabla}^2 \left( \psi_{a_+a_-b}(x^+)^{a_+}(x^-)^{-b} - (x^3)^b \right) = \psi_{a_+a_-b}(x^+)^{a_+}(x^-)^{-b} - (x^3)^b = 0. \tag{4.5.438}
\]

This explicitly demonstrates that the Laplacian acting on a homogeneous polynomial of degree \(\ell\) is a homogeneous polynomial of degree \(\ell - 2\). Since the latter has \(N_{\ell-2}\) independent terms (by eq. (4.5.436)), that means eq. (4.5.438) provides us \(N_{\ell-2}\) constraints to be obeyed by the \(N_\ell\) independent terms of eq. (4.5.433). Therefore, there must actually be

\[
N_\ell - N_{\ell-2} = \frac{(\ell + 1)(\ell + 2)}{2} - \frac{(\ell - 1)\ell}{2} = 2\ell + 1 \tag{4.5.439}
\]

independent terms in the most general homogeneous polynomial of degree \(\ell\) that solves eq. (4.5.314) in 3D.

But, as we have already discovered, \(2\ell + 1\) is exactly the number of linearly independent spherical harmonics \(\{Y^m_\ell | m = -\ell, \ldots, +\ell\}\) for a fixed \(\ell\). This indicates the solutions of eq. (4.5.314) in 3D must, up to an overall multiplicative constant, be the \(Y^m_\ell\) themselves. In fact, let us define

\[
m \equiv a_+ - a_- \tag{4.5.440}
\]

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

\[
\max(a_+ - a_-) = \ell \quad \text{and} \quad \min(a_+ - a_-) = -\ell. \tag{4.5.441}
\]

In the other words,

\[
-\ell \leq m \leq +\ell. \tag{4.5.442}
\]

By taking into account equations (4.5.435) and (4.5.440), eq. (4.5.433) now reads

\[
Y^m_\ell(\theta, \phi) = \frac{1}{r^{\ell}} \sum_{a_+ + a_- + b = \ell, a_+ - a_- = m} \psi'_b(x^+)^{a_+}(x^-)^{-b} - (x^3)^b \left( a_- = \frac{\ell - b \pm m}{2} \right) \tag{4.5.443}
\]

\[
= \sum_b \psi'_b \cdot \sin^{\ell-b}(\theta) \cos^b(\theta)e^{im\phi}, \tag{4.5.444}
\]

for appropriate coefficients \(\{\psi'_b\}\). The \(\exp(im\phi)\) indicates it obeys the equivalent of eq. (4.5.421), namely

\[
-i\partial_\phi Y^m_\ell = mY^m_\ell. \tag{4.5.445}
\]

Furthermore, from our analysis, these \(\{Y^m_\ell(\theta, \phi)\}\) must be proportional to the corresponding \(\{Y^m_\ell(\theta, \phi)\}\); since they correspond to the same number of independent solutions to

\[
-\tilde{\nabla}^2 Y^m_\ell = \ell(\ell + 1)Y^m_\ell. \tag{4.5.446}
\]
To sum: in 3D, the \( r^\ell Y^m_\ell (\theta, \phi) \), when expressed in Cartesian coordinates \( \vec{x} \), are homogeneous polynomials of degree \( \ell \) satisfying equations (4.5.445) and (4.5.446).

**Example: \( \ell = 0 \)** For \( \ell = 0 \), this corresponds to having zero powers of the coordinates — i.e., a constant: i.e., \( Y^0_0 = \text{constant} \).

**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

\[
r Y^{\pm 1}_1 \propto x^\pm = r \sin(\theta) e^{\pm i\phi}, \quad r Y^0_1 \propto x^3 = r \sin(\theta).
\] (4.5.447)

**Example: \( \ell = 2 \)** For \( \ell = 2 \), we have the possibilities

\[
(a_+, a_-, b) = (2, 0, 0) \Rightarrow m = 2
\]

(4.5.448)

\[
(a_+, a_-, b) = (0, 2, 0) \Rightarrow m = -2
\]

(4.5.449)

\[
(a_+, a_-, b) = (0, 0, 2) \Rightarrow m = 0
\]

(4.5.450)

\[
(a_+, a_-, b) = (1, 1, 0) \Rightarrow m = 0
\]

(4.5.451)

\[
(a_+, a_-, b) = (1, 0, 1) \Rightarrow m = 1
\]

(4.5.452)

\[
(a_+, a_-, b) = (0, 1, 1) \Rightarrow m = -1.
\]

(4.5.453)

Here, the “\( \Rightarrow \)” means the term \((x^+)^a (x^-)^b (x^3)^b\) 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 4.79.** Normalize the spherical harmonics to unity on the sphere, i.e.,

\[
\int_{-1}^{+1} d(cos \theta) \int_0^{2\pi} d\phi |Y^m_\ell (\theta, \phi)|^2 = 1.
\] (4.5.454)

Proceed to compute \( Y^m_\ell \) (up to a multiplicative phase) for \( \ell = 0, 1, 2 \) by demanding they satisfy the homogeneous equations (4.5.314) and (4.5.316). Hint: The answers can be found in equations (9.2.67), (9.2.68) and (9.2.69) below.

### 4.5.7 Rotations in 3 Spatial Dimensions: Half Integer Spin, SU\(_2\)

**Spin-1/2** From the Pauli matrices in eq. (3.2.17), which obey the algebra in eq. (4.3.113), we see that

\[
\left[ \frac{\sigma^i}{2}, \frac{\sigma^j}{2} \right] = \frac{i}{4} (\epsilon^{ijk} - \epsilon^{jik}) \sigma^k = i \epsilon^{ijk} \frac{\sigma^k}{2}.
\] (4.5.455)

We see that, the identification

\[
J^i \equiv \frac{\sigma^i}{2}, \quad i \in \{1, 2, 3\}
\] (4.5.456)

may be used as a 2D representation of the Lie algebra of the rotation group in 3 spatial dimensions. **YZ: Topology. Spinors.**
Using raising/lowering operators,
\[ J^+ \left| \frac{1}{2}, \frac{1}{2} \right\rangle = 0 \] (4.5.457)
\[ J^- \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = 0 \] (4.5.458)
we may obtain
\[ \langle \theta, \phi \left| \frac{1}{2}, \pm \frac{1}{2} \right\rangle \propto \sqrt{\sin(\theta)} \exp(\pm i\phi/2). \] (4.5.459)
But
\[ \langle \theta, \phi \left| J^- \left| \frac{1}{2}, \frac{1}{2} \right\rangle \propto e^{-\frac{i}{2} \phi \cot(\theta)} \sqrt{\sin(\theta)} \] (4.5.460)

4.5.8 Rotations in 3 Spatial Dimensions: ‘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 \{\left| \ell_1 m_1, \ell_2 m_2 \right\rangle\} formed from the tensor product of two separate angular momentum spaces:
\[ \left| \ell_1 m_1, \ell_2 m_2 \right\rangle \equiv \left| \ell_1 m_1 \right\rangle \otimes \left| \ell_2 m_2 \right\rangle; \] (4.5.461)
where the \left| \ell_1 m_1 \right\rangle and \left| \ell_2 m_2 \right\rangle are the eigenvectors of two separate sets of angular momentum operators \{\vec{J}^{\ell_1} = J^{\ell_1}, J^{\ell_3}\} and \{\vec{J}^{\ell_2} = J^{\ell_2}, J^{\ell_3}\}. Namely, we have 2 distinct sets of eq. (4.5.359):
\[ \vec{J}^{\ell_1} \left| \ell_1 m_1 \right\rangle = \ell_1 (\ell_1 + 1) \left| \ell_1 m_1 \right\rangle \] (4.5.462)
\[ J^{\ell_3} \left| \ell_1 m_1 \right\rangle = m_1 \left| \ell_1 m_1 \right\rangle \] (4.5.463)
and
\[ \vec{J}^{\ell_2} \left| \ell_2 m_2 \right\rangle = \ell_2 (\ell_2 + 1) \left| \ell_2 m_2 \right\rangle \] (4.5.464)
\[ J^{\ell_3} \left| \ell_2 m_2 \right\rangle = m_2 \left| \ell_2 m_2 \right\rangle. \] (4.5.465)

The situation represented by eq. (4.5.461) 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
\[ J^i \equiv J^{\ell_1} \otimes \mathbb{I} + \mathbb{I} \otimes J^{\ell_2}. \] (4.5.466)
Oftentimes, the \otimes is dropped for notational convenience.
\[ J^i \equiv J^{\ell_1} + J^{\ell_2}. \] (4.5.467)
By assumption, the 2 separate sets of angular momentum operators commute

\[ [J^a, J^b] = 0. \]  

(4.5.468)

This allows us to see that, the exponential of the total angular momentum operator yields the rotation operator that implements rotations on the states \( \{|\ell_1 m_1, \ell_2 m_2\}\).

\[ \exp(-i\theta^a J^a) |\ell_1 m_1, \ell_2 m_2\rangle = \exp(-i\theta^a (J^a + J^b)) |\ell_1 m_1, \ell_2 m_2\rangle \]  

(4.5.469)

\[ = \left( \exp\left(-i\theta^a J^a\right) |\ell_1 m_1\rangle \right) \otimes \left( \exp\left(-i\theta^b J^b\right) |\ell_2 m_2\rangle \right) \]  

(4.5.470)

\[ = \left( D(\hat{R}(\theta)) |\ell_1 m_1\rangle \right) \otimes \left( D(\hat{R}(\theta')) |\ell_2 m_2\rangle \right). \]  

(4.5.471)

In words: rotating the state \( |\ell_1 m_1, \ell_2 m_2\rangle \) means simultaneously rotating the \( |\ell_1 m_1\rangle \) and \( |\ell_2 m_2\rangle \); this is precisely what exponential of the total angular momentum operator does. Note that \( \exp(-i\theta^a J^a + J^b) = \exp(-i\theta^a J^a) \exp(-i\theta^b J^b) \) because \( J^a \) commutes with \( J^b \).

**Eigensystems** We may see from equations (4.5.462) through (4.5.465) that the tensor product state \( |\ell_1 m_1, \ell_2 m_2\rangle \) are, too, eigenstates of

\[ \{\hat{J}^2, J^3, \hat{J}^2, J^3\}. \]  

(4.5.472)

For example,

\[ \hat{J}^2 |\ell_1 m_1, \ell_2 m_2\rangle = \hat{J}^2 |\ell_1 m_1\rangle \otimes |\ell_2 m_2\rangle \]

\[ = \ell_1(\ell_1 + 1) |\ell_1 m_1, \ell_2 m_2\rangle \]  

(4.5.473)

and

\[ J^3 |\ell_1 m_1, \ell_2 m_2\rangle = m_1 |\ell_1 m_1, \ell_2 m_2\rangle. \]  

(4.5.474)

Likewise

\[ \hat{J}^2 |\ell_1 m_1, \ell_2 m_2\rangle = \ell_2(\ell_2 + 1) |\ell_1 m_1, \ell_2 m_2\rangle \]  

(4.5.475)

\[ \hat{J}^2 |\ell_1 m_1, \ell_2 m_2\rangle = m_2 |\ell_1 m_1, \ell_2 m_2\rangle. \]  

(4.5.476)

We will now proceed to argue that, instead of the mutually compatible observables in eq. (4.5.472), one may also pick the set

\[ \{\hat{J}^2, J^3, \hat{J}^2, J^3\}. \]  

(4.5.477)

Their simultaneous eigenstates will be denoted as \( \{|j m; \ell_1 \ell_2\}\), obeying the relations

\[ \hat{J}^2 |j m; \ell_1 \ell_2\rangle = j(j + 1) |j m; \ell_1 \ell_2\rangle, \]  

(4.5.478)

\[ J^3 |j m; \ell_1 \ell_2\rangle = m |j m; \ell_1 \ell_2\rangle. \]  

(4.5.479)

\[ \hat{J}^2 |j m; \ell_1 \ell_2\rangle = \ell_1(\ell_1 + 1) |j m; \ell_1 \ell_2\rangle \]  

(4.5.480)

\[ \hat{J}^2 |j m; \ell_1 \ell_2\rangle = \ell_2(\ell_2 + 1) |j m; \ell_1 \ell_2\rangle. \]  

(4.5.481)

The total angular momentum \( j \) will turn out to be restricted within the range

\[ j \in \{|\ell_1 - \ell_2|, |\ell_1 - \ell_2| + 1, \ldots, \ell_1 + \ell_2 - 1, \ell_1 + \ell_2\}. \]  

(4.5.482)

and, of course,

\[ m \in \{-j, -j + 1, \ldots, j - 1, j\}. \]  

(4.5.483)
Problem 4.80. Explain why the total angular momentum generators still obey the Lie Algebra in eq. (4.5.327). That is, verify
\[ [J^a, J^b] = i\epsilon^{abc} J^c. \]  
(4.5.484)

From the discussions in the previous sections, we see that upon diagonalization, equations (4.5.478), (4.5.479) and (4.5.483) follow. \[ \square \]

There are \( \binom{4}{2} = 6 \) commutators to check, in order to confirm eq. (4.5.477) consists of a set of mutually compatible observables:
\[ [\vec{J}_2, J^3], [J^2, \vec{J}^2], [\vec{J}^2, \vec{J}^3], [J^3, \vec{J}^2], [\vec{J}^3, \vec{J}^2], \text{ and } [\vec{J}^2, \vec{J}^2]. \]  
(4.5.485)

That \( [\vec{J}^2, J^3] = 0 \) follows from the SO\(_3\) Lie algebra in eq. (4.5.484); \( \vec{J}^2 \) is a scalar under rotating along the 3–axis. Next, we may readily see that \( [\vec{J}^2, \vec{J}^2] = 0 = [\vec{J}^2, \vec{J}^{n2}] \). Simply expand out \( \vec{J}^2 = (\vec{J} + \vec{J}^n)^2 \); for instance,
\[ [\vec{J}^2, \vec{J}^2] = [\vec{J}^2 + \vec{J}^{n2} + 2\vec{J}^i \cdot \vec{J}^n, \vec{J}^2] \]  
(4.5.486)
\[ = 2 [\vec{J}^i \cdot \vec{J}^n, \vec{J}^2] = 0. \]  
(4.5.487)

The last equality follows because of eq. (4.5.386); i.e., \( \vec{J}^2 \) is a scalar under rotations generated by \( \vec{J} \). A similar line of reasoning leads to
\[ [\vec{J}^2, \vec{J}^{n2}] = 0. \]  
(4.5.488)

That \( [J^3, \vec{J}^i] = 0 = [J^3, \vec{J}^n] \) follows from the same reason why \( [\vec{J}^2, J^3] = 0 \), since the prime and double-primed operators are already assumed to commute.
\[ [J^3, \vec{J}^i] = [J^3, \vec{J}^n] = 0 \text{ and } [J^3, J^m] = [J^{n3}, \vec{J}^i] = 0. \]  
(4.5.489)

Finally, \( [\vec{J}^2, \vec{J}^{n2}] = 0 \) also by assumption.

Problem 4.81. Note, however, that none of the individual components of \( J^n \) or \( J^{m} \) commute with \( \vec{J}^2 \). Show that
\[ [\vec{J}^2, J^n] = -2i(\vec{J}^n \times \vec{J}^i), \]  
(4.5.490)
\[ [\vec{J}^2, J^m] = -2i(\vec{J} \times \vec{J}^n)^i; \]  
(4.5.491)
where, for vector operators \( \vec{A} \) and \( \vec{B} \), we have defined
\[ (\vec{A} \times \vec{B})^i \equiv \epsilon^{iab} A^a B^b. \]  
(4.5.492)

Recalling the discussion in Problem (4.5.386), we see these commutators are non-zero because \( J^n \) generates rotation only on the \( |\ell_1, m_1\rangle \) space; and \( J^{m} \) only the \( |\ell_2, m_2\rangle \) space. Hence, only the \( \vec{J}^i \) operators in \( \vec{J}^2 \) are altered for the former; and only the \( \vec{J}^n \) 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

\[ | j m; \ell_1 \ell_2 \rangle = \sum_{-\ell_1 \leq m_1 \leq \ell_1} \sum_{-\ell_2 \leq m_2 \leq \ell_2} | \ell_1 m_1, \ell_2 m_2 \rangle \langle \ell_1 m_1, \ell_2 m_2 | j m; \ell_1 \ell_2 \rangle. \tag{4.5.493} \]

There is no sum over the \( \ell \)'s, 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 \( (\ell_1, \ell_2) \) subspace, we therefore have

\[ \sum_{-\ell_1 \leq m_1 \leq \ell_1} \sum_{-\ell_2 \leq m_2 \leq \ell_2} | \ell_1 m_1, \ell_2 m_2 \rangle \langle \ell_1 m_1, \ell_2 m_2 | = I. \tag{4.5.494} \]

To begin, let us first notice that

\[ J^3 | j m; \ell_1 \ell_2 \rangle = m | j m; \ell_1 \ell_2 \rangle \tag{4.5.495} \]

\[ = \sum_{m'_{1,2}} (J^3 + J'^3) | \ell'_1 m'_1, \ell'_2 m'_2 \rangle \langle \ell'_1 m'_1, \ell'_2 m'_2 | j m; \ell_1 \ell_2 \rangle \tag{4.5.496} \]

\[ = \sum_{m'_{1,2}} (m'_{1} + m'_{2}) | \ell'_1 m'_1, \ell'_2 m'_2 \rangle \langle \ell'_1 m'_1, \ell'_2 m'_2 | j m; \ell_1 \ell_2 \rangle. \tag{4.5.497} \]

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. (4.5.493) must be constrained by

\[ m = m_1 + m_2. \tag{4.5.498} \]

Now, the largest possible \( m \), which is also the maximum \( j \) (cf. (4.5.483)), is gotten from \( \max(m_1 + m_2) = \max m_1 + \max m_2 = \ell_1 + \ell_2 \).

\[ \max m = \ell_1 + \ell_2 = \max j. \tag{4.5.499} \]

A similar argument informs us, \( \min m = \min m_1 + \min m_2 = -(\ell_1 + \ell_2) \). Altogether,

\[ | j = \ell_1 + \ell_2 \mid m = \pm(\ell_1 + \ell_2); \ell_1 \ell_2 \rangle = | \ell_1 \pm \ell_1, \ell_2 \pm \ell_2 \rangle = | \ell_1, \pm \ell_1 \rangle \otimes | \ell_2, \pm \ell_2 \rangle. \tag{4.5.500} \]

**Problem 4.82. Checking eq. (4.5.500)**

Defining the total raising (+) and lowering (−) operator as

\[ J^\pm \equiv J^1 \pm iJ^2 = J'^\pm + J''\pm, \tag{4.5.501} \]

verify the relation

\[ \vec{J}^2 = \vec{J}^2 + \vec{J}'^2 + 2(J^3)(J'^3) + J'^+ J'^- + J^+ J'^-. \tag{4.5.502} \]

Use it to directly calculate the result of acting \( \vec{J}^2 \) on both sides of eq. (4.5.500). \( \Box \)
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. (4.5.501). We recall eq. (4.5.366):

\[
J^\pm \ket{m; \ell_1 \ell_2} = \sqrt{j(j+1)-m(m\pm1)} \ket{m; \ell_1 \ell_2} \quad \text{with} \quad j \equiv \ell_1 + \ell_2.
\]

(4.5.503)

On the other hand, using \(J^\pm = J'^\pm + J''^\pm\),

\[
\sqrt{(j \mp m)(j \pm m+1)} \ket{m; \ell_1 \ell_2} = \sum_{m_1, m_2} (J'^\pm + J''^\pm) \ket{m_1, \ell_1 \ell_2} \langle m_1, \ell_1 \ell_2 | m; \ell_1 \ell_2 \rangle
\]

(4.5.505)

Let us now apply the lowering operator to the maximum possible \(m\) state in eq. (4.5.500). For \(j = \ell_1 + \ell_2\),

\[
\sqrt{(j + j)(j - j + 1)} \ket{j; \ell_1 \ell_2} = \sqrt{(\ell_1 + \ell_1)(\ell_1 - \ell_1 + 1)} \ket{\ell_1 \ell_1 - 1, \ell_2 \ell_2}
\]

(4.5.507)

\[
+ \sqrt{(\ell_2 + \ell_2)(\ell_2 - \ell_2 + 1)} \ket{\ell_1 \ell_1, \ell_2 \ell_2 - 1}.
\]

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:

\[
\ket{j - 1; \ell_1 \ell_2} = \frac{1}{\sqrt{j}} \left( \sqrt{\ell_1} \ket{\ell_1 \ell_1 - 1, \ell_2 \ell_2} + \sqrt{\ell_2} \ket{\ell_1 \ell_1, \ell_2 \ell_2 - 1} \right),
\]

(4.5.508)

(4.5.509)

We may continue this ‘lowering procedure’ to obtain all the \(m\) states \(|j = \ell_1 + \ell_2; \ell_1 \ell_2\rangle\) until we reach \(|j - j; \ell_1 \ell_2\rangle \propto (J^-)^2 |j; \ell_1 \ell_2\rangle\).

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. (4.5.498) 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. (4.5.498), 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:

\[
|j = \ell_1 + \ell_2 - 1; \ell_1 \ell_2\rangle = |\ell_1 \ell_1 - 1, \ell_2 \ell_2\rangle \langle \ell_1 \ell_1 - 1, \ell_2 \ell_2 | j; \ell_1 \ell_2 \rangle
\]

(4.5.510)

\[
+ |\ell_1 \ell_1, \ell_2 \ell_2 - 1\rangle \langle \ell_1 \ell_1, \ell_2 \ell_2 - 1 | j; \ell_1 \ell_2 \rangle.
\]

\(\text{Above in eq. (4.5.508), we have already constructed the highest-}\) j state with the same \(m\) value as the next-to-highest-\(j\) state in eq. (4.5.510), namely \(|j = \ell_1 + \ell_2 j - 1; \ell_1 \ell_2\rangle\), which has the

---

\(^{34}\text{Actually, we know from the preceding arguments that }|j = \ell_1 + \ell_2 - j; \ell_1 \ell_2\rangle = |\ell_1 - \ell_1, \ell_2 - \ell_2\rangle.\text{ But if you do push the analysis all the way till } (J^-)^2 |j; \ell_1 \ell_2\rangle,\text{ this would serve as a consistency check.}\)
same \( m \) value as \( |j = \ell_1 + \ell_2 - 1 \, j; \ell_1 \ell_2 \rangle \). These two states must be orthogonal because they have different \( \vec{J}^2 \) eigenvalues. Taking their inner product and setting it to zero,

\[
\sqrt{\ell_1} \langle \ell_1 \, \ell_1 - 1, \ell_2 \, \ell_2 | j; \ell_1 \ell_2 \rangle + \sqrt{\ell_2} \langle \ell_1 \, \ell_1, \ell_2 \, \ell_2 - 1 | j; \ell_1 \ell_2 \rangle = 0. \tag{4.5.511}
\]

Inserting it back into eq. (4.5.510),

\[
| j = \ell_1 + \ell_2 - 1 \, j; \ell_1 \ell_2 \rangle = \left( -\sqrt{\frac{\ell_2}{\ell_1}} | \ell_1 \, \ell_1 - 1, \ell_2 \, \ell_2 \rangle + \sqrt{\frac{\ell_1}{\ell_2}} | \ell_1 \, \ell_1, \ell_2 \, \ell_2 - 1 \rangle \right) \times \langle \ell_1 \, \ell_1, \ell_2 \, \ell_2 - 1 | j; \ell_1 \ell_2 \rangle. \tag{4.5.512}
\]

Since this state needs to be normalized to unity, we have up to an arbitrary phase \( e^{i\delta_{\ell_1+\ell_2-1}} \),

\[
| j = \ell_1 + \ell_2 - 1 \, j; \ell_1 \ell_2 \rangle = \frac{e^{i\delta_{\ell_1+\ell_2-1}}}{\sqrt{\ell_1 + \ell_2}} \left( -\sqrt{\frac{\ell_2}{\ell_1}} | \ell_1 \, \ell_1 - 1, \ell_2 \, \ell_2 \rangle + \sqrt{\frac{\ell_1}{\ell_2}} | \ell_1 \, \ell_1, \ell_2 \, \ell_2 - 1 \rangle \right). \tag{4.5.513}
\]

As before, we may then apply the lowering operator repeatedly to obtain all the \( j = \ell_1 + \ell_2 - 1 \) states, namely

\[
| j = \ell_1 + \ell_2 - 1 \, j - s; \ell_1 \ell_2 \rangle \propto (J^-)^s | j = \ell_1 + \ell_2 - 1 \, j; \ell_1 \ell_2 \rangle. \tag{4.5.514}
\]

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 \((m_1, m_2)s – \)

\[
(m_1 = \ell_1 - 2, m_2 = \ell_2), \quad (m_1 = \ell_1 - 1, m_2 = \ell_2 - 1), \quad (m_1 = \ell_1, m_2 = \ell_2 - 2). \tag{4.5.515}
\]

The \( | j = \ell_1 + \ell_2 - 2 \, j; \ell_1 \ell_2 \rangle \) must be perpendicular to both

\[
| j = \ell_1 + \ell_2 \, m = j - 2 \rangle \text{ and } | j = \ell_1 + \ell_2 - 1 \, m = j - 1 \rangle \tag{4.5.516}
\]

because they have different \( \vec{J}^2 \) eigenvalues. Setting to zero

\[
\langle j = \ell_1 + \ell_2 \, m = j - 2 | j' = \ell_1 + \ell_2 - 2 \, m' = j' \rangle \text{ and } \langle j = \ell_1 + \ell_2 - 1 \, m = j - 1 | j' = \ell_1 + \ell_2 - 2 \, m' = j' \rangle \tag{4.5.517}
\]

yields 2 equations for 3 unknown Clebsch-Gordan coefficients

\[
\langle \ell_1 \, \ell_1 - 2, \ell_2 \, \ell_2 | j = \ell_1 + \ell_2 - 2 \, j; \ell_1 \ell_2 \rangle, \tag{4.5.518}
\]

\[
\langle \ell_1 \, \ell_1 - 1, \ell_2 \, \ell_2 - 1 | j = \ell_1 + \ell_2 - 2 \, j; \ell_1 \ell_2 \rangle, \tag{4.5.519}
\]

\[
\langle \ell_1 \, \ell_1, \ell_2 \, \ell_2 - 2 | j = \ell_1 + \ell_2 - 2 \, j; \ell_1 \ell_2 \rangle. \tag{4.5.520}
\]

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

\[
|\ell_1, m_1 \rangle \otimes |\ell_2, m_2 \rangle \equiv |\ell_1, m_1, \ell_2, m_2 \rangle
\]

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 \( \ell_1, \ell_2 \), 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

\[
N = (2\ell_1 + 1)(2\ell_2 + 1) = \sum_{i=0}^{s} (2\ell_1 + 2\ell_2 - 2i + 1) = (2\ell_1 + 2\ell_2 + 1)(s + 1) - s(s + 1).
\]

This quadratic equation for \( s \) has two solutions

\[
s = 2\ell_1 \quad \Rightarrow \quad j = \ell_2 - \ell_1,
\]

or \( s = 2\ell_2 \) \( \Rightarrow \) \( j = \ell_1 - \ell_2 \).

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,

\[
\min j = |\ell_1 - \ell_2|.
\]

**Problem 4.83. From \(|j j - 1\rangle \) to \(|j - 1 j - 1\rangle\)** \( \text{YZ: Ignore this problem for now.} \)

Use equations (4.5.490) and (4.5.491) to show that

\[
\begin{align*}
[\vec{J}^2, J^\pm] &= \mp 2J^3 J^\pm \pm 2J^\pm J^3, \\
[\vec{J}^2, J^\mp] &= \mp 2J^3 J^\mp \pm 2J^\pm J^3.
\end{align*}
\]

Consider

\[
\begin{align*}
\vec{J}^2 J^\pm &= [\vec{J}^2, J^\pm] + J^\pm \vec{J}^2 \\
&= \mp 2J^3 J^\pm \pm 2J^\pm J^3 + J^\pm \vec{J}^2, \\
\vec{J}^2 J^\mp &= [\vec{J}^2, J^\mp] + J^\mp \vec{J}^2 \\
&= \mp 2J^3 J^\mp \pm 2J^\pm J^3 + J^\pm \vec{J}^2.
\end{align*}
\]

**Example: Tensor product of spin-1/2 systems** 
Consider the tensor product of two spin-1/2 systems. Eq. (4.5.482) informs us, the total angular momentum \( j \) runs from \(|1/2 - 1/2| = 0 \) to \( 1/2 + 1/2 = 1 \).

\[
j \in \{0, 1\}.
\]
To save notation-baggage, let us denote
\[ |++\rangle \equiv \left| \frac{1}{2} \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2} \frac{1}{2} \right\rangle, \tag{4.5.534} \]
\[ |+-\rangle \equiv \left| \frac{1}{2} \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2} \frac{1}{2} \right\rangle, \tag{4.5.535} \]
\[ |--\rangle \equiv \left| \frac{1}{2} \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2} \frac{1}{2} \right\rangle, \tag{4.5.536} \]
\[ |--\rangle \equiv \left| \frac{1}{2} \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2} \frac{1}{2} \right\rangle. \tag{4.5.537} \]

We will also suppress the \((1/2)s\) in the total \(j\) states; i.e.,
\[ |j m; \frac{1}{2} \frac{1}{2}\rangle \equiv |j m\rangle. \tag{4.5.538} \]

Let us start with \(j = 1\). The highest \(m\) state is
\[ |j = 1 m = 1\rangle = |++\rangle. \tag{4.5.539} \]

Applying the lowering operator gives
\[ \sqrt{2} |1 0\rangle = |--\rangle + |+-\rangle \tag{4.5.540} \]
\[ |1 0\rangle = \frac{|+-\rangle + |+-\rangle}{\sqrt{2}}. \tag{4.5.541} \]

Applying the lowering operator once more,
\[ \sqrt{1(1+1)} |1 -1\rangle \]
\[ = \frac{1}{\sqrt{2}} (J^- + J''^-) |--\rangle + \frac{1}{\sqrt{2}} (J^- + J''^-) |+-\rangle \tag{4.5.542} \]
\[ = \frac{1}{\sqrt{2}} J''^- |--\rangle + \frac{1}{\sqrt{2}} J^- |+-\rangle \tag{4.5.543} \]
\[ = \frac{1}{\sqrt{2}} \sqrt{\frac{1}{2} \left( \frac{1}{2} + 1 \right) - \frac{1}{2} \left( \frac{1}{2} - 1 \right)} |--\rangle + \frac{1}{\sqrt{2}} \sqrt{\frac{1}{2} \left( \frac{1}{2} + 1 \right) - \frac{1}{2} \left( \frac{1}{2} - 1 \right)} |--\rangle. \]
\[ = \sqrt{2} |--\rangle. \tag{4.5.544} \]

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.
\[ |1 1; \frac{1}{2} \frac{1}{2}\rangle = \left| \frac{1}{2} \frac{1}{2}; \frac{1}{2} \frac{1}{2} \right\rangle = |++\rangle, \tag{4.5.545} \]
\[ |1 0; \frac{1}{2} \frac{1}{2}\rangle = \frac{1}{\sqrt{2}} \left| \frac{1}{2} \frac{1}{2} - \frac{1}{2} \frac{1}{2} \right\rangle + \frac{1}{\sqrt{2}} \left| \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} \right\rangle = \frac{|+-\rangle + |+-\rangle}{\sqrt{2}}. \tag{4.5.546} \]
\[ |1 - 1; \frac{1}{2} \frac{1}{2}\rangle = |\frac{1}{2} - \frac{1}{2} \frac{1}{2} - \frac{1}{2}\rangle = |--\rangle. \]  \hspace{1cm} (4.5.547)

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

\[ (m_1 = \pm \frac{1}{2}, m_2 = \mp \frac{1}{2}). \]  \hspace{1cm} (4.5.548)

Hence,

\[ |0 \ 0\rangle = |--\rangle \langle ++| 0 \ 0\rangle + |+-\rangle \langle +--| 0 \ 0\rangle. \]  \hspace{1cm} (4.5.549)

This state must be perpendicular to \(|1 0\rangle\) in eq. (4.5.546), because they have distinct \(J^2\) eigenvalues (1(1 + 1) vs. 0). Taking their inner product,

\[ \langle -+| 0 \ 0\rangle + \langle +--| 0 \ 0\rangle = 0. \]  \hspace{1cm} (4.5.550)

At this point,

\[ |0 \ 0\rangle = (|--\rangle - |--\rangle) \langle --| 0 \ 0\rangle. \]  \hspace{1cm} (4.5.551)

Because the state has to be normalized to unity, we have now determined it up to a phase \(e^{i\delta_0}\):

\[ \begin{array}{c}
|0 \ 0; \frac{1}{2} \frac{1}{2}\rangle = \frac{e^{i\delta_0}}{\sqrt{2}} \left( |\frac{1}{2} - \frac{1}{2} \frac{1}{2} - \frac{1}{2}\rangle - |\frac{1}{2} \frac{1}{2} \frac{1}{2} - \frac{1}{2}\rangle \right) \\
= \frac{e^{i\delta_0}}{\sqrt{2}} (-|--\rangle - |+-\rangle). \end{array} \]  \hspace{1cm} (4.5.552)

Example: ‘Orbital’ angular momentum and spin-half

Let us now consider taking the tensor product

\[ |\ell, m\rangle \otimes |\frac{1}{2}; \pm \frac{1}{2}\rangle; \]  \hspace{1cm} (4.5.554)

for integer \(\ell = 0, 1, 2, \ldots\) 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,

\[ |\frac{1}{2}; \pm \frac{1}{2}\rangle = |0, 0\rangle \otimes |\frac{1}{2}; \pm \frac{1}{2}\rangle. \]  \hspace{1cm} (4.5.555)

\(\ell \geq 1\)

For non-zero \(\ell\), eq. (4.5.482) says we must have \(j\) running from \(\ell - 1/2\) to \(\ell + 1/2\):

\[ j = \ell \pm \frac{1}{2}. \]  \hspace{1cm} (4.5.556)

We start from the highest possible \(m\) value.

\[ |j = \ell + \frac{1}{2}; m = j; \ell \frac{1}{2}\rangle = |\ell, \ell\rangle \otimes |\frac{1}{2}; \frac{1}{2}\rangle. \]  \hspace{1cm} (4.5.557)
Applying the lowering operator \( s \) times, we have on the left hand side

\[
(J^-)^s \left| j = \ell + \frac{1}{2} m = j; \ell \frac{1}{2} \right> = A_s^{\ell+\frac{1}{2}} \left| j = \ell + \frac{1}{2} m = j - s \right>,
\]

(4.5.58)

where the constant \( A_s^{\ell+\frac{1}{2}} \) follows from repeated application of eq. (4.5.366)

\[
A_s^{\ell+\frac{1}{2}} = \prod_{i=0}^{s-1} \sqrt{(2\ell + 1 - i)(i + 1)}.
\]

(4.5.59)

Whereas on the right hand side, \((J^-)^s = (J'^- + J''^-)^s\) may be expanded using the binomial theorem since \([J'-, J''-] = 0\). Altogether,

\[
A_s^{\ell+\frac{1}{2}} \left| j = \ell + \frac{1}{2} m = j - s; \ell \frac{1}{2} \right> = \sum_{i=0}^{s} \binom{s}{i} (J^-)^{s-i} |\ell, \ell \rangle \otimes (J''^-)^i \left| \frac{1}{2}, \frac{1}{2} \right>. \]

(4.5.60)

But \((J''^-)^i \left| \frac{1}{2}, \frac{1}{2} \right> = 0\) 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.

\[
A_s^{\ell+\frac{1}{2}} \left| j = \ell + \frac{1}{2} m = j - s; \ell \frac{1}{2} \right> = (J^-)^s |\ell, \ell \rangle \otimes \left| \frac{1}{2}, \frac{1}{2} \right> \nonumber
\]

\[
= \frac{s!}{(s-1)!} \sqrt{\left( \frac{1}{2} + \frac{1}{2} \right) \left( \frac{1}{2} - \frac{1}{2} + 1 \right)} (J^-)^{s-1} |\ell, \ell \rangle \otimes \left| \frac{1}{2}, -\frac{1}{2} \right> \nonumber
\]

\[
= A_s^{\ell} |\ell, \ell - s \rangle \otimes \left| \frac{1}{2}, \frac{1}{2} \right> + s \cdot A_{s-1}^{\ell} |\ell, \ell - s + 1 \rangle \otimes \left| \frac{1}{2}, -\frac{1}{2} \right>.
\]

Here, the constants are

\[
A_s^{\ell} = \prod_{i=0}^{s-1} \sqrt{(2\ell - i)(i + 1)},
\]

(4.5.62)

\[
A_{s-1}^{\ell} = \prod_{i=0}^{s-2} \sqrt{(2\ell - i)(i + 1)}.
\]

(4.5.63)

Writing them out more explicitly,

\[
\sqrt{2\ell + 1} \sqrt{2\ell - 1} \sqrt{2\ell - 2} \sqrt{2\ell - 3} \ldots \sqrt{2\ell - (s - 2)} \sqrt{s} \left| j = \ell + \frac{1}{2} m = j - s; \ell \frac{1}{2} \right>,
\]

(4.5.64)

\[
= \sqrt{2\ell} \sqrt{2\ell - 1} \sqrt{2\ell - 2} \sqrt{2\ell - 3} \ldots \sqrt{2\ell - (s - 1)} \sqrt{s} |\ell, \ell - s \rangle \otimes \left| \frac{1}{2}, \frac{1}{2} \right> \nonumber
\]

\[
+ (\sqrt{3})^2 \sqrt{2\ell} \sqrt{2\ell - 1} \sqrt{2\ell - 2} \sqrt{2\ell - 3} \ldots \sqrt{2\ell - (s - 2)} \sqrt{s - 1} |\ell, \ell - s + 1 \rangle \otimes \left| \frac{1}{2}, -\frac{1}{2} \right>.
\]
The factors $\sqrt{2\ell} \ldots \sqrt{2\ell - (s - 2)}$ and $\sqrt{1} \ldots \sqrt{s}$ 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 $m$.

$$\left| j = \ell + \frac{1}{2}, m; \ell \frac{1}{2} \right\rangle = \frac{1}{\sqrt{2\sqrt{2\ell + 1}}} \left( \sqrt{2\ell + 2m + 1} \left| \ell, m - \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2}, \frac{1}{2} \right\rangle + \sqrt{2\ell - 2m + 1} \left| \ell, m + \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right)$$

(Remember $\ell \pm 1/2$ is half-integer, since $\ell$ is integer; so the azimuthal number $m \pm 1/2$ itself is an integer.) For the states $\left| j = \ell - (1/2), m \right\rangle$, we will again see that there are only two terms in the superposition over the tensor product states. For a fixed $m$, $\left| j = \ell - (1/2), m \right\rangle$ must be perpendicular to $\left| j = \ell + (1/2), m \right\rangle$. This allows us to write down its solution (up to an arbitrary phase) by inspecting eq. (4.5.565):

$$\left| j = \ell - \frac{1}{2}, m; \ell \frac{1}{2} \right\rangle = e^{i\delta_{\ell, \ell - 1/2}} \sqrt{2\sqrt{2\ell + 1}} \left( \sqrt{2\ell - 2m + 1} \left| \ell, m - \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2}, \frac{1}{2} \right\rangle - \sqrt{2\ell + 2m + 1} \left| \ell, m + \frac{1}{2} \right\rangle \otimes \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right)$$

**Problem 4.84.** Use the form of $\vec{J}^2$ in eq. (4.5.502) to confirm the right hand sides of equations (4.5.565) and (4.5.566) are indeed its eigenvectors, with respective eigenvalues of $(\ell + 1/2)(\ell + 1/2 + 1)$ and $(\ell - 1/2)(\ell - 1/2 + 1)$.

**Problem 4.85.** Express the states $\left| j m; \ell \frac{3}{2}, 1 \right\rangle$ in terms of the basis $\{|\ell \frac{3}{2}, m_1 \rangle \otimes |1, m_2 \rangle\}$. 

### Invariant Subspaces & Clebsch-Gordan Unitarity

Among the mutually compatible observables in eq. (4.5.477), we highlight

$$[\vec{J}^2, J^3] = 0. \quad (4.5.567)$$

Because the rotation operator involves the exponential of the generators $\vec{J}$, that means it also commutes with $\vec{J}^2$.

$$[\vec{J}^2, D(\hat{R})] = [\vec{J}^2, \exp(-i\vec{\theta} \cdot \vec{J})] = 0 \quad (4.5.568)$$

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 $(j, \ell_1, \ell_2)$ – an invariant subspace under rotations. For, we may utilize eq. (4.5.568) to compute

$$\vec{J}^2 \left(D(\hat{R}) \left| j m; \ell_1 \ell_2 \right\rangle\right) = D(\hat{R})\vec{J}^2 \left| j m; \ell_1 \ell_2 \right\rangle \quad (4.5.569)$$
\[ = j(j + 1)D(\hat{R}) \langle j m; \ell_1 \ell_2 \rangle. \quad (4.5.70) \]

In words: both \(|j m; \ell_1 \ell_2\rangle\) and \(D(\hat{R}) \langle j m; \ell_1 \ell_2 \rangle\) are eigenvectors of \(\hat{J}_z\), with the same eigenvalue \(j(j + 1)\). Therefore, under an arbitrary rotation \(D(\hat{R})\), the vector space spanned by \(|j m; \ell_1 \ell_2\rangle\) gets rotated into itself – the matrix element

\[ \left\langle j' m'; \ell_1 \ell_2 \right| D(\hat{R}) \left| j m; \ell_1 \ell_2 \right\rangle \propto \delta^{j'}_j \quad (4.5.71) \]

is zero unless \(j' = j\). In fact, for fixed \((\ell_1, \ell_2)\) the matrix of eq. \((4.5.71)\) 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,

\[ \{|j = \ell_1 + \ell_2 - j,m \leq m \leq j; \ell_1 \ell_2\rangle\} \quad (4.5.72) \]
\[ \{|j = \ell_1 + \ell_2 - 1 - j,m \leq m \leq j; \ell_1 \ell_2\rangle\} \quad (4.5.73) \]
\[ \ldots \]
\[ \{|j = |\ell_1 - \ell_2| + 1 - j,m \leq m \leq j; \ell_1 \ell_2\rangle\} \quad (4.5.75) \]
\[ \{|j = |\ell_1 - \ell_2| - j,m \leq m \leq j; \ell_1 \ell_2\rangle\} \quad (4.5.76) \]

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 \(|j m; \ell_1 \ell_2\rangle\) to \(|j m; \ell_1 \ell_2\rangle\) and vice versa. In particular, the inverse relation to eq. \((4.5.493)\) is

\[ |\ell_1 m_1, \ell_2 m_2\rangle = \sum_{j=|\ell_1 - \ell_2|}^{\ell_1 + \ell_2} \sum_{m=-j}^{+j} |j m; \ell_1 \ell_2\rangle \langle j m; \ell_1 \ell_2| \ell_1 m_1, \ell_2 m_2\rangle, \quad (4.5.77) \]

with the associated completeness relation

\[ \sum_{j=|\ell_1 - \ell_2|}^{\ell_1 + \ell_2} \sum_{m=-j}^{+j} |j m; \ell_1 \ell_2\rangle \langle j m; \ell_1 \ell_2| = \mathbb{I}. \quad (4.5.78) \]

The unitary character of these Clebsch-Gordan coefficients follow from the completeness relation in equations \((4.5.494)\) and \((4.5.578)\).

\[ \sum_{m_1=-\ell_1}^{+\ell_2} \sum_{m_2=-\ell_2}^{+\ell_2} \langle j' m'; \ell_1 \ell_2| \ell_1 m_1, \ell_2 m_2\rangle \langle \ell_1 m_1, \ell_2 m_2| j m; \ell_1 \ell_2\rangle = \delta^{j'}_j \delta^{m'}_m \quad (4.5.79) \]
\[ \sum_{j=|\ell_1 - \ell_2|}^{\ell_1 + \ell_2} \sum_{m=-j}^{+j} \langle \ell_1 m_1', \ell_2 m_2'| j m; \ell_1 \ell_2\rangle \langle j m; \ell_1 \ell_2| \ell_1 m_1, \ell_2 m_2\rangle = \delta^{m_1'}_{m_1} \delta^{m_2'}_{m_2}. \quad (4.5.80) \]
(Irreducible) Spherical Vector & Tensor Operators

We may generalize the definition of a vector operator in eq. (4.5.384) to a higher rank tensor $T_{i_1 i_2 \ldots i_N}$.

$$D(\hat{R})^\dagger T_{i_1 i_2 \ldots i_N} D(\hat{R}) = \hat{R}_{i_1}^{j_1} \hat{R}_{i_2}^{j_2} \ldots \hat{R}_{i_N}^{j_N} T_{j_1 j_2 \ldots j_N}$$  \quad (4.5.581)

But we may also remember eq. (4.5.368), where we found a different representation for the rotation operation, one based on the angular momentum eigenkets themselves. Specifically, because the rotation operator $D(\hat{R})$ leaves invariant the space spanned by $\{|\ell, m\rangle\}$ for a fixed $\ell$ and this is the smallest such space (i.e., $D(\hat{R})$ mixes all the $m$ values in general), this basis is said to provide an irreducible representation for the rotation operator in eq. (4.5.369). 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. (4.5.368); but one that would involve the angular spherical harmonics $Y_{\ell}^{m}(\hat{\mathbf{r}})$. We first define the spherical harmonic of the 3D position operator $X^i$ by

$$Y_{\ell}^{m}(\hat{X}) |\mathbf{x}\rangle \equiv r^\ell Y_{\ell}^{m}(\hat{r}) |\mathbf{x}\rangle \equiv r^\ell \hat{r} |\mathbf{x}\rangle ; \quad (4.5.582)$$

where $r\hat{r}$ is simply the Cartesian coordinates $\mathbf{x}$ expressed in spherical coordinates, with $r \equiv |\mathbf{x}|$ and $\hat{r} = \mathbf{x}/r$. In practice, this means, for instance:

$$Y_{0}^{0}(\hat{X}) = \frac{1}{\sqrt{4\pi}}$$  \quad (4.5.583)

and

$$\sqrt{\frac{3}{4\pi}} X^\pm \equiv Y_{1}^{\pm}(\hat{X}) = \mp \sqrt{\frac{3}{4\pi}} X^\pm \pm iX^\mp \frac{\sqrt{2}}{\sqrt{2}}$$,  \quad (4.5.584)

$$\sqrt{\frac{3}{4\pi}} X^0 \equiv Y_{1}^{0}(\hat{X}) = \sqrt{\frac{3}{4\pi}} X^0$$  \quad (4.5.585)

Now, on the one hand

$$D(\hat{R})^\dagger Y_{\ell}^{m}(\hat{X}) D(\hat{R}) |\mathbf{x}\rangle = D(\hat{R})^\dagger Y_{\ell}^{m}(\hat{X}) |\hat{R}\mathbf{x}\rangle = D(\hat{R})^\dagger Y_{\ell}^{m}(\hat{R}) |\hat{R}\mathbf{x}\rangle = r^\ell Y_{\ell}^{m}(\hat{R}\hat{r}) |\mathbf{x}\rangle ; \quad (4.5.586)$$

which holds for arbitrary $|\mathbf{x}\rangle$ and hence

$$D(\hat{R})^\dagger Y_{\ell}^{m}(\hat{X}) D(\hat{R}) = r^\ell Y_{\ell}^{m}(\hat{R}\hat{r})$$.

On the other hand,

$$D(\hat{R}^T) |\ell, m\rangle = D(\hat{R})^\dagger |\ell, m\rangle = \sum_{m'} |\ell, m\rangle \overline{D_{(\ell)}^m}_{m'}$$.

(4.5.588)

Upon acting $\langle \theta, \phi | \equiv \langle \hat{\mathbf{r}} |$ from the left on both sides, and recognizing $\langle \hat{\mathbf{r}} | D(\hat{R})^\dagger = \langle \hat{R}\hat{r} |$ and $\langle \hat{\mathbf{r}} | \ell, m\rangle = Y_{\ell}^{m}(\theta, \phi)$,

$$Y_{\ell}^{m}(\hat{R}\hat{r}) = \sum_{m'} \overline{D_{(\ell)}^m}_{m'}(\hat{R}) Y_{\ell}^{m'}(\hat{\mathbf{r}})$$

(4.5.589)
\[ Y^m_\ell(\hat{R}\hat{X}) |\vec{x}\rangle = \sum_{m'} D_{m'}^{m}(\hat{R}) Y^{m'}_\ell(\hat{X}) |\vec{x}\rangle. \] (4.5.590)

Comparing equations (4.5.587) and (4.5.590),

\[ D(\hat{R})^\dagger Y^m_\ell(\hat{X}) D(\hat{R}) = \sum_{m'=-\ell}^{+\ell} Y^{m'}_\ell(\hat{X}) \bar{D}_{\ell}^{m'}(\hat{R}). \] (4.5.591)

Swapping \( \hat{R} \leftrightarrow \hat{R}^T \) and recalling \( D(\hat{R}^T) = D(\hat{R})^\dagger \), we see that

\[ D(\hat{R}) Y^m_\ell(\hat{X}) D(\hat{R})^\dagger = Y^m_\ell(\hat{R}^T \hat{X}) = \sum_{m'=-\ell}^{+\ell} Y^{m'}_\ell(\hat{X}) \langle \ell, m' | D(\hat{R}) | \ell, m \rangle. \] (4.5.592)

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

\[ D(\hat{R}) |\ell, m\rangle = \sum_{m'} |\ell, m'\rangle \langle \ell, m' | D(\hat{R}) | \ell, m \rangle. \] (4.5.593)

The relation in eq. (4.5.592) really does not depend on \( \hat{X} \) being the position operator; rather, it is really due to \( \hat{X} \) being a vector operator. That is, eq. (4.5.592) would still hold if we replaced \( \hat{X} \) with any vector operator \( V^i \) obeying

\[ D(\hat{R})^\dagger V^i D(\hat{R}) = R^i_j V^j. \]

Spherical Tensor: Definition

A spherical tensor \( O^m_j \) of rank \( j \) with \( 2j+1 \) components is defined as a linear operator obeying

\[ D(\hat{R}) O^m_j D(\hat{R})^\dagger = \sum_{m'=-j}^{+j} O^{m'}_j \bar{D}^{m'}_j(\hat{R}). \] (4.5.595)

The equivalent infinitesimal version is provided by the equations

\[ \left[ J^i, O^m_j \right] = \sum_{m'=-j}^{+j} O^{m'}_j \langle j, m' | J^i | j, m \rangle; \] (4.5.596)

where \( J^i = (J^1, J^2, J^3) \) refers to the Cartesian components of the rotation generators.
**Problem 4.86.** Derive eq. \((4.5.596)\) from eq. \((4.5.595)\). Then explain why
\[
\left[ J^3, O^m_j \right] = mO^m_j, \quad (4.5.597)
\]
\[
\left[ J^\pm, O^m_j \right] = \sqrt{(j \mp m)(j \pm m + 1)}O^{m \pm 1}_j; \quad (4.5.598)
\]
where \(J^\pm \equiv J^1 \pm iJ^2\) are the raising/lowering angular momentum operators.

**Example** We may immediately generalize the results in equations \((4.5.584)\) and \((4.5.585)\) to an arbitrary vector operator \(V^i\). We define
\[
V^{1\pm} \equiv \pm \frac{V^1 \pm iV^2}{\sqrt{2}}, \quad (4.5.599)
\]
\[
V^0 \equiv V^3. \quad (4.5.600)
\]
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
\[
\vec{V} = \frac{V^{1-} - V^{1+}}{\sqrt{2}} \hat{e}_1 + \frac{i}{\sqrt{2}} (V^{-1} + V^{1+}) \hat{e}_2 + V^0 \hat{e}_3, \quad (4.5.601)
\]
where \(\hat{e}_i\) is the unit vector along the \(i\)th axis.

In particular, the angular momentum operators themselves can be expressed as
\[
J^{1\pm} \equiv \pm \frac{J^\pm}{\sqrt{2}} \quad \text{and} \quad J^0 = J^3. \quad (4.5.602)
\]

**Problem 4.87. 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
\[
Q^m_j = \sum_{|j_1 - j_2| \leq j_1 + j_2} \sum_{m=m_1+m_2} A_{j_1}^{m_1} B_{j_2}^{m_2} \langle j_1 m_1, j_2 m_2 | j m; j_1 j_2 \rangle \quad (4.5.603)
\]
produces a spherical tensor \(Q^m_j\). 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 4.88. Irreducible decomposition of Vector \(\otimes\) Vector** Via a direct calculation, show that the following trace, antisymmetric, and symmetric-trace-free decomposition of a product of two vectors, namely
\[
V^i W^i = \frac{1}{3} \delta^{ij} \vec{V} \cdot \vec{W} + \frac{V^i W^j - V^j W^i}{2} + \left( \frac{V^i W^j + V^j W^i}{2} - \frac{1}{3} \delta^{ij} \vec{V} \cdot \vec{W} \right); \quad (4.5.604)
\]
admits the following irreducible decomposition. The trace portion is
\[
\vec{V} \cdot \vec{W} = V^0 W^0 - V^{1+} W_{1-}^{-1} - V^{-1} W_{1+}^{+1}; \quad (4.5.605)
\]

136
the antisymmetric sector is

\[ \frac{V^i W^j - V^j W^i}{2} = \frac{i}{2} \left( V^{-1}_1 W^{+1}_1 - V^{+1}_1 W^{1}_1 \right) (\hat{e}_1^i \hat{e}_2^j - \hat{e}_1^j \hat{e}_2^i) \tag{4.5.066} \]

\[ + \frac{1}{2} \frac{V^{-1}_1 W^0_1 - V^0 W^{1}_1}{\sqrt{2}} \left\{ (\hat{e}_1^i \hat{e}_3^j - \hat{e}_1^j \hat{e}_3^i) + i (\hat{e}_2^i \hat{e}_3^j - \hat{e}_2^j \hat{e}_3^i) \right\} \]

\[ - \frac{1}{2} \frac{V^{+1}_1 W^0_1 - V^0 W^{-1}_1}{\sqrt{2}} \left\{ (\hat{e}_1^i \hat{e}_3^j - \hat{e}_1^j \hat{e}_3^i) - i (\hat{e}_2^i \hat{e}_3^j - \hat{e}_2^j \hat{e}_3^i) \right\} \]

\[ = \frac{1}{2} \left( V^{-1}_1 W^{+1}_1 - V^{+1}_1 W^{1}_1 \right) (\hat{e}_+^i \hat{e}_-^j - \hat{e}_-^i \hat{e}_+^j) \]

\[ + \frac{1}{2} \left( V^{-1}_1 W^0_1 - V^0 W^{1}_1 \right) (\hat{e}_+^i \hat{e}_3^j - \hat{e}_+^j \hat{e}_3^i) \]

\[ - \frac{1}{2} \left( V^{+1}_1 W^0_1 - V^0 W^{-1}_1 \right) (\hat{e}_-^i \hat{e}_3^j - \hat{e}_-^j \hat{e}_3^i) \]

and the symmetric and trace-less part is

\[\frac{V^i W^j + V^j W^i}{2} - \frac{1}{3} \delta^{ij} \vec{V} \cdot \vec{W} \]

\[= \frac{V^{-1}_1 W^{1}_1}{2} \left\{ (\hat{e}_1^i \hat{e}_2^j - \hat{e}_1^j \hat{e}_2^i) + i (\hat{e}_1^i \hat{e}_2^j + \hat{e}_2^i \hat{e}_1^j) \right\} \tag{4.5.067} \]

\[+ \frac{V^{+1}_1 W^{+1}_1}{2} \left\{ (\hat{e}_1^i \hat{e}_2^j - \hat{e}_2^i \hat{e}_1^j) - i (\hat{e}_1^i \hat{e}_2^j + \hat{e}_2^i \hat{e}_1^j) \right\} \]

\[+ \frac{1}{2} \frac{V^{-1}_1 W^0_1 + V^0 W^{1}_1}{\sqrt{2}} \left\{ (\hat{e}_1^i \hat{e}_3^j + \hat{e}_1^j \hat{e}_3^i) + i (\hat{e}_2^i \hat{e}_3^j + \hat{e}_2^j \hat{e}_3^i) \right\} \]

\[+ \frac{1}{2} \frac{V^{+1}_1 W^0_1 + V^0 W^{-1}_1}{\sqrt{2}} \left\{ (\hat{e}_1^i \hat{e}_3^j + \hat{e}_1^j \hat{e}_3^i) - i (\hat{e}_2^i \hat{e}_3^j + \hat{e}_2^j \hat{e}_3^i) \right\} \]

\[= \frac{1}{6} \left( 2V^0_1 W^0_1 + V^{+1}_1 W^{-1}_1 + V^{-1}_1 W^{+1}_1 \right) (\hat{e}_1^i \hat{e}_2^j + \hat{e}_1^j \hat{e}_2^i - 2\hat{e}_2^i \hat{e}_3^j) \]

\[= V^{+1}_1 W^{+1}_1 \hat{e}_+^i \hat{e}_+^j + V^{-1}_1 W^{-1}_1 \hat{e}_-^i \hat{e}_-^j \]

\[= \frac{V^{-1}_1 W^0_1 + V^0 W^{1}_1}{2} \left( \hat{e}_+^i \hat{e}_3^j + \hat{e}_+^j \hat{e}_3^i \right) - \frac{V^{+1}_1 W^0_1 + V^0 W^{-1}_1}{2} \left( \hat{e}_-^i \hat{e}_3^j + \hat{e}_-^j \hat{e}_3^i \right) \]

\[= \frac{1}{6} \left( 2V^0_1 W^0_1 + V^{+1}_1 W^{-1}_1 + V^{-1}_1 W^{+1}_1 \right) \left( \hat{e}_+^i \hat{e}_-^j + \hat{e}_+^j \hat{e}_-^i - 2\hat{e}_3^i \hat{e}_3^j \right) . \]

We have defined

\[\hat{e}_\pm \equiv \hat{e}_1 \pm i\hat{e}_2. \tag{4.5.068} \]

Identity all the distinct (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 \{\hat{e}_i\} are orthonormal vectors, \(\delta^{ab} \hat{e}_a \hat{e}_b = \delta^{ij}\). (Can you explain why?)
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
\[
J^i |j, m; \Psi\rangle = \sum_{m'} |j, m'; \Psi\rangle \langle j, m' | J^i | j, m\rangle,
\]
\[
J^i |j, m; \Phi\rangle = \sum_{m'} |j, m'; \Phi\rangle \langle j, m' | J^i | j, m\rangle;
\]
then the matrix element
\[
\langle j, m; \Psi | Q | j, m; \Phi\rangle
\]
is in fact independent of \( m \), as long as the operator \( Q \) commutes with \( \{J^i\} \).

We will use the raising/lowering operators. Consider
\[
\langle j, m \pm 1; \Psi | Q J^\pm | j, m; \Phi\rangle = \sqrt{(j \mp m)(j \pm m + 1)} \langle j, m \pm 1; \Psi | Q | j, m \pm 1; \Phi\rangle. \tag{4.5.612}
\]
By assumption, the \( J^\pm \) may also be moved to the left of \( Q \),
\[
\langle j, m \pm 1; \Psi | Q J^\pm | j, m; \Phi\rangle = \langle j, m \pm 1; \Psi | (J^\pm)^\dagger Q | j, m; \Phi\rangle = \sqrt{(j \mp (m \pm 1))(j \mp (m \pm 1) + 1)} \langle j, m; \Psi | Q | j, m; \Phi\rangle. \tag{4.5.613}
\]
Comparing the two results tells us
\[
\langle j, m \pm 1; \Psi | Q | j, m \pm 1; \Phi\rangle = \langle j, m; \Psi | Q | j, m; \Phi\rangle. \tag{4.5.614}
\]
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 \( m, m', m'' \).
\[
\langle j'' m''; \Phi | O_j^m | j'm'; \Psi \rangle = \langle j''; \Phi | O_j | j'; \Psi \rangle \langle j'' m''; j j' | j m, j' m' \rangle. \tag{4.5.615}
\]
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 \( \Psi \)).
\[
\vec{J}^2 |j'm'; \Psi\rangle = j' (j' + 1) |j'm'; \Psi\rangle, \tag{4.5.616}
\]
138
\[ J^3 |j'm'; \Psi \rangle = m' |j'm'; \Psi \rangle. \quad (4.5.617) \]

We see that
\[ D(\hat{R}) (O_j^n |j'm'; \Psi \rangle) = D(\hat{R}) O_j^n D(\hat{R})^\dagger D(\hat{R}) |j'm'; \Psi \rangle \]
\[ = \sum_{n,n'} O_j^n |j'n'; \Psi \rangle \langle j, n | D(\hat{R}) |j', m\rangle \langle j', n' | D(\hat{R}) |j', m'\rangle. \quad (4.5.618) \]

In other words, this \( O_j^n |j'm'; \Psi \rangle \) transforms in the same manner under rotations as the tensor product state \(|j m\rangle \otimes |j'm'\rangle\).

\[ O_j^n |j'm'; \Psi \rangle \leftrightarrow |j m\rangle \otimes |j'm'\rangle \quad (4.5.619) \]

Hence it must be possible to use the Clebsch-Gordan coefficients to construct the analog of the ‘total angular momentum’ state
\[ |j''m''; j j'; O, \Psi \rangle \equiv \sum_{m+m'=m''} O_j^n |j' m'; \Psi \rangle \langle j, m, j', m' | j'' m''; j j' \rangle. \quad (4.5.620) \]

By construction, this state obeys
\[ \hat{J}^2 |j''m''; j j'; O, \Psi \rangle = j'' (j'' + 1) |j''m''; j j'; O, \Psi \rangle, \quad (4.5.621) \]
\[ J^3 |j''m''; j j'; O, \Psi \rangle = m'' |j''m''; j j'; O, \Psi \rangle. \quad (4.5.622) \]

This also implies we should be able to invert this relation and solve for
\[ O_j^n |j'm'; \Psi \rangle = \sum_{j'' \in \{ |j-j'|, |j-j'|+1, ..., j+j'\} \atop m'' = m+m'} |j''m''; j j'; O, \Psi \rangle \langle j''m''; j j' | j, m, j' m' \rangle. \quad (4.5.623) \]

If \(|j''m''; \Phi \rangle\) is another eigenstate of angular momentum (which may depend on other variables, collectively denoted as \( \Phi \); then we may project both sides of eq. (4.5.623) with it.

\[ \langle j'' m''; j j' | O_j^n |j'm'; \Psi \rangle = \langle j'' m''; j j' | \Phi | j'' m''; j j' | O, \Psi \rangle \langle j'' m''; j j' | j, m, j' m' \rangle \quad (4.5.624) \]

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:

\[ \langle \ell, m; \alpha | V_1^n | \ell', m'; \beta \rangle = \langle \ell, m; 1 \ell' | 1 n, \ell' m' \rangle \langle \ell, \alpha | V_1 | \ell'; \beta \rangle \]
\[ n \in \{ \pm 1, 0 \}. \quad (4.5.625) \]

Likewise for \( W^i \),

\[ \langle \ell, m; \alpha | W_1^n | \ell', m'; \beta \rangle = \langle \ell, m; 1 \ell' | 1 n, \ell' m' \rangle \langle \ell, \alpha | W_1 | \ell'; \beta \rangle \]
\[ n \in \{ \pm 1, 0 \}. \quad (4.5.627) \]
Since the Clebsch-Gordan coefficients are common between the two, this means the ratio of the matrix elements in equations (4.5.625) and (4.5.627) only depends on the $m$-independent matrix elements.

\[
\frac{\langle \ell, m; \alpha | V^n_1 | \ell', m'; \beta \rangle}{\langle \ell, m; \alpha | W^n_1 | \ell', m'; \beta \rangle} = \frac{\langle \ell; \alpha | V^n_1 | \ell'; \beta \rangle}{\langle \ell; \alpha | W^n_1 | \ell'; \beta \rangle}.
\] (4.5.629)

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.

\[
\frac{\langle \ell, m; \alpha | V^i | \ell', m'; \beta \rangle}{\langle \ell, m; \alpha | W^i | \ell', m'; \beta \rangle} = \frac{\langle \ell; \alpha | V^1 | \ell'; \beta \rangle}{\langle \ell; \alpha | W^1 | \ell'; \beta \rangle}.
\] (4.5.630)

**Selection rules**

We see that such matrix elements in equations (4.5.625) and (4.5.627) are non-zero only when the following selection rules are satisfied, as dictated by the Clebsch-Gordan coefficient $\langle \ell m; 1 | n, \ell' m' \rangle$.

\[
\ell \in \{|\ell' - 1|, \ldots, \ell' + 1\}, \quad m' + n = m
\] (4.5.631)

In other words, $\ell$ cannot differ from $\ell'$ by more than one; and similarly for $m$ and $m'$ since $n = \pm 1, 0$.

\[
|\ell - \ell'| \leq 1, \quad |m - m'| \leq 1
\] (4.5.632)

When $\ell' = 0$ then $\ell > 0$; i.e., $\ell = 0 = \ell'$ is forbidden. Therefore, for integer $\ell$ and $\ell'$, note that

\[
\ell + \ell' \geq 1,
\] (4.5.633)

Electric dipole transitions in quantum mechanics are described by replacing $\vec{V}$ with the position operator $\vec{X}$ in eq. (4.5.625).

**Projection theorem**

We now turn to using the lemma enveloping eq. (4.5.611) and the Wigner-Eckart theorem itself to deduce, the matrix element within a spin $j$ subspace of the rank 1 spherical tensor $V^n_1$ is related to its angular momentum counterpart $J^n_1$ via the relation

\[
\langle j, m; \Psi | V^n_1 | j, m'; \Phi \rangle = \frac{\langle j, m; \Psi | \vec{J} \cdot \vec{V} | j, m; \Phi \rangle}{j(j + 1)}
\] (4.5.634)

\[
= \frac{\langle j, m'; \Phi | \vec{J} \cdot \vec{V} | j, m; \Psi \rangle}{j(j + 1)}
\] (4.5.635)

\[
= \frac{j(j + 1)}{\langle j, m | J^n_1 | j, m' \rangle}
\] (4.5.636)

where $\vec{J} \cdot \vec{V}$ is the Cartesian dot product; and the $\langle j, m | J^n_1 | j, m' \rangle$ is simply the matrix element of $J^n_1$ in the angular momentum eigenstate basis, and does not involve $\Psi$ nor $\Phi$.

By the Wigner-Eckart theorem, we know that

\[
\langle j, m; \Psi | V^n_1 | j, m'; \Phi \rangle = \langle j m; 1 | j 1 n; j m \rangle \langle j | V^n_1 | j \rangle \langle j | \Phi \rangle,
\] (4.5.637)

\[
\langle j, m | J^n_1 | j, m' \rangle = \langle j m; 1 | j 1 n; j m \rangle \langle j | J^n_1 | j \rangle.
\] (4.5.638)
Therefore, we may establish eq. (4.5.634) once we can show

\[
\frac{\langle j, m; \Psi \mid \vec{J} \cdot \vec{V} \mid j, m; \Phi \rangle}{j(j+1)} = \frac{\langle j, m; \Psi \mid \vec{J} \cdot \vec{V} \mid j, m; \Phi \rangle}{\langle j \mid J \mid j \rangle} = \frac{\langle j; \Psi \parallel V_1 \parallel j; \Phi \rangle}{\langle j \parallel J_1 \parallel j \rangle},
\]

(4.5.637)

since eq. (4.5.634) will read

\[
\langle j \parallel j \parallel j \rangle < n \parallel j \parallel j m \rangle \langle j; \Psi \parallel V_1 \parallel j; \Phi \rangle = \frac{\langle j; \Psi \parallel V_1 \parallel j; \Phi \rangle}{\langle j \parallel J \parallel j \rangle} \langle j \parallel j \parallel j \rangle < n \parallel j \parallel j \parallel j m \rangle \langle j \parallel J_1 \parallel j \rangle.
\]

(4.5.638)

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 \( \vec{J}^i \) itself. By the lemma surrounding eq. (4.5.611), we see that both the numerator and denominator after the first equality of eq. (4.5.637) are \( m \) independent. In particular, we may exploit the decomposition in eq. (4.5.605) (and eq. (4.5.601)),

\[
\langle j, m; \Psi \parallel J \cdot V \parallel j, m; \Phi \rangle = \langle j, m; \Psi \parallel J_1^0 V_1^0 + 2^{-1/2} J^+ V_1^{-1} - 2^{-1/2} J^- V_1^{+1} \parallel j, m; \Phi \rangle
\]

(4.5.639)

\[
= m \langle j, m; \Psi \parallel V_1^0 \parallel j, m; \Phi \rangle + \sqrt{(j + m)(j - m + 1)/2} \langle j, m - 1; \Psi \parallel V_1^{-1} \parallel j, m; \Phi \rangle - \sqrt{(j - m)(j + m + 1)/2} \langle j, m + 1; \Psi \parallel V_1^{+1} \parallel j, m; \Phi \rangle
\]

(4.5.640)

\[
= (m \langle j \parallel j \parallel 1, 0; j, m \rangle + \sqrt{(j + m)(j - m + 1)/2} \langle j - 1, 1 \parallel j \parallel 1, -1; j, m \rangle - \sqrt{(j - m)(j + m + 1)/2} \langle j + 1, 1 \parallel j \parallel 1, +1, j, m \rangle) \langle j; \Psi \parallel V_1 \parallel j; \Phi \rangle
\]

(4.5.641)

\[
\equiv \chi_j \langle j; \Psi \parallel V_1 \parallel j; \Phi \rangle.
\]

(4.5.642)

(This \( \chi_j \) actually does not depend on \( m \) – can you explain why?) By replacing \( V \rightarrow \vec{J} \) we may immediately write down

\[
j(j+1) = \langle j, m \parallel \vec{J}^2 \parallel j, m \rangle
\]

(4.5.643)

\[
= (m \langle j \parallel j \parallel 1, 0; j, m \rangle + \sqrt{(j + m)(j - m + 1)/2} \langle j - 1, 1 \parallel j \parallel 1, -1, j, m \rangle - \sqrt{(j - m)(j + m + 1)/2} \langle j + 1, 1 \parallel j \parallel 1, +1, j, m \rangle) \langle j \parallel J_1 \parallel j \rangle
\]

(4.5.644)

\[
= \chi_j \langle j \parallel J_1 \parallel j \rangle.
\]

(4.5.645)

Dividing equations (4.5.642) by (4.5.645) lead us to eq. (4.5.637). This proves the projection theorem.

### 4.6 Special Topic: Clebsch-Gordan Coefficients

In this section, we compile for the reader’s reference, Clebsch-Gordan coefficients

\[
\{ \langle \ell_1 m_1, \ell_2 m_2 \parallel j m \parallel \ell_1 \ell_2 \rangle \}
\]

(4.6.1)
for adding the spins $\ell_1$ and $\ell_2$, 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

$$|\ell_1, m_1\rangle \otimes |\ell_2, m_2\rangle \equiv |\ell_1 m_1, \ell_2 m_2\rangle.$$  \hfill (4.6.2)

for the tensor product states; and

$$|j m; \ell_1 \ell_2\rangle$$  \hfill (4.6.3)

for the total angular momentum states arising from adding spin $\ell_1$ and $\ell_2$.

Remember the constraint

$$m_1 + m_2 = m$$  \hfill (4.6.4)

and admissible range of $j$:

$$j \in \{|\ell_1 - \ell_2|, |\ell_1 - \ell_2| + 1, |\ell_1 - \ell_2| + 2, \ldots, \ell_1 + \ell_2 - 2, \ell_1 + \ell_2 - 1, \ell_1 + \ell_2\}.$$  \hfill (4.6.5)

Adding $|\frac{1}{2}, m_1\rangle \otimes |\frac{1}{2}, m_2\rangle$ to yield spin 0

$$\begin{align*}
\langle \frac{1}{2}  0  | \frac{1}{2}  0 \rangle &= \frac{1}{\sqrt{2}}, \\
\langle \frac{1}{2}  1  | \frac{1}{2}  1 \rangle &= -\frac{1}{\sqrt{2}}.
\end{align*}$$  \hfill (4.6.6)

Adding $|\frac{1}{2}, m_1\rangle \otimes |\frac{1}{2}, m_2\rangle$ to yield spin 1

$$\begin{align*}
\langle \frac{1}{2}  0  | \frac{1}{2}  1 \rangle &= 1, \\
\langle \frac{1}{2}  1  | \frac{1}{2}  0 \rangle &= \frac{1}{\sqrt{2}}, \\
\langle \frac{1}{2}  1  | \frac{1}{2}  1 \rangle &= \frac{1}{\sqrt{2}}, \\
\langle \frac{1}{2}  0  | \frac{1}{2}  1 \rangle &= 1.
\end{align*}$$  \hfill (4.6.8)

Adding $|\frac{1}{2}, m_1\rangle \otimes |1, m_2\rangle$ or $|1, m_1\rangle \otimes |\frac{1}{2}, m_2\rangle$ to yield spin 1/2

$$\begin{align*}
\langle \frac{1}{2}  0  | \frac{1}{2}  1 \rangle &= \frac{\sqrt{2}}{3}, \\
\langle \frac{1}{2}  0  | \frac{1}{2}  1 \rangle &= \frac{\sqrt{1}}{3}, \\
\langle \frac{1}{2}  0  | \frac{1}{2}  1 \rangle &= -\frac{\sqrt{1}}{3}.
\end{align*}$$  \hfill (4.6.12)

142
\[
\langle \frac{1}{2}, \frac{1}{2}, 1 | \frac{1}{2}, \frac{1}{2}, 1, 1 \rangle = \sqrt{\frac{2}{3}}; \tag{4.6.15}
\]

and
\[
\langle 1, 1, \frac{1}{2}, \frac{1}{2} | 1, \frac{1}{2}, 1, 1 \rangle = \sqrt{\frac{2}{3}}; \tag{4.6.16}
\]
\[
\langle 1, 0, \frac{1}{2}, \frac{1}{2} | 1, \frac{1}{2}, 1, 1 \rangle = -\sqrt{\frac{1}{3}}; \tag{4.6.17}
\]
\[
\langle 1, 0, \frac{1}{2}, \frac{1}{2} | 1, \frac{1}{2}, 1, 1 \rangle = \sqrt{\frac{1}{3}}; \tag{4.6.18}
\]
\[
\langle 1, 1, \frac{1}{2}, \frac{1}{2} | 1, \frac{1}{2}, 1, 1 \rangle = -\sqrt{\frac{2}{3}}. \tag{4.6.19}
\]

Adding \(|\frac{1}{2}, m_1 \rangle \otimes |1, m_2 \rangle\) or \(|1, m_1 \rangle \otimes |\frac{1}{2}, m_2 \rangle\) to yield spin 3/2

### 4.7 Special Topic: Approximation Methods for Eigensystems

#### 4.7.1 Rayleigh-Schrödinger Perturbation Theory

Suppose we know how to diagonalize some Hermitian operator \(H_0\) exactly.

\[
H_0 | \bar{E} \rangle = \bar{E} | \bar{E} \rangle \tag{4.7.1}
\]

In this section\[35\] we will address how to diagonalize a \(H\), namely

\[
H | E \rangle = E | E \rangle ; \tag{4.7.2}
\]

in the situation where it is a small perturbation of the \(H_0\), in the following sense:

\[
H \equiv H_0 + \epsilon H_1 + \epsilon^2 H_2 + \mathcal{O} (\epsilon^3) \tag{4.7.3}
\]
\[
= H_0 + \sum_{\ell=1}^{+\infty} \epsilon^\ell H_\ell. \tag{4.7.4}
\]

Here, \(0 < \epsilon \ll 1\) is oftentimes fictitious parameter indicating the ‘smallness’ of the \(H_\ell\)s; so the \(\epsilon\) in \(\epsilon H_1\) reminds us \(H_1\) is to be considered first order in perturbation, \(\epsilon^2 H_2\) second order, etc. and the \(\{ \delta_\ell H | \ell = 1, 2, 3, \ldots \}\) 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 \(\epsilon\) of an operator may often be identified with ratios of important dimensionful quantities of the setup at hand. Moreover, we have implicitly assumed a single parameter \(\epsilon\) here; while in

\[35\]The discussion here is inspired by §4.11 of Byron and Fuller [14], Sakurai and Napolitano [12], and Weinberg [13] Chapter 5.
The goal is therefore to compute the perturbations of the eigenstate along the unperturbed ones itself, a power series

\[
\langle E_b | E_a \rangle = \delta_a^b, \quad (4.7.5)
\]

\[
\sum_a |E_a\rangle \langle E_a| = \mathbb{I}. \quad (4.7.6)
\]

Then, we assert

\[
|E_a\rangle = |\bar{E}_a\rangle + \sum_{\ell = 1}^{+\infty} \epsilon^\ell |\ell E_a\rangle \quad (4.7.7)
\]

\[
= |\bar{E}_a\rangle + \sum_{\ell = 1}^{+\infty} \sum_s \epsilon^\ell |\bar{E}_s\rangle \langle \bar{E}_s| \ell E_a\rangle. \quad (4.7.8)
\]

where \(|\ell E_a\rangle\) is the \(\ell\)th correction to the \(a\)th energy eigenstate. The \(a\)th energy eigenvalue is, itself, a power series

\[
E_a = \bar{E}_a + \sum_{\ell = 1}^{+\infty} \epsilon^\ell \delta_\ell E_a. \quad (4.7.9)
\]

The goal is therefore to compute the perturbations of the eigenstate along the unperturbed ones \(\{ \langle \bar{E}_s| \ell E_a\rangle \}\) and of the eigenvalues \(\{\delta_\ell E_a\}\) in terms of the unperturbed ones \(\{\bar{E}_a\}\).

Now, the eigenvalue problem is given by \(H |E_a\rangle = E_a |E_a\rangle\). Expanding in powers of \(\epsilon\),

\[
(H_0 + \epsilon H_1 + \epsilon^2 H_2 + \ldots) (|\bar{E}_a\rangle + \epsilon |1 E_a\rangle + \epsilon^2 |2 E_a\rangle + \ldots)
\]

\[
= (\bar{E}_a + \epsilon \delta_1 E_a + \epsilon^2 \delta_2 E_a + \ldots) (|\bar{E}_a\rangle + \epsilon |1 E_a\rangle + \epsilon^2 |2 E_a\rangle + \ldots); \quad (4.7.10)
\]

we may collect powers of \(\epsilon\) in the following manner:

\[
H_0 |\bar{E}_a\rangle + \epsilon H_0 |1 E_a\rangle + \epsilon^2 H_0 |2 E_a\rangle + \ldots \quad (4.7.11)
\]

\[
+ \epsilon H_1 |\bar{E}_a\rangle + \epsilon^2 H_1 |1 E_a\rangle + \epsilon^2 H_1 |2 E_a\rangle + \ldots \quad (4.7.12)
\]

\[
+ \epsilon^2 H_2 |\bar{E}_a\rangle + \ldots \quad (4.7.13)
\]

\[
= \bar{E}_a |\bar{E}_a\rangle + \epsilon \bar{E}_a |1 E_a\rangle + \epsilon^2 \bar{E}_a |2 E_a\rangle + \ldots \quad (4.7.14)
\]

\[36\] The completeness relation of eq. \([4.7.6]\) 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 \(\{|\bar{E}_b\rangle\}\) are allowed to be degenerate. Strictly speaking, in such a situation we ought to introduce a degeneracy label, for e.g., \(|\bar{E}_b; i\rangle\), but prefer not to do so to avoid notation overload.
\[ + \epsilon \delta_1 E_a |\tilde{E}_a\rangle + \epsilon^2 \delta_1 E_a |1 E_a\rangle + \ldots \quad (4.7.15) \]
\[ + \epsilon^2 \delta_2 E_a |\tilde{E}_a\rangle + \ldots \quad (4.7.16) \]

The \(O(\epsilon^0)\) terms on both sides cancel out because they simply amount to eq. (4.7.1). The \(O(\epsilon, \epsilon^2)\) terms are

\[
(H_0 - \tilde{E}_a) |1 E_a\rangle = - (H_1 - \delta_1 E_a) |\tilde{E}_a\rangle, \quad (4.7.17)
\]
\[
(H_0 - \tilde{E}_a) |2 E_a\rangle = - (H_1 - \delta_1 E_a) |1 E_a\rangle - (H_2 - \delta_2 E_a) |\tilde{E}_a\rangle. \quad (4.7.18)
\]

More generally, at the \(O(\epsilon^{\ell+1})\) level,

\[
(H_0 - \tilde{E}_a) |\ell E_a\rangle
= - (H_1 - \delta_1 E_a) |\ell-1 E_a\rangle - (H_2 - \delta_2 E_a) |\ell-2 E_a\rangle
\]
\[\ldots - (H_{\ell-2} - \delta_{\ell-2} E_a) |2 E_a\rangle - (H_{\ell-1} - \delta_{\ell-1} E_a) |1 E_a\rangle - (H_\ell - \delta_\ell E_a) |\tilde{E}_a\rangle \quad (4.7.19)\]

Due to the hermitian character of \(H_0\), eq. (4.7.1) may be expressed as

\[
\langle \tilde{E}_a | (H_0 - \tilde{E}_a) = 0. \quad (4.7.20)\]

Therefore, \(\langle \tilde{E}_a \rangle\) acting on both sides of equations (4.7.17), (4.7.18), (4.7.19), etc. would yield zero on their left hand sides and in turn lead to

\[
0 = - \langle \tilde{E}_a | H_1 |\tilde{E}_a\rangle + \delta_1 E_a, \quad (4.7.21)\]
\[
0 = - \langle \tilde{E}_a | H_1 - \delta_1 E_a |1 E_a\rangle - \langle \tilde{E}_a | H_2 |\tilde{E}_a\rangle + \delta_2 E_a, \quad (4.7.22)\]
\[\ldots\]
\[
0 = - \langle \tilde{E}_a | H_1 - \delta_1 E_a |\ell-1 E_a\rangle - \langle \tilde{E}_a | H_2 - \delta_2 E_a |\ell-2 E_a\rangle \quad (4.7.23)\]
\[\ldots - \langle \tilde{E}_a | H_{\ell-2} - \delta_{\ell-2} E_a |2 E_a\rangle - \langle \tilde{E}_a | H_{\ell-1} - \delta_{\ell-1} E_a |1 E_a\rangle - \langle \tilde{E}_a | H_\ell |\tilde{E}_a\rangle + \delta_\ell E_a.\]

At this juncture, let us observe it is always possible to render

\[
\langle \tilde{E}_a | \ell E_a\rangle = 0 \quad (4.7.24)\]

for all \(\ell \geq 1\) simply by choosing to normalize our eigenstates as

\[
\langle \tilde{E}_a | \ell E_a\rangle = 1. \quad (4.7.25)\]

To this end, let us first recall that an eigenvector |\(\lambda\)\rangle is only defined up to an overall multiplicative complex amplitude \(\chi\); i.e., if \(A|\lambda\rangle = \lambda|\lambda\rangle\) so does \(A(\chi|\lambda\rangle) = \lambda(\chi|\lambda\rangle)\). Therefore, since \(\chi\) multiplies every coefficient when we expand \(\chi|\lambda\rangle\) as a superposition over basis vectors \(\{|j\}\}\), 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 \(|\lambda\rangle \rightarrow \chi|\lambda\rangle\), \(\langle i| \lambda\rangle \rightarrow \chi \langle i| \lambda\rangle\).
This is precisely the case in eq. (4.7.25), where we know $|\bar{E}_a\rangle$ must have significant overlap with the exact eigenstate $|E_a\rangle$. Expanding eq. (4.7.25),

$$\langle \bar{E}_a | \bar{E}_a \rangle + \sum_{\ell=1}^{+\infty} \epsilon^\ell \langle \bar{E}_a | \epsilon E_a \rangle = 1$$  \hspace{1cm} (4.7.26)

$$\sum_{\ell=1}^{+\infty} \epsilon^\ell \langle \bar{E}_a | \epsilon E_a \rangle = 0; \hspace{1cm} (4.7.27)$$

followed by setting to zero the coefficient of each $\epsilon^{\ell\geq 1}$, we arrive at eq. (4.7.24).

Additionally, starting with $|1E_a\rangle$, notice eq. (4.7.17) is invariant under the replacement $|1E_a\rangle \rightarrow |1E_a\rangle + \chi_1 \bar{E}_a$ – for arbitrary complex number $\chi_1$ — because of the eigen-equation (4.7.1). In other words, if we found a solution $|1E_a\rangle = |\psi_1\rangle$; then is $|1E_a\rangle = |\psi_1\rangle + \chi_1 |\bar{E}_a\rangle$. Hence, if $\langle \bar{E}_a | \epsilon E_a \rangle = \langle \bar{E}_a | \psi_1 \rangle \neq 0$, we may simply choose $\chi_1$ such that the new solution $|1E_a\rangle_{\text{new}} \equiv |\psi_1\rangle + \chi E_a$ satisfies $\langle \bar{E}_a | 1E_a \rangle_{\text{new}} = \langle \bar{E}_a | \psi_1 \rangle + \chi = 0$. Now, suppose we have solved $|1E_a\rangle$ from $i = 1$ up to $i = \ell - 1$. Then we see, just like the $\ell = 1$ case, both $|\epsilon E_a\rangle = |\psi_{\ell}\rangle$ and $|\epsilon E_a\rangle = |\psi_{\ell}\rangle + \chi \ell |\bar{E}_a\rangle$ solve eq. (4.7.19) as long as $|\psi_{\ell}\rangle$ is a solution. Therefore if $\langle \bar{E}_a | \epsilon E_a \rangle$ were not zero, then the ‘new’ solution $|\epsilon E_a\rangle_{\text{new}} \equiv |\psi_{\ell}\rangle + \chi \ell |\bar{E}_a\rangle$ can be made to satisfy $0 = \langle \bar{E}_a | \epsilon E_a \rangle_{\text{new}} = \langle \bar{E}_a | \psi_{\ell}\rangle + \chi$ simply by choosing $\chi = -\langle \bar{E}_a | \psi_{\ell}\rangle$.

The freedom to shift the perturbation by a constant multiple of $|\bar{E}_a\rangle$ at each step of the construction is related to the freedom to re-scale the eigenket $|E_a\rangle$ itself. For, at the $\ell$th step, when we perform $|\epsilon E_a\rangle \rightarrow |\epsilon E_a\rangle + \chi \ell |E_a\rangle$, the normalization condition in eq. (4.7.25) is altered into

$$\langle \bar{E}_a | E_a \rangle = \langle \bar{E}_a | \sum_{k=1}^{\ell-1} \epsilon^k |kE_a\rangle + \epsilon^\ell (|\epsilon E_a\rangle + \chi \ell |\bar{E}_a\rangle) + \mathcal{O}(\epsilon^{\ell+1}) \rangle$$

$$= 1 + \epsilon^\ell \chi \ell + \mathcal{O}(\epsilon^{\ell+1})$$  \hspace{1cm} (4.7.28)

As you will witness in Problem (4.90) 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 (4.7.21)–(4.7.23), we may thus employ eq. (4.7.24) to gather:

$$0 = -\langle \bar{E}_a | H_1 | \bar{E}_a \rangle + \delta_1 E_a,$$  \hspace{1cm} (4.7.29)

$$0 = -\langle \bar{E}_a | H_1 | 1E_a \rangle - \langle \bar{E}_a | H_2 | E_a \rangle + \delta_2 E_a,$$  \hspace{1cm} (4.7.30)

$$\text{......}$$

$$0 = -\langle \bar{E}_a | H_1 | \ell-1E_a \rangle - \langle \bar{E}_a | H_2 | \ell-2E_a \rangle$$

$$\text{......} - \langle \bar{E}_a | H_\ell-2 | 2E_a \rangle - \langle \bar{E}_a | H_\ell-1 | 1E_a \rangle - \langle \bar{E}_a | H_\ell | \bar{E}_a \rangle + \delta_\ell E_a.$$  \hspace{1cm} (4.7.31)

The eigensystem of eq. (4.7.1) also tells us, for $b \neq a$,

$$\langle \bar{E}_b | (H_0 - \bar{E}_a) = (\bar{E}_b - \bar{E}_a) \langle \bar{E}_b |.$$  \hspace{1cm} (4.7.32)

Therefore, $\langle \bar{E}_{b \neq a} |$ acting on both sides of equations (4.7.17), (4.7.18), (4.7.19), etc. would produce

$$(\bar{E}_b - \bar{E}_a) \langle \bar{E}_b | 1E_a \rangle = -\langle \bar{E}_b | H_1 | \bar{E}_a \rangle,$$  \hspace{1cm} (4.7.33)
\[
(\tilde{E}_b - \tilde{E}_a) \langle \tilde{E}_b | 2E_a \rangle = -\langle \tilde{E}_b | H_1 - \delta_1 E_a | 1E_a \rangle - \langle \tilde{E}_b | H_2 | \tilde{E}_a \rangle, \quad (4.7.34)
\]

\[
\ldots \ldots \quad (4.7.35)
\]

\[
(\tilde{E}_b - \tilde{E}_a) \langle \tilde{E}_b | \epsilon E_a \rangle = -\langle \tilde{E}_b | H_1 - \delta_1 E_a | \epsilon - 1E_a \rangle - \langle \tilde{E}_b | H_2 - \delta_2 E_a | \epsilon - 2E_a \rangle
\]

\[
\quad \cdots - \langle \tilde{E}_b | H_{\epsilon - 2} - \delta_{\epsilon - 2} E_a | 2E_a \rangle - \langle \tilde{E}_b | H_{\epsilon - 1} - \delta_{\epsilon - 1} E_a | 1E_a \rangle
\]

\[
- \langle \tilde{E}_b | H_\ell | \tilde{E}_a \rangle.
\]

From eq. (4.7.29), we see that the first order correction to the energy is the expectation value of the first correction to the Hamiltonian:

\[
\delta_1 E_a = \langle \tilde{E}_a | H_1 | \tilde{E}_a \rangle. \quad (4.7.36)
\]

The off-diagonal term in eq. (4.7.33) allows us to infer, the first-order correction to the \(a\)th eigenstate along the \(|\tilde{E}_b \neq a\rangle\)-direction is

\[
\langle \tilde{E}_b \neq a | 1E_a \rangle = \frac{\langle \tilde{E}_b | H_1 | \tilde{E}_a \rangle}{E_a - E_b}. \quad (4.7.37)
\]

Turning to the second order corrections, the completeness relation \(\mathbb{I} = \sum_c |\tilde{E}_c\rangle \langle \tilde{E}_c|\), together with equations (4.7.5), (4.7.24), (4.7.36), and (4.7.37), allow us to massage equations (4.7.30) and (4.7.34):

\[
0 = - \sum_b \langle \tilde{E}_a | H_1 | \tilde{E}_b \rangle \langle \tilde{E}_b | 1E_a \rangle - \langle \tilde{E}_a | H_2 | \tilde{E}_a \rangle + \delta_2 E_a \quad (4.7.38)
\]

\[
\delta_2 E_a = \sum_{b \neq a} \langle \tilde{E}_a | H_1 | \tilde{E}_b \rangle \langle \tilde{E}_b | 1E_a \rangle + \langle \tilde{E}_a | H_2 | \tilde{E}_a \rangle
\]

\[
= \sum_{b \neq a} \langle \tilde{E}_a | H_1 | \tilde{E}_b \rangle \frac{\langle \tilde{E}_b | H_1 | \tilde{E}_a \rangle}{E_a - E_b} + \langle \tilde{E}_a | H_2 | \tilde{E}_a \rangle. \quad (4.7.40)
\]

\[
(\tilde{E}_b - \tilde{E}_a) \langle \tilde{E}_b | 2E_a \rangle = - \sum_c \langle \tilde{E}_b | H_1 - \delta_1 E_a | \tilde{E}_c \rangle \langle \tilde{E}_c | 1E_a \rangle - \langle \tilde{E}_b | H_2 | \tilde{E}_a \rangle \quad (4.7.41)
\]

\[
= - \sum_{c \neq a} \langle \tilde{E}_b | H_1 - \langle E_a | H_1 | E_a \rangle | E_c \rangle \frac{\langle E_c | H_1 | \tilde{E}_a \rangle}{E_a - E_c} - \langle \tilde{E}_b | H_2 | \tilde{E}_a \rangle
\]

\[
\langle \tilde{E}_b | 2E_a \rangle = (\tilde{E}_a - \tilde{E}_b)^{-1} \left( \sum_{c \neq a} \langle \tilde{E}_b | H_1 | \tilde{E}_c \rangle \frac{\langle E_c | H_1 | \tilde{E}_a \rangle}{E_a - E_c} \right)
\]

\[
- \langle \tilde{E}_a | H_1 | \tilde{E}_a \rangle \frac{\langle \tilde{E}_b | H_1 | \tilde{E}_a \rangle}{E_a - E_b} + \langle \tilde{E}_b | H_2 | \tilde{E}_a \rangle \right). \quad (4.7.42)
\]

We now have the necessary ingredients to construct both the eigenvalue and eigenstate perturbations \(|\epsilon E_a\rangle = \sum_b |\tilde{E}_b\rangle \langle \tilde{E}_b | \epsilon E_a \rangle\) up to second order. Collecting the first order results from equations (4.7.36) and (4.7.37); the second order ones from equations (4.7.40) and (4.7.42); and remembering we have chosen to satisfy the constraint in eq. (4.7.24) that eigenket perturbations are orthogonal to \(|\tilde{E}_a\rangle\):

147
Non-Degenerate PT

The eigensystem of

\[ H = H_0 + \epsilon H_1 + \epsilon^2 H_2 + \mathcal{O}(\epsilon^3), \]  

expressed in terms of the unperturbed ones – i.e., \( \{|E_a\rangle\} \) obeying \( H_0 |E_a\rangle = E_a |E_a\rangle \) – are given by

\[ H |E_a\rangle = E_a |E_a\rangle, \]  

\[ E_a = \bar{E}_a + \epsilon \langle \bar{E}_a | H_1 | \bar{E}_a \rangle + \epsilon^2 \left( \sum_{b \neq a} \langle \bar{E}_a | H_1 | \bar{E}_b \rangle \frac{\langle \bar{E}_b | H_1 | E_a \rangle}{E_a - E_b} + \langle \bar{E}_a | H_2 | E_a \rangle \right) + \mathcal{O}(\epsilon^3), \]  

\[ |E_a\rangle = |E_a\rangle + \epsilon \sum_{b \neq a} |E_b\rangle \frac{\langle \bar{E}_b | H_1 | E_a \rangle}{E_a - E_b} + \epsilon^2 \sum_{b \neq a} |E_b\rangle \left( \sum_{c \neq a} \langle \bar{E}_b | H_1 | E_c \rangle \frac{\langle \bar{E}_c | H_1 | E_a \rangle}{E_a - E_c} \right) - \langle \bar{E}_a | H_1 | E_a \rangle \frac{\langle \bar{E}_a | H_1 | E_a \rangle}{E_a - E_b} + \langle \bar{E}_a | H_2 | E_a \rangle \right) + \mathcal{O}(\epsilon^3). \]

Problem 4.89. ‘Inverse’ of \( H_0 - \bar{E}_a \)

The operator \( H_0 - \bar{E}_a \) on the left hand sides of equations (4.7.17), (4.7.18) and (4.7.19), etc. has no inverse because it has a null eigenvector \( |\bar{E}_a\rangle\); i.e., \( (H_0 - \bar{E}_a) |E_a\rangle = 0 \). 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:

\[ (H_0 - \bar{E}_a)_{\perp}^{-1} \equiv \sum_{b \neq a} |\bar{E}_b\rangle \langle \bar{E}_b | \frac{E_a - E_b}{|\bar{E}_a\rangle \langle \bar{E}_a |}. \]

Verify that

\[ (H_0 - \bar{E}_a)_{\perp}^{-1}(H_0 - \bar{E}_a) = (H_0 - \bar{E}_a)(H_0 - \bar{E}_a)_{\perp}^{-1} \]

\[ = \sum_{b \neq a} |\bar{E}_b\rangle \langle \bar{E}_b | = 1 - |\bar{E}_a\rangle \langle \bar{E}_a |. \]

Let \( |\psi\rangle \) be orthogonal to \( |\bar{E}_a\rangle \) but otherwise arbitrary. Explain why

\[ (H_0 - \bar{E}_a)_{\perp}^{-1}(H_0 - \bar{E}_a) |\psi\rangle = (H_0 - \bar{E}_a)(H_0 - \bar{E}_a)_{\perp}^{-1} |\psi\rangle = |\psi\rangle. \]

Now explain how we may solve for \( \{|i E_a\rangle | i = 1, 2, 3, \ldots \} \) from (4.7.17), (4.7.18) and (4.7.19), etc. through the pseudo-inverse \( (H_0 - E_a)_{\perp}^{-1} \). □

Problem 4.90. Unit Norm Eigenket

Above, we have argued that, if we had already solved \( |i E_a\rangle \) from \( i = 1 \) up to \( i = \ell - 1 \), then if \( |i E_a\rangle \) solves eq. (4.7.19) – so does \( |i E_a\rangle + \chi |E_a\rangle \). For example, if \( |1 E_a\rangle \) is given by equations (4.7.24) and (4.7.37); then both \( |E_a\rangle + \epsilon |1 E_a\rangle + \ldots \)
\(\epsilon^2 \ket{2E_a} \) and \((1 + \epsilon^2 \chi_2) \ket{\tilde{E}_a} + \epsilon \ket{1E_a} + \epsilon^2 \ket{2E_a}\) are eigenkets of \(H = H_0 + \epsilon H_1 + \epsilon^2 H_2\) up to quadratic order in \(\epsilon\); as long as \(\ket{2E_a}\) solves eq. (4.7.18).

Demonstrate that we may normalize eq. (4.7.46) to unity, up to \(O(\epsilon^2)\), by shifting the second order correction by

\[
\ket{2E_a} \to \ket{2E_a} - \frac{\ket{\tilde{E}_a}}{2} \sum_{c \neq a} \left| \frac{\langle \tilde{E}_c | H_1 | \tilde{E}_a \rangle}{\tilde{E}_a - \tilde{E}_c} \right|^2. \tag{4.7.51}
\]

Hence, up to second order in perturbation theory,

\[
\ket{E_a} = \ket{\tilde{E}_a} + \epsilon \sum_{b \neq a} \frac{\langle \tilde{E}_b | H_1 | \tilde{E}_a \rangle}{\tilde{E}_a - \tilde{E}_b} + \epsilon^2 \left\{ \sum_{b \neq a} \frac{\langle \tilde{E}_b | H_1 | \tilde{E}_a \rangle}{\tilde{E}_a - \tilde{E}_b} \left( \sum_{c \neq a} \frac{\langle \tilde{E}_c | H_1 | \tilde{E}_a \rangle}{\tilde{E}_a - \tilde{E}_c} \right) - \frac{\ket{\tilde{E}_a}}{2} \sum_{c \neq a} \left| \frac{\langle \tilde{E}_c | H_1 | \tilde{E}_a \rangle}{\tilde{E}_a - \tilde{E}_c} \right|^2 \right\} \tag{4.7.52}
\]

is not only an eigenket of \(H = H_0 + \epsilon H_1 + \epsilon^2 H_2\) it is also unit norm. \(\square\)

**Degenerate Case** If the eigenvalue \(\tilde{E}_a\) in eq. (4.7.1) (and (4.7.20)) is degenerate, the preceding discussion goes through, up to equations (4.7.17), (4.7.18), and (4.7.19); but we now need to add an enumeration label \(j\) to the eigenstate – namely, \(\ket{\tilde{E}_a; j}\) – that runs from 1 through \(N\), the dimension of this degenerate subspace. Beginning with eq. (4.7.17), we have

\[
(H_0 - \tilde{E}_a) \ket{1E_a; j} = -(H_1 - \delta_1 E_a) \ket{\tilde{E}_a; j}. \tag{4.7.53}
\]

Let us now act \(\langle \tilde{E}_a; i | \) on both sides of this equations, keeping in mind eq. (4.7.20).

\[
0 = - \langle \tilde{E}_a; i | H_1 | \tilde{E}_a; j \rangle + \delta_1 E_a \langle \tilde{E}_a; i | \tilde{E}_a; j \rangle. \tag{4.7.54}
\]

Within the degenerate subspace, we may of course choose the \(\{ \ket{\tilde{E}_a; i} \}\) to be orthonormal,

\[
\langle \tilde{E}_a; i | \tilde{E}_a; j \rangle = \delta^i_j. \tag{4.7.55}
\]

Eq. (4.7.54) then reads

\[
\delta_1 E_a \cdot \delta^i_j = \langle \tilde{E}_a; i | H_1 | \tilde{E}_a; j \rangle. \tag{4.7.56}
\]

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. (4.7.36). 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. (4.7.36), the off-diagonal \(i \neq j\) equations appear to tell us \(H_1\) must have trivial off-diagonal components,

\[
\langle \tilde{E}_a; i | H_1 | \tilde{E}_a; j \rangle = 0, \quad (i \neq j). \tag{4.7.57}
\]

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. (4.7.56) as an instruction to choose the basis of this degenerate
Then, the expansion in eq. (4.7.58) tells us

\[
|E_a; j\rangle = |\bar{E}_a; j\rangle + \sum_{\ell=1}^{+\infty} \sum_{s \neq a} \epsilon^\ell |\bar{E}_s\rangle \langle E_s \; \ell E_a; j\rangle + \sum_{\ell=1}^{+\infty} \sum_{i=1}^{N} \epsilon^\ell |\bar{E}_a; i\rangle \langle \bar{E}_a \; i E_a; j\rangle,
\]

(4.7.58)

and eq. (4.7.53) takes the form

\[
(H_0 - \bar{E}_a) |E_a; j\rangle = - (H_1 - \delta_{1,j} E_a) |\bar{E}_a; j\rangle.
\]

(4.7.60)

Like the non-degenerate case, we now require that the eigenket be normalized as

\[
\langle \bar{E}_a; j \; | E_a; j \rangle = 1.
\]

(4.7.61)

Then, the expansion in eq. (4.7.58) tells us

\[
\langle \bar{E}_a; j \; | \bar{E}_a; j \rangle + \sum_{\ell=1}^{\infty} \epsilon^\ell \langle \bar{E}_a; j \; \ell E_a; j \rangle = 1
\]

(4.7.62)

\[
\sum_{\ell=1}^{\infty} \epsilon^\ell \langle \bar{E}_a; j \; \ell E_a; j \rangle = 0.
\]

(4.7.63)

Setting the coefficient of \(\epsilon^\ell\) to zero,

\[
\langle \bar{E}_a; j \; | \ell \geq 1 \; E_a; j \rangle = 0.
\]

(4.7.64)

Even though the left hand side of eq. (4.7.60) appears to be invariant under the replacement

\[
|E_a; j\rangle \rightarrow |E_a; j\rangle + \lambda_1 |\bar{E}_a; i\rangle,
\]

we will discover from the \(O(\epsilon^2)\) equations below that, it is inconsistent to set \(\langle \bar{E}_a; i \neq j \; | \ell E_a; j \rangle = 0\). However, eq. (4.7.64) will remain consistent.

For now, let us apply \(\langle \bar{E}_b \rangle\), for \(\bar{E}_b \neq \bar{E}_a\), on both sides of eq. (4.7.60).

\[
(\bar{E}_b - \bar{E}_a) \langle \bar{E}_b \neq a \; \ell E_a; j \rangle = - \langle \bar{E}_b \neq a \; | H_1 \; \bar{E}_a; j \rangle
\]

(4.7.65)

Since \(\bar{E}_b \neq \bar{E}_a\), we have eliminated the \(\delta_{1,j} E_a\) term in eq. (4.7.60) via the orthogonality condition

\[
\langle \bar{E}_b \neq a \; \bar{E}_a; j \rangle = 0.
\]

(4.7.66)
Eq. (4.7.65) returns us the component of the first order eigenstate correction along \( |E_{b \neq a}\rangle \).

\[
\langle E_{b \neq a} | I_{E_{a}; j} \rangle = - \langle \bar{E}_{b \neq a} | H_{1} | E_{a}; j \rangle \quad (4.7.67)
\]

As already alluded to, in order to determine \( \langle E_{a}; i | I_{E_{a}; j} \rangle \), we need to turn to the \( \mathcal{O}(\epsilon^{2}) \) eq. (4.7.18).

\[
(H_{0} - \bar{E}_{a}) | 2E_{a}; j \rangle = - (H_{1} - \delta_{1,j}E_{a}) | 1E_{a}; j \rangle - (H_{2} - \delta_{2,j}E_{a}) | \bar{E}_{a}; j \rangle . \quad (4.7.68)
\]

Keeping in mind eq. (4.7.56), applying \( \langle E_{a}; i | \) and \( \langle \bar{E}_{b \neq a} | \) on both sides now yield, respectively

\[
0 = - \langle E_{a}; i | H_{1} | 1E_{a}; j \rangle + \delta_{1,j}E_{a} \langle \bar{E}_{a}; i | 1E_{a}; j \rangle - \langle \bar{E}_{a}; i | H_{2} | \bar{E}_{a}; j \rangle + \delta_{2,j}E_{a}\delta_{i}^{j} \quad (4.7.69)
\]

\[
0 = - \sum_{b \neq a} \langle E_{a}; i | H_{1} | \bar{E}_{b} \rangle \frac{\langle \bar{E}_{b} | H_{1} | E_{a}; j \rangle}{E_{a} - E_{b}} - (\delta_{1,i}E_{a} - \delta_{1,j}E_{a}) \langle \bar{E}_{a}; i | 1E_{a}; j \rangle - \langle \bar{E}_{a}; i | H_{2} | \bar{E}_{a}; j \rangle + \delta_{2,j}E_{a}\delta_{i}^{j} \quad (4.7.70)
\]

and

\[
(E_{b} - \bar{E}_{a}) \langle \bar{E}_{b \neq a} | 2E_{a}; j \rangle = - \langle \bar{E}_{b \neq a} | H_{1} | 1E_{a}; j \rangle + \delta_{1,j}E_{a} \langle \bar{E}_{b \neq a} | 1E_{a}; j \rangle - \langle \bar{E}_{b \neq a} | H_{2} | \bar{E}_{a}; j \rangle + \delta_{2,j}E_{a} \langle \bar{E}_{b \neq a} | \bar{E}_{a}; j \rangle \quad (4.7.71)
\]

\[
\langle \bar{E}_{b} | 2E_{a}; j \rangle = (E_{a} - E_{b})^{-1} \left( \sum_{c \neq a} \langle \bar{E}_{b} | H_{1} | \bar{E}_{c} \rangle \frac{\langle \bar{E}_{c} | H_{1} | \bar{E}_{a}; j \rangle}{E_{a} - E_{c}} \right.
\]

\[
+ \sum_{k=1}^{N} \langle \bar{E}_{b \neq a} | H_{1} | \bar{E}_{a}; k \rangle \langle \bar{E}_{a}; k | 1E_{a}; j \rangle - \delta_{1,j}E_{a} \frac{\langle \bar{E}_{b} | H_{1} | \bar{E}_{a}; j \rangle}{E_{a} - E_{b}} + \langle \bar{E}_{b} | H_{2} | \bar{E}_{a}; j \rangle \right) . \quad (4.7.72)
\]

By setting \( i = j \) in eq. (4.7.70) we immediately obtain the second order corrections to the eigenvalue

\[
\delta_{2,j}E_{a} = \sum_{b \neq a} \langle \bar{E}_{a}; j | H_{1} | \bar{E}_{b} \rangle \frac{\langle \bar{E}_{b} | H_{1} | \bar{E}_{a}; j \rangle}{E_{a} - E_{b}} + \langle \bar{E}_{a}; j | H_{2} | \bar{E}_{a}; j \rangle . \quad (4.7.73)
\]

whereas the \( i \neq j \) equations hand us the components of \( | 1E_{a}; j \rangle \) along \( | \bar{E}_{a}; i \neq j \rangle \),

\[
\langle \bar{E}_{a}; i | 1E_{a}; j \rangle = - (\delta_{1,i}E_{a} - \delta_{1,j}E_{a})^{-1}
\]

\[
\times \left( \sum_{b \neq a} \langle \bar{E}_{a}; i | H_{1} | \bar{E}_{b} \rangle \frac{\langle \bar{E}_{b} | H_{1} | \bar{E}_{a}; j \rangle}{E_{a} - E_{b}} \right. + \langle \bar{E}_{a}; i | H_{2} | \bar{E}_{a}; j \rangle \right) . \quad (4.7.74)
\]

This explicitly demonstrates, as long as \( i \neq j \), it is inconsistent to set \( \langle \bar{E}_{a}; i | 1E_{a}; j \rangle = 0 \).
Inserting eq. (4.7.74) into eq. (4.7.72),

\[
\langle \bar{E}_b |_a 2E_a; j \rangle = (\bar{E}_a - \bar{E}_b)^{-1}\left( \sum_{c \neq a} \langle \bar{E}_b | H_1 | \bar{E}_c \rangle \frac{\langle E_c | H_1 | E_a; j \rangle}{\bar{E}_a - \bar{E}_c} \right) - \frac{1}{\bar{E}_a - \bar{E}_b} 
\]

\[
- \sum_{k=1}^{N} \left( \frac{\langle \bar{E}_b | H_1 | \bar{E}_a; k \rangle}{\delta_{1,k} E_a - \delta_{1,j} E_a} \left( \sum_{c \neq a} \langle \bar{E}_a; k | H_1 | \bar{E}_c \rangle \frac{\langle E_c | H_1 | E_a; j \rangle}{\bar{E}_a - \bar{E}_c} + \langle E_a; k | H_2 | E_a; j \rangle \right) \right)
\]

\[
- \langle \bar{E}_a; j | H_1 | \bar{E}_a; j \rangle \left( \frac{\langle \bar{E}_b | H_1 | \bar{E}_a; j \rangle}{\bar{E}_a - \bar{E}_b} + \langle \bar{E}_b | H_2 | \bar{E}_a; j \rangle \right) \right).
\]

To derive \( \langle \bar{E}_a; i | 2E_a; j \rangle \) we need to move on to the \( O(\epsilon^3) \) equations. Applying \( \langle \bar{E}_a; j \rangle \) on both sides of \( \ell = 3 \) version of eq. (4.7.35),

\[
(H_0 - \bar{E}_a) \langle 3E_a; j \rangle = - \sum_{s=1}^{2} (H_s - \delta_{s,j} E_a) \langle \ell_s E_a; j \rangle - (H_3 - \delta_{3,j} E_a) \langle \bar{E}_a; j \rangle.
\]

Because we are only after \( \langle E_a; i | 2E_a; j \rangle \), let us apply \( \langle E_a; i \rangle \) on eq. (4.7.76) to eliminate \( \langle 3E_a; j \rangle \) (cf. eq. (4.7.20)).

\[
0 = - \left( \langle E_a; i | H_1 | 2E_a; j \rangle - \delta_{1,j} E_a \langle E_a; i | 2E_a; j \rangle \right) - \left( \langle E_a; i | H_2 | 1E_a; j \rangle - \delta_{2,j} E_a \langle E_a; i | 1E_a; j \rangle \right) - \langle \bar{E}_a; i | H_3 | \bar{E}_a; j \rangle \right).
\]

Inserting the zeroth order completeness relations,

\[
\sum_{c \neq a} \langle \bar{E}_a; i | H_1 | \bar{E}_c \rangle \langle \bar{E}_c | 2E_a; j \rangle + \sum_{k=1}^{N} \langle \bar{E}_a; i | H_1 | \bar{E}_a; k \rangle \langle \bar{E}_a; k | 2E_a; j \rangle - \delta_{1,j} E_a \langle \bar{E}_a; i | 2E_a; j \rangle \right)
\]

\[
= - \sum_{c \neq a} \langle \bar{E}_a; i | H_2 | \bar{E}_c \rangle \langle \bar{E}_c | 1E_a; j \rangle - \sum_{k=1}^{N} \langle \bar{E}_a; i | H_2 | \bar{E}_a; k \rangle \langle \bar{E}_a; k | 1E_a; j \rangle + \delta_{2,j} E_a \langle \bar{E}_a; i | 1E_a; j \rangle \right)
\]

\[
\left( \delta_{1,j} E_a - \delta_{1,j} E_a \right) \langle \bar{E}_a; i | 2E_a; j \rangle \right) \right)
\]

\[
- \sum_{f \neq a} \langle \bar{E}_a; i | H_1 | \bar{E}_f \rangle \left( \sum_{c \neq a} \langle \bar{E}_f | H_1 | \bar{E}_c \rangle \langle \bar{E}_c | H_1 | \bar{E}_a; j \rangle \right)
\]

\[
- \sum_{k=1}^{N} \left( \frac{\langle \bar{E}_f | H_1 | \bar{E}_a; k \rangle}{\delta_{1,k} E_a - \delta_{1,j} E_a} \left( \sum_{c \neq a} \langle \bar{E}_a; k | H_1 | \bar{E}_c \rangle \frac{\langle E_c | H_1 | E_a; j \rangle}{\bar{E}_a - \bar{E}_c} + \langle E_a; k | H_2 | E_a; j \rangle \right) \right)
\]

\[
- \langle \bar{E}_a; j | H_1 | \bar{E}_a; j \rangle \left( \frac{\langle \bar{E}_f | H_1 | \bar{E}_a; j \rangle}{\bar{E}_a - \bar{E}_b} + \langle \bar{E}_f | H_2 | \bar{E}_a; j \rangle \right) \right)
\]

\[
- \sum_{c \neq a} \langle \bar{E}_a; i | H_2 | \bar{E}_c \rangle \langle \bar{E}_c | H_1 | \bar{E}_a; j \rangle \frac{\langle \bar{E}_c | H_1 | \bar{E}_a; j \rangle}{\bar{E}_a - \bar{E}_c}
\]

152
\[
+ \sum_{k=1}^{N} \frac{\langle \tilde{E}_a; i \mid H_2 \mid \tilde{E}_a; k \rangle}{\delta_{1,k} E_a - \delta_{1,i} E_a} \left( \sum_{b \neq a} \frac{\langle \tilde{E}_b; k \mid H_1 \mid \tilde{E}_a; j \rangle}{E_a - E_b} \right) + \langle \tilde{E}_a; k \mid H_2 \mid \tilde{E}_a; j \rangle \\
- \left( \sum_{b \neq a} \frac{\langle \tilde{E}_a; j \mid H_1 \mid \tilde{E}_b \rangle}{E_a - E_b} \right) \left( \sum_{i} \frac{\langle \tilde{E}_i \mid H_1 \mid \tilde{E}_a; j \rangle}{E_a - E_b} \right) \left( \sum_{b \neq a} \frac{\langle \tilde{E}_b \mid H_1 \mid \tilde{E}_a; j \rangle}{E_a - E_b} \right) \left( \delta_{1,i} E_a - \delta_{1,j} E_a \right)^{-1} \\
\times \left( \sum_{b \neq a} \frac{\langle \tilde{E}_a; i \mid H_1 \mid \tilde{E}_b \rangle}{E_a - E_b} \right) \left( \sum_{i} \frac{\langle \tilde{E}_i \mid H_1 \mid \tilde{E}_a; j \rangle}{E_a - E_b} \right) - \langle \tilde{E}_a; i \mid H_3 \mid \tilde{E}_a; j \rangle.
\]

Let us summarize the situation thus far.

- When a given set of eigenkets \( \{ | \tilde{E}_a; j \rangle \mid i = 1, 2, \ldots, N \} \) of \( H_0 \) is degenerate, to find the corresponding eigenvectors of the perturbed operator
  \[
  H = H_0 + \epsilon H_1 + \epsilon^2 H_2 + \ldots,
  \]  
  first ensure these \( \{ | \tilde{E}_a; j \rangle \} \) have been chosen such that \( H_1 \) is diagonal within this subspace (cf. eq. (4.7.56)):
  \[
  \langle \tilde{E}_a; i \mid H_1 \mid \tilde{E}_a; j \rangle = \delta_{1,j} E_a \delta_{i,j}.
  \]  

- With respect to such a basis, the perturbed eigenvalue up to second order then reads (cf. equations (4.7.56) and (4.7.73)):
  \[
  E_{a,j} = \tilde{E}_a + \epsilon \langle \tilde{E}_a; j \mid H_1 \mid \tilde{E}_a; j \rangle \\
  + \epsilon^2 \left( \sum_{E_b \neq E_a} \frac{\langle \tilde{E}_b; j \mid H_1 \mid \tilde{E}_a \rangle^2}{E_a - E_b} + \langle \tilde{E}_a; j \mid H_2 \mid \tilde{E}_a; j \rangle \right) + O(\epsilon^3).
  \]  

- If the first order corrections completely break the degeneracy, then the eigenkets of \( H \) up to second order, namely
  \[
  | E_a; j \rangle = | \tilde{E}_a; j \rangle + \epsilon \sum_{E_b \neq E_a} \frac{\langle \tilde{E}_b \mid H_1 \mid \tilde{E}_a; j \rangle}{E_a - E_b} \\
  - \sum_{i=1}^{N} \frac{| \tilde{E}_a; i \rangle}{\delta_{1,i} E_a - \delta_{1,j} E_a} \left( \sum_{b \neq a} \frac{\langle \tilde{E}_a; i \mid H_1 \mid \tilde{E}_b \rangle}{E_a - E_b} \right) \left( \sum_{b \neq a} \frac{\langle \tilde{E}_a; j \mid H_1 \mid \tilde{E}_b \rangle}{E_a - E_b} \right) \\
  + \epsilon^2 \left( \sum_{i=1}^{N} \frac{| \tilde{E}_a; i \rangle \langle \tilde{E}_a; i \mid 2E_a; j \rangle}{2 E_a; j \rangle} + \sum_{E_b \neq E_a} \langle \tilde{E}_b \mid 2E_a; j \rangle \right) + O(\epsilon^3),
  \]  
  may be constructed using the coefficients in equations (4.7.67), (4.7.74), (4.7.75), and (4.7.79).

Examples
4.7.2 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.

$$E_0 \leq \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle}, \quad \forall |\psi\rangle. \quad (4.7.84)$$

To see this, we exploit the fact that $H$ is Hermitian to insert a complete set of its eigenstates $\{|E_n\rangle\}$ on the right hand side.

$$\frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} = \sum_n E_n \frac{\langle \psi | E_n \rangle \langle E_n | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (4.7.85)$$

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,

$$E_0 \langle \psi | E_n \rangle \langle E_n | \psi \rangle = E_0 |\langle \psi | E_n \rangle |^2 \leq E_n |\langle \psi | E_n \rangle |^2 = E_n \langle \psi | E_n \rangle \langle E_n | \psi \rangle \quad (4.7.86)$$

for all $n \neq 0$. Therefore

$$\sum_n E_n \frac{\langle \psi | E_n \rangle \langle E_n | \psi \rangle}{\langle \psi | \psi \rangle} = \sum_0 E_0 \frac{\langle \psi | E_0 \rangle \langle E_0 | \psi \rangle}{\langle \psi | \psi \rangle} + \sum_{n \neq 0} E_n |\langle \psi | E_n \rangle |^2 \geq E_0 \frac{\sum_n \langle \psi | E_n \rangle \langle E_n | \psi \rangle}{\langle \psi | \psi \rangle} = E_0 \frac{\langle \psi | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (4.7.87)$$

**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 $\{|\alpha_i\rangle\}$, then

$$\mathcal{E}(\alpha_1) \equiv \frac{\langle \psi; \{\alpha_1\} | H | \psi; \{\alpha_1\} \rangle}{\langle \psi; \{\alpha_1\} | \psi; \{\alpha_1\} \rangle} \quad (4.7.89)$$

is necessarily a function of these $\alpha$s. We may then search for its minimum – i.e., evaluate it at the $\alpha_1$ obeying $\partial \mathcal{E} / \partial \alpha_1 = 0$. This must yield, at least within this class of states $|\psi; \{\alpha_1\}\rangle$, a number closest to $E_0$ (from above): $E_0 \leq \mathcal{E}(\tilde{\alpha}_1)$. There is a certain art here, to cook up the right family of states, with parameters $\{|\alpha_1\}\}$ introduced, so that one may obtain a good enough estimate for the application at hand.

**Extremization**  We also note that, the first order perturbation of the normalized expectation value obtained by varying the state $|\psi\rangle$ is

$$\delta \left( \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \right) = \frac{\langle \delta \psi | H | \psi \rangle + \langle \psi | H | \delta \psi \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \delta \psi | \psi \rangle \langle \psi | \delta \psi \rangle}{\langle \psi | \psi \rangle^2} \quad (4.7.90)$$
\[
\langle \delta \psi \rangle \left( H |\psi\rangle - |\psi\rangle \frac{\langle \psi |H| \psi\rangle}{\langle \psi |\psi\rangle} \right) + \text{h.c.} \quad (4.7.91)
\]

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

\[
\delta \left( \frac{\langle \psi |H| \psi\rangle}{\langle \psi |\psi\rangle} \right) = 0. \quad (4.7.92)
\]

On the other hand, if the first order variation of this normalized expectation value is zero for all variations \(|\delta \psi\rangle\), then

\[
0 = \langle \delta \psi \rangle \left( H |\psi\rangle - |\psi\rangle \frac{\langle \psi |H| \psi\rangle}{\langle \psi |\psi\rangle} \right) + \text{h.c.}
\]

\[
= 2\text{Re} \langle \delta \psi \rangle \left( H |\psi\rangle - |\psi\rangle \frac{\langle \psi |H| \psi\rangle}{\langle \psi |\psi\rangle} \right). \quad (4.7.93)
\]

Since \(|\delta \psi\rangle\) is arbitrary, we conclude that the eigensystem equation

\[
H |\psi\rangle = |\psi\rangle \frac{\langle \psi |H| \psi\rangle}{\langle \psi |\psi\rangle} \quad (4.7.94)
\]

must hold. To sum:

**Eigenvectors extremize** \( E\) 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 \(\lambda\).

**Example** Let us consider the Hermitian operator consisting of the unit radial vector \(\hat{r}\) dotted into the Pauli matrices in eq. (3.2.17):

\[
\hat{r} \cdot \vec{\sigma} \equiv \delta_{ij} \hat{r}^i \sigma^j \quad (4.7.95)
\]

\[
\hat{r}(0 \leq \theta \leq \pi, 0 < \phi \leq 2\pi) \equiv (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta). \quad (4.7.96)
\]

Now, up to an overall (irrelevant) multiplicative phase \(e^{i\delta}\), the most general unit norm \(\xi^\dagger \xi = 1\) 2-component object \(\xi\) can be parametrized as

\[
\xi(\alpha, \beta) = (e^{i\beta} \sin \alpha, \cos \alpha)^T, \quad (4.7.97)
\]

where \(\alpha, \beta\) are real angles. We are going to extremize the expectation value

\[
E(\alpha, \beta) \equiv \frac{\langle \xi | \hat{r} \cdot \vec{\sigma} | \xi \rangle}{\langle \xi | \xi \rangle} = \xi^\dagger (\hat{r} \cdot \vec{\sigma}) \xi
\]

\[
= \sin(2\alpha) \sin(\theta) \cos(\beta + \phi) - \cos(2\alpha) \cos(\theta). \quad (4.7.99)
\]

Because eq. (4.7.97) is the most general 2-component object, the extremum of \(E\) through the variation of \(\xi\) 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 $\alpha$ and $\beta$, and setting the results to zero yield the following relations.

$$0 = \frac{\partial \mathcal{E}}{\partial \alpha}$$

$$= 2 \cos(2\alpha) \sin(\theta) \cos(\beta + \phi) + 2 \sin(2\alpha) \cos(\theta) \quad (4.7.100)$$

$$0 = \frac{\partial \mathcal{E}}{\partial \beta}$$

$$= - \sin(2\alpha) \sin(\theta) \sin(\beta + \phi). \quad (4.7.101)$$

Suppose $\sin \theta = 0$, then eq. (4.7.101) becomes trivial; while equations (4.7.99) and (4.7.100) become instead

$$\mathcal{E} = -(2 \cos(\alpha)^2 - 1) \cos(\theta) \quad (4.7.102)$$

$$0 = \frac{\partial \mathcal{E}}{\partial \alpha} = \sin(\alpha) \cos(\alpha) \cos(\theta). \quad (4.7.103)$$

If $\sin(2\alpha) = 0 = \sin \theta$, the possible solutions are

$$(\alpha, \cos \alpha, \theta, \cos \theta, \mathcal{E}) = (0, 1, 0, 1, -1), \quad (4.7.104)$$

$$(\alpha, \cos \alpha, \theta, \cos \theta, \mathcal{E}) = (0, 1, \pi, -1, +1), \quad (4.7.105)$$

$$(\alpha, \cos \alpha, \theta, \cos \theta, \mathcal{E}) = (\pi, -1, 0, 1, -1), \quad (4.7.106)$$

$$(\alpha, \cos \alpha, \theta, \cos \theta, \mathcal{E}) = (\pi, -1, \pi, -1, +1). \quad (4.7.107)$$

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 $\beta$ can be anything; while $\alpha = 0, \pm \pi/2, \pm \pi, \pm (3/2)\pi, \cdots = (n/2)\pi$.

$$\mathcal{E}(\alpha = (n/2)\pi, \beta) = -(-)^n \quad (4.7.108)$$

$$\xi(\alpha = n\pi, \beta) = (0, 1)^T \quad (4.7.109)$$
5 Calculus on the Complex Plane

5.1 Differentiation

The derivative of a complex function \( f(z) \) is defined in a similar way as its real counterpart:

\[
f'(z) \equiv \frac{df(z)}{dz} \equiv \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}.
\] (5.1.1)

However, the meaning is more subtle because \( \Delta z \) (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,

\[
f'(z) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x + iy) - f(x + iy)}{\Delta x} = \partial_x f(z) \]
\[
= \lim_{\Delta y \to 0} \frac{f(x + i(y + \Delta y)) - f(x + iy)}{i\Delta y} = \frac{\partial f(z)}{\partial(iy)} = \frac{1}{i} \partial_y f(z),
\] (5.1.2)

where \( x, y, \Delta x \) and \( \Delta y \) 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.

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 of \( z \) if and only if the following (Cauchy-Riemann) equations are satisfied by the real and imaginary parts of \( f \):

\[
\partial_x u = \partial_y v, \quad \partial_y u = -\partial_x v.
\] (5.1.3)

To understand these Cauchy-Riemann conditions, we first consider differentiating along the (real) \( x \) direction,

\[
\frac{df}{dz} = \frac{\partial f}{\partial x} = \partial_x u + i\partial_x v.
\] (5.1.4)

If we instead differentiate along the (imaginary) \( iy \) direction,

\[
\frac{df}{dz} = \frac{1}{i} \frac{\partial f}{\partial y} = \frac{1}{i} \partial_y u + \partial_y v = \partial_y v - i\partial_y u.
\] (5.1.5)

\(^{37}\)Much of the material here on complex analysis is based on Arfken et al’s Mathematical Methods for Physicists, which in turn is very similar in spirit to Brown and Churchill [12].
Since these two results must be the same, we may equate their real and imaginary parts to obtain eq. (5.1.3). (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. (5.1.3) are necessary but not necessarily sufficient ones for analyticity.) Altogether, eq. (5.1.3) is equivalent to:

\[
\left( \frac{\partial}{\partial x} - \frac{1}{i} \frac{\partial}{\partial y} \right) f(x, y) = 0.
\]

(5.1.6)

Conversely, if \(u\) and \(v\) have continuous first derivatives, we may consider an arbitrary variation of the function \(f\), and thus deduce it must be complex differentiable – i.e., analytic – whenever the Cauchy-Riemann relations of eq. (5.1.3) (or, equivalently, eq. (5.1.6)) are satisfied. To perform this analysis, it is useful to recast the Cartesian coordinates \((x, y)\) in terms of the complex coordinate \(z = x + iy\) and its complex conjugate \(\bar{z} = x - iy\) through the relations

\[
x = \frac{z + \bar{z}}{2} \quad \text{and} \quad y = \frac{z - \bar{z}}{2i}.
\]

(5.1.7)

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. (5.1.7),

\[
df(z, \bar{z}) = du + idv = \partial_x f(z, \bar{z}) \frac{dz + d\bar{z}}{2} + \partial_y f(z, \bar{z}) \frac{dz - d\bar{z}}{2i} = \frac{1}{2} \left( \frac{\partial f(z, \bar{z})}{\partial x} + \frac{\partial f(z, \bar{z})}{\partial (iy)} \right)dz + \frac{1}{2} \left( \frac{\partial f(z, \bar{z})}{\partial x} - \frac{\partial f(z, \bar{z})}{\partial (iy)} \right)d\bar{z} = \frac{\partial f(z, \bar{z})}{\partial z}dz + \frac{\partial f(z, \bar{z})}{\partial \bar{z}}d\bar{z}.
\]

(5.1.8)

\[38\]

Using the version of Cauchy-Riemann relations in eq. (5.1.6), the \(d\bar{z}\) term in eq. (5.1.8) is set to zero and we infer

\[
\frac{df}{dz} = \frac{\partial f}{\partial x} = \frac{1}{i} \frac{\partial f}{\partial y} \quad \text{and} \quad \frac{\partial f}{\partial \bar{z}} = 0.
\]

(5.1.9)

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, \(\text{Re} \ z = (z + \bar{z})/2\) is not analytic because it depends on both \(z\) and \(z^*\).

**Polar coordinates**

It is also useful to express the Cauchy-Riemann conditions in polar coordinates \((x, y) = r(\cos \theta, \sin \theta)\). We have

\[
\partial_r = \frac{\partial}{\partial r} \partial_x x + \frac{\partial}{\partial r} \partial_y y = \cos \theta \partial_x x + \sin \theta \partial_y y \quad \text{and} \quad \partial_{\theta} = \frac{\partial}{\partial \theta} \partial_x x + \frac{\partial}{\partial \theta} \partial_y y = -r \sin \theta \partial_x x + r \cos \theta \partial_y y.
\]

(5.1.11)

(5.1.12)

\[38\] 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 \(df\) (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.
By viewing this as a matrix equation \((\partial_r, \partial_\theta)^T = M(\partial_x, \partial_y)^T\), we may multiply \(M^{-1}\) on both sides and obtain the \((\partial_x, \partial_y)\) in terms of the \((\partial_r, \partial_\theta)\).

\[
\begin{align*}
\partial_x &= \cos \theta \partial_r - \frac{\sin \theta}{r} \partial_\theta \\
\partial_y &= \sin \theta \partial_r + \frac{\cos \theta}{r} \partial_\theta.
\end{align*}
\]

(5.1.13)

(5.1.14)

The Cauchy-Riemann conditions in eq. (5.1.3) can now be manipulated by replacing the \(\partial_x\) and \(\partial_y\) with the right hand sides above. Denoting \(c \equiv \cos \theta\) and \(s \equiv \sin \theta\),

\[
\begin{align*}
\left(cs\partial_r - \frac{s^2}{r} \partial_\theta\right) u &= \left(s^2 \partial_r + \frac{cs}{r} \partial_\theta\right) v, \\
\left(sc\partial_r + \frac{c^2}{r} \partial_\theta\right) u &= - \left(c^2 \partial_r - \frac{sc}{r} \partial_\theta\right) v,
\end{align*}
\]

(5.1.15)

(5.1.16)

and

\[
\begin{align*}
\left(c^2 \partial_r - \frac{sc}{r} \partial_\theta\right) u &= \left(sc\partial_r + \frac{c^2}{r} \partial_\theta\right) v, \\
\left(s^2 \partial_r + \frac{sc}{r} \partial_\theta\right) u &= - \left(cs\partial_r - \frac{s^2}{r} \partial_\theta\right) v.
\end{align*}
\]

(5.1.17)

(5.1.18)

(We have multiplied both sides of eq. (5.1.3) with appropriate factors of sines and cosines.) Subtracting the first pair and adding the second pair of equations, we arrive at the polar coordinates version of Cauchy-Riemann:

\[
\begin{align*}
\frac{1}{r} \partial_\theta u &= - \partial_r v, \\
\partial_r u &= \frac{1}{r} \partial_\theta v.
\end{align*}
\]

(5.1.19)

**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 \(z = x + iy\),

\[
f(z) = |z| = \sqrt{x^2 + y^2} = u(x, y) + iv(x, y) \Rightarrow v(x, y) = 0.
\]

(5.1.20)

It’s not hard to see that the Cauchy-Riemann conditions in eq. (5.1.3) 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.

\[
f(x, y) = u(x, y) \Rightarrow \partial_x u = \partial_y v = 0, \quad \partial_y u = -\partial_x v = 0.
\]

(5.1.21)

Similarly, if \(f\) were purely imaginary across the complex \(z\) plane, it is not differentiable unless it is constant.

\[
f(x, y) = iv(x, y) \Rightarrow 0 = \partial_x u = \partial_y v, \quad 0 = -\partial_y u = \partial_x v.
\]

(5.1.22)
Differentiation rules 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

$$\frac{d}{dz} z^\alpha = \alpha z^{\alpha-1}, \quad \alpha \in \mathbb{R},$$

(5.1.23)

and

$$\frac{d \sin(z)}{dz} = \cos z, \quad \frac{d a^z}{dz} = \frac{d e^{z \ln a}}{dz} = a^z \ln a.$$  

(5.1.24)

It is not difficult to check the first derivatives of $z^\alpha$, $\sin(z)$ and $a^z$ are continuous; and the Cauchy-Riemann conditions are satisfied. For instance, $z^\alpha = r^\alpha e^{i\alpha \theta} = r^\alpha \cos(\alpha \theta) + ir^\alpha \sin(\alpha \theta)$ and eq. (5.1.19) can be verified.

$$r^{\alpha-1} \partial_\theta \cos(\alpha \theta) = -\alpha r^{\alpha-1} \sin(\alpha \theta) - \sin(\alpha \theta) \partial_\theta r, r^\alpha = -\alpha r^{\alpha-1} \sin(\alpha \theta),$$  

(5.1.25)

$$\cos(\alpha \theta) \partial_\theta r, r^\alpha = \alpha r^{\alpha-1} \cos(\alpha \theta) = r^{\alpha-1} \partial_\theta \sin(\alpha \theta) = \alpha r^{\alpha-1} \cos(\alpha \theta).$$  

(5.1.26)

(This proof that $z^\alpha$ is analytic fails at $r = 0$; in fact, for $\alpha < 1$, we see that $z^\alpha$ 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,

$$(fg)' = f'g + fg'.$$  

(5.1.27)

because

$$
(fg)' = \lim_{\Delta z \to 0} \frac{f(z + \Delta z)g(z + \Delta z) - f(z)g(z)}{\Delta z}
= \lim_{\Delta z \to 0} \frac{(f(z) + f' \cdot \Delta z)(g(z) + g' \Delta z) - f(z)g(z)}{\Delta z}
= \lim_{\Delta z \to 0} \frac{fg + fg' \Delta z + f'g \Delta z + \mathcal{O}(\Delta z^2) - fg}{\Delta z}
= f'g + fg'.
$$

(5.1.28)

We will have more to say later about carrying over properties of real differentiable functions to their complex counterparts.

Problem 5.1. Use the Cauchy-Riemann conditions to verify that $\ln z$ is an analytic function.

Problem 5.2. Conformal transformations Complex functions can be thought of as a map from one 2D plane to another. In this problem, we will see how they define angle preserving transformations. Consider two paths on a complex plane $z = x + iy$ that intersects at some point $z_0$. Let the angle between the two lines at $z_0$ be $\theta$. Given some complex function $f(z) = u(x, y) + iv(x, y)$, this allows us to map the two lines on the $(x, y)$ plane into two lines on the $(u, v)$ plane. Show that, as long as $df(z)/dz \neq 0$, the angle between these two lines on the $(u, v)$ plane at $f(z_0)$, is still $\theta$. Hint: imagine parametrizing the two lines with $\lambda$, where the first line is $\xi_1(\lambda) = x_1(\lambda) + iy_1(\lambda)$ while the second line is $\xi_2(\lambda) = x_2(\lambda) + iy_2(\lambda)$. Let their intersection point be $\xi_1(\lambda_0) = \xi_2(\lambda_0)$. Now also consider the two lines on the $(u, v)$ plane: $f(\xi_1(\lambda)) = u(\xi_1(\lambda)) + iv(\xi_1(\lambda))$ and $f(\xi_2(\lambda)) = u(\xi_2(\lambda)) + iv(\xi_2(\lambda))$. On the $(x, y)$-plane, consider $\arg([d\xi_1/d\lambda]/[d\xi_2/d\lambda])$; whereas on the $(u, v)$-plane consider $\arg([df(\xi_1)/d\lambda]/[df(\xi_2)/d\lambda])$.  

160
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. (5.1.3) imply that both the real and imaginary parts of a complex function obey Laplace’s equation, namely

\[
(\partial_x^2 + \partial_y^2)u(x, y) = (\partial_x^2 + \partial_y^2)v(x, y) = 0. \tag{5.1.29}
\]

To see this we differentiate eq. (5.1.3) appropriately,

\[
\begin{align*}
\partial_x \partial_y u &= \partial_y^2 v, \\
\partial_x \partial_y u &= -\partial_x^2 v, \\
\partial_x^2 u &= \partial_x \partial_y v, \\
\partial_x^2 u &= \partial_x \partial_y v.
\end{align*} \tag{5.1.30}
\]

We now can equate the right hand sides of the first line; and the left hand sides of the second line. This leads to (5.1.29).

Because of eq. (5.1.29), 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 \( \partial_x^2 v \) are non-zero. In particular, the determinants of the \( 2 \times 2 \) 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,

\[
\begin{align*}
\det \frac{\partial^2 u}{\partial (x, y)^i \partial (x, y)^j} &= \det \begin{bmatrix} \partial_x^2 u & \partial_x \partial_y u \\ \partial_x \partial_y u & \partial_y^2 u \end{bmatrix} \\
&= \partial_x^2 u \partial_y^2 u - (\partial_x \partial_y u)^2 = -(\partial_x^2 u)^2 - (\partial_y^2 v)^2 \leq 0, \tag{5.1.32}
\end{align*}
\]

\[
\begin{align*}
\det \frac{\partial^2 v}{\partial (x, y)^i \partial (x, y)^j} &= \det \begin{bmatrix} \partial_x^2 v & \partial_x \partial_y v \\ \partial_x \partial_y v & \partial_y^2 v \end{bmatrix} \\
&= \partial_x^2 v \partial_y^2 v - (\partial_x \partial_y v)^2 = -(\partial_x^2 v)^2 - (\partial_y^2 u)^2 \leq 0, \tag{5.1.33}
\end{align*}
\]

where both equations (5.1.29) and (5.1.30) were employed.

5.2 Cauchy’s integral theorems, Laurent Series, Analytic Continuation

Complex integration is really a line integral \( \int \vec{\xi} \cdot (dx, dy) \) 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 \( z(\lambda_2) = z_2 \),

\[
\int_C dz f(z) = \int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (dx + idy) (u(x, y) + iv(x, y)) \\
= \int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (udx - vdy) + i \int_{z(\lambda_1 \leq \lambda \leq \lambda_2)} (vdx +udy) \\
= \int_{\lambda_1}^{\lambda_2} d\lambda \left( u \frac{dx(\lambda)}{d\lambda} - v \frac{dy(\lambda)}{d\lambda} \right) + i \int_{\lambda_1}^{\lambda_2} d\lambda \left( v \frac{dx(\lambda)}{d\lambda} + u \frac{dy(\lambda)}{d\lambda} \right). \tag{5.2.1}
\]

The real part of the line integral involves \( \text{Re} \vec{\xi} = (u, -v) \) and its imaginary part \( \text{Im} \vec{\xi} = (v, u) \).
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.

$$\int_{-C} dz f(z) = - \int_C dz f(z) \quad (5.2.2)$$

Remark II  The complex version of the fundamental theorem of calculus has to hold, in that

$$\int_C dz f'(z) = \int_C df = f(\text{“upper” end point of } C) - f(\text{“lower” end point of } C) = \int_{z_1}^{z_2} dz f'(z) = f(z_2) - f(z_1). \quad (5.2.3)$$

Cauchy’s integral theorem  In introducing the contour integral in eq. (5.2.1), 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 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.

$$\oint_C f(z) dz = 0 \quad (5.2.4)$$

Unfortunately the detailed proof will take up too much time and effort, but the mathematically minded can consult, for example, Brown and Churchill’s Complex Variables and Applications.

Problem 5.3.  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 $\oint_C f(z) dz$ is a (complex) line integral around a closed loop. Then apply Stokes’ theorem followed by the Cauchy-Riemann conditions in eq. (5.1.3). Can you fill in the details?  

Important Remarks  Cauchy’s theorem has an important implication. Suppose we have a contour integral $\int_C g(z) dz$, 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

$$\int_C g(z) dz = \int_{C'} g(z) dz. \quad (5.2.5)$$

The reason is that, by subtracting these two integrals, say $(\int_C - \int_{C'}) g(z) dz$, the $-$ sign can be absorbed by reversing the direction of the $C'$ integral. We then have a closed contour integral $(\int_C - \int_{C'}) g(z) dz = \oint g(z) dz$ and Cauchy’s theorem in eq. (5.2.4) applies.

This 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 possible type of singularity is known as a branch point, which will then require us to introduce a branch cut.

---

A simply connected region is one where every closed loop in it can be shrunk to a point.
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

\[
\oint_{C_1} f(z)\,dz = \oint_{C_2} f(z)\,dz. \tag{5.2.6}
\]

**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 \( z = 0 \).

\[
I \equiv \oint_C \frac{dz}{z} \tag{5.2.7}
\]

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

\[
\oint_C \frac{dz}{z} = \oint_{C'} \frac{dz}{z}. \tag{5.2.8}
\]

\[\text{We may then employ polar coordinates, so that the path } C' \text{ could be described as } z = Re^{i\theta}, \text{ where } \theta \text{ would run from } 0 \text{ to } 2\pi.\]

\[
\oint_C \frac{dz}{z} = \int_0^{2\pi} \frac{i\,d\theta}{Re^{i\theta}} = \int_0^{2\pi} i\,d\theta = 2\pi i. \tag{5.2.9}
\]

**Example II** Let’s evaluate \( \oint_C zdz \) and \( \oint_C dz \) directly and by using Cauchy’s integral theorem. Here, \( C \) is some closed contour on the complex plane. Directly:

\[
\oint_C zdz = \left. \frac{z^2}{2} \right|_{z=z_0}^{z=z_0} = 0, \quad \oint_C dz = \left. z \right|_{z=z_0}^{z=z_0} = 0. \tag{5.2.10}
\]

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.

**Problem 5.4.** For some contour \( C \), let \( M \) be the maximum of \( |f(z)| \) along it and \( L \equiv \int_C \sqrt{dx^2 + dy^2} \) be the length of the contour itself, where \( z = x + iy \) (for \( x \) and \( y \) real). Argue that

\[
\left| \oint_C f(z)\,dz \right| \leq \int_C |f(z)||dz| \leq M \cdot L. \tag{5.2.11}
\]

Note: \( |dz| = \sqrt{dx^2 + dy^2} \). (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.

\[\text{This is where drawing a picture would help: for simplicity, if } C' \text{ lies entirely within } C, \text{ the first portion of the cut lines would begin anywhere from } C' \text{ to anywhere to } C, \text{ followed by the reverse trajectory from } C \text{ to } C' \text{ that runs infinitesimally close to the first portion. Because they are infinitesimally close, the contributions of these two portions cancel; but we now have a simply connected closed contour integral that amounts to } 0 = (\int_C - \int_{C'})dz/z.\]
Problem 5.5. Evaluate
\[ \oint_C \frac{dz}{z(z + 1)}, \]  \hspace{1cm} (5.2.12)
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
\[ \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{z' - z} = \begin{cases} f(z) & \text{if } z \text{ lies inside } C \\ 0 & \text{if } z \text{ lies outside } C. \end{cases} \]  \hspace{1cm} (5.2.13)

**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 \( z' \approx z \),
\[ z' \equiv z + \epsilon e^{i\theta}, \quad 0 < \epsilon \ll 1. \]  \hspace{1cm} (5.2.14)
We then have
\[ \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{z' - z} = \frac{1}{2\pi i} \int_0^{2\pi} \frac{\epsilon e^{i\theta} i d\theta}{\epsilon e^{i\theta}} f(z + \epsilon e^{i\theta}) \]
\[ = \int_0^{2\pi} \frac{d\theta}{2\pi} f(z + \epsilon e^{i\theta}). \]  \hspace{1cm} (5.2.15)
By taking the limit \( \epsilon \to 0^+ \), we get \( f(z) \), since \( f(z') \) is analytic and thus continuous at \( z' = z \).

**Cauchy’s integral formula for derivatives** By applying the limit definition of the derivative, we may obtain an analogous definition for the \( n \)th derivative of \( f(z) \). For some closed counterclockwise contour \( C \),
\[ \oint_C \frac{dz'}{2\pi i} \frac{f(z')}{(z' - z)^{n+1}} = \begin{cases} \frac{f^{(n)}(z)}{n!} & \text{if } z \text{ lies inside } C \\ 0 & \text{if } z \text{ lies outside } C. \end{cases} \]  \hspace{1cm} (5.2.16)
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 \( \oint_C f(z) dz = 0 \) 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. (5.2.16)), \( f(z) \) 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 blow up somewhere on the complex plane.

**Proof**  
To prove this result we employ eq. (5.2.16). Choose a counterclockwise circular contour \( C \) that encloses some arbitrary point \( z \),

\[
|f^{(n)}(z)| \leq n! \oint_C \frac{|dz'|}{2\pi} \frac{|f(z')|}{|(z'-z)^{n+1}|} \leq n! \frac{M}{2\pi r^{n+1}} \oint_C |dz'| = n! \frac{M}{r^n}.
\]

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 \( n \)th 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 \( \text{Re} \, z \to \infty \). 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 \( R \in \mathbb{R} \),

\[
\sin(iR) = \frac{e^{-R} - e^R}{2i}, \quad \cos(iR) = \frac{e^{-R} + e^R}{2}.
\]

Both \( \sin(iR) \) and \( \cos(iR) \) blow up as \( R \to \pm\infty \).

**Problem 5.6. Fundamental theorem of algebra.**  
Let \( P(z) = p_0 + p_1 z + \ldots + p_n z^n \) be an \( n \)th degree polynomial, where \( n \) is an integer greater or equal to 1. By considering \( f(z) = 1/P(z) \), 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 \( n \)th 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 5.7.**  
By starting with the \( N \)th partial sum,

\[
S_N \equiv \sum_{\ell=0}^{N} t^\ell, \quad (5.2.20)
\]
prove that, as long as $|t| < 1$,

$$\frac{1}{1-t} = \sum_{\ell=0}^{\infty} t^{\ell}. \quad (5.2.21)$$

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. (5.2.13), and deduce a series expansion about $z_0$:

$$f(z) = \int\limits_{C} \frac{dz'}{2\pi i} \frac{f(z')}{z' - z} = \int\limits_{C} \frac{dz'}{2\pi i} \frac{f(z')}{(z' - z_0)(1 - (z - z_0)/(z' - z_0))}$$

$$= \sum_{\ell=0}^{\infty} \int\limits_{C} \frac{dz'}{2\pi i} \frac{f(z')}{(z' - z_0)^{\ell+1}} (z - z_0)^{\ell}. \quad (5.2.22)$$

We have used the geometric series in eq. (5.2.21) 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 $n$th derivative of an analytic function, eq. (5.2.16), we have arrived at its Taylor series.

**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,

$$f(z) = \sum_{\ell=0}^{\infty} (z - z_0)^{\ell} \frac{f^{(\ell)}(z_0)}{\ell!}, \quad (5.2.23)$$

where $f^{(\ell)}(z)/\ell!$ is given by eq. (5.2.16).

**Problem 5.8.** Complex binomial theorem. For $p$ any real number and $z$ any complex number obeying $|z| < 1$, prove the complex binomial theorem using eq. (5.2.23),

$$(1 + z)^p = \sum_{\ell=0}^{\infty} \binom{p}{\ell} z^{\ell}, \quad \binom{p}{0} \equiv 1, \quad \binom{p}{\ell} = \frac{p(p-1)\ldots(p-(\ell-1))}{\ell!}. \quad (5.2.24)$$

**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

$$f(z) = \int_{C_1} + \int_{C_2} \frac{dz'}{2\pi i} \frac{f(z')}{z' - z}. \quad (5.2.25)$$
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. (5.2.21) 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

$$f(z) = \sum_{\ell=0}^{\infty} (z - z_0)^{\ell} \int_{C_2} \frac{dz'}{2\pi i} \frac{f(z')}{(z' - z_0)^{\ell+1}} - \sum_{\ell=0}^{\infty} \frac{1}{(z - z_0)^{\ell+1}} \int_{C_1} \frac{dz'}{2\pi i} (z' - z_0)^{\ell} f(z').$$  \hspace{1cm} (5.2.26)

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 $C_2$,

$$\int_{C'_1} \frac{dz'}{2\pi i} (z' - z_0)^{\ell} f(z') = \int_{C_2} \frac{dz'}{2\pi i} (z' - z_0)^{\ell} f(z'),$$  \hspace{1cm} (5.2.27)

because for positive $\ell$ 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

$$f(z) = \sum_{\ell=0}^{\infty} \int_{C_2} \frac{dz'}{2\pi i} \left( (z - z_0)^{\ell} \frac{f(z')}{(z' - z_0)^{\ell+1}} + \frac{1}{(z - z_0)^{\ell+1}} (z' - z_0)^{\ell} f(z') \right).$$  \hspace{1cm} (5.2.28)

Proceeding to re-label the second series by replacing $\ell + 1 \rightarrow -\ell'$, so that the summation then runs from $-1$ through $-\infty$, the Laurent series emerges.

**Laurent series** 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

$$f(z) = \sum_{\ell=-\infty}^{\infty} L_\ell(z_0) \cdot (z - z_0)^\ell,$$  \hspace{1cm} (5.2.29)

$$L_\ell(z_0) \equiv \int_C \frac{dz'}{2\pi i} \frac{f(z')}{(z' - z_0)^{\ell+1}}.$$  \hspace{1cm} (5.2.30)

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. (5.2.29) to obtain the expansion coefficients $L_\ell$. 

167
Problem 5.9. For complex $z$, $a$ and $b$, obtain the Laurent expansion of
\[ f(z) \equiv \frac{1}{(z-a)(z-b)}, \quad a \neq b, \]  
(5.2.31)
about $z = a$, in the region $0 < |z - a| < |a - b|$ using eq. (5.2.29). Check your result either by writing
\[ \frac{1}{z - b} = -\frac{1}{1 - (z - a)/(b-a)} \cdot \frac{1}{b - a}, \]  
(5.2.32)
and employing the geometric series in eq. (5.2.21), or directly performing a Taylor expansion of $1/(z-b)$ about $z = a$.

Problem 5.10. 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
\[ (f(z))^* = u(x, y) - iv(x, y) = f(z^*) = u(x, -y) + iv(x, -y). \]  
(5.2.33)
Comment on why this is called the “reflection principle”.

We now turn to an important result that allows us to extend the definitions of complex differentiable functions beyond their original range of validity.

Analytic continuation An analytic function $f(z)$ is fixed uniquely throughout a given region $\Sigma$ on the complex plane, once its value is specified on a line segment lying within $\Sigma$.

This in turn means, suppose we have an analytic function $f_1(z)$ defined in a region $\Sigma_1$ on the complex plane, and suppose we found another analytic function $f_2(z)$ defined in some region $\Sigma_2$ such that $f_2(z)$ agrees with $f_1(z)$ in their common region of intersection. (It is important that $\Sigma_2$ does have some overlap with $\Sigma_1$.) 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 $\Sigma_1$ and $\Sigma_2$, yet behave different in the rest of $\Sigma_2$.

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
\[ f_1(z) \equiv \sum_{\ell=0}^{\infty} z^{\ell}, \quad |z| < 1, \]  
(5.2.34)
and
\[ f_2(z) \equiv \frac{1}{1 - z}, \]  
(5.2.35)
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 $|z| < 1$.

Problem 5.11. 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

$$\, _2F_1(\alpha, \beta; \gamma; z) = \sum_{\ell=0}^{\infty} C_\ell(\alpha, \beta; \gamma) \frac{z^\ell}{\Gamma(\ell+1)},$$

$$C_0(\alpha, \beta; \gamma) \equiv 1,$$

$$C_{\ell \geq 1}(\alpha, \beta; \gamma) \equiv \frac{\alpha(\alpha + 1) \ldots (\alpha + (\ell - 1)) \cdot \beta(\beta + 1) \ldots (\beta + (\ell - 1))}{\gamma(\gamma + 1) \ldots (\gamma + (\ell - 1))}. \tag{5.2.36}$$

On the other hand, it also has the following integral representation,

$$\, _2F_1(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\gamma - \beta) \Gamma(\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}dt, \quad \text{Re}(\gamma) > \text{Re}(\beta) > 0. \tag{5.2.37}$$

(Here, $\Gamma(z)$ is known as the Gamma function; see http://dlmf.nist.gov/5.) Show that eq. (5.2.37) does in fact agree with eq. (5.2.36) for $|z| < 1$. You can apply the binomial expansion in eq. (5.2.24) to $(1 - tz)^{-\alpha}$, followed by result

$$\int_0^1 dt(1 - t)^{\alpha-1}t^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad \text{Re}(\alpha), \text{Re}(\beta) > 0. \tag{5.2.39}$$

You may also need the property

$$z\Gamma(z) = \Gamma(z + 1). \tag{5.2.40}$$

Therefore eq. (5.2.37) extends eq. (5.2.36) into the region $|z| > 1$.

5.3 Poles and Residues

In this section we will consider the closed counterclockwise contour integral

$$\oint_C \frac{dz}{2\pi i} f(z), \tag{5.3.1}$$
where \( f(z) \) is analytic everywhere on and within \( C \) except at isolated singular points of \( f(z) \) — which we will denote as \( \{ z_1, \ldots, z_n \} \), 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,

\[
f(z) = \sum_{\ell=-m}^{\infty} L_\ell \cdot (z - z_0)^\ell, \tag{5.3.2}
\]

we say the function has a pole of order \( m \) at \( z = z_0 \). 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 5.12.** If \( n \) is an arbitrary integer, show that

\[
\oint_C (z' - z)^n \frac{dz'}{2\pi i} = 1, \quad \text{when } n = -1,
\]

\[
= 0, \quad \text{when } n \neq -1, \tag{5.3.3}
\]

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 \( \{C'_i|i = 1, 2, \ldots, n\} \) around each and every isolated point. This means

\[
\oint_C f(z') \frac{dz'}{2\pi i} = \sum_i \oint_{C'_i} f(z') \frac{dz'}{2\pi i}. \tag{5.3.4}
\]

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 \( C'_i \rightarrow C'_{i+1} \) and the other \( C'_{i+1} \rightarrow C'_i \). 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 \( i \)th singular point, and then proceed to integrate the series term-by-term using eq. (5.3.3).

\[
\oint_{C'_i} f(z') \frac{dz'}{2\pi i} = \int_{C'_i} \sum_{\ell=-m_i}^{\infty} L_\ell^{(i)} \cdot (z' - z_i)\ell \frac{dz'}{2\pi i} = L_{-1}^{(i)}. \tag{5.3.5}
\]

**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 \( \{z_i\} \), yields the sum of the residues at each of these points. In equation form,

\[
\oint_C f(z') \frac{dz'}{2\pi i} = \sum_i L_{-1}^{(i)}, \tag{5.3.6}
\]

where \( L_{-1}^{(i)} \) is the residue at the \( i \)th singular point \( z_i \).
Let us start with a simple application of this result. Let $C$ be some closed counterclockwise contour containing the points $z = 0, a, b$.

$$ I = \oint_C \frac{dz}{2\pi i \ z(z-a)(z-b)}. \tag{5.3.7} $$

One way to do this is to perform a partial fractions expansion first.

$$ I = \oint_C \frac{dz}{2\pi i \left( \frac{1}{ab} + \frac{1}{a(a-b)(z-a)} + \frac{1}{b(b-a)(z-b)} \right)}. \tag{5.3.8} $$

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

$$ I = \frac{1}{ab} + \frac{1}{a(a-b)} + \frac{1}{b(b-a)} = \frac{(a-b)+b-a}{ab(a-b)} = 0. \tag{5.3.9} $$

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.,

$$ \frac{1}{z(z-a)(z-b)} = \frac{1}{z} \left( \frac{1}{ab} + \mathcal{O}(z) \right). \tag{5.3.10} $$

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.

$$ \oint_C \frac{dz}{2\pi i \ z(z-a)(z-b)} = \sum_{z_i=0,a,b} \left( \text{Residue of } \frac{1}{z(z-a)(z-b)} \text{ at } z_i \right) \nonumber \\
= \frac{1}{ab} + \frac{1}{a(a-b)} + \frac{1}{b(b-a)} = 0. \tag{5.3.11} $$

The reason why the result is zero can actually be understood via contour integration as well. If you now consider a closed clockwise contour $C_\infty$ at infinity and view the integral $(\int_C + \int_{C_\infty}) f(z)dz$, you will be able to convert it into a closed contour integral by linking $C$ and $C_\infty$ via two infinitesimally close radial lines which would not actually contribute to the answer. But $(\int_C + \int_{C_\infty}) f(z)dz = \int_{C_\infty} f(z)dz$ because $C_\infty$ does not contribute either – why? Therefore, since there are no poles in the region enclosed by $C_\infty$ 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

$$ I \equiv \oint_C \exp(1/z^2)dz. \tag{5.3.12} $$
We Taylor expand the exp, and notice there is no term that goes as $1/z$. Hence,

$$I = \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \oint_C \frac{dz}{z^{2\ell}} = 0. \quad (5.3.13)$$

A major application of contour integration is to that of integrals involving real variables.

**Application I: Trigonometric integrals** If we have an integral of the form

$$\int_0^{2\pi} d\theta f(\cos \theta, \sin \theta), \quad (5.3.14)$$

then it may help to change from $\theta$ to

$$z \equiv e^{i\theta} \quad \Rightarrow \quad dz = id\theta \cdot e^{i\theta} = id\theta \cdot z, \quad (5.3.15)$$

and

$$\sin \theta = \frac{z - 1/z}{2i}, \quad \cos \theta = \frac{z + 1/z}{2}. \quad (5.3.16)$$

The integral is converted into a sum over residues:

$$\int_0^{2\pi} d\theta f(\cos \theta, \sin \theta) = 2\pi \sum_j \left( j\text{th residue of } \frac{f\left(\frac{z+1/z}{2}, \frac{z-1/z}{2i}\right)}{z} \text{ for } |z| < 1 \right). \quad (5.3.17)$$

**Example** For $a \in \mathbb{R}$,

$$I = \int_0^{2\pi} \frac{d\theta}{a + \cos \theta} = \oint_{|z|=1} \frac{dz}{iz} \frac{1}{a + (1/2)(z + 1/z)} = \oint_{|z|=1} \frac{dz}{i} \frac{1}{az + (1/2)(z^2 + 1)}$$

$$= 4\pi \oint_{|z|=1} \frac{dz}{2\pi i} \frac{1}{(z - z_+)(z - z_-)}, \quad z_+ \equiv -a + \sqrt{a^2 - 1}. \quad (5.3.18)$$

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 $-a + \sqrt{a^2 - 1}$ runs from $-\infty$ to 0 because $a > \sqrt{a^2 - 1}$. When $-a$ runs from 1 to $\infty$, on the other hand, $-a - \sqrt{a^2 - 1}$ runs from 1 to 0; while $-a + \sqrt{a^2 - 1}$ runs from 1 to $\infty$. In other words, for $a > 1$, $z_+ = -a + \sqrt{a^2 - 1}$ lies within the unit circle and the relevant residue is $1/(z_+ - z_-) = 1/(\sqrt{a^2 - 1}) = \text{sgn}(a)/(2\sqrt{a^2 - 1})$. For $a < -1$ it is $z_- = -a - \sqrt{a^2 - 1}$ that lies within the unit circle and the relevant residue is $1/(z_- - z_+) = -1/(2\sqrt{a^2 - 1}) = \text{sgn}(a)/(2\sqrt{a^2 - 1})$. Therefore,

$$\int_0^{2\pi} \frac{d\theta}{a + \cos \theta} = \frac{2\pi \text{sgn}(a)}{\sqrt{a^2 - 1}}, \quad a \in \mathbb{R}, \quad |a| > 1. \quad (5.3.19)$$
Application II: Integrals along the real line If you need to do \( \int_{-\infty}^{+\infty} f(z)dz \), 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
\[
I \equiv \int_{-\infty}^{\infty} \frac{dz}{z^2 + z + 1}. \tag{5.3.20}
\]
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).
\[
\lim_{\rho \to \infty} \left| \frac{id\theta \cdot \rho e^{i\theta}}{\rho^2 e^{i2\theta} + \rho e^{i\theta} + 1} \right| \to \lim_{\rho \to \infty} \frac{d\theta}{\rho} = 0. \tag{5.3.21}
\]
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
\[
I = 2\pi i \oint_C \frac{dz}{2\pi i (z - e^{-i2\pi/3})(z - e^{i2\pi/3})}. \tag{5.3.22}
\]
Closing the contour in the upper half plane yields a counterclockwise path, which yields
\[
I = \frac{2\pi i}{e^{i2\pi/3} - e^{-i2\pi/3}} = \frac{\pi}{\sin(2\pi/3)}. \tag{5.3.23}
\]
Closing the contour in the lower half plane yields a clockwise path, which yields
\[
I = \frac{-2\pi i}{e^{-i2\pi/3} - e^{i2\pi/3}} = \frac{\pi}{\sin(2\pi/3)}. \tag{5.3.24}
\]
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.
\[
f(t) = \int_{-\infty}^{\infty} \tilde{f}(\omega) e^{i\omega t} d\omega \frac{1}{2\pi}, \quad t \in \mathbb{R}. \tag{5.3.25}
\]
We will assume \( t \) is real and \( \tilde{f} \) has only isolated singularities. Let \( C \) be a large semi-circular path, either in the upper or lower complex plane; consider the following integral along \( C \).
\[
I' \equiv \oint_C \tilde{f}(\omega) e^{i\omega t} d\omega \frac{1}{2\pi} = \lim_{\rho \to \infty} \oint \tilde{f}(\rho e^{i\theta}) e^{i\rho(cos \theta) t} e^{-\rho(sin \theta) t} \frac{id\theta \cdot \rho e^{i\theta}}{2\pi} \tag{5.3.26}
\]
\(^{41}\)In physical applications \( \tilde{f} \) may have branch cuts; this will be dealt with in the next section.
At this point we see that, for $t < 0$, unless $\tilde{f}$ goes to zero much faster than the $e^{-\rho(\sin \theta)t}$ for large $\rho$, 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 $\rho$, 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 $|\tilde{f}| \leq M$ on the semi-circle and consider the magnitude of this integral,

$$|I'| \leq \lim_{\rho \to \infty} \left( \rho M \int e^{-\rho(\sin \theta)t} \frac{d\theta}{2\pi} \right), \quad (5.3.27)$$

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

$$|I'| \leq \lim_{\rho \to \infty} \left( 2\rho M \int_0^{\pi/2} e^{-\rho(\sin \theta)|t|} \frac{d\theta}{2\pi} \right), \quad (5.3.28)$$

because

$$\int_0^\pi F(\sin(\theta))d\theta = 2 \int_0^{\pi/2} F(\sin(\theta))d\theta \quad (5.3.29)$$

for any function $F$. The next observation is that, over the range $\theta \in [0, \pi/2],

$$\frac{2\theta}{\pi} \leq \sin \theta, \quad (5.3.30)$$

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

$$|I'| \leq \lim_{\rho \to \infty} \left( 2\rho M \int_0^{\pi/2} e^{-2\rho|t|/\pi} \frac{d\theta}{2\pi} \right) \quad (5.3.31)$$

$$= \lim_{\rho \to \infty} \left( \rho M \frac{e^{-\rho\pi|t|/\pi} - 1}{-2\rho|t|} \right) = \frac{1}{2|t|} \lim_{\rho \to \infty} M \quad (5.3.32)$$

As long as $|\tilde{f}(\omega)|$ 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 $\tilde{f}(\omega)$ 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. (5.3.25),

- If $t > 0$ and if $|\tilde{f}(\omega)|$ 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 $f(t) = \int_{-\infty}^{\infty} \tilde{f}(\omega)e^{i\omega t} \frac{d\omega}{2\pi}$ to $i$ times the sum of the residues of $\tilde{f}(\omega)e^{i\omega t}$ for $\text{Im}(\omega) > 0$ – provided the function $\tilde{f}(\omega)$ is analytic except at isolated points there.
• If \( t < 0 \) and if \(|\tilde{f}(\omega)|\) 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 \( f(t) = \int_{-\infty}^{\infty} \tilde{f}(\omega) e^{i\omega t} \frac{d\omega}{2\pi} \) to \(-i\) times the sum of the residues of \( \tilde{f}(\omega)e^{i\omega t} \) for \( \text{Im}(\omega) < 0 \) provided the function \( \tilde{f}(\omega) \) is analytic except at isolated points there.

• A quick guide to how to close the contour is to evaluate the exponential on the imaginary \( \omega \) 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 \( \tilde{f}(\omega) \) 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

\[
I(t) \equiv \int_{-\infty}^{+\infty} d\omega \frac{e^{i\omega t}}{2\pi (\omega + i)^2(\omega - 2i)}, \quad t \in \mathbb{R}.
\] (5.3.33)

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. (5.2.13).

\[
I(t > 0) = i \frac{e^{i(2i)t}}{(2i + i)^2} = -i e^{-2t} = \frac{e^{-2t}}{9}.
\] (5.3.34)

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. (5.2.16),

\[
I(t < 0) = i \int_{C} d\omega \frac{e^{i\omega t}}{2\pi i (\omega + i)^2(\omega - 2i)} = -i \frac{d}{d\omega} \left( \frac{e^{i\omega t}}{\omega - 2i} \right)_{\omega = -i}
= -ie^t \frac{1 - 3t}{9}
\] (5.3.35)

To summarize,

\[
\int_{-\infty}^{+\infty} d\omega \frac{e^{i\omega t}}{2\pi (\omega + i)^2(\omega - 2i)} = -i \frac{e^{-2t}}{9} \Theta(t) - ie^t \frac{1 - 3t}{9} \Theta(-t),
\] (5.3.36)

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 \( \frac{d}{dt} \) amounts to a \( i\omega \). Therefore, we can check the following differential equations obeyed by \( I(t) \):

\[
\left( \frac{1}{i} \frac{d}{dt} + i \right)^2 \left( \frac{1}{i} \frac{d}{dt} - 2i \right) I(t) = \delta(t),
\]

(5.3.37)

\[
\left( \frac{1}{i} \frac{d}{dt} + i \right)^2 I(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega t}}{\omega - 2i} = i\Theta(t)e^{-2t},
\]

(5.3.38)

\[
\left( \frac{1}{i} \frac{d}{dt} - 2i \right) I(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega t}}{\omega + 2i} = -i\Theta(-t)ite^t = \Theta(-t)te^t.
\]

(5.3.39)

**Problem 5.13.** Evaluate

\[
\int_{-\infty}^{\infty} \frac{dz}{z^3 + i}.
\]

(5.3.40)

**Problem 5.14.** Show that the integral representation of the step function \( \Theta(t) \) is

\[
\Theta(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi i} \frac{e^{i\omega t}}{\omega - i0^+}.
\]

(5.3.41)

The \( \omega - i0^+ \) 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.

Next, let \( a \) and \( b \) be non-zero real numbers. Evaluate

\[
I(a, b) \equiv \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi i} \frac{e^{i\omega a}}{\omega + ib}.
\]

(5.3.42)

**Problem 5.15.** (From Arfken et al.) Sometimes this “closing-the-contour” trick need not involve closing the contour at infinity. Show by contour integration that

\[
I \equiv \int_{0}^{\infty} \frac{(\ln x)^2}{1 + x^2} \, dx = \frac{\pi^3}{8}.
\]

(5.3.43)

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 5.16.** Evaluate

\[
I(a) \equiv \int_{-\infty}^{\infty} \frac{\sin(ax)}{x} \, dx, \quad a \in \mathbb{R}.
\]

(5.3.44)

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 5.17. Evaluate

\[ I(t) = \int_{-\infty}^{+\infty} \frac{e^{-i\omega t}}{2\pi} \frac{e^{-i(\omega + ib)t}}{(\omega - ia)^2(\omega + ib)^2}, \quad t \in \mathbb{R}; \quad a, b > 0. \]  

(5.3.45)

Application III: Counting zeros of analytic functions

Within a simply connected domain \( \mathcal{D} \) on the complex plane, such that \( C \) denotes the counterclockwise path along its boundary, let us show that the following integral

\[ N = \frac{1}{2\pi i} \oint_C \frac{f'(z)}{f(z)} \, dz \]  

(5.3.46)

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

\[ f(z) = c_n (z - z_0)^n + c_{n+1} (z - z_0)^{n+1} + \ldots \]

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.

5.4 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,

\[ f \left( z_0 + \epsilon \cdot e^{i\theta} \right) \neq f \left( z_0 + \epsilon \cdot e^{i(\theta + 2\pi)} \right), \quad 0 < \epsilon \ll 1. \]  

(5.4.1)

Example I

One example is the power \( z^\alpha \), for \( \alpha \) non-integer. Zero is a branch point because, for \( 0 < \epsilon \ll 1 \), we may considering circling it \( n \in \mathbb{Z}^+ \) times.

\[ (\epsilon e^{2\pi ni})^\alpha = e^{\alpha 2\pi ni} \neq e^\alpha. \]  

(5.4.2)

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 \( \theta \) runs beyond \( 2\pi \), 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 \( \theta \) 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 \( n \)th sheet our function reads \( z = |z|^{1/m} e^{i(2\pi n/m) + i\theta} \). It is the \( m \)th sheet that needs to be
joined with the very first sheet, because by the \( m \)th sheet we have covered all the \( m \) solutions of what we mean by taking the \( m \)th 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 \( m \)th sheet, then to the \((m - 1)\)th sheet and so on.) Finally, if \( \alpha \) were not rational – it is not the ratio of two integers – we would need an infinite number of Riemann sheets to fully describe \( z^\alpha \) 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 \( n \)th 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 \( n \)th 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

\[
(e^{\ln z})' = e^{\ln z} \cdot (\ln z)' = z \cdot (\ln z)' = 1. \tag{5.4.3}
\]

This implies,

\[
\frac{d \ln z}{dz} = \frac{1}{z}, \quad z \neq 0, \tag{5.4.4}
\]

which in turn says \( \ln z \) is analytic away from the origin. We may now consider making \( m \) infinitesimal circular trips around \( z = 0 \).

\[
\ln(\epsilon e^{i2\pi m}) = \ln(\epsilon e^{i2\pi m}) = \ln \epsilon + i2\pi m \neq \ln \epsilon. \tag{5.4.5}
\]

Just as for \( f(z) = z^\alpha \) when \( \alpha \) 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 \( m \)th Riemann sheet to a horizontal band on the complex plane, lying between \( 2\pi(m - 1) \leq \text{Im}(\ln(z)) \leq 2\pi m \).

**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. (5.2.29). 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: \( f(z = x + \Im 0^+) = x^\alpha \neq f(z = x - \Im 0^+) = x^\alpha e^{i2\pi \alpha} \), for instance. Therefore the derivation there would not go through, and a Laurent series for either \( z^\alpha \) 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 \( z^\alpha \) and \( \ln z \) are analytic once all relevant Riemann sheets have been taken into
account. For example, let us do \( \oint_C \ln(z) \, dz \), 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 \( z \approx 0 \), \( \ln(z) \) is analytic. We may therefore invoke Cauchy’s theorem in eq. (5.2.4) to deduce the result depends on the path only through its ‘winding number’ \( n \). Because \( (z \ln(z) - z)' = \ln z \),

\[
\int_{z_1}^{z_2} \ln(z) \, dz = r_2 e^{i\theta_2} \left( \ln r_2 + i(\theta_2 + 2\pi n) - 1 \right) - r_1 e^{i\theta_1} \left( \ln r_1 + i\theta_1 - 1 \right). \tag{5.4.6}
\]

Likewise, for the same integration contour \( C \),

\[
\int_{z_1}^{z_2} z^\alpha \, dz = \frac{r_2^{\alpha+1}}{\alpha+1} e^{i(\alpha+1)(\theta_2 + 2\pi n)} - \frac{r_1^{\alpha+1}}{\alpha+1} e^{i(\alpha+1)\theta_1}. \tag{5.4.7}
\]

**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 \( \sqrt{z^2 - 1} \):

\[
\sqrt{z + 1} \sqrt{z - 1} = \sqrt{r_1 r_2} e^{i(\theta_1 + \theta_2)/2}, \tag{5.4.8}
\]

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 \((\theta_1 + \theta_2)/2 = 2\pi \frac{\theta_1}{\theta_2} \) 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

\[
Q_0(z) = \ln \left[ \frac{z + 1}{z - 1} \right]. \tag{5.4.9}
\]

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

\[
z + 1 \equiv r_1 e^{i\theta_1} \quad \text{and} \quad z - 1 \equiv r_2 e^{i\theta_2} \tag{5.4.10}
\]
as before. Then,

\[
Q_0(z) = \ln \left[ \frac{z + 1}{z - 1} \right] = \ln \frac{r_1}{r_2} + i(\theta_1 - \theta_2). \tag{5.4.11}
\]

\[42\text{Arfken et al. goes through various points along this circling-the-}(z = \pm 1)\text{process, but the main point is that there is no jump after a complete circle, unlike what you’d get circling the branch point of, say } z^{1/3}. \text{ On the other hand, you may want to use the } z + 1 \equiv r_1 e^{i\theta_1} \text{ and } z - 1 \equiv r_2 e^{i\theta_2} \text{ parametrization here and understand how many Riemann sheets it would take define the whole } \sqrt{z^2 - 1}.\]
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

\[
Q_0(x) = \frac{1}{2}Q_0(x + i0^+) + \frac{1}{2}Q_0(x - i0^+),
\]

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

\[
Q_0(x) = \ln \left[ \frac{1 + x}{1 - x} \right], \quad -1 < x < +1.
\]

because the imaginary parts cancel while \(|z + 1| = x + 1 \) and \(|z - 1| = 1 - x \) in this region.

**Example**
Let us exploit the following branch of natural log

\[
\ln z = \ln r + i\theta, \quad z = re^{i\theta}, \quad 0 \leq \theta < 2\pi
\]

to evaluate the integral encountered in eq. (5.3.43).

\[
I \equiv \int_0^\infty \frac{(\ln x)^2}{1 + x^2} dx = \frac{\pi^3}{8}.
\]

To begin we will actually consider

\[
I' \equiv \lim_{\epsilon \to 0} \int_{C_1 + C_2 + C_3 + C_4} \frac{(\ln z)^2}{1 + z^2} dz,
\]

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.

\[
\left| \lim_{\epsilon \to 0} \int_{C_2} \frac{(\ln z)^2}{1 + z^2} dz \right| = \left| \lim_{\epsilon \to 0} \int_0^\pi i\theta e^{i\theta} \frac{(\ln \epsilon + i\theta)^2}{1 + \epsilon^2 e^{2i\theta}} \right| \\
\leq \lim_{\epsilon \to 0} \int_0^\pi d\theta |\ln \epsilon + i\theta|^2 = 0.
\]

and

\[
\left| \lim_{R \to \infty} \int_{C_4} \frac{(\ln z)^2}{1 + z^2} dz \right| = \left| \lim_{R \to \infty} \int_0^\pi i\theta Re^{i\theta} \frac{(\ln R + i\theta)^2}{1 + R^2 e^{2i\theta}} \right| \\
\leq \lim_{R \to \infty} \int_0^\pi d\theta |\ln R + i\theta|^2 / R = 0.
\]
Moreover, $I'$ can be evaluated via the residue theorem; within the closed contour, the integrand blows up at $z = i$.

\[
I' = 2\pi i \lim_{R \to \infty} \oint_{C_1+C_2+C_3+C_4} \frac{(\ln z)^2}{(z+i)(z-i)} \frac{dz}{2\pi i} = 2\pi i \frac{(\ln i)^2}{2i} = \pi(\ln(1) + i(\pi/2))^2 = -\frac{\pi^3}{4}.
\]

(5.4.19)

This means the sum of the integral along $C_1$ and $C_3$ yields $-\frac{\pi^3}{4}$. If we use polar coordinates along both $C_1$ and $C_2$, namely $z = re^{i\theta}$,

\[
\int_0^\infty d\theta e^{i\theta} \frac{(\ln r + i\pi)^2}{1 + r^2 e^{i2\pi}} + \int_0^\infty \frac{(\ln r)^2}{1 + r^2} d\theta = -\frac{\pi^3}{4} \quad (5.4.20)
\]

\[
\int_0^\infty d\theta \frac{2(\ln r)^2 + i2\pi \ln r - \pi^2}{1 + r^2} = -\frac{\pi^3}{4} \quad (5.4.21)
\]

We may equate the real and imaginary parts of both sides. The imaginary one, in particular, says

\[
\int_0^\infty d\theta \frac{\ln r}{1 + r^2} = 0, \quad (5.4.22)
\]

while the real part now hands us

\[
2I = \pi^2 \int_0^\infty \frac{dr}{1 + r^2} - \frac{\pi^3}{4}
\]

\[
= \pi^2 [\arctan(r)]_{r=0}^{r=\infty} - \frac{\pi^3}{4} = \frac{\pi^3(2 - 1)}{4} = \frac{\pi^3}{4} \quad (5.4.23)
\]

We have managed to solve for the integral $I$.

**Problem 5.18.** If $x$ is a non-zero real number, justify the identity

\[
\ln(x + i0^+) = \ln|\imath| + i\pi\Theta(-x), \quad (5.4.24)
\]

where $\Theta$ is the step function.

**Problem 5.19.** (From Arfken et al.) For $-1 < a < 1$, show that

\[
\int_0^\infty dx \frac{x^a}{(x+1)^2} = \frac{\pi a}{\sin(\pi a)}. \quad (5.4.25)
\]

Hint: Complexify the integrand, then define a branch cut along the positive real line. Consider the closed counterclockwise contour that starts at the origin $z = 0$, goes along the positive real line, sweeps out an infinite counterclockwise circle which returns to the positive infinity end of the real line, then runs along the positive real axis back to $z = 0$. 

\[\square\]
5.5 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

$$f(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \tilde{f}(\vec{k}) e^{i\vec{k} \cdot \vec{x}} \quad (5.5.1)$$

$$\tilde{f}(\vec{k}) = \int_{\mathbb{R}^D} d^D\vec{x} f(\vec{x}) e^{-i\vec{k} \cdot \vec{x}} \quad (5.5.2)$$

Note that we have normalized our integrals differently from the linear algebra discussion. There, we had a $1/(2\pi)^D/2$ in both integrals, but here we have a $1/(2\pi)^D$ in the momentum space integrals and no $(2\pi)s$ in the position space ones. Always check the Fourier conventions of the literature you are reading. By inserting eq. (5.5.2) into eq. (5.5.1) we may obtain the integral representation of the $\delta$-function

$$\delta^{(D)}(\vec{x} - \vec{x}') = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} \quad (5.5.3)$$

In physical applications, almost any function residing in infinite space can be Fourier transformed. The meaning of the Fourier expansion in eq. (5.5.1) 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, $|\vec{k}| \sim 1/\lambda$. 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 $|\vec{k}| \sim \Lambda_{UV}$. (This in turn means $\Lambda_{UV}$ lets us define what we mean by coarse ($\equiv |\vec{k}| < \Lambda_{UV}$) and fine ($\equiv |\vec{k}| > \Lambda_{UV}$) 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 5.20.** 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 $\sigma$, but with some small oscillations added on top of it.

$$f(x) = \exp \left( -\frac{1}{2} \left( \frac{x - x_0}{\sigma} \right)^2 \right) (1 + \epsilon \sin(\omega x)) , \quad |\epsilon| \ll 1. \quad (5.5.4)$$
Assume that the wavelength of the oscillations is much shorter than the width of the Gaussian, $1/\omega \ll \sigma$. Find the Fourier transform $\tilde{f}(k)$ 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 $\epsilon$ is not small, then the oscillations – and hence the higher $|\vec{k}|$ modes – cannot be ignored.

**Problem 5.21.** Find the inverse Fourier transform of the “top hat” in 3 dimensions:

$$\tilde{f}(k) \equiv \Theta \left( \Lambda - |\vec{k}| \right) \quad (5.5.5)$$

$$f(\vec{x}) = ? \quad (5.5.6)$$

**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 5.22.** What is the Fourier transform of a multidimensional Gaussian

$$f(\vec{x}) = \exp \left( -x^i M_{ij} x^j \right) \quad (5.5.7)$$

where $M_{ij}$ is a real symmetric matrix? (You may assume all its eigenvalues are strictly positive.) Hint: You need to diagonalize $M_{ij}$. The Fourier transform result would involve both its inverse and determinant. Furthermore, your result should justify the statement: “The Fourier transform of a Gaussian is another Gaussian”.

**Problem 5.23.** If $f(\vec{x})$ is real, show that $\tilde{f}(\vec{k})^* = \tilde{f}(\vec{-k})$. Similarly, if $f(\vec{x})$ is a real periodic function in $D$-space, show that the Fourier series coefficients in eq. (4.5.165) and (4.5.166) obey $\tilde{f}(n^1, \ldots, n^D)^* = \tilde{f}(-n^1, \ldots, -n^D)$.

Suppose we restrict the space of functions on infinite $\mathbb{R}^D$ to those that are even under parity, $f(\vec{x}) = f(-\vec{x})$. Show that

$$f(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^D k}{(2\pi)^D} \cos (\vec{k} \cdot \vec{x}) \tilde{f}(\vec{k}). \quad (5.5.8)$$

What’s the inverse Fourier transform? If instead we restrict to the space of odd parity functions, $f(-\vec{x}) = -f(\vec{x})$, show that

$$f(\vec{x}) = i \int_{\mathbb{R}^D} \frac{d^D k}{(2\pi)^D} \sin (\vec{k} \cdot \vec{x}) \tilde{f}(\vec{k}). \quad (5.5.9)$$

Again, write down the inverse Fourier transform. Can you write down the analogous Fourier/inverse Fourier series for even and odd parity periodic functions on $\mathbb{R}^D$?

**Problem 5.24.** For a complex $f(\vec{x})$, show that

$$\int_{\mathbb{R}^D} d^D x |f(\vec{x})|^2 = \int_{\mathbb{R}^D} \frac{d^D k}{(2\pi)^D} |\tilde{f}(\vec{k})|^2, \quad (5.5.10)$$

$$\int_{\mathbb{R}^D} d^D x M^{ij} \partial_i f(\vec{x})^* \partial_j f(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^D k}{(2\pi)^D} M^{ij} k_i k_j |\tilde{f}(\vec{k})|^2, \quad (5.5.11)$$
where you should assume the matrix $M_{ij}$ does not depend on position $\vec{x}$.

Next, prove the convolution theorem: the Fourier transform of the convolution of two functions $F$ and $G$

$$f(\vec{x}) \equiv \int_{\mathbb{R}^D} d^Dy F(\vec{x} - \vec{y}) G(\vec{y})$$  \hspace{1cm} (5.5.12)

is the product of their Fourier transforms

$$\tilde{f}(\vec{k}) = \tilde{F}(\vec{k})\tilde{G}(\vec{k}).$$  \hspace{1cm} (5.5.13)

You may need to employ the integral representation of the $\delta$-function; or invoke linear algebraic arguments.

**5.5.1 Application: 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.\(^{43}\) We will study it in the form of the 2nd order ordinary differential equation (ODE)

$$m \ddot{x}(t) + f \dot{x}(t) + k x(t) = F(t), \hspace{0.5cm} f, k > 0,$$  \hspace{1cm} (5.5.14)

where each dot represents a time derivative; for e.g., $\ddot{x} \equiv d^2x/dt^2$. 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

$$\ddot{x}(t) + 2\gamma \dot{x}(t) + \Omega^2 x(t) = F(t), \hspace{0.5cm} \Omega \geq \gamma > 0.$$  \hspace{1cm} (5.5.15)

(For technical convenience, we have further restricted $\Omega$ to be greater or equal to $\gamma$.) We will perform a Fourier analysis of this problem by transforming both the trajectory and the external force,

$$x(t) = \int_{-\infty}^{+\infty} \tilde{x}(\omega)e^{i\omega t} \frac{d\omega}{2\pi}, \hspace{0.5cm} F(t) = \int_{-\infty}^{+\infty} \tilde{F}(\omega)e^{i\omega t} \frac{d\omega}{2\pi}.$$  \hspace{1cm} (5.5.16)

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 $(d/dt)^2 + 2\gamma(d/dt) + \Omega^2$. I will then show the fundamental importance of the Green’s function by showing how you

\(^{43}\)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.
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 5.25.** Show that, in frequency space, eq. (5.5.15) is

\[
(-\omega^2 + 2i\omega\gamma + \Omega^2) \tilde{x}(\omega) = \tilde{F}(\omega).
\]  

(5.5.17)

In effect, each time derivative \(d/dt\) is replaced with \(i\omega\). We see that the differential equation in eq. (5.5.15) is converted into an algebraic one in eq. (5.5.17).

**Inhomogeneous (particular) solution** For \(F \neq 0\), we may infer from eq. (5.5.17) that the particular solution — the part of \(\tilde{x}(\omega)\) that is due to \(\tilde{F}(\omega)\) — is

\[
\tilde{x}_p(\omega) = \frac{\tilde{F}(\omega)}{-\omega^2 + 2i\omega\gamma + \Omega^2},
\]  

(5.5.18)

which in turn implies

\[
x_p(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega t} \frac{\tilde{F}(\omega)}{-\omega^2 + 2i\omega\gamma + \Omega^2} = \int_{-\infty}^{+\infty} dt' F(t') G(t - t')
\]  

(5.5.19)

where

\[
G(t - t') = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{e^{i\omega(t-t')}}{-\omega^2 + 2i\omega\gamma + \Omega^2}.
\]  

(5.5.20)

To get to eq. (5.5.19) we have inserted the inverse Fourier transform

\[
\tilde{F}(\omega) = \int_{-\infty}^{+\infty} dt' F(t') e^{-i\omega t'}.
\]  

(5.5.21)

**Problem 5.26.** Show that the Green’s function in eq. (5.5.20) obeys the damped harmonic oscillator equation eq. (5.5.15), but driven by a impulsive force (“point-source-at-time \(t''\)"

\[
\left(\frac{d^2}{dt'^2} + 2\gamma \frac{d}{dt'} + \Omega^2\right) G(t - t') = \left(\frac{d^2}{dt'^2} - 2\gamma \frac{d}{dt'} + \Omega^2\right) G(t - t') = \delta(t - t'),
\]  

(5.5.22)

so that eq. (5.5.19) 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 \(2\gamma\) term. By “closing the contour” appropriately, verify that eq. (5.5.20) yields

\[
G(t - t') = \Theta(t - t') e^{-\gamma(t-t')} \frac{\sin \left(\sqrt{\Omega^2 - \gamma^2}(t - t')\right)}{\sqrt{\Omega^2 - \gamma^2}}.
\]  

(5.5.23)
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. (5.5.19) due to the step function $\Theta(t - t')$ in eq. (5.5.23). That is,

$$x_p(t) = \int_{-\infty}^{t} dt' F(t') G(t - t'). \quad (5.5.24)$$

**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

$$\left(\frac{d^2}{dt^2} + 2\gamma \frac{d}{dt} + \Omega^2\right) x_h(t) = 0. \quad (5.5.25)$$

We then start by examining the following integral

$$I(t, t') \equiv \int_{t'}^{\infty} dt'' \left\{ x_h(t'') \left( \frac{d^2}{dt''^2} - 2\gamma \frac{d}{dt''} + \Omega^2 \right) G(t - t'') - G(t - t'') \left( \frac{d^2}{dt''^2} + 2\gamma \frac{d}{dt''} + \Omega^2 \right) x_h(t'') \right\}. \quad (5.5.26)$$

Using the equations (5.5.22) and (5.5.25) obeyed by $G(t - t')$ and $x_h(t)$, we may immediately infer that

$$I(t, t') = \int_{t'}^{\infty} dt'' x_h(t'') \delta(t - t'') = \Theta(t - t') x_h(t). \quad (5.5.27)$$

(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 $\delta$-function contribution and the result would be zero.) On the other hand, we may in eq. (5.5.26) cancel the $\Omega^2$ terms, and then integrate-by-parts one of the derivatives from the $G$, $\dot{G}$, and $\ddot{x}_h$ terms.

$$I(t, t') = \left[ x_h(t'') \left( \frac{d}{dt''} - 2\gamma \right) G(t - t'') - G(t - t'') \frac{dx_h(t'')}{dt''} \right]_{t'' = t'} \int_{t'}^{\infty} dt'' + \int_{t'}^{\infty} dt'' \left( - \frac{dx_h(t'')}{dt''} \frac{dG(t - t'')}{dt''} + 2\gamma \frac{dx_h(t'')}{dt''} G(t - t'') \right.$$

$$\left. + \frac{dG(t - t'')}{dt''} \frac{dx_h(t'')}{dt''} - 2\gamma G(t - t'') \frac{dx_h(t'')}{dt''} \right). \quad (5.5.28)$$

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. (5.5.27), we have arrived at

$$\Theta(t - t') x_h(t) = G(t - t') \frac{dx_h(t')}{dt'} + \left( 2\gamma G(t - t') + \frac{dG(t - t')}{dt} \right) x_h(t'). \quad (5.5.29)$$
Because we have not made any assumptions about our trajectory – except it satisfies the homogeneous equation in eq. (5.5.25) – 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. (5.5.29) must span the 2-dimensional space of solutions to eq. (5.5.25). Therefore, the coefficients of \( x_h(t') \) and \( \dot{x}_h(t') \) must in fact be the two linearly independent homogeneous solutions to \( x_h(t') \),

\[
x^{(1)}_h(t) = G(t > t') = e^{-\gamma(t-t')} \frac{\sin \left( \sqrt{\Omega^2 - \gamma^2}(t-t') \right)}{\sqrt{\Omega^2 - \gamma^2}},
\]

\[
x^{(2)}_h(t) = 2\gamma G(t > t') + \partial_t G(t > t') = e^{-\gamma(t-t')} \left( \gamma \cdot \frac{\sin \left( \sqrt{\Omega^2 - \gamma^2}(t-t') \right)}{\sqrt{\Omega^2 - \gamma^2}} + \cos \left( \sqrt{\Omega^2 - \gamma^2}(t-t') \right) \right).
\]

(5.5.30) 5.5.31

\[\text{That } x^{(1,2)}_h \text{ must be independent for any } \gamma > 0 \text{ and } \Omega^2 \text{ is worth reiterating, because this is a potential issue for the damped harmonic oscillator equation when } \gamma = \Omega. \text{ We can check directly that, in this limit, } x^{(1,2)}_h \text{ 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 }

\[x_h(t) = \text{Re } e^{i\omega t},\]

we would find, upon inserting eq. (5.5.33) into eq. (5.5.25), that

\[\omega = \omega_\pm \equiv i\gamma \pm \sqrt{\Omega^2 - \gamma^2}.\]

(5.5.34)

This means, when \( \Omega = \gamma \), we obtain repeated roots and the otherwise linearly independent solutions

\[x_\pm(t) = \text{Re } e^{-\gamma t \pm i\sqrt{\Omega^2 - \gamma^2}t}\]

(5.5.35)

become linearly dependent there – both \( x_\pm(t) = e^{-\gamma t} \).

**Problem 5.27.** 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. (5.5.33) and verify that eq. (5.5.35) are indeed solutions to eq. (5.5.25) for \( \Omega \neq \gamma \). Comment on why the presence of \( t' \) in equations (5.5.30) and (5.5.31) amount to arbitrary constants multiplying the homogeneous solutions in eq. (5.5.35).

\[\text{Note that}
\]

\[\frac{dG(t - t')}{dt} = \Theta(t - t') \frac{d}{dt} \left( e^{-\gamma(t-t')} \frac{\sin \left( \sqrt{\Omega^2 - \gamma^2}(t-t') \right)}{\sqrt{\Omega^2 - \gamma^2}} \right).\]

(5.5.32)

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 5.28. 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

\[
F(t) = \text{Im} \left( e^{-(t/\tau)^2} e^{i\mu t} \right), \quad \text{for} \ -2\pi n/\mu \leq t \leq 2\pi n/\mu, \quad \mu > 0, \quad . \tag{5.5.36}
\]

and \( F(t) = 0 \) otherwise. \((n \text{ is an integer greater than } 1)\). Solve for the motion \( x(t > t_0) \) of the damped simple harmonic oscillator, in terms of \( t_0, V_0, \tau, \mu \text{ and } n. \)

5.6 Fourier Series

Consider a periodic function \( f(x) \) with period \( L \), meaning

\[
f(x + L) = f(x). \tag{5.6.1}
\]

Then its Fourier series representation is given by

\[
f(x) = \sum_{n=\infty}^{\infty} C_n e^{i2\pi n/L} x, \tag{5.6.2}
\]

\[
C_n = \frac{1}{L} \int_{\text{one period}} dx' f(x') e^{-i2\pi n/L} x'.
\]

(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 \( \psi \) has to vanish there

\[
\psi(x = 0) = \psi(x = L) = 0. \tag{5.6.3}
\]

Even though the Fourier series is supposed to represent the profile \( \psi \) 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 (5.6.2), applies, and is simply the superposition of possible vibrational modes allowed on the string itself.

Problem 5.29. \((\text{From Riley et al.})\) Find the Fourier series representation of the Dirac comb, i.e., find the \( \{C_n\} \) in

\[
\sum_{n=-\infty}^{\infty} \delta(x + nL) = \sum_{n=-\infty}^{\infty} C_n e^{i2\pi n/L} x, \quad x \in \mathbb{R}. \tag{5.6.4}
\]

Then prove the Poisson summation formula; where for an arbitrary function \( f(x) \) and its Fourier transform \( \tilde{f} \),

\[
\sum_{n=-\infty}^{\infty} f(x + nL) = \frac{1}{L} \sum_{n=-\infty}^{\infty} \tilde{f} \left( \frac{2\pi n}{L} \right) e^{i2\pi n/L} x. \tag{5.6.5}
\]
Hint: Note that
\[ f(x + nL) = \int_{-\infty}^{+\infty} dx' f(x') \delta(x - x' + nL). \]  \hspace{1cm} (5.6.6)

**Problem 5.30. 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 \( x \in [0, L) \),

\[
\begin{align*}
  f(x) &= -1, \quad -L/2 \leq x \leq 0 \\
  &= 1, \quad 0 \leq x \leq L/2. 
\end{align*}
\]  \hspace{1cm} (5.6.7) (5.6.8)

Find its Fourier series representation

\[ f(x) = \sum_{n=-\infty}^{\infty} C_n e^{i \frac{2 \pi n x}{L}}. \]  \hspace{1cm} (5.6.9)

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

\[ f_N(x) \equiv \sum_{n=-N}^{N} C_n e^{i \frac{2 \pi n x}{L}}, \]  \hspace{1cm} (5.6.10)

and find a computer program to plot \( f_N(x) \) as well as \( f(x) \) in eq. (5.6.7). You should see the \( f_N(x) \) over/undershooting the \( f(x) \) near the latter’s discontinuities, even for very large \( N \gg 1 \). \hfill \Box

\footnote{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.}
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. 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.

\[
\int f(x)\,dx = \int g'(x)\,dx = g(x) + \text{constant} \tag{6.0.1}
\]

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.

### 6.1 Gaussian integrals

As a start, let us consider the following “Gaussian” integral:

\[
I_G(a) \equiv \int_{-\infty}^{+\infty} e^{-ax^2}\,dx, \tag{6.1.1}
\]

where \( \text{Re}(a) > 0 \). (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:

\[
(I_G(a))^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-ax^2}e^{-ay^2}\,dx\,dy. \tag{6.1.2}
\]

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 \(dx\,dy = d\phi\,dr \cdot r\).

\[
(I_G(a))^2 = \int_{-\infty}^{+\infty} e^{-a(x^2+y^2)}\,dx\,dy = \int_0^{2\pi} d\phi \int_0^{+\infty} dr \cdot re^{-ar^2} \tag{6.1.3}
\]

The integral over \( \phi \) is straightforward; whereas the radial one now contains an additional \( r \) in the integrand – this is exactly what makes the integral do-able.

\[
(I_G(a))^2 = 2\pi \int_0^{+\infty} dr \frac{1}{-2a} \partial_re^{-ar^2} = \left[ -\frac{\pi}{a} e^{-ar^2} \right]_r^{r=\infty} = \frac{\pi}{a} \tag{6.1.4}
\]

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 \( \nu \).
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
\[
\int_{-\infty}^{+\infty} e^{-ax^2} \, dx = \sqrt{\frac{\pi}{a}}, \quad \text{Re}(a) > 0. \tag{6.1.5}
\]
Because both sides of eq. (6.1.5) 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 $\text{Re}(a) > 0$. 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?)

\[
I_G(a) = \int_{\xi=-\infty}^{\xi=+\infty} e^{-\rho e^{i\delta}(e^{-i\delta/2}\xi)^2} d(e^{-i\delta/2}\xi) = \frac{1}{e^{i\delta/2}} \int_{\xi=-\infty}^{\xi=+\infty} e^{-\rho \xi^2} \, d\xi \tag{6.1.6}
\]

At this point, since $\rho > 0$ we may refer to our result for $I_G(a > 0)$ and conclude
\[
\int_{-\infty}^{+\infty} e^{-ax^2} \, dx = \frac{1}{e^{i\delta/2}} \sqrt{\frac{\pi}{\rho}} = \sqrt{\frac{\pi}{\rho e^{i\delta}}} = \sqrt{\frac{\pi}{a}}, \quad -\frac{\pi}{2} < (\delta \equiv \text{arg}[a]) < \frac{\pi}{2}. \tag{6.1.7}
\]

**Problem 6.1.** Compute, for $\text{Re}(a) > 0$,

\[
\begin{align*}
\int_{0}^{+\infty} e^{-ax^2} \, dx, & \quad \text{for } \text{Re}(a) > 0 \tag{6.1.8} \\
\int_{-\infty}^{+\infty} e^{-ax^2} x^n \, dx, & \quad \text{for } n \text{ odd} \tag{6.1.9} \\
\int_{-\infty}^{+\infty} e^{-ax^2} x^n \, dx, & \quad \text{for } n \text{ even} \tag{6.1.10} \\
\int_{0}^{+\infty} e^{-ax^2} x^\beta \, dx, & \quad \text{for } \text{Re}(\beta) > -1 \tag{6.1.11}
\end{align*}
\]

Hint: For the very last integral, consider the change of variables $x' \equiv \sqrt{ax}$, and refer to eq. 5.2.1 of the NIST page [here](#).

**Problem 6.2. 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 $\Omega_{D-1}$ subtended by a sphere of radius $r$ is defined through the relation

\[
\text{Surface area of sphere } \equiv \Omega_{D-1} \cdot r^{D-1}. \tag{6.1.12}
\]

Since $r$ is the only length scale in the problem, and since area in $D$-space has to scale as $[\text{Length}^{D-1}]$, we see that $\Omega_{D-1}$ is independent of the radius $r$. Moreover, the volume of a
spherical shell of radius $r$ and thickness $dr$ must be the area of the sphere times $dr$. Now, argue that the $D$ dimensional integral in spherical coordinates becomes

$$\left(I_G(a = 1)\right)^D = \int_{\mathbb{R}^D} d^D\mathbf{x} e^{-\mathbf{x}^2} = \Omega_{D-1} \int_0^{\infty} dr \cdot r^{D-1} e^{-r^2}. \quad (6.1.13)$$

Next, evaluate $\left(I_G(a = 1)\right)^D$ directly. Then use the results of the previous problem to compute the last equality of eq. (6.1.13). At this point you should arrive at

$$\Omega_{D-1} = \frac{2\pi^{D/2}}{\Gamma(D/2)}, \quad (6.1.14)$$

where $\Gamma$ is the Gamma function.

6.2 Complexification

Sometimes complexifying the integral makes it easier. Here’s a simple example from Matthews and Walker [15].

$$I = \int_0^{\infty} dx e^{-ax} \cos(\lambda x), \quad a > 0, \; \lambda \in \mathbb{R}. \quad (6.2.1)$$

If we regard $\cos(\lambda x)$ as the real part of $e^{i\lambda x}$,

$$I = \text{Re} \int_0^{\infty} dx e^{-(a-i\lambda)x}$$
$$= \text{Re} \left[ e^{-(a-i\lambda)x} \right]_{x=0}^{x=\infty}$$
$$= \text{Re} \frac{1}{a - i\lambda} = \text{Re} \frac{a + i\lambda}{a^2 + \lambda^2} = \frac{a}{a^2 + \lambda^2} \quad (6.2.2)$$

Problem 6.3. What is

$$\int_0^{\infty} dx e^{-ax} \sin(\lambda x), \quad a > 0, \; \lambda \in \mathbb{R}? \quad (6.2.3)$$

6.3 Differentiation under the integral sign (Leibniz’s theorem)

Differentiation under the integral sign, or Leibniz’s theorem, is the result

$$\frac{d}{dz} \int_{a(z)}^{b(z)} ds F(z, s) = b'(z) F(z, b(z)) - a'(z) F(z, a(z)) + \int_{a(z)}^{b(z)} ds \frac{\partial F(z, s)}{\partial z}. \quad (6.3.1)$$

Problem 6.4. By using the limit definition of the derivative, i.e.,

$$\frac{d}{dz} H(z) = \lim_{\delta \to 0} \frac{H(z + \delta) - H(z)}{\delta}, \quad (6.3.2)$$

argue the validity of eq. (6.3.1).
Why this result is useful for integration can be illustrated by some examples. The art involves creative insertion of some auxiliary parameter $\alpha$ in the integrand. Let’s start with

$$\Gamma(n+1) = \int_0^\infty dt t^n e^{-t}, \quad n \text{ a positive integer.} \tag{6.3.3}$$

For $\Re(n) > -1$ this is in fact the definition of the Gamma function. We introduce the parameter as follows

$$I_n(\alpha) = \int_0^\infty dt t^n e^{-\alpha t}, \quad \alpha > 0, \tag{6.3.4}$$

and notice

$$I_n(\alpha) = (-\partial_\alpha)^n \int_0^\infty dt e^{-\alpha t} = (-\partial_\alpha)^n \frac{1}{\alpha}$$

$$= (-\alpha)^n (-1)(-2) \ldots (-n)\alpha^{-1-n} = n!\alpha^{-1-n} \tag{6.3.5}$$

By setting $\alpha = 1$, we see that the Gamma function $\Gamma(z)$ evaluated at integer values of $z$ returns the factorial.

$$\Gamma(n+1) = I_n(\alpha = 1) = n! \tag{6.3.6}$$

Next, we consider a trickier example:

$$\int_{-\infty}^{\infty} \frac{\sin(x)}{x} \, dx. \tag{6.3.7}$$

This can be evaluated via a contour integral. But here we do so by introducing a $\alpha \in \mathbb{R}$,

$$I(\alpha) \equiv \int_{-\infty}^{\infty} \frac{\sin(\alpha x)}{x} \, dx. \tag{6.3.8}$$

Observe that the integral is odd with respect to $\alpha$, $I(-\alpha) = -I(\alpha)$. Differentiating once,

$$I'(\alpha) = \int_{-\infty}^{\infty} \cos(\alpha x) \, dx = \int_{-\infty}^{\infty} e^{i\alpha x} \, dx = 2\pi \delta(\alpha). \tag{6.3.9}$$

($\cos(\alpha x)$ 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 $\delta$-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

$$I(\alpha) = \pi (\Theta(\alpha) - \Theta(-\alpha)) = \pi \text{sgn}(\alpha), \tag{6.3.10}$$

There is no integration constant here because it will spoil the property $I(-\alpha) = -I(\alpha)$. What remains is to choose $\alpha = 1$,

$$I(1) = \int_{-\infty}^{\infty} \frac{\sin(x)}{x} \, dx = \pi. \tag{6.3.11}$$

**Problem 6.5.** Evaluate the following integral

$$I(\alpha) = \int_0^\pi \ln \left[1 - 2\alpha \cos(x) + \alpha^2\right] \, dx, \quad |\alpha| \neq 1, \tag{6.3.12}$$

by differentiating once with respect to $\alpha$, 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.
6.4 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

\[ I(\vec{k}) \equiv \int_{S^2} \frac{d\Omega_{\hat{n}}}{4\pi} e^{i\vec{k} \cdot \hat{n}}. \] (6.4.1)

The \( \int_{S^2} d\Omega \) means we are integrating the unit radial vector \( \hat{n} \) 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

\[ \hat{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \] (6.4.2)

and \( d\Omega = d(\cos \theta)d\phi \). The key point here is that we have a rotationally invariant integral. In particular, the \((\theta, \phi)\) 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 \( R \),

\[ \hat{n}'(\theta, \phi) = R^j_i \hat{n}'(\theta', \phi'), \] (6.4.3)

where \( \det R = 1 \); in matrix notation \( \hat{n} = R^j_i \hat{n}' \) and \( R^T R = I \). Then \( d(\cos \theta)d\phi = d\Omega' \) \( \det R = d\Omega' = d(\cos \theta')d\phi' \), and

\[ I(R\vec{k}) = \int_{S^2} \frac{d\Omega_{\hat{n}}}{4\pi} e^{i\vec{k} \cdot \hat{n}} = \int_{S^2} \frac{d\Omega'_{\hat{n}'}}{4\pi} e^{i\vec{k} \cdot \hat{n}'} = I(\vec{k}). \] (6.4.4)

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 \( \vec{k} \cdot \hat{n}' \) only involved the \( \cos \theta \equiv \hat{n}' \cdot \hat{e}_3 \).

\[ I(|\vec{k}|) = \int_{0}^{2\pi} d\phi \int_{-1}^{+1} \frac{d(\cos \theta)}{4\pi} e^{i|\vec{k}| \cos \theta} = \frac{e^{i|\vec{k}|} - e^{-i|\vec{k}|}}{2i|\vec{k}|}. \] (6.4.5)

We arrive at

\[ \int_{S^2} \frac{d\Omega_{\hat{n}}}{4\pi} e^{i\vec{k} \cdot \hat{n}} = \sin |\vec{k}| \frac{|\vec{n}|}{|\vec{k}|}. \] (6.4.6)

**Problem 6.6.** With \( \hat{n} \) denoting the unit radial vector in 3–space, evaluate

\[ I(\vec{x}) = \int_{S^2} \frac{d\Omega_{\hat{n}}}{|\vec{x} - \vec{r}|}, \quad \vec{r} \equiv r\hat{n}. \] (6.4.7)

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 6.7. A problem that combines both rotational symmetry and the higher dimensional version of “differentiation under the integral sign” is the (tensorial) integral

$$
\int_{S^{2}} \frac{d\Omega}{4\pi} \hat{n}^{i_1} \hat{n}^{i_2} \ldots \hat{n}^{i_N},
$$

where $\hat{n}$ 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 $d\Omega \hat{n}$ transform under a parity flip of the coordinate system, namely under $\hat{n} \rightarrow -\hat{n}$? What’s the answer for even $N$? Hint: consider differentiating eq. (6.4.6) with respect to $k^{i_1}, \ldots, k^{i_N}$; 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 6.8. Can you generalize eq. (6.4.6) to $D$ spatial dimensions, namely

$$
\int_{S^{D-1}} d\Omega \hat{n} e^{i\vec{k} \cdot \hat{n}} = ?
$$

The $\vec{k}$ is an arbitrary vector in $D$-space and $\hat{n}$ is the unit radial vector in the same. Hint: You should find

$$
\int_{S^{D-1}} d\Omega \hat{n} e^{i\vec{k} \cdot \hat{n}} = \left( \int_{S^{D-2}} d\Omega \hat{n} \right) \left( \int_{0}^{\pi} (\sin \theta)^{D-2} e^{i|\vec{k}| \cos \theta} d\theta \right).
$$

Then refer to eq. 10.9.4 of the NIST page [here].

Tensor integrals. Next, we consider the following integral involving two arbitrary vectors $\vec{a}$ and $\vec{k}$ in 3D space.

$$
I(\vec{a}, \vec{k}) = \int_{S^2} d\Omega \frac{\vec{a} \cdot \hat{n}}{1 + \vec{k} \cdot \hat{n}}.
$$

First, we write it as $\vec{a}$ dotted into a vector integral $\vec{J}$, namely

$$
I(\vec{a}, \vec{k}) = \vec{\vec{a}} \cdot \vec{J}, \quad \vec{J}(\vec{k}) \equiv \int_{S^2} d\Omega \frac{\hat{n}}{1 + \vec{k} \cdot \hat{n}}.
$$

Let us now consider replacing $\vec{k}$ with a rotated version of $\vec{k}$. This amounts to replacing $\vec{k} \rightarrow R\vec{k}$, where $R$ is an orthogonal $3 \times 3$ matrix of unit determinant, with $R^T R = R R^T = I$. We shall see that $\vec{J}$ transforms as a vector $\vec{J} \rightarrow R \vec{J}$ under this same rotation. This is because $\int d\Omega \hat{n} \rightarrow \int d\Omega \hat{n}'$, for $\hat{n}' \equiv R^T \hat{n}$, and

$$
\vec{J}(R\vec{k}) = \int_{S^2} d\Omega \frac{R(R^T \hat{n})}{1 + R \hat{n}' \cdot (R^T \hat{n})}
= R \int_{S^2} d\Omega \frac{\hat{n}'}{1 + \vec{k} \cdot \hat{n}'} = R \vec{J}(\vec{k}).
$$

---

47 This example was taken from Matthews and Walker [15].
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 $\vec{k}$.

$$\vec{J} = f \cdot \vec{k}, \quad \Rightarrow \quad I \left( \vec{a}, \vec{k} \right) = f \vec{a} \cdot \vec{k}. \quad (6.4.14)$$

To calculate $f$ we now dot both sides with $\vec{k}$.

$$f = \frac{\vec{J} \cdot \vec{k}}{k^2} = \frac{1}{k^2} \int_{S^2} d\Omega_{\vec{n}} \frac{\vec{k} \cdot \hat{n}}{1 + \vec{k} \cdot \hat{n}} \quad (6.4.15)$$

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 $\hat{e}_3$ axis,

$$f = \frac{2\pi}{k^2} \int_{-1}^{+1} d\cos \theta \frac{|\vec{k}| \cos \theta}{1 + |\vec{k}| \cos \theta} = \frac{2\pi}{k^2} \int_{-1}^{+1} d\cos \theta \left( 1 - \frac{1}{1 + |\vec{k}| \cos \theta} \right) = \frac{4\pi}{k^2} \left( 1 - \frac{1}{2|\vec{k}|} \ln \left( 1 + \frac{|\vec{k}|}{1 - |\vec{k}|} \right) \right). \quad (6.4.16)$$

Therefore,

$$\int_{S^2} d\Omega_{\vec{n}} \frac{\vec{a} \cdot \hat{n}}{1 + \vec{k} \cdot \hat{n}} = \frac{4\pi}{k^2} \left( 1 - \frac{1}{2|\vec{k}|} \ln \left( \frac{1 + |\vec{k}|}{1 - |\vec{k}|} \right) \right). \quad (6.4.17)$$

This technique of reducing tensor integrals into scalar ones find applications even in quantum field theory calculations.

**Problem 6.9.** Calculate

$$A^{ij}(\vec{a}) \equiv \int \frac{d^3k}{(2\pi)^3} \frac{k^i k^j}{k^2 + (\vec{k} \cdot \vec{a})^2}, \quad (6.4.18)$$

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

$$a^i \rightarrow R^i_j a^j, \quad (6.4.19)$$

we have

$$A^{ij}(R^k_i a^l) = R^i_l R^j_k A^{kl}(\vec{a}). \quad (6.4.20)$$

Hint: The only rank-2 tensors available here are $\delta^{ij}$ and $a^i a^j$, so we must have

$$A^{ij} = f_1 \delta^{ij} + f_2 a^i a^j. \quad (6.4.21)$$

To find $f_{1,2}$ take the trace and also consider $A^{ij} a_i a_j$. 

196
6.5 Asymptotic expansion of integrals

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 are well studied – Bessel, Legendre, hypergeometric, etc. – all have integral representations. Often we wish to study these functions when their arguments are either very small or very large, and it is then useful to have techniques to extract an answer from these integrals in such limits. This topic is known as the “asymptotic expansion of integrals”.

6.5.1 Integration-by-parts (IBP)

In this section we will discuss how to use integration-by-parts (IBP) to approximate integrals. Previously we evaluated

\[
\frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-t^2} \, dt = 1. \tag{6.5.1}
\]

The erf function is defined as

\[
erf(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}. \tag{6.5.2}
\]

Its small argument limit can be obtained by Taylor expansion,

\[
erf(x \ll 1) = \frac{2}{\sqrt{\pi}} \int_0^x dt \left(1 - t^2 + \frac{t^4}{2!} - \frac{t^6}{3!} + \ldots\right) = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \ldots\right). \tag{6.5.3}
\]

But what about its large argument limit erf\(x \gg 1\)? We may write

\[
erf(x) = \frac{2}{\sqrt{\pi}} \left(\int_0^\infty dt - \int_x^\infty dt\right) e^{-t^2} = 1 - \frac{2}{\sqrt{\pi}} I(x), \quad I(x) \equiv \int_x^\infty dt e^{-t^2}. \tag{6.5.4}
\]

Integration-by-parts may be employed as follows.

\[
I(x) = \int_x^\infty dt \frac{1}{-2t} \partial_t e^{-t^2} = \left[\frac{e^{-t^2}}{-2t}\right]_{t=x}^{t=\infty} - \int_x^\infty dt \partial_t \left(\frac{1}{-2t}\right) e^{-t^2}
\]

\[
= \frac{e^{-x^2}}{2x} - \int_x^\infty dt \frac{e^{-t^2}}{2t^2} = \frac{e^{-x^2}}{2x} - \int_x^\infty dt \frac{1}{2t^2(-2t)} \partial_x e^{-t^2}
\]

\[
= \frac{e^{-x^2}}{2x} - \frac{e^{-x^2}}{4x^3} + \int_x^\infty dt \frac{3}{4t^4} e^{-t^2}. \tag{6.5.5}
\]

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].
Problem 6.10. After $n$ integration by parts,
\[
\int_x^\infty dt e^{-t^2} = e^{-x^2} \sum_{\ell=1}^n (-1)^{\ell-1} \frac{1 \cdot 3 \cdot 5 \ldots (2\ell - 3)}{2^\ell x^{2\ell-1}} - (-1)^n \frac{1 \cdot 3 \cdot 5 \ldots (2n - 1)}{2^n} \int_x^\infty dt \frac{e^{-t^2}}{t^{2n}}. \tag{6.5.6}
\]
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. (6.5.6), 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.49

Problem 6.11. Using integration-by-parts, develop a large $x \gg 1$ expansion for
\[
I(x) \equiv \int_x^\infty dt \frac{\sin(t)}{t}. \tag{6.5.7}
\]

Hint: Consider instead $\int_x^\infty dt \frac{\exp(it)}{t}$.

What is an asymptotic series? A Taylor expansion of say $e^x$
\[
e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{6.5.8}
\]
converges for all $|x|$. In fact, for a fixed $|x|$, we know summing up more terms of the series
\[
\sum_{\ell=0}^N \frac{x^{\ell}}{\ell!}, \tag{6.5.9}
\]
– 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
\[
S_N(x) = A_0 + \frac{A_1}{x} + \frac{A_2}{x^2} + \cdots + \frac{A_N}{x^N}. \tag{6.5.10}
\]
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.”

49In 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.
6.5.2 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 region. Outside of the relevant region the integrand itself is highly exponentially suppressed – a basic illustration of this is

\[ I(x) = \int_0^x e^{-t} = 1 - e^{-x}. \] (6.5.11)

As \( x \to \infty \) 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 \( t \sim O(10) \). 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.

Laplace’s Method We now turn to integrals of the form

\[ I(x) = \int_a^b f(t) e^{x\phi(t)} dt \] (6.5.12)

where both \( f \) and \( \phi \) 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 \( x \to +\infty \) 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 \( a < c < b \):

\[ I(x) \sim e^{x\phi(c)} \int_{c-\kappa}^{c+\kappa} (f(c) + (t-c)f'(c) + \ldots) \exp \left( x \left\{ \frac{\phi^{(p)}(c)}{p!} (t-c)^p + \ldots \right\} \right) dt, \] (6.5.13)

where we have assumed the first non-zero derivative of \( \phi \) is at the \( p \)th order, and \( \kappa \) 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

\[ \Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} dt, \quad \text{Re}(z) > 0, \] (6.5.14)

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. (6.5.13), is usually justified because the errors incurred are again exponentially small.

\[ ^{50}\text{In the Fourier transform section I pointed out how, if you merely need to resolve the coarser features of your wave profile, then provided the short wavelength modes do not have very large amplitudes, only the coefficients of the modes with longer wavelengths need to be known accurately. Here, we shall see some integrals only require us to know their integrands in a small region, if all we need is an approximate (but oftentimes highly accurate) answer. This is a good rule of thumb to keep in mind when tackling difficult, apparently complicated, problems in physics: focus on the most relevant contributions to the final answer, and often this will simplify the problem-solving process.} \]
Examples

The first example, where \( \phi'(c) \neq 0 \), is related to the integral representation of the parabolic cylinder function; for \( \text{Re}(\nu) > 0 \),

\[
I(x) = \int_0^{100} t^{\nu-1} e^{-t^2/2} e^{-xt} dt.
\]

(6.5.15)

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 \( x \gg 1 \).

\[
I(x \to \infty) \sim \int_0^\infty t^{\nu-1} \left( 1 - \frac{t^2}{2} + \ldots \right) e^{-xt} dt
= \int_0^\infty \frac{(xt)^{\nu-1}}{x^{\nu-1}} \left( 1 - \frac{(xt)^2}{2x^2} + \ldots \right) e^{-(xt)} \frac{d(xt)}{x}
= \frac{\Gamma(\nu)}{x^\nu} \left( 1 + O\left(x^{-2}\right) \right).
\]

(6.5.16)

The second example is

\[
I(x \to \infty) = \int_0^{88} \frac{\exp(-x \cosh(t))}{\sqrt{\sinh(t)}} dt
\sim \int_0^\infty \frac{\exp \left( -x \left( 1 + \frac{t^2}{2} + \ldots \right) \right)}{\sqrt{t} \sqrt{1 + t^2/6 + \ldots}} dt
\sim e^{-x} \int_0^\infty \frac{(x/2)^{1/4} \exp \left( -(\sqrt{x/2}t)^2 \right)}{\sqrt{\sqrt{x/2}t}} \frac{d(\sqrt{x/2}t)}{\sqrt{x/2}}.
\]

(6.5.17)

To obtain higher order corrections to this integral, we would have to expand both the \( \exp \) and the square root in the denominator. But the \( t^2/2 + \ldots \) comes multiplied with a \( x \) whereas the denominator is \( x \)-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 \( z \equiv t^2 \Rightarrow dz = 2tdt = 2\sqrt{z}dt \).

\[
\int_0^{88} \frac{\exp(-x \cosh(t))}{\sqrt{\sinh(t)}} dt \sim \frac{e^{-x}}{(x/2)^{1/4}} \int_0^\infty z^{\left(1-\frac{1}{4}-\frac{1}{4}\right)} - e^{-z} \frac{dz}{2}
= e^{-x} \frac{\Gamma(1/4)}{2^{3/4}x^{1/4}}.
\]

(6.5.18)

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 \( x \), 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 Can Laplace’s method apply to obtain a large $x \gg 1$ limit representation of the Gamma function itself?

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt = \int_0^\infty e^{(x-1)\ln(t)} e^{-t} dt \quad (6.5.19)$$

It does not appear so because here $\phi(t) = \ln(t)$ and the maximum is at $t = \infty$. Actually, the maximum of the exponent is at

$$\frac{d}{dt} ((x-1)\ln(t) - t) = \frac{x-1}{t} - 1 = 0 \quad \Rightarrow \quad t = x - 1. \quad (6.5.20)$$

Re-scale $t \rightarrow (x-1)t$:

$$\Gamma(x) = (x-1)e^{(x-1)\ln(x-1)} \int_0^\infty e^{(x-1)(\ln(t)-t)} dt. \quad (6.5.21)$$

Comparison with eq. (6.5.12) tells us $\phi(t) = \ln(t) - t$ and $f(t) = 1$. We may now expand the exponent about its maximum at 1:

$$\ln(t) - t = -1 - \frac{(t-1)^2}{2} + \frac{(t-1)^3}{3} + \ldots. \quad (6.5.22)$$

This means

$$\Gamma(x) \sim \sqrt{\frac{2}{x-1}} (x-1)^x e^{-(x-1)} \quad (6.5.23)$$

$$\times \int_{-\infty}^{+\infty} \exp \left( - \left( \sqrt{x-1} - \frac{t-1}{\sqrt{2}} \right)^2 + O((t-1)^3) \right) d(\sqrt{x-1}t/\sqrt{2}).$$

Noting $x - 1 \approx x$ for large $x$; we arrive at Stirling’s formula,

$$\Gamma(x \rightarrow \infty) \sim \sqrt{\frac{2\pi x^x}{x e^x}}. \quad (6.5.24)$$

Problem 6.12. What is the leading behavior of

$$I(x) \equiv \int_0^{50.12345 + \epsilon \sqrt{x} + \pi \sqrt{x}} e^{-x t^2} \sqrt{1 + \sqrt{t}} dt \quad (6.5.25)$$

in the limit $x \rightarrow +\infty$? And, how does the first correction scale with $x$?

Problem 6.13. What is the leading behavior of

$$I(x) = \int_{-\pi/2}^{\pi/2} e^{-x \cos(t)^2} \frac{dt}{\cos(t)^p}, \quad (6.5.26)$$

for $0 \leq p < 1$, in the limit $x \rightarrow +\infty$? Note that there are two maximums of $\phi(t)$ here.
Method of Stationary Phase  We now consider the case where the exponent is purely imaginary,

\[ I(x) = \int_a^b f(t)e^{ix\phi(t)}dt. \]  \hspace{1cm} (6.5.27)

Here, both \( f \) and \( \phi \) are real. As we did previously, we will consider the \( x \to +\infty \) 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 \textbf{Riemann-Lebesgue lemma} states that \( I(x \to \infty) \) in eq. (6.5.27) goes to zero provided: (I) \( \int_a^b |f(t)|dt < \infty \); (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 \( e^{ix\phi(t)} \) fluctuates wildly as \( x \to +\infty \) on the \( t \in [a, b] \) interval. For large enough \( x \), \( f(t)e^{ix\phi(t)} \) 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. (6.5.27) has to scale as \( 1/x \) as \( x \to \infty \).

\[
I(x) = \int_a^b \frac{f(t)}{ix\phi'(t)} \frac{d}{dt}e^{ix\phi(t)}dt \\
= \frac{1}{ix} \left\{ \left[ \frac{f(t)}{\phi'(t)} e^{ix\phi(t)} \right]_a^b - \int_a^b e^{ix\phi(t)} \frac{d}{dt} \left( \frac{f(t)}{\phi'(t)} \right) dt \right\}. \hspace{1cm} (6.5.28)
\]

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 \( x \to \infty \). Therefore what remains is

\[
\int_a^b f(t)e^{ix\phi(t)}dt \sim \frac{1}{ix} \left[ \frac{f(t)}{\phi'(t)} e^{ix\phi(t)} \right]_a^b, \quad x \to +\infty, \quad \phi'(a \leq t \leq b) \neq 0. \hspace{1cm} (6.5.29)
\]

**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. (6.5.27) scales as \( 1/x^{1/p} \), where \( p \) is the lowest order derivative of \( \phi \) that is non-zero at \( t = c \). Because \( 1/p < 1 \), the \( 1/x \) behavior we found above is sub-dominant to \( 1/x^{1/p} \) – 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

\[
\int_a^b f(t)e^{ix\phi(t)}dt \sim f(a) \exp \left( ix\phi(a) \pm i \frac{\pi}{2p} \right) \frac{\Gamma(1/p)}{p} \left( \frac{p!}{x^{\phi(p)(a)}} \right)^{1/p}, \hspace{1cm} (6.5.30)
\]
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. (6.5.30), we decompose the integral into

$$I(x) = \int_a^{a+\kappa} f(t) e^{ix\phi(t)} dt + \int_{a+\kappa}^b f(t) e^{ix\phi(t)} dt. \quad (6.5.31)$$

The second integral scales as $1/x$, as already discussed, since we assume there are no stationary points there. The first integral, which we shall denote as $S(x)$, may be expanded in the following way provided $\kappa$ is chosen appropriately:

$$S(x) = \int_a^{a+\kappa} (f(a) + \ldots) e^{ix\phi(a)} \exp \left( \frac{ix}{p!} (t-a)^p \phi^{(p)}(a) + \ldots \right) dt. \quad (6.5.32)$$

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 $\rho > 0$.

$$S(x \to \infty) \sim f(a) e^{ix\phi(a)} e^{\pm i\pi/(2p)} \int_0^{+\infty} \exp \left( -\frac{x}{p!} |\phi^{(p)}(a)| \rho^p \right) \frac{d(\rho^p)}{p \cdot \rho^{p-1}} \quad (6.5.33)$$

$$\sim f(a) e^{ix\phi(a)} \frac{e^{\pm i\pi/(2p)}}{p(\frac{\pi}{2}) |\phi^{(p)}(a)|^{1/p}} \int_0^{+\infty} \left( \frac{x}{p! |\phi^{(p)}(a)| s} \right)^{\frac{1}{p} - 1} \exp \left( -\frac{x}{p! |\phi^{(p)}(a)| s} \right) d \left( \frac{x}{p! |\phi^{(p)}(a)| s} \right).$$

This establishes the result in eq. (6.5.30).

**Problem 6.14.** Starting from the following integral representation of the Bessel function

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos (n\theta - x \sin \theta) d\theta \quad (6.5.34)$$

where $n = 0, 1, 2, 3, \ldots$, show that the leading behavior as $x \to +\infty$ is

$$J_n(x) \sim \sqrt{\frac{2}{\pi x}} \cos \left( x - \frac{n \pi}{2} - \frac{\pi}{4} \right). \quad (6.5.35)$$

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 to be complex.

$$I(x) = \int_C f(t) e^{xu(t)} e^{ixv(t)} dt, \quad (6.5.36)$$

\[\text{If } p \text{ is even, and if the stationary point is not one of the end points, observe that we can choose } \theta = \pm(\pi/(2p) + \pi) \Rightarrow e^{i\rho\theta} = \pm i \text{ for the } \rho < 0 \text{ portion of the contour – i.e., run a straight line rotated by } \theta \text{ through the stationary point – and the final result would simply be twice of eq. (6.5.30).} \]
The $f$, $u$ and $v$ are real; $C$ is some contour on the complex $t$ plane; and as before we will study the $x \to \infty$ 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 – where the imaginary part of the exponent does not change along it.

$$I(x) = e^{ixv} \int_{C'} f(t)e^{xu(t)} dt \quad (6.5.37)$$

One reason for doing so is that the constant phase contour also coincides with the steepest descent one of the real part of the exponent – unless the contour passes through a saddle point, where more than one steepest descent paths can intersect. Along a steepest descent path, Laplace’s method can then be employed to obtain an asymptotic series.

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

$$\partial_x u = \partial_y v, \quad \partial_y u = -\partial_x v \quad (6.5.38)$$

means the dot product of their gradients is zero:

$$\vec{\nabla}u \cdot \vec{\nabla}v = \partial_x u \partial_x v + \partial_y u \partial_y v = \partial_y v \partial_x v - \partial_x v \partial_y v = 0. \quad (6.5.39)$$

To sum:

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 $e^{xu}$.

We will examine the following simple example:

$$I(x) = \int_0^1 \ln(t) e^{ixt} dt. \quad (6.5.40)$$

We 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 $\ln(i\infty)e^{-x\infty}$, which is zero for positive $x$.

$$I(x) = i \int_0^{\infty} \ln(it)e^{-xt}dt - i \int_0^{\infty} \ln(1 + it)e^{ix(1+it)}dt$$

$$= i \int_0^{\infty} \ln(it)e^{-xt}dt - ie^{ix} \int_0^{\infty} \ln(1 + it)e^{-xt}dt. \quad (6.5.41)$$
Notice the exponents in both integrands have now zero (and therefore constant) phases.

\[
I(x) = i \int_0^\infty \ln(i(xt)/x)e^{-(xt)} \frac{d(xt)}{x} - ie^{ix} \int_0^\infty \ln(1 + i(xt)/x)e^{-(xt)} \frac{d(xt)}{x}
\]

\[
= i \int_0^\infty \ln(z) - \ln(x) + i\pi/2)e^{-z} \frac{dz}{x} - ie^{ix} \int_0^\infty \left(\frac{z}{x} + O(x^{-2})\right)e^{-z} \frac{dz}{x}.
\]  

(6.5.42)

The only integral that remains unfamiliar is the first one

\[
\int_0^\infty e^{-z} \ln(z) = \frac{\partial}{\partial \mu} \bigg|_{\mu=1} \int_0^\infty e^{-z} \mu^{(\mu-1)} \ln(z) = \frac{\partial}{\partial \mu} \bigg|_{\mu=1} \int_0^\infty e^{-z} z^{\mu-1}
\]

\[
= \Gamma'(1) = -\gamma_E
\]  

(6.5.43)

The \(\gamma_E = 0.577216\ldots\) is known as the Euler-Mascheroni constant. At this point,

\[
\int_0^1 \ln(t)e^{ix} dt \sim \frac{i}{x} \left( -\gamma_E - \ln(x) + i\pi/2 - \frac{ie^{ix}}{x} + O(x^{-2}) \right), \quad x \to +\infty.
\]  

(6.5.44)

**Problem 6.15.** Perform an asymptotic expansion of

\[
I(k) \equiv \int_{-1}^{+1} e^{ikx^2} dx
\]  

(6.5.45)

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? \(\Box\)

### 6.6 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):

\[
-\epsilon^2 \psi''(x) + U(x)\psi(x) = 0,
\]  

(6.6.1)

where \(\epsilon\) 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
ODEs may be removed via a redefinition of $\psi$.) 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
\[ 0 = \sum_{\ell=0}^{\infty} \epsilon^\ell \left( \alpha_\ell(x) \left( S'(x)^2 + U(x) \right) - i \left( \alpha_{\ell-1}(x) S''(x) + 2 S'(x) \alpha'_\ell(x) \right) - \alpha''_{\ell-2}(x) \right) \] (6.6.2)

with the understanding that $\alpha_{-2}(x) = \alpha_{-1}(x) = 0$. We need to set the coefficients of $\epsilon^\ell$ to zero. The first two terms ($\ell = 0, 1$) give us solutions to $S(x)$ and $\alpha_0(x)$.

\[ 0 = a_0 \left( S'(x)^2 + U(x) \right) \Rightarrow S_{\pm}(x) = \sigma_0 \pm i \int^x dx' \sqrt{U(x')}; \quad \sigma_0 = \text{const.} \]

\[ 0 = -i \epsilon \left( 2 \alpha_0'(x) S'(x) + \alpha_0(x) S''(x) \right), \quad \Rightarrow \alpha_0(x) = \frac{C_0}{U(x)^{1/4}} \]

(While the solutions $S_{\pm}(x)$ contains two possible signs, the $\pm$ in $S'$ and $S''$ factors out of the second equation and thus $\alpha_0$ does not have two possible signs.)

**Problem 6.16.** *Recursion relation for higher order terms*  
By considering the $\ell \geq 2$ terms in eq. (6.6.2), 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?

\[ 0 = -\epsilon^2 \psi_{\pm}''(x) + U(x) \psi_{\pm}(x) \] (6.6.3)

\[ \psi_{\pm}(x) = \frac{1}{U(x)^{1/4}} \exp \left[ \mp \frac{1}{\epsilon} \int^x dx' \sqrt{U(x')} \right] \sum_{\ell=0}^{\infty} \epsilon^\ell Q_{(\ell\pm)}(x), \] (6.6.4)

\[ Q_{(\ell\pm)}(x) = \pm \frac{1}{2} \int^x \frac{d x'}{U(x')^{1/4}} \frac{d^2}{d x'^2} \left( \frac{Q_{(\ell-1\pm)}(x')}{U(x')^{1/4}} \right), \quad Q_{(0\pm)}(x) \equiv 1 \] (6.6.5)

To lowest order

\[ \psi_{\pm}(x) = \frac{1}{U^{1/4}(x)} \exp \left[ \mp \frac{1}{\epsilon} \int^x dx' \sqrt{U(x')} \right] \left( 1 + \mathcal{O}[\epsilon] \right). \] (6.6.6)

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 $d/dx$ acting on some power of $U(x)$ yields a smaller quantity, i.e., roughly speaking $d \ln U(x)/dx \sim \epsilon \ll 1$; 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 fictitious 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. (6.6.4), 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, $U(x) \rightarrow ((x-x_0)^n/n!)U^{(n)}(x_0)$; 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.

6.7 **Calculus of Variation**
7 Differential Geometry of Curved Spaces

7.1 Preliminaries, Tangent Vectors, Metric, 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 learned, and will allow you to generalize it readily to 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 \( (d\ell)^2 \) between \( (x,y,z) \) and \( (x+dx, y+dy, z+dz) \) in flat (aka Euclidean) space is

\[
(d\ell)^2 = (dx)^2 + (dy)^2 + (dz)^2.
\]  

(7.1.1)

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. (7.1.1) into spherical coordinates,

\[
(x, y, z) \equiv r (\sin \theta \cdot \cos \phi, \sin \theta \cdot \sin \phi, \cos \theta),
\]

(7.1.2)

and find

\[
(d\ell)^2 = dr^2 + r^2(d\theta^2 + \sin(\theta)^2d\phi^2).
\]

(7.1.3)

The geometries in eq. (7.1.1) and eq. (7.1.3) 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 convention 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 with respect to the \( i \)th coordinate is \( \partial_i \equiv \partial/\partial x^i \). The advantage of such a notation is its

52 In 4-dimensional flat spacetime, 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: \( ds^2 \equiv (dt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \). (The opposite sign convention, i.e., \( ds^2 \equiv -(dt)^2 + (dx)^2 + (dy)^2 + (dz)^2 \), 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.
compactness: we can say we are using coordinates \( \{ x^i \} \), where \( i \in \{ 1, 2, 3 \} \). 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. (7.1.1) now reads:

\[
(dx^1)^2 + (dx^2)^2 + (dx^3)^2 = \delta_{ij} dx^i dx^j \equiv \sum_{1 \leq i,j \leq 3} \delta_{ij} dx^i dx^j. 
\]

(We say the indices of the \( \{ dx^i \} \) are being contracted with that of \( \delta_{ij} \).) The symbol \( \delta_{ij} \) is known as the Kronecker delta, defined as

\[
\delta_{ij} = 1, \quad i = j, \\
0, \quad i \neq j.
\]

Of course, \( \delta_{ij} \) is simply the \( ij \) component of the identity matrix. Already, we can see \( \delta_{ij} \) 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. (7.1.1) and eq. (7.1.3) as follows. First write

\[
\xi^i \equiv (r, \theta, \phi);
\]
which are subject to the restrictions

\[
r \geq 0, \quad 0 \leq \theta \leq \pi, \quad \text{and} \quad 0 \leq \phi < 2\pi.
\]

Then (7.1.1) becomes

\[
\delta_{ij} dx^i dx^j = \delta_{ab} \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} d\xi^i d\xi^j = \frac{\partial \vec{x}}{\partial \xi^i} \cdot \frac{\partial \vec{x}}{\partial \xi^j} d\xi^i d\xi^j,
\]

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, \( \delta_{ij} dx^i dx^j \equiv g_{ij} dx^i dx^j \), we have

\[
g_{i'j'}(\vec{\xi}) = \frac{\partial \vec{x}}{\partial \xi^i} \cdot \frac{\partial \vec{x}}{\partial \xi^j},
\]

where 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. In fact, what we are finding in eq. (7.1.9) is

\[
g_{i'j'}(\vec{\xi}) = g_{ab}(\vec{x}) \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j}.
\]

Let’s proceed to work out the above dot products. Firstly,

\[
\frac{\partial \vec{x}}{\partial r} = (\sin \theta \cdot \cos \phi, \sin \theta \cdot \sin \phi, \cos \theta),
\]

53It is common to use the English alphabets to denote space coordinates and Greek letters to denote spacetime ones. We will adopt this convention in these notes, but note that it is not a universal one; so be sure to check the notation of the book you are reading.
\[
\frac{\partial \vec{x}}{\partial \theta} = r \left(\cos \theta \cdot \cos \phi, \cos \theta \cdot \sin \phi, -\sin \theta\right), \\
\frac{\partial \vec{x}}{\partial \phi} = r \left(-\sin \theta \cdot \sin \phi, \sin \theta \cdot \cos \phi, 0\right).
\] (7.1.13)

\[
\frac{\partial \vec{x}}{\partial \phi} = r \left(-\sin \theta \cdot \sin \phi, \sin \theta \cdot \cos \phi, 0\right). \\
\frac{\partial \vec{x}}{\partial \phi} = r \left(-\sin \theta \cdot \sin \phi, \sin \theta \cdot \cos \phi, 0\right).
\] (7.1.14)

A direct calculation should return the results

\[
g_{r\theta} = g_{\theta r} = \partial \vec{x} / \partial r \cdot \partial \vec{x} / \partial \theta = 0, \\
g_{r\phi} = g_{\phi r} = \partial \vec{x} / \partial r \cdot \partial \vec{x} / \partial \phi = 0, \\
g_{\theta\phi} = g_{\phi \theta} = \partial \vec{x} / \partial \theta \cdot \partial \vec{x} / \partial \phi = 0;
\] (7.1.15)

and

\[
g_{rr} = \frac{\partial \vec{x}}{\partial r} \cdot \frac{\partial \vec{x}}{\partial r} \equiv \left(\frac{\partial \vec{x}}{\partial r}\right)^2 = 1, \\
g_{\theta\theta} = \left(\frac{\partial \vec{x}}{\partial \theta}\right)^2 = r^2, \\
g_{\phi\phi} = \left(\frac{\partial \vec{x}}{\partial \phi}\right)^2 = r^2 \sin^2(\theta).
\] (7.1.16) (7.1.17) (7.1.18)

Altogether, these yield eq. (7.1.3).

If the \( g_{ab}(\vec{x}) \) in eq. (7.1.11) were not simply \( \delta_{ab} \), 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. (7.1.11) is then the \( ij \) component of

\[
\hat{g}(\vec{\xi}) = \left(\frac{\partial x}{\partial \xi}\right)^T \hat{g}(\vec{x}) \frac{\partial x}{\partial \xi}.
\] (7.1.19)

**Problem 7.1.** Verify that the Jacobian matrix \( \partial x^i / \partial (r, \theta, \phi)^a \) encountered above can be cast as the following product

\[
\frac{\partial x^i}{\partial (r, \theta, \phi)^a} = \begin{bmatrix} \hat{r} & \hat{\theta} & \hat{\phi} \end{bmatrix} \text{diag} (1, r, r \sin \theta).
\] (7.1.20)

The \( \hat{r}, \hat{\theta}, \) and \( \hat{\phi} \) are the unit vectors pointing along the \( r- \), \( \theta- \) and \( \phi- \) coordinate lines at a given point in space. Use this result to carry out the matrix multiplication in eq. (7.1.19), so as to verify that eq. (7.1.3) follows from eq. (7.1.1).

**Tangent vectors** 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 \( \lambda \), transversing through a \((D \geq 2)\)–dimensional space. (The \( \vec{x} \) need not be Cartesian coordinates.) We may then define the tangent vector \( v^i(\lambda) \equiv \frac{d\vec{x}(\lambda)}{d\lambda} \). Conversely, given a vector field \( v^i(\vec{x}) \), i.e., a \((D \geq 2)\)–component object defined at every point in space, we may find a trajectory \( \vec{x}(\lambda) \) such that \( \frac{d\vec{x}}{d\lambda} = v^i(\vec{x}(\lambda)) \). (This amounts to integrating an ODE, and in this context is why \( \vec{x}(\lambda) \) is called the \textit{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 \( i = 1, 2, \ldots, D \), 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:

\[
v \equiv v^i(\vec{x}) \frac{\partial}{\partial x^i} \equiv v^i(x^1, x^2, \ldots, x^D) \frac{\partial}{\partial x^i} \equiv v^i \partial_i. \tag{7.1.21}
\]

As already alluded to, given these components \( \{v^i\} \), the vector \( v \) can be thought of as the velocity with respect to some (fictitious) time \( \lambda \) by solving the ordinary differential equation \( v^i = \frac{d\vec{x}^i(\lambda)}{d\lambda} \). We may now see this more explicitly; \( v^i \partial_i f(\vec{x}) \) is the time derivative of \( f \) along the integral curve of \( \vec{v} \) because

\[
v^i \partial_i f(\vec{x}(\lambda)) = \frac{dx^i(\lambda)}{d\lambda} \partial_i f(\vec{x}) = \frac{df(\lambda)}{d\lambda}. \tag{7.1.22}
\]

To sum: the \( \{\partial_i\} \) are the basis kets based at a given point \( p \) in the curved space, allowing us to enumerate all the independent directions along which we may compute the ‘time derivative’ of \( f \) at the same point \( p \).

**General spatial metric** In a generic curved space, the square of the infinitesimal distance between the neighboring points \( \vec{x} \) and \( \vec{x} + d\vec{x} \), which we will continue to denote as \((d\ell)^2\), is no longer given by eq. (7.1.1) – because we cannot expect Pythagoras’ theorem to apply. But by scaling arguments it should still be quadratic in the infinitesimal distances \( \{dx^i\} \). The most general of such expression is

\[
(d\ell)^2 = g_{ij}(\vec{x})dx^i dx^j. \tag{7.1.23}
\]

Since it measures distances, \( g_{ij} \) needs to be real. It is also symmetric, since any antisymmetric portion would drop out of the summation in eq. (7.1.23) 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 coordinate system, as the matrix inverse of \( g_{ij} \):

\[
g^{ij} g_{jl} \equiv \delta^i_l, \quad \Leftrightarrow \quad g^{ij} \equiv (g^{-1})_{ij}. \tag{7.1.24}
\]
Everything else in a differential geometric calculation follows from the curved metric in eq. (7.1.23), once it is specified for a given setup[^54] the ensuing Christoffel symbols, Riemann/Ricci tensors, covariant derivatives/curl/divergence; what defines straight lines; parallel transport; 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 the distance swept out by this path is given by the integral

\[
\ell = \int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} \sqrt{g_{ij}(\vec{x}(\lambda))} \, dx^i dx^j = \int_{\lambda_1}^{\lambda_2} d\lambda \sqrt{g_{ij}(\vec{x}(\lambda))} \frac{dx^i(\lambda) \, dx^j(\lambda)}{d\lambda}. \tag{7.1.25}
\]

**Problem 7.2. Affine Parameterization**

Show that the definition in eq. (7.1.25) yields an infinitesimal distance that is invariant under an arbitrary change of the parameter \( \lambda \), as long as the transformation is orientation preserving. That is, suppose we replace \( \lambda \rightarrow \lambda'(\lambda) \) and thus \( d\lambda = (d\lambda'/d\lambda) \, d\lambda' \) — then as long as \( d\lambda'/d\lambda > 0 \), we have

\[
d\ell = d\lambda \sqrt{g_{ij}(\vec{x})} \frac{dx^i \, dx^j}{d\lambda} = d\lambda' \sqrt{g_{ij}(\vec{x})} \frac{dx^i \, dx^j}{d\lambda'/d\lambda}; \tag{7.1.26}
\]

and hence

\[
\ell = \int_{\lambda_1}^{\lambda_2} d\lambda' \sqrt{g_{ij}(\vec{x}(\lambda'))} \frac{dx^i(\lambda') \, dx^j(\lambda')}{d\lambda'/d\lambda}, \tag{7.1.27}
\]

where \( \lambda(\lambda_1, \lambda_2) = \lambda_{1,2} \). The parameter \( \lambda \) is really a coordinate in the 1D swept out by the path \( \vec{x}(\lambda) \); parameterization invariance here simply amounts to the statement that any 1D coordinate may be used to describe distances/paths.

Why can we always choose \( \lambda \) such that

\[
\sqrt{g_{ij}(\vec{x}(\lambda))} \frac{dx^i(\lambda) \, dx^j(\lambda)}{d\lambda} = \text{constant}, \tag{7.1.28}
\]

i.e., the square root factor can be made constant along the entire path linking \( \vec{x}_1 \) to \( \vec{x}_2 \)? (Hint: Up to a re-scaling and a 1D translation, this amounts to using the path length itself as the parameter \( \lambda \).) Now, suppose eq. (7.1.28) holds, explain why the square of the distance integral in eq. (7.1.25) may then be expressed as

\[
\ell^2 = (\lambda_2 - \lambda_1) \int_{\lambda_1}^{\lambda_2} g_{ij}(\vec{z}(\lambda)) \frac{dz^i \, dz^j}{d\lambda \, d\lambda} \, d\lambda. \tag{7.1.29}
\]

Hint: Use the constancy of the square root factor to solve \( \lambda_2 - \lambda_1 \) in terms of \( \ell \).

**Kets and Bras**

Earlier, while discussing tangent vectors, we stated that the \( \{\partial_i\} \) are the ket’s, the basis tangent vectors at a given point in space. The infinitesimal distances \( \{dx^i\} \) can now, in turn, be thought of as the basis dual vectors (the bra’s) — through the definition

\[
\langle dx^i | \partial_j \rangle = \delta^i_j. \tag{7.1.30}
\]

[^54]: As with most physics texts on differential geometry, we will ignore torsion (but will discuss it briefly in §8).
Why this is a useful perspective is due to the following. Let us consider an infinitesimal variation of our arbitrary function at $\vec{x}$:

$$df = \partial_if(\vec{x})dx^i. \quad (7.1.31)$$

Then, given a vector field $v$, we can employ eq. (7.1.30) to construct the derivative of the latter along the former, at some point $\vec{x}$, by

$$\langle df|v \rangle = v^j \partial_i f(\vec{x}) \langle dx^i|\partial_j \rangle = v^j \partial_i f(\vec{x}). \quad (7.1.32)$$

What about the inner products $\langle dx^i|dx^j \rangle$ and $\langle \partial_i|\partial_j \rangle$? They are

$$\langle dx^i|dx^j \rangle = g^{ij} \quad \text{and} \quad \langle \partial_i|\partial_j \rangle = g_{ij}. \quad (7.1.33)$$

This is because

$$g_{ij} \langle dx^j \rangle \equiv \langle \partial_i \rangle \quad \Leftrightarrow \quad g_{ij} \langle dx^i \rangle \equiv \langle \partial_i \rangle; \quad (7.1.34)$$

or, equivalently,

$$\langle dx^j \rangle \equiv g^{ij} \langle \partial_i \rangle \quad \Leftrightarrow \quad \langle dx^i \rangle \equiv g^{ij} \langle \partial_i \rangle. \quad (7.1.35)$$

In other words,

At a given point in a curved space, one may define two different vector spaces – one spanned by the basis tangent vectors $\{\langle \partial_i \rangle\}$ and another by its dual ‘bras’ $\{\langle dx^i \rangle\}$. Moreover, these two vector spaces are connected through the metric $g_{ij}$ and its inverse.

**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 angles of an extended triangle is not $\pi$.

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 $\vec{x} + d\vec{y}$; then to $\vec{x} + d\vec{y} + d\vec{z}$; then to $\vec{x} + d\vec{z}$; then back to $\vec{x}$. Then, denoting the end result as $v'^i$, we would find that

$$v'^i - v^i \propto R^i_{jkl}v^jdy^kdz^l. \quad (7.1.36)$$

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.
A common textbook example of a curved space is that of a 2-sphere of some fixed radius, sitting in 3D flat space, parametrized by the usual spherical coordinates \(0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi\). 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 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 \(\pi + |\phi_0 - \phi'_0| > \pi\).

Comparing tangent vectors at different places That tangent vectors cannot, in general, be parallel transported in a curved space also tells us comparing tangent vectors based at different locations is not a straightforward procedure, especially compared to the situation in flat Euclidean space. This is because, if \(\vec{v}(\vec{x})\) is to be compared to \(\vec{v}(\vec{x}')\) by parallel transporting \(\vec{v}(\vec{x})\) to \(\vec{x}'\); different results will 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.

7.2 Locally Flat Coordinates & Symmetries, Infinitesimal Volumes, General Tensors, Orthonormal Basis

Locally flat 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:

\[
\lim_{\vec{y} \to \vec{y}_0} g_{ij}(\vec{y}) = \delta_{ij} - \frac{1}{3} R_{ikjl}(\vec{y}_0) (y - y_0)^k(y - y_0)^l + \ldots, \quad (7.2.1)
\]

Any curved space can in fact always be viewed as a curved surface residing in a higher dimensional flat space. The 2-sphere has positive curvature; whereas a saddle has negative curvature, and would support a triangle whose angles add up to less than \(\pi\). 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.

Also known as Riemann normal coordinates.
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 spacetime, 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 \( \{ \hat{R}_{ij} \} \), which obeys \( \hat{R}_a^i \hat{R}_b^j \delta_{ab} = \delta_{ij} \), if we define in Euclidean space the following change-of-Cartesian coordinates (from \( \vec{x} \) to \( \vec{x}' \))

\[
x^i \equiv \hat{R}_i^j x'^j;
\]

the flat metric would retain the same form

\[
\delta_{ij} dx^i dx^j = \delta_{ab} \hat{R}_a^i \hat{R}_b^j dx'^i dx'^j = \delta_{ij} dx'^i dx'^j.
\]

A similar calculation would tell us flat Euclidean space is invariant under parity flips, i.e., \( x'^k \equiv -x^k \) for some fixed \( k \), as well as spatial translations \( \vec{x}' \equiv \vec{x} + \vec{a} \), for constant \( \vec{a} \). 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 \( \mathcal{O}(1/|\max R_{ijkl}(\vec{y}_0)|^{1/2}) \), 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!

**Coordinate-transforming the metric** Note that, in the context of eq. (7.1.23), \( \vec{x} \) is not a vector in Euclidean space, but rather another way of denoting \( x^a \) without introducing too many dummy indices \( \{ a, b, \ldots, i, j, \ldots \} \). Also, \( x^i \) in eq. (7.1.23) 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

\[
g_{ij}(\vec{x}) dx^i dx^j = g_{ij}(\vec{\xi}) \left( \frac{\partial x^i(\vec{\xi})}{\partial \xi^a} \right) \left( \frac{\partial x^j(\vec{\xi})}{\partial \xi^b} \right) d\xi^a d\xi^b \equiv g_{ab}(\vec{\xi}) d\xi^a d\xi^b.
\]

We can attribute all the coordinate transformation to how it affects the components of the metric:

\[
g_{ab}(\vec{\xi}) = g_{ij}(\vec{x}) \left( \frac{\partial x^i(\vec{x})}{\partial \xi^a} \right) \left( \frac{\partial x^j(\vec{x})}{\partial \xi^b} \right).
\]

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. (7.1.19) that eq. (7.2.5) may be calculated via matrix multiplication.
Inverse metric

Previously, we defined $g^{ij}$ to be the matrix inverse of the metric tensor $g_{ij}$. We can also view $g^{ij}$ as components of the tensor

$$g^{ij}(\vec{x}) \partial_i \otimes \partial_j,$$  

(7.2.6)

where we have now used $\otimes$ to indicate we are taking the tensor product of the partial derivatives $\partial_i$ and $\partial_j$. In $g_{ij}(\vec{x}) \, dx^i \, dx^j$ we really should also have $dx^i \otimes dx^j$, 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 $dx^i$ transforms under $\vec{x} \rightarrow \vec{x}(\vec{\xi})$, we also can work out how the partial derivatives transform.

$$g^{ij}(\vec{\xi}) \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} = g^{ab}(\vec{x}(\vec{\xi})) \frac{\partial \xi^i}{\partial x^a} \frac{\partial \xi^j}{\partial x^b} \otimes \frac{\partial}{\partial \xi^a} \otimes \frac{\partial}{\partial \xi^b}$$  

(7.2.7)

In terms of its components, we can read off their transformation rules:

$$g^{ij}(\vec{\xi}) = g^{ab}(\vec{x}(\vec{\xi})) \frac{\partial \xi^i}{\partial x^a} \frac{\partial \xi^j}{\partial x^b}.$$  

(7.2.8)

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 $\vec{\xi}$.

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,

$$\frac{\partial x^i}{\partial \xi^j} \frac{\partial \xi^j}{\partial x^a} = \delta_i^a, \quad \text{and} \quad \frac{\partial x^i}{\partial \xi^a} \frac{\partial \xi^a}{\partial \xi^j} = \delta_i^j.$$  

(7.2.9)

In other words, given $\vec{x} \rightarrow \vec{x}(\vec{\xi})$, we can compute $J_i^a \equiv \partial x^a/\partial \xi^i$ in terms of $\vec{\xi}$, with $a$ being the row number and $i$ as the column number. Then find the inverse, i.e., $(J^{-1})_i^a$, and identify it with $\partial \xi^a/\partial x^i$ in terms of $\vec{\xi}$.

**Problem 7.3.** Let $x^i$ be Cartesian coordinates and

$$\xi^i \equiv (r, \theta, \phi)$$  

(7.2.10)

be the usual spherical coordinates; see eq. (7.1.7). Calculate $\partial \xi^i/\partial x^a$ in terms of $\vec{\xi}$ and thereby, from the flat inverse metric $\delta^{ij}$ in Cartesian coordinates, find the inverse metric $g^{ij}(\vec{\xi})$ in the spherical coordinate system. Hint: Compute $\partial x^a/\partial (r, \theta, \phi)^a$. How do you get $\partial (r, \theta, \phi)^a/\partial x^i$ from it? You may also find the form of the Jacobian matrix in eq. (7.1.20) to be useful here.

**General tensor**

A *scalar* $\varphi$ is an object with no indices that transforms as

$$\varphi(\vec{\xi}) = \varphi \left( \vec{x}(\vec{\xi}) \right).$$  

(7.2.11)

That is, take $\varphi(\vec{x})$ and simply replace $\vec{x} \rightarrow \vec{x}(\vec{\xi})$ 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 \((r, \theta, \phi)\), the equipotential surfaces of the radial coordinate are simply the surface of a sphere with radius \(r\). Their intersection with constant \(\theta\) surfaces form latitude lines; and with constant \(\phi\) surfaces form longitude ones.

A vector \(v^i(\vec{x})\partial_i\) transforms as, by the chain rule,
\[
v^i(\vec{x}) \frac{\partial}{\partial x^i} = v^i(\vec{\xi}(\vec{x})) \frac{\partial \xi^j}{\partial x^i} \frac{\partial}{\partial \xi^j} \equiv v^j(\vec{\xi}) \frac{\partial}{\partial \xi^j}.
\] (7.2.12)

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 \(\vec{\xi}\)-coordinate system \(v^j(\vec{\xi})\) through the relation
\[
v^j(\vec{\xi}) = v^a(\vec{x}(\vec{\xi})) \frac{\partial \xi^j}{\partial x^a}.
\] (7.2.13)

Similarly, a 1-form \(A_i dx^i\) transforms, by the chain rule,
\[
A_i(\vec{x}) dx^i = A_i(\vec{\xi}(\vec{x})) \frac{\partial x^i}{\partial \xi^j} d\xi^j \equiv A_j(\vec{\xi}) d\xi^j.
\] (7.2.14)

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 \(A_j(\vec{\xi})\) through
\[
A_j(\vec{\xi}) = A_i(\vec{x}(\vec{\xi})) \frac{\partial x^i}{\partial \xi^j}.
\] (7.2.15)

By taking tensor products of \(\{\partial_i\}\) and \(\{dx^i\}\), we may define a rank \((N_M)\) tensor \(T\) as an object with \(N\) “upper indices” and \(M\) “lower indices” that transforms as
\[
T^{i_1 i_2 ... i_N}_{j_1 j_2 ... j_M}(\vec{x}) = T^{a_1 a_2 ... a_N}_{b_1 b_2 ... b_M}(\vec{x}(\vec{\xi})) \frac{\partial \xi^{i_1}}{\partial x^{a_1}} \cdots \frac{\partial \xi^{i_N}}{\partial x^{a_N}} \frac{\partial x^{b_1}}{\partial \xi^{j_1}} \cdots \frac{\partial x^{b_M}}{\partial \xi^{j_M}}. \] (7.2.16)

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
\[
T^{i_1 i_2 ... i_N}_{j_1 j_2 ... j_M}(\vec{x}) \left( \frac{\partial}{\partial x^{i_1}} \right) \otimes \ldots \otimes \left( \frac{\partial}{\partial x^{i_N}} \right) \otimes \left( dx^{j_1} \right) \otimes \ldots \otimes \left( dx^{j_M} \right).
\] (7.2.17)

58 Above, we only considered \(T\) with all upper indices followed by all lower indices. Suppose we had \(T^{i \, k}_{\, j}\); it is the components of
\[
T^{i \, k}_{\, j}(\vec{x}) |\partial_i \rangle \otimes \langle dx^j | \otimes |\partial_k \rangle.
\] (7.2.18)

58 Strictly speaking, when discussing the metric and its inverse above, we should also have respectively expressed them as \(g_{ij} \langle dx^i | \otimes \langle dx^j |\) and \(g^{ij} |\partial_i \rangle \otimes |\partial_j \rangle\), with the appropriate bras and kets enveloping the \(\{dx^i\}\) and \(\{\partial_i\}\). We did not do so because we wanted to highlight the geometric interpretation of \(g_{ij}dx^i dx^j\) as the square of the distance between \(\vec{x}\) and \(\vec{x} + d\vec{x}\), where the notion of \(dx^i\) as (a component of) an infinitesimal ‘vector’ – as opposed
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,

\[ T^{m_1...m_a \ n_1...n_b}_i = g_{ij} T^{m_1...m_a j n_1...n_b}, \]  
\[ T_{m_1...m_a \ n_1...n_b}^i = g^{ij} T_{m_1...m_a j n_1...n_b}. \]

Because upper indices transform oppositely from lower indices – see eq. (7.2.9) – when we contract a upper and lower index, it now transforms as a scalar. For example,

\[ A_i^i B^{ij}(\xi) = \frac{\partial \xi^i}{\partial x^m} A^m_a \left( \vec{x}(\xi) \right) \frac{\partial x^a}{\partial \xi^c} \frac{\partial \xi^c}{\partial x^n} B^{cn} \left( \vec{x}(\xi) \right) \frac{\partial \xi^j}{\partial x^n}. \]

Hence, \( v_i = g_{ij} v^j \) automatically converts the vector \( v^i \) into a tensor that transforms properly as a 1-form; and similarly, \( v^i = g^{ij} v_j \) automatically produces a vector from a 1-form \( v_i \). (Recall the “Kets and Bras” discussion above.) Moreover, we have the following equivalent scalars

\[ v^i w_i = g_{ij} v^i w^j = g^{ij} v_i w_j. \]

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.

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. (7.2.16) that defines a tensor is actually the statement that, the mathematical description of the physical world (the tensors themselves in eq. (7.2.17) 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. An example is, it makes no sense to assert that your new-found law of physics depends on \( g^{ij} \), 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?

Below, we will show that the infinitesimal volume in curved space is given by \( d^D \vec{x} \sqrt{g(\vec{x})} \), where \( g(\vec{x}) \) is the determinant of the metric in the \( \vec{x} \)–coordinate basis. For this to make sense 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. (7.2.11) 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, \( \xi \)) one would, in fact, do \( T(\vec{x}) \equiv T(\vec{x}(\xi)) \), because you’d want \( \xi \) to refer to the same point in space as \( \vec{x} = \vec{x}(\xi) \). For a general tensor in eq. (7.2.17), the tensor components \( T^{a_1a_2...a_N}_{b_1b_2...b_M} \) 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. (7.2.16) are really a shorthand for the lazy physicist who does not want to carry the basis vectors/1-forms around in his/her calculations.

\(^{59}\)You may also demand your equations/quantities to be tensors/scalars under group transformations.
geometrically, you will show in Problem [7.4] below that it is in fact generally covariant – i.e., it takes the same form in any coordinate system:

$$d^D\vec{x}\sqrt{g(\vec{x})} = d^D\vec{\xi}\sqrt{g(\vec{\xi})};$$

(7.2.23)

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

$$S^{ii...} = T^{ii...},$$

(7.2.24)

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. [7.2.1], 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 $d^Dx$ in Cartesian coordinates?

**Determinant of metric and the infinitesimal volume**

The determinant of the metric transforms as

$$\det g_{ij}(\vec{\xi}) = \det \left[ g_{ab}\left(\vec{x}(\vec{\xi})\right)\right] \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j}. $$

(7.2.25)

Using the properties $\det A \cdot B = \det A \det B$ and $\det A^T = \det A$, for any two square matrices $A$ and $B$,

$$\det g_{ij}(\vec{\xi}) = \left(\det \frac{\partial x^a}{\partial \xi^b}(\vec{\xi})\right)^2 \det g_{ij}\left(\vec{x}(\vec{\xi})\right).$$

(7.2.26)

The square root of the determinant of the metric is often denoted as $\sqrt{|g|}$. It transforms as

$$\sqrt{|g(\vec{\xi})|} = \sqrt{\left|g\left(\vec{x}(\vec{\xi})\right)\right|} \left|\det \frac{\partial x^a}{\partial \xi^b}(\vec{\xi})\right|. $$

(7.2.27)

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 $d^D\vec{x}$ and $g_{ij}(\vec{x}_0) = 1$. Recall from multi-variable calculus that, whenever we transform $\vec{x} \rightarrow \vec{x}(\vec{\xi})$, the integration measure would correspondingly transform as

$$d^D\vec{x} = d^D\vec{\xi}\left|\det \frac{\partial x^i}{\partial \xi^a}\right|,$$

(7.2.28)
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. (7.2.27) 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 \( \vec{\xi} \),

\[
\sqrt{|g(\vec{\xi})|} = \left( \sqrt{|g(\vec{x}(\vec{\xi}))|} \right) \left| \det \frac{\partial x^i(\vec{\xi})}{\partial \xi^a} \right|_{\vec{x} = \vec{x}_0} = \left| \det \frac{\partial x^i(\vec{\xi})}{\partial \xi^a} \right|_{\vec{x} = \vec{x}_0}. \tag{7.2.29}
\]

In flat space and by employing Cartesian coordinates \( \{ \vec{x} \} \), the infinitesimal volume (at some location \( \vec{x} = \vec{x}_0 \)) is \( d^D \vec{x} \). What is its curved analog? What we have just shown is that, by going from \( \vec{\xi} \) to a locally flat coordinate system \( \{ \vec{x} \} \),

\[
d^D \vec{x} = d^D \vec{\xi} \left| \det \frac{\partial x^i(\vec{\xi})}{\partial \xi^a} \right|_{\vec{x} = \vec{x}_0} = d^D \vec{\xi} \sqrt{|g(\vec{\xi})|}. \tag{7.2.30}
\]

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

\[
d^D \vec{\xi} \sqrt{|g(\vec{\xi})|} \equiv d\xi^1 \ldots d\xi^D \sqrt{|g(\vec{\xi})|}. \tag{7.2.31}
\]

**Problem 7.4.** Upon an orientation preserving change of coordinates \( \vec{y} \rightarrow \vec{y}(\vec{\xi}) \), where \( \det \partial y / \partial \xi > 0 \), show that

\[
d^D \vec{y} \sqrt{|g(\vec{y})|} = d^D \vec{\xi} \sqrt{|g(\vec{\xi})|}. \tag{7.2.32}
\]

Therefore calling \( d^D \vec{x} \sqrt{|g(\vec{x})|} \) 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 \( \vec{y}(\vec{\xi}) \); i.e., be careful that the determinant is not a scalar.

**Volume integrals** If \( \varphi(\vec{x}) \) is some scalar quantity, finding its volume integral within some domain \( \mathcal{D} \) in a generally covariant way can be now carried out using the infinitesimal volume we have uncovered; it reads

\[
I \equiv \int_{\mathcal{D}} d^D \vec{x} \sqrt{|g(\vec{x})|} \varphi(\vec{x}). \tag{7.2.33}
\]

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 \( d\ell^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \). Therefore \( \det g_{ij} = r^4(\sin \theta)^2 \) and the sphere’s volume reads

\[
\text{Vol}(r \leq R) = \int_{r \leq R} d^3 \vec{\xi} \sqrt{|g(\vec{\xi})|}, \quad \xi^i \equiv (r, \theta, \phi) \tag{7.2.34}
\]

\[
= \int_0^R dr \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi = \frac{4}{3} \pi R^3. \tag{7.2.35}
\]
Problem 7.5. Spherical coordinates in $D$ space dimensions. In $D$ space dimensions, we may denote the $D$-th unit vector as $\hat{e}_D$; and $\hat{n}_{D-1}$ as the unit radial vector, parametrized by the angles $\{0 \leq \theta^1 < 2\pi, 0 \leq \theta^2 \leq \pi, \ldots, 0 \leq \theta^{D-2} \leq \pi\}$, in the plane perpendicular to $\hat{e}_D$. Let $r \equiv |\vec{x}|$ and $\hat{n}_D$ be the unit radial vector in the $D$ space. Any vector $\vec{x}$ in this space can thus be expressed as

$$\vec{x} = r\hat{n}(\hat{\theta}) = r \cos(\theta^{D-1})\hat{e}_D + r \sin(\theta^{D-1})\hat{n}_{D-1}, \quad 0 \leq \theta^{D-1} \leq \pi.$$  

(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 $\hat{e}_D$ direction; while $r \sin(\theta^{D-1})$ is that along the radial direction in the plane perpendicular to $\hat{e}_D$.

First show that the Cartesian metric $\delta_{ij}$ in $D$-space transforms to

$$(d\ell)^2 = dr^2 + r^2 d\Omega_D^2 = dr^2 + r^2 \left((d\theta^{D-1})^2 + (\sin \theta^{D-1})^2 d\Omega_{D-1}^2\right),$$

where $d\Omega_D^2$ is the square of the infinitesimal solid angle in $N$ spatial dimensions, and is given by

$$d\Omega_D^2 \equiv \sum_{I,J=1}^{N-1} \Omega_{1J}^{(N)} d\theta^I d\theta^J, \quad \Omega_{1J}^{(N)} \equiv \sum_{i,j=1}^N \delta_{ij} \frac{\partial \hat{n}_N}{\partial \theta^i} \frac{\partial \hat{n}_N}{\partial \theta^j}. \quad (7.2.38)$$

Proceed to argue that the full $D$-metric in spherical coordinates is

$$d\ell^2 = dr^2 + r^2 \left((d\theta^{D-1})^2 + \sum_{I=2}^{D-1} s_{D-I+1}^2 (d\theta^{D-1})^2\right),$$

$$\theta^1 \in [0, 2\pi), \quad \theta^2, \ldots, \theta^{D-1} \in [0, \pi]. \quad (7.2.39)$$

(Here, $s_1 \equiv \sin \theta^1$.) Show that the determinant of the angular metric $\Omega_{1J}^{(N)}$ obeys a recursion relation

$$\det \Omega_{1J}^{(N)} = (\sin \theta^{N-1})^{2(N-2)} \cdot \det \Omega_{1J}^{(N-1)}. \quad (7.2.41)$$

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 $d^D\vec{x}$ in Cartesian coordinates then becomes, in spherical coordinates,

$$d^D\vec{x} = dr \cdot r^{D-1} \cdot d\theta^1 \ldots d\theta^{D-1} (\sin \theta^{D-1})^{D-2} \sqrt{\det \Omega_{1J}^{(D-1)}}. \quad (7.2.42)$$

Bonus: Can you use these results to find the solid angle subtended by a unit sphere in $D$ spatial dimensions? (Hint: You may find this page useful.)

Symmetries (aka Isometries), infinitesimal displacements, Killing vectors In some Cartesian coordinates $\{x^i\}$ the flat space metric is $\delta_{ij} dx^i dx^j$. Suppose we chose a different set of axes for new Cartesian coordinates $\{x'^n\}$, the metric will still take the same form, namely $\delta_{ij} dx'^i dx'^j$. Likewise, on a 2-sphere the metric is $d\theta^2 + (\sin \theta)^2 d\phi^2$ 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 $(\theta', \phi')$, the 2-sphere metric is still of the same form $d\theta'^2 + (\sin \theta')^2 d\phi'^2$. 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 – 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 $\vec{\xi}(\vec{x})$,

$$\vec{x}' \equiv \vec{x} - \vec{\xi}(\vec{x}). \tag{7.2.43}$$

(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 $\vec{\xi}$ is assumed to be “small,” we may replace in the above equation, $\vec{\xi}(\vec{x})$ with $\vec{\xi}'(\vec{x} \to \vec{x}')$. This is because the error incurred would be of $O(\xi^2)$.

$$\vec{x} = \vec{x}' + \vec{\xi}(\vec{x}') + O(\xi^2) \Rightarrow \frac{\partial x^i}{\partial x'^a} = \delta^i_a + \partial_a \xi^i(\vec{x}') + O(\partial \xi) \tag{7.2.44}$$

How does this change our metric?

$$g_{ij}(\vec{x}) \, dx^i dx^j = g_{ij}(\vec{x}') \left( \vec{x}' + \vec{\xi}(\vec{x}') + \ldots \right) \left( \delta^a_i + \partial_a \xi^i + \ldots \right) \left( \delta^b_j + \partial_b \xi^j + \ldots \right) dx^a dx^b$$
$$= (g_{ij}(\vec{x}') + \xi^i \partial_j g_{ij}(\vec{x}') + \ldots) \left( \delta^a_i + \partial_a \xi^i + \ldots \right) \left( \delta^b_j + \partial_b \xi^j + \ldots \right) dx^a dx^b$$
$$= (g_{ij}(\vec{x}') + \delta \xi g_{ij}(\vec{x}') + O(\xi^2)) \, dx^i dx^j, \tag{7.2.45}$$

where

$$\delta \xi g_{ij}(\vec{x}') \equiv \xi^e(\vec{x}') \frac{\partial g_{ij}(\vec{x}')}{\partial x^e} + g_{ia}(\vec{x}') \frac{\partial \xi^a(\vec{x}')}{\partial x^j} + g_{ja}(\vec{x}') \frac{\partial \xi^a(\vec{x}')}{\partial x^i}. \tag{7.2.46}$$

A point of clarification might be helpful. In eq. (7.2.45), we are not merely asking “What is $d\ell^2$ at $\vec{x}'$?” The answer to that question would be $d\ell^2 = g_{ij}(\vec{x} - \vec{\xi}(\vec{x})) \, dx^i dx^j$, with no need to transform the $dx^i$. Rather, here, we are performing a coordinate transformation from $\vec{x}$ to $\vec{x}'$, induced by an infinitesimal displacement via $\vec{x}' = \vec{x} - \vec{\xi}(\vec{x}) \leftrightarrow \vec{x} = \vec{x}' + \vec{\xi}(\vec{x}') + \ldots$ – 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 $\epsilon$. 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 $g_{ij}(\vec{x}) \, dx^i dx^j = g_{ij}(\vec{x}') \, dx'^i dx'^j$ and therefore

$$\delta \xi g_{ij} = 0, \quad \text{isometry along } \vec{\xi}. \quad \tag{7.2.47}$$

60 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.

61 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 $g_{33'}(r', \theta') = (r' \sin \theta')^2$ in the 2-sphere metric.
Conversely, if $\delta_\xi g_{ij} = 0$ 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 retains its form along its entirety. Now, a vector $\vec{\xi}$ that satisfies $\delta_\xi g_{ij} = 0$ is called a Killing vector and eq. (7.2.47) 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. (7.2.47) everywhere in space.

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 (7.2.46) and (7.2.47) 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 (7.15) for an answer.

Problem 7.6. Can you justify the statement: “If the metric $g_{ij}$ is independent of one of the coordinates, say $x^k$, then $\partial_k$ is a Killing vector of the geometry”?

Problem 7.7. Angular momentum ‘generators’ & 2-sphere isometries

The generators of rotation in 3D space are proportional to the following vectors:

\[ J^x = \sin(\phi)\partial_\theta + \cos(\phi)\cot(\theta)\partial_\phi, \]
\[ J^y = -\cos(\phi)\partial_\theta + \sin(\phi)\cot(\theta)\partial_\phi, \]
\[ J^z = \partial_\phi. \]  

(See §(4.5.6) for a discussion. Briefly: $J^x$ generates rotations on the $(y, z)$ plane; $J^y$ on the $(x, z)$ plane; and $J^z$ on the $(x, y)$ plane.) Verify directly that they satisfy the Killing equation (7.2.47) on the metric of the unit 2–sphere centered at $\vec{x} = 0$ in 3D flat space:

\[ d\ell^2 = d\theta^2 + (\sin \theta)^2 d\phi^2. \]  

Since these $(J^x, J^y, J^z)$ produce infinitesimal rotations, this problem is a direct verification that the geometry of the 2–sphere is invariant under rotations.

Remark The maximum number of linearly independent Killing vectors in $D$ dimensions is $D(D+1)/2$. See Chapter 13 of Weinberg’s *Gravitation and Cosmology* [20] for a discussion.

Orthonormal frame So far, we have been writing tensors in the coordinate basis – the basis vectors of our tensors are formed out of tensor products of $\{dx^i\}$ 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.

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 $\vec{x} = (x, y)$ flowing in the $x$- and $y$-directions respectively. Suppose we used polar coordinates, however,

\[ \xi^i = r(\cos \phi, \sin \phi). \]

$^{62}$\(\delta_\xi g_{ij}\) is known as the Lie derivative of the metric along $\xi$, and is more commonly denoted as $(\mathcal{L}_\xi g)_{ij}$. 

223
The metric would read
\[(d\ell)^2 = dr^2 + r^2d\phi^2.\] (7.2.53)

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 \[v^\phi\] is not even the same as that of \[v^r;\] if \[v^r\] were of \([\text{Length/Time}]\), then \[v^\phi\] is of \([1/\text{Time}]\). At this point we recall – just as \[dr\] (which is dual to \[\partial_r\]) can be interpreted as an infinitesimal length in the radial direction, the arc length \[r d\phi\] (which is dual to \((1/r)\partial_\phi\)) is the corresponding one in the perpendicular azimuthal direction. Moreover,
\[\langle \partial_r \mid \partial_r \rangle = g_{rr} = 1 \quad \text{and} \quad \langle r^{-1} \partial_\phi \mid r^{-1} \partial_\phi \rangle = r^{-2} \langle \partial_\phi \mid \partial_\phi \rangle = \frac{g_{\phi\phi}}{r^2} = 1.\] (7.2.54)

Using the above considerations as a guide, we would now express the velocity at \(\vec{\xi}\) as
\[v = v^r \frac{\partial}{\partial r} + (r \cdot v^\phi) \left(\frac{1}{r} \frac{\partial}{\partial \phi}\right) = \frac{dr}{d\lambda} \frac{\partial}{\partial r} + \left(r \cdot \frac{d\phi}{d\lambda}\right) \left(\frac{1}{r} \frac{\partial}{\partial \phi}\right),\] (7.2.55)
\[\text{so that now } v^\phi \equiv r \cdot v^\phi \text{ may be interpreted as the velocity in the azimuthal direction.}\]

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:
\[g_{ij} dx^i dx^j = \sum_{a,b} (O^a_i \cdot \lambda_a \delta_{ab} \cdot O^b_j) dx^i dx^j\]
\[= \sum_{a,b} \left(\sqrt{\lambda_a} O^a_i \cdot \lambda_a O^b_j\right) dx^i dx^j\]
\[= \left(\delta_{ab} \varepsilon^a_i \varepsilon^b_j\right) dx^i dx^j = \delta_{ab} \left(\varepsilon^a_i dx^i\right) \left(\varepsilon^b_j dx^j\right),\] (7.2.57)
\[\varepsilon^a_i \equiv \sqrt{\lambda_a} O^a_i, \quad \text{(no sum over } a).\] (7.2.58)

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 the second we have exploited the assumption that we are working in Riemannian spaces, where all eigenvalues of the metric are positive\(^{63}\) to take the positive square roots of the eigenvalues; in the third we have defined the orthonormal frame vector fields as \(\varepsilon^a_i = \sqrt{\lambda_a} O^a_i\), with no sum over \(a\). Finally, from eq. (7.2.57) and by defining the infinitesimal lengths \(\varepsilon^a_i \equiv \varepsilon^a_i dx^i\), we arrive at the following curved space parallel to Pythagoras’ theorem in flat space:
\[(d\ell)^2 = g_{ij} dx^i dx^j = (\varepsilon^1)^2 + (\varepsilon^2)^2 + \cdots + (\varepsilon^D)^2.\] (7.2.59)

\(^{63}\)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
\[ g_{ij} = \delta_{ab} \varepsilon_{i}^{\hat{a}} \varepsilon_{j}^{\hat{b}}, \]  
(7.2.60)

Whereas the metric determinant reads
\[ \det g_{ij} = (\det \varepsilon_{i}^{\hat{a}})^2. \]  
(7.2.61)

We say the metric on the right hand side of eq. (7.2.57) is written in an orthonormal frame, because in this basis \( \{\varepsilon_{a}^i dx^i | a = 1, 2, \ldots, D\} \), the metric components are identical to the flat Cartesian ones. We have put a \( \hat{\cdot} \) over the \( a \)-index, to distinguish from the \( i \)-index, because the latter transforms as a tensor
\[ \varepsilon_{i}^{\hat{a}}(\xi) = \varepsilon_{i}^{\hat{a}} \left( \frac{\partial x^j(\xi)}{\partial \xi^i} \right). \]  
(7.2.62)

This also implies the \( i \)-index can be moved using the metric:
\[ \varepsilon_{i}^{\hat{a}}(\xi) = g_{ij}(\xi) \varepsilon_{j}^{\hat{a}}(\xi), \quad \varepsilon_{i}^{\hat{a}}(\xi) = g_{ij}(\xi) \varepsilon_{j}^{\hat{a}}(\xi). \]  
(7.2.63)

We may readily check that eq. (7.2.62) is the correct transformation rule because it is equivalent to eq. (7.2.5).
\[ g_{\hat{a}\hat{b}}(\xi) = \delta_{mn} \varepsilon_{\hat{a}}^{\hat{m}}(\xi) \varepsilon_{\hat{b}}^{\hat{n}}(\xi) = \delta_{mn} \varepsilon_{i}^{\hat{m}}(\xi) \varepsilon_{j}^{\hat{n}}(\xi) \frac{\partial x^i}{\partial \xi^m} \frac{\partial x^j}{\partial \xi^n} = g_{ij}(\xi) \frac{\partial x^i}{\partial \xi^a} \frac{\partial x^j}{\partial \xi^b}. \]  
(7.2.64)

The \( \hat{a} \) index does not transform under coordinate transformations. But it can be rotated by an orthogonal matrix \( \tilde{R}_{\hat{a}\hat{b}}(\xi) \), which itself can depend on the space coordinates, while keeping the metric in eq. (7.2.57) the same object. By orthogonal matrix, we mean any \( R \) that obeys
\[ \tilde{R}_{\hat{a}\hat{b}} \varepsilon_{\hat{b}} \varepsilon_{\hat{b}} \hat{f} = \delta_{cf} \]  
(7.2.65)
\[ \tilde{R}^T \tilde{R} = I. \]  
(7.2.66)

Upon the replacement
\[ \varepsilon_{i}^{\hat{a}}(\xi) \rightarrow \tilde{R}_{\hat{a}\hat{b}}(\xi) \varepsilon_{i}^{\hat{b}}(\xi), \]  
(7.2.67)
we have
\[ g_{ij} dx^i dx^j \rightarrow \left( \delta_{ab} \tilde{R}_{\hat{a}\hat{b}} \tilde{R}_{\hat{a}\hat{b}} \right) \varepsilon_{i}^{\hat{a}} \varepsilon_{j}^{\hat{b}} dx^i dx^j = g_{ij} dx^i dx^j. \]  
(7.2.68)

The interpretation of eq. (7.2.67) 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 \( R \) obeying \( R^T R = I \), these local axes can also be rotated freely. 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
\[ V_{\hat{a}} = \tilde{R}_{\hat{a}\hat{b}} V_{\hat{b}}, \quad V_{\hat{a}} = (\tilde{R}^{-1})_{\hat{a}} \tilde{V}_{\hat{f}} \]  
(7.2.69)
so that
\[ V^\hat{a} V_{\hat{a}'} = V^\hat{a} V_{\hat{a}}. \] (7.2.70)

But \( \hat{R}^T \hat{R} = I \) means \( \hat{R}^{-1} = \hat{R}^T \) and thus the \( \hat{a} \)th row and \( \hat{c} \)th column of the inverse, namely \( (\hat{R}^{-1})^\hat{a}_{\hat{c}} \), is equal to the \( \hat{c} \)th row and \( \hat{a} \)th column of \( \hat{R} \) itself: \( (\hat{R}^{-1})^\hat{a}_{\hat{c}} = \hat{R}^\hat{c}_{\hat{a}} \).

\[ V_{\hat{a}'} = \sum_f \hat{R}^\hat{a}_f V_f. \] (7.2.71)

In other words, \( V_{\hat{a}} \) transforms just like \( V^\hat{a} \).

To sum, we have shown that the orthonormal frame index is moved by the Kronecker delta;
\[ V_{\hat{a}'} = V^\hat{a} \text{ for any vector written in an orthonormal frame, and in particular,} \]
\[ \varepsilon^\hat{a}_i(\vec{x}) = \delta^{\hat{a}\hat{b}} \varepsilon^\hat{b}_i(\vec{x}) = \varepsilon_{\hat{a}i}(\vec{x}). \] (7.2.72)

Next, we also demonstrate that these vector fields are indeed of unit length.
\[ \varepsilon^\hat{b}_j \varepsilon^\hat{a}_i \varepsilon^\hat{b}_j = \varepsilon^\hat{b}_j \varepsilon^\hat{a}_i \varepsilon^\hat{a}_j = \varepsilon^\hat{a}_i. \] (7.2.73)

To understand this we begin with the diagonalization of the metric, \( \delta_{\hat{c}\hat{f}} \varepsilon^\hat{c}_i \varepsilon^\hat{f}_i = g_{ij} \). Contracting both sides with the orthonormal frame vector \( \varepsilon^\hat{b}_i \),
\[ \delta_{\hat{c}\hat{f}} \varepsilon^\hat{c}_i \varepsilon^\hat{f}_i \varepsilon^\hat{b}_j = \varepsilon^\hat{b}_i, \] (7.2.75)
\[ (\varepsilon^\hat{b}_j \varepsilon^\hat{f}_j) \varepsilon^\hat{b}_i = \varepsilon^\hat{b}_i. \] (7.2.76)

If we let \( M \) denote the matrix \( M^\hat{f}_j \equiv (\varepsilon^\hat{b}_j \varepsilon^\hat{f}_j) \), then we have \( i = 1, 2, \ldots, D \) matrix equations \( M \cdot \varepsilon_i = \varepsilon_i \). As long as the determinant of \( g_{\hat{a}\hat{b}} \) is non-zero, then \( \{\varepsilon_i\} \) are linearly independent vectors spanning \( \mathbb{R}^D \) (see eq. (7.2.61)). Since every \( \varepsilon_i \) is an eigenvector of \( M \) with eigenvalue one, that means \( M = I \), and we have proved eq. (7.2.73).

To summarize,
\[ g_{ij} = \delta_{\hat{a}\hat{b}} \varepsilon^\hat{a}_i \varepsilon^\hat{b}_j, \quad g^{ij} = \delta^{\hat{a}\hat{b}} \varepsilon^\hat{a}_i \varepsilon^\hat{b}_j, \]
\[ \delta_{\hat{a}\hat{b}} = g_{ij} \varepsilon^\hat{a}_i \varepsilon^\hat{b}_j, \quad \delta^{\hat{a}\hat{b}} = g^{ij} \varepsilon^\hat{a}_i \varepsilon^\hat{b}_j. \] (7.2.77)

**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 \( \varepsilon^\hat{a}_i \) in eq. (7.2.57). For example, the velocity field in an orthonormal frame is
\[ v^\hat{a} = \varepsilon^\hat{a}_i v^i. \] (7.2.78)

For the two dimension example above,
\[ (dr)^2 + (rd\phi)^2 = \delta_{rr}(dr)^2 + \delta_{r\phi}(rd\phi)^2, \] (7.2.79)
allowing us to read off the only non-zero components of the orthonormal frame fields are

\[ \varepsilon_r^\hat{r} = 1, \quad \varepsilon_\phi^\hat{\phi} = r; \quad (7.2.80) \]

which in turn implies

\[ v^\hat{r} = \varepsilon_r^\hat{r} v^r, \quad v^\hat{\phi} = \varepsilon_\phi^\hat{\phi} v^\phi = r v^\phi. \quad (7.2.81) \]

More generally, what we are doing here is really switching from writing the same tensor in coordinates basis \( \{dx^i\} \) and \( \{\partial_i\} \) to an orthonormal basis \( \{\varepsilon_a^\hat{a} dx^i\} \) and \( \{\varepsilon_i^a \partial_a\} \). For example,

\[ T_{ijk}^l \langle dx^i \rangle \otimes \langle dx^j \rangle \otimes \langle dx^k \rangle \otimes |\partial_l\rangle = T_{\hat{a} \hat{b} \hat{c}}^{\hat{l}} \langle \varepsilon_{\hat{a}}^i \rangle \otimes \langle \varepsilon_{\hat{b}}^j \rangle \otimes \langle \varepsilon_{\hat{c}}^k \rangle \otimes |\varepsilon_{\hat{l}}\rangle \quad (7.2.82) \]

\[ \varepsilon_i^a \equiv \varepsilon_\hat{a}^i dx^a, \quad \varepsilon_i^a \partial_a. \quad (7.2.83) \]

To sum: the formula that converts a general tensor in a coordinate basis to the same in an orthonormal one is

\[ T_{\hat{a}_1 \ldots \hat{a}_M}^{\hat{b}_1 \ldots \hat{b}_N} = T^{i_1 \ldots i_M}_{\hat{b}_1 \ldots \hat{b}_N} \varepsilon_{\hat{a}_1}^{i_1} \ldots \varepsilon_{\hat{a}_M}^{i_M} \varepsilon_{\hat{b}_1}^{j_1} \ldots \varepsilon_{\hat{b}_N}^{j_N}. \quad (7.2.84) \]

**Problem 7.8.** Explain why the ‘inverse’ transformation of eq. (7.2.84) is

\[ T_{\hat{a}_1 \ldots \hat{a}_M}^{\hat{b}_1 \ldots \hat{b}_N} \varepsilon_{\hat{a}_1}^{i_1} \ldots \varepsilon_{\hat{a}_M}^{i_M} \varepsilon_{\hat{b}_1}^{j_1} \ldots \varepsilon_{\hat{b}_N}^{j_N} = T_{ijk}^l \langle dx^i \rangle \otimes \langle dx^j \rangle \otimes \langle dx^k \rangle \otimes |\partial_l\rangle. \quad (7.2.85) \]

Hint: Insert eq. (7.2.84) into the left hand side of (7.2.85), followed by using the appropriate relation in eq. (7.2.77).

Even though the physical dimension of the whole tensor \([T]\) is necessarily consistent, because the \( \{dx^i\} \) and \( \{\partial_i\} \) do not have the same dimensions – compare, for e.g., \( dr \) versus \( d\theta \) 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. (7.2.83), we see that

\[ \sum_a (\varepsilon_a^\hat{a})^2 = \delta_{ab} \varepsilon_a^\hat{a} \varepsilon_b^\hat{b} dx^i dx^j g_{ij} = g_{ij} dx^i dx^j \quad (7.2.86) \]

\[ [\varepsilon_a^\hat{a}] = \text{Length}; \quad (7.2.87) \]

and

\[ \sum_a (\varepsilon_a^i)^2 = \delta^{ab} \varepsilon_a^i \varepsilon_b^j \partial_i \partial_j = g^{ij} \partial_i \partial_j \quad (7.2.88) \]

\[ [\varepsilon_a^i] = 1/\text{Length}; \quad (7.2.89) \]

which in turn implies, for instance, the consistency of the physical dimensions of the orthonormal components \( T_{ijk}^l \) in eq. (7.2.82):

\[ [T_{ijk}^l] [\varepsilon_i^\hat{a}]^3 [\varepsilon_j^\hat{b}] = [T], \quad (7.2.90) \]

\[ [T_{ijk}^l] = \frac{[T]}{\text{Length}^2}. \quad (7.2.91) \]
Problem 7.9. Find the orthonormal frame fields \( \{ \hat{e}_i^a \} \) in 3-dimensional Cartesian, Spherical and Cylindrical coordinate systems. Hint: Just like the 2D case above, by packaging the metric \( g_{ij}dx^i dx^j \) appropriately, you can read off the frame fields without further work.

(Curved) Dot Product So far we have viewed the metric \((d\ell)^2\) as the square of the distance between \(\vec{x}\) and \(\vec{x} + d\vec{x}\), generalizing Pythagoras’ theorem in flat space. The generalization of the dot product between two (tangent) vectors \(U\) and \(V\) at some location \(\vec{x}\) is
\[
U(\vec{x}) \cdot V(\vec{x}) \equiv g_{ij}(\vec{x}) U^i(\vec{x}) V^j(\vec{x}).
\] (7.2.92)
(We have already noted above, this defines a coordinate scalar.) That this is in fact the analogy of the dot product in Euclidean space can be readily seen by going to the orthonormal frame:
\[
U(\vec{x}) \cdot V(\vec{x}) = \delta_{ij} \hat{U}_i(\vec{x}) \hat{V}_j(\vec{x}).
\] (7.2.93)

Line integral The line integral that occurs in 3D vector calculus, is commonly written as \(\int \vec{A} \cdot d\vec{x}\). While the dot product notation is very convenient and oftentimes quite intuitive, there is an implicit assumption that the underlying coordinate system is Cartesian in flat space. The integrand that actually transforms covariantly is the tensor \(A_i dx^i\), where the \(\{x^i\}\) are no longer necessarily Cartesian. The line integral itself then consists of integrating this over a prescribed path \(\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)\), namely
\[
\int_{\vec{x}(\lambda_1 \leq \lambda \leq \lambda_2)} A_i dx^i = \int_{\lambda_1}^{\lambda_2} A_i(\vec{x}(\lambda)) \frac{dx^i(\lambda)}{d\lambda} d\lambda. \tag{7.2.94}
\]

7.3 Covariant derivatives, Parallel Transport, Geodesics, Levi-Civita, Hodge Dual

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
\[
\frac{\partial T_j(\vec{x})}{\partial \xi^i} = \frac{\partial x^a}{\partial \xi^i} \frac{\partial}{\partial x^a} \left( T_b(\vec{x}(\xi)) \frac{\partial x^b}{\partial \xi^j} \right)
\]
\[
= \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} \frac{\partial T_b(\vec{x}(\xi))}{\partial x^a} + \frac{\partial^2 x^b}{\partial \xi^i \partial \xi^j} T_b(\vec{x}(\xi)). \tag{7.3.1}
\]
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 \(\nabla\), 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,
\[
\Gamma^i_{jk} = \frac{1}{2} g^{il} (\partial_j g_{kl} + \partial_k g_{jl} - \partial_l g_{jk}). \tag{7.3.2}
\]
Notice the Christoffel symbol is symmetric in its lower indices: \(\Gamma^i_{jk} = \Gamma^i_{kj}\).
Comparing equations (7.3.9) and (7.3.10) leads us to relate the Christoffel symbol written in \( \vec{\xi} \) coordinates \( \Gamma^{ij}_{\ell} \) to the one written in \( \vec{x} \) coordinates \( \Gamma^{ij}_{\ell} \).

On the other hand,

\[
\nabla_{\xi} T_j (\xi) = \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} T_a (\xi) \nabla_{\xi} T_b (\xi).
\]

For a general \((N)\) tensor, we have

\[
\nabla_{\xi} T_{i_1 \ldots i_N} (\xi) = \frac{\partial x^a}{\partial \xi^i} \frac{\partial x^b}{\partial \xi^j} T_{i_1 \ldots i_N} (\xi) \nabla_{\xi} T_{a} (\xi).
\]

By using eq. (7.3.1) we may infer how the Christoffel symbols themselves must transform — they are not tensors. Firstly,

\[
\nabla_{\xi} T_j (\xi) = \partial_{\xi} T_j (\xi) - \Gamma^l_{ij}(\xi) T_l (\xi).
\]

Comparing equations (7.3.9) and (7.3.10) leads us to relate the Christoffel symbol written in \( \vec{\xi} \) coordinates \( \Gamma^{ij}_{\ell} \) and that written in \( \vec{x} \) coordinates \( \Gamma^i_{\ell j} \).
On the right hand side, all $\vec{x}$ have been replaced with $\vec{x}(\xi)$.\footnote{We note in passing that in gauge theory – which encompasses humanity’s current description of the non-gravitational forces (electromagnetic-weak (SU$_2$) left-handed fermions $\times$ (U$_1$) hypercharge and strong nuclear (SU$_3$) color) – the fundamental fields there $\{A^a_{\mu}\}$ transforms (in a group theory sense) in a very similar fashion as the Christoffel symbols do (under a coordinate transformation) in eq. (7.3.11).}

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

$$\nabla (T_1 T_2) = (\nabla T_1) T_2 + T_1 (\nabla T_2).$$

(7.3.12)

Unlike partial derivatives, repeated covariant derivatives do not commute; hence, make sure you keep track of the order of operations. For instance,

$$\nabla_a \nabla_b T^{ij} \neq \nabla_b \nabla_a T^{ij}.$$  

(7.3.13)

As you will see below, the metric is parallel transported in all directions,

$$\nabla_i g_{jk} = \nabla_i g_{jk} = 0.$$  

(7.3.14)

Combined with the product rule in eq. (7.3.12), 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 $\nabla$. For example,

$$g_{ia} \nabla_j T^{kal} = \nabla_j g_{ia} \cdot T^{kal} + g_{ia} \nabla_j T^{kal} = \nabla_j (g_{ia} T^{kal})$$

$$= \nabla_j T^{k \ell}_{\ i}.$$  

(7.3.15)

\textbf{Remark} 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 $\{\langle dx^i \rangle\}$ and kets $\{|\partial_i\rangle\}$ as they are moved from one point in space to a neighboring point infinitesimally far away. Specifically,

$$\nabla_i \langle dx^j \rangle = -\Gamma^j_{\ ik} \langle dx^k \rangle \quad \text{and} \quad \nabla_i |\partial_j\rangle = \Gamma^l_{\ ij} |\partial_l\rangle.$$  

(7.3.16)

Within this perspective, the tensor components are scalars. The product rule then yields, for instance,

$$\nabla_i (V_a \langle dx^a \rangle) = (\nabla_i V_a) \langle dx^a \rangle + V_a \nabla_i \langle dx^a \rangle$$

$$= (\partial_i V_a - V_a \Gamma^a_{\ ij}) \langle dx^j \rangle.$$  

(7.3.17)

\textbf{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 $T^{j_1...j_N}$ 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 $v$ when

$$v^i \nabla_i T^{j_1...j_N} = 0.$$  

(7.3.18)
Problem 7.10. As an example, let’s calculate the Christoffel symbols of the metric on the 2-sphere with unit radius,

$$(\diff{\ell})^2 = \diff{\theta}^2 + (\sin \theta)^2 \diff{\phi}^2.$$  

(7.3.19)

Do not calculate from scratch – remember you have already computed the Christoffel symbols in 3D Euclidean space. How do you extract the 2-sphere Christoffel symbols from that calculation?

In the coordinate system $(\theta, \phi)$, 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

$$v^i \nabla_i v^j = \nabla_\theta v^j = 0.$$  

(7.3.20)

Also calculate $\nabla_\phi v^j$; comment on the result at $\theta = \pi/2$. Hint: recall our earlier 2-sphere discussion, where we considered parallel transporting a tangent vector from the North pole to the equator, along the equator, then back up to the North pole.

Riemann and Ricci tensors I will not use them very much in the rest of our discussion in this section (§7), 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

$$R^i_{jkl} = \partial_k \Gamma^i_{lj} - \partial_l \Gamma^i_{kj} + \Gamma^i_{sk} \Gamma^s_{lj} - \Gamma^i_{sl} \Gamma^s_{kj}.$$  

(7.3.21)

The failure of parallel transport of some vector $V^i$ around an infinitesimally small loop, is characterized by

$$[\nabla_k, \nabla_l]V^i \equiv (\nabla_k \nabla_l - \nabla_l \nabla_k)V^i = R^i_{jkl} V^j,$$  

$$[\nabla_k, \nabla_l]V^j \equiv (\nabla_k \nabla_l - \nabla_l \nabla_k)V^j = -R^i_{jkl} V_i.$$  

(7.3.22)

(7.3.23)

The generalization to higher rank tensors is

$$[\nabla_i, \nabla_j]T^{k_1 \ldots k_N}_{l_1 \ldots l_M} = R^{k_1}_{aij} T^{a k_2 \ldots k_N}_{l_1 \ldots l_M} + R^{k_2}_{aij} T^{k_1 a k_3 \ldots k_N}_{l_1 \ldots l_M} + \cdots + R^{k_N}_{aij} T^{k_1 \ldots k_{N-1} a}_{l_1 \ldots l_M} - R^a_{l_1 ij} T^{k_1 \ldots k_N}_{a l_2 \ldots l_M} - R^a_{l_2 ij} T^{k_1 \ldots k_N}_{l_1 a l_3 \ldots l_M} - \cdots - R^a_{l_M ij} T^{k_1 \ldots k_N}_{l_1 \ldots l_{M-1} a}.$$  

(7.3.24)

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.

$$R_{ijab} = R_{abij}, \quad R_{ijab} = -R_{jiba}, \quad R_{abij} = -R_{abji}.$$  

(7.3.25)

The Riemann tensor also obeys the Bianchi identities$^{[66]}$

$$R^i_{[jkl]} = \nabla_i R^{jk}_{lm} = 0.$$  

(7.3.26)

$^{[66]}$The symbol $[\ldots]$ 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.
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 $R_{1111} = -R_{1111} = 0$.

The Ricci tensor is defined as the non-trivial contraction of a pair of the Riemann tensor’s indices.

$$R_{jl} \equiv R^i_{jil}. \quad (7.3.27)$$

It is symmetric

$$R_{ij} = R_{ji}. \quad (7.3.28)$$

Finally the Ricci scalar results from a contraction of the Ricci tensor’s indices.

$$\mathcal{R} \equiv g^{jl}R_{jl}. \quad (7.3.29)$$

Contracting eq. (7.3.26) appropriately yields the Bianchi identities involving the Ricci tensor and scalar

$$\nabla^i \left( R_{ij} - \frac{g_{ij}}{2} \mathcal{R} \right) = 0. \quad (7.3.30)$$

This is a good place to pause and state, the Christoffel symbols in eq. (7.3.2), 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. (7.1.25). 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. (7.1.25). This is a ‘calculus of variation’ problem. The argument runs as follows. Suppose you found the path $\vec{z}(\lambda)$ that yields the shortest $\ell$. Then, if you consider a slight variation $\delta \vec{z}$ of the path, namely consider

$$\vec{z}(\lambda) = \vec{z}(\lambda) + \delta \vec{z}(\lambda), \quad (7.3.31)$$

we must find the contribution to $\ell$ 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. In other words, in the integrand of eq. (7.1.25) we must replace

$$g_{ij} \left( \vec{x}(\lambda) \right) \rightarrow g_{ij} \left( \vec{z}(\lambda) + \delta \vec{z}(\lambda) \right) = g_{ij} \left( \vec{z}(\lambda) \right) + \delta z^k(\lambda) \frac{\partial g_{ij}(\vec{z}(\lambda))}{\partial z^k} + O(\delta z^2) \quad (7.3.32)$$

$$\frac{dx^i(\lambda)}{d\lambda} \rightarrow \frac{dz^i(\lambda)}{d\lambda} + \frac{d\delta z^i(\lambda)}{d\lambda}. \quad (7.3.33)$$

Since $\delta \vec{z}$ was arbitrary, at first order, its coefficient within the integrand must vanish. If we further specialize to affine parameters $\lambda$ – i.e., such that

$$\sqrt{g_{ij}(dz^i/d\lambda)(dz^j/d\lambda)} = \text{constant along the entire path } \vec{z}(\lambda) \quad (7.3.34)$$

---

67 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 \( \ell \) between \( \vec{x}_1 \) and \( \vec{x}_2 \) leads to the shortest path \( \vec{z}(\lambda) \) (\( \equiv \) geodesic) obeying:

\[
0 = \frac{d^2 z^i}{d\lambda^2} + \Gamma^i_{jk}(g_{ab}(\vec{z})) \frac{dz^j}{d\lambda} \frac{dz^k}{d\lambda},
\]

with the boundary conditions

\[
\vec{z}(\lambda_1) = \vec{x}_1, \quad \vec{z}(\lambda_2) = \vec{x}_2.
\]

You will verify this discussion in Problem (7.11) below.

The converse is also true, in that – if the geodesic equation in eq. (7.3.35) holds, then \( g_{ij} \left( \frac{dz^i}{d\lambda} \right) \left( \frac{dz^j}{d\lambda} \right) \) is a constant along the entire geodesic. Denoting \( \ddot{z}^i \equiv \frac{d^2 z^i}{d\lambda^2} \) and \( \dot{z}^i \equiv \frac{dz^i}{d\lambda} \),

\[
\frac{d}{d\lambda} \left( g_{ij} \ddot{z}^i \dot{z}^j \right) = 2\dot{z}^i \dot{z}^j g_{ij} + \dot{z}^k \partial_k g_{ij} \dot{z}^i \dot{z}^j \\
= 2\dot{z}^i \dot{z}^j g_{ij} + \dot{z}^k \dot{z}^j (\partial_k g_{ij} + \partial_i g_{kj} - \partial_j g_{ik})
\]

(7.3.37)

Note that the last two terms inside the parenthesis of the second equality cancels. The reason for inserting them is because the expression contained within the parenthesis is related to the Christoffel symbol; keeping in mind eq. (7.3.2),

\[
\frac{d}{d\lambda} \left( g_{ij} \ddot{z}^i \dot{z}^j \right) = 2\dot{z}^i \left\{ \dddot{z}^j g_{ij} + \dot{z}^k \ddot{z}^j g_{il} g^{lm} \left( \partial_k g_{jm} + \partial_j g_{km} - \partial_m g_{jk} \right) \right\} = 0.
\]

(7.3.38)

The last equality follows because the expression in the \( \{ \ldots \} \) is the left hand side of eq. (7.3.35). This constancy of \( g_{ij} \left( \frac{dz^i}{d\lambda} \right) \left( \frac{dz^j}{d\lambda} \right) \) is useful for solving the geodesic equation itself.

**Problem 7.11. Noether’s theorem for Lagrangian mechanics**

Show that the affine parameter form of the geodesic (7.3.35) follows from demanding the distance-squared integral of eq. (7.1.29) be extremized:

\[
\ell^2 = (\lambda_2 - \lambda_1) \int_{\lambda_1}^{\lambda_2} d\lambda g_{ij} \left( \vec{z}(\lambda) \right) \frac{dz^i}{d\lambda} \frac{dz^j}{d\lambda}.
\]

(7.3.39)

That is, show that eq. (7.3.35) follows from applying the Euler-Lagrange equations

\[
\frac{d}{d\lambda} \frac{\partial L}{\partial \dot{z}^i} = \frac{\partial L}{\partial z^i}
\]

(7.40)

to the Lagrangian

\[
L \equiv \frac{1}{2} g_{ij} \ddot{z}^i \dot{z}^j, \quad \dot{z}^i \equiv \frac{dz^i}{d\lambda}.
\]

(7.41)

\[\text{Some jargon: In the General Relativity literature, } \ell^2/2 \text{ (half of eq. (7.3.39)) is known as Synge's world function.}\]
If the geodesic equation (7.3.35) is satisfied by \( \vec{z}(\lambda) \), argue that the integral in eq. (7.3.39) yields the square of the geodesic distance between \( \vec{x}_1 \equiv \vec{z}(\lambda_1) \) and \( \vec{x}_2 \equiv \vec{z}(\lambda_2) \).

**Conserved quantities from symmetries** Finally, suppose \( \partial_k \) is a Killing vector. Explain why

\[
\frac{\partial L}{\partial \dot{z}^k} = \text{constant.} \tag{7.3.42}
\]

This is an example of Noether’s theorem. For example, in flat Euclidean space, since the metric in Cartesian coordinates is a constant \( \delta_{ij} \), all the \( \{\partial_i| i = 1, 2, \ldots, D\} \) are Killing vectors. Therefore, from \( L = (1/2)\delta_{ij}\dot{z}^i\dot{z}^j \), and we have

\[
\frac{d}{d\lambda} \frac{dz^i}{d\lambda} = 0 \quad \Rightarrow \quad \frac{dz^i}{d\lambda} = \text{constant.} \tag{7.3.43}
\]

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.

**Geodesics on a 2–sphere** How many geodesics are there joining any two points on the 2–sphere? How many geodesics are there joining the North Pole and South Pole? Solve the geodesic equation (cf. eq. (7.3.35)) on the unit 2–sphere described by

\[
d\ell^2 = d\theta^2 + \sin(\theta)^2d\phi^2. \tag{7.3.44}
\]

Explain how your answer would change if the sphere were of radius \( R \) instead. Hint: To solve the geodesic equation it helps to exploit the spherical symmetry of the problem; for e.g., which coordinate is the metric independent of?

**Christoffel symbols from Lagrangian** Instead of computing the Christoffel symbols using the formula in eq. (7.3.2), we may instead use the variational principle encoded eq. (7.3.39) to obtain its components. That is, starting from the Lagrangian in eq. (7.3.41), one may compute the geodesic equation (7.3.35) and read off \( \Gamma_{ab}^c \) as the coefficient of \( \dot{z}^a \dot{z}^b \) for \( a = b \); and half of the coefficient of \( \dot{z}^a \dot{\phi} \) for \( a \neq b \).

**Example I** As a first example, let us extract the Christoffel symbols of the 2D flat metric in polar coordinates

\[
d\ell^2 = dr^2 + r^2d\phi^2. \tag{7.3.45}
\]

The Lagrangian in eq. (7.3.41) is

\[
L_g = \frac{1}{2}r^2 + \frac{1}{2}r^2\dot{\phi}^2. \tag{7.3.46}
\]

The Euler-Lagrange equations are

\[
\frac{d}{d\lambda} \frac{\partial L_g}{\partial \dot{r}} = \frac{\partial L_g}{\partial r} \tag{7.3.47}
\]

\[
\ddot{r} = r\dot{\phi}^2 \tag{7.3.48}
\]
\[ r - r\dot{\theta}^2 = r + \Gamma^r_{\phi\phi} \dot{\phi}^2 = 0; \quad (7.3.49) \]

and
\[ \frac{d}{d\lambda} \frac{\partial L_g}{\partial \dot{\phi}} = \frac{\partial L_g}{\partial \phi} = 0; \quad (7.3.50) \]
\[ \frac{d}{d\lambda} (r^2 \ddot{\phi}) = 0; \quad (7.3.51) \]
\[ \ddot{\phi} + \frac{2}{r} \dot{r} \dot{\phi} = \ddot{\phi} + \Gamma^\phi_{r\phi} \dot{r} \dot{\phi} + \Gamma^\phi_{\phi r} \dot{\phi} \dot{r} = 0. \quad (7.3.52) \]

We see that \( \Gamma^r_{\phi\phi} = -r; \) whereas, due to its symmetric character, \( \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = 1/r. \) The latter is a technical point worth reiterating: for \( a \neq b, \) the coefficient of \( \dot{z}^a \dot{z}^b \) in the geodesic equation \( \ddot{z}^i + (\ldots) \dot{z}^a \dot{z}^b + \cdots = 0 \) is \textit{twice} of \( \Gamma^i_{ab}, \) because – with no sum over \( a \) and \( b – \)
\[ \Gamma^i_{ab} \dot{z}^a \dot{z}^b + \Gamma^i_{ba} \dot{z}^b \dot{z}^a = 2 \Gamma^i_{ab} \dot{z}^a \dot{z}^b. \quad (7.3.53) \]

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:
\[ \ell^2 \equiv a(\vec{x})^2 d\vec{x} \cdot d\vec{x}, \quad (7.3.54) \]
where \( a(\vec{x}) \) is an arbitrary function. The Lagrangian in eq. (7.3.41) is now
\[ L = \frac{1}{2} a^2 \delta_{ij} \dot{z}^i \dot{z}^j, \quad (7.3.55) \]
Applying the Euler-Lagrange equations,
\[ \frac{d}{d\lambda} \frac{\partial L}{\partial \dot{z}^i} - \frac{\partial L}{\partial z^i} = 0 \quad (7.3.56) \]
\[ \frac{d}{d\lambda} \left( a^2 \dot{z}^i \right) - a \ddot{z}^i = 0 \quad (7.3.57) \]
\[ 2a \dot{z}^i \partial_j a \dot{z}^j + a^2 \ddot{z}^i - a \partial_i a \dot{z}^2 = 0 \quad (7.3.58) \]
\[ \dot{z}^i + \left( \frac{\partial_j a}{a} \delta^i_j + \frac{\partial_i a}{a} \delta^j_i - \frac{\partial_i a}{a} \delta^j_k \right) \dot{z}^j \dot{z}^i = \ddot{z}^i + \Gamma^i_{lj} \dot{z}^j \dot{z}^j = 0. \quad (7.3.59) \]

Using \( \{\ldots\} \) to indicate symmetrization of the indices, we have derived
\[ \Gamma^i_{lj} = \frac{1}{a} \left( \partial_j a \delta^i_l - \partial_i a \delta^j_l \right) \quad (7.3.60) \]

Problem 7.12. 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
\[ p_i \equiv \frac{\partial L}{\partial \dot{z}^i} = g_{ij} \dot{z}^j, \quad (7.3.61) \]
where \(L\) is the Lagrangian in eq. (7.3.41); and further define the Hamiltonian \(H\) through the Legendre transform
\[
H(\vec{z}, \vec{p}) \equiv p_i \dot{z}^i(\vec{z}, \vec{p}) - L(\vec{z}, \vec{p}).
\]
(7.3.62)

This relation between \(H\) and \(L\) assumes all the \(\{\dot{z}^i \equiv \frac{dz^i}{d\lambda}\}\) 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
\[
H = \frac{1}{2} g^{ij} p_i p_j.
\]
(7.3.63)

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 \(\lambda\), the right hand side of eq. (7.3.62) is necessarily a constant of motion.)

**Geodesic Equations**

Show that Hamilton’s equations
\[
\begin{align*}
\frac{dz^i}{d\lambda} &= \frac{\partial H}{\partial p_i} = g^{ij} p_j, \\
\frac{dp_i}{d\lambda} &= -\frac{\partial H}{\partial z^i} = -\frac{1}{2} (\partial_i g^{ab}) p_a p_b
\end{align*}
\]
(7.3.64) (7.3.65)

are equivalent to the geodesic equation (7.3.35).

**Problem 7.13.** 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
\[
\Gamma^k_{ij}(\vec{y}_0) = 0.
\]
(7.3.66)

Can you demonstrate why this is true from the equivalence principle encoded in eq. (7.2.1)? 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. (7.2.57) – as well as the existence of a locally flat coordinate system, is why the measure of curvature, in particular the Riemann tensor in eq. (7.3.21), depends on first and second derivatives of the metric. Specifically, when eq. (7.3.66) holds but space is curved, we would have from eq. (7.3.21),
\[
R^i_{jmn}(\vec{y}_0) = \partial_m \Gamma^i_{nj}(\vec{y}_0) - \partial_n \Gamma^i_{mj}(\vec{y}_0).
\]
(7.3.67)

**Problem 7.14.** Christoffel \(\Gamma^i_{jk}\) contains as much information as \(\partial_i g_{ab}\). Why do the Christoffel symbols take on the form in eq. (7.3.2)? It comes from assuming that the Christoffel symbol obeys the symmetry \(\Gamma^i_{jk} = \Gamma^i_{kj}\) – this is the torsion-free condition – and demanding that the covariant derivative of a metric is a zero tensor,
\[
\nabla_i g_{jk} = 0.
\]
(7.3.68)
This can be expanded as
\[ \nabla_i g_{jk} = 0 = \partial_i g_{jk} - \Gamma^l_{ij} g_{lk} - \Gamma^l_{ik} g_{jl}. \] (7.3.69)

Expand also \( \nabla_j g_{ki} \) and \( \nabla_k g_{ij} \), and show that
\[ 2\Gamma^l_{ij} g_{lk} = \partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}. \] (7.3.70)

Divide both sides by 2 and contract both sides with \( g^{km} \) to obtain \( \Gamma^m_{ij} \) in eq. (7.3.2).

**Remark**  Incidentally, while eq. (7.3.2) tells us the Christoffel symbol can be written in terms of the first derivatives of the metric; eq. (7.3.69) indicates the first derivative of the metric can also always be expressed in terms of the Christoffel symbols. In other words, \( \partial g_{ab} \) contains as much information as \( \Gamma^i_{jk} \), provided of course that \( g_{ij} \) itself is known.

**Problem 7.15.** Can you show that the \( \delta \xi g_{ij} \) in eq. (7.2.46) can be re-written in a more covariant looking expression
\[ \delta \xi g_{ij}(\vec{x}') = \nabla_i \xi_j + \nabla_j \xi_i. \] (7.3.71)

\( \delta \xi g_{ij} = \nabla_i \xi_j + \nabla_j \xi_i = 0 \) is known as Killing’s equation, and a vector that satisfies Killing’s equation is called a Killing vector. Showing that \( \delta \xi g_{ij} \) is a tensor indicate 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 \( \partial_a \xi^i = \nabla_a \xi^i - \Gamma^i_{ab} \xi^b. \) 

**Problem 7.16.** Argue that, if a tensor \( T_{i_1i_2\ldots i_N} \) is zero in some coordinate system, it must be zero in any other coordinate system.

**Problem 7.17.** Prove that the tensor \( T_{i_1i_2\ldots i_N} \) is zero if and only if the corresponding tensor \( T_{i_1i_2\ldots i_N} \) 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 7.18.** Calculate the Christoffel symbols of the 3-dimensional Euclidean metric in Cartesian coordinates \( \delta_{ij} \). Then calculate the Christoffel symbols for the same space, but in spherical coordinates: \( (dl)^2 = dr^2 + r^2(d\theta^2 + (\sin \theta)^2 d\phi^2) \). To start you off, the non-zero components of the metric are
\[ g_{rr} = 1, \quad g_{\theta\theta} = r^2, \quad g_{\phi\phi} = r^2(\sin \theta)^2; \] (7.3.72)
\[ g^{rr} = 1, \quad g^{\theta\theta} = r^{-2}, \quad g^{\phi\phi} = \frac{1}{r^2(\sin \theta)^2}. \] (7.3.73)

Also derive the Christoffel symbols in spherical coordinates from their Cartesian counterparts using eq. (7.3.11). This lets you cross-check your results; you should also feel free to use software to help. Partial answer: the non-zero components in spherical coordinates are
\[ \Gamma^r_{\theta\theta} = -r, \quad \Gamma^r_{\phi\phi} = -r(\sin \theta)^2, \] (7.3.74)
\[ \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \frac{1}{r}, \quad \Gamma^\theta_{\phi\phi} = -\cos \theta \cdot \sin \theta, \] (7.3.75)
\[ \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{1}{r}, \quad \Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi \theta} = \cot \theta. \quad (7.3.76) \]

To provide an example, let us calculate the Christoffel symbols of 2D flat space written in cylindrical coordinates \( \xi^i \equiv (r, \phi) \),

\[ d\ell^2 = dr^2 + r^2 d\phi, \quad r \geq 0, \ \phi \in [0, 2\pi). \quad (7.3.77) \]

This means the non-zero components of the metric are

\[ g_{rr} = 1, \quad g_{\phi\phi} = r^2, \quad g^{rr} = 1, \quad g^{\phi\phi} = r^{-2}. \quad (7.3.78) \]

Keeping the diagonal nature of the metric in mind, let us start with

\[ \Gamma^r_{ij} = \frac{1}{2} g^{rk} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) = \frac{1}{2} g^{rr} (\partial_i g_{jr} + \partial_j g_{ir} - \partial_r g_{ij}) \]

\[ = \frac{1}{2} \left( \delta^r_j \partial_i g_{rr} + \delta^r_i \partial_j g_{rr} - \delta^r_i \delta^r_j \partial_r r^2 \right) = -\delta^r_i \delta^r_j r. \quad (7.3.79) \]

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

\[ \Gamma^\phi_{ij} = \frac{1}{2} g^{\phi\phi} (\partial_i g_{\phi j} + \partial_j g_{\phi i} - \partial_{\phi} g_{ij}) \]

\[ = \frac{1}{2r^2} \left( \delta^\phi_j \partial_i g_{\phi\phi} + \delta^\phi_i \partial_j g_{\phi\phi} \right) = \frac{1}{2r^2} \left( \delta^\phi_j \delta^r_i \partial_r r^2 + \delta^\phi_i \delta^r_j \partial_r r^2 \right) \]

\[ = \frac{1}{r} \left( \delta^\phi_j \delta^r_i + \delta^\phi_i \delta^r_j \right). \quad (7.3.80) \]

If we had started from Cartesian coordinates \( x^i \),

\[ x^i = r(\cos \phi, \sin \phi), \quad (7.3.81) \]

we know the Christoffel symbols in Cartesian coordinates are all zero, since the metric components are constant. If we wish to use eq. (7.3.11) to calculate the Christoffel symbols in \( (r, \phi) \), 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

\[ \frac{\partial x^i}{\partial \xi^j} = \begin{bmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{bmatrix}^i_j, \quad (7.3.82) \]

\[ \frac{\partial \xi^i}{\partial x^j} = \left( \frac{\partial x}{\partial \xi} \right)^{-1} = \begin{bmatrix} \cos \phi & \sin \phi \\ -r^{-1} \sin \phi & r^{-1} \cos \phi \end{bmatrix}^i_j, \quad (7.3.83) \]

whereas the second derivative matrices are

\[ \frac{\partial^2 x^1}{\partial \xi^i \xi^j} = \begin{bmatrix} 0 & \sin \phi \\ -\sin \phi & -r \cos \phi \end{bmatrix}, \quad (7.3.84) \]

\[ \frac{\partial^2 x^2}{\partial \xi^i \xi^j} = \begin{bmatrix} 0 & \cos \phi \\ \cos \phi & -r \sin \phi \end{bmatrix}. \quad (7.3.85) \]
Therefore, from eq. (7.3.11),
\[
\Gamma^r_{ij}(r, \phi) = \frac{\partial r}{\partial x^k} \frac{\partial x^k}{\partial \xi^i} \frac{\partial x^k}{\partial \xi^j} = \cos \phi \begin{bmatrix} 0 & -\sin \phi \\ -\sin \phi & -r \cos \phi \end{bmatrix} + \sin \phi \begin{bmatrix} 0 & \cos \phi \\ \cos \phi & -r \sin \phi \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -r \end{bmatrix}. \tag{7.3.86}
\]
Similarly,
\[
\Gamma^\phi_{ij}(r, \phi) = \frac{\partial \phi}{\partial x^k} \frac{\partial x^k}{\partial \xi^i} \frac{\partial x^k}{\partial \xi^j} = -r^{-1} \sin \phi \begin{bmatrix} 0 & -\sin \phi \\ -\sin \phi & -r \cos \phi \end{bmatrix} + r^{-1} \cos \phi \begin{bmatrix} 0 & \cos \phi \\ \cos \phi & -r \sin \phi \end{bmatrix} = \begin{bmatrix} 0 & r^{-1} \\ -r^{-1} & 0 \end{bmatrix}. \tag{7.3.87}
\]
At this juncture, we may summarize the 3 methods of calculating Christoffel symbols.

- Do it by brute force, using eq. (7.3.2).
- Use the Lagrangian method: apply the Euler-Lagrangian equations to the Lagrangian
  \[ L_g = \frac{1}{2} g_{ij} \dot{z}^i \dot{z}^j, \]
  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 can be obtained through the Hessian (second derivatives matrix) terms of eq. (7.3.11).

**Variation of the metric & divergence of tensors**

Suppose we perturb the metric slightly
\[
g_{ij} \rightarrow g_{ij} + h_{ij}, \tag{7.3.88}
\]
where the components of \( h_{ij} \) are to be viewed as “small”, and its indices are moved with the metric; for e.g.,
\[
h^i_j = g^{ia} h_{aj}. \tag{7.3.89}
\]
The inverse metric will become
\[
g^{ij} \rightarrow g^{ij} - h^{ij} + h^{ik} h_k^j + O(h^3), \tag{7.3.90}
\]
then the square root of the determinant of the metric will change as
\[
\sqrt{|g|} \rightarrow \sqrt{|g|} \left( 1 + \frac{1}{2} g^{ab} h_{ab} + O(h^2) \right). \tag{7.3.91}
\]

**Problem 7.19.** Use the matrix identity, where for any square matrix \( X \),
\[
\det e^X = e^{\text{Tr}[X]}, \tag{7.3.92}
\]
69 To prove eq. (7.3.91). (The \( \text{Tr} \) \( X \) means the trace of the matrix \( X \) – sum over its diagonal terms.) Hint: Start with \( \det(g_{ij} + h_{ij}) = \det(g_{ij}) \cdot \det(\delta_{ij} + h_{ij}) \). Then massage \( \delta_{ij} + h_{ij} = \exp(\ln(\delta_{ij} + h_{ij})). \)
Problem 7.20. Use eq. [7.3.91] and the definition of the Christoffel symbol to show that

\[ \partial_i \ln \sqrt{|g|} = \frac{1}{2} g^{ab} \partial_i g_{ab} = \Gamma^s_{is}. \]  

(7.3.93)

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.

Problem 7.21. Divergence of tensors. Verify the following formulas for the divergence of a vector \( V^i \), a fully antisymmetric rank-\((N \leq D)\) tensor \( F^{i_1i_2...i_N} \) and a symmetric tensor \( S^{ij} = S^{ji} \),

\[ \nabla_i V^i = \frac{\partial_i (\sqrt{|g|} V^i)}{\sqrt{|g|}}, \]  

(7.3.94)

\[ \nabla_j F^{i_1i_2...i_N} = \frac{\partial_j (\sqrt{|g|} F^{i_1i_2...i_N})}{\sqrt{|g|}}, \]  

(7.3.95)

\[ \nabla_i S^{ij} = \frac{\partial_i (\sqrt{|g|} S^{ij})}{\sqrt{|g|}} + \Gamma^j_{ab} S^{ab}. \]  

(7.3.96)

Note that, fully antisymmetric means, swapping any pair of indices costs a minus sign,

\[ F^{i_1i_2...i_N} = -F^{i_1...i_{a-1}i_a i_{a+1}...i_N}. \]  

(7.3.97)

Comment on how these expressions, equations (7.3.94)- (7.3.96), transform under a coordinate transformation, i.e., \( \vec{x} \rightarrow \vec{x}(\vec{\xi}) \).

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.

\[ \nabla^i \varphi = g^{ij} \nabla_j \varphi = g^{ij} \partial_j \varphi. \]  

(7.3.98)

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 \((r, \theta, \phi)\), where

\[ g^{ij} = \text{diag}(1, r^{-2}, r^{-2}(\sin \theta)^{-2}); \]  

(7.3.99)

the gradient of a scalar is

\[ \nabla^i \varphi = (\partial_r \varphi, r^{-2} \partial_\theta \varphi, r^{-2}(\sin \theta)^{-2} \partial_\phi \varphi). \]  

(7.3.100)

while the same object with lower indices is simply

\[ \nabla_i \varphi = (\partial_r \varphi, \partial_\theta \varphi, \partial_\phi \varphi). \]  

(7.3.101)

Laplacian of a scalar. The Laplacian of a scalar \( \psi \) 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 \( \Box \) or \( \nabla_i \nabla^i \).

\[ \Box \psi \equiv \vec{\nabla}^2 \psi = \nabla_i \nabla^i \psi = g^{ij} \nabla_i \nabla_j \psi. \]  

(7.3.102)
Consider transforming equalities of eq. (7.3.106), let us specialize to flat space Cartesian coordinates: \( \text{d} \ell (7.3.106) \). Hint: Refer to equations (7.3.11) and (7.3.93).

Explain why the final equality is, within this context, equivalent to the third equality of eq. Have the relation but if we now compute the scalar Laplacian acting on it with respect to \( \vec{y} \) instead, then we must have the relation

\[
\nabla^2 A_i = \frac{1}{\sqrt{|g'(\vec{y})|}} \frac{\partial}{\partial x^a} \left( \sqrt{|g'(\vec{y})|} g^{ab}(\vec{x}) \partial_{x^b} A_i(\vec{x}) \right) = \frac{1}{\sqrt{|g'(\vec{y})|}} \frac{\partial}{\partial x^m} \left( \sqrt{|g'(\vec{y})|} g^{mn}(\vec{y}) \partial_{y^n} A_i(\vec{y}) \right);
\]

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. That the 2nd equality of eq. (7.3.106) has to follow from its 1st, is because we are now effectively treating \( A_i \) as a scalar under coordinate transformations.

**Problem 7.23.** To further understand the transformation from \( \vec{x} \to \vec{y} \) in the second and third equalities of eq. (7.3.106), let us specialize to flat space Cartesian coordinates: \( \text{d} \ell^2 = \delta_{ij} \text{d}x^i \text{d}x^j \). Consider transforming \( \vec{x} \) to some other coordinate system \( \vec{x} = \vec{x}(\vec{y}) \). Calculus tells us,

\[
\nabla^2 A_i = \delta^{ab} \partial_{x^a} \partial_{x^b} A_i(\vec{x}) = \delta^{ab} \partial_{x^a} \left( \frac{\partial y^n}{\partial x^a} \partial_{y^n} A_i \right) = \frac{\partial y^m}{\partial x^a} \frac{\partial y^n}{\partial x^b} \partial_{y^m} \partial_{y^n} A_i(\vec{x}(\vec{y})) + \frac{\partial y^n}{\partial x^a} \frac{\partial y^n}{\partial x^b} \partial_{y^m} A_i(\vec{x}(\vec{y})).
\]

Explain why the final equality is, within this context, equivalent to the third equality of eq. (7.3.106). Hint: Refer to equations (7.3.11) and (7.3.93).
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

\[ \epsilon_{i_1i_2...i_{D-1}i_D}. \]  

(7.3.108)

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

\[ \epsilon_{i_1i_2...i_{a-1}i_ai_{a+1}...i_{D-1}i_D} = -\epsilon_{i_1i_2...i_{a-1}i_{b}i_{a+1}...i_{D-1}i_D}. \]  

(7.3.109)

- For a given ordering of the $D$ distinct coordinates \( \{ x^i | i = 1, 2, 3, ..., D \} \), \( \epsilon_{123...D} \equiv 1 \).

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, \( \epsilon_{i_1i_2...i_{D-1}i_D} \) is +1 if it takes even number of swaps to go from \( \{ 1, ..., D \} \) to \( \{ i_1, ..., i_D \} \); and is a $-1$ if it takes an odd number of swaps to do the same.

For example, in the (perhaps familiar) 3 dimensional case, in Cartesian coordinates \( (x^1, x^2, x^3) \),

\[ 1 = \epsilon_{123} = -\epsilon_{213} = -\epsilon_{321} = -\epsilon_{132} = \epsilon_{231} = \epsilon_{312}. \]  

(7.3.110)

The Levi-Civita tensor \( \tilde{\epsilon}_{i_1...i_D} \) is defined as

\[ \tilde{\epsilon}_{i_{i_1}...i_{i_D}} \equiv \sqrt{|g|} \epsilon_{i_{i_1}...i_{i_D}}. \]  

(7.3.111)

Let us understand why it is a (pseudo-)tensor. Because the Levi-Civita symbol is just a multi-index array of $\pm 1$ and 0, it does not change under coordinate transformations. Equation (7.2.27) then implies

\[ \sqrt{|g(\xi)|} \epsilon_{a_1a_2...a_D} = \sqrt{|g(\bar{x}(\xi))|} \left| \det \frac{\partial x^i(\xi)}{\partial \xi^j} \right| \epsilon_{a_1a_2...a_D}. \]  

(7.3.112)

On the right hand side, \( |g\left( \bar{x}(\xi) \right)| \) is the absolute value of the determinant of $g_{ij}$ written in the coordinates $\bar{x}$ but with $\bar{x}$ replaced with $\bar{x}(\xi)$.

If $\tilde{\epsilon}_{i_{i_1}...i_{i_D}}$ were a tensor, on the other hand, it must obey eq. (7.2.16),

\[ \sqrt{|g(\xi)|} \epsilon_{a_1a_2...a_D} \equiv \sqrt{|g(\bar{x}(\xi))|} \epsilon_{i_{i_1}...i_{i_D}} \frac{\partial x^{i_1}}{\partial \xi^{i_1}} \cdots \frac{\partial x^{i_D}}{\partial \xi^{i_D}}, \]

\[ = \sqrt{|g\left( \bar{x}(\xi) \right)|} \left( \det \frac{\partial x^i}{\partial \xi^j} \right) \epsilon_{a_1...a_D}, \]  

(7.3.113)
where in the second line we have recalled the co-factor expansion determinant of any matrix \( M \),
\[
\epsilon_{a_1 \ldots a_D} \det M = \epsilon_{i_1 \ldots i_D} M^{i_1}_{a_1} \ldots M^{i_D}_{a_D}.
\] (7.3.114)
Comparing equations [7.3.112] and [7.3.113] tells us the Levi-Civita \( \tilde{\epsilon}_{a_1 \ldots a_D} \) transforms as a tensor for orientation-preserving coordinate transformations, namely for all coordinate transformations obeying
\[
\det \frac{\partial x^i}{\partial \tilde{x}^j} = \epsilon_{i_1 i_2 \ldots i_D} \frac{\partial x^{i_1}}{\partial \tilde{x}^1} \frac{\partial x^{i_2}}{\partial \tilde{x}^2} \ldots \frac{\partial x^{i_D}}{\partial \tilde{x}^D} > 0.
\] (7.3.115)

**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,
\[
\sqrt{g(x)} \epsilon_{i_1 \ldots i_D} = \sqrt{g(\tilde{x})} \left| \text{det diag}[1, \ldots, 1, -1, 1, \ldots, 1] \right| \epsilon_{i_1 \ldots i_D} = \sqrt{g(\tilde{x})} \epsilon_{i_1 \ldots i_D};
\] (7.3.116)
whereas, under the usual rules of coordinate transformations (eq. [7.3.110]) we would have expected a ‘true’ tensor \( T_{i_1 \ldots i_D} \) to behave, for instance, as
\[
T_{(1)(2)\ldots(i-1)(i+1)\ldots(D)}(x) \frac{\partial x^i}{\partial \xi^j} = -T_{(1)(2)\ldots(i-1)(i+1)\ldots(D)}(\tilde{x}).
\] (7.3.117)

**Orientation of coordinate system**  What is orientation? It is the choice of how one orders the coordinates in use, say \((x^1, x^2, \ldots, x^D)\), together with the convention that \( \epsilon_{12 \ldots D} = 1 \).

In 2D flat spacetime, for example, we may choose the ‘right-handed’ \((x^1, x^2)\) as Cartesian coordinates, \( \epsilon_{12} = 1 \), and obtain the infinitesimal volume \( d^2x = dx^1 dx^2 \). We can switch to cylindrical coordinates
\[
\tilde{x}(\tilde{\xi}) = r(\cos \phi, \sin \phi).
\] (7.3.118)
so that
\[
\frac{\partial x^i}{\partial r} = (\cos \phi, \sin \phi), \quad \frac{\partial x^i}{\partial \phi} = (\cos \phi, \sin \phi), \quad r \geq 0, \ \phi \in [0, 2\pi).
\] (7.3.119)
If we ordered \((\xi^1, \xi^2) = (r, \phi)\), we would have
\[
\epsilon_{i_1 i_2} \frac{\partial x^{i_1}}{\partial r} \frac{\partial x^{i_2}}{\partial \phi} = \text{det} \begin{bmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{bmatrix} = r(\cos \phi)^2 + r(\sin \phi)^2 = r.
\] (7.3.120)
If we instead ordered \((\xi^1, \xi^2) = (\phi, r)\), we would have
\[
\epsilon_{i_1 i_2} \frac{\partial x^{i_1}}{\partial \phi} \frac{\partial x^{i_2}}{\partial r} = \text{det} \begin{bmatrix} -r \sin \phi & \cos \phi \\ r \cos \phi & \sin \phi \end{bmatrix} = -r(\sin \phi)^2 - r(\cos \phi)^2 = -r.
\] (7.3.121)
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 \( \epsilon_{r\phi} = 1 \).\(^{70}\)

\(^70\)We have gone from a ‘right-handed’ coordinate system \((x^1, x^2)\) to a ‘right-handed’ \((r, \phi)\); we could also have gone from a ‘left-handed’ one \((x^1, x^2)\) to a ‘left-handed’ \((\phi, r)\) and this would still be orientation-preserving.
By going from Cartesian coordinates \((x^1, x^2, x^3)\) to spherical ones,
\[
\vec{x}(\vec{\xi}) = r(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta),
\]
(7.3.122)
determine what is the orientation preserving ordering of the coordinates of \(\vec{\xi}\), and is \(\epsilon_{r\theta\phi}\) equal +1 or −1? □

**Infinitesimal volume re-visited** The infinitesimal volume we encountered earlier can really be written as
\[
d(\text{vol.}) = d^D\vec{x}\sqrt{|g(\vec{x})|} \epsilon_{12\ldots D} = d^D\vec{x}\sqrt{|g(\vec{x})|},
\]
(7.3.123)
so that under a coordinate transformation \(\vec{x} \to \vec{\xi}(\vec{\xi})\), the necessarily positive infinitesimal volume written in \(\vec{x}\) transforms into another positive infinitesimal volume, but written in \(\vec{\xi}\):
\[
d^D\vec{x}\sqrt{|g(\vec{x})|} \epsilon_{12\ldots D} = d^D\vec{\xi}\sqrt{|g(\vec{\xi})|} \epsilon_{12\ldots D}.
\]
(7.3.124)
Below, we will see that \(d^D\vec{x}\sqrt{|g(\vec{x})|}\) in modern integration theory is viewed as a differential \(D\)–form.

**Problem 7.25.** We may consider the infinitesimal volume in 3D flat space in Cartesian coordinates
\[
d(\text{vol.}) = dx^1 dx^2 dx^3.
\]
(7.3.125)
Now, let us switch to spherical coordinates \(\vec{\xi}\), with the ordering in the previous problem. Show that it is given by
\[
dx^1 dx^2 dx^3 = d^3\vec{\xi}\sqrt{|g(\vec{\xi})|}, \quad \sqrt{|g(\vec{\xi})|} = \epsilon_{i_1i_2i_3} \frac{\partial x^{i_1}}{\partial \xi^1} \frac{\partial x^{i_2}}{\partial \xi^2} \frac{\partial x^{i_3}}{\partial \xi^3}.
\]
(7.3.126)
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 \(\partial_{\xi^3} x^i\)? □

**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
\[
(\vec{X} \times \vec{Y})^k = \epsilon^{ijk} X^i Y^j.
\]
(7.3.127)
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
\[
(\vec{X} \times \vec{Y})^k \to |\vec{X}| |\vec{Y}| \epsilon^{12k} = |\vec{X}| |\vec{Y}| \delta^k_3,
\]
(7.3.128)
\[\text{Because of the existence of locally flat coordinates \(\{y^i\}\), the interpretation of }\sqrt{|g(\vec{\xi})|} \text{ as the volume of parallelepiped formed by }\{\partial_{\xi^1} y^i, \ldots, \partial_{\xi^3} y^i\} \text{ actually holds very generally.}\]
i.e., it is parallel to the positive $x^3$ axis. (Remember $k$ cannot be either 1 or 2 because $\epsilon_{ijk}$ 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 7.26.** Show that the Levi-Civita tensor with all upper indices is given by

$$\tilde{\epsilon}^{i_1 i_2 \ldots i_D} = \frac{\text{sgn} \det(g_{ab})}{\sqrt{|g|}} \epsilon_{i_1 i_2 \ldots i_D}. \quad (7.3.129)$$

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.\footnote{See eq. (7.2.61) to understand why the sign of the determinant of the metric is always determined by the sign of the determinant of its flat counterpart.} Hint: Raise the indices by contracting with inverse metrics, then recall the cofactor expansion definition of the determinant.

**Problem 7.27.** Show that the covariant derivative of the Levi-Civita tensor is zero.

$$\nabla_j \tilde{\epsilon}^{i_1 i_2 \ldots i_D} = 0. \quad (7.3.130)$$

(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 some vector $V$,

$$\nabla_j (\tilde{\epsilon}^{i_1 i_2 \ldots i_D - 2j^k} V_k) = \tilde{\epsilon}^{i_1 i_2 \ldots i_D - 2j^k} \nabla_j V_k = \tilde{\epsilon}^{i_1 i_2 \ldots i_D - 2j^k} \partial_j V_k. \quad (7.3.131)$$

Why is $\nabla_i V_j - \nabla_j V_i = \partial_i V_j - \partial_j V_i$ for any $V_j$? 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 $T_{i_1 \ldots i_N}$. Denoting $\Pi(i_1 \ldots i_N)$ to be a permutation of the indices $\{i_1 \ldots i_N\}$, the antisymmetrization procedure is given by

$$T_{[i_1 \ldots i_N]} = \sum_{\text{permutations } \Pi \text{ of } \{i_1, i_2, \ldots, i_N\}} \sigma_\Pi \cdot T_{\Pi(i_1 \ldots i_N)} \quad (7.3.132)$$

$$= \sum_{\text{even permutations } \Pi \text{ of } \{i_1, i_2, \ldots, i_N\}} T_{\Pi(i_1 \ldots i_N)} - \sum_{\text{odd permutations } \Pi \text{ of } \{i_1, i_2, \ldots, i_N\}} T_{\Pi(i_1 \ldots i_N)}.$$

In words: for a rank $-N$ tensor, $T_{[i_1 \ldots i_N]}$ consists of a sum of $N!$ terms. The first is $T_{i_1 \ldots i_N}$. 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 $\sigma_\Pi$ is often called the sign of the permutation $\Pi$.) For example,

$$T_{[ij]} = T_{ij} - T_{ji}, \quad T_{[ijk]} = T_{ijk} - T_{ikj} - T_{jik} + T_{kij} + T_{kji} - T_{kji}. \quad (7.3.133)$$

Can you see why eq. (7.3.132) 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. (7.3.132) into pairs, say \( \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} - \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} \). That is, for a given term \( \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} \) there must be a counterpart with \( i_{a} \leftrightarrow i_{b} \) swapped, multiplied 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. (7.3.132),

\[
T_{[i_{1},\ldots,i_{a},\ldots,i_{b},\ldots,i_{N}]} = \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} - \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} + \cdots,
\]

(7.3.134)

\[
T_{[i_{1},\ldots,i_{a},\ldots,i_{N}]} = - (\prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} - \prod_{i} \prod_{J} T_{j_{1},\ldots,i_{a},\ldots,i_{b}} + \cdots).
\]

(7.3.135)

**Problem 7.28.** Given \( T_{i_{1}i_{2},\ldots,i_{N}} \), how do we construct a fully symmetric object from it, i.e., such that swapping any two indices returns the same object? □

**Problem 7.29.** 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 \( \delta \)s.

\[
\epsilon_{i_{1}i_{2}\ldots i_{D}} = \delta_{i_{1}}^{1} \delta_{i_{2}}^{2} \ldots \delta_{i_{D-1}}^{D-1} \delta_{i_{D}}^{D}
\]

(7.3.136)

Can you also explain why the following is true?

\[
\epsilon_{a_{1}a_{2}\ldots a_{D-1}a_{D}} \det A = \epsilon_{i_{1}i_{2}\ldots i_{D-1}i_{D}} A_{a_{1}i_{1}}^{i_{1}} A_{a_{2}i_{2}}^{i_{2}} \cdots A_{a_{D-1}i_{D-1}}^{i_{D-1}} A_{a_{D}i_{D}}^{i_{D}}
\]

(7.3.137)

**Problem 7.30.** Argue that

\[
T_{[i_{1}\ldots i_{N}]} = T_{[i_{1}\ldots i_{N-1}]} - T_{[i_{N}i_{1}\ldots i_{N-1}]} + T_{[i_{1}i_{2}\ldots i_{N-1}]} - T_{[i_{1}i_{2}\ldots i_{N}]} + \cdots - T_{[i_{1}\ldots i_{N-2}i_{N}]}.
\]

(7.3.138)

In words: to construct the fully anti-symmetric combination of \( N \) indices, anti-symmetrize the first \( N-1 \) indices. Then swap the first and \( N \)th index of the this first group; then swap the second and \( N \) index; etc. □

**Product of Levi-Civita tensors** The product of two Levi-Civita tensors will be important for the discussions to come. We have

\[
\varepsilon^{i_{1}\ldots i_{N}k_{1}\ldots k_{D-N}} \varepsilon_{j_{1}\ldots j_{N}k_{1}\ldots k_{D-N}} = \text{sgn det}(g_{ab}) \cdot A_{N} \delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{N}]}^{i_{N}}, \quad 1 \leq N \leq D,
\]

(7.3.139)

\[
\varepsilon^{k_{1}\ldots k_{D}} \varepsilon_{k_{1}\ldots k_{D}} = \text{sgn det}(g_{ab}) \cdot A_{0}, \quad A_{N\geq0} \equiv (D-N)!. \]

(7.3.140)

(Remember \( 0! = 1! = 1 \); also, \( \delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{N}]}^{i_{N}} = \delta_{j_{1}}^{[i_{1}} \ldots \delta_{j_{N}}^{i_{N}]} \) ) Let us first understand why there are a bunch of Kronecker deltas on the right hand side, starting from the \( N = D \) case — where no indices are contracted.

\[
\text{sgn det}(g_{ab}) \varepsilon^{i_{1}\ldots i_{D}} \varepsilon_{j_{1}\ldots j_{D}} = \epsilon_{i_{1}i_{D}} \epsilon_{j_{1}j_{D}} = \delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{D}]}^{i_{D}}
\]

(7.3.141)

(This means \( A_{D} = 1 \).) The first equality follows from eq. (7.3.129). The second may seem a bit surprising, because the indices \( \{i_{1},\ldots,i_{D}\} \) are attached to a completely different \( \varepsilon \) tensor from the \( \{j_{1},\ldots,j_{D}\} \). However, if we manipulate

\[
\delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{D}]}^{i_{D}} = \delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{D}]}^{i_{D}} \sigma_{j_{1}} \sigma_{j_{D}} = \delta_{[j_{1}}^{i_{1}} \ldots \delta_{j_{D}]}^{i_{D}} \sigma_{i} \sigma_{j} = \epsilon_{i_{1}i_{D}} \epsilon_{j_{1}j_{D}},
\]

(7.3.142)
where \( \sigma_i = 1 \) if it took even number of swaps to re-arrange \( \{i_1, \ldots, i_D\} \) to \( \{1, \ldots, D\} \) 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 \( \{j_1, \ldots, j_D\} \) to \( \{1, \ldots, D\} \) and \( \sigma_j = -1 \) if it took odd number of swaps. But \( \sigma_i \) is precisely the Levi-Civita symbol \( \epsilon_{i_1 \ldots i_D} \) and likewise \( \sigma_j = \epsilon_{j_1 \ldots j_D} \). The \((\geq 1)\)-contractions between the \( \tilde{\epsilon} \)s can, in principle, be obtained by contracting the right hand side of (7.3.141). 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. (7.3.139) 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 \( A_N \)s. Starting from eq. (7.3.139), and employing eq. (7.3.138),

\[
\tilde{\epsilon}^{i_1 \ldots i_{N-1} \sigma k_1 \ldots k_{D-N}}_{j_1 \ldots j_{N-1} \sigma k_1 \ldots k_{D-N}} = A_N \delta^{i_1}_{[j_1} \ldots \delta^{i_{N-1}}_{(j_N-1)} \delta^{\sigma}_{\sigma]} = A_N \left( \delta^{i_1}_{[j_1} \ldots \delta^{i_{N-1}}_{(j_N-1)} \delta^{\sigma}_{\sigma]} - \delta^{i_1}_{[j_1} \delta^{i_2}_{j_2} \ldots \delta^{i_{N-1}}_{(j_N-1)} \delta^{\sigma}_{\sigma]} \right) = A_N \cdot (D - (N - 1)) \delta^{i_1}_{[j_1} \ldots \delta^{i_{N-1}}_{(j_N-1)} \equiv A_{N-1} \delta^{i_1}_{[j_1} \ldots \delta^{i_{N-1}}_{(j_N-1)}.
\]

(The last equality is a definition, because \( A_{N-1} \) is the coefficient of \( \delta^{i_1}_{[j_1} \ldots \delta^{i_{N-1}}_{(j_N-1)} \)\) We have the relationship

\[
A_N = \frac{A_{N-1}}{D - (N - 1)}.
\]

If we contract every index, we have to sum over all the \( D! \) (non-zero components of the Levi-Civita symbol)\(^2\),

\[
\tilde{\epsilon}^{i_1 \ldots i_D} \tilde{\epsilon}_{i_1 \ldots i_D} = \text{sgn det}(g_{ab}) \cdot \sum_{i_1, \ldots, i_D} (\epsilon_{i_1 \ldots i_D})^2 = \text{sgn det}(g_{ab}) \cdot D!
\]

That means \( A_0 = D! \). If we contracted every index but one,

\[
\tilde{\epsilon}^{i_k \ldots i_D} \tilde{\epsilon}_{j_k \ldots i_D} = \text{sgn det}(g_{ab}) A_1 \delta^{j}_j.
\]

Contracting the \( i \) and \( j \) indices, and invoking eq. (7.3.145),

\[
\text{sgn det}(g_{ab}) \cdot D! = \text{sgn det}(g_{ab}) A_1 \cdot D \quad \Rightarrow \quad A_1 = (D - 1)!.
\]

That means we may use \( A_1 \) (or, actually, \( A_0 \)) to generate all other \( A_{N\geq 0} \)s,

\[
A_N = \frac{A_{N-1}}{(D - (N - 1))} = \frac{1}{D - (N - 1)} A_N - 2 = \ldots = \frac{A_{1}}{(D - (N - 1))} = \frac{(D - 1)!}{(D - 1)(D - 2)(D - 3) \ldots (D - (N - 1))} = (D - 1)! = (D - N)!.
\]

Note that \( 0! = 1 \), so \( A_D = 1 \) as we have found earlier.
Problem 7.31. Matrix determinants revisited

Explain why the cofactor expansion definition of a square matrix in eq. (3.2.1) can also be expressed as

\[ \det A = \epsilon_{i_1 i_2 \ldots i_D} A_{i_1}^1 A_{i_2}^2 \ldots A_{i_{D-1}}^{D-1} A_{i_D}^D \] (7.3.149)

provided we define \( \epsilon_{i_1 i_2 \ldots i_D} \) in the same way we defined its lower index counterpart, including \( \epsilon_{123} \equiv 1 \). 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 \( \det A^T = \det A \)? Can you also prove, using our result for the product of two Levi-Civita symbols, that \( \det(A \cdot B) = (\det A)(\det B) \)?

Problem 7.32. In 3D vector calculus, the curl of a gradient of a scalar is zero – how would you express that using the \( \tilde{\epsilon} \) tensor? What about the statement that the divergence of a curl of a vector field is zero? Can you also derive, using the \( \tilde{\epsilon} \) tensor in Cartesian coordinates and eq. (7.3.139), the 3D vector cross product identity

\[ \vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C} \] (7.3.150)

Hodge dual

We are now ready to define the Hodge dual. Given a fully antisymmetric rank-\( N \) tensor \( T_{i_1 \ldots i_N} \), its Hodge dual – which I shall denote as \( \tilde{T}^{j_1 \ldots j_{D-N}} \) – is a fully antisymmetric rank-(\( D-N \)) tensor whose components are

\[ \tilde{T}^{j_1 \ldots j_{D-N}} \equiv \frac{1}{N!} \epsilon^{j_1 \ldots j_{D-N} i_1 \ldots i_N} T_{i_1 \ldots i_N}. \] (7.3.151)

Invertible

Note that the Hodge dual is an invertible operation, as long as we are dealing with fully antisymmetric tensors, in that given \( \tilde{T}^{j_1 \ldots j_{D-N}} \) we can recover \( T_{i_1 \ldots i_N} \) and vice versa. All you have to do is contract both sides with the Levi-Civita tensor, namely

\[ T_{i_1 \ldots i_N} = \text{sgn}(\det g_{ab}) \frac{(-)^{N(D-N)}}{(D-N)!} \epsilon^{i_1 \ldots i_N j_1 \ldots j_{D-N}} \tilde{T}^{j_1 \ldots j_{D-N}}. \] (7.3.152)

In other words \( \tilde{T}^{j_1 \ldots j_{D-N}} \) and \( T_{i_1 \ldots i_N} \) contain the same amount of information.

Problem 7.33. Using eq. (7.3.139), verify the proportionality constant \( (-)^{N(D-N)} \text{sgn} g \) in the inverse Hodge dual of eq. (7.3.152), and thereby prove that the Hodge dual is indeed invertible for fully antisymmetric tensors.

Curl

The curl of a vector field \( A_i \) can now either be defined as the antisymmetric rank-2 tensor

\[ F_{ij} \equiv \partial_i A_j \] (7.3.153)
or its rank-$(D - 2)$ Hodge dual

$$
\tilde{F}^{i_1 i_2 \ldots i_{D-2}} \equiv \frac{1}{2} \varepsilon^{i_1 i_2 \ldots i_{D-2} j k} \partial_j A_k.
$$

$(D = 3)$-dimensional space is a special case where both the original vector field $A^i$ and the Hodge dual $\tilde{F}^i$ 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 $A_i$:

$$
B^i = \frac{1}{2} \varepsilon^{ijk} \partial_j A_k = \varepsilon^{ijk} \partial_j A_k.
$$

In particular, when we specialize to 3D flat space with Cartesian coordinates:

$$
(\nabla \times \vec{A})^i = \varepsilon^{ijk} \partial_j A_k, \quad \text{(Flat 3D Cartesian).}
$$

$$
\left(\nabla \times \vec{A} \right)^1 = \varepsilon^{123} \partial_2 A_3 + \varepsilon^{132} \partial_3 A_2 = \partial_2 A_3 - \partial_3 A_2, \quad \text{etc.}
$$

By setting $i = 1, 2, 3$ we can recover the usual definition of the curl in 3D vector calculus. But you may have noticed from equations (7.3.153) and (7.3.154), 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.

**Divergence versus Curl**

We can extend the definition of a curl of a vector field to that of a rank$-N \leq D - 1$ fully antisymmetric $B_{i_1 \ldots i_N}$ as

$$
\nabla_{[\sigma} B_{i_1 \ldots i_N]} = \partial_{[\sigma} B_{i_1 \ldots i_N]}.
$$

(Can you explain why the $\nabla$ can be replaced with $\partial$? Notice, this definition of the curl does not involve the metric.) With the Levi-Civita tensor, we can convert the curl of an antisymmetric tensor into the divergence of its dual,

$$
\nabla_{\ell} \tilde{B}^{\ell_{j_1 \ldots j_{D-N-1}} \cdot \ell} = \frac{1}{N!} \tilde{\varepsilon}^{\ell_{j_1 \ldots j_{D-N-1}} \cdot \ell j_1 \ldots i_N} \nabla_{\ell} B_{i_1 \ldots i_N}
$$

$$
= (N + 1) \cdot \tilde{\varepsilon}^{\ell_{j_1 \ldots j_{D-N-1}} \cdot \ell j_1 \ldots i_N} \partial_{[\ell} B_{i_1 \ldots i_N]}.
$$

In the first equality, we have used the fact that the Levi-Civita tensor is covariantly constant (cf. eq. (7.3.130)). Since $\partial_{[\sigma} B_{i_1 \ldots i_N]}$ and its Hodge dual contains the same information, we may proceed to identify the two objects,

$$
\nabla_{\ell} \tilde{B}^{\ell_{j_1 \ldots j_{D-N-1}} \cdot \ell} \leftrightarrow \partial_{[\ell} B_{i_1 \ldots i_N]}.
$$

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:

$$
\nabla_j \tilde{A}^{ij} = \varepsilon^{ijk} \nabla_j A_k = B^i.
$$

The divergence of the dual of $A_i$ is the (negative) curl of $A_i$.

Let us take the anti-symmetric derivative of $F_{ij} \equiv \partial_{[i} A_{j]}$.

$$
\partial_{[i} F_{jk]} = \partial_{[i} \partial_{j]} A_{k]} = 2 \partial_{[i} \partial_{j]} A_{k]}.
$$
\[ = \partial_i \partial_j A_k - \partial_j \partial_i A_k = 0. \quad (7.3.163) \]

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 \( v_i \equiv \partial_i \varphi \),

\[ \partial_i v_j = \partial_i \partial_j \varphi \]
\[ = (\partial_i \partial_j - \partial_j \partial_i) \varphi = 0. \quad (7.3.164) \]

**Problem 7.34.** In 3D vector calculus, we learn that the divergence of a curl is zero

\[ \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0; \quad (7.3.165) \]

and the curl of a gradient is zero

\[ \vec{\nabla} \times \vec{\nabla} \varphi = 0. \quad (7.3.166) \]

In 3D curved space, verify that equations (7.3.163) and (7.3.164) are simply the Hodge dual versions of equations (7.3.165) and (7.3.166).

**Problem 7.35.** Prove the following \( D = 3 \) identity,

\[ \left( \vec{\nabla} \times \left( \vec{\nabla} \times \vec{A} \right) \right)^i = \nabla^i (\nabla_j A^j) - \nabla_j \nabla^j A^i; \quad (7.3.167) \]

which holds in arbitrary curved spaces.

**Problem 7.36.** Show, by contracting both sides of eq. (7.3.155) with an appropriate \( \tilde{\epsilon} \)-tensor, that

\[ \tilde{\epsilon}_{ijk} B^k = \frac{1}{2} \partial_i A_j. \quad (7.3.168) \]

Assume \( \text{sgn } \det(g_{ab}) = 1. \)

**Problem 7.37.** In \( D \)-dimensional space, is the Hodge dual of a rank-\( D \) fully antisymmetric tensor \( F_{i_1 \ldots i_D} \) invertible? Hint: If \( F_{i_1 \ldots i_D} \) is fully antisymmetric, how many independent components does it have? Can you use that observation to relate \( \tilde{F} \) and \( F_{i_1 \ldots i_D} \) in

\[ \tilde{F} \equiv \frac{1}{D!} \tilde{\epsilon}^{i_1 \ldots i_D} F_{i_1 \ldots i_D}? \quad (7.3.169) \]

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?

**Problem 7.38. All 2D Metrics Are Conformally Flat**

A metric \( g_{ij} \) is said to be conformally flat if it is equal to the flat metric multiplied by a scalar function (which we shall denote as \( \Omega^2 \) – not to be confused with the solid angle):

\[ g_{ij} = \Omega^2 \bar{g}_{ij}. \quad (7.3.170) \]

Here, \( \bar{g}_{ij} = \text{diag}[1,1] \) if we are working with a curved space; whereas (in the following Chapter) \( \bar{g}_{ij} = \text{diag}[1,-1] \) if we are dealing with a curved spacetime instead.

In this problem, we will prove that:

---

\(^{74}\)This problem is based on appendix 11C of [22].
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. (7.3.170).

Suppose we begin with the metric \( g_{ij'}(\vec{x'})dx^i'dx^j' \). To show that we can find a coordinate transformation \( \vec{x}'(\vec{x}) \) such that eq. (7.3.170) is achieved, explain why

\[
\frac{\partial x^1}{\partial x'^m} \frac{\partial x^2}{\partial x'^n} g^{mn'}(\vec{x}') = 0. \tag{7.3.171}
\]

If we view \( x^1 \) and \( x^2 \) as scalar fields in the curved space(time), then \( \partial x^1/\partial x^m \) are 1-forms (for e.g., \( dx^1 = (\partial x^1/\partial x^m)dx^m \)), and eq. (7.3.171) tells us they are orthogonal. Show that eq. (7.3.171) may be solved by demanding one is the Hodge dual of the other – namely,

\[
\frac{\partial x^1}{\partial x'^m} = \tilde{\epsilon}^{m'n'} \frac{\partial x^2}{\partial x'^n}. \tag{7.3.172}
\]

Provided eq. (7.3.172) holds, next show that

\[
\frac{\partial x^1}{\partial x'^m} \frac{\partial x^2}{\partial x'^n} g^{mn'} = (\text{sgn det } g) \frac{\partial x^2}{\partial x'^m} \frac{\partial x^1}{\partial x'^n} g^{mn'}. \tag{7.3.173}
\]

Now explain why eq. (7.3.171) implies eq. (7.3.170) for both curved space and spacetime.

**Problem 7.39. 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

\[
d^2 = \delta_{ij}dx^idx^j. \tag{7.3.174}
\]

Show that the gradient of a scalar \( \psi \) is

\[
\vec{\nabla}\psi = (\partial_1\psi, \partial_2\psi, \partial_3\psi) = (\partial^1\psi, \partial^2\psi, \partial^3\psi); \tag{7.3.175}
\]

the Laplacian of a scalar \( \psi \) is

\[
\nabla_i \nabla^i \psi = \delta^{ij}\partial_i\partial_j\psi = \left(\partial^1_2 + \partial^2_1 + \partial^3_3\right)\psi; \tag{7.3.176}
\]

the divergence of a vector \( A \) is

\[
\nabla_i A^i = \partial_i A^i; \tag{7.3.177}
\]

and the curl of a vector \( A \) is

\[
(\vec{\nabla} \times \vec{A})^i = \epsilon^{ijk}\partial_j A_k. \tag{7.3.178}
\]

**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

\[
\vec{x} = (\rho \cos \phi, \rho \sin \phi, z) \tag{7.3.179}
\]
to argue that the flat metric is translated from \( g_{ij} = \delta_{ij} \) to
\[
d\ell^2 = d\rho^2 + \rho^2 d\phi^2 + dz^2. \tag{7.3.180}
\]

Show that the gradient of a scalar \( \psi \) is
\[
\nabla \hat{\rho} \psi = \partial_\rho \psi, \quad \nabla \hat{\phi} \psi = \frac{1}{\rho} \partial_\phi \psi, \quad \nabla \hat{z} \psi = \partial_z \psi; \tag{7.3.181}
\]

the Laplacian of a scalar \( \psi \) is
\[
\nabla_i \nabla^i \psi = \frac{1}{\rho} \partial_\rho (\rho \partial_\rho \psi) + \frac{1}{\rho^2} \partial^2_\phi \psi + \partial^2_z \psi;
\tag{7.3.182}
\]

the divergence of a vector \( A \) is
\[
\nabla_i A^i = \frac{1}{\rho} \left( \partial_\rho (\rho A^\rho) + \partial_\phi A^\phi \right) + \partial_z A^z; \tag{7.3.183}
\]

and the curl of a vector \( A \) is given by
\[
\tilde{\epsilon}^{ijk} \partial_j A_k = \frac{1}{\rho \sin \theta} \left( \partial_\theta (\sin \theta \cdot A^\phi) - \partial_\phi A^\theta \right), \quad \tilde{\epsilon}^{\phi jk} \partial_j A_k = \frac{1}{r \sin \theta} \left( \partial_\phi (\sin \theta \cdot A^\rho) - \partial_\rho A^\phi \right), \quad \tilde{\epsilon}^{\phi jk} \partial_j A_k = \frac{1}{r} \left( \partial_r (r A^\phi) - \partial_\theta A^\rho \right). \tag{7.3.184}
\]

**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
\[
\vec{x} = (r \sin(\theta) \cos(\phi), r \sin(\theta) \sin(\phi), r \cos(\theta)). \tag{7.3.185}
\]

Show that the flat metric is now
\[
d\ell^2 = dr^2 + r^2 (d\theta^2 + (\sin \theta)^2 d\phi^2); \tag{7.3.186}
\]

the gradient of a scalar \( \psi \) is
\[
\nabla \hat{\rho} \psi = \partial_\rho \psi, \quad \nabla \hat{\theta} \psi = \frac{1}{r} \partial_\theta \psi, \quad \nabla \hat{\phi} \psi = \frac{1}{r \sin \theta} \partial_\phi \psi; \tag{7.3.187}
\]

the Laplacian of a scalar \( \psi \) is
\[
\nabla_i \nabla^i \psi = \frac{1}{r^2} \partial_r (r^2 \partial_\rho \psi) + \frac{1}{r^2 \sin \theta} \partial_\theta (\sin \theta \cdot \partial_\theta \psi) + \frac{1}{r^2 (\sin \theta)^2} \partial^2_\phi \psi; \tag{7.3.188}
\]

the divergence of a vector \( A \) reads
\[
\nabla_i A^i = \frac{1}{r} \partial_r (r^2 A^\rho) + \frac{1}{r \sin \theta} \partial_\theta (\sin \theta \cdot A^\phi) + \frac{1}{r \sin \theta} \partial_\phi A^\phi; \tag{7.3.189}
\]

and the curl of a vector \( A \) is given by
\[
\tilde{\epsilon}^{\rho jk} \partial_j A_k = \frac{1}{r \sin \theta} \left( \partial_\theta (\sin \theta \cdot A^\rho) - \partial_\rho A^\theta \right), \quad \tilde{\epsilon}^{\phi jk} \partial_j A_k = \frac{1}{r} \left( \partial_r (r A^\phi) - \partial_\theta A^\rho \right). \tag{7.3.190}
\]

\[\square\]
Problem 7.40. Translation operator in infinite curved space. When discussing the translation operator in, say eq. (4.5.72), 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 \( \langle \vec{x} | \vec{x} \prime \rangle \); for it to be generally covariant we require that is a coordinate scalar,

\[
\langle \vec{x} | \vec{x} \prime \rangle = \delta^{(D)}(\vec{x} - \vec{x} \prime) \frac{1}{\sqrt{|g(\vec{x})g(\vec{x} \prime)|}}.
\] (7.3.191)

Argue that any state \( | f \rangle \) can now be expressed through the superposition

\[
| f \rangle = \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(\vec{x} \prime)|} | \vec{x} \prime \rangle \langle \vec{x} \prime | f \rangle ;
\] (7.3.192)

and the completeness relation is therefore

\[
\mathbb{I} = \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(\vec{x} \prime)|} | \vec{x} \prime \rangle \langle \vec{x} \prime | .
\] (7.3.193)

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

\[
\mathcal{T}(\vec{\xi}) = \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(\vec{x} \prime)|} | \vec{x} \prime \rangle \langle \vec{x} + \vec{\xi} | \vec{x} \prime | .
\] (7.3.194)

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. (7.3.194)) for the inverse of \( \mathcal{T}(\vec{d}) \)? What happens when \( \vec{\xi} \) in eq. (7.3.194) is infinitesimal and satisfies Killing’s equation (cf. eq. (8.5.46))? Specifically, show that

\[
\mathcal{T}(\vec{\xi})^\dagger \mathcal{T}(\vec{\xi}) = \mathbb{I} + O \left( \vec{\xi}^2 \right) ;
\] (7.3.195)

namely, the translation operator acting along \( \vec{\xi} \) is unitary up to first order in the infinitesimal displacement. \( \square \)

7.4 Hypersurfaces

7.4.1 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 \( \{ \xi^I | I = 1, 2, \ldots, N \} \), 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.,

\[
g_{ij} dx^i dx^j \rightarrow g_{ij} \left( \vec{x}(\vec{\xi}) \right) \frac{\partial x^i(\vec{\xi})}{\partial \xi^I} \frac{\partial x^j(\vec{\xi})}{\partial \xi^J} d\xi^I d\xi^J \equiv H_{ij}(\vec{\xi}) d\xi^I d\xi^J .
\] (7.4.1)
The $H_{IJ}$ is the (induced) metric on the hypersurface.\(^{75}\)

Observe that the $N$ vectors

$$\left\{ \frac{\partial x^i}{\partial \xi^I} \bigg| I = 1, 2, \ldots, N \right\},$$

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 $\partial/\partial \xi^I$ themselves form a basis set of tangent vectors, from the perspective of an observer confined to live on this hypersurface.

**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 $\xi^I \equiv (0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$, such that from the ambient perspective, the sphere is described by

$$x^i(\vec{\xi}) = R(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \quad \text{(Cartesian components).} \quad (7.4.3)$$

(Remember $R$ is a fixed quantity here.) The induced metric on the sphere itself, according to eq. (7.4.1), will lead us to the expected result

$$H_{IJ}(\vec{\xi})d\xi^I d\xi^J = R^2 \left( d\theta^2 + (\sin \theta)^2 d\phi^2 \right). \quad (7.4.4)$$

**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

$$X^i \equiv (x, y, z) = (x, y, f(x, y)). \quad (7.4.5)$$

The tangent vectors $(\partial X^i/\partial \xi^I)$ are

$$\frac{\partial X^i}{\partial x} = (1, 0, \partial_x f), \quad \frac{\partial X^i}{\partial y} = (0, 1, \partial_y f). \quad (7.4.6)$$

The induced metric, according to eq. (7.4.1), is given by

$$H_{IJ}(\vec{\xi})d\xi^I d\xi^J = \delta_{ij} \left( \frac{\partial X^i}{\partial x} \frac{\partial X^j}{\partial x} (dx)^2 + \frac{\partial X^i}{\partial y} \frac{\partial X^j}{\partial y} (dy)^2 + 2 \frac{\partial X^i}{\partial x} \frac{\partial X^j}{\partial y} dx dy \right),$$

$$H_{IJ}(\vec{\xi}) = \begin{bmatrix} 1 + (\partial_x f)^2 & \partial_x f \partial_y f \\ 1 + (\partial_y f)^2 & \partial_y f \partial_y f \end{bmatrix}, \quad \xi^I \equiv (x, y), \quad (7.4.7)$$

where on the second line the “$\equiv$” 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.

---

\(^{75}\)The Lorentzian signature of curved spacetimes, 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.
components. Recall that the infinitesimal volume (= 2D area) is given in any coordinate system \( \xi \) by \( d^2 \xi \sqrt{\det H_{ij}(\xi)} \). That means from taking the det of eq. (7.4.7), if the domain on \((x, y)\) is denoted as \( \mathcal{D} \), the corresponding area swept out by \( f \) is given by the 2D integral

\[
\int_{\mathcal{D}} dx dy \sqrt{\det H_{ij}(x, y)} = \int_{\mathcal{D}} dx dy \sqrt{(1 + (\partial_x f)^2)(1 + (\partial_y f)^2) - (\partial_x f \partial_y f)^2} = \int_{\mathcal{D}} dx dy \sqrt{1 + (\partial_x f(x, y))^2 + (\partial_y f(x, y))^2}.
\]

### Differential Forms and Volume

Although we have not (and shall not) employ differential forms very much, it is very much part of modern integration theory. One no longer writes \( \int d^3 \vec{x} f(\vec{x}) \) in flat space, for instance, but rather

\[
\int f(\vec{x}) dx^1 \wedge dx^2 \wedge dx^3.
\]

More generally, whenever the following \( N \)-form occur under an integral sign, we have the definition

\[
\frac{dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{N-1} \wedge dx^N}{(\text{Differential form notation})} = \frac{d^N x}{\text{Physicists' colloquial math-speak}}.
\]

(Here \( N \leq D \), where \( D \) is the dimension of space.) This needs to be supplemented with the constraint that it is a fully antisymmetric object:

\[
dx[i_1] \wedge dx[i_2] \wedge \cdots \wedge dx[i_{N-1}] \wedge dx[i_N] = \epsilon_{i_1 \cdots i_N} dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{N-1} \wedge dx^N.
\]

The superposition of rank-(\( N \leq D \)) differential forms spanned by \( \{(1/N!)F_{i_1 \cdots i_N} dx[i_1] \wedge \cdots dx[i_N]\} \), for arbitrary but fully antisymmetric \( \{F_{i_1 \cdots i_N}\} \), forms a vector space.

Why differential forms are fundamental to integration theory is because, it is this antisymmetry that allows its proper definition as the volume spanned by an \( N \)-parallelpiped. For one, the antisymmetric nature of forms is responsible for the Jacobian upon a change-of-variables \( \vec{x}(\vec{y}) \) familiar from multi-variable calculus – using eq. (7.4.11):

\[
dx^1 \wedge dx^2 \wedge \cdots \wedge dx^{N-1} \wedge dx^N = \frac{\partial x^1}{\partial y^{i_1}} \frac{\partial x^2}{\partial y^{i_2}} \cdots \frac{\partial x^N}{\partial y^{i_N}} \, dy^{i_1} \wedge dy^{i_2} \wedge \cdots \wedge dy^{i_{N-1}} \wedge dy^{i_N}
\]

\[
= \frac{\partial x^1}{\partial y^{i_1}} \frac{\partial x^2}{\partial y^{i_2}} \cdots \frac{\partial x^N}{\partial y^{i_N}} \epsilon^{i_1 \cdots i_N} \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^{N-1} \wedge dy^N
\]

\[
= \left( \det \frac{\partial x^i}{\partial y^j} \right) \, dy^1 \wedge dy^2 \wedge \cdots \wedge dy^{N-1} \wedge dy^N.
\]

In a \((D \geq 2)\)-dimensional flat space, you might be familiar with the statement that \( D \) linearly independent vectors define a \( D \)-parallelpiped. 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 \( \{\Phi^l\} | l = 1, 2, \ldots, N \) , i.e.,

\[
d\Phi^1 \wedge \cdots \wedge d\Phi^N.
\]

255
let us see how it defines an infinitesimal $N$–volume by generalizing the notion of volume-as-
determinants.\footnote{These scalar fields \{\Phi^I\} can also be thought of as coordinates parametrizing some $N$–dimensional sub-space of the ambient $D$–dimensional space.} Focusing on the $N = 2$ case, if \( \vec{v} \equiv (p_1 dx^1, \ldots, p_D dx^D) \) and \( \vec{w} \equiv (q_1 dx^1, \ldots, q_D dx^D) \) are two linearly independent vectors formed from \( p_i = \partial_i \Phi^1 \) and \( q_i = \partial_i \Phi^2 \), then

\[
d\Phi^1 \wedge d\Phi^2 = (p_i dx^i) \wedge (q_j dx^j) = p_i q_j dx^i \wedge dx^j \tag{7.4.14}
\]
is in fact the 2D area spanned by the parallelepiped defined by \( \vec{v} \) and \( \vec{w} \). For, since \( d\Phi^1 \wedge d\Phi^2 \) is a coordinate scalar, we may choose a locally flat coordinate system \( \{y^i\} \) such that \( p_i \) and \( q_i \) lie on the \((1, 2)\)–plane; i.e., \( p_{i>2} = q_{i>2} = 0 \) and

\[
d\Phi^1 \wedge d\Phi^2 = (p_i dy^i) \wedge (q_j dy^j) = p_1 q_2 dy^1 \wedge dy^2 + p_2 q_1 dy^2 \wedge dy^1 \\
= (p_1 q_2 - p_2 q_1) dy^1 dy^2 = \det \begin{bmatrix} \vec{v} & \vec{w} \end{bmatrix} \tag{7.4.15}
\]

where now

\[
\vec{v} = \left( \partial_1 \Phi^1 dy^1, \partial_2 \Phi^1 dy^2, 0 \right)^T, \tag{7.4.16}
\]

\[
\vec{w} = \left( \partial_1 \Phi^2 dy^1, \partial_2 \Phi^2 dy^2, 0 \right)^T. \tag{7.4.17}
\]

This argument can be readily extended to higher \( 2 < N \leq D \).

### 7.4.2 Fluxes, Gauss-Stokes’ theorems, Poincaré lemma

**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-\( f \) surface coincides with the hypersurface,

\[
f(\vec{x}) = C \equiv \text{constant}. \tag{7.4.18}
\]

For example, a 2-sphere of radius \( R \) can be defined in Cartesian coordinates \( \vec{x} \) as

\[
f(\vec{x}) = R^2, \quad \text{where} \quad f(\vec{x}) = \vec{x}^2. \tag{7.4.19}
\]

Given the function \( f \), we now show that \( df = 0 \) can be used to define a unit normal \( n^i \) through

\[
n^i \equiv \frac{\nabla^i f}{\sqrt{\nabla^j f \nabla_j f}} = \frac{g^{ik} \partial_k f}{\sqrt{g^{lm} \nabla_l f \nabla_m f}}. \tag{7.4.20}
\]

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 from our previous discussion, it returns zero:

\[
\left. \frac{\partial x^i(\vec{\xi})}{\partial \xi^1} \partial_1 f(\vec{x}) \right|_{\text{on hypersurface}} = \frac{\partial}{\partial \xi^1} f \left( \vec{x}(\vec{\xi}) \right) = \partial_1 f(\vec{\xi}) = 0. \tag{7.4.21}
\]
That \( \partial f(\xi) = 0 \) is just a re-statement that \( f \) is constant on our hypersurface \( \bar{x}(\xi) \). Using \( n^i \) we can also write down the induced metric on the hypersurface as

\[
H_{ij} = g_{ij} - n_in_j. \tag{7.4.22}
\]

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

\[
H_{ij}v^iw^j = g_{ij}v_i^w_j; \tag{7.4.23}
\]

where \( v^i_\parallel \) and \( w^i_\parallel \) are \( v^i \) and \( w^i \) projected along the hyper-surface at hand. In words: \( H_{ij}v^iw^j \) is the dot product computed using the ambient metric but with the components of \( v \) and \( w \) orthogonal to the hypersurface removed. Now,

\[
v^i_\parallel = H^i_jv^j \quad \text{and} \quad w^i_\parallel = H^i_jw^j. \tag{7.4.24}
\]

That this construction of \( v^i_\parallel \) and \( w^i_\parallel \) yields vectors perpendicular to \( n^i \) is because

\[
H_{ij}n_j = (g_{ij} - n_in_j)n^j = n_i - n_i = 0. \tag{7.4.25}
\]

Furthermore, because

\[
H^i_lH^l_j = H^i_j, \tag{7.4.26}
\]

we deduce

\[
H_{ij}v^iw^j = g_{ij}H^i_ah^j_\alpha v^\alpha w^\beta. \tag{7.4.27}
\]

**Problem 7.41.** For the 2-sphere in 3-dimensional flat space, defined by eq. \( (7.4.19) \), calculate the components of the induced metric \( H_{ij} \) in eq. \( (7.4.22) \) and compare it that in eq. \( (7.4.4) \). Hint: compute \( d\sqrt{x^2} \) in terms of \( \{dx^i\} \) and exploit the constraint \( x^2 = R^2 \); then consider what is the \(- (n_i dx_i)^2 \) occurring in \( H_{ij}dx^i dx^j \), when written in spherical coordinates? \( \square \)

**Problem 7.42.** Consider some 2-dimensional surface parametrized by \( \xi^1 = (\sigma, \rho) \), whose trajectory in \( D \)-dimensional flat space is provided by the Cartesian coordinates \( \bar{x}(\sigma, \rho) \). What is the formula analogous to eq. \( (7.4.8) \), which yields the area of this 2D surface over some domain \( \mathcal{D} \) on the \((\sigma, \rho) \) plane? Hint: First ask, “what is the 2D induced metric?” Answer:

\[
\text{Area} = \int_\mathcal{D} d\sigma d\rho \sqrt{(\partial_\sigma \bar{x})^2(\partial_\rho \bar{x})^2 - (\partial_\sigma \bar{x} \cdot \partial_\rho \bar{x})^2}, \quad (\partial_i \bar{x})^2 \equiv \partial_1 x^i_\partial_1 x^j_\delta_{ij}. \tag{7.4.28}
\]

(This is not too far from the Nambu-Goto action of string theory.) \( \square \)

**Directed surface elements** What is the analog of \( d(\text{Area}) \) 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 \( H_{ij}(\xi^K) \), determinant \( H \equiv \det H_{ij} \), and a unit normal \( n^i \) to it, then the answer is

\[
d^{D-1}\Sigma_i \equiv d^{D-1}\xi \sqrt{|H(\xi)|} n_i \left( \bar{x}(\xi) \right). \tag{7.4.29}
\]

257
\[ = d^{D-1} \xi \tilde{\epsilon}_{ij_1 j_2 \ldots j_{D-1}} \left( \vec{x}(\vec{\xi}) \right) \frac{\partial x^{i_1}(\vec{\xi})}{\partial \xi^1} \frac{\partial x^{i_2}(\vec{\xi})}{\partial \xi^2} \ldots \frac{\partial x^{i_{D-1}}(\vec{\xi})}{\partial \xi^{D-1}}. \quad (7.4.30) \]

The difference between equations \[ (7.4.29) \] and \[ (7.4.30) \] is that the first requires knowing the normal vector beforehand, while the second 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 \( \{ \xi^i \} \) should be consistent with that of the ambient \( \{ \vec{x} \} \) and the infinitesimal volume \( d^D \vec{x} \sqrt{|g|} \). The \( d^{D-1} \xi \sqrt{|H|} \) is the (scalar) infinitesimal area (= \( (D - 1) \)-volume) and \( n_i \) provides the direction. The second equality requires justification. Let's define \( \{ \mathcal{E}^i_1 \}_{i = 1, 2, 3, \ldots, D - 1} \) to be the \( (D - 1) \) vector fields

\[ \mathcal{E}^i_1(\vec{\xi}) \equiv \frac{\partial x^i(\vec{\xi})}{\partial \xi^1}. \quad (7.4.31) \]

\textbf{Problem 7.43.} Show that the tensor in eq. \[ (7.4.30) \],

\[ \tilde{n}_i \equiv \tilde{\epsilon}_{ij_1 j_2 \ldots j_{D-1}} \mathcal{E}^j_1 \ldots \mathcal{E}^{j_{D-1}} \]

\[ (7.4.32) \]

is orthogonal to all the \( (D - 1) \) vectors \( \{ \mathcal{E}^i_1 \} \). Since \( n_i \) is the sole remaining direction in the \( D \) space, \( \tilde{n}_i \) must be proportional to \( n_i \)

\[ \tilde{n}_i = \varphi \cdot n_i. \quad (7.4.33) \]

To find \( \varphi \) we merely have to dot both sides with \( n^i \),

\[ \varphi(\vec{\xi}) = \sqrt{|g(\vec{x}(\vec{\xi}))|} \epsilon_{ij_1 j_2 \ldots j_{D-1}} n^i \frac{\partial x^{i_1}(\vec{\xi})}{\partial \xi^1} \ldots \frac{\partial x^{i_{D-1}}(\vec{\xi})}{\partial \xi^{D-1}}. \quad (7.4.34) \]

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 \( \{ x^2, x^3, \ldots, x^D \} \) lie on the surface itself. Argue that, in this coordinate system, eq. \[ (7.4.20) \] becomes

\[ n^i = \frac{g^{(i)(1)}}{\sqrt{g^{(1)(1)}}}. \quad (7.4.35) \]

and therefore eq. \[ (7.4.34) \] reads

\[ \varphi(\vec{\xi}) = \sqrt{|\epsilon(\vec{x}(\vec{\xi}))|} \sqrt{g^{(1)(1)}}. \quad (7.4.36) \]

Cramer's rule (cf. \[ (3.2.15) \]) from matrix algebra reads: the \( ij \) component (the \( i \)th row and \( j \)th column) of the inverse of a matrix \( (A^{-1})_{ij} \) is \( (\det A)^{-1} \) times the determinant of \( A \) with the \( j \)th row and \( i \)th column removed. Use this and the definition of the induced metric to conclude that

\[ \varphi(\vec{\xi}) = \sqrt{|H(\vec{\xi})|}, \quad (7.4.37) \]

thereby proving the equality of equations \[ (7.4.29) \] and \[ (7.4.30) \]. \( \square \)
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 \( D \), of the divergence of a vector field is equal to the flux of the same vector field across the boundary \( \partial D \) of the domain. Exactly the same statement applies in a \( D \) dimensional ambient curved space with some closed \((D - 1)\) dimensional hypersurface that defines \( \partial D \).

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 \( d^{D-1}\Sigma_i \) given by equations (7.4.29) and (7.4.30). Then

\[
\int_D d^D x \sqrt{|g(\vec{x})|} \nabla_i V^i(\vec{x}) = \int_{\partial D} d^{D-1}\Sigma_i V^i(\vec{x}(\vec{\xi})) . \tag{7.4.38}
\]

Flux  Just as in 3D vector calculus, the \( d^{D-1}\Sigma_i V^i \) can be viewed as the flux of some fluid described by \( V^i \) across an infinitesimal element of the hypersurface \( \partial D \).

Remark  Gauss’ theorem is not terribly surprising if you recognize the integrand as a total derivative,

\[
\sqrt{|g|}\nabla_i V^i = \partial_i(\sqrt{|g|} V^i) \tag{7.4.39}
\]

(recall eq. (7.3.94)) and therefore it should integrate to become a surface term (\( \equiv (D - 1)\)-dimensional integral). The right hand side of eq. (7.4.38) merely makes this surface integral explicit, in terms of the coordinates \( \vec{\xi} \) describing the boundary \( \partial D \).

Closed surface  Note that if you apply Gauss’ theorem eq. (7.4.38), 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. (7.4.38) is automatically zero.

Problem 7.44.  We may see this directly for the 2-sphere case. The metric on the 2-sphere of radius \( R \) is

\[
d\ell^2 = R^2(d\theta^2 + (\sin \theta)^2d\phi^2), \quad \theta \in [0, \pi], \ \phi \in [0, 2\pi] . \tag{7.4.40}
\]

Let \( V^i \) be an arbitrary smooth vector field on the 2-sphere. Show explicitly – namely, do the integral – that

\[
\int_{S^2} d^2 x \sqrt{|g(\vec{x})|} \nabla_i V^i = 0 . \tag{7.4.41}
\]

Hint: For the \( \phi \)-integral, remember that \( \phi = 0 \) and \( \phi = 2\pi \) refer to the same point, for a fixed \( \theta \).

Problem 7.45.  Hodge dual formulation of Gauss’ theorem in \( D \)-space.  Let us consider the Hodge dual of the vector field in eq. (7.4.38),

\[
\widetilde{V}_{i_1...i_{D-1}} \equiv \tilde{\epsilon}_{i_1...i_{D-1}j} V^j . \tag{7.4.42}
\]
First show that
\[ \mathring{\eta}_{i_1 \ldots i_{D-1}} \nabla_j \mathring{V}_{i_1 \ldots i_{D-1}} \propto \partial_{[1} \mathring{V}_{23 \ldots D]} \propto \nabla_i V^i. \] (7.4.43)

(Find the proportionality factors.) Then deduce the dual formulation of Gauss’ theorem, namely, the relationship between
\[ \int_D dx \partial_{[1} \mathring{V}_{23 \ldots D]} \] and
\[ \int_{\partial D} d\xi \mathring{V}_{i_1 \ldots i_{D-1}} \left( \mathring{x}(\xi) \right) \frac{\partial x^{i_1}(\xi)}{\partial \xi^1} \ldots \frac{\partial x^{i_{D-1}}(\xi)}{\partial \xi^{D-1}}. \] (7.4.44)

The \( \mathring{V}_{i_1 \ldots i_{D-1}} \partial_\xi x^{i_1} \ldots \partial_\xi x^{i_{D-1}} \) can be viewed as the original tensor \( \mathring{V}_{i_1 \ldots i_{D-1}} \), but projected onto the boundary \( \partial D \).

In passing, I should point out, what you have shown in eq. (7.4.44) can be written in a compact manner using differential forms notation:
\[ \int_D d\mathring{V} = \int_{\partial D} \mathring{V}, \] (7.4.45)
by viewing the fully antisymmetric object \( \mathring{V} \) as a differential \((D-1)\)-form.

**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
\[ \nabla_i \nabla^i \psi = -\delta^{(D)}(\mathring{x} - \mathring{x}'). \] (7.4.46)
in flat space. Let us consider as domain \( D \) the sphere of radius \( r \) centered at the point charge at \( \mathring{x}' \). Using spherical coordinates, \( \mathring{x} = r \mathring{n}(\xi) \), where \( \mathring{n} \) is the unit radial vector emanating from \( \mathring{x}' \), the induced metric on the boundary \( \partial D \) is simply the metric of the \((D-1)\)-sphere. We now identify in eq. (7.4.38) \( V^i = \nabla^i \psi \). The normal vector is simply \( n^i \partial_i = \partial_r \), and so Gauss’ law using eq. (7.4.29) reads
\[ -1 = \int_{S^{D-1}} d^{D-1} \mathring{x} \sqrt{|H|} r^{D-1} \partial_r \psi(r). \] (7.4.47)
The \( \int_{S^{D-1}} d^{D-1} \mathring{x} \sqrt{|H|} = 2\pi^{D/2} / \Gamma(D/2) \) is simply the solid angle subtended by the \((D-1)\)-sphere \((\equiv \text{volume of the } (D-1)\text{-sphere of unit radius})\). So at this point we have
\[ \partial_r \psi(r) = -\frac{\Gamma(D/2)}{2\pi^{D/2} r^{D-1}} \quad \Rightarrow \quad \psi(r) = \frac{\Gamma(D/2)}{4((D-2)/2)\pi^{D/2} r^{D-2}} = \frac{\Gamma\left(\frac{D}{2} - 1\right)}{4\pi^{D/2} r^{D-2}}. \] (7.4.48)
I have used the Gamma-function identity \( \Gamma(z)z = \Gamma(z+1) \). Replacing \( r \rightarrow |\mathring{x} - \mathring{x}'| \), we conclude that the Coulomb potential due to a unit strength electric charge is
\[ \psi(\mathring{x}) = \frac{\Gamma\left(\frac{D}{2} - 1\right)}{4\pi^{D/2} |\mathring{x} - \mathring{x}'|^{D-2}}, \quad D \geq 3. \] (7.4.49)
It is instructive to also use Gauss’ law using eq. (7.4.30).

\[-1 = \int_{S^{D-1}} d^{D-1}\xi\epsilon_{i_1\ldots i_{D-1}} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} g^{jk}(\vec{x}(\vec{\xi})) \partial_k \psi(r = \sqrt{\vec{x}^2}). \tag{7.4.50}\]

On the surface of the sphere, we have the completeness relation (cf. (4.3.24)):

\[g^{jk}(\vec{x}(\vec{\xi})) = \delta^{ij} \frac{\partial x^i}{\partial \xi^1} \frac{\partial x^k}{\partial \xi^j} + \frac{\partial x^i}{\partial r} \frac{\partial x^k}{\partial r}. \tag{7.4.51}\]

(This is also the coordinate transformation for the inverse metric from Cartesian to Spherical coordinates.) At this point,

\[-1 = \int_{S^{D-1}} d^{D-1}\xi\epsilon_{i_1\ldots i_{D-1}} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} \left( \delta^{ij} \frac{\partial x^j}{\partial \xi^1} \frac{\partial x^k}{\partial \xi^j} + \frac{\partial x^i}{\partial r} \frac{\partial x^k}{\partial r} \right) \partial_k \psi(r = \sqrt{\vec{x}^2})
= \int_{S^{D-1}} d^{D-1}\xi\epsilon_{i_1\ldots i_{D-1}} \frac{\partial x^{i_1}}{\partial \xi^1} \cdots \frac{\partial x^{i_{D-1}}}{\partial \xi^{D-1}} \frac{\partial x^j}{\partial \xi^j} \left( \frac{\partial x^k}{\partial r} \partial_k \psi(r = \sqrt{\vec{x}^2}) \right). \tag{7.4.52}\]

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. (7.4.29).

**Problem 7.46. Coulomb Potential in 2D**  
Use the above arguments to show, the solution to

\[\nabla_i \nabla^i \psi = -\delta^{(2)}(\vec{x} - \vec{x}') \tag{7.4.53}\]

is

\[\psi(\vec{x}) = -\frac{\ln(L^{-1}|\vec{x} - \vec{x}'|)}{2\pi}. \tag{7.4.54}\]

Here, $L$ is an arbitrary length scale. Why is there is a restriction $D \geq 3$ in eq. (7.4.49)?

**Tensor elements**  
Suppose we have a $(N < D)$-dimensional domain $\mathfrak{D}$ parametrized by $\{\vec{x}(\xi)| I = 1, 2, \ldots, N\}$ whose boundary $\partial \mathfrak{D}$ is parametrized by $\{\vec{x}(\theta)| \theta : 1, 2, \ldots, N-1\}$. We may define a $(D - N)$-tensor element that generalizes the one in eq. (7.4.30)

\[d^N\Sigma_{i_1\ldots i_{D-N}} \equiv d^N\xi \epsilon_{i_1\ldots i_{D-N} j_1\ldots j_N} \left( \vec{x}(\vec{\xi}) \right) \frac{\partial x^{j_1}(\vec{\xi})}{\partial \xi^1} \frac{\partial x^{j_2}(\vec{\xi})}{\partial \xi^2} \cdots \frac{\partial x^{j_N}(\vec{\xi})}{\partial \xi^N}. \tag{7.4.55}\]

We may further define the boundary surface element

\[d^{N-1}\Sigma_{i_1\ldots i_{D-N} k} \equiv d^{N-1}\theta \epsilon_{i_1\ldots i_{D-N} k j_1\ldots j_{N-1}} \left( \vec{x}(\vec{\theta}) \right) \frac{\partial x^{j_1}(\vec{\theta})}{\partial \theta^1} \frac{\partial x^{j_2}(\vec{\theta})}{\partial \theta^2} \cdots \frac{\partial x^{j_{N-1}}(\vec{\theta})}{\partial \theta^{N-1}}. \tag{7.4.56}\]
Stokes’ theorem\footnote{Just like for the Gauss’ theorem case, in equations \(7.4.55\) and \(7.4.56\), the \(\xi\) and \(\theta\) coordinate systems need to be defined with orientations consistent with the ambient \(d^D \vec{x} \sqrt{|g(\vec{x})|} \epsilon_{12\ldots D}\) one.}  Stokes’ theorem is the assertion that, in a \((N < D)\)-dimensional simply connected subregion \(\mathcal{D}\) of some \(D\)-dimensional ambient space, the divergence of a fully antisymmetric rank \((D - N + 1)\) tensor field \(B_{i_1\ldots i_{D-N}k}\) integrated over the domain \(\mathcal{D}\) can also be expressed as the integral of \(B_{i_1\ldots i_{D-N}k}\) over its boundary \(\partial \mathcal{D}\). Namely,

\[
\int_{\mathcal{D}} d^N \sum_{i_1\ldots i_{D-N}} \nabla_k B_{i_1\ldots i_{D-N}k} = \frac{1}{D - N + 1} \int_{\partial \mathcal{D}} d^{N-1} \sum_{i_1\ldots i_{D-N}k} B_{i_1\ldots i_{D-N}k}, \quad (7.4.57)
\]

\[N < D, \quad B_{i_1\ldots i_{D-N}k} = (D - N + 1)! B_{i_1\ldots i_{D-N}k}.\]

Problem 7.47. Hodge dual formulation of Stokes’ theorem. Define

\[
\tilde{B}_{j_1\ldots j_{N-1}} \equiv \frac{1}{(D - N + 1)!} \epsilon_{j_1\ldots j_{N-1} i_1\ldots i_{D-N}k} B_{i_1\ldots i_{D-N}k}. \quad (7.4.58)
\]

Can you convert eq. \((7.4.57)\) into a relationship between

\[
\int_{\mathcal{D}} d^N \tilde{\xi} \partial_{[i_1} \tilde{B}_{i_2\ldots i_{N-1}]} \frac{\partial x^{i_1}}{\partial \tilde{\xi}^1} \ldots \frac{\partial x^{i_{N-1}}}{\partial \tilde{\xi}^N} \text{ and } \int_{\partial \mathcal{D}} d^{N-1} \tilde{\theta} \tilde{B}_{i_1\ldots i_{N-1}} \frac{\partial x^{i_1}}{\partial \tilde{\theta}^1} \ldots \frac{\partial x^{i_{N-1}}}{\partial \tilde{\theta}^{N-1}}? \quad (7.4.59)
\]

Furthermore, explain why the Jacobians can be “brought inside the derivative”.

\[
\partial_{[i_1} \tilde{B}_{i_2\ldots i_{N-1}]} \frac{\partial x^{i_1}}{\partial \tilde{\xi}^1} \ldots \frac{\partial x^{i_{N-1}}}{\partial \tilde{\xi}^N} = \left( \frac{\partial x^{i_1}}{\partial \xi^1} \right) \ldots \left( \frac{\partial x^{i_{N-1}}}{\partial \xi^N} \right) \tilde{B}_{i_2\ldots i_{N-1}}. \quad (7.4.60)
\]

The \(|·|\) around \(i_1\) indicate it is not to be part of the anti-symmetrization; only do so for the \(\xi\)-indices.

Like for Gauss’ theorem, we point out that – by viewing \(\tilde{B}_{j_1\ldots j_{N-1}}\) as components of a \((N - 1)\)-form, Stokes’ theorem in eq. \((7.4.57)\) reduces to the simple expression

\[
\int_{\mathcal{D}} d\tilde{B} = \int_{\partial \mathcal{D}} \tilde{B}. \quad (7.4.61)
\]

Relation to 3D vector calculus. Stokes’ theorem in vector calculus states that the flux of the curl of a vector field over some 2D domain \(\mathcal{D}\) sitting in the ambient 3D space, is equal to the line integral of the same vector field along the boundary \(\partial \mathcal{D}\) of the domain. Because eq. \((7.4.57)\) may not appear, at first sight, to be related to the Stokes’ theorem from 3D vector calculus, we shall work it out in some detail.

Problem 7.48. Consider some 2D hypersurface \(\mathcal{D}\) residing in a 3D curved space. For simplicity, let us foliate \(\mathcal{D}\) with constant \(\rho\) surfaces; let the other coordinate be \(\phi\), so \(\vec{x}(0 \leq \rho \leq \rho_>, 0 \leq \phi \leq 2\pi)\) describes a given point on \(\mathcal{D}\) and the boundary \(\partial \mathcal{D}\) is given by the closed loop \(\vec{x}(\rho = \rho_>, 0 \leq \phi \leq 2\pi)\). Let

\[
B_{ik} \equiv \tilde{\epsilon}^{ikj} A_j \quad (7.4.62)
\]
for some vector field \( A^j \). This implies in Cartesian coordinates,

\[
\nabla_k B^{ik} = \left( \vec{\nabla} \times \vec{A} \right)^i.
\]

(7.4.63)

Denote \( \vec{\xi} = (\rho, \phi) \). Show that Stokes’ theorem in eq. (7.4.57) reduces to the \( N = 2 \) vector calculus case:

\[
\int_0^{\rho_0} d\rho \int_0^{2\pi} d\phi \sqrt{|H(\vec{\xi})|} \vec{n} \cdot \left( \vec{\nabla} \times \vec{A} \right) = \int_0^{2\pi} d\phi \frac{\partial \vec{x}(\rho_0, \phi)}{\partial \phi} \cdot \vec{A}(\vec{x}(\rho_0, \phi)).
\]

(7.4.64)

where the unit normal vector is given by

\[
\vec{n} = \frac{\left( \frac{\partial \vec{x}(\vec{\xi})}{\partial \rho} \right) \times \left( \frac{\partial \vec{x}(\vec{\xi})}{\partial \phi} \right)}{|\left( \frac{\partial \vec{x}(\vec{\xi})}{\partial \rho} \right) \times \left( \frac{\partial \vec{x}(\vec{\xi})}{\partial \phi} \right)|}.
\]

(7.4.65)

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.,

\[
\int \mathcal{D} d^2 \xi \sqrt{|H(\vec{\xi})|} n_i \tilde{\epsilon}_{ijk} \partial_x^j A_k = \int_{\partial \mathcal{D}} A_i dx^i.
\]

(7.4.66)

**Step-by-step guide:** Start with eq. (7.4.30), and show that in a Cartesian basis,

\[
d^2 \Sigma_i = d^2 \xi \left( \frac{\partial \vec{x}}{\partial \rho} \times \frac{\partial \vec{x}}{\partial \phi} \right)^i.
\]

(7.4.67)

The induced metric on the 2D domain \( \mathcal{D} \) is

\[
H_{ij} = \delta_{ij} \partial_1 x^i \partial_1 x^j.
\]

(7.4.68)

Work out its determinant. Then work out

\[
|(\partial \vec{x}/\partial \rho) \times (\partial \vec{x}/\partial \phi)|^2
\]

using the identity

\[
\tilde{\epsilon}^{ijk} \tilde{\epsilon}_{lmk} = \delta^j_l \delta^i_m - \delta^i_l \delta^j_m.
\]

(7.4.70)

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. (7.4.57) yields the left hand side of (7.4.64)?

For the right hand side of eq. (7.4.64), begin by arguing that the boundary (line) element in eq. (7.4.56) becomes

\[
d\Sigma_{ki} = d\phi \tilde{\epsilon}_{kij} \frac{\partial x^j}{\partial \phi}.
\]

(7.4.71)

Then use \( \tilde{\epsilon}^{ij} \tilde{\epsilon}_{kij} = 2 \delta^i_k \) to then show that the right hand side of eq. (7.4.57) is now that of eq. (7.4.64).

\[
\square
\]

263
Problem 7.49. Discuss how the tensor element in eq. (7.4.55) transforms under a change of hypersurface coordinates $\xi \rightarrow \xi'(\tilde{\theta})$. Do the same for the tensor element in eq. (7.4.56): how does it transforms under a change of hypersurface coordinates $\tilde{\theta} \rightarrow \tilde{\theta}'$?

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 $\vec{A}$.

$$\nabla \cdot \vec{B} = 0 \quad \Leftrightarrow \quad \vec{B} = \nabla \times \vec{A}. \quad (7.4.72)$$

And, the curl of a vector $\vec{B}$ is zero everywhere in space iff it is the gradient of scalar $\psi$.

$$\nabla \times \vec{B} = 0 \quad \Leftrightarrow \quad \vec{B} = \nabla \psi. \quad (7.4.73)$$

Here, we shall see that these statements are special cases of the following.

Poincaré Lemma In an arbitrary $D$ dimensional curved space, let $B_{i_1...i_N}(x)$ be a fully antisymmetric rank-$N$ tensor field, with $N \leq D$. Then, everywhere within a simply connected region of space,

$$B_{i_1...i_N} = \partial_{[i_1} C_{i_2...i_N]}, \quad (7.4.74)$$

– i.e., $B$ is the “curl” of a fully antisymmetric rank-$(N-1)$ tensor $C$ – if and only if

$$\partial_{[i} B_{i_1...i_N]} = 0. \quad (7.4.75)$$

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,

$$dB = 0 \text{ iff } B = dC. \quad (7.4.76)$$

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 $B^i$:

$$\tilde{B}^{ij} \equiv \tilde{\epsilon}^{ijk} B_k. \quad (7.4.77)$$

The Poincaré Lemma says $\tilde{B}^{ij} = \partial_{[i} A_{j]}$ if and only if $\partial_{[k} \tilde{B}^{ij]} = 0$ everywhere in space. We shall proceed to take the dual of these two conditions. Via eq. (7.3.139), the first is equivalent to

$$\tilde{\epsilon}^{kij} \tilde{B}^{ij} = \tilde{\epsilon}^{kij} \partial_{[i} A_{j]},
= 2\tilde{\epsilon}^{kij} \partial_{[i} A_{j]}, \quad (7.4.78)$$

On the other hand, employing eq. (7.3.139),

$$\tilde{\epsilon}^{kij} \tilde{B}^{ij} = \tilde{\epsilon}^{kij} \tilde{\epsilon}_{ij} B^i = 2 B^k, \quad (7.4.79)$$

and therefore $\tilde{B}$ is the curl of $A_i$:

$$B^k = \tilde{\epsilon}^{kij} \partial_{i} A_{j}. \quad (7.4.80)$$
While the latter condition $d\vec{B} = 0$ is, again utilizing eq. (7.3.139), equivalent to

$$0 = \varepsilon^{kij} \partial_k \tilde{B}_{ij}$$

$$= \varepsilon_{kij} \varepsilon^{ijl} \nabla_k B_l = 2 \nabla_l B^l. \quad (7.4.81)$$

That is, the divergence of $\vec{B}$ is zero.

**Example II** A simple application is that of the line integral

$$I(\vec{x}, \vec{x}'; \mathfrak{P}) \equiv \int_{\mathfrak{P}} A_i dx^i, \quad (7.4.82)$$

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 $\varphi$. Then $A_i dx^i = \partial_i \varphi dx^i = d\varphi$, and the integral itself is actually path independent – it depends only on the end points:

$$\int_{\vec{x}'}^{\vec{x}} A_i dx^i = \int_{\mathfrak{P}} d\varphi = \varphi(\vec{x}) - \varphi(\vec{x}'), \quad \text{whenever } \partial_i A_j = 0. \quad (7.4.83)$$

**Problem 7.50.** Make a similar translation, from the Poincaré Lemma, to the 3D vector calculus statement that a vector $B$ is curl-less if and only if it is a pure gradient everywhere.

**Problem 7.51.** Consider the vector potential, written in 3D Cartesian coordinates,

$$A_i dx^i = \frac{x^1 dx^2 - x^2 dx^1}{(x^1)^2 + (x^2)^2}. \quad (7.4.84)$$

Can you calculate

$$F_{ij} = \partial_i A_j? \quad (7.4.85)$$

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 to show that

$$F_{ij} = 2\pi \epsilon_{i3j} \delta(x^1) \delta(x^2)\? \quad (7.4.86)$$

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. (7.4.66) 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.
Differential Geometry In Curved Spacetimes

We now move on to differential geometry in curved spacetimes. I assume the reader is familiar with basic elements of Special Relativity and with the discussion in §7 – in many instances, I will simply bring over the results from there to the curved spacetime context. In §8.1 I discuss Lorentz/Poincaré symmetry in flat spacetime, since it is fundamental to both Special and General Relativity. I then cover curved spacetime differential geometry proper from §8.3 through §8.5, focusing on issues not well developed in §7. These three sections, together with §7, are intended to form the first portion – the kinematics of curved space(time) – of a course on gravitation. Following that, §8.6 contains somewhat specialized content regarding the expansion of geometric quantities off some fixed ‘background’ geometry; and finally, in §8.7 we compile conformal transformation properties of geometric objects.

8.1 Constancy of \(c\), Poincaré and Lorentz symmetry

We begin in flat (aka Minkowski) spacetime written in Cartesian coordinates \(\{x^\mu \equiv (t, \vec{x})\}\). The ‘square’ of the distance between \(x^\mu\) and \(x^\mu + dx^\mu\), is given by a ‘modified Pythagoras’ theorem’ of sorts:

\[
ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = (dt)^2 - \delta_{ij} dx^i dx^j; \tag{8.1.1}
\]

where the Minkowski metric tensor reads

\[
\eta_{\mu\nu} \equiv \text{diag}[1, -1, \ldots, -1]. \tag{8.1.2}
\]

The inverse metric \(\eta^{\mu\nu}\) is simply the matrix inverse, \(\eta^{\mu\sigma} \eta_{\sigma\beta} = \delta^{\mu}_{\beta}\); it is numerically equal to the flat metric itself:

\[
\eta^{\mu\nu} \equiv \text{diag}[1, -1, \ldots, -1]. \tag{8.1.3}
\]

Strictly speaking we should be writing eq. (8.1.1) in the ‘dimensionally-correct’ form

\[
ds^2 = c^2 dt^2 - d\vec{x} \cdot d\vec{x}; \tag{8.1.4}
\]

where \(c\) is the speed of light and \([ds^2] = [\text{Length}^2]\). 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 §[7], say \(i, j, k, \ldots \in \{1, 2, 3, \ldots, D\}\) to Greek ones \(\mu, \nu, \ldots \in \{0, 1, 2, \ldots, D \equiv d - 1\}\); the former run over the spatial coordinates while the latter over time (0th) and space (1, \ldots, \(D\)). Also note that the opposite ‘mostly plus’ sign convention \(\eta_{\mu\nu} = \text{diag}[-1, +1, \ldots, +1]\) is equally valid and, in fact, more popular in the contemporary physics literature.

\(^{78}\)As opposed to the dynamics of spacetime, which involves studying General Relativity, Einstein’s field equations for the metric, and its applications.
**Constancy of c**   
One of the primary motivations that led Einstein to recognize eq. (8.1.1) 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. That is, for transformation matrices $\{\Lambda\}$ satisfying

$$\Lambda^\alpha_\mu \Lambda^\beta_\nu \eta_{\alpha\beta} = \eta_{\mu\nu}, \quad (8.1.5)$$

and constant vectors $\{a^\mu\}$ we have

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = \eta_{\mu\nu} dx'^\mu dx'^\nu \quad (8.1.6)$$

whenever

$$x^\alpha = \Lambda^\alpha_\mu x'^\mu + a^\alpha. \quad (8.1.7)$$

The physical interpretation is that the frames parametrized by $\{x^\mu = (t, \vec{x})\}$ and $\{x'^\mu = (t', \vec{x}')\}$ are inertial frames: compact bodies with no external forces acting on them will sweep out geodesics $d^2 x^\mu/d\tau^2 = 0 = d^2 x'^\mu/d\tau'^2$, where the proper times $\tau$ and $\tau'$ are defined through the relations $d\tau = dt \sqrt{1 - (d\vec{x}/dt)^2}$ and $d\tau' = dt' \sqrt{1 - (d\vec{x}'/dt')^2}$. To interpret physical phenomenon taking place in one frame from the other frame’s perspective, one would first have to figure out how to translate between $x$ and $x'$.

Let $x^\mu$ be the spacetime Cartesian coordinates of a single photon; in a different Lorentz frame it has Cartesian coordinates $x'^\mu$. Invoking its null character, namely $ds^2 = 0$ – which holds in any inertial frame – we have $(dx^0)^2 = d\vec{x} \cdot d\vec{x}$ and $(dx'^0)^2 = d\vec{x}' \cdot d\vec{x}'$. This in turn tells us the speeds in both frames are unity:

$$\frac{|d\vec{x}|}{dx^0} = \frac{|d\vec{x}'|}{dx'^0} = 1. \quad (8.1.8)$$

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. (8.1.1) – together with its Lorentz invariance – that gives rise to the wave equations obeyed by the photon. Equation (8.1.8) then follows as a consequence.

**Problem 8.1.** Explain why eq. (8.1.5) is equivalent to the matrix equation

$$\Lambda^T \eta \Lambda = \eta. \quad (8.1.9)$$

Hint: What are $\eta_{\mu\nu} \Lambda^\nu_\beta$ and $A^\nu_\beta B_{\nu\gamma}$ in matrix notation?

**Moving indices**   
Just like in curved/flat space, tensor indices in flat spacetime are moved with the metric $\eta_{\mu\nu}$ and its inverse $\eta^{\mu\nu}$. For example,

$$v^\mu = \eta^{\mu\nu} v_\nu, \quad v_\mu = \eta_{\mu\nu} v^{\nu}; \quad (8.1.10)$$

$$T_{\mu\nu} = \eta_{\mu\alpha} \eta_{\nu\beta} T^{\alpha\beta}, \quad T^{\mu\nu} = \eta^{\mu\alpha} \eta^{\nu\beta} T_{\alpha\beta}. \quad (8.1.11)$$

**Symmetries**   
We shall define Poincaré transformations $\{x(x')\}$ to be the set of all coordinate transformations that leave the flat spacetime metric invariant (cf. eq. (8.1.6)). Poincaré and

---

79Poincaré transformations are also sometimes known as inhomogeneous Lorentz transformations.
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 study it in some detail. As we will now proceed to demonstrate, the most general invertible Poincaré transformation is in fact the one in eq. (8.1.7).

### Derivation of eq. (8.1.6)

Now, under a coordinate transformation, eq. (8.1.6) reads

$$\eta_{\mu \nu} dx^\mu dx^\nu = \eta_{\alpha \beta} \frac{\partial x^\mu}{\partial x'^\alpha} dx'^\alpha dx'^\beta = \eta_{\alpha' \beta'} dx'^\alpha dx'^\beta.$$  \hspace{1cm} (8.1.12)

Let us differentiate both sides of eq. (8.1.12) with respect to $x'^\sigma$.

$$\eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x'^\beta} + \eta_{\mu \nu} \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial^2 x'^\nu}{\partial x'^\alpha \partial x'^\beta} = 0.$$  \hspace{1cm} (8.1.13)

Next, consider symmetrizing $\sigma\alpha$ and anti-symmetrizing $\sigma\beta$.

$$2\eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x'^\beta} + \eta_{\mu \nu} \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial^2 x'^\nu}{\partial x'^\alpha \partial x'^\beta} = 0.$$  \hspace{1cm} (8.1.14)

$$\eta_{\mu \nu} \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial^2 x'^\nu}{\partial x'^\alpha \partial x'^\beta} - \eta_{\mu \nu} \frac{\partial^2 x'^\nu}{\partial x'^\alpha \partial x'^\beta} = 0.$$  \hspace{1cm} (8.1.15)

Since partial derivatives commute, the second term from the left of eq. (8.1.13) vanishes upon anti-symmetrization of $\sigma\beta$. Adding equations (8.1.14) and (8.1.15) hands us

$$3\eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x'^\beta} + \eta_{\mu \nu} \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial^2 x'^\nu}{\partial x'^\alpha \partial x'^\beta} = 0.$$  \hspace{1cm} (8.1.16)

Finally, subtracting eq. (8.1.13) from eq. (8.1.16) produces

$$2\eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x'^\beta} = 0.$$  \hspace{1cm} (8.1.17)

Because we have assumed Poincaré transformations are invertible, we may contract both sides with $\partial x'^\beta / \partial x^\kappa$.

$$\eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x'^\beta} \frac{\partial x'^\beta}{\partial x^\kappa} = \eta_{\mu \nu} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} \delta^\nu_\kappa = 0.$$  \hspace{1cm} (8.1.18)

Finally, we contract both sides with $\eta^{\kappa \rho}$:

$$\eta_{\mu \nu} \eta^{\kappa \rho} \frac{\partial^2 x^\mu}{\partial x'^\sigma \partial x'^\alpha} = \frac{\partial^2 x^\rho}{\partial x'^\sigma \partial x'^\alpha} = 0.$$  \hspace{1cm} (8.1.19)

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. (8.1.7). Plugging eq. (8.1.7) into eq. (8.1.12), we recover the necessary definition of the Lorentz transformation in eq. (8.1.5).

---

80 This argument can be found in Weinberg [20].
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 (8.1.5) plus constant spacetime translations.

(Homogeneous) Lorentz Transformations form a Group

If \( \Lambda^\mu_\alpha \) and \( \Lambda'^\mu_\alpha \) denotes different Lorentz transformations, then notice the composition

\[
\Lambda''^\mu_\alpha \equiv \Lambda^\mu_\sigma \Lambda'^\sigma_\alpha
\]

(8.1.20)

is also a Lorentz transformation. For, keeping in mind the fundamental definition in eq. (8.1.5), we may directly compute

\[
\Lambda''^\mu_\alpha \Lambda''^\nu_\beta \eta_{\mu \nu} = \Lambda^\mu_\sigma \Lambda'^\sigma_\alpha \Lambda'^\nu_\rho \Lambda'^\rho_\beta \eta_{\mu \nu} = \Lambda'^\sigma_\alpha \Lambda'^\rho_\beta \eta_{\sigma \rho} = \eta_{\alpha \beta}.
\]

(8.1.21)

To summarize:

The set of all Lorentz transformations \( \{ \Lambda^\mu_\alpha \} \) satisfying eq. (8.1.5), together with the composition law in eq. (8.1.20) 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 \( \Lambda^\mu_\sigma \Lambda'^\rho_\rho \Lambda''^\rho_\nu \) 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:

\[
\Lambda \cdot \Lambda' \cdot \Lambda'' = \Lambda \cdot (\Lambda' \cdot \Lambda'') = (\Lambda \cdot \Lambda') \cdot \Lambda''.
\]

(8.1.22)

• Identity \( \delta^\mu_\alpha \) is the identity Lorentz transformation:

\[
\delta^\mu_\sigma \Lambda^\sigma_\nu = \Lambda^\mu_\sigma \delta^\sigma_\nu = \Lambda^\mu_\nu, \tag{8.1.23}
\]

and

\[
\delta^\mu_\alpha \delta^\nu_\beta \eta_{\mu \nu} = \eta_{\alpha \beta}. \tag{8.1.24}
\]

• Inverse Let us take the determinant of both sides of eq. (8.1.5) – by viewing the latter as matrix multiplication, we have \( \Lambda^T \cdot \eta \cdot \Lambda = \eta \), which in turn means

\[
(\det \Lambda)^2 = 1 \quad \Rightarrow \quad \det \Lambda = \pm 1. \tag{8.1.25}
\]

81 Refer to §[B] for the defining axioms of a Group.
Here, we have recalled $\det A^T = \det A$ for any square matrix $A$. Since the determinant of $\Lambda$ is strictly non-zero, what eq. (8.1.25) teaches us is that $\Lambda$ is always invertible: $\Lambda^{-1}$ is guaranteed to exist. What remains is to check that, if $\Lambda$ is a Lorentz transformation, so is $\Lambda^{-1}$. Starting with the matrix form of eq. (8.1.9), and utilizing $(\Lambda^{-1})^T = (\Lambda^T)^{-1}$,

\begin{align}
\Lambda^T \eta \Lambda &= \eta \\
(\Lambda^T)^{-1} \Lambda^T \eta \Lambda^{-1} &= (\Lambda^T)^{-1} \cdot \eta \cdot \Lambda^{-1} \\
\eta &= (\Lambda^{-1})^T \cdot \eta \cdot \Lambda^{-1}.
\end{align}

**Problem 8.2.** Remember that indices are moved with the metric, so for example,

$\Lambda^\mu_\alpha \eta_{\mu\nu} = \Lambda^\nu_\alpha$. 

(8.1.29)

First explain how to go from eq. (8.1.5) to

$\Lambda^{\sigma}_\alpha \Lambda^{\sigma}_\beta = \delta^\alpha_\beta$ 

(8.1.30)

and deduce the inverse Lorentz transformation

$(\Lambda^{-1})^\alpha_\beta = \Lambda^\beta_\alpha = \eta_{\beta\nu} \eta^{\alpha\mu} \Lambda^\nu_\mu$. 

(8.1.31)

(Recall the inverse always exists because $\det \Lambda = \pm 1$.)

**Problem 8.3.** Explain why

$\Lambda^\mu_\alpha \Lambda^\nu_\beta \eta^{\alpha\beta} = \eta^{\mu\nu}$. 

(8.1.36)

Hint: Start from eq. (8.1.28).

**Problem 8.4.** Under the Poincaré transformation in eq. (8.1.7), show that

$\eta^{\mu\nu} \partial^\mu \partial_\nu = \eta^{\mu\nu} \partial^\mu \partial_\nu$;

(8.1.37)

where $\partial_\mu \equiv \partial / \partial x^\mu$ and $\partial_{\mu'} \equiv \partial / \partial x^{\mu'}$. How does

$\partial^\mu \equiv \eta^{\mu\nu} \partial_\nu$ 

(8.1.38)

transform under eq. (8.1.7)?
Problem 8.5. Prove that the Poincaré transformation in eq. (8.1.7) also defines a group. To systemize the discussion, first promote the spacetime coordinates to \(d + 1\) dimensional objects: \(x^\alpha = (x^\mu, 1)\) and \(x'^\alpha = (x'^\mu, 1)\), with \(\alpha = 0, 1, 2, \ldots, d - 1, d\). Then define the matrix

\[
\Pi_\alpha^\beta [\Lambda, a] = \begin{bmatrix}
\Lambda_\mu^\nu & a^\mu \\
0 & 1
\end{bmatrix};
\]  

(8.1.39)

namely, its upper left \(d \times d\) block is simply the Lorentz transformation \(\Lambda_\mu^\nu\); while its rightmost column is \((a^\mu, 1)^T\) and its bottom row is \((0, \ldots, 0, 1)\). First check that \(x^\alpha = \Pi_\alpha^\beta [\Lambda, a] x'^\beta\) is equivalent to eq. (8.1.7). Then proceed to verify that these set of matrices \(\{\Pi_\alpha^\beta [\Lambda, a]\}\) for different Lorentz transformations \(\Lambda\) and translation vectors \(a\), with the usual rules of matrix multiplication, together define a group.

Lorentzian ‘inner product’ is preserved

That \(\Lambda\) is a Lorentz transformation means it is a linear operator that preserves the Lorentzian inner product. For suppose \(v\) and \(w\) are arbitrary vectors, the inner product of \(v' \equiv \Lambda v\) and \(w' \equiv \Lambda w\) is that between \(v\) and \(w\).

\[
v' \cdot w' \equiv \eta_{\alpha\beta} v'^\alpha w'^\beta = \eta_{\alpha\beta} \Lambda_\mu^\alpha \Lambda_\nu^\beta v^\mu w^\nu
\]

\[
= \eta_{\mu\nu} v^\mu w^\nu = v \cdot w.
\]  

(8.1.40)

(8.1.41)

This 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 \(\hat{R}^T \hat{R} = I_{D \times D}\), where \(\vec{v}' \equiv \hat{R} \vec{v}\) and \(\vec{w}' \equiv \hat{R} \vec{w}\).

We wish to study in some detail what the most general form \(\Lambda_\mu^\alpha\) may take. To this end, we shall do so by examining how it acts on some arbitrary vector field \(v^\mu\). Even though this section deals with Minkowski spacetime, this \(v^\mu\) may also be viewed as a vector in a curved spacetime written in an orthonormal basis.

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 \(\hat{R}\), obeying \(\hat{R}^T \hat{R} = I\), such that

\[
\hat{R}^i_j v^j \equiv \pm |\vec{v}| (1, 0, \ldots, 0)^T, \quad |\vec{v}| \equiv \sqrt{\delta_{ij} v^i v^j}.
\]  

(8.1.42)

Conversely, since \(\hat{R}\) is necessarily invertible, any spatial vector \(v^i\) can be obtained by rotating it from \(|\vec{v}|(1, 0^T)\). Moreover, in \(D + 1\) notation, these rotation matrices can be written as

\[
\hat{R}^\mu_\nu = \begin{bmatrix}
1 & 0^T \\
0 & \hat{R}^i_j
\end{bmatrix}
\]

(8.1.43)

\[
\hat{R}^0_\nu v^\nu = v^0,
\]

(8.1.44)

\[
\hat{R}^i_\nu v^\nu = \hat{R}^i_j v^j = (\pm |\vec{v}|, 0, \ldots, 0)^T.
\]  

(8.1.45)

\[\text{This } \hat{R} \text{ is not unique: for example, by choosing another rotation matrix } \hat{R}'' \text{ that only rotates the space orthogonal to } v^i, \hat{R}'' \hat{R} v \text{ and } \hat{R} v \text{ both yield the same result.}\]
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

\[
\Lambda \begin{bmatrix}
v_0 \\
v_1 \\
\vdots \\
v_D \\
\end{bmatrix} = \Lambda \begin{bmatrix}
\frac{1}{0} \\
0 \\
\end{bmatrix} \begin{bmatrix}
\frac{v_0}{0} \\
\pm |\vec{v}| \\
\end{bmatrix}.
\]

(8.1.46)

And if the result of this matrix multiplication yields non-zero spatial components, namely \((v_0', v_1', \ldots, v_D')^T\), we may again find a rotation matrix \(\hat{R}'\) such that

\[
\Lambda \begin{bmatrix}
v_0 \\
v_1 \\
\vdots \\
v_D \\
\end{bmatrix} = \Lambda \begin{bmatrix}
\frac{v_0}{0} \\
\pm |\vec{v}'| \\
\end{bmatrix}.
\]

(8.1.47)

At this point, we have reduced our study of Lorentz transformations to

\[
\begin{bmatrix}
1 \\
0 \\
\end{bmatrix} \begin{bmatrix}
\hat{0}^T \\
\hat{R}'^T \\
\end{bmatrix} \Lambda \begin{bmatrix}
\frac{1}{0} \\
\hat{0} \\
\end{bmatrix} \begin{bmatrix}
v_0 \\
v_1 \\
\vdots \\
v_D \\
\end{bmatrix} = \Lambda' \begin{bmatrix}
v_0 \\
v_1 \\
\vdots \\
v_D \\
\end{bmatrix} = \begin{bmatrix}
v_0 \\
v_1 \\
\vdots \\
v_D \\
\end{bmatrix}.
\]

(8.1.48)

Because \(\Lambda\) was arbitrary so is \(\Lambda'\), since one can be gotten from another via rotations.

**Time Reversal & Parity Flips**

Suppose the time component of the vector \(v^\mu\) were negative \((v^0 < 0)\), we may write it as

\[
\begin{bmatrix}
-|v_0| \\
\vec{v} \\
\end{bmatrix} = \hat{T} \begin{bmatrix}
|v_0| \\
\vec{v} \\
\end{bmatrix}, \quad \hat{T} \equiv \begin{bmatrix}
-1 & \frac{0^T}{0} \\
0 & I_{D \times D} \\
\end{bmatrix};
\]

(8.1.49)

where \(\hat{T}\) is the time reversal matrix since it reverses the sign of the time component of the vector. You may readily check that \(\hat{T}\) itself is a Lorentz transformation in that it satisfies \(\hat{T}^T \eta \hat{T} = \eta\).

**Problem 8.6. Parity flip of the \(i\)th axis**

Suppose we wish to flip the sign of the \(i\)th spatial component of the vector, namely \(v^i \rightarrow -v^i\). You can probably guess, this may be implemented via the diagonal matrix with all entries set to unity, except the \(i\)th component which is set instead to \(-1\).

\[
i \hat{P}_\mu v^\nu = v^\mu, \quad \mu \neq i, \quad \text{(8.1.50)}
\]

\[
i \hat{P}_\mu v^\nu = -v^i, \quad \text{(8.1.51)}
\]

\[
i \hat{P} \equiv \text{diag}[1, 1, \ldots, 1, -1, 1, \ldots, 1]. \quad \text{(8.1.52)}
\]

Define the rotation matrix \(\hat{R}_\mu^\nu\) such that it leaves all the axes orthogonal to the 1st and \(i\)th invariant, namely

\[
\hat{R}_\mu^\nu \vec{e}_\ell = \vec{e}_\ell',
\]

(8.1.53)
while rotating the \((1, i)\)-plane clockwise by \(\pi/2\):

\[
\hat{R} \cdot \hat{e}_1 = -\hat{e}_i, \quad \hat{R} \cdot \hat{e}_i = +\hat{e}_1. \tag{8.1.55}
\]

Now argue that

\[
i \hat{P} = \hat{R}^T \cdot 1 \hat{P} \cdot \hat{R}. \tag{8.1.56}
\]

Is \(i \hat{P}\) a Lorentz transformation?

**Lorentz Boosts**

As already discussed, we may focus on the 2D case to elucidate the form of the most general Lorentz boost. This is the transformations that would mix time and space components, and yet leave the metric of spacetime \(\eta_{\mu\nu} = \text{diag}[1, -1]\) invariant. (Neither time reversal, parity flips, nor spatial rotations mix time and space.) 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 translated 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 8.7.** We wish to find a \(2 \times 2\) matrix \(\Lambda\) that obeys \(\Lambda^T \cdot \eta \cdot \Lambda = \eta\), where \(\eta_{\mu\nu} = \text{diag}[1, -1]\). By examining the diagonal terms of \(\Lambda^T \cdot \eta \cdot \Lambda = \eta\), show that

\[
\Lambda = \begin{bmatrix}
\sigma_1 \cosh(\xi_1) & \sigma_2 \sinh(\xi_2) \\
\sigma_3 \sinh(\xi_1) & \sigma_4 \cosh(\xi_2)
\end{bmatrix}, \tag{8.1.57}
\]

where the \(\sigma_{1,2,3,4}\) are either +1 or -1; altogether, there are 16 choices of signs. (Hint: \(x^2 - y^2 = c^2\), for constant \(c\), describes a hyperbola on the \((x, y)\) plane.) From the off diagonal terms of \(\Lambda^T \cdot \eta \cdot \Lambda = \eta\), argue that either \(\xi_1 = \xi_2 \equiv \xi\) or \(\xi_1 = -\xi_2 \equiv \xi\). Then explain why, if \(\Lambda^0_0\) were not positive, we can always multiply it by a time reversal matrix to render it so; and likewise \(\Lambda^1_1\) can always be rendered positive by multiplying it by a parity flip. By requiring \(\Lambda^0_0\) and \(\Lambda^1_1\) be both positive, therefore, prove that the resulting 2D Lorentz boost is

\[
\Lambda^\mu_\nu(\xi) = \begin{bmatrix}
\cosh(\xi) & \sinh(\xi) \\
\sinh(\xi) & \cosh(\xi)
\end{bmatrix}. \tag{8.1.58}
\]

This \(\xi\) is known as *rapidity*. In 2D, the rotation matrix is

\[
\hat{R}^i_j(\theta) = \begin{bmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{bmatrix}; \tag{8.1.59}
\]

and therefore rapidity \(\xi\) is to the Lorentz boost in eq. (8.1.58) what the angle \(\theta\) is to rotation \(\hat{R}^i_j(\theta)\) in eq. (8.1.59).
2D Lorentz Group: In (1+1)D, the continuous boost in \( \Lambda_{\mu}^{\nu}(\xi) \) in eq. (8.1.58) and the discrete time reversal and spatial reflection operators

\[
\hat{T} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \hat{P} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix};
\]

(altogether form the full set of Lorentz transformations – i.e., all solutions to eq. (8.1.5) consist of products of these three matrices.

To understand the meaning of the rapidity \( \xi \), let us consider applying it to an arbitrary 2D vector \( U^\mu \).

\[
U' \equiv \Lambda \cdot U = \begin{bmatrix} U^0 \cosh(\xi) + U^1 \sinh(\xi) \\ U^1 \cosh(\xi) + U^0 \sinh(\xi) \end{bmatrix}.
\] (8.1.61)

**Lorentz Boost: Timelike case** A vector \( U^\mu \) is timelike if \( U^2 \equiv \eta_{\mu\nu} U^\mu U^\nu > 0 \); this often corresponds to vector tangent to the worldline of some material object. We will now show that it is always possible to Lorentz boost to its ‘rest frame’, namely \( U'^\mu = \Lambda_{\mu}^{\nu} U^\nu = (U^0, 0) \).

In 2D, \( U^2 > 0 \Rightarrow (U^0)^2 > (U^1)^2 \Rightarrow |U^0/U^1| > 1 \). Then it is not possible to find a finite \( \xi \) such that \( U'^0 = 0 \), because that would amount to solving \( \tanh(\xi) = -U^0/U^1 \) but \( \tanh(\xi) \) lies between \(-1\) and \(+1\) while \(-U^0/U^1\) is either less than \(-1\) or greater than \(+1\). On the other hand, it does mean we may solve for \( \xi \) that would set the spatial component to zero: \( \tanh(\xi) = -v \). Recalling that tangent vectors may be interpreted as the derivative of the spacetime coordinates with respect to some parameter \( \lambda \), namely \( U^\mu \equiv dx^\mu/d\lambda \). Therefore

\[
\frac{U^1}{U^0} = \frac{dx^1}{d\lambda} \frac{dx^0}{d\lambda} = \frac{d\lambda}{dx^0} \equiv v
\] (8.1.62)

is the velocity associated with \( U^\mu \) in the frame \( \{x^\mu\} \). Starting from \( \tanh(\xi) = -v \), some algebra would then hand us (cf. eq. (8.1.58))

\[
\cosh(\xi) = \gamma \equiv \frac{1}{\sqrt{1-v^2}},
\] (8.1.63)

\[
\sinh(\xi) = -\gamma \cdot v = -\frac{v}{\sqrt{1-v^2}},
\] (8.1.64)

\[
\Lambda_{\mu}^{\nu} = \begin{bmatrix} \gamma & -\gamma \cdot v \\ -\gamma \cdot v & \gamma \end{bmatrix}.
\] (8.1.65)

This in turn yields

\[
U' = \left( \text{sgn}(U^0) \sqrt{\eta_{\mu\nu} U^\mu U^\nu}, 0 \right)^T
\] (8.1.66)

leading us to interpret the \( \Lambda_{\mu}^{\nu} \) we have found in eq. (8.1.65) as the boost that brings observers to the frame where the flow associated with \( U^\mu \) is ‘at rest’. (Note that, if \( U^\mu = dx^\mu/d\tau \), where \( \tau \) is proper time, then \( \eta_{\mu\nu} U^\mu U^\nu = 1 \).)
As an important aside, we may generalize the two-dimensional Lorentz boost in eq. (8.1.65) 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:

\[
\Lambda^\mu_\nu = \begin{pmatrix}
\gamma & -\gamma \cdot v & 0 \\
-\gamma \cdot v & \gamma & 0 \\
0 & 0 & I_{(D-2) \times (D-2)}
\end{pmatrix}.
\] (8.1.67)

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. (8.1.65) to the spatial vectors \(\Lambda^0_i\) and \(\Lambda^i_0\) parallel to \(v^i\).

Whereas the \((1,1)\) component of eq. (8.1.65) 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 \(I_{(D-2) \times (D-2)}\) in eq. (8.1.67), 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

\[
\Lambda^\mu_\nu(\vec{v}) = \begin{pmatrix}
\gamma & -\gamma \cdot v & \gamma v^i \\
-\gamma \cdot v & \gamma & \gamma v^j \\
\gamma \cdot v^i & \gamma \cdot v^j & \delta_{ij} - v^i v^j / \vec{v}^2
\end{pmatrix}, \quad \vec{v}^2 \equiv \delta_{ab} v^a v^b.
\] (8.1.68)

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. (8.1.68) describes

\[
U'^\mu = \Lambda^\mu_\nu(\vec{v}) U^\nu, \quad (8.1.69)
\]

\[
U^\mu = \frac{dx^\mu}{d\lambda}, \quad U'^\mu = \frac{dx'^\mu}{d\lambda} = \left( \text{sgn}(U^0) \sqrt{\eta_{\mu\nu}U^\mu U^\nu}, 0 \right)^T. \quad (8.1.70)
\]

\(\lambda\) is the intrinsic 1D coordinate parametrizing the worldlines, and by definition does not alter under Lorentz boost. The above statement is therefore equivalent to

\[
dx'^\mu = \Lambda^\mu_\nu(\vec{v}) dx^\nu, \quad (8.1.71)
\]

\[
x'^\mu = \Lambda^\mu_\nu(\vec{v}) x^\nu + a^\mu, \quad (8.1.72)
\]

where the spacetime translation \(a^\mu\) shows up here as integration constants.

**Problem 8.8. 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. (8.1.68) produces a \(\Lambda^\mu_\nu v^\nu\) that has no non-trivial spatial components. Also check that eq. (8.1.68) is, in fact, a Lorentz transformation. What is \(\Lambda^\mu_\sigma(\vec{v}) \Lambda^\sigma_\nu(-\vec{v})\)?

**Lorentz Boost: Spacelike case**  A vector \(U^\mu\) is spacelike if \(U^2 \equiv \eta_{\mu\nu}U^\mu U^\nu < 0\). As we will now show, it is always possible to find a Lorentz boost so that \(U'^\mu = \Lambda^\mu_\nu U^\nu = (0, \vec{U}')\) 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, \( U^2 < 0 \Rightarrow (U^0)^2 < (U^1)^2 \Rightarrow |U^1/U^0| = |dx^1/dx^0| \equiv |v| > 1. \) Then, recalling eq. (8.1.61), it is not possible to find a finite \( \xi \) such that \( U^0' = 0 \), 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

\[
\begin{align*}
\cosh(\xi) &= (1 - v^{-2})^{-1/2} = \frac{|v|}{\sqrt{v^2 - 1}}, \\
\sinh(\xi) &= -(1/v) (1 - v^{-2})^{-1/2} = -\frac{\text{sgn}(v)}{\sqrt{v^2 - 1}}, \\
U' &= \left(0, \text{sgn}(v)\sqrt{-\eta_{\mu\nu}U^\mu U^\nu}\right)^T, \quad v \equiv \frac{U^1}{U^0}.
\end{align*}
\]

(8.1.73)

(8.1.74)

(8.1.75)

We may interpret \( U^\mu' \) and \( U^\mu \) as infinitesimal vectors joining the same pair of spacetime points but in their respective frames. Specifically, \( U^\mu' \) are the components in the frame where the pair lies on the same constant-time surface (\( U^0 = 0 \)). While \( U^\mu \) are the components in a boosted frame.

**Lorentz Boost: Null (aka lightlike) case** The vector \( U^\mu \) is null if \( U^2 = \eta_{\mu\nu}U^\mu U^\nu = 0 \). If \( U \) were null in 2D, that means \((U^0)^2 = (U^1)^2\), which in turn implies

\[
U^\mu = \omega(1, \pm 1)
\]

(8.1.76)

for some real number \( \omega \). Upon a Lorentz boost, eq. (8.1.61) tells us

\[
U' \equiv \Lambda \cdot U = \omega \begin{bmatrix} \cosh(\xi) \pm \sinh(\xi) \\ \sinh(\xi) \pm \cosh(\xi) \end{bmatrix}.
\]

(8.1.77)

As we shall see below, if \( U^\mu \) describes the \( d \)--momentum of a photon, so that \( |\omega| \) is its frequency in the un-boosted frame, the \( U^0'/U^0 = \cosh(\xi) \pm \sinh(\xi) \) describes the photon’s red- or blue-shift in the boosted frame. Notice it is not possible to set either the time nor the space component to zero, unless \( \xi \to \pm \infty \).

**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. 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 8.9. Null, spacelike vs. timelike** Do null vectors form a vector space? Similarly, do spacelike or timelike vectors form a vector space? Hint: Check for closure.

**Problem 8.10. Geodesics in Inertial & Rotating Frames** For a massive \( m > 0 \) point particle, its trajectory

\[
z^\mu(t) = (t, \bar{z}(t))
\]

(8.1.78)
over an infinitesimal period of time \(dz^\mu = \dot{z}^\mu(t)\,dt \equiv (dz^\mu/dt)\,dt\) is timelike, as discussed above. This means \(\eta_{\mu\nu}dz^\mu dz^\nu > 0\); and, in particular, there must be a frame \(\{dz^\mu\}\) related to the \(\{dz^\alpha\}\) via

\[
dz^\mu = \Lambda^\mu_\nu dz^\nu \quad (8.1.79)
\]

(i.e., it must be possible to find some Lorentz transformation \(\Lambda^\mu_\nu\)) such that

\[
dz^\mu = (d\tau, \vec{0}) \quad (8.1.80)
\]

This means \(\eta_{\mu\nu}dz^\mu dz^\nu > 0\); and, in particular, there must be a frame \(\{dz^\mu\}\) related to the \(\{dz^\alpha\}\) via \(\Lambda_{\mu\nu}\) such that

\[
dz^\mu = (d\tau, \vec{0}) \quad (8.1.80)
\]

This is, of course, simply the instantaneous rest frame of the point particle and \(d\tau\) is its infinitesimal proper time – the time read off a Cesium atom attached to the point particle, say. From equations (8.1.5), (8.1.79) and (8.1.80), what we have managed to argue is – for a timelike worldline – the spacetime counterpart to eq. (7.1.25) reads

\[
\tau (\vec{z}(t_1) \rightarrow \vec{z}(t_2)) \equiv \int_{t_1}^{t_2} d\tau = \int_{t_1}^{t_2} \sqrt{\eta_{\mu\nu}dz^\mu dz^\nu} \quad (8.1.81)
\]

\[
= \int_{t_1}^{t_2} \sqrt{\eta_{\mu\nu}\dot{z}^\mu \dot{z}^\nu} \, dt = \int_{t_1}^{t_2} \sqrt{1 - \vec{v}^2} \, dt; (8.1.82)
\]

where the metric is \(ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu \equiv dt^2 - \vec{d}x \cdot \vec{d}x\).

By demanding that this proper time be extremized (usually maximized), for some fixed end points \(\vec{z}(t_1) = \vec{x}_1\) and \(\vec{z}(t_2) = \vec{x}_2\), show that geodesic motion in Minkowski spacetime corresponds to the Special Relativistic version of Newton’s 2nd law for a free particle:

\[
\frac{d}{dt} \left( \frac{\dot{z}}{\sqrt{1 - \vec{v}^2}} \right) = 0. \quad (8.1.83)
\]

**Proper vs ‘Global’-Inertial Time**

On a related note, for a generic timelike trajectory \(z^\mu(\tau)\) in Minkowski spacetime parametrized by Cartesian coordinates \(x^\mu = (t, \vec{x})\), let us use its proper time \(\tau\) as the 1D coordinate parametrizing the worldline itself, namely

\[
d\tau = \left( \sqrt{\eta_{\mu\nu}u^\mu u^\nu} \, d\lambda \right)_{\lambda=\tau}, \quad u^\mu \equiv \frac{dz^\mu}{d\tau}. (8.1.84)
\]

Recall \(\sqrt{g_{\mu\nu}(dz^\mu/d\lambda)(dz^\nu/d\lambda)}d\lambda = \sqrt{g_{\mu\nu}(dz^\mu/d\lambda')(dz^\nu/d\lambda')}d\lambda'\) is an object that takes the same form no matter the 1D coordinate \(\lambda = \lambda(\lambda')\) used. If we do use \(\lambda = \tau\), the square root in eq. (8.1.84) must be unity. Since \(u^\mu\) is timelike, this tells us

\[
\eta_{\mu\nu}u^\mu u^\nu = (u^0)^2 - \vec{u}^2 = +1. \quad (8.1.85)
\]

Because the time component of \(z^\mu(\tau) = (t(\tau), \vec{z}(\tau))\) is simply the global time \(t\) in the inertial frame \(\{x^\mu\}\), explain why – along a given timelike worldline –

\[
\frac{d\tau}{dt} = \sqrt{1 - \vec{v}^2}, \quad \vec{v} \equiv \frac{d\vec{z}}{dt}. \quad (8.1.86)
\]
Next, let us see that the non-relativistic Newton’s 2nd law of motion in a (3+1)D rotating frame may be recovered by starting from such a spacetime perspective. For concreteness, we will let the inertial frame be \( x^\mu = (t, \vec{x}) \) and the rotating frame be \( x'^\mu = (t, \vec{x}') \). We will assume the rotating frame is revolving counterclockwise at an angular frequency \( \omega \) around the \( x^3 = z' \) axis with respect to the inertial one; namely,

\[
\begin{bmatrix}
x'^1 \\
x'^2 \\
x'^3
\end{bmatrix} = \begin{bmatrix}
\cos(\omega t) & -\sin(\omega t) & 0 \\
\sin(\omega t) & \cos(\omega t) & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
x^1 \\
x^2 \\
x^3
\end{bmatrix}.
\]

(For instance, if an observer is at rest in the rotating frame on its 1-axis; i.e., \( \vec{x}' = (1, 0, 0) \), then \( \vec{x}(t) = (\cos(\omega t), \sin(\omega t), 0)^T \).) Denoting \( x'^i \equiv (\vec{x}' \perp, x'^3) \),

\[
d s^2 = g_{\mu\nu} dx'^\mu dx'^\nu = (1 - \omega^2 x'^2 \perp) dt^2 - 2dt(\vec{\omega} \times \vec{x}') \cdot d\vec{x}' - d\vec{x}' \perp \cdot d\vec{x}' \perp - (dx'^3)^2,
\]

\[
= (1 - (\vec{\omega} \times \vec{x}')^2) dt^2 - 2dt(\vec{\omega} \times \vec{x}') \cdot d\vec{x}' - d\vec{x}'^2,
\]

\[
(8.1.89)
\]

Remember \( |\omega\vec{x}' \perp| \) is the speed \( v \) in the inertial frame. Argue that the non-relativistic limit of the proper time is

\[
\tau (\vec{z}'(t_1) \to \vec{z}'(t_2)) = \int_{t_1}^{t_2} d\tau = \int_{t_1}^{t_2} \sqrt{g_{\mu\nu} \frac{dz'^\mu}{dt} \frac{dz'^\nu}{dt}} dt
\]

\[
= \int_{t_1}^{t_2} (1 - L_{\text{NR}} + O(v^3)) dt;
\]

\[
(8.1.92)
\]

where the \( O(v^2) \) Lagrangian for the rotating frame is

\[
L_{\text{NR}} = \frac{1}{2} \dot{\vec{z}}'^2 + \frac{1}{2} \omega^2 z'^2 \perp + (\vec{\omega} \times \vec{z}') \cdot \ddot{\vec{z}}',
\]

\[
= \frac{1}{2} \dot{\vec{z}}'^2 + \frac{1}{2} (\vec{\omega} \times \vec{z}')^2 + (\vec{\omega} \times \vec{z}') \cdot \ddot{\vec{z}}';
\]

\[
(8.1.94)
\]

By minimizing the proper time, show that the resulting non-relativistic ‘2nd law’ is

\[
m \frac{d^2 \vec{z}'}{dt^2} = F_{\text{Coriolis}} + F_{\text{Centrifugal}};
\]

\[
(8.1.97)
\]

with the Coriolis and Centrifugal forces taking, respectively, the forms

\[
F_{\text{Coriolis}} = -2m\vec{\omega} \times \ddot{\vec{z}}'
\]

\[
(8.1.98)
\]

and

\[
F_{\text{Centrifugal}} = -m\vec{\omega} \times (\vec{\omega} \times \vec{z}').
\]

\[
(8.1.99)
\]

Recall that \( \vec{\omega} \) is given in eq. (8.1.91).
Exponential Form  

Lorentz transformations have continuous parameters that tell us how large/small a rotation and/or boost is being performed. Whenever these parameters may be tuned to set the said Lorentz transformation $\Lambda$ to the identity, we may write it in an exponential form:

$$\Lambda^\mu_\nu = (e^X)^\mu_\nu,$$  \hspace{1cm} (8.1.100)

where the exponential of the matrix $X$ is defined through its power series, $\exp X = \sum_{\ell=0}^{\infty} X^\ell / \ell!$. 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:

$$1/n! X^\mu_{\sigma_1} X^\sigma_1_{\sigma_2} X^\sigma_2_{\sigma_3} \cdots X^\sigma_{n-2}_{\sigma_{n-1}} X^\sigma_{n-1}_\nu.$$  \hspace{1cm} (8.1.101)

If we use the defining relation in eq. (8.1.5), but consider it for small $X$ only,

$$\left(\delta^\mu_\alpha + X^\mu_\alpha + O(X^2)\right) \eta^\mu_\nu \left(\delta^\nu_\beta + X^\nu_\beta + O(X^2)\right)$$  \hspace{1cm} (8.1.102)

$$= \eta^\alpha_\beta + \delta^\mu_\alpha \eta^\mu_\nu X^\nu_\beta + X^\mu_\alpha \eta^\mu_\nu \delta^\nu_\beta + O(X^2)$$  \hspace{1cm} (8.1.103)

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:

$$X^{\alpha\beta} = -X^{\beta\alpha}.$$  \hspace{1cm} (8.1.104)

The general Lorentz transformation continuously connected to the identity must therefore be the exponential of the superposition of the basis of anti-symmetric matrices:

$$\Lambda^{\alpha}_\beta = \left(\exp \left(-\frac{i}{2} \omega^\mu_\nu J^{\mu\nu}\right) \right)^{\alpha}_\beta,$$ \hspace{1cm} (Boosts & Rotations),  \hspace{1cm} (8.1.105)

$$-i (J^{\mu\nu})^{\alpha}_\beta = \eta^{\mu\alpha} \delta^\nu_\beta - \eta^{\nu\alpha} \delta^\mu_\beta = +i (J^{\nu\mu})^{\alpha}_\beta,$$  \hspace{1cm} $\omega^\mu_\nu = -\omega^\nu_\mu \in \mathbb{R}.$  \hspace{1cm} (8.1.106)

Some words on the indices: $(J^{\mu\nu})^{\alpha}_\beta$ is the $\alpha$-th row and $\beta$-th column of the $(\mu, \nu)$-th basis anti-symmetric matrix; with $\mu \neq \nu$. $\omega^\mu_\nu = -\omega^\nu_\mu$ are the parameters controlling the size of the rotations and boosts; they need to be real because $\Lambda^{\alpha}_\beta$ is real.

**Problem 8.11.** From eq. (8.1.106), write down $(J^{\mu\nu})^{\alpha}_\beta$ and explain why these form a complete set of basis matrices for the generators of the Lorentz group.  

**Generators**  
To understand the geometric meaning of eq. (8.1.106), let us figure out the form of $X$ in eq. (8.1.100) that would generate individual Lorentz boosts and rotations in $(2 + 1)$D. The boost along the $1-$axis, according to eq. (8.1.58) is

$$\Lambda^{\mu}_\nu (\xi) = \begin{bmatrix} \cosh(\xi) & \sinh(\xi) & 0 \\ \sinh(\xi) & \cosh(\xi) & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_{3 \times 3} - i \xi \begin{bmatrix} 0 & i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + O(\xi^2).$$  \hspace{1cm} (8.1.107)
The boost along the $2$–axis is
\[
\Lambda^\mu_\nu(\xi) = \begin{bmatrix}
\cosh(\xi) & 0 & \sinh(\xi) \\
0 & 1 & 0 \\
\sinh(\xi) & 0 & \cosh(\xi)
\end{bmatrix} = \mathbb{I}_{3 \times 3} - i\xi \begin{bmatrix}
0 & i & 0 \\
0 & 0 & 0 \\
i & 0 & 0
\end{bmatrix} + \mathcal{O}(\xi^2) .
\] (8.1.108)

Equations (8.1.107) and (8.1.108) tell us the generators of Lorentz boost, assuming $\Lambda^\mu_\nu(\xi)$ take the form $\exp(-i\xi K)$, is then
\[
K^1 \equiv J^{01} \equiv -J^{10} \equiv \begin{bmatrix}
0 & i & 0 \\
i & 0 & 0 \\
0 & 0 & 0
\end{bmatrix} \hat{=} i(\eta^{\mu 0}\delta^1_\nu - \eta^{\mu 1}\delta^0_\nu) ,
\] (8.1.109)
and
\[
K^2 \equiv J^{02} \equiv -J^{20} \equiv \begin{bmatrix}
0 & 0 & i \\
i & 0 & 0 \\
0 & 0 & 0
\end{bmatrix} \hat{=} i(\eta^{\mu 0}\delta^2_\nu - \eta^{\mu 2}\delta^0_\nu) .
\] (8.1.110)

The counter-clockwise rotation on the $(1,2)$ plane, according to eq. (8.1.59), is
\[
\Lambda^\mu_\nu(\theta) = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \theta & -\sin \theta \\
0 & \sin \theta & \cos \theta
\end{bmatrix} = \mathbb{I}_{3 \times 3} - i\theta \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & -i \\
i & 0 & 0
\end{bmatrix} + \mathcal{O}(\theta^2) .
\] (8.1.111)

Assuming this rotation is $\Lambda^\mu_\nu(\theta) = \exp(-i\theta J^{12})$, the generator is.
\[
J^{12} \equiv -J^{21} \equiv i(\eta^{\mu 1}\delta^2_\nu - \eta^{\mu 2}\delta^1_\nu) .
\] (8.1.112)

We may gather, from equations (8.1.109), (8.1.110), and (8.1.112), the generators of boosts and rotations are in fact the ones in eq. (8.1.106).

**Problem 8.12.** Show, by a direct calculation, that $\exp(-i\xi K^1)$ and $\exp(-i\xi K^2)$ do indeed yield the boosts in equations (8.1.107) and (8.1.108) respectively. Show that $\exp(-i\theta J^{12})$ does indeed yield the rotation in eq. (8.1.111). 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 $\exp(-i\xi K^j)$.

**Problem 8.13.** We have only seen that eq. (8.1.106) generates individual boosts and rotations in $(2 + 1)$D. Explain why eq. (8.1.106) does in fact generalize to the generators of boosts and rotations in all dimensions $d \geq 3$. Hint: See previous problem.

**Determinants, Discontinuities**

By taking the determinant of eq. (8.1.5), and utilizing $\det(AB) = \det A \det B$ and $\det A^T = \det A$,
\[
\det \Lambda^\nu_\mu \cdot \det \eta \cdot \det \Lambda = \det \eta \quad (\det \Lambda)^2 = 1 \quad \det \Lambda = \pm 1
\] (8.1.113) (8.1.114) (8.1.115)

Notice the time reversal $\hat{T}$ and parity flips $\{ (\cdot) \hat{P} \}$ matrices each has determinant $-1$. On the other hand, Lorentz boosts and rotations that may be tuned to the identity transformation must have determinant $+1$. This is because the identity itself has det $+1$ and since boosts and rotations depend continuously on their parameters, their determinant cannot jump abruptly from $+1$ and $-1$. 

280
Problem 8.14. The determinant is a tool that can tell us there are certain Lorentz transformations that are disconnected from the identity – examples are
\[
\begin{pmatrix}
\cos \theta & -\sin \theta \\
-\sin \theta & -\cos \theta \\
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
-\cosh \xi & \sinh \xi \\
-\sinh \xi & \cosh \xi \\
\end{pmatrix}.
\] (8.1.116)

You can explain why these are disconnected from \( I \)? □

Group multiplication Because matrices do not commute, it is not true in general that \( e^X e^Y = e^{X+Y} \). Instead, the the Baker-Campbell-Hausdorff formula tells us
\[
e^X e^Y = \exp \left( X + Y + \frac{1}{2} [X, Y] + \frac{1}{12} [X, [X, Y]] - \frac{1}{12} [Y, [X, Y]] + \ldots \right),
\] (8.1.117)
\[
[A, B] \equiv AB - BA;
\] (8.1.118)

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. (8.1.100) holds, we would expect that the product of two Lorentz transformations to yield the same exponential form:
\[
\exp \left( -\frac{i}{2} a_{\mu \nu} J^{\mu \nu} \right) \exp \left( -\frac{i}{2} b_{\alpha \beta} J^{\alpha \beta} \right) = \exp \left( -\frac{i}{2} c_{\delta \gamma} J^{\delta \gamma} \right).
\] (8.1.119)

Comparison with eq. (4.5.36) 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. (8.1.119). More specifically, according to eq. (4.5.36), the first commutator would yield
\[
e^{-\frac{i}{2} a_{\mu \nu} J^{\mu \nu}} e^{-\frac{i}{2} b_{\mu \nu} J^{\mu \nu}} = \exp \left[ -\frac{i}{2} (a_{\mu \nu} + b_{\mu \nu}) J^{\mu \nu} + \frac{1}{2} \left( -\frac{i}{2} \right)^2 a_{\mu \nu} b_{\alpha \beta} [J^{\mu \nu}, J^{\alpha \beta}] \\
+ \frac{1}{12} \left( -\frac{i}{2} \right)^3 a_{\sigma \rho} a_{\mu \nu} b_{\alpha \beta} [J^{\sigma \rho}, [J^{\mu \nu}, J^{\alpha \beta}]] + \ldots \right]
\] (8.1.120)
\[
= \exp \left[ -\frac{i}{2} (a_{\mu \nu} + b_{\mu \nu}) J^{\mu \nu} + \frac{1}{2} \left( -\frac{i}{2} \right)^2 a_{\mu \nu} b_{\alpha \beta} Q^{\mu \nu \alpha \beta \kappa \xi} \kappa J^{\kappa \xi} \\
+ \frac{1}{12} \left( -\frac{i}{2} \right)^3 a_{\sigma \rho} a_{\mu \nu} b_{\alpha \beta} Q^{\mu \nu \alpha \beta \kappa \xi} Q^{\rho \kappa \omega \lambda} J^{\omega \lambda} + \ldots \right],
\] (8.1.121)

for appropriate complex numbers \( \{ Q^{\mu \nu \alpha \beta \kappa \xi} \} \).

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. (8.1.106) would return:
Lie Algebra for $\text{SO}_{D,1}$

\[
[J^{\mu\nu},J^{\rho\sigma}] = i \left( \eta^{\nu\rho} J^{\mu\sigma} - \eta^{\mu\rho} J^{\nu\sigma} + \eta^{\mu\sigma} J^{\nu\rho} - \eta^{\nu\sigma} J^{\mu\rho} \right). \tag{8.1.122}
\]

**Problem 8.15.** Remember that linear operators acting on a Hilbert space themselves form a vector space. Consider a collection of linearly independent linear operators \( \{L_1, L_2, \ldots, L_N\} \). Suppose they are closed under commutation, namely

\[
[L_i, L_j] = \sum_{k=1}^{N} c_{ijk} L_k; \tag{8.1.123}
\]

for any \( i \) and \( j \); and the \( c_{ijk} \) here are (complex) numbers. Prove that these \( N \) operators form a vector space.

---

### 8.2 Lorentz Transformations in 4 Dimensions

**Lie Algebra for $\text{SO}(3,1)$** 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. \((8.1.122)\) more explicitly. Denoting the boost generator as

\[
K^i \equiv J^0^i \tag{8.2.1}
\]

and the rotation generators as

\[
J^i \equiv \frac{1}{2} \epsilon^{imn} J^{mn} \quad \Leftrightarrow \quad \epsilon^{imn} J^i \equiv J^{mn}, \tag{8.2.2}
\]

with \( \epsilon^{123} = \epsilon_{123} \equiv 1 \). The generic Lorentz transformation continuously connected to the identity is

\[
\Lambda(\vec{\xi}, \vec{\theta}) = \exp \left( -i \xi^j K^j - i \theta^j J^j \right). \tag{8.2.3}
\]

At this point, these \( \{\Lambda(\vec{\xi}, \vec{\theta})\} \) are not necessarily the \( 4 \times 4 \) matrices obeying \( \Lambda^T \eta \Lambda = \eta \). Rather, their generators simply need to obey the same commutation relations in eq. \((8.1.122)\).

We may compute from eq. \((8.1.122)\) that

\[
[J^m, J^n] = i \epsilon^{mnl} J^l, \tag{8.2.4}
\]

\[
[K^m, J^n] = i \epsilon^{mnl} K^l, \tag{8.2.5}
\]

\[
[K^m, K^n] = -i \epsilon^{mnl} J^l. \tag{8.2.6}
\]

**Problem 8.16. $\text{SU}(2)_L \times \text{SU}(2)_R$** Let us next define

\[
J^i_+ \equiv \frac{1}{2} (J^i + iK^i), \tag{8.2.7}
\]

\[
J^i_- \equiv \frac{1}{2} (J^i - iK^i). \tag{8.2.8}
\]
Use equations (8.2.4) through (8.2.6) to show that
\[
\begin{align*}
[J^i_+, J^j_+] &= i\epsilon^{ijk} J^k_+ , \\
[J^i_-, J^j_-] &= i\epsilon^{ijk} J^k_- , \\
[J^i_+, J^j_-] &= 0 .
\end{align*}
\]  
(8.2.9)
(8.2.10)
(8.2.11)

Equations (8.2.9) and (8.2.10) tell us the \(J^i_{\pm}\) obey the same algebra as the angular momentum ones in eq. (8.2.4); and eq. (8.2.11) says the two sets \(\{\vec{J}_+, \vec{J}_-\}\) commute.

### 8.2.1 SL\(_{2,C}\) Spinors and Spin-Half

To describe spin–1/2 fermions in Nature – leptons (electrons, muons and taus) and quarks – one has to employ spinors. We will now build spinors in 4D Minkowski spacetime by viewing them as representations of the SL\(_{2,C}\) group.

**Basic Properties of \(\{\sigma^\mu\}\)** We begin by collecting the results in Problems (3.10) and (4.26) as well as the ‘Pauli matrices from their algebra’ discussion in §(4.3.2). A basis set of orthonormal 2 × 2 complex matrices is provided by \(\{\sigma^\mu | \mu = 0, 1, 2, 3\}\), the 2 × 2 identity matrix \(\sigma^0 \equiv I_{2\times2}\) (8.2.12) together with the Hermitian Pauli matrices \(\{\sigma^i\}\). The \(\{\sigma^i| i = 1, 2, 3\}\) may be viewed as arising from the algebra

\[
\sigma^i \sigma^j = \delta^{ij} I_{2\times2} + i\epsilon^{ijk} \sigma^k ,
\]  
(8.2.13)

which immediately implies the respective anti-commutator and commutator results:

\[
\{\sigma^i, \sigma^j\} = 2\delta^{ij} \quad \text{and} \quad [\sigma^i, \sigma^j] = 2i\epsilon^{ijk} \sigma^k .
\]  
(8.2.14)

As a result of eq. (8.2.13), the Pauli matrices have eigenvalues ±1, namely

\[
\sigma^i |\pm; i\rangle = \pm |\pm; i\rangle ;
\]  
(8.2.15)

and thus −1 determinant (i.e., product of eigenvalues) and zero trace (i.e., sum of eigenvalues):

\[
\det \sigma^i = -1 , \quad \text{Tr} \sigma^i = 0 .
\]  
(8.2.16)

An equivalent way of writing eq. (8.2.13) is to employ arbitrary complex vectors \(\vec{a}, \vec{b}\) and \(\vec{c}\). Denoting \(\vec{a} \cdot \vec{\sigma} \equiv a_i \sigma^i\),

\[
(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma}) = \vec{a} \cdot \vec{b} + i(\vec{a} \times \vec{b}) \cdot \vec{\sigma} , \quad (\vec{a} \times \vec{b})^k = \epsilon^{ijk} a_i b_j .
\]  
(8.2.17)

We may multiply by \((\vec{c} \cdot \vec{\sigma})\) from the right on both sides:

\[
(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma})(\vec{c} \cdot \vec{\sigma}) = i(\vec{a} \times \vec{b}) \cdot \vec{c} + \{ (\vec{c} \cdot \vec{b})\vec{a} - (\vec{c} \cdot \vec{a})\vec{b} + (\vec{a} \cdot \vec{b})\vec{c}\} \cdot \vec{\sigma} .
\]  
(8.2.18)

**Problem 8.17.** Verify eq. (8.2.18).
In the representation where \( \sigma^3 \) is diagonal,
\[
\sigma^0 \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma^1 \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (8.2.19)
\]
The inner product of \( \{ \sigma^\mu \} \) is provided by the matrix trace,
\[
\langle \sigma^\mu | \sigma^\nu \rangle \equiv \frac{1}{2} \text{Tr} [\sigma^\mu \sigma^\nu] = \delta^{\mu\nu}. \quad (8.2.20)
\]
Since the \( \{ \sigma^\mu \} \) form a basis, any \( 2 \times 2 \) 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
\[
\bar{\sigma}^\mu \equiv (I_{2\times2}, -\sigma^i) = \sigma^\mu. \quad (8.2.21)
\]
84 We also need the 2D Levi-Civita symbol \( \epsilon \). Since \( \epsilon \) is real and antisymmetric,
\[
\epsilon^\dagger = \epsilon^T = -\epsilon, \quad (8.2.22)
\]
a direct calculation would reveal
\[
\epsilon \cdot \epsilon^\dagger = -\epsilon^2 = I. \quad (8.2.23)
\]
According to eq. (8.2.13), because \( \sigma^i \sigma^i = I \) (for fixed \( i \)) that implies \( \sigma^i \) is its own inverse. We may then invoke eq. (3.2.8) to state
\[
(\sigma^i)^{-1} = \sigma^i = -\epsilon (\sigma^i)^T \epsilon \quad \text{det} \sigma^i = \frac{\epsilon (\sigma^i)^T \epsilon}{\epsilon (\sigma^i)^T \epsilon} = \frac{\epsilon (\sigma^i)^T \epsilon}{\epsilon (\sigma^i)^T \epsilon}. \quad (8.2.24)
\]
Since \( \epsilon \) is real, \( \text{det} \sigma^i = -1 \) (cf. eq. (8.2.16)), and \( \sigma^i \) is Hermitian, we may take the complex conjugate on both sides and deduce
\[
(\sigma^i)^* = \epsilon \cdot \sigma^i \cdot \epsilon = \epsilon^\dagger (\sigma^i)^* \epsilon = \epsilon (\sigma^i) \epsilon^\dagger. \quad (8.2.25)
\]
Since \( \epsilon^2 = -I \), we may multiply both sides by \( \epsilon \) on the left and right,
\[
\epsilon \cdot (\sigma^i)^* \cdot \epsilon = \epsilon^\dagger \cdot (\sigma^i)^* \cdot \epsilon = \epsilon \epsilon (\sigma^i) \epsilon^\dagger = \sigma^i. \quad (8.2.26)
\]
**Problem 8.18.** Using the notation in eq. (8.2.21), explain why
\[
\epsilon \cdot (\bar{\sigma}^\mu)^* \cdot \epsilon^\dagger = \epsilon^\dagger \cdot (\bar{\sigma}^\mu)^* \cdot \epsilon = \sigma^\mu; \quad (8.2.27)
\]
\[
\epsilon \cdot (\sigma^\mu)^* \cdot \epsilon^\dagger = \epsilon^\dagger \cdot (\sigma^\mu)^* \cdot \epsilon = \bar{\sigma}^\mu; \quad (8.2.28)
\]
and therefore
\[
\epsilon \cdot \bar{\sigma}^\mu \cdot \epsilon^\dagger = \epsilon^\dagger \cdot \bar{\sigma}^\mu \cdot \epsilon = (\sigma^\mu)^*; \quad (8.2.29)
\]
\[
\epsilon \cdot \sigma^\mu \cdot \epsilon^\dagger = \epsilon^\dagger \cdot \sigma^\mu \cdot \epsilon = (\bar{\sigma}^\mu)^*. \quad (8.2.30)
\]
**Hint:** Remember the properties of \( \epsilon \) and \( \sigma^0 \).

84 Caution: The over-bar on \( \bar{\sigma} \) is not complex conjugation.
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
\[
\sigma^\mu = -\epsilon \cdot (\bar{\sigma}^\mu)^T \cdot \epsilon = -\epsilon \cdot (\sigma_\mu)^T \cdot \epsilon \tag{8.2.31}
\]
\[
= \epsilon^\dagger \cdot (\sigma^\mu)^T \cdot \epsilon = \epsilon^\dagger \cdot (\sigma_\mu)^T \cdot \epsilon \tag{8.2.32}
\]
That \(\bar{\sigma}^\mu = \sigma_\mu\) is because lowering the spatial components costs a minus sign.

**Problem 8.19.** Explain why
\[
\bar{\sigma}^\mu = -\epsilon \cdot (\sigma^\mu)^T \cdot \epsilon = -\epsilon \cdot (\bar{\sigma}_\mu)^T \cdot \epsilon \tag{8.2.33}
\]
\[
= \epsilon^\dagger \cdot (\sigma^\mu)^T \cdot \epsilon = \epsilon^\dagger \cdot (\bar{\sigma}_\mu)^T \cdot \epsilon \tag{8.2.34}
\]

**Exponential form**

Finally, for any complex \(\{\psi_i\}\), we have from eq. (3.2.23),
\[
\exp \left( -\frac{i}{2} \psi_i \sigma^i \right) = \cos \left( \frac{\bar{\psi}}{2} \right) - i \frac{\bar{\psi} \cdot \bar{\sigma}}{\bar{\psi}} \sin \left( \frac{\bar{\psi}}{2} \right), \tag{8.2.35}
\]
\[
\bar{\psi} \cdot \bar{\sigma} \equiv \psi_j \sigma^j, \quad |\bar{\psi}| \equiv \sqrt{\psi_i \psi_i}. \tag{8.2.36}
\]

One may readily check that its inverse is
\[
\left( \exp \left( -\frac{i}{2} \psi_i \sigma^i \right) \right)^{-1} = \exp \left( +\frac{i}{2} \psi_i \sigma^i \right) = \cos \left( \frac{|\bar{\psi}|}{2} \right) + i \frac{\bar{\psi} \cdot \bar{\sigma}}{|\bar{\psi}|} \sin \left( \frac{|\bar{\psi}|}{2} \right). \tag{8.2.37}
\]

(We will take the \(\sqrt{\cdot}\) in the definition of \(|\bar{\psi}|\) to be the positive square root.) Observe that the relation in eq. (3.2.23) is basis independent; namely, if we found a different representation of the Pauli matrices
\[
\sigma^i = U \sigma^i U^{-1} \quad \Leftrightarrow \quad U^{-1} \sigma^i U = \sigma^i \tag{8.2.38}
\]
then the algebra in eq. (8.2.13) and the exponential result in eq. (8.2.35) would respectively become
\[
U^{-1} \sigma^i U U^{-1} \sigma^j U = U^{-1} (\delta^{ij} + i \epsilon^{ijk} \sigma^k) U, \tag{8.2.39}
\]
\[
\sigma^i \sigma^j = \delta^{ij} + i \epsilon^{ijk} \sigma^k \tag{8.2.40}
\]
and
\[
\exp \left( -\frac{i}{2} \psi_i U^{-1} \sigma^i U \right) = U^{-1} \exp \left( -\frac{i}{2} \psi_i \sigma^i \right) U = U^{-1} \left( \cos \left( \frac{|\bar{\psi}|}{2} \right) - i \frac{\bar{\psi} \sigma^j}{|\bar{\psi}|} \sin \left( \frac{|\bar{\psi}|}{2} \right) \right),
\]
\[
\exp \left( -\frac{i}{2} \psi_i \sigma^i \right) = \cos \left( \frac{|\bar{\psi}|}{2} \right) - i \frac{\bar{\psi} \sigma^j}{|\bar{\psi}|} \sin \left( \frac{|\bar{\psi}|}{2} \right). \tag{8.2.41}
\]
Lorentz Invariant $p^2$ & Helicity Eigenstates

We now move on to the discussion of $\text{SL}_2\mathbb{C}$ proper. 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 $p^2$:

$$\det p_\mu \sigma^\mu = \eta^{\mu\nu} p_\mu p_\nu \equiv p^2. \quad (8.2.42)$$

If we exploited the representation in eq. (8.2.19),

$$p_\mu \sigma^\mu = \begin{bmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{bmatrix}, \quad (8.2.43)$$

from which eq. (8.2.42) may be readily verified. Furthermore, if we now multiply a $2 \times 2$ complex matrix $L$ to the left and $L^\dagger$ to the right of the matrix $p_\mu \sigma^\mu$, namely

$$p_\mu \sigma^\mu \rightarrow L \cdot p_\mu \sigma^\mu \cdot L^\dagger; \quad (8.2.44)$$

– this transformation preserves the Hermitian nature of $p_\mu \sigma^\mu$ for real $p_\mu$ – then its determinant will transform as

$$p^2 = \det[p_\mu \sigma^\mu] \rightarrow \det \left[ L \cdot p_\mu \sigma^\mu \cdot L^\dagger \right] = |\det L|^2 p^2. \quad (8.2.45)$$

If we choose

$$\det L = 1 \quad (8.2.46)$$

– this is the “$S$” $\equiv$ “special” $\equiv$ “unit determinant” in the $\text{SL}_2\mathbb{C}$ – then we see from eq. (8.2.45) that such a transformation would preserve the inner product $p^2 \rightarrow p^2$. Therefore, we expect the group of $\text{SL}_2\mathbb{C}$ matrices $\{L\}$ to implement Lorentz transformations via eq. (8.2.44).

We first note that the Hermitian object $(p_i/|\vec{p}|) \sigma^i$, for real $p_i$ and $|\vec{p}| \equiv \sqrt{\delta^{ij} p_i p_j}$, may be diagonalized as

$$\frac{p_i}{|\vec{p}|} \left( \sigma^i \right)_{\text{AB}} = \xi^+_A \xi^-_B - \xi^-_A \xi^+_B; \quad (8.2.47)$$

$$\frac{p_i}{|\vec{p}|} \sigma^i \xi^\pm = \pm \xi^\pm. \quad (8.2.48)$$

In the representation of the Pauli matrices in eq. (8.2.19), the unit norm helicity eigenstates are, up to overall phases,

$$\xi^+_A \equiv \begin{pmatrix} e^{-i\theta_p} \cos \left( \frac{\theta_p}{2} \right), \sin \left( \frac{\theta_p}{2} \right) \end{pmatrix}^T \quad (8.2.49)$$

$$\xi^-_A \equiv \frac{1}{\sqrt{2}} \sqrt{1 - \frac{p_3}{|\vec{p}|}} \begin{pmatrix} \frac{|\vec{p}| + p_3}{p_1 + ip_2}, 1 \end{pmatrix}^T \quad (8.2.50)$$

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.

This dotted/un-dotted notation will be explained shortly.
and

\[ \xi^-_A = \left( -e^{-i\phi_p} \sin \left( \frac{\theta_p}{2} \right), \cos \left( \frac{\theta_p}{2} \right) \right)^T \]  

(8.2.51)

\[ \xi^-_A = \frac{1}{\sqrt{2}} \sqrt{1 + \frac{p_3}{p}} \begin{pmatrix} -|p| - p_3 \\ p_1 + ip_2 \end{pmatrix}^T. \]  

(8.2.52)

Note that we have switched to spherical coordinates in momentum space, namely

\[ p_i \equiv p(\sin \theta_p \cos \phi_p, \sin \theta_p \sin \phi_p, \cos \theta_p). \]  

(8.2.53)

Also notice, under parity

\[ \phi_p \to \phi_p + \pi \quad \text{and} \quad \theta_p \to \pi - \theta_p, \]  

(8.2.54)

the helicity eigenstates in equations (8.2.49) and (8.2.51) transform into each other:

\[ \xi^+ \to \xi^- \quad \text{and} \quad \xi^- \to \xi^+. \]  

(8.2.55)

These eigenstates \( \xi^\pm \) of the Hermitian \( p_i \sigma^i \), in equations (8.2.49) and (8.2.51), span the 2D complex vector space, so their completeness relation is

\[ I_{AB} = \xi^+_A \xi^+_B + \xi^-_A \xi^-_B; \]  

(8.2.56)

Therefore, \( p_\mu \sigma^\mu = p_0 I + p_i \sigma^i \) itself must be \( p_0 \) times of eq. (8.2.56) plus \( |\vec{p}| \) times of eq. (8.2.47).

\[ p_\mu (\sigma^\mu)_{AB} \equiv p_{AB} = \lambda_+ \xi^+_A \xi^+_B + \lambda_- \xi^-_A \xi^-_B; \quad \lambda_\pm \equiv p_0 \pm |\vec{p}|. \]  

(8.2.57)

**Massive particles**

If we define \( \sqrt{p_\mu \sigma^\mu} \) to be the solution to \( \sqrt{p_\mu \sigma^\mu} \sqrt{p_\mu \sigma^\mu} = p_\mu \sigma^\mu \), then

\[ \sqrt{p \cdot \sigma} = \sqrt{p_\mu \sigma^\mu} = \sqrt{\lambda_+ \xi^+_A \xi^+_B} + \sqrt{\lambda_- \xi^-_A \xi^-_B}. \]  

(8.2.58)

In physical applications where \( p_\mu \) is the momentum of a particle with mass \( m \), \( p_0 \geq |\vec{p}| \) and \( p^2 = m^2 \), the \( \sqrt{\cdot} \) will often be chosen to the positive one – in the following sense. Firstly, the \( \lambda_\pm \) in eq. (8.2.57), could have either positive or negative energy:

\[ p^2 = m^2 \quad \Rightarrow \quad p_0 = \pm E_{\vec{p}} \equiv \pm \sqrt{\vec{p}^2 + m^2}. \]  

(8.2.59)

We shall choose, for positive energy,

\[ \sqrt{\lambda_\pm} = \sqrt{E_{\vec{p}} \pm |\vec{p}|} > 0; \]  

(8.2.60)

and, for negative energy,

\[ \sqrt{\lambda_\mp} = i \sqrt{E_{\vec{p}} \mp |\vec{p}|}, \]  

(8.2.61)

where the \( \sqrt{\cdot} \) is the positive one.
With such a choice, positive energy solutions obey
\[
\sqrt{p \cdot \sigma} \sqrt{p \cdot \bar{\sigma}} \equiv \sqrt{p_\mu \sigma^\mu \sqrt{p_\nu \bar{\sigma}^\nu}} = \sqrt{\lambda_+ \lambda_-} \left( \xi_+^A \bar{\xi}_+^B + \xi_-^A \bar{\xi}_-^B \right) = \sqrt{p^2 I_{2\times2}} = m \cdot I_{2\times2},
\] (8.2.62)

where the orthonormality and completeness of the helicity eigenstates \(\xi^\pm\) were used.

Whereas, negative energy solutions obey
\[
\sqrt{p \cdot \sigma} \sqrt{p \cdot \bar{\sigma}} \equiv \sqrt{p_\mu \sigma^\mu \sqrt{p_\nu \bar{\sigma}^\nu}} = \sqrt{\lambda_+ \lambda_-} \left( \xi_+^A \bar{\xi}_+^B + \xi_-^A \bar{\xi}_-^B \right) = i^2 \sqrt{E_p^2 - p^2 I_{2\times2}} = \imath m \cdot I_{2\times2}.
\] (8.2.63)

Additionally, since \((\sqrt{\lambda_\pm})^* = -\imath \sqrt{E_p \mp |p|}\), we have
\[
\sqrt{p \cdot \sigma} \sqrt{p \cdot \bar{\sigma}} = \left( \sqrt{\lambda_+} \right)^* \sqrt{\lambda_-} \left( \xi_+^A \bar{\xi}_+^B + \xi_-^A \bar{\xi}_-^B \right) = \sqrt{E_p^2 - p^2 I_{2\times2}} = m \cdot I_{2\times2},
\] (8.2.64)

\[
\sqrt{p \cdot \sigma} \sqrt{p \cdot \bar{\sigma}} = \sqrt{E_p^2 - p^2 I_{2\times2}} = m \cdot I_{2\times2}.
\] (8.2.65)

**Massless particles**

For massless particles, \(m = 0\) and \(p_0 = \pm |p|\).

For positive energy \(p_0 = |p|\), the \(\xi^-\) mode becomes a null eigenvector because \(\lambda_- = 0\). Whereas, eq. (8.2.57) now reads
\[
p_{AB} = \xi_+^A \bar{\xi}_+^B, \quad \xi_+ \equiv \sqrt{2|p|} \xi_+^A.
\] (8.2.69)

For negative energy \(p_0 = -|p|\), the \(\xi^+\) mode becomes a null eigenvector because \(\lambda_+ = 0\). Whereas, eq. (8.2.57) now reads
\[
p_{AB} = -\xi_-^A \bar{\xi}_-^B, \quad \xi_- \equiv \sqrt{2|p|} \xi_-^A.
\] (8.2.70)

**Construction of \(L\)**

We have discussed in §(4.5.2), any operator that is continuously connected to the identity can be written in the form \(\exp X\). Since \(L\) has unit determinant (cf. (8.2.46)), let us focus on the case where it is continuously connected to the identity whenever it does depend on a set of complex parameters \(\{q_\mu\}\), say:

\[
L = e^{X(q)}.
\] (8.2.71)

Now, if we use eq. (7.3.92), \(\det e^X = e^{\text{Tr}[X]}\), we find that
\[
\det L = e^{\text{Tr}[X(q)]} = 1.
\] (8.2.72)

This implies
\[
\text{Tr}[X(q)] = 2\pi i n, \quad n = 0, \pm 1, \pm 2, \ldots.
\] (8.2.73)

Recalling that the \(\{\sigma^\mu\}\) form a complete set, we may express
\[
X(q) = q_\mu \sigma^\mu
\] (8.2.74)
and using the trace properties in eq. (8.2.16), we see that \( \text{Tr} \, X(q) = 2q_0 = 2\pi i n \). Since this \( q_0 \sigma^0 = i\pi n I_{2 \times 2} \), which commutes with all the other Pauli matrices, we have at this point

\[
L = e^{i\pi n} e^{q_j \sigma^j} = (-)^n e^{q_j \sigma^j}
\]

(8.2.75)

\[
= (-)^n \left( \cos (i|q|) - i \frac{q_j \sigma^j}{|q|} \sin (i|q|) \right)
\]

(8.2.76)

\[
= (-)^n \left( \cosh (|q|) + \frac{q_j \sigma^j}{|q|} \sinh (|q|) \right)
\]

(8.2.77)

Here, we have replaced \( \theta_j \rightarrow 2iq_j \) in eq. (8.2.35); and note that \( \sqrt{\theta_i \theta_i} = 2i \sqrt{q_i q_i} \) because we have defined the square root to be the positive one. To connect \( L \) to the identity, we need to set the \( q_j \sigma^j \) terms to zero, since the Pauli matrices \( \{\sigma^i\} \) are linearly independent and perpendicular to the identity \( I_{2 \times 2} \). This is accomplished by putting \( \vec{q} = \vec{0} \); which in turn means \( n \) must be even since the cosine/cosh would be unity. To summarize, at this stage:

We have deduced that the most general unit determinant \( 2 \times 2 \) complex matrix that is continuously connected to the identity is, in fact, given by eq. (8.2.35) for arbitrary complex \( \vec{\psi} \), which we shall in turn parametrize as

\[
L = \exp \left( \frac{1}{2} (\xi_j - i\theta_j) \sigma^j \right),
\]

(8.2.78)

where the \( \{\xi_j\} \) and \( \{\theta_j\} \) are real (i.e., \( q_j \equiv (1/2)(\xi_j - i\theta_j) \)). Its inverse is

\[
L^{-1} = \exp \left( -\frac{1}{2} (\xi_j - i\theta_j) \sigma^j \right).
\]

(8.2.79)

By returning to the transformation in eq. (8.2.44), we will now demonstrate the \( \{\xi_j\} \) correspond to Lorentz boosts and \( \{\theta_j\} \) spatial rotations.

**Problem 8.20.** Use eq. (3.2.8) to argue that, for \( L \) belonging to the SL\(_{2,\mathbb{C}} \) group, it obeys

\[
L^{-1} = -\epsilon \cdot L^T \cdot \epsilon.
\]

(8.2.80)

Therefore

\[
(L^{-1})^\dagger = -\epsilon \cdot L^* \cdot \epsilon = e^\dagger \cdot L^* \cdot \epsilon = \epsilon \cdot L^* \cdot \epsilon^\dagger.
\]

(8.2.81)

\[\square\]

**Rotations**

Set \( \vec{\xi} = 0 \) and focus on the case

\[
\theta_j \sigma^j \rightarrow \theta \sigma^k
\]

(8.2.82)

for a fixed \( 1 \leq k \leq 3 \); so that eq. (8.2.78) is now

\[
L = \exp \left( -\frac{i}{2} \theta \sigma^k \right) = \cos(\theta/2) - i\sigma^k \sin(\theta/2);
\]

(8.2.83)
which may be directly inferred from eq. (8.2.35) by setting \( \xi_j = 0 \). Eq. (8.2.44), in turn, now reads

\[
p_\mu \sigma^\mu \rightarrow e^{-(i/2)\theta \sigma^k} p_0 e^{(i/2)\theta \sigma^k} + (\cos(\theta/2) - i\sigma^k \sin(\theta/2)) \ p_i \sigma^i \ (\cos(\theta/2) + i\sigma^k \sin(\theta/2))
\]

\[
= p_0 + p'_i \sigma^i.
\]

(8.2.84)

If \( k = 1 \), we have \( p_i \) rotated on the \((2, 3)\) plane:

\[
p'_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} j \\ i \end{bmatrix} p_j.
\]

(8.2.85)

If \( k = 2 \), we have \( p_i \) rotated on the \((1, 3)\) plane:

\[
p'_i = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} j \\ i \end{bmatrix} p_j.
\]

(8.2.86)

If \( k = 3 \), we have \( p_i \) rotated on the \((1, 2)\) plane:

\[
p'_i = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} j \\ i \end{bmatrix} p_j.
\]

(8.2.87)

**Spin Half** Note that the presence of the generators of rotation, namely \( \sigma^k/2 \) in eq. (8.2.83), with eigenvalues \( \pm 1/2 \), confirms we are dealing with spin\( -1/2 \) systems.

**Problem 8.21.** Verify eq. (8.2.84) for any one of the \( k = 1, 2, 3 \).

**Boosts** Next, we set \( \vec{\theta} = 0 \) and focus on the case

\[
\xi_j \sigma^j \rightarrow \xi \sigma^k,
\]

(8.2.88)

again for a fixed \( k = 1, 2, 3 \). Setting eq. (8.2.77), and remembering \( n \) is even,

\[
L = \exp \left( \frac{1}{2} \xi \sigma^k \right) = \cosh(\xi/2) + \sigma^k \sinh(\xi/2).
\]

(8.2.89)

Eq. (8.2.78) is now

\[
p_\mu \sigma^\mu \rightarrow (\cosh(\xi/2) + \sigma^k \sinh(\xi/2)) \ p_\mu \sigma^\mu \ (\cosh(\xi/2) + \sigma^k \sinh(\xi/2))
\]

\[
= p'_\mu \sigma^\mu.
\]

(8.2.90)

If \( k = 1 \), we have \( p_\mu \) boosted in the \( 1 \)--direction:

\[
p'_\mu = \begin{bmatrix} \cosh \xi & \sinh \xi & 0 & 0 \\ \sinh \xi & \cosh \xi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \nu \\ \mu \end{bmatrix} p_\nu.
\]

(8.2.91)
If \( k = 2 \), we have \( p_\mu \) boosted in the 2–direction:

\[
p'_{\mu} = \begin{bmatrix} \cosh \xi & 0 & \sinh \xi & 0 \\ 0 & 1 & 0 & 0 \\ \sinh \xi & 0 & \cosh \xi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sigma_{\mu \nu} p_{\nu}.
\]

If \( k = 3 \), we have \( p_\mu \) boosted in the 3–direction:

\[
p'_{\mu} = \begin{bmatrix} \cosh \xi & 0 & 0 & \sinh \xi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh \xi & 0 & 0 & \cosh \xi \end{bmatrix} \sigma_{\mu \nu} p_{\nu}.
\]

**Problem 8.22.** Verify eq. (8.2.90) for any one of the \( k = 1, 2, 3 \).  

**Boosts/Rotations & SL\(_{2,\mathbb{C}}\) Spinors**
To summarize, we have discovered that the group of \( 2 \times 2 \) matrices \( \{ L \} \) continuously connected to the identity obeying

\[
\epsilon^{AB} L_A^I L_B^J = \epsilon^{IJ}
\]

implies Lorentz transformations

\[
L_A^I L_B^J \sigma_{\mu \nu}^{\dagger} = \sigma_{\nu \mu}^{\dagger} A_{\nu \mu}.
\]

In terms of matrix multiplication,

\[
L \sigma^\mu \sigma^\nu = \sigma^\sigma \Lambda_{\nu \mu}^\sigma;
\]

where the \( \Lambda_{\nu \mu}^\sigma \) is the \( 4 \times 4 \) Lorentz transformations parametrized by \( \{ \vec{\xi}, \vec{\theta} \} \) satisfying eq. (8.1.5).

Observe that we can take the complex conjugate of equations (8.2.94) and (8.2.96) to deduce that, for the same \( L \) in eq. (8.2.96) – \( L^* \) not only belongs to SL\(_{2,\mathbb{C}}\), it also generates exactly the same Lorentz transformation \( \Lambda_{\nu \mu}^\sigma \) in eq. (8.2.96).

\[
\epsilon^{AB} L_A^I L_B^J = \epsilon^{IJ},
\]

\[
L^* (\sigma^\mu)^* (L^*)^\dagger = (\sigma^\nu)^* \Lambda_{\nu \mu}^\sigma.
\]

For real \( p_\mu \), notice that \( \det p_\mu \sigma^\mu = p^2 = \det p_\mu (\sigma^\mu)^\dagger \). Despite generating the same Lorentz transformation, we shall see below, \( L \) and \( L^* \) are inequivalent representations of SL\(_{2,\mathbb{C}}\) – i.e., there is no change-of-basis \( U \) such that \( U L U^{-1} = L^* \).

Using the dotted and un-dotted index notation in eq. (8.2.57),

\[
L_A^M L_B^N p_{MN} = (\sigma^\nu)_{AB} \Lambda_{\nu \mu} p_{\mu} \equiv p'_{AB} = \lambda_+ \xi_{A}^{\nu \mu} \xi_{B}^{\nu \mu} + \lambda_- \xi_{A}^{\nu \mu} \xi_{B}^{\nu \mu};
\]

where the ‘new’ but un-normalized eigenvectors and eigenvalues are

\[
\xi_{A}^{\nu \mu} (p'_\mu = \Lambda_{\nu \mu} p_{\nu}) = L_A^B \xi_{B}^{\nu \mu} (p_{\mu}) \quad \text{and} \quad \xi_{A}^{\nu \mu} (p'_\mu = \Lambda_{\nu \mu} p_{\nu}) = L_A^B \xi_{B}^{\nu \mu} (p_{\mu})
\]

\( 291 \)
with the old eigenvalues

$$\lambda_\pm \equiv p_0 \pm |\vec{p}|.$$  \hfill (8.2.102)

Any 2-component object that transforms according to eq. (8.2.101), where the $L^A_B$ are SL$_{2,\mathbb{C}}$ matrices, is said to be a spinor. As already alluded to, in the context of $p_\mu \sigma^\mu$, they are also helicity eigenstates of $p_\sigma^i$.

If we normalize the spinors to unity

$$\xi'^{\pm}_A = \xi'^{\pm}_A \left\{ (\xi'^{\pm}_A)^\dagger \xi'^{\pm}_A \right\}^{-\frac{1}{2}}; \hfill (8.2.103)$$

then eq. (8.2.99) now reads

$$L^M_A M^N_B p_{MN} = p'_A B = \lambda'_+ \xi'^+_A \xi'^+_B + \lambda'_- \xi'^-_A \xi'^-_B; \hfill (8.2.104)$$

with the new eigenvalues

$$\lambda'_\pm \equiv p'_0 \pm |\vec{p}'|.$$  \hfill (8.2.105)

$L$ vs. $L^*$ Furthermore, note that the $L$ and its complex conjugate $\overline{L} = L^*$ are not equivalent transformations once Lorentz boosts are included; i.e., once $\xi \neq 0$. To see this, we first recall, for any Taylor-expandable function $f$, $U f(A) U^{-1} = f(U A U^{-1})$ for arbitrary operators $A$ and (invertible) $U$. Remembering the form of $L$ in (8.2.78), let us consider

$$U L^* U^{-1} = \exp \left( \frac{1}{2} U (\xi_j + i \theta_j) (\sigma^j)^* U^{-1} \right). \hfill (8.2.106)$$

Suppose it is possible to find a change-of-basis such that $L^*$ becomes $L$ in eq. (8.2.78), that means we must have for a given $j$,

$$U \cdot \rho_j e^{-i \theta_j} (\sigma^j)^* U^{-1} = \rho_j e^{i \theta_j} \sigma^j; \hfill (8.2.107)$$

$$\rho_j e^{i \theta_j} \equiv \xi_j - i \theta_j; \hfill (8.2.108)$$

$$\rho_j = \sqrt{\xi_j^2 + \theta_j^2}, \quad \tan \theta_j = -\frac{\theta_j}{\xi_j}. \hfill (8.2.109)$$

Taking the determinant on both sides of the first line, for a fixed $j$,

$$\det \left[ e^{-2i \theta_j} (\sigma^j)^* \right] = \det \left[ \sigma^j \right] \hfill (8.2.110)$$

$$e^{-4i \theta_j} \det \left[ \sigma^j \right] = -e^{-4i \theta_j} = \det \left[ \sigma^j \right] = -1. \hfill (8.2.111)$$

(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 $\theta_j = 2\pi n/4 = \pi n/2$ for integer $n$. For even $n$, this corresponds to pure boosts, because

$$\xi_j - i \theta_j = \rho_j e^{\frac{\pi}{4} n} = \pm \rho_j. \hfill (8.2.112)$$
For odd $n$, this corresponds to pure rotations, because
\[ \xi_j - i\theta_j = \rho_j e^{i\frac{\pi}{2}n} = \pm i\rho_j. \quad (8.2.113) \]

However, as we shall show below, there is no transformation $U$ that could bring a pure boost $L^*$ back to $L$:
\[ U \cdot L[\vec{\theta} = \vec{0}] \cdot U^{-1} = U \cdot e^{\frac{i}{2}\xi_j(\sigma^j)^*} \cdot U^{-1} \neq e^{\frac{i}{2}\xi_j\sigma^j} = L[\vec{\theta} = \vec{0}]. \quad (8.2.114) \]

In other words, only the complex conjugate of a pure rotation may be mapped into the same pure rotation. In fact, using $\epsilon(\sigma^i)^*\epsilon^\dagger = -\sigma^i$ in eq. (8.2.26),
\[ \epsilon \cdot L[\vec{\xi} = 0] \cdot \epsilon^\dagger = \epsilon e^{-(i/2)\theta_j(\sigma^j)^*} \epsilon^\dagger = \epsilon^{+(i/2)\theta_j\epsilon(\sigma^j)^*}\epsilon \]
\[ = \epsilon^{-(-i/2)\theta_j\sigma^j} = L[\vec{\xi} = 0]. \quad (8.2.115) \]

But – to reiterate – once $\vec{\xi} \neq 0$, there is no $U$ such that $UL[\vec{\xi},\vec{\theta}] U^{-1} = L[\vec{\xi},\vec{\theta}]$.

Generators:

That $L$ and $L^*$ are generically inequivalent transformations is why the former corresponds to un-dotted indices and the latter to dotted ones – the notation helps distinguishes between them. At this point, let us write
\[ L = \exp \left( -i\xi_j \frac{\sigma^j}{2} - i\theta_j \frac{\sigma^j}{2} \right); \quad (8.2.116) \]

and by referring to generic Lorentz transformation in eq. (8.2.3), we may identify the boost and rotation generators as, respectively,
\[ K^i_R = i\frac{\sigma^i}{2} \quad \text{and} \quad J^i_R = \frac{\sigma^i}{2}. \quad (8.2.118) \]

In this representation, therefore, the Lie algebra in equations (8.2.7) and (8.2.8) read
\[ J^i_+ = \frac{1}{4} (\sigma^i + i^2 \sigma^i) = 0 \quad (8.2.119) \]
\[ J^i_- = \frac{1}{4} (\sigma^i - i^2 \sigma^i) = \frac{\sigma^i}{2}. \quad (8.2.120) \]

The $J^i_+$ generators describe spin $j_+$ zero; whereas the $J^i_-$ ones spin $j_-$ one-half (since the Pauli matrices have eigenvalues $\pm 1$). We therefore label this is as the $\left( j_+, j_- \right) = (0, 1/2)$ representation.

As for the $L^*$, we may express it as
\[ L^* = \exp \left( -i\xi_j (\sigma^j)^* - i\theta_j \frac{-(\sigma^j)^*}{2} \right); \quad (8.2.121) \]

and again referring to eq. (8.2.3),
\[ K^i = i\frac{(\sigma^i)^*}{2} \quad \text{and} \quad J^i = -\frac{(\sigma^i)^*}{2}. \quad (8.2.122) \]
In this case, we may compute the Lie algebra in equations (8.2.7) and (8.2.8):
\[
J^i_+ = \frac{1}{4} \left( - (\sigma^i)^* + i^2 (\sigma^i)^* \right) = - \frac{(\sigma^i)^*}{2}
\]
\[
J^i_- = \frac{1}{4} \left( - (\sigma^i)^* - i^2 (\sigma^i)^* \right) = 0.
\]

This is the \((j_+, j_-) = (1/2, 0)\) representation. We may also recall eq. (8.2.25) and discover that eq. (8.2.122) is equivalent to
\[
K^i = \epsilon^+ \left( - \frac{i}{2} \sigma^i \right) \epsilon \quad \text{and} \quad J^i = \epsilon^+ \left( \frac{1}{2} \sigma^i \right) \epsilon;
\]
which in turn implies we must also obtain an equivalent \((j_+, j_-) = (1/2, 0)\) representation using
\[
K^i_L = - \frac{i}{2} \sigma^i \quad \text{and} \quad J^i_L = \frac{1}{2} \sigma^i.
\]

At this point, eq. (8.2.125) applied to eq. (8.2.121) hands us
\[
L^* = \epsilon^+ \exp \left( - \frac{1}{2} \left( \vec{\xi} + i \vec{\theta} \right) \cdot \vec{\sigma} \right) \epsilon
\]
\[
= \epsilon^+ (L^*)^{-1} \epsilon,
\]
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. (8.2.21) and record here that eq. (8.2.126) may be obtained through
\[
J^{\mu\nu}_L \equiv \frac{i}{4} \sigma^{[\mu} \vec{\sigma}^{\nu]},
\]
\[
J^{0i}_L = \frac{i}{4} \left( \sigma^0 (-) \sigma^i - \sigma^i \sigma^0 \right)
\]
\[
= - \frac{i}{2} \sigma^i = K^i_L;
\]
\[
J^{ab}_L = \frac{i}{4} \left( \sigma^a (-) \sigma^b - \sigma^b (-) \sigma^a \right)
\]
\[
= - \frac{i}{4} [\sigma^a, \sigma^b] = \frac{1}{2} \epsilon^{abc} \sigma^c = \epsilon^{abc} J^c_L.
\]

This is consistent with equations (8.2.1) and (8.2.2). Similarly, eq. (8.2.122) may be obtained through
\[
J^{\mu\nu}_R \equiv \frac{i}{4} \sigma^{[\mu} \vec{\sigma}^{\nu]},
\]
\[
J^{0i}_R = \frac{i}{4} \left( \sigma^0 \sigma^i - (-) \sigma^i \sigma^0 \right)
\]
\[
= + \frac{i}{2} \sigma^i = K^i_R;
\]

294
\[ J^{ab}_R = \frac{i}{4} \left( (-\sigma^a\sigma^b) - (-\sigma^b\sigma^a) \right) \]  \hspace{1cm} (8.2.137)

\[ = -\frac{i}{4} \left[ \sigma^a, \sigma^b \right] = \frac{1}{2} \epsilon^{abc} \sigma^c = \epsilon^{abc} J^c_R. \]  \hspace{1cm} (8.2.138)

**Summary**

For the same set of real boost \( \{\xi_j\} \) and rotation \( \{\theta_j\} \) parameters, the \((j_+, j_-) = (0, 1/2)\) representation of SL\(_{2,\mathbb{C}}\) is provided by the transformation

\[
L \left( \vec{\xi}, \vec{\theta} \right) = \exp \left( -i \vec{\xi} \cdot \vec{K}_R - i \vec{\theta} \cdot \vec{J}_R \right) = e^{\frac{i}{2} (\xi \cdot \theta) \vec{\sigma}}, \]  \hspace{1cm} (8.2.139)

\[
\vec{\xi} \cdot \vec{K}_R \equiv \xi^j K^j_R, \quad \vec{\theta} \cdot \vec{J}_R \equiv \theta^i J^i_R, \]  \hspace{1cm} (8.2.140)

\[
K^i_R = \frac{i}{2} \sigma^i = \frac{i}{4} \bar{\sigma}^{[0} \sigma^{i]}, \quad J^i_R = \frac{1}{2} \sigma^i = \frac{1}{2} \epsilon^{imn} \frac{i}{4} \bar{\sigma}^{[m} \sigma^{n]}; \]  \hspace{1cm} (8.2.141)

whereas the inequivalent \((j_+, j_-) = (1/2, 0)\) representation of SL\(_{2,\mathbb{C}}\) is provided by its complex conjugate

\[
L \left( \vec{\xi}, \vec{\theta} \right) = \epsilon^T \exp \left( -i \vec{\xi} \cdot \vec{K}_L - i \vec{\theta} \cdot \vec{J}_L \right) \epsilon = e^{\frac{i}{2} (\xi \cdot \theta) \vec{\sigma}} \epsilon \]  \hspace{1cm} (8.2.142)

\[
\vec{\xi} \cdot \vec{K}_L \equiv \xi^j K^j_L, \quad \vec{\theta} \cdot \vec{J}_L \equiv \theta^i J^i_L, \]  \hspace{1cm} (8.2.143)

\[
K^i_L = -\frac{i}{2} \sigma^i = \frac{i}{4} \sigma^{[0} \sigma^i], \quad J^i_L = \frac{1}{2} \sigma^i = \frac{1}{2} \epsilon^{imn} \frac{i}{4} \bar{\sigma}^{[m} \sigma^{n]}. \]  \hspace{1cm} (8.2.144)

**Problem 8.23.** Consider the infinitesimal SL\(_{2,\mathbb{C}}\) transformation

\[ L^B_A = \delta^B_A + \omega^B_A. \]  \hspace{1cm} (8.2.146)

Show that, by viewing \( \epsilon^{AB} \) and \( \omega^A_B \) as matrices,

\[ \epsilon \cdot \omega = (\epsilon \cdot \omega)^T. \]  \hspace{1cm} (8.2.147)

From this, deduce

\[ \omega^A_B = \begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix}, \]  \hspace{1cm} (8.2.148)

where \( \alpha, \beta, \) and \( \gamma \) are arbitrary complex parameters. Notice this yields 6 real parameters – in accordance to the 3 directions for boosts plus the 3 directions for rotations we uncovered in eq. \(8.2.78\).  

**Problem 8.24.** Check that the \( J^i \) and \( K^i \) in eq. \(8.2.118\), \(8.2.122\), and \(8.2.126\) satisfy the SO\(_{3,1}\) Lie Algebra \(8.2.4\), \(8.2.5\) and \(8.2.6\).
Transformation of Chiral Spinors

Even though we have defined spinors to be any 2 component object $\xi_A$ that transforms as $\xi \rightarrow L \xi$ for all $L \in \text{SL}_2, C$, our discovery of two inequivalent representations demand that we sharpen this notion further.

Specifically, for the $L(\vec{\xi}, \vec{\theta})$ in eq. (8.2.139), we would say that the spinor transforming as

$$\xi'_A \left( p'_\mu = \Lambda_\mu^\nu p_\nu \right) = L_A^1 \xi_1 \left( p_\nu \right)$$

and

$$\xi'_A \left( p'_\mu = p_\nu \Lambda_\nu^\mu \right) = \left( L^{-1} \right)_A^1 \xi_1 \left( p_\nu \right),$$

is a $(j_+, j_-) = (0, 1/2)$ one; or “right-handed chiral spinor'. Whereas – recalling eq. (8.2.98) – for the same $L(\vec{\xi}, \vec{\theta})$ in eq. (8.2.139), we would say that the spinor transforming as

$$\kappa'_A \left( p'_\mu = \Lambda_\mu^\nu p_\nu \right) = L_A^I \kappa_I \left( p_\nu \right)$$

and

$$\kappa'_A \left( p'_\mu = p_\nu \Lambda_\nu^\mu \right) = \left( L^{-1} \right)_A^I \kappa_I \left( p_\nu \right),$$

is a $(j_+, j_-) = (1/2, 0)$ one; or “left-handed chiral spinor'. We next turn to a different basis to express eq. (8.2.151).

Problem 8.25. Explain why

$$L = \epsilon^\dagger \cdot (L^{-1})^T \cdot \epsilon = \epsilon^\dagger \cdot (L^T)^{-1} \cdot \epsilon,$$

$$L^* = \epsilon^\dagger \cdot (L^{-1})^\dagger \cdot \epsilon = \epsilon^\dagger \cdot (L^\dagger)^{-1} \cdot \epsilon.$$  

(Hint: Recall eq. (3.2.8).) Since $L^*$ is inequivalent to $L$, this shows that $(L^{-1})^\dagger$ is also inequivalent to $L$.

Then show that

$$(L^{-1})^\dagger \bar{\sigma}^\mu L^{-1} = \bar{\sigma}^\nu \Lambda_\nu^\mu;$$

followed by

$$L^\dagger \bar{\sigma}^\mu L = \Lambda_\nu^\mu \bar{\sigma}^\nu.$$  

We see from equations (8.2.98), (8.2.154) and (8.2.155) that, since $(L(\vec{\xi}, \vec{\theta}))^{-1 \dagger}$ is equivalent to $L(\vec{\xi}, \vec{\theta})^*$, and since $L(\vec{\xi}, \vec{\theta})^*$ implements the same Lorentz transformation $\Lambda_\nu^\mu$ as $L(\vec{\xi}, \vec{\theta})$, the $(L(\vec{\xi}, \vec{\theta})^{-1})^\dagger$ also implements on the right-handed spinor the same $\Lambda_\nu^\mu$. Whereas, $L(\vec{\xi}, \vec{\theta})^\dagger$ implements on the right handed spinor the inverse Lorentz transformation $\Lambda_\nu^\mu$.  

For the same $L(\vec{\xi}, \vec{\theta})$ in eq. (8.2.139), we would say that the spinor transforming as

$$\eta'_A \left( p'_\mu = \Lambda_\mu^\nu p_\nu \right) = \left( (L^{-1})^\dagger \right)_A^1 \eta_1 \left( p_\nu \right)$$

and

$$\eta'_A \left( p'_\mu = p_\nu \Lambda_\nu^\mu \right) = \left( L^\dagger \right)_A^1 \eta_1 \left( p_\nu \right),$$

is a $(j_+, j_-) = (1/2, 0)$ one; or “left-handed chiral spinor'; where the $\kappa$ in eq. (8.2.151) and $\eta$ in eq. (8.2.157) are related through the change-of-basis

$$\eta' = \epsilon \cdot \kappa'$$

and

$$\eta = \epsilon \cdot \kappa.$$  

296
Because det $p \cdot \bar{\sigma} = \det p \sigma^\mu = p^2$, we see the spinor $\eta$ obeying equations (8.2.157) and (8.2.158) must yield
\[
\bar{p}_{AB} \equiv p_\mu \bar{\sigma}^\mu_{AB} = \lambda_+ \eta^\dagger_A \eta^\dagger_B + \lambda_- \eta^\dagger_A \eta^\dagger_B; \tag{8.2.160}
\]
\[
\lambda_\pm \equiv p_0 \pm \lvert \vec{p} \rvert. \tag{8.2.161}
\]

**Problem 8.26. SL$_{2,\mathbb{C}}$ Covariant and Invariant Objects**

Suppose the spinor $\xi$ is a right-handed spinor (i.e., subject to equations (8.2.149) and (8.2.150)) and $q_\mu$ is a Lorentz spacetime tensor that obeys
\[
q'_\mu = \Lambda_\mu^\nu q_\nu; \tag{8.2.162}
\]
show that
\[
(\bar{\sigma}^\mu q'_\mu)(L\xi) = (L^\dagger)^{-1}(\bar{\sigma}^\mu q_\mu)\xi, \tag{8.2.163}
\]
\[
\epsilon \cdot (L\xi)^* = (L^\dagger)^{-1}\epsilon \cdot \xi^*. \tag{8.2.164}
\]
Likewise, suppose $u_\mu$ is a Lorentz spacetime tensor that obeys
\[
u'_\mu = u_\nu \Lambda^\nu_\mu; \tag{8.2.165}
\]
show that
\[
(\sigma^\mu u'_\mu)(L^\dagger \eta) = L^{-1}(\sigma^\mu u_\mu)\eta, \tag{8.2.166}
\]
\[
\epsilon \cdot (L^\dagger \eta)^* = L^{-1}\epsilon \cdot \eta^*. \tag{8.2.167}
\]
Roughly speaking, $(\bar{\sigma} \cdot q)\xi$ and $\epsilon \cdot \xi^*$ transform like the left-handed spinor $\eta$; while $(\sigma \cdot u)\eta$ and $\epsilon^\dagger \eta^*$ transform like the right-handed spinor $\xi$.

Next, explain how
\[
\xi^\dagger \bar{\sigma}^\mu \xi \quad \text{and} \quad \eta^\dagger \sigma^\mu \eta \tag{8.2.168}
\]
transform under their relevant SL$_{2,\mathbb{C}}$ transformations. Are
\[
\xi^\dagger \xi \quad \text{and} \quad \eta^\dagger \eta \tag{8.2.169}
\]
scalars under their relevant SL$_{2,\mathbb{C}}$ transformations? Are
\[
\xi^\dagger \eta \quad \text{and} \quad \eta^\dagger \xi \tag{8.2.170}
\]
scalars under their relevant SL$_{2,\mathbb{C}}$ transformations?

**PDEs for Spinors**

To form partial differential equations (PDEs) for spinor fields, the guiding principle is that they transform covariantly under Lorentz (i.e., SL$_{2,\mathbb{C}}$) transformations, so they take the same form in all inertial frames.

**Majorana Equations**

Firstly, recalling the momentum $p_\mu$ dependence in the transformation rule of $\xi$ in eq. (8.2.149), we see that $q_\mu$ in eq. (8.2.163) may be replaced with it: $q_\mu = p_\mu$. 297
If $\xi$ is now viewed as the Fourier coefficient of its position spacetime counterpart, we may now recognize
\[
(\bar{\sigma}^{\mu} p_\mu) \xi(p) e^{-ip\cdot x} = i(\bar{\sigma}^{\mu} \partial_\mu)(\xi(p) e^{-ip\cdot x}).
\] (8.2.171)

Because the terms in equations (8.2.163) and (8.2.164) transform the same way, under $\xi' = L\xi$, we may immediate write down the $(0, 1/2)$ Majorana equation in position space:
\[
i\bar{\sigma}^{\mu} \partial_\mu \xi(x) = m \epsilon \cdot \xi(x)^*.
\] (8.2.172)

The $m$ here is of dimensions mass, because the left hand side involves a derivative, i.e., 1/length. A similar discussion will let us write down the $(1/2, 0)$ counterpart from the terms in equations (8.2.166) and (8.2.167):
\[
i\sigma^{\mu} \partial_\mu \eta(x) = m \epsilon \cdot \eta(x)^*.
\] (8.2.173)

\textit{Weyl Equations} Setting $m = 0$ in equations (8.2.172) and (8.2.173) hands us the Weyl equations
\[
i\bar{\sigma}^{\mu} \partial_\mu \xi = 0 \quad \text{and} \quad i\sigma^{\mu} \partial_\mu \eta = 0.
\] (8.2.174)

\textit{Dirac Equations} Under the transformation $\xi' = L\xi$, eq. (8.2.163) transforms as $(\bar{\sigma}^{\mu} q_\mu)\xi' = (L^\dagger)^{-1}(\bar{\sigma}^{\mu} q_\mu)\xi$, which thus transforms in the same manner as $\eta' = (L^\dagger)^{-1}\eta$. (Recall too, eq. (8.2.154) tells us $(L^\dagger)^{-1}$ is equivalent to $L^*$..) In a similar vein, under the transformation $\eta' = (L^\dagger)^{-1}\eta$, eq. (8.2.166) transforms as $(\sigma^{\mu} u_\mu)\eta' = L(\sigma^{\mu} u_\mu)\eta$, which thus transforms in the same manner as $\xi' = L\xi$. Since $L$ and $L^*$ implement the same Lorentz transformation, we may write down the following pair of Lorentz covariant PDEs:
\[
i\bar{\sigma}^{\mu} \partial_\mu \xi = m \eta \quad \text{and} \quad i\sigma^{\mu} \partial_\mu \eta = m \xi.
\] (8.2.175)

The pair of PDEs in eq. (8.2.175) is known as the Dirac equation(s).

\textbf{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. (4.3.24). Now, according to eq. (8.2.20), $c_\mu \sigma^{\mu} \leftrightarrows c_\mu = (1/2) \text{Tr}[(c_\nu \sigma^{\nu})\sigma^{\mu}]$ for any complex coefficients $\{c_\nu\}$. (We will not distinguish between dotted and un-dotted indices for now.) Consider
\[
c_\mu(\sigma^{\mu})_{AB} = \sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^{\mu})_{AB} \text{Tr}[(\sigma^{\mu})^T c_\nu \sigma^{\nu}]
\] (8.2.176)
\[
= \sum_{1 \leq C,D \leq 2} \left( \sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^{\mu})_{AB} (\sigma^{\mu})^T_{CD} \right) c_\nu(\sigma^{\nu})_{CD}
\] (8.2.177)
\[
= \sum_{1 \leq C,D \leq 2} \left( \sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^{\mu})_{AB} (\sigma^{\mu})_{CD} \right) c_\nu(\sigma^{\nu})_{CD}.
\] (8.2.178)

That the same mass $m$ appears in both equations, follows from the demand that the associated Lagrangian density $L_{\text{Dirac mass}} = -m(\eta^\dagger \xi + \xi^\dagger \eta)$ (and hence its resulting contribution to the Hamiltonian) be Hermitian.
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 \(c_\nu\) was arbitrary, it must act on each and every \(\sigma^\nu\) to return \(\sigma^\nu\), 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.

\[
\sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^\mu)_{AB} \overline{(\sigma^\mu)}_{CD} = \sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^\mu)_{AB} (\sigma^\mu)^T_{CD} = \delta^C_A \delta^D_B \tag{8.2.179}
\]

In the second equality we have used the Hermitian nature of \(\sigma^\mu\) to deduce \((\sigma^\mu)^{\dagger}_{AB} = (\sigma^\mu)^T_{AB} = (\sigma^\mu)^*_{AB}\). If we further employ \((\sigma^\mu)^* = \epsilon \cdot \sigma^\mu \cdot \epsilon^\dagger = \epsilon \cdot \sigma^\mu \cdot \epsilon^T\) in eq. (8.2.30) within the leftmost expression,

\[
\sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^\mu)_{AB} \overline{(\sigma^\mu)}_{CD} = \sum_{0 \leq \mu \leq 3} \frac{1}{2} (\sigma^\mu)_{AB} \epsilon^{CM} \epsilon^{DN} (\overline{\sigma^\mu})_{MN}. \tag{8.2.180}
\]

If we now restore the dotted notation on the right index, so that

\[
\epsilon^{CM} \epsilon^{DN} (\overline{\sigma^\mu})_{MN} \equiv (\overline{\sigma^\mu})^{MN}, \tag{8.2.181}
\]

then eq. (8.2.179), with Einstein summation in force, becomes

\[
\frac{1}{2} \sigma^\mu_{AB} \overline{\sigma^\mu}^{CD} = \delta^C_A \delta^D_B \tag{8.2.182}
\]

with

\[
\sigma^\mu_{CD} \equiv (\sigma^\mu)_{EF} \epsilon^{EC} \epsilon^{FD}. \tag{8.2.183}
\]

Next, consider

\[
(\sigma^\mu)^{MN}(\sigma^\nu)^{MN} = (\sigma^\mu)^{MN} \epsilon^{MA} \epsilon^{NB} (\sigma^\nu)_{AB} = (\sigma^\mu)^T_{NM} \epsilon^{MA} (\sigma^\nu)_{AB} (\epsilon^T)^{BN} \tag{8.2.184}
\]

\[
= \text{Tr} \left[ (\sigma^\mu)^T \cdot \epsilon \cdot \sigma^\nu \cdot \epsilon^T \right] = \text{Tr} \left[ \epsilon^T \cdot (\sigma^\mu)^T \cdot \epsilon \cdot \sigma^\nu \right] = \text{Tr} \left[ (\sigma^\mu)^T \sigma^\nu \right] = \text{Tr} \left[ (\sigma^\mu)\sigma^\nu \right]. \tag{8.2.185}
\]

Invoking the orthonormality of the \(\{\sigma^\mu\}\) in eq. (8.2.20),

\[
\frac{1}{2} (\sigma^\mu)^{MN}(\sigma^\nu)^{MN} = \delta^\mu_{\nu}. \tag{8.2.187}
\]

Equation (8.2.187) tell us we may view the spacetime Lorentz index \(\mu\) and the pair of spinor indices \(A \dot{B}\) as different basis for describing tensors. For example, we may now switch between the momentum \(p_\mu\) and \(p_{AB}\) via:

\[
p_\mu \sigma^\mu_{AB} = p_{AB} \quad \Leftrightarrow \quad p_\mu = \frac{1}{2} \sigma^\mu_{AB} p_{AB}, \tag{8.2.188}
\]

where the latter relation is a direct consequence of eq. (8.2.187),

\[
p_\mu = \frac{1}{2} \sigma^\mu_{AB} \sigma^\nu_{AB} p_\nu = \delta^\nu_\mu p_\nu = p_\mu, \tag{8.2.189}
\]

299
Levi-Civita as Spinor-Metric  

The Levi-Civita symbol may be viewed as the ‘metric’ for the both the dotted and un-dotted spinor. We will move the un-dotted indices as follows:

\[ \xi_A = \epsilon_{AB} \xi^B \quad \text{and} \quad \xi^A = \xi_B \epsilon^{BA}. \]  

(8.2.190)

Numerically, \( \xi_1 = \epsilon_{12} \xi^2 = \xi^2 \) while \( \xi_2 = \epsilon_{21} \xi^1 = -\xi^1 \). Notice we contract with the right index of \( \epsilon \) when lowering the spinor index; but with the left index when raising. This is because the Levi-Civita symbol is anti-symmetric, and this distinction is necessary for consistency:

\[ \xi_A = \epsilon_{AB} \xi^B = \epsilon_{AB} \epsilon^{CB} \xi_C = -\epsilon_{AB} \epsilon^{BC} \xi_C = \delta^C_A \xi_C = \xi_A. \]  

(8.2.191)

Similarly,

\[ \xi_A = \epsilon_{AB} \xi^B \quad \text{and} \quad \xi^A = \xi_B \epsilon^{BA}. \]  

(8.2.193)

We may even move the indices of \( \epsilon \); for instance, keeping in mind \( \epsilon^2 = -1 \),

\[ \epsilon_{AB} = \epsilon_{AM} \epsilon_{BN} \epsilon^{MN} \]  

(8.2.194)

\[ = -\delta^N_A \epsilon_{BN} = -\epsilon_{BA}. \]  

(8.2.195)

The primary reason why we may move these indices with \( \epsilon \) and view the latter as a metric, is because the ‘scalar product’

\[ \xi \cdot \eta \equiv \epsilon^{IJ} \xi_I \eta_J = \xi^J \eta_J = -\epsilon^{JI} \eta_I \xi_I = -\eta \cdot \xi \]  

(8.2.196)

is invariant under Lorentz SL\(_{2,C} \) transformations. For, under the replacement

\[ \xi_I \rightarrow L_1^A \xi_A \quad \text{and} \quad \eta_I \rightarrow L_1^A \eta_A, \]  

(8.2.197)

the ‘scalar product’ transforms as

\[ \xi \cdot \eta \rightarrow \epsilon^{IJ} L_1^A L_1^B \xi_A \eta_B \]  

(8.2.198)

\[ = (\det L) \epsilon^{AB} \xi_A \eta_B = \xi \cdot \eta. \]  

(8.2.199)

The second equality is due to the defining condition of the SL\(_{2,C} \) group, \( \det L = 1 \), as expressed in eq. (8.2.94). Likewise,

\[ \epsilon^{\dot{A} \dot{B}} \xi_{\dot{A}} \eta_{\dot{B}} \rightarrow \epsilon^{\dot{I} \dot{J}} L_1^{\dot{A}} L_1^{\dot{B}} \xi_{\dot{A}} \eta_{\dot{B}} = \epsilon^{\dot{A} \dot{B}} \xi_{\dot{A}} \eta_{\dot{B}}. \]  

(8.2.200)

Note that the scalar product between a dotted and un-dotted spinor \( \epsilon^{\dot{A} \dot{B}} \xi_{\dot{A}} \eta_{\dot{B}} \) would not, in general, be an invariant because its transformation will involve both \( L \) and \( L^* \).

Since eq. (8.2.198) informs us that \( \xi^I \eta_I \) is a SL\(_{2,C} \) scalar, it must be that the upper index spinor transforms oppositely from its lower index counterpart. Let’s see this explicitly.

\[ \xi'^A = \xi'^B \epsilon^{BA} = -\epsilon^{AB} L_B^C \xi_C \]  

(8.2.201)

\[ = -\epsilon^{AB} L_B^C \epsilon_{CD} \xi^D. \]  

(8.2.202)
Recalling eq. (3.2.8) and eq. (8.2.45),

\[ \xi^A = ((L^{-1})^T)^A_D \xi^D = \xi^D(L^{-1})_D^A. \]  

(8.2.203)

**Parity in 2D SL\(_{2,\mathbb{C}}\)**

We will now demonstrate that the parity operator does not exist within the SL\(_{2,\mathbb{C}}\) representations we have been studying. This has important consequences for constructing the parity covariant Dirac equation, for instance. Now, by parity, we mean the transformation

\[ P \in SL_{2,\mathbb{C}} \]

that would – for arbitrary \( p_\mu \) – flip the sign of its spatial components, namely

\[ P(p_0 \mathbb{I} + \vec{p} \cdot \vec{\sigma}) P^\dagger = p_0 \mathbb{I} - \vec{p} \cdot \vec{\sigma} = p_\mu \bar{\sigma}^\mu. \]

(8.2.204)

In fact, since this is for arbitrary \( p_\mu \), we may put \( p_0 = 0 \), \( p_i = \delta^j_i \) (for fixed \( j \)), and see that

\[ P \sigma^j P^\dagger = -\sigma^j, \quad j \in \{1, 2, 3\}. \]

(8.2.205)

We may also set \( p_i = 0 \) in eq. (8.2.204) and observe that \( P \) needs to be unitary if it is to be a representation of SL\(_{2,\mathbb{C}}\):

\[ P (p_0 \mathbb{I}) P^\dagger = p_0 \mathbb{I} \iff P \cdot P^\dagger = \mathbb{I}. \]

(8.2.207)

Remember eq. (8.2.35) is in fact the most general form of an SL\(_{2,\mathbb{C}}\) transformation. We may therefore take its dagger and set it equal to its inverse in (8.2.37).

\[ \cos \left( \frac{\|\vec{\psi}\|^*}{2} \right) + i \frac{\vec{\psi}^* \cdot \vec{\sigma}}{\|\vec{\psi}\|^*} \sin \left( \frac{\|\vec{\psi}\|^*}{2} \right) = \cos \left( \frac{\|\vec{\psi}\|}{2} \right) + i \frac{\vec{\psi} \cdot \vec{\sigma}}{\|\vec{\psi}\|} \sin \left( \frac{\|\vec{\psi}\|}{2} \right). \]

(8.2.208)

Since the Pauli matrices are linearly independent and orthogonal to the identity, we must have \( \psi \) real; i.e., the most general unitary operator that is also an SL\(_{2,\mathbb{C}}\) transformation is thus nothing but the rotation operator

\[ P = \exp \left( -\frac{i}{2} \vec{\theta} \cdot \vec{\sigma} \right), \quad \vec{\theta} \in \mathbb{R}. \]

(8.2.209)

Returning to eq. (8.2.204), and recalling it must hold for arbitrary \( \vec{p} \), we may now set \( p_j = \theta_j \):

\[ P \left( p_0 + \vec{\theta} \cdot \vec{\sigma} \right) P^\dagger = p_0 + \exp \left( -\frac{i}{2} \vec{\theta} \cdot \vec{\sigma} \right) (\vec{\theta} \cdot \vec{\sigma}) \exp \left( +\frac{i}{2} \vec{\theta} \cdot \vec{\sigma} \right) \]

\[ = p_0 + \exp \left( -\frac{i}{2} \vec{\theta} \cdot \vec{\sigma} \right) \exp \left( +\frac{i}{2} \vec{\theta} \cdot \vec{\sigma} \right) (\vec{\theta} \cdot \vec{\sigma}) = p_0 + \vec{\theta} \cdot \vec{\sigma}. \]

(8.2.210)

(8.2.211)

\[ 88 \] It is, of course, possible to find the parity operator that works for a given \( \vec{p} \); it is given by

\[ P = [\xi^+ | 1 \right] \left[ \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right] \left[ \begin{array}{cc} 1 & \xi^- \\ \xi^+ & \xi^- \end{array} \right]^\dagger, \quad P \xi^\pm = \xi^\mp, \quad P(\vec{p} \cdot \vec{\sigma}) P^\dagger = P(\vec{p} \cdot \vec{\sigma}) P = -\vec{p} \cdot \vec{\sigma}. \]

(8.2.206)

Here, the \( \xi^\pm \) are the helicity eigenstates in equations (8.2.49) and (8.2.51). But since this is a \( \vec{p} \) specific operator, that would not be a parity operation on the whole space \( \vec{x} \to -\vec{x} \); for that to be true we need it to be applicable for all \( \vec{p} \) and, from eq. (8.2.205), therefore \( \vec{p} \)-independent.

\[ 89 \] For a function like sine or cosine which may be Taylor expanded on the complex plane, \( f(z)^* = f(z^*) \).

301
In other words, it is impossible to construct a parity operator, for otherwise we would find the end result to be \( p_0 - \vec{\theta} \cdot \vec{\sigma} \).

Before we moving on, let us prove the statement alluded to earlier, that the complex conjugate of pure boost, i.e., \( \exp((1/2)\vec{\xi} \cdot \vec{\sigma}^*) \) (with real \( \vec{\xi} \)), cannot be equivalent to the pure boost itself. (The proof is very similar to the one we just delineated for the non-existence of the parity operator.) Suppose a \( U \) existed, such that

\[
U \exp \left( \frac{1}{2} \vec{\xi} \cdot \vec{\sigma} \right) U^{-1} = \exp \left( \frac{1}{2} \vec{\xi} \cdot \vec{\sigma} \right).
\]  

(8.2.212)

According to equations (8.2.23) and (8.2.25), we may convert this into

\[
(U \cdot \epsilon) \exp \left( -\frac{1}{2} \vec{\xi} \cdot \vec{\sigma} \right) (U \cdot \epsilon)^{-1} = \exp \left( -\frac{1}{2} (U \cdot \epsilon)(\vec{\xi} \cdot \vec{\sigma})(U \cdot \epsilon)^{-1} \right) = \exp \left( \frac{1}{2} \vec{\xi} \cdot \vec{\sigma} \right).
\]  

(8.2.213)

Since \( \vec{\xi} \) is arbitrary we must have

\[
U' \sigma^j U'^{-1} = -\sigma^j, \quad U' \equiv U \cdot \epsilon.
\]  

(8.2.214)

Now, if \( \det U' \neq 1 \), we may define \( U'' \equiv U'/(\det U')^{1/2} \Rightarrow \det U'' = 1 \), i.e., \( U'' \in \text{SL}_2, \mathbb{C} \), so that

\[
U' \sigma^j U'^{-1} = (\det U')^{1/2} U'' \sigma^j U''^{-1}(\det U')^{-1/2} = U'' \sigma^j U''^{-1} = -\sigma^j.
\]  

(8.2.215)

(8.2.216)

Since \( U'' \) is a \( \text{SL}_2, \mathbb{C} \) transformation, we may use its form in eq. (8.2.35) and its inverse in eq. (8.2.37). If we first contract eq. (8.2.216) with the same \( \vec{\psi} \) in eq. (8.2.35), we arrive at the following contradiction:

\[
\exp \left( -\frac{i}{2} \vec{\psi} \cdot \vec{\sigma} \right) \left( \vec{\psi} \cdot \vec{\sigma} \right) \exp \left( +\frac{i}{2} \vec{\psi} \cdot \vec{\sigma} \right) = \exp \left( -\frac{i}{2} \vec{\psi} \cdot \vec{\sigma} \right) \exp \left( +\frac{i}{2} \vec{\psi} \cdot \vec{\sigma} \right) \left( \vec{\psi} \cdot \vec{\sigma} \right) = \vec{\psi} \cdot \vec{\sigma} = -\vec{\psi} \cdot \vec{\sigma}.
\]  

(8.2.217)

(8.2.218)

To discuss parity for spinors, we therefore need to go beyond these 2 component ones.

**Parity & Clifford Algebra**

To be continued . . .

### 8.3 Curved Metrics, Orthonormal Frames & Volume; Timelike, Space-like vs. Null Vectors; Gravitational Time Dilation

**Curved Spacetime, Spacetime Volume**

The generalization of the ‘distance-squared’ between \( x^\mu \) to \( x^\mu + dx^\mu \), from the Minkowski to the curved case, is the following “line element”:

\[
ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu,
\]  

(8.3.1)

where \( x \) is simply shorthand for the spacetime coordinates \( \{x^\mu\} \), which we emphasize may no longer be Cartesian. We also 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. The infinitesimal spacetime volume continues to take the form
\[
d(\text{vol.}) = d^d x \sqrt{|g(x)|},
\]
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 \(\delta_{ij}\) 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 \(x^\mu\), we introduce an orthonormal basis \(\{\hat{\epsilon}^\alpha_\mu\}\) through the ‘diagonalization’ process:
\[
g_{\mu\nu}(x) = \eta_{\alpha\beta} \hat{\epsilon}^\alpha_\mu(x) \hat{\epsilon}^\beta_\nu(x).
\]
By defining \(\hat{\epsilon} = \hat{\epsilon}^\alpha_\mu \, dx^\mu\), the analog to achieving a Cartesian-like expression for the spacetime metric is
\[
ds^2 = \left(\hat{\epsilon}^0\right)^2 - \sum_{i=1}^{D} \left(\hat{\epsilon}^i\right)^2 = \eta_{\mu\nu} \hat{\epsilon}^\mu \hat{\epsilon}^\nu.\]

This means under a local Lorentz transformation – i.e., for all \(\Lambda^\mu_\alpha(x) \Lambda^\nu_\beta(x) \eta_{\mu\nu} = \eta_{\alpha\beta}\)
\[
\hat{\epsilon}^\mu(x) = \Lambda^\mu_\alpha(x) \hat{\epsilon}^\alpha(x)
\]
– the metric remains the same:
\[
ds^2 = \eta_{\mu\nu} \hat{\epsilon}^\mu \hat{\epsilon}^\nu = \eta_{\mu\nu} \hat{\epsilon}^\mu \hat{\epsilon}^\nu.\]

By viewing \(\hat{\epsilon}\) as the matrix with the \(\alpha\)th row and \(\mu\)th column given by \(\hat{\epsilon}^\alpha_\mu\), the determinant of the metric \(g_{\mu\nu}\) can be written as
\[
\det g_{\mu\nu}(x) = (\det \hat{\epsilon})^2 \det \eta_{\mu\nu}.
\]
The infinitesimal spacetime volume in eq. (8.3.2) now can be expressed as
\[
d^d x \sqrt{|g(x)|} = d^d x \det \hat{\epsilon} = \hat{\epsilon}^0 \wedge \hat{\epsilon}^1 \wedge \cdots \wedge \hat{\epsilon}^{d-1}.
\]
The second equality follows because
\[
\hat{\epsilon}^0 \wedge \cdots \wedge \hat{\epsilon}^{d-1} = \hat{\epsilon}_\mu^0 \, dx^\mu_1 \wedge \cdots \wedge \hat{\epsilon}_\mu^d \, dx^\mu_d.
\]
\[ = \epsilon_{\mu_1...\mu_d} \hat{\epsilon}_{\mu_1} \cdots \hat{\epsilon}_{\mu_{d-1}} \, dx^0 \wedge \cdots \wedge dx^{d-1} = (\det \hat{\epsilon}) dx^d. \]  

(8.3.11)

Of course, that \( g_{\mu\nu} \) may be 'diagonalized' follows from the fact that \( g_{\mu\nu} \) is a real symmetric matrix:

\[ g_{\mu\nu} = \sum_{\alpha,\beta} O^\alpha_\mu \lambda_\alpha \eta_{\alpha\beta} O^\beta_\nu = \sum_{\alpha,\beta} \hat{\epsilon}_\mu^\alpha \eta_{\alpha\beta} \hat{\epsilon}_\nu^\beta, \]  

(8.3.12)

where all \( \{ \lambda_\alpha \} \) are positive by assumption, so we may take their positive root:

\[ \hat{\epsilon}_\mu^\alpha = \sqrt{\lambda_\alpha} O^\alpha_\mu, \quad \{ \lambda_\alpha > 0 \}, \quad \text{(No sum over } \alpha). \]  

(8.3.13)

That \( \hat{\epsilon}_\mu^0 \) acts as 'standard clock' and \( \{ \hat{\epsilon}_\mu^i | i = 1, 2, \ldots, D \} \) act as 'standard rulers' is because they are of unit length:

\[ g^{\mu\nu} \hat{\epsilon}_\mu^\alpha \hat{\epsilon}_\nu^\beta = \eta^{\alpha\beta}. \]  

(8.3.14)

The \( \hat{\cdot} \) on the index indicates it is to be moved with the flat metric, namely

\[ \hat{\epsilon}_\mu^\alpha = \eta^{\alpha\beta} \hat{\epsilon}_\mu^\beta \quad \text{and} \quad \hat{\epsilon}_\mu^\alpha = \eta_{\alpha\beta} \hat{\epsilon}_\mu^\beta; \]  

(8.3.15)

while the spacetime index is to be moved with the spacetime metric

\[ \hat{\epsilon}_\mu^\mu = g^{\mu\nu} \hat{\epsilon}_\nu^\nu \quad \text{and} \quad \hat{\epsilon}_\mu^\nu = g_{\mu\nu} \hat{\epsilon}_\nu^\nu. \]  

(8.3.16)

In other words, we view \( \hat{\epsilon}_\mu^\alpha \) as the \( \mu \)th spacetime component of the \( \alpha \)th vector field in the basis set \( \{ \hat{\epsilon}_\mu^\alpha | \alpha = 0, 1, 2, \ldots, D \equiv d - 1 \} \). We may elaborate on the interpretation that \( \{ \hat{\epsilon}_\mu^\alpha \} \) act as 'standard clock/rulers' as follows. For a test (scalar) function \( f(x) \) defined throughout spacetime, the rate of change of \( f \) along \( \hat{\epsilon}_0^0 \) is

\[ \langle df | \hat{\epsilon}_0^0 \rangle = \hat{\epsilon}_0^\mu \partial_\mu f \equiv \frac{df}{dy^0}; \]  

(8.3.17)

whereas that along \( \hat{\epsilon}_i^i \) is

\[ \langle df | \hat{\epsilon}_i^i \rangle = \hat{\epsilon}_i^\mu \partial_\mu f \equiv \frac{df}{dy^i}; \]  

(8.3.18)

where \( y^0 \) and \( \{ y^i \} \) are to be viewed as 'time' and 'spatial' parameters along the integral curves of \( \{ \hat{\epsilon}_\mu^\alpha \} \). That these are Cartesian-like can now be expressed as

\[ \left\langle \frac{d}{dy^\mu} \left| \frac{d}{dy^\nu} \right. \right\rangle = \hat{\epsilon}_\mu^\alpha \hat{\epsilon}_\nu^\beta \langle \partial_\alpha | \partial_\beta \rangle = \hat{\epsilon}_\mu^\alpha \hat{\epsilon}_\nu^\beta g_{\alpha\beta} = \eta_{\mu\nu}. \]  

(8.3.19)

It is worth reiterating that the first equalities of eq. \[8.3.12\] 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. \[90\]

---

\[90\] 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.
**Commutators & Coordinates**  
Note that the \( \frac{d}{dy^\mu} \) in eq. (8.3.19) do not, generically, commute. For instance, acting on a scalar function,

\[
\left[ \frac{d}{dy^\mu}, \frac{d}{dy^\nu} \right] f(x) = \left( \frac{d}{dy^\mu} \frac{d}{dy^\nu} - \frac{d}{dy^\nu} \frac{d}{dy^\mu} \right) f(x) = \left( \varepsilon_\mu^\alpha \partial_\alpha \varepsilon_\nu^\beta - \varepsilon_\nu^\alpha \partial_\alpha \varepsilon_\mu^\beta \right) \partial_\beta f(x) \neq 0. \tag{8.3.20}
\]

More generally, for any two vector fields \( V^\mu \) and \( W^\mu \), their commutator is

\[
[V, W]^\mu = V^\sigma \nabla_\sigma W^\mu - W^\sigma \nabla_\sigma V^\mu \tag{8.3.22}
\]

\[
= V^\sigma \partial_\sigma W^\mu - W^\sigma \partial_\sigma V^\mu. \tag{8.3.23}
\]

(Can you explain why the covariant derivatives can be replaced with partial ones?) A theorem in differential geometry\(^{91}\) tells us:

A set of \( 1 < N \leq d \) vector fields \( \{ \frac{d}{dy^\mu} \} \) 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 \( \frac{d}{dy^\mu}, \frac{d}{dy^\nu} = 0 \) in eq. (8.3.19), 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 \( \frac{d}{dy^0}, \frac{d}{dy^1} \neq 0 \), this means it is not possible to vary the ‘coordinate’ \( y^0 \) (i.e., along the integral curve of \( \frac{d}{dy^0} \)) without holding the ‘coordinate’ \( y^1 \) fixed; or, it is not possible to hold \( y^0 \) fixed while moving along the integral curve of \( \frac{d}{dy^1} \). More generally, in a \( d \)-dimensional space(time), if \( x^\mu \) is a coordinate parametrizing space(time), then it must be possible to vary it while keeping fixed the rest of its counterparts \( \{ x^\nu \} \neq \nu = \mu \).

**Problem 8.27. Example of non-commuting vector fields on \( S^2 \) (Schutz [19] Exercise 2.1)** In 2D flat space, starting from Cartesian coordinates \( x^i \), we may convert to cylindrical coordinates

\[
(x^1, x^2) = r(\cos \phi, \sin \phi). \tag{8.3.24}
\]

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 \( \partial_\phi \) and vice versa. However, show via a direct calculation that the following commutator involving the unit vector fields \( \hat{r} \) and \( \hat{\phi} \) is not zero:

\[
[\hat{r}, \hat{\phi}] f(r, \phi) \neq 0; \tag{8.3.25}
\]

where

\[
\hat{r} \equiv \cos(\phi) \partial_{x^1} + \sin(\phi) \partial_{x^2}, \tag{8.3.26}
\]

\[
\hat{\phi} \equiv -\sin(\phi) \partial_{x^1} + \cos(\phi) \partial_{x^2}. \tag{8.3.27}
\]

Therefore \( \hat{r} \) and \( \hat{\phi} \) do not form a coordinate basis.

\(^{91}\)See, for instance, Schutz [19] for a pedagogical discussion.
Timelike, Spacelike, and Null Distances/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 $i$th orthonormal space component – involves positive, Zero, and negative ‘distance-squared’.

With our ‘mostly minus’ sign convention (cf. eq. (8.1.1)), a vector $v^\mu$ is:

- **Time-like** if $v^2 \equiv \eta_{\mu\nu}v^\mu v^\nu > 0$. We have seen in §(8.1): if $v^2 > 0$, it is always possible to find a Lorentz transformation $\Lambda$ (cf. eq. (8.1.5)) such that $\Lambda^\mu_\alpha v^\alpha = (v^0, \vec{0})$. In flat spacetime, if $ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu > 0$ then this result indicates it is always possible to find an inertial frame where $ds^2 = dt^2$: hence the phrase ‘timelike’. (Also see Problem (8.10).)

More generally, for a timelike trajectory $z^\mu(\lambda)$ in curved spacetime – i.e.,

$$g_{\mu\nu}(dz^\mu/d\lambda)(dz^\nu/d\lambda) > 0,$$

we may identify

$$d\tau \equiv d\lambda \sqrt{g_{\mu\nu}(z(\lambda)) \frac{dz^\mu}{d\lambda} \frac{dz^\nu}{d\lambda}},$$

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 (8.5.6) through (8.5.8) 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 $g_{\mu\nu}dz^\mu dz^\nu = \eta_{00}(dz^0)^2 = (d\tau)^2$.

More generally, using the orthonormal frame fields in eq. (8.3.12),

$$d\tau = d\lambda \sqrt{\eta_{\alpha\beta} \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda}}, \quad \frac{dz^\alpha}{d\lambda} \equiv \varepsilon^\alpha_\mu(z(\lambda)) \frac{dz^\mu}{d\lambda}. \tag{8.3.30}$$

Since $v^\mu \equiv dz^\mu/d\lambda$ is assumed to be timelike, it must be possible to find a local Lorentz transformation $\Lambda^\mu_\nu(z)$ such that $\Lambda^\mu_\nu v^\nu = (v^0, \vec{0})$. Assuming $d\lambda > 0$,

$$d\tau = d\lambda \sqrt{\eta_{\mu\nu} \Lambda^\mu_\alpha \Lambda^\nu_\beta \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda}},$$

$$= d\lambda \sqrt{(\frac{dz^0}{d\lambda})^2 = |dz^0|}. \tag{8.3.31}$$

The generalization of the discussion in Problem (8.10) to timelike trajectories $z^\mu(\tau)$ in generic curved spacetimes is as follows. If $\tau$ refers to its proper time and $u^\mu \equiv dz^\mu/d\tau$, then $u^0$ cannot be arbitrary but is related to the proper spatial velocity $\vec{u}$ via

$$g_{00}(u^0)^2 + 2g_{0i}u^0 u^i + g_{ij}u^i u^j = +1. \tag{8.3.32}$$
Multiplying throughout by \(1/(u^0)^2 = (dx^0/d\tau)^2\),
\[
    g_{00} + 2g_{0i} \frac{dx^i}{d\tau} + g_{ij} \left( \frac{dx^i}{d\tau} \right) \left( \frac{dx^j}{d\tau} \right) = \left( \frac{dx^0}{d\tau} \right)^2,
\]

and
\[
    g_{\mu\nu} \frac{dz^\mu}{dx^0} \frac{dz^\nu}{dx^0} = \left( \frac{d\tau}{dx^0} \right)^2. \tag{8.3.34}
\]

Furthermore, if the trajectory is moving forward in time, then \(u^0 = dx^0/d\tau > 0\) and the positive square root is to be chosen:
\[
    \frac{d\tau}{dx^0} = \sqrt{g_{\mu\nu} \frac{dz^\mu}{dx^0} \frac{dz^\nu}{dx^0}}. \tag{8.3.35}
\]

- **Space-like** if \(v^2 \equiv \eta_{\mu\nu}v^\mu v^\nu < 0\). We have seen in §(8.1): if \(v^2 < 0\), it is always possible to find a Lorentz transformation \(\Lambda\) such that \(\Lambda^\mu_\alpha v^\alpha = (0, v^i)\). In flat spacetime, if \(ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu < 0\) then this result indicates it is always possible to find an inertial frame where \(ds^2 = -dx^2\): hence the phrase ‘spacelike’.

More generally, for a spacelike trajectory \(z^\mu(\lambda)\) in curved spacetime – i.e., \(g_{\mu\nu}(dz^\mu/d\lambda)(dz^\nu/d\lambda) < 0\), we may identify
\[
    d\ell \equiv d\lambda \sqrt{g_{\mu\nu}(z(\lambda)) \frac{dz^\mu}{d\lambda} \frac{dz^\nu}{d\lambda}}, \tag{8.3.36}
\]

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 \(dt = 0\), i.e., the rod is lying on the constant-\(t\) surface, then \(d\ell = |d\vec{x} \cdot d\vec{x}|^{1/2}\).) Using the orthonormal frame fields in eq. (8.3.12),
\[
    d\ell = d\lambda \sqrt{\left| \eta_{\alpha\beta} \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda} \right|}, \quad \frac{dz^\alpha}{d\lambda} \equiv \varepsilon^{\alpha\mu} \frac{dz^\mu}{d\lambda}. \tag{8.3.37}
\]

Furthermore, since \(v^\mu \equiv dz^\mu/d\lambda\) is assumed to be spacelike, it must be possible to find a local Lorentz transformation \(\Lambda^\mu_\nu(z)\) such that \(\Lambda^\mu_\nu v^\nu = (0, v^\mu)\); assuming \(d\lambda > 0\),
\[
    d\ell = d\lambda \sqrt{\eta_{\mu\nu} \Lambda^\mu_\alpha \Lambda^\nu_\beta \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda}} = |dz^\nu|; \tag{8.3.38}
\]
\[
    d\vec{z}^\nu \equiv \Lambda^\nu_\mu \varepsilon^{\mu\mu} v dz^\nu. \tag{8.3.39}
\]

- **Null** if \(v^2 \equiv \eta_{\mu\nu}v^\mu v^\nu = 0\). We have already seen, in flat spacetime, if \(ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu = 0\) then \(|d\vec{x}|/dx^0 = |dx^\nu|/dx^0 = 1\) 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’

\[ v^2 \equiv \eta_{\mu\nu} v^\mu v^\nu = \eta_{\alpha\beta} \Lambda^\alpha_{\mu} \Lambda^\beta_{\nu} v^\mu v^\nu = \eta_{\alpha\beta} v^\alpha v^\beta \]  

(8.3.40)

has to be invariant under all Lorentz transformations \( v^\beta = \Lambda^\beta_{\mu} v^\mu \). 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.

**Problem 8.28. Orthonormal Frames in Kerr-Schild Spacetimes**  A special class of geometries, known as *Kerr-Schild* spacetimes, take the following form.

\[ g_{\mu\nu} = \bar{g}_{\mu\nu} + H k_{\mu} k_{\nu} \]  

(8.3.41)

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\(^{92}\). The \( k_{\mu} \) is null with respect to \( \bar{g}_{\mu\nu} \), i.e.,

\[ \bar{g}_{\alpha\beta} k^\alpha k^\beta = 0, \]  

(8.3.42)

and we shall move its indices with \( \bar{g}_{\mu\nu} \).

Verify that the inverse metric is

\[ g^{\mu\nu} = \bar{g}^{\mu\nu} - H k^\mu k^\nu, \]  

(8.3.43)

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

\[ \bar{g}_{\mu\nu} = \eta_{\alpha\beta} \varepsilon^\alpha_{\mu} \varepsilon^\beta_{\nu}; \]  

(8.3.44)

verify that the orthonormal frame fields are

\[ \varepsilon^\hat{\alpha}_{\mu} = \varepsilon^\hat{\alpha}_{\sigma} \left( \delta^\sigma_{\mu} + \frac{1}{2} H k^\sigma k_{\mu} \right). \]  

(8.3.45)

Can you explain why \( k^\mu \) is also null with respect to the full metric \( g_{\mu\nu} \)?

**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

\[ d\tau_y \equiv dt \sqrt{g_{\mu\nu}(y(t)) y'^\mu y'^\nu}, \quad \dot{y}^\mu \equiv \frac{dy^\mu}{dt}; \]  

(8.3.46)

\(^{92}\)See Gibbons et al. arXiv: hep-th/0404008. The special property of Kerr-Schild coordinates is that Einstein’s equations become *linear* in these coordinates.
\[\text{Problem 8.29. Example}\]

The spacetime geometry around the Earth itself can be approximated by the line element
\[
ds^2 = \left(1 - \frac{r_{s,E}}{r}\right)dt^2 - \frac{dr^2}{1 - r_{s,E}/r} - r^2 \left( d\theta^2 + \sin(\theta)^2 d\phi^2 \right), \quad (8.3.49)
\]

where \(t\) is the time coordinate and \((r, \theta, \phi)\) 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
\[
r_{s,E} \equiv 2G_NM_E. \quad (8.3.50)
\]

Find the 4–beins of the geometry in eq. (8.3.49). Then find the numerical value of \(r_{s,E}/R_E\), where \(R_E\) is the radius of the Earth. Explain why this means we may – for practical purposes – expand the metric in eq. (8.3.50) as
\[
ds^2 = \left(1 - \frac{r_{s,E}}{r}\right)dt^2 - \frac{dr^2}{1 - r_{s,E}/r} - r^2 \left( d\theta^2 + \sin(\theta)^2 d\phi^2 \right) - r^2 \left( d\theta^2 + \sin(\theta)^2 d\phi^2 \right). \quad (8.3.51)
\]

Since we are not in flat spacetime, the \((t, r, \theta, \phi)\) are no longer subject to the same interpretation. However, use your computation of \(r_{s,E}/R_E\) 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.

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 \(\Delta\tau_B\); that at the tip as \(\Delta\tau_T\); and assuming for simplicity the Earth is a perfect sphere – show that eq. (8.3.48) translates to
\[
\frac{\Delta\tau_B}{\Delta\tau_T} = \sqrt{\frac{g_{00}(R_E)}{g_{00}(R_E + h_{101})}} \approx 1 + \frac{1}{2} \left( \frac{r_{s,E}}{R_E + h_{101}} - \frac{r_{s,E}}{R_E} \right). \quad (8.3.52)
\]
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. [8.3.52] analytically, then plug in the relevant numbers to compute the numerical value of \( \Delta \tau_B / \Delta \tau_T \). Does the clock at the base of Taipei 101 or that on its tip tick more slowly?

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 [31].

### 8.4 Connections, Curvature, Geodesics

#### Connections & Christoffel Symbols

The partial derivative on a scalar \( \varphi \) is a rank-1 tensor, so we shall simply define the covariant derivative acting on \( \varphi \) to be

\[
\nabla_\alpha \varphi = \partial_\alpha \varphi. \tag{8.4.1}
\]

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.,

\[
\nabla_\alpha V^\mu = \partial_\alpha V^\mu + \Gamma^\mu_{\sigma\rho} V^\rho. \tag{8.4.2}
\]

Under a coordinate transformation of the partial derivatives and \( V^\mu \), say going from \( x \) to \( x' \),

\[
\partial_\sigma V^\mu + \Gamma^\mu_{\rho\sigma} V^\rho = \frac{\partial x^\lambda}{\partial x'^\sigma} \frac{\partial x^\mu}{\partial x'^\nu} \partial_\lambda V^\nu + \left( \frac{\partial x^\lambda}{\partial x'^\sigma} \frac{\partial^2 x^\mu}{\partial x'^\lambda \partial x'^\nu} + \Gamma^\mu_{\rho\sigma} \frac{\partial x^\rho}{\partial x'^\nu} \right) V^\nu. \tag{8.4.3}
\]

On the other hand, if \( \nabla_\sigma V^\mu \) were to transform as a tensor,

\[
\partial_\sigma V^\mu + \Gamma^\mu_{\rho\sigma} V^\rho = \frac{\partial x'^\lambda}{\partial x^\sigma} \frac{\partial x^\mu}{\partial x'^\nu} \partial_\lambda V^\nu + \frac{\partial x'^\lambda}{\partial x^\sigma} \frac{\partial x^\mu}{\partial x'^\nu} \Gamma'^\rho_{\lambda\nu} V^\nu. \tag{8.4.4}
\]

\[93\] Since \( V^\nu \) is an arbitrary vector, we may read off its coefficient on the right hand sides of equations [8.4.3] and [8.4.4], and deduce the connection has to transform as

\[
\frac{\partial x'^\lambda}{\partial x^\sigma} \frac{\partial^2 x^\mu}{\partial x'^\lambda \partial x'^\nu} + \Gamma^\mu_{\rho\sigma}(x) \frac{\partial x^\rho}{\partial x'^\tau} \Gamma'^\tau_{\lambda\nu} V^\nu = \frac{\partial x'^\lambda}{\partial x^\sigma} \frac{\partial x^\mu}{\partial x'^\nu} \Gamma'^\rho_{\lambda\nu}(x'). \tag{8.4.5}
\]

Moving all the Jacobians onto the connection written in the \( \{ x^\mu \} \) frame,

\[
\Gamma'^\rho_{\kappa\nu}(x') = \frac{\partial x'^\tau}{\partial x^\mu} \frac{\partial^2 x^\mu}{\partial x'^\kappa \partial x'^\nu} + \frac{\partial x'^\tau}{\partial x^\mu} \Gamma^\mu_{\sigma\rho}(x) \frac{\partial x^\sigma}{\partial x'^\kappa} \frac{\partial x^\rho}{\partial x'^\nu}. \tag{8.4.6}
\]

All connections have to satisfy this non-tensorial transformation law. On the other hand, if we found an object that transforms according to eq. [8.4.6], and if one employs it in eq. [8.4.2], then the resulting \( \nabla_\alpha V^\mu \) would transform as a tensor.

---

\[93\] All un-primed indices represent tensor components in the \( x \)-system; while all primed indices those in the \( x' \) system.
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,

$$\nabla(T_1T_2) = (\nabla T_1)T_2 + T_1(\nabla T_2).$$  \hspace{1cm} (8.4.7)

**Problem 8.30. Covariant Derivative on 1-form**  Let us take the covariant derivative of a 1-form:

$$\nabla_\alpha V_\mu = \partial_\alpha V_\mu + \Gamma^\sigma_{\alpha\mu} V_\sigma.$$  \hspace{1cm} (8.4.8)

Can you prove that this connection is negative of the vector one in eq. (8.4.2)?

$$\Gamma^\sigma_{\alpha\mu} = -\Gamma^\sigma_{\alpha\mu},$$  \hspace{1cm} (8.4.9)

where $\Gamma^\sigma_{\alpha\mu}$ is the connection in eq. (8.4.2) – if we define the covariant derivative of a scalar to be simply the partial derivative acting on the same, i.e.,

$$\nabla_\alpha (V^\mu W_\mu) = \partial_\alpha (V^\mu W_\mu)?$$  \hspace{1cm} (8.4.10)

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. (8.4.9).

Suppose we found two such connections, \( (1) \Gamma_{\kappa\nu}(x) \) and \( (2) \Gamma_{\kappa\nu}(x) \). 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:

$$\left( (1) \Gamma_{\kappa\nu}(x') - (2) \Gamma_{\kappa\nu}(x') \right) = \frac{\partial x'^\kappa}{\partial x^\mu} \left( (1) \Gamma_{\sigma\rho}(x) - (2) \Gamma_{\sigma\rho}(x) \right) \frac{\partial x^{\sigma}}{\partial x'^\kappa} \frac{\partial x^{\rho}}{\partial x'^\nu}. \hspace{1cm} (8.4.11)$$

Now, any connection can be decomposed into its symmetric and antisymmetric parts in the following sense:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} \Gamma^\mu_{\{\alpha\beta\}} + \frac{1}{2} \Gamma^\mu_{[\alpha\beta]}.$$  \hspace{1cm} (8.4.12)

This is, of course, mere tautology. However, let us denote

$$\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2} \Gamma^\mu_{\alpha\beta} \quad \text{and} \quad \Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2} \Gamma^\mu_{\beta\alpha};$$  \hspace{1cm} (8.4.13)

so that

$$\frac{1}{2} \Gamma^\mu_{[\alpha\beta]} = (1) \Gamma^\mu_{\alpha\beta} - (2) \Gamma^\mu_{\alpha\beta} \equiv T^\mu_{\alpha\beta}.$$  \hspace{1cm} (8.4.14)

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

$$\nabla_\alpha V^\mu = \partial_\alpha V^\mu + \Gamma^\mu_{\alpha\beta} V^\beta + T^\mu_{\alpha\beta} V^\beta.$$  \hspace{1cm} (8.4.15)
As is common within the physics literature, we proceed to set to zero the torsion term: $T_{\alpha\beta}^\mu \to 0$. If we further impose the metric compatibility condition,

$$\nabla_\mu g_{\alpha\beta} = 0,$$

then we have already seen in §(7) this (together with the zero torsion assumption) implies

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\sigma}(\partial_\alpha g_{\beta\sigma} + \partial_\beta g_{\alpha\sigma} - \partial_\sigma g_{\alpha\beta}).$$

(8.4.17)

**Parallel Transport & Riemann Tensor**

Along a curve $z^\mu(\lambda)$ such that one end is $z^\mu(\lambda_1) = x'^\mu$ and the other end is $z^\mu(\lambda_2) = x^\mu$, we may parallel transport some vector $V^\alpha$ 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 $V^\alpha$ evaluated at $x'$ – we have

$$V^\alpha\left(x' \xrightarrow{z(\lambda)} x\right) = \exp\left[(\lambda_2 - \lambda_1)\dot{z}^\mu(\lambda_1)\nabla_\mu\right]V^\alpha(x').$$

(8.4.18)

This is the covariant derivative analog of the Taylor expansion of a scalar function – where, translation by a constant spacetime vector $a^\mu$ may be implemented as

$$f(x^\mu + a^\mu) = \exp(a^\nu\partial_\nu)f(x^\mu).$$

(8.4.19)

Eq. (8.4.18) is also consistent with the discussion leading up to eq. (7.3.18), which in the curved spacetime context would be: a spacetime tensor $T_{\mu_1\cdots\mu_N}^\mu$ is invariant under parallel transport along some curve whose tangent vector is $v^\mu$, whenever

$$v^\sigma\nabla_\sigma T_{\mu_1\cdots\mu_N}^\mu = 0$$

(8.4.20)

along the entire curve. For, once $\dot{z}^\mu(\lambda_1)$ in eq. (8.4.18) is identified with $v^\mu$, if eq. (8.4.20) is satisfied then

$$\exp\left[(\lambda_2 - \lambda_1)v^\mu(x')\nabla_\mu\right]V^\alpha(x') = V^\alpha(x'),$$

(8.4.21)

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 $V^\alpha$ 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 $\epsilon$. We will now start from one of its 4 corners, which we will denote as $x$. $V^\alpha$ 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

Note that if we were to relax both the zero torsion and metric compatibility conditions, this amounts to introducing two new tensors: $(1/2)\Gamma^\mu_{[\alpha\beta]} = T_{\alpha\beta}^\mu$ 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 $T_{\alpha\beta}^\mu$ and $Q_{\mu\alpha\beta}$ obey; and, what are they sourced by?
to $x + \epsilon B - \epsilon B = x$. Let us first work out the parallel transport along the ‘side’ $A$ using eq. (8.4.18). Denoting $\nabla_A \equiv A^\mu \nabla_\mu$, $\nabla_B \equiv B^\mu \nabla_\mu$, etc.,

$$V^\alpha(x \to x + \epsilon A) = \exp(\epsilon \nabla_A) V^\alpha(x),$$

$$= V^\alpha(x) + \epsilon \nabla_A V^\alpha(x) + \frac{\epsilon^2}{2} \nabla_A^2 V^\alpha(x) + O(\epsilon^3). \tag{8.4.22}$$

We then parallel transport this result from $x + \epsilon A$ to $x + \epsilon A + \epsilon B$.

$$V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B)$$

$$= \exp(\epsilon \nabla_B) \exp(\epsilon \nabla_A) V^\alpha(x),$$

$$= V^\alpha(x) + \epsilon \nabla_A V^\alpha(x) + \frac{\epsilon^2}{2} \nabla_A^2 V^\alpha(x)$$

$$+ \epsilon \nabla_B V^\alpha(x) + \epsilon^2 \nabla_B \nabla_A V^\alpha(x)$$

$$+ \frac{\epsilon^2}{2} \nabla_B^2 V^\alpha(x) + O(\epsilon^3)$$

$$= V^\alpha(x) + \epsilon (\nabla_A + \nabla_B) V^\alpha(x) + \frac{\epsilon^2}{2} \left( \nabla_A^2 + \nabla_B^2 + 2 \nabla_B \nabla_A \right) V^\alpha(x) + O(\epsilon^3). \tag{8.4.23}$$

Pressing on, we now parallel transport this result from $x + \epsilon A + \epsilon B$ to $x + \epsilon B$.

$$V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x + \epsilon B)$$

$$= \exp(-\epsilon \nabla_A) \exp(\epsilon \nabla_B) \exp(\epsilon \nabla_A) V^\alpha(x),$$

$$= V^\alpha(x) + \epsilon (\nabla_A + \nabla_B) V^\alpha(x) + \frac{\epsilon^2}{2} \left( \nabla_A^2 + \nabla_B^2 + 2 \nabla_B \nabla_A \right) V^\alpha(x)$$

$$- \epsilon \nabla_A V^\alpha(x) - \epsilon^2 \left( \nabla_A^2 + \nabla_A \nabla_B \right) V^\alpha(x)$$

$$+ \frac{\epsilon^2}{2} \nabla_A^2 V^\alpha(x) + O(\epsilon^3)$$

$$= V^\alpha(x) + \epsilon \nabla_B V^\alpha(x) + \epsilon^2 \left( \frac{1}{2} \nabla_B^2 + \nabla_B \nabla_A - \nabla_A \nabla_B \right) V^\alpha(x) + O(\epsilon^3). \tag{8.4.24}$$

Finally, we parallel transport this back to $x + \epsilon B - \epsilon B = x$.

$$V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x + \epsilon B \to x)$$

$$= \exp(-\epsilon \nabla_B) \exp(-\epsilon \nabla_A) \exp(\epsilon \nabla_B) \exp(\epsilon \nabla_A) V^\alpha(x),$$

$$= V^\alpha(x) + \epsilon \nabla_B V^\alpha(x) + \epsilon^2 \left( \frac{1}{2} \nabla_B^2 + \nabla_B \nabla_A - \nabla_A \nabla_B \right) V^\alpha(x)$$

$$- \epsilon \nabla_B V^\alpha(x) - \epsilon^2 \nabla_B^2 V^\alpha(x)$$

$$+ \frac{\epsilon^2}{2} \nabla_B^2 V^\alpha(x) + O(\epsilon^3)$$

$$= V^\alpha(x) + \epsilon^2 \left( \nabla_B \nabla_A - \nabla_A \nabla_B \right) V^\alpha(x) + O(\epsilon^3). \tag{8.4.25}$$

\footnote{We have arrived at the central characterization of \textit{local} geometric curvature. By parallel transporting a vector around an infinitesimal parallelogram, we see the parallel transported \ref{8.4.18} The careful reader may complain, we should have evaluated the covariant derivatives at the \ref{8.4.22} The \ref{8.4.23} The \ref{8.4.24} The \ref{8.4.25} The}
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 parallelogram, i.e., it scales as $\mathcal{O}(\epsilon^2)$ for infinitesimal $\epsilon$.

\[ V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x + \epsilon B \to x) - V^\alpha(x) = \epsilon^2 [\nabla_B, \nabla_A] V^\alpha(x) + \mathcal{O}(\epsilon^3), \]

We shall proceed to calculate the commutator in a coordinate basis.

\[ [\nabla_A, \nabla_B] V^\mu \equiv A^\sigma [\nabla_A (B^\sigma \nabla_\rho V^\mu) - B^\sigma \nabla_A (A^\sigma \nabla_\rho V^\mu)] \]

\[ = (A^\sigma \partial_\sigma B^\rho - B^\sigma \partial_\sigma A^\rho) \nabla_\rho V^\mu + A^\sigma B^\rho [\nabla_\sigma, \nabla_\rho] V^\mu. \]

Let us tackle the two groups separately. Firstly,

\[ [A, B]^\rho \nabla_\rho V^\mu \equiv (A^\sigma \partial_\sigma B^\rho - B^\sigma \partial_\sigma A^\rho) \partial_\rho V^\mu \]

\[ = (A^\sigma \partial_\sigma B^\rho + \Gamma^\rho_{\sigma\lambda} A^\sigma B^\lambda - B^\sigma \partial_\sigma A^\rho - \Gamma^\rho_{\sigma\lambda} B^\sigma A^\lambda) \partial_\rho V^\mu \]

\[ = (A^\sigma \partial_\sigma B^\rho - B^\sigma \partial_\sigma A^\rho) \nabla_\rho V^\mu. \]

Next, we need $A^\rho B^\sigma [\nabla_\sigma, \nabla_\rho] V^\mu = A^\rho B^\sigma (\nabla_\sigma \nabla_\rho - \nabla_\rho \nabla_\sigma) V^\mu$. The first term is

\[ A^\sigma B^\rho \nabla_\sigma \nabla_\rho V^\mu = A^\sigma B^\rho (\partial_\sigma \nabla_\rho V^\mu - \Gamma^\lambda_{\sigma\rho} \nabla_\lambda V^\mu + \Gamma^\mu_{\sigma\lambda} \partial_\rho V^\lambda) \]

\[ = A^\sigma B^\rho (\partial_\sigma (\partial_\rho V^\mu) + \Gamma^\mu_{\rho\lambda} \partial_\sigma V^\lambda - \Gamma^\lambda_{\sigma\rho} (\partial_\lambda V^\mu + \Gamma^\mu_{\lambda\omega} V^\omega) + \Gamma^\mu_{\sigma\lambda} (\partial_\lambda V^\mu + \Gamma^\mu_{\lambda\omega} V^\omega)) \]

\[ = A^\sigma B^\rho \left\{ \partial_\sigma \partial_\rho V^\mu + \partial_\sigma \Gamma^\mu_{\rho\lambda} V^\lambda + \Gamma^\mu_{\rho\lambda} \partial_\sigma V^\lambda - \Gamma^\lambda_{\sigma\rho} (\partial_\lambda V^\mu + \Gamma^\mu_{\lambda\omega} V^\omega) \right\} \]

\[ + \Gamma^\mu_{\sigma\lambda} (\partial_\lambda V^\mu + \Gamma^\mu_{\lambda\omega} V^\omega) \].

Swapping ($\sigma \leftrightarrow \rho$) within the parenthesis $\{\ldots\}$ and subtract the two results, we gather

\[ A^\rho B^\sigma [\nabla_\sigma, \nabla_\rho] V^\mu = A^\rho B^\sigma \left\{ \partial_\sigma \Gamma^\mu_{\rho\lambda} V^\lambda + \Gamma^\mu_{\lambda\rho} \partial_\sigma V^\lambda - \Gamma^\lambda_{\rho\sigma} (\partial_\lambda V^\mu + \Gamma^\mu_{\lambda\omega} V^\omega) \right\} \]

\[ + \Gamma^\mu_{\rho\sigma} (\partial_\sigma \Gamma^\lambda_{\rho\omega} V^\omega) \].

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:

\[ [\nabla_\mu, \nabla_\nu] V^\alpha = R^\alpha_{\beta\mu\nu} V^\beta; \]

\[ R^\alpha_{\beta\mu\nu} \equiv \partial_\mu \Gamma^\alpha_{\nu\beta} + \Gamma^\alpha_{\sigma\mu} \partial_\nu \Gamma^\sigma_{\beta\nu} \]

various corners of the parallelogram — namely, $\exp(-\epsilon \nabla B(x + \epsilon A(x)) - \epsilon A(x + \epsilon A(x))) - \epsilon A(x + \epsilon A(x)) + \epsilon B(x + \epsilon A(x)))$ — 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 $(1 + \mathcal{O}(\epsilon))$. 314
\[ = \partial_\mu \Gamma^\alpha_{\nu\beta} - \partial_\nu \Gamma^\alpha_{\mu\beta} + \Gamma^\alpha_{\sigma\mu} \Gamma^\sigma_{\nu\beta} - \Gamma^\alpha_{\sigma\nu} \Gamma^\sigma_{\mu\beta}. \quad (8.4.35) \]

Inserting the results in equations (8.4.29) and (8.4.32) into eq. (8.4.28) – we gather, for arbitrary vector fields \(A\) and \(B\):

\[ (\nabla_A, \nabla_B) V^\mu = R^\mu_{\nu\alpha\beta} V^\nu A^\alpha B^\beta. \quad (8.4.36) \]

Moreover, we may return to eq. (8.4.26) and re-express it as

\[ V^\alpha(x \to x + \epsilon A \to x + \epsilon A + \epsilon B \to x) = V^\alpha(x) - \epsilon^2 R^\alpha_{\beta\mu\nu}(x) V^\beta(x) A^\mu(x) + O(\epsilon^3). \quad (8.4.37) \]

When \(A = \partial_\mu\) and \(B = \partial_\nu\) are coordinate basis vectors themselves, \([A, B] = [\partial_\mu, \partial_\nu] = 0\), and eq. (8.4.36) then coincides with eq. (8.4.33). 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. (8.4.37) 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.

**Problem 8.31. 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\(^{96}\) and show that

\[ R_{\alpha\beta\mu\nu} = \frac{1}{2} (\partial_\beta \partial_\mu g_{\nu\alpha} - \partial_\alpha \partial_\mu g_{\nu\beta}). \quad (8.4.39) \]

From this result, argue that Riemann has the following symmetries:

\[ R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}, \quad R_{\mu\nu\alpha\beta} = -R_{\nu\alpha\beta\mu}, \quad R_{\mu\nu\alpha\beta} = -R_{\mu\nu\beta\alpha}. \quad (8.4.40) \]

This indicates the components of the Riemann tensor are not all independent. 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

\[ (\nabla_\alpha, \nabla_\beta) V_\nu = -R^\mu_{\nu\alpha\beta} V_\mu. \quad (8.4.41) \]

**Hint:** Start with \(\nabla_\alpha, \nabla_\beta)(g_{\nu\sigma} V^\sigma)\).

**Ricci tensor and scalar**

Because of the symmetries of Riemann in eq. (8.4.40), 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, \(R_{\alpha\beta\mu} = 0\). In fact, the Ricci tensor is defined as the sole distinct and non-zero contraction of Riemann:

\[ R_{\mu\nu} \equiv R^\sigma_{\mu\sigma\nu}. \quad (8.4.42) \]

\(^{96}\)See equations (8.5.6) through (8.5.8) below.
This is a symmetric tensor, \( R_{\mu\nu} = R_{\nu\mu} \), because of eq. \( (8.4.40) \); for,

\[
R_{\mu\nu} = g^{\sigma\rho}R_{\sigma\mu\rho\nu} = g^{\sigma\rho}R_{\rho\nu\sigma\mu} = R_{\nu\mu}.
\] (8.4.43)

Its contraction yields the Ricci scalar

\[
\mathcal{R} \equiv g^{\mu\nu}R_{\mu\nu}.
\] (8.4.44)

**Problem 8.32. Commutator of covariant derivatives on higher rank tensor**

Prove that

\[
\left[ \nabla_\alpha, \nabla_\beta \right] T^{\alpha_1...\alpha_N}_{\beta_1...\beta_M} = \sum \left( R_{\alpha_1}^{\alpha_2} \sigma_{\mu\nu} T^{\alpha_1...\alpha_{N-1}}_{\beta_1...\beta_M} + R_{\beta_1}^{\alpha_1} \sigma_{\mu\nu} T^{\alpha_2...\alpha_N}_{\beta_1...\beta_M} + \cdots \right)
\] (8.4.45)

Also verify that

\[
\left[ \nabla_\alpha, \nabla_\beta \right] \varphi = 0,
\] (8.4.46)

where \( \varphi \) is a scalar.

**Problem 8.33. Differential Bianchi identities I**

Show that

\[
R^\mu_{\ [\alpha\beta\delta]} = 0.
\] (8.4.47)

Hint: Use eq. \( (8.4.39) \), the Riemann tensor expressed in a locally flat coordinate system.

**Problem 8.34. Differential Bianchi identities II**

If \([A, B] \equiv AB - BA\), can you show that the differential operator

\[
\left[ \nabla_\alpha, \left[ \nabla_\beta, \nabla_\delta \right] \right] + \left[ \nabla_\beta, \left[ \nabla_\delta, \nabla_\alpha \right] \right] + \left[ \nabla_\delta, \left[ \nabla_\alpha, \nabla_\beta \right] \right]
\] (8.4.48)

is actually zero? (Hint: Just expand out the commutators.) Why does that imply

\[
\nabla_{[\alpha} R^\mu_{\beta\delta]} = 0?
\] (8.4.49)

Using this result, show that

\[
\nabla_\sigma R^\sigma_\mu = \nabla_{[\mu} R^\beta_{\nu]}.
\] (8.4.50)

The *Einstein tensor* is defined as

\[
G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R}.
\] (8.4.51)

From eq. \( (8.4.50) \) can you show the divergence-less property of the Einstein tensor, i.e.,

\[
\nabla^\mu G_{\mu\nu} = \nabla^\mu \left( R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} \right) = 0?
\] (8.4.52)

This is an important property when understanding Einstein’s equations of General Relativity,

\[
G_{\mu\nu} = 8\pi G N T_{\mu\nu};
\] (8.4.53)

where \( T_{\mu\nu} \) encodes the energy-momentum-stress-shear of matter. By employing eq. \( (8.4.52) \), we see that taking the divergence of eq. \( (8.4.53) \) leads us to the conservation of energy-momentum-stress-shear: \( \nabla^\mu T_{\mu\nu} = 0 \).
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 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 (7.18). On the other hand, in a locally flat or Fermi-Normal-Coordinate system – see equations (7.2.1) in the previous Chapter; and (8.5.6)–(8.5.8) 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, $d s^2$ is not positive-definite (cf. (8.1.1)), 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:

$$s = \int_{\lambda_1}^{\lambda_2} \left| g_{\mu\nu}(z(\lambda)) \frac{dz^\mu}{d\lambda} \frac{dz^\nu}{d\lambda} \right|^{1/2} d\lambda. \tag{8.4.54}$$

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 $\lambda$ such that $\sqrt{|g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu|} = \text{constant}$, this amounts to imposing the principle of stationary action on Synge’s world function (recall eq. (7.3.39)):

$$\sigma(x, x') \equiv \frac{1}{2} (\lambda_2 - \lambda_1) \int_{\lambda_1}^{\lambda_2} g_{\alpha\beta}(z(\lambda)) \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda} d\lambda, \tag{8.4.55}$$

$$z^\mu(\lambda_1) = x'^\mu, \quad z^\mu(\lambda_2) = x^\mu. \tag{8.4.56}$$

When evaluated on geodesics, eq. (8.4.55) is half the square of the geodesic distance between $x$ and $x'$. The curved spacetime geodesic equation in affine-parameter form which follows from eq. (8.4.55), is

$$D^2 z^\mu \equiv \frac{d^2 z^\mu}{d\lambda^2} + \Gamma^\mu_{\alpha\beta} \frac{dz^\alpha}{d\lambda} \frac{dz^\beta}{d\lambda} = 0. \tag{8.4.57}$$

Problem 8.35. Choice of ‘units’ for affine parameter  

Show that eq. (8.4.57) takes the same form under re-scaling and constant shifts of the parameter $\lambda$. That is, if

$$\lambda = a\lambda' + b, \tag{8.4.58}$$

for constants $a$ and $b$, then eq. (8.4.57) becomes

$$D^2 z^\mu \equiv \frac{d^2 z^\mu}{d\lambda'^2} + \Gamma^\mu_{\alpha\beta} \frac{dz^\alpha}{d\lambda'} \frac{dz^\beta}{d\lambda'} = 0. \tag{8.4.59}$$

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 §(7.3) had already informed us, the Lagrangian associated with eq. (8.4.55),
\[ L_g \equiv \frac{1}{2} g_{\mu\nu} (z(\lambda)) \dot{z}^\mu \dot{z}^\nu, \quad \dot{z}^\mu \equiv \frac{dz^\mu}{d\lambda}, \quad (8.4.60) \]
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),
\[ L_g[\lambda] = \frac{1}{2} g_{\mu\nu} (z(\lambda)) \frac{dz^\nu}{d\lambda} \frac{dz^\nu}{d\lambda} = \frac{1}{2} g_{\mu\nu} (z(\lambda')) \frac{dz^\nu}{d\lambda'} \frac{dz^\nu}{d\lambda'} \frac{1}{a^2} = \frac{L_g[\lambda']}{a^2}. \quad (8.4.61) \]
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 \( d\lambda \sqrt{g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu} = d\tau \), we see that the Lagrangian is then set to \( L_g = 1/2 \).
- If \( \dot{z}^\mu \) is spacelike, then by choosing the affine parameter to be proper length \( d\lambda \sqrt{|g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu|} = d\ell \), we see that the Lagrangian is then set to \( L_g = -1/2 \).
- If \( \dot{z}^\mu \) is null, then the Lagrangian is zero: \( L_g = 0 \). Since both sides of eq. (8.4.61) will remain zero under re-scaling, there is always a freedom to rescale the affine parameter by a constant:
\[ L_g[\lambda] = 0 = L_g[\lambda'], \quad (8.4.62) \]
whenever \( \lambda = (\) 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 zero proper time (from above).\(^{97}\) 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 \( x^\mu \) and orthonormal frame fields defined through \( g_{\mu\nu} = \eta_{\alpha\beta} \varepsilon^\alpha_{\mu} \varepsilon^\beta_{\nu} \), the proper time must be bounded by
\[ \int d\tau = \int \sqrt{(dz^0)^2 - \delta_{ij} dz^i dz^j} \leq \int |dz^0|, \quad dz^\tilde{\alpha} \equiv \varepsilon^\tilde{\alpha}_{\alpha} dz^\alpha. \quad (8.4.63) \]

**YZ: Need to check the above argument.** Therefore, at least locally,\(^{98}\) A timelike extremum must be a maximum.

\(^{97}\) A version of this argument may be found in Carroll’s lecture notes \[^{23}\].

\(^{98}\) 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 \( z^\mu(0 \leq \lambda \leq 1) = (0, \vec{x}) \).
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 99:

A spacelike extremum must be a minimum.

**Hamiltonian Dynamics of Geodesics** In §(7.3), we also delineated an alternate but equivalent Hamiltonian formulation for geodesic motion. The conjugate momentum $p_\mu$ to the coordinate $z^\mu$ is

$$p_\mu \equiv \frac{\partial L_g}{\partial \dot{z}^\mu} = g_{\mu\nu}\dot{z}^\nu.$$  
(8.4.64)

The Hamiltonian is

$$H(z, p) = \frac{1}{2} g^{\alpha\beta} (z(\lambda)) p_\alpha(\lambda) p_\beta(\lambda);$$  
(8.4.65)

and the associated Hamilton’s equations are

$$\frac{dz^\mu}{d\lambda} = \frac{\partial H}{\partial p_\mu} = g^{\mu\nu} p_\nu, \quad (8.4.66)$$

$$\frac{dp_\mu}{d\lambda} = -\frac{\partial H}{\partial z^\mu} = -\frac{1}{2} (\partial_\mu g^{\alpha\beta}) p_\alpha p_\beta. \quad (8.4.67)$$

Together, equations (8.4.66) and (8.4.67) are equivalent to eq. (8.4.57).

**Example** In flat spacetime, the Hamiltonian would read

$$H = \frac{1}{2} \eta^{\alpha\beta} p_\alpha p_\beta.$$  
(8.4.68)

Since $\eta^{\alpha\beta}$ is a constant matrix, we infer from equations (8.4.66) and (8.4.67) the conservation of linear momentum:

$$\dot{z}^\mu = \eta^{\mu\nu} p_\nu = p^\mu, \quad (8.4.69)$$

$$\dot{p}_\mu = 0. \quad (8.4.70)$$

**Formal solution to geodesic equation** We may re-write eq. (8.4.57) into an integral equation by simply integrating both sides with respect to the affine parameter $\lambda$:

$$v^\mu(\lambda) = v^\mu(\lambda_1) - \int_{z(\lambda_1)}^{z(\lambda)} \Gamma^\mu_{\alpha\beta} v^\alpha dz^\beta; \quad (8.4.71)$$

where $v^\mu \equiv dz^\mu/d\lambda$; 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.

99Globally, 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.
Problem 8.36. Synge’s World Function In Minkowski

Verify that Synge’s world function (cf. (8.4.55)) in Minkowski spacetime is

\[ \bar{\sigma}(x, x') = \frac{1}{2}(x - x')^2 \equiv \frac{1}{2}\eta_{\mu\nu}(x - x')^\mu(x - x')^\nu, \]  
(8.4.72)

\[ (x - x')^\mu \equiv x^\mu - x'^\mu. \]  
(8.4.73)

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

\[ z^\mu(0 \leq \lambda \leq 1) = x'^\mu + \lambda(x - x')^\mu. \]  
(8.4.74)

This is, of course, the ‘constant velocity’ solution of classical kinematics if we identify \( \lambda \) as a fictitious time.

Problem 8.37. Geodesic Vector Fields

Let \( v^\mu(x) \) be a vector field defined throughout a given spacetime. Show that the geodesic equation (8.4.57) follows from

\[ v^\sigma \nabla_\sigma v^\mu = 0, \]  
(8.4.75)

i.e., \( v^\mu \) is parallel transported along itself – provided we recall the ‘velocity flow’ interpretation of a vector field:

\[ v^\mu(z(s)) = \frac{dz^\mu}{ds}. \]  
(8.4.76)

Parallel transport preserves norm-squared

The metric compatibility condition in eq. (8.4.16) obeyed by the covariant derivative \( \nabla_\alpha \) 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 \( (v^\mu \) subject to eq. (8.4.75)) be preserved under parallel transport. Can you explain this statement using the appropriate equations?

Non-affine form of geodesic equation

Suppose instead

\[ v^\sigma \nabla_\sigma v^\mu = \kappa v^\mu. \]  
(8.4.77)

This is the more general form of the geodesic equation, where the parameter \( \lambda \) 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. (8.4.77) can be transformed into the form in eq. (8.4.75); that is, identify an appropriate \( v'^\mu \) such that

\[ v'^\sigma \nabla_\sigma v'^\mu = 0. \]  
(8.4.78)

You should comment on how this re-scaling fails when \( v^\mu \) is null.

Starting from the finite distance integral

\[ s \equiv \int_{\lambda_1}^{\lambda_2} d\lambda \sqrt{|g_{\mu\nu}(z(\lambda))\dot{z}^\mu \dot{z}^\nu|}, \]  
\[ \dot{z}^\mu \equiv \frac{dz^\mu}{d\lambda}, \]  
(8.4.79)

\[ z^\mu(\lambda_1) = x', \]  
\[ z^\mu(\lambda_2) = x; \]  
(8.4.80)

show that demanding \( s \) be extremized leads to the non-affine geodesic equation

\[ \ddot{z}^\mu + \Gamma^\mu_{\alpha\beta} \dot{z}^\alpha \dot{z}^\beta = \frac{d}{d\lambda} \ln \sqrt{g_{\alpha\beta} \dot{z}^\alpha \dot{z}^\beta}. \]  
(8.4.81)
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:

$$ds^2 = dt^2 - a(t)^2 d\vec{x} \cdot d\vec{x}.$$  \hfill (8.4.82)

Let us demonstrate that

$$U^\mu = \delta_0^\mu$$  \hfill (8.4.83)

is in fact a timelike geodesic vector field. Firstly,

$$g_{\mu\nu}U^\mu U^\nu = g_{00} = 1 > 0.$$  \hfill (8.4.84)

Next, keeping in mind eq. (8.4.83), we compute

$$U^\mu \nabla_\mu U^\nu = \nabla_\nu U^\alpha = \partial_\nu \delta_0^\alpha + \Gamma_\alpha^{\nu\alpha} = \frac{1}{2} g^{\alpha\sigma} (\partial_\alpha g_{\nu\sigma} + \partial_\nu g_{\alpha\sigma} - \partial_{\sigma} g_{\alpha\nu}) = g^{\alpha\sigma} \partial_\alpha g_{\nu\sigma} = 0.$$  \hfill (8.4.85)

The interpretation is that $U^\mu = \delta_0^\mu$ 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. (7.4.22))

$$H_{\mu\nu} = g_{\mu\nu} - U_\mu U_\nu$$  \hfill (8.4.87)

is the metric orthogonal to $U^\mu$ itself; namely,

$$H_{\mu\nu} U^\nu = U_\mu - U_\mu (U_\nu U^\nu) = 0$$  \hfill (8.4.88)

because eq. (8.4.84) tells us $U_\nu U^\nu = 1$. The space orthogonal to $U_\mu$ reads

$$dl^2 = -H_{\mu\nu} dx^\mu dx^\nu = -(dt^2 - a^2 d\vec{x} \cdot d\vec{x} - (U_\mu dx^\mu)^2) = a(t)^2 d\vec{x} \cdot d\vec{x},$$  \hfill (8.4.89)

as $(U_\mu dx^\mu)^2 = (\delta_0^\mu dx^\mu)^2 = dt^2$. It is expanding/contracting, with relative $t$—dependent size governed by $a(t)$.

**Problem 8.38. Null Geodesics & Weyl Transformations** Suppose two geometries $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$ are related via a Weyl transformation

$$g_{\mu\nu}(x) = \Omega(x)^2 \bar{g}_{\mu\nu}(x).$$  \hfill (8.4.90)

Consider the null geodesic equation in the geometry $g_{\mu\nu}(x)$,

$$k^\nu \nabla_\sigma k^\mu = 0, \quad g_{\mu\nu} k^\mu k^\nu = 0$$  \hfill (8.4.91)

where $\nabla$ is the covariant derivative with respect to $g_{\mu\nu}$; as well as the null geodesic equation in $\bar{g}_{\mu\nu}(x)$,

$$k^\sigma \nabla_\sigma k^\mu = 0, \quad \bar{g}_{\mu\nu} k^\mu k^\nu = 0;$$  \hfill (8.4.92)
where $\nabla$ is the covariant derivative with respect to $\bar{g}_{\mu\nu}$. Show that

$$k^\mu = \Omega^2 \cdot k'{}^\mu. \quad (8.4.93)$$

Hint: First show that the Christoffel symbol $\Gamma^\mu{}_{\alpha\beta}[\bar{g}]$ 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

$$\Gamma^\mu{}_{\alpha\beta}[g] = \bar{\Gamma}^\mu{}_{\alpha\beta}[\bar{g}] + \delta^\mu_{\{\beta} \nabla_{\alpha\}} \ln \Omega - \bar{g}_{\alpha\beta} \nabla^\mu \ln \Omega. \quad (8.4.94)$$

Then remember to use the constraint $g_{\mu\nu}k'^\mu k'^\nu = 0 = \bar{g}_{\mu\nu}k^\mu k^\nu$.

A spacetime is said to be conformally flat if it takes the form

$$g_{\mu\nu} = \Omega(x)^2 \eta_{\mu\nu}. \quad (8.4.95)$$

Solve the null geodesic equation explicitly in such a spacetime.

**Problem 8.39. 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 $\Phi$, through the metric

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\Phi(\vec{x}) \delta_{\mu\nu}. \quad (8.4.96)$$

Within 4D Linearized General Relativity, we will find that the Newtonian potential is sourced by the astrophysical energy density $\rho$ via Poisson’s equation:

$$\vec{\nabla}^2 \Phi(\vec{x}) = 4\pi G_N \rho(\vec{x}). \quad (8.4.97)$$

In §[9] below, we shall solve this equation through the Euclidean Green’s function.

$$\Phi(\vec{x}) = -G_N \int_{\mathbb{R}^3} \frac{\rho(\vec{x})}{|\vec{x} - \vec{x}'|} d^3\vec{x}'. \quad (8.4.98)$$

Let us shoot a beam of light from one side of the astrophysical system to opposite side, through its central region where $\Phi$ 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 $\Phi$ 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 $t_e$.

First show that, by virtue of being a null signal, eq. (8.6.85) leads to the expression

$$T^2 = R^2 - 2(T^2 + R^2) \int_0^1 \Phi(\vec{x}_e + \lambda(\vec{x}_r - \vec{x}_e)) d\lambda + \mathcal{O}(\Phi^2). \quad (8.4.99)$$

where

$$T \equiv t_r - t_e \quad \text{and} \quad R \equiv |\vec{x}_r - \vec{x}_e|. \quad (8.4.100)$$
According to eq. (8.4.99), $T^2$ goes as $R^2$ plus an order $\Phi$ correction. Therefore, replacing the $T^2$ on the right hand side of eq. (8.4.99) with $R^2$ would incur an error of order $\Phi^2$. Explain why the time elapsed $T = t_r - t_e$ is thus

$$T = R \left( 1 - 2 \int_0^1 \Phi (\vec{x}_e + \lambda (\vec{x}_r - \vec{x}_e)) \, d\lambda \right) + O (\Phi^2).$$  (8.4.101)

Why is this a time delay? Hint: What sign is the gravitational potential $\Phi$? 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 [33, 34, 35]. To date, the most precise Shapiro time-delay measurement is from the Doppler tracking of the Cassini spacecraft; see §4.1.2 of [32].

8.5 Equivalence Principles, Geometry-Induced Tidal Forces, Isometries & Geometric Tensors

Weak Equivalence Principle, ‘Free-Fall’ & 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 inhabits 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

$$M_1 \frac{d^2 \vec{x}_1}{dt^2} = -\hat{n} G_m M_1 M_2 \frac{1}{|\vec{x}_1 - \vec{x}_2|^2}, \quad \hat{n} \equiv \frac{\vec{x}_1 - \vec{x}_2}{|\vec{x}_1 - \vec{x}_2|}.$$  (8.5.1)

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:

$$\frac{d^2 \vec{x}_1}{dt^2} = -\hat{n} G_m M_2 \frac{1}{|\vec{x}_1 - \vec{x}_2|^2}. \quad (8.5.2)$$

Similarly, the acceleration of body 2 due to the gravitational force exerted by body 1 is independent of $M_2$:

$$\frac{d^2 \vec{x}_2}{dt^2} = +\hat{n} G_m M_1 \frac{1}{|\vec{x}_1 - \vec{x}_2|^2}. \quad (8.5.3)$$

This Weak Equivalence Principle$^{100}$ 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 (timelike) geodesics \( z^\mu(\tau) \) in curved spacetimes:

\[
a^\mu \equiv \frac{D^2z^\mu}{d\tau^2} \equiv \frac{d^2z^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta} \frac{dz^\alpha}{d\tau} \frac{dz^\beta}{d\tau} = 0; \quad g_{\mu\nu} \left(z(\lambda)\right) \frac{dz^\mu}{d\lambda} \frac{dz^\nu}{d\lambda} > 0; \quad (8.5.4)
\]

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 \( f^\mu \), then the covariant Newton’s second law for a system of mass \( M \) would read: \( MD^2z^\mu/d\tau^2 = f^\mu \).

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. (8.5.1) 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 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. (8.5.1), 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, photons in fact sweep out null geodesics \( z^\mu(\lambda) \) in curved spacetimes:

\[
a^\mu \equiv \frac{D^2z^\mu}{d\lambda^2} = 0, \quad g_{\mu\nu} \left(z(\lambda)\right) \frac{dz^\mu}{d\lambda} \frac{dz^\nu}{d\lambda} = 0. \quad (8.5.5)
\]

**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 \( \vec{x} - \vec{y} \):

\[
g_{00}(\tau, \vec{x}) = 1 - R_{0ab}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}((x - y)^3), \quad (8.5.6)
\]

\[
g_{0i}(\tau, \vec{x}) = -\frac{2}{3} R_{0ab}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}((x - y)^3), \quad (8.5.7)
\]

\[
g_{ij}(\tau, \vec{x}) = \eta_{ij} - \frac{1}{3} R_{aib}(\tau) \cdot (x^a - y^a)(x^b - y^b) + \mathcal{O}((x - y)^3). \quad (8.5.8)
\]
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 $\tau$.) 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.101

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 (8.5.6) through (8.5.8) 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 that in flat Minkowski spacetime up to corrections that go at least as

$$\epsilon_E \equiv \frac{\text{Length or inverse mass scale of system}}{\text{Length scale of the spacetime geometric curvature}}. \quad (8.5.9)$$

In particular, since the Christoffel symbols on the world line vanishes, the geodesic $y^\mu$ itself obeys the free-particle version of Newton’s 2nd law: $d^2 y^\mu / ds^2 = 0$.

More generally, because material bodies (with mass $> 0$) sweep out geodesics according to eq. (8.5.4), 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 8.40.** 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 $y^\alpha$ to conclude $dy^\alpha / d\tau$ are constants. If $\tau$ refers to the proper time of the freely falling observer at $y^\alpha(\tau)$, then explain why

$$\eta_{\alpha\beta} \frac{dy^\alpha}{d\tau} \frac{dy^\beta}{d\tau} = 1. \quad (8.5.10)$$

Since this is a Lorentz invariant condition, $\{y^\alpha\}$ can be Lorentz boosted $y^\alpha \rightarrow \Lambda^\alpha_\mu y^\mu$ to the rest frame such that

$$\frac{dy^\alpha}{ds} \rightarrow \Lambda^\alpha_\mu \frac{dy^\mu}{ds} = \left(1, \vec{0}\right); \quad (8.5.11)$$

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 $\tau$ dependence in equations (8.5.6), (8.5.7) and (8.5.8) occur in the Riemann tensor.101

---

101 The primary difference between eq. (7.2.1) and equations (8.5.6)-(8.5.8), 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.
Problem 8.41. Verify that the coefficients in front of the Riemann tensor in equations (8.5.6), (8.5.7) and (8.5.8) are independent of the spacetime dimension. That is, starting with
\[ g_{00}(x) = 1 - A \cdot R_{0ab}(\tau) \cdot (x - y)^a (x - y)^b + O((x - y)^3), \] (8.5.12)
\[ g_{0i}(x) = -B \cdot R_{0ai}(\tau) \cdot (x - y)^a (x - y)^b + O((x - y)^3), \] (8.5.13)
\[ g_{ij}(x) = \eta_{ij} - C \cdot R_{iajb}(\tau) \cdot (x - y)^a (x - y)^b + O((x - y)^3), \] (8.5.14)
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 (8.5.6), (8.5.7), and (8.5.8) – for that, see Poisson [18] §1.6 – but merely a consistency check.

Fermi versus Riemann Normal Coordinates
The Riemann normal coordinate system \( \{ y^a \} \) version of eq. (7.2.1) but in curved spacetimes reads
\[ g_{\mu\nu}(y \rightarrow y_0) = \eta_{\mu\nu} - \frac{1}{3} R_{\mu\alpha\nu\beta}(y_0) \cdot (y - y_0)^\alpha (y - y_0)^\beta + O((y - y_0)^3). \] (8.5.15)
This is to be contrasted with equations (8.5.6), (8.5.7), and (8.5.8). The latter holds along the entire ‘free-falling’ geodesic; where eq. (8.5.15) only holds in the neighborhood around \( y \approx y_0 \). In particular, the Riemann tensor in eq. (8.5.15) should be viewed as a constant; while the Riemann in equations (8.5.6), (8.5.7), and (8.5.8) is a function of time, since curvature can change along the geodesic.

Problem 8.42. Gravitational force in a weak gravitational field
Consider the following metric:
\[ g_{\mu\nu}(t, \vec{x}) = \eta_{\mu\nu} + 2\Phi(\vec{x})\delta_{\mu\nu}, \] (8.5.16)
where \( \Phi(\vec{x}) \) is time-independent. Assume this is a weak gravitational field, in that \( |\Phi| \ll 1 \) everywhere in spacetime, and there are no non-gravitational forces. (Linearized General Relativity reduces to the familiar Poisson equation \( \vec{\nabla}^2 \Phi = 4\pi G N \rho \), where \( \rho(\vec{x}) \) is the mass/energy density of matter.) Starting from the non-affine form of the action principle
\[ -M s = -M \int_{t_1}^{t_2} dt \sqrt{g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu}, \quad \dot{z}^\mu \equiv \frac{dz^\mu}{d\ell} \]
\[ = -M \int_{t_1}^{t_2} dt \sqrt{1 - \vec{v}^2 + 2\Phi(1 + \vec{v}^2)}, \quad \vec{v}^2 \equiv \delta_{ij} \dot{z}^i \dot{z}^j; \] (8.5.17)
expand this action to lowest order in \( \vec{v}^2 \) and \( \Phi \) and work out the geodesic equation of a ‘test mass’ \( M \) sweeping out some worldline \( z^\mu \) in such a spacetime. (You should find something very familiar from Classical Mechanics.) Show that, in this non-relativistic limit, Newton’s law of gravitation is recovered:
\[ \frac{d^2 z^i}{dt^2} = -\partial_i \Phi. \] (8.5.18)
We see that, in the weakly curved spacetime of eq. (8.5.16), \( \Phi \) may indeed be identified as the Newtonian potential.
Geodesic Deviation & 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 \( \lambda \), so that we may connect one geodesic to the other at the same \( \lambda \) via an infinitesimal vector \( \xi^\mu \). We will denote the tangent vector to one of geodesics to be \( U^\mu \), such that

\[
U^\sigma \nabla_\sigma U^\mu = 0. \tag{8.5.19}
\]

Furthermore, we will assume that \([U, \xi] = 0\), i.e., \( U \) and \( \xi \) 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

\[
U^\alpha \nabla_\alpha \left( U^\beta \nabla_\beta \xi^\mu \right) = \nabla_U \nabla_U \xi^\mu = -R^\mu_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta. \tag{8.5.20}
\]

As its name suggests, this equation tells us how the deviation vector \( \xi^\mu \) 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 \( U^\mu \) is a timelike vector tangent to, say, the geodesic trajectory of its center-of-mass, the geodesic deviation equation (8.5.20) 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. 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. (8.5.20)**  
We start by noting \([\xi, U] = (\xi^\alpha \partial_\alpha U^\mu - U^\alpha \partial_\alpha \xi^\mu) \partial_\mu = 0\) translates to

\[
\nabla_\xi U = \nabla_U \xi; \tag{8.5.21}
\]

because \( \nabla_\xi U^\mu = \xi^\alpha \partial_\alpha U^\mu + \Gamma^\mu_{\sigma\kappa} \xi^\sigma U^\kappa \) and \( \nabla_U \xi^\mu = U^\alpha \partial_\alpha \xi^\mu + \Gamma^\mu_{\sigma\kappa} \xi^\sigma U^\kappa \); 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 \( \nabla_\xi \) upon it.

\[
\nabla_\xi \nabla_U U^\mu = 0 \tag{8.5.22}
\]

\[
\nabla_U \nabla_\xi U^\mu + [\nabla_\xi, \nabla_U] U^\mu = 0 \tag{8.5.23}
\]

\[
\nabla_U \nabla_U \xi^\mu = -R^\mu_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta \tag{8.5.24}
\]

On the last line, we have exploited the assumption that \([U, \xi] = 0\) to say \([\nabla_\xi, \nabla_U] U^\mu = R^\mu_{\nu\alpha\beta} U^\nu \xi^\alpha U^\beta \) – recall eq. (8.4.36).

\[\text{102}\] 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 8.43. Alternate Derivation of Geodesic Deviation Equation  A less geometric but equally valid manner to derive eq. (8.5.20) 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

\[
\frac{d^2(y^\alpha(\tau) + \xi^\alpha(\tau))}{d\tau^2} + \Gamma^\alpha_{\alpha\beta}(y + \xi) \frac{dy^\beta(\tau)}{d\tau} \frac{dy^\beta(\tau)}{d\tau} = 0. \tag{8.5.25}
\]

If the components \( \xi^\mu \) may be considered ‘small,’ expand the above up to linear order in \( \xi^\mu \) and show that

\[
\frac{d^2\xi^\mu}{d\tau^2} + 2\Gamma^\mu_{\alpha\beta}(y) \frac{d\xi^\alpha}{d\tau} \frac{dy^\beta}{d\tau} + \xi^\sigma \partial_{\sigma} \Gamma^\mu_{\alpha\beta}(y) \frac{dy^\alpha}{d\tau} \frac{dy^\beta}{d\tau} = 0. \tag{8.5.26}
\]

Now proceed to demonstrate that equations (8.5.20) and (8.5.26) are equivalent.

Problem 8.44. Geodesic Deviation & FNC  Argue that all the Christoffel symbols \( \Gamma^\alpha_{\mu\nu} \) evaluated along the free-falling geodesic in equations (8.5.6)-(8.5.8), namely when \( x = y \), vanish. Then argue that all the time derivatives of the Christoffel symbols vanish along \( y \) too: \( \partial^\tau \Gamma^\alpha_{\mu\nu} = 0 \). (Hints: Recall from Problem (7.14) that, specifying the first derivatives of the metric is equivalent to specifying the Christoffel symbols. Why is \( \partial^\tau g^\alpha\beta(x = y) = 0 \)? Why is \( \partial^\tau \partial_i g^\alpha\beta(x = y) = 0 \)? Why does this imply, denoting \( U^\mu \equiv \frac{dy^\mu}{d\tau}, \) the geodesic equation

\[
U^\nu \nabla_\nu U^\mu = \frac{dU^\mu}{d\tau} = 0? \tag{8.5.27}
\]

Next, evaluate the geodesic deviation equation in these Fermi Normal Coordinates (FNC) system. Specifically, show that

\[
U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^\mu = \frac{d^2\xi^\mu}{d\tau^2} = -R^\mu_{\alpha0\beta} \xi^\nu. \tag{8.5.28}
\]

Why does this imply, if the deviation vector is purely spatial at a given \( s = s_0 \), specifically \( \xi^\nu(\tau_0) = 0 = d\xi^\nu/d\tau_0 \), then it remains so for all time? (Hint: In an FNC system and on the world line of the free-falling observer, \( R^\nu_{\alpha0\beta} = R^\nu_{0\alpha\beta} \). What do the (anti)symmetries of the Riemann tensor say about the right hand side?)

Problem 8.45. A Common Error  Eq. (8.5.28) says that the acceleration of the deviation vector within the FNC system is simply the ordinary one: i.e.,

\[
U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^\mu = \frac{d^2\xi^\mu}{d\tau^2}. \tag{8.5.29}
\]

Thus, eq. (8.5.28) 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. (8.5.29) 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,

\[
g_{\mu\nu}dx^\mu dx^\nu = d\tau^2 - (\delta_{ij} - h_{ij}(\tau, \vec{x})) dx^i dx^j, \quad |h_{ij}| \ll 1, \quad U^\mu = \delta^\mu_0. \tag{8.5.30}
\]

Show that eq. (8.5.29) is no longer true; but up to first order in \( h_{ij} \) it reads instead

\[
U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^\mu = \ddot{\xi}^\mu + \eta^{\mu j} \dot{\xi}^k h_{jk} + \frac{1}{2} \eta^{\mu j} \xi^k \dddot{h}_{jk} + O(h^2), \tag{8.5.31}
\]

where all the overdot(s) are partial derivative(s) with respect to proper time \( \tau \).

**Problem 8.46. 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 \( r_\epsilon \), 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 \( \rho(\vec{x}) \).

Assume that the weak field metric generated by \( \rho \) is given by eq. (8.5.16); it is possible to justify this statement by using the linearized Einstein’s equations. Show that the vector field

\[
U^\mu(t, \vec{x}) \equiv \delta^\mu_0 (1 - \Phi(\vec{x})) - t \delta^\mu_i \partial_i \Phi(\vec{x}) \tag{8.5.32}
\]

is a timelike geodesic up to linear order in the Newtonian potential \( \Phi \). This \( U^\mu \) 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 \( U^\alpha \) to show that \( \dot{} \) up to first order in \( \Phi \) – the right hand sides of its geodesic deviation equations are

\[
U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^0 = 0, \tag{8.5.33}
\]

\[
U^\alpha U^\beta \nabla_\alpha \nabla_\beta \xi^i = R_{i0j0} \xi^j; \tag{8.5.34}
\]

where the linearized Riemann tensor reads

\[
R_{i0j0} = -\partial_i \partial_j \Phi(\vec{x}). \tag{8.5.35}
\]

Assuming that the monopole contribution dominates,

\[
\Phi(\vec{x}) \approx \Phi(r) = -\frac{G_N M}{r} = -\frac{r_s}{2r}, \tag{8.5.36}
\]

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

\[
R_{i0j0} \approx -\left( \delta^{ij} - \vec{\eta}^i \vec{\eta}^j \right) \frac{\Phi'(r)}{r} - \vec{\eta}^i \vec{\eta}^j \Phi''(r), \quad \vec{\eta}^i \equiv \frac{x^i}{r}, \tag{8.5.37}
\]
so that the result follows simply from counting the powers of $1/r$ from $\Phi'(r)/r$ and $\Phi''(r)$. By setting $\xi$ to be (anti-)parallel and perpendicular to the radial direction $\hat{r}$, argue that the test masses lying on the radial line emanating from the body centered at $x = \vec{0}$ will be stretched apart while the test masses lying on the plane perpendicular to $\hat{r}$ will be squeezed together. (Hint: You should be able to see that $\delta^{ij} - \hat{r}^i \hat{r}^j$ is the Euclidean space orthogonal to $\hat{r}$.)

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.

**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. (8.4.43)? One such geometric meaning can be found within the Raychaudhuri equation, which describes the rate of expansion or contraction of a bundle of integral curves; see Poisson [18] for a discussion. Another (related) perspective is its relation to the local volume of spacetime relative that of Minkowski. For, we may identify in equations (8.5.15),

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}; \quad (8.5.38)$$

where

$$h_{\mu\nu}(y) = -\frac{1}{3} R_{\mu\alpha\nu\beta}(y_0) \cdot (y - y_0)^{\alpha}(y - y_0)^{\beta} + \mathcal{O} \left((y - y_0)^3\right). \quad (8.5.39)$$

This in turn implies, from eq. (8.4.43),

$$h(y) \equiv \eta^{\alpha\beta}(y) h_{\alpha\beta}(y) = -\frac{1}{3} R_{\alpha\beta}(y_0) \cdot (y - y_0)^{\alpha}(y - y_0)^{\beta} + \mathcal{O} \left((y - y_0)^3\right). \quad (8.5.40)$$

At this point, we may invoke the spacetime version of the discussion leading up to eq. (7.3.91),

to deduce the infinitesimal spacetime volume element around $y = y_0$ is given by

$$d^d y \sqrt{|g(y \approx y_0)|} = d^d y \left(1 - \frac{1}{6} R_{\alpha\beta}(y_0) \cdot (y - y_0)^{\alpha}(y - y_0)^{\beta} + \mathcal{O} \left((y - y_0)^3\right)\right). \quad (8.5.41)$$

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 $y_0$.

**Interlude** Let us pause to summarize the physics we have revealed thus far.

In a curved spacetime, the collective motion of a system of mass $M$ sweeps out a timelike geodesic – recall equations (8.4.57), (8.4.75), and (8.4.81) – 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 $1/\max R_{\alpha\beta\mu\nu}^{1/2}$. 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. (8.5.28).

---

A shorter version of this discussion may be found on Wikipedia. A closely related explanation of the meaning of Einstein’s equation (8.4.53) using the Raychaudhuri equation may be found in Baez and Bunn [26].
**Killing Vectors**  
A geometry is said to enjoy an isometry – or, symmetry – when we perform the following infinitesimal displacement

\[ x^\mu \rightarrow x^\mu + \xi^\mu(x) \]  
(8.5.42)

and find that the geometry is unchanged

\[ g_{\mu\nu}(x) \rightarrow g_{\mu\nu}(x) + O(\xi^2). \]  
(8.5.43)

Generically, under the infinitesimal transformation of eq. (8.5.42),

\[ g_{\mu\nu}(x) \rightarrow g_{\mu\nu}(x) + \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu. \]  
(8.5.44)

where

\[ \nabla_{(\mu} \xi_{\nu)} = \xi^\sigma \partial_\sigma g_{\mu\nu} + g_{\sigma(\mu} \partial_{\nu)} \xi^\sigma. \]  
(8.5.45)

If an isometry exists along the integral curve of \( \xi^\mu \), it has to obey Killing’s equation – recall equations (7.2.46) and (7.2.47) –

\[ \nabla_{(\mu} \xi_{\nu)} = \xi^\sigma \partial_\sigma g_{\mu\nu} + \partial_{(\mu} \xi^\sigma g_{\nu)} = 0. \]  
(8.5.46)

In fact, by exponentiating the infinitesimal coordinate transformation, it is possible to show that – if \( \xi^\mu \) is a Killing vector (i.e., it satisfies eq. (8.5.46)), then an isometry exists along its integral curve. In other words,

A spacetime geometry enjoys an isometry (aka symmetry) along the integral curve of \( \xi^\mu \) 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 [20] for a discussion.)

**Problem 8.47. Conserved quantities along geodesics**  
(I of II)  
If \( p_\mu \) denotes the ‘momentum’ variable of a geodesic

\[ p_\mu \equiv \frac{\partial L_g}{\partial \dot{z}^\mu}, \]  
(8.5.47)

where \( L_g \) is defined in eq. (8.4.60), and if \( \xi^\mu \) is a Killing vector of the same geometry \( \nabla_{(\alpha} \xi_{\beta)} = 0 \), show that

\[ \xi^\mu(z(\lambda)) p_\mu(\lambda) \]  
(8.5.48)

is a constant along the geodesic \( z^\mu(\lambda) \). Hints: If you perturb the coordinates by the Killing vector \( \xi^\mu \), namely \( x^\mu \rightarrow x^\mu + \xi^\mu \), then you should be able to obtain – to first order in \( \xi \) –

\[ \dot{z}^\mu \rightarrow \dot{z}^\mu + \dot{z}^\sigma \partial_\sigma \xi^\mu = \frac{d}{d\lambda} \left( z^\mu(\lambda) + \xi^\mu(z(\lambda)) \right), \]  
(8.5.49)

\[ L_g \rightarrow L_g; \]  
(8.5.50)

i.e., the Lagrangian is invariant if you recall eq. (8.5.46). On the other hand, varying the Lagrangian to first order yields

\[ \delta L_g = \frac{\partial L_g}{\partial \dot{z}^\sigma} \dot{\xi}^\sigma + \frac{\partial L_g}{\partial z^\sigma} \xi^\sigma + O(\xi^2). \]  
(8.5.51)

(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 \( \xi^\mu \) is a Killing vector, then \( \xi_\nu v^\nu \) is conserved along the integral curve of \( v^\mu \).

Can you demonstrate the validity of this statement?

**Second Derivatives of Killing Vectors**

Now let us also consider the second derivatives of \( \xi^\mu \). In particular, we will now explain why

\[
\nabla_\alpha \nabla_\beta \xi_\delta = R^\lambda_{\alpha \beta \delta} \xi_\lambda. 
\]

Consider

\[
0 = \nabla_\delta \nabla_{\{\alpha} \xi_{\beta\}} \tag{8.5.53}
= [\nabla_\delta, \nabla_\alpha] \xi_\beta + \nabla_\alpha \nabla_\delta \xi_\beta + [\nabla_\delta, \nabla_\beta] \xi_\alpha + \nabla_\beta \nabla_\delta \xi_\alpha \tag{8.5.54}
= -R^\lambda_{\beta \delta \alpha} \xi_\lambda - \nabla_\alpha \nabla_\beta \xi_\delta - R^\lambda_{\alpha \delta \beta} \xi_\lambda - \nabla_\beta \nabla_\alpha \xi_\delta \tag{8.5.55}
\]

Because Bianchi says \( 0 = R^\lambda_{[\alpha \beta \delta]} \Rightarrow R^\lambda_{\alpha \beta \delta} = R^\lambda_{\beta \alpha \delta} + R^\lambda_{\delta \beta \alpha} \).

\[
0 = -R^\lambda_{\beta \delta \alpha} \xi_\lambda - \nabla_\alpha \nabla_\beta \xi_\delta + (R^\lambda_{\beta \alpha \delta} + R^\lambda_{\delta \beta \alpha}) \xi_\lambda - \nabla_\beta \nabla_\alpha \xi_\delta \tag{8.5.56}
0 = -2R^\lambda_{\beta \delta \alpha} \xi_\lambda - \nabla_{\{\beta} \nabla_\alpha\} \xi_\delta - [\nabla_\beta, \nabla_\alpha] \xi_\delta \tag{8.5.57}
0 = -2R^\lambda_{\beta \delta \alpha} \xi_\lambda - 2\nabla_\beta \nabla_\alpha \xi_\delta \tag{8.5.58}
\]

This proves eq. (8.5.52).

**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

\[
\nabla_{\{\alpha} \xi_{\beta\}} = \nabla_{\{\alpha} [U, V]_{\beta\}} = 0. 
\]

More explicitly, let us compute:

\[
\nabla_\alpha (U^\mu \nabla_\mu V_\beta - V^\mu \nabla_\mu U_\beta) + (\alpha \leftrightarrow \beta) \\
= \nabla_\alpha U^\mu \nabla_\mu V_\beta - \nabla_\alpha V^\mu \nabla_\mu U_\beta + U^\mu \nabla_\alpha \nabla_\mu V_\beta - V^\mu \nabla_\alpha \nabla_\mu U_\beta + (\alpha \leftrightarrow \beta) \\
= -\nabla_\mu U_\alpha \nabla^\mu V_\beta + \nabla_\mu V_\alpha \nabla^\mu U_\beta + U^\mu \nabla_{[\alpha} \nabla_\mu] V_\beta + U^\mu \nabla_\alpha \nabla_\mu V_\beta - V^\mu \nabla_{[\alpha} \nabla_\mu] U_\beta - V^\mu \nabla_\alpha \nabla_\mu U_\beta + (\alpha \leftrightarrow \beta) \\
= -U^\mu R^\sigma_{\beta \alpha \mu} V_\sigma + V^\mu R^\sigma_{\beta \alpha \mu} U_\sigma + (\alpha \leftrightarrow \beta) \\
= -U^\mu [V^\sigma] R^\sigma_{\beta \alpha \mu} + (\alpha \leftrightarrow \beta)
\]

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 8.48. Killing Vectors in Minkowski**

In Minkowski spacetime \( g_{\mu \nu} = \eta_{\mu \nu} \), with Cartesian coordinates \(\{x^\mu\}\), use eq. (8.5.52) to argue that the most general Killing vector takes the form

\[
\xi_\mu = \ell_\mu + \omega_{\mu \nu} x^\nu, 
\]

(8.5.60)
for constant $\ell_\mu$ and $\omega_{\mu\nu}$. (Hint: Think about Taylor expansions; use eq. (8.5.52) to show that the 2nd and higher partial derivatives of $\xi_\delta$ are zero.) Then use the Killing equation (8.5.46) to infer that

$$\omega_{\mu\nu} = -\omega_{\nu\mu}. \tag{8.5.61}$$

The $\ell_\mu$ 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:

$$\partial_\mu \quad (\text{Generators of spacetime translations}) \tag{8.5.62}$$

and

$$x^{[\mu} \partial^{\nu]} \quad (\text{Generators of Lorentz boosts or rotations}). \tag{8.5.63}$$

There are $d$ distinct $\partial_\mu$’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 $\theta$, the 2D rotation matrix in eq. (8.1.59) reads

$$\hat{R}^i_j(\theta) = \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix} + \mathcal{O}(\theta^2). \tag{8.5.64}$$

In other words, $\hat{R}^i_j(\theta) = \delta_{ij} - \theta \epsilon_{ij}$, where $\epsilon_{ij}$ 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 $f$,

$$f(x^i) \to f(\hat{R}^i_j x^j) = f \left( x^i - \theta \epsilon_{ij} x^j \right) + \mathcal{O}(\theta^2)$$

$$= f(x) - \theta \epsilon_{ij} x^j \partial_i f(x) + \mathcal{O}(\theta^2). \tag{8.5.65}$$

Since $\theta$ is arbitrary, the basic differential operator that implements an infinitesimal rotation of the coordinate system on any Minkowski scalar is

$$-\epsilon_{ij} x^j \partial_i = x^1 \partial_2 - x^2 \partial_1. \tag{8.5.66}$$

This is the 2D version of eq. (8.5.63) for rotations. As for 2D Lorentz boosts, eq. (8.1.58) tells us

$$\Lambda^{\mu}_\nu(\xi) = \begin{bmatrix} 1 & \xi \\ \xi & 1 \end{bmatrix} + \mathcal{O}(\xi^2). \tag{8.5.67}$$

(This $\xi$ 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

$$f(x^\mu) \to f \left( \Lambda^{\mu}_\nu x^\nu \right) = f \left( x^\mu + \xi \epsilon^{\mu}_\nu x^\nu + \mathcal{O}(\xi^2) \right) \tag{8.5.68}$$
\[ f(x) - \xi \epsilon_{\nu \mu} x^{\nu} \partial^{\mu} f(x) + O(\xi^2). \] (8.5.70)

Since \( \xi \) is arbitrary, to implement a Lorentz boost of the coordinate system on any Minkowski scalar, the appropriate differential operator is

\[ \epsilon_{\mu \nu} x^\mu \partial^\nu = x^0 \partial^1 - x^1 \partial^0; \] (8.5.71)

which again is encoded within eq. (8.5.63).

**Problem 8.49.** Verify that Lie Algebra of \( \text{SO}_{D,1} \) in (8.1.122) is recovered if we exploit eq. (8.5.63) to define

\[ J^{\mu \nu} = i(x^\mu \partial^\nu - x^\nu \partial^\mu), \] (8.5.72)

where \( \partial^\mu \equiv \eta^{\mu \nu} \partial_\nu \). This tells us, under a Lorentz boost or rotation \( f(x) \to \exp\left(-i/2\omega_{\mu \nu} J^{\mu \nu}\right)f(x). \)

**Problem 8.50. 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-Lemaitre-Robertson-Walker (FLRW) metric

\[ ds^2 = dt^2 - a(t)^2 d\vec{x} \cdot d\vec{x}; \] (8.5.73)

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., \( d\vec{x} = 0 \). 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 8.51. Killing identities involving Ricci**

Prove the following results. If \( \xi^\mu \) is a Killing vector and \( R_{\alpha \beta} \) and \( R \) are the Ricci tensor and scalar respectively, then

\[ \xi^\alpha \nabla^\beta R_{\alpha \beta} = 0 \quad \text{and} \quad \xi^\alpha \nabla_\alpha R = 0. \] (8.5.74)

Hints: First use eq. (8.5.52) to show that

\[ \Box \xi_\delta = -R^\lambda_\beta \xi_\lambda; \] (8.5.75)

\[ \Box \equiv g^{\alpha \beta} \nabla_\alpha \nabla_\beta = \nabla_\alpha \nabla^\alpha. \] (8.5.76)

Then take the divergence on both sides, and commute the covariant derivatives until you obtain the term \( \Box \nabla^\beta \xi_\delta \) — what is \( \nabla^\beta \xi_\delta \) equal to? Argue why \( \xi^\alpha \nabla^\beta R_{\alpha \beta} = \nabla^\beta(\xi^\alpha R_{\alpha \beta}) \). You may also need to employ the Einstein tensor Bianchi identity \( \nabla^\mu G_{\mu \nu} = 0 \) to infer that \( \xi^\alpha \nabla_\alpha R = 0. \)

**Problem 8.52. In \( d \) spacetime dimensions, show that**

\[ \partial_{[\alpha_1} \left( J^{\mu} \epsilon_{\alpha_2 \ldots \alpha_d] \mu} \right) \] (8.5.77)
is proportional to $\nabla \sigma J^\sigma$. What is the proportionality factor? (This discussion provides a differential forms based language to write $d^d x \sqrt{|g|} \nabla \sigma J^\sigma$.) If $\nabla \sigma J^\sigma = 0$, what does the Poincaré lemma tell us about eq. (8.5.77)? Find the dual of your result and argue there must an antisymmetric tensor $\Sigma^{\mu \nu}$ such that

$$J^\mu = \nabla_\nu \Sigma^{\mu \nu}.$$  \hspace{1cm} (8.5.78)

Hint: For the first step, explain why eq. (8.5.77) is proportional to the Levi-Civita symbol $\epsilon_{\alpha_1 ... \alpha_d}$.

### Problem 8.53. 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)

$$g_{\mu \nu} = \eta_{\mu \nu} + 2\Phi \delta_{\mu \nu} = \eta_{\mu \nu} - \frac{r_s}{r} \delta_{\mu \nu},$$  \hspace{1cm} (8.5.79)

where we assume $|\Phi| = r_s/r \ll 1$ and

$$r_s \equiv 2G_NM.$$  \hspace{1cm} (8.5.80)

Now, the metric of an isolated static non-rotating black hole – i.e., the Schwarzschild black hole – in isotropic coordinates is

$$ds^2 = \left(1 - \frac{r_s}{2r} \right)^2 dt^2 - \left(1 + \frac{r_s}{4r} \right)^4 \mathbf{d}\mathbf{x} \cdot \mathbf{d}\mathbf{x}, \quad r \equiv \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$  \hspace{1cm} (8.5.81)

The $r_s \equiv 2G_NM$ 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. (8.5.81) up to first order $r_s/r$ and verify this yields eq. (8.5.79). We may therefore identify eq. (8.5.79) as either the metric due to the monopole moment of some static mass density $\rho(x)$ 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 $\Delta \phi$ it suffers due to the presence of a mass monopole. Express the answer $\Delta \phi$ 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. (8.5.79) 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. (8.5.79) in spherical coordinates

$$\mathbf{x} = r \left( \sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta) \right).$$  \hspace{1cm} (8.5.82)

**Spherical Symmetry and $\theta$** Because of the spherical symmetry of the problem, we may always assume that all geodesic motion takes place on the equatorial plane:

$$\theta = \frac{\pi}{2}.$$  \hspace{1cm} (8.5.83)
If you wish to provide a more robust argument, verify that the \( \theta \) variable obeys the following geodesic equation:

\[
2 \dot{\theta} \dot{r} + r \left( \ddot{\theta} - \sin(\theta) \cos(\theta) \dot{\phi}^2 \right) = 0; \quad (8.5.84)
\]

where all overdots are derivatives with respect to \( \lambda \). Notice \( \theta = \pi/2 \Rightarrow \dot{\theta} = 0 = \ddot{\theta} \) satisfies eq. (8.5.84).

\textit{Energy’ Conservation} \quad \text{Proceed to use the} \ t—\text{independence of the metric, together with the invariance of the null geodesic Lagrangian} \ L_g \ \text{under constant re-scaling of its affine parameter} \ \lambda, \ \text{to argue that} \ \lambda \ \text{itself can always be chosen such that}

\[
\dot{t} = \left( 1 - \frac{r_s}{r} \right)^{-1}. \quad (8.5.85)
\]

\textit{Angular Momentum conservation} \quad \text{Next, use the} \ \phi—\text{independence of the metric to show that angular momentum conservation} \ -\partial L_g/\partial \dot{\phi} \equiv \ell \ \text{(constant) yields}

\[
\dot{\phi} = \frac{\ell}{r^2} \left( 1 + \frac{r_s}{r} \right)^{-1}. \quad (8.5.86)
\]

We are primarily interested in the trajectory as a function of angle, so we may eliminate all \( \dot{r} \equiv \frac{dr}{d\lambda} \) as

\[
\dot{r} = \frac{d\phi}{d\lambda} r^\prime(\phi) = \frac{\ell}{r^2} \left( 1 + \frac{r_s}{r} \right)^{-1} r^\prime(\phi), \quad (8.5.87)
\]

where eq. (8.5.86) was employed in the second equality. At this point, by utilizing equations (8.5.83), (8.5.85), (8.5.86) and (8.5.87), verify that the geodesic Lagrangian now takes the form

\[
L_g = \frac{1}{2} \left( \frac{r}{r - r_s} - \frac{\ell^2}{r^2(1 + \frac{r_s}{r})} \left( 1 + \left( \frac{r^\prime(\phi)}{r} \right)^2 \right) \right). \quad (8.5.88)
\]

\textit{Closest approach vs angular momentum} \quad \text{If} \ r_0 \text{ is the coordinate radius of closest approach, which we shall assume is appreciably larger than the Schwarzschild radius} \ r_0 \gg r_s, \ \text{that means} \ r^\prime(\phi) = 0 \ \text{when} \ r = r_0. \ \text{Show that}

\[
\ell = r_0 \sqrt{\frac{r_0 + r_s}{r_0 - r_s}}. \quad (8.5.89)
\]

\textit{An ODE} \quad \text{Since null geodesics render} \ L_g = 0, \ \text{utilize eq. (8.5.89)} \ \text{in eq. (8.5.88), and proceed to show that} \ -\ \text{to first order in} \ r_s -

\[
\frac{d\phi}{dr} = \frac{1}{\sqrt{r^2 - r_0^2}} \left( \frac{r_0}{r} + \frac{r_s}{r + r_0} \right) + \mathcal{O}(r_s^2). \quad (8.5.90)
\]

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

\[
\Delta \phi = \frac{2r_s}{r_0}. \quad (8.5.91)
\]
Note that, if the photon were undeflected, the total change in angle \((\int_{r_0}^{\infty} dr + \int_{r_0}^{\infty} dr)(d\phi/dr)\) would be \(\pi\). Therefore, the total deflection angle is

\[
\Delta \phi = 2 \left| \int_{r_\infty}^{r_0} \frac{d\phi}{dr} dr \right| - \pi.
\] (8.5.92)

**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 I** For further help on this problem, consult §8.5 of Weinberg [20].

**Remark II** The geometry of eq. (8.3.49) is in fact the same as that in eq. (8.5.81). More specifically,

\[
ds^2 = \left(1 - \frac{r_s}{r'} \right) dt^2 - \frac{dr'^2}{1 - r_s/r'} - r'^2 \left(d\theta^2 + \sin(\theta)^2 d\phi^2\right)
\]

\[
= \left(1 - \frac{r_s}{4r'} \right)^2 dt^2 - \left(1 + \frac{r_s}{4r'} \right)^4 r'^2 \left(d\theta^2 + \sin(\theta)^2 d\phi^2\right);
\] (8.5.93)

where the \(d\vec{x} \cdot d\vec{x}\) in eq. (8.5.81) has been converted into the equivalent expression in spherical coordinates. You may verify, identifying the coordinate transformation rule \(r'^2 = (1 + r_s/(4r))^4 r^2\) brings one from the first line to the second, or vice versa.

**Problem 8.54. Gauge-covariant derivative**

Let \(\psi\) be a vector under group transformations. By this we mean that, if \(\psi^a\) corresponds to the \(a\)th component of \(\psi\), then given some matrix \(U^a_b\), \(\psi\) transforms as

\[
\psi'^a = U^a_b \psi^b \quad \text{or, } \psi' = U \psi.
\] (8.5.95)

Compare eq. (8.5.95) to how a spacetime vector transforms under coordinate transformations:

\[
V^\mu(x') = \mathcal{J}^\mu_\sigma V^\sigma(x), \quad \mathcal{J}^\mu_\sigma \equiv \frac{\partial x'^\mu}{\partial x^\sigma}.
\] (8.5.96)

Now, let us consider taking the gauge-covariant derivative \(\tilde{D}\) of \(\psi\) such that it still transforms 'covariantly' under group transformations, namely

\[
\tilde{D}_a \psi' = \tilde{D}_a (U \psi) = U (\tilde{D}_a \psi).
\] (8.5.97)

Crucially:

We shall now demand that the gauge-covariant derivative transforms covariantly – i.e., eq. (8.5.97) 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.,

\[
\nabla_a \psi' \neq \tilde{D}_a \psi'.
\] (8.5.98)
by showing that eq. (8.5.97) 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:

\[
(D_\mu)^{\bar{a}}_{\bar{b}} \equiv \delta^{\bar{a}}_{\bar{b}} \nabla_\mu + (A_\mu)^{\bar{a}}_{\bar{b}}. \tag{8.5.99}
\]

Or, in gauge-index-free notation,

\[
\hat{D}_\mu \equiv \nabla_\mu + A_\mu. \tag{8.5.100}
\]

With the definition in eq. (8.5.99), how must the gauge potential \( A_\mu \) (or, equivalently, \((A_\mu)^{\bar{a}}_{\bar{b}}\)) transform so that eq. (8.5.97) is satisfied? Compare the answer to the transformation properties of the Christoffel symbol in eq. (8.4.6). (Since the answer can be found in most Quantum Field Theory textbooks, make sure you verify the covariance explicitly!)

**Bonus:** Here, we have treated \( \psi \) as a spacetime scalar and the gauge-covariant derivative \( \hat{D}_\alpha \) itself as a scalar under group transformations. Can you generalize the analysis here to the higher-rank tensor case?

### 8.6 Special Topic: 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:

\[
g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{8.6.1}
\]

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.

**Metric inverse, Determinant** Whenever performing a perturbative analysis, we shall agree to move all tensor indices – including that of \( h_{\mu\nu} \) – with the \( \bar{g}_{\alpha\beta} \). For example,

\[
h^{\alpha\beta} \equiv \bar{g}^{\alpha\sigma} h_{\sigma\beta}, \quad \text{and} \quad h^{\alpha\beta} \equiv \bar{g}^{\alpha\sigma} \bar{g}^{\beta\rho} h_{\sigma\rho}. \tag{8.6.2}
\]

With this convention in place, let us note that the inverse metric is a geometric series. Firstly,

\[
g_{\mu\nu} = \bar{g}_{\mu\sigma} (\delta^\sigma_\nu + h^\sigma_\nu) \equiv \bar{g} \cdot (\mathbb{I} + h). \tag{8.6.3}
\]

(Here, \( h \) is a matrix, whose \( \mu \)th row and \( \nu \)th 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

\[
g^{-1} = (\mathbb{I} + h)^{-1} \cdot \bar{g}^{-1}. \tag{8.6.4}
\]

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.

\[
g^{\mu\nu} = \left( \delta^\mu_\lambda + \sum_{\ell=1}^{\infty} (-)^\ell h^\mu_\sigma_1 h^\sigma_1_\sigma_2 \ldots h^\sigma_\ell-2_\sigma_\ell-1 h^\sigma_\ell-1_\lambda \right) \bar{g}^{\lambda\nu} \tag{8.6.5}
\]
\[ g^{\mu\nu} + \sum_{\ell=1}^{\infty} (-)^{\ell} h^{\mu}_{\sigma_1} h^{\sigma_1}_{\sigma_2} \cdots h^{\sigma_{\ell-2}}_{\sigma_{\ell-1}} h^{\sigma_{\ell-1\nu}} \]  
(8.6.6)

\[ g^{\mu\nu} - h^{\mu}_{\sigma_1} h^{\sigma_1}_{\nu} - h^{\nu}_{\sigma_1} h^{\sigma_1}_{\mu} h^{\sigma_2\nu} + \ldots. \]  
(8.6.7)

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 \( A \),

\[ \det A = \exp \left[ \text{Tr} [\ln A] \right], \]  
(8.6.8)

where \( \text{Tr} \) is the matrix trace; for e.g., \( \text{Tr} [h] = h^\sigma_{\sigma} \). Taking the determinant of both sides of eq. (8.6.3), and using the property \( \det [A \cdot B] = \det A \cdot \det B \),

\[ \det g_{\alpha\beta} = \det \bar{g}_{\alpha\beta} \cdot \det [I + h], \]  
(8.6.9)

so that eq. (8.6.8) can be employed to state

\[ \sqrt{|g|} = \sqrt{|\bar{g}|} \cdot \exp \left[ \frac{1}{2} \text{Tr} [\ln [I + h]] \right]. \]  
(8.6.10)

The first few terms read

\[ \sqrt{|g|} = \sqrt{|\bar{g}|} \left( 1 + \frac{1}{2} h + \frac{1}{8} h^2 - \frac{1}{4} h^\sigma_{\rho} h_{\sigma\rho} \right. \]
\[ \left. + \frac{1}{48} h^3 - \frac{1}{8} h \cdot h^\sigma_{\rho} h_{\sigma\rho} + \frac{1}{6} h^\sigma_{\rho} h_{\rho\kappa} h^\kappa_{\sigma} + O[h^4] \right) \]  
(8.6.11)

\( h \equiv h^\sigma_{\sigma} \).  
(8.6.12)

**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

\[ g_{\alpha'\beta'}(x') = (\bar{g}_{\mu\nu}(x(x')) + h_{\mu\nu}(x(x'))) \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta}. \]  
(8.6.13)

If we define the ‘background metric’ to transform covariantly; namely

\[ \bar{g}_{\alpha'\beta'}(x') \equiv \bar{g}_{\mu\nu}(x(x')) \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta}; \]  
(8.6.14)

then, from eq. (8.6.13), the perturbation itself can be treated as a tensor

\[ h_{\alpha'\beta'}(x') = h_{\mu\nu}(x(x')) \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta}. \]  
(8.6.15)

These will now guide us to construct the geometric tensors – the full Riemann tensor, Ricci tensor and Ricci scalar – using the covariant derivative \( \nabla \) 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. (8.6.1):

\[ \nabla_\alpha g_{\mu\nu} = \nabla_\alpha (\bar{g}_{\mu\nu} + h_{\mu\nu}) = \nabla_\alpha h_{\mu\nu}. \]  
(8.6.16)
On the other hand, the usual rules of covariant differentiation tell us

$$\nabla_{\alpha}g_{\mu\nu} = \partial_{\alpha}g_{\mu\nu} - \Gamma^\sigma_{\alpha\mu}g_{\sigma\nu} - \Gamma^\sigma_{\alpha\nu}g_{\mu\sigma}; \quad (8.6.17)$$

where the Christoffel symbols here are built out of the ‘background metric’,

$$\Gamma^\sigma_{\alpha\mu} = \frac{1}{2} \bar{g}^{\sigma\lambda}(\partial_{\alpha}\bar{g}_{\mu\lambda} + \partial_{\mu}\bar{g}_{\alpha\lambda} - \partial_{\lambda}\bar{g}_{\mu\alpha}). \quad (8.6.18)$$

**Problem 8.55. Relation between ‘background’ and ‘full’ Christoffel**

Show that equations (8.6.16) and (8.6.17) can be used to deduce that the full Christoffel symbol

$$\Gamma^\alpha_{\mu\nu}[g] = \frac{1}{2} g^{\alpha\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) \quad (8.6.19)$$

can be related to that of its background counterpart through the relation

$$\Gamma^\alpha_{\mu\nu}[\bar{g}] + \delta\Gamma^\alpha_{\mu\nu}. \quad (8.6.20)$$

Here,

$$\delta\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\alpha\sigma} H_{\sigma\mu\nu}, \quad (8.6.21)$$

$$H_{\sigma\mu\nu} = \nabla_\mu h_{\nu\sigma} + \nabla_\nu h_{\mu\sigma} - \nabla_\sigma h_{\mu\nu}. \quad (8.6.22)$$

Notice the difference between the ‘full’ and ‘background’ Christoffel symbols, namely $\Gamma^\mu_{\alpha\beta} - \Gamma^\mu_{\alpha\beta}$, is a tensor.

**Problem 8.56. Geometric tensors**

With the result in eq. (8.6.20), show that for an arbitrary 1-form $V_\beta$,

$$\nabla_\alpha V_\beta = \bar{\nabla}_\alpha V_\beta - \delta\Gamma^\sigma_{\alpha\beta}V_\sigma. \quad (8.6.23)$$

Use this to compute $[\nabla_\alpha, \nabla_\beta]V_\lambda$ and proceed to show that the exact Riemann tensor is

$$R^\alpha_{\beta\mu\nu}[g] = \bar{R}^\alpha_{\beta\mu\nu}[\bar{g}] + \delta R^\alpha_{\beta\mu\nu}, \quad (8.6.24)$$

$$\delta R^\alpha_{\beta\mu\nu} \equiv \nabla_{[\mu}\delta\Gamma^\alpha_{\nu]\beta} + \delta\Gamma^\sigma_{\alpha[\mu}\delta\Gamma^\sigma_{\nu]\beta} \quad (8.6.25)$$

$$= \frac{1}{2} \nabla_\mu (g^{\alpha\lambda}H_{\lambda\nu\beta}) - \frac{1}{2} \nabla_\nu (g^{\alpha\lambda}H_{\lambda\mu\beta}) + \frac{1}{4} g^{\alpha\lambda}g^{\sigma\rho}(H_{\lambda\mu\sigma}H_{\rho\nu\beta} - H_{\lambda\nu\sigma}H_{\rho\mu\beta}). \quad (8.6.26)$$

where $R^\alpha_{\beta\mu\nu}[\bar{g}]$ is the Riemann tensor built entirely out of the background metric $\bar{g}_{\alpha\lambda}$.

From eq. (8.6.24), the Ricci tensor and scalars can be written down:

$$R_{\mu\nu}[g] = R^\sigma_{\mu\sigma\nu}, \quad \text{and} \quad R[g] = g^{\mu\nu}R_{\mu\nu}. \quad (8.6.27)$$

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. (8.6.6). 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 $n$th order piece of the Ricci scalar is

$$\delta_n \mathcal{R} = \sum_{\ell=0}^{n} \delta_{\ell} g^{\mu\nu} \delta_{n-\ell} R_{\mu\nu}, \quad (8.6.28)$$

where $\delta_{\ell} g^{\mu\nu}$ is the piece of the full inverse metric containing exactly $\ell$ powers of $h_{\mu\nu}$ and $\delta_{n-\ell} R_{\mu\nu}$ is that containing precisely $n - \ell$ powers of the same.

**Problem 8.57. Linearized geometric tensors** The Riemann tensor that contains up to one power of $h_{\mu\nu}$ can be obtained readily from eq. (8.6.24). The $H^2$ terms begin at order $h^2$, so we may drop them; and since $H$ is already linear in $\bar{h}$, the $g^{-1}$ contracted into it can be set to the background metric.

$$R^\alpha_{\beta\mu\nu}[\bar{g}] + \frac{1}{2} \nabla_\mu (\nabla_\nu h^\beta_{\alpha} + \nabla_{[\beta}h_{\alpha]}^\beta - \nabla^\alpha h_{\nu\beta}) + O(h^2) = R^\alpha_{\beta\mu\nu}[\bar{g}] + \frac{1}{2} \left( [\nabla_\mu, \nabla_\nu] h^\beta_{\alpha} + \nabla_\mu \nabla_\beta h^\alpha_{\nu} - \nabla_\nu \nabla_\beta h^\alpha_{\mu} - \nabla_\mu \nabla^\alpha h_{\nu\beta} + \nabla_\nu \nabla^\alpha h_{\mu\beta} \right) + O(h^2). \quad (8.6.29)$$

(The $|\beta|$ on the first line indicates the $\beta$ is not to be antisymmetrized.) Starting from the linearized Riemann tensor in eq. (8.6.29), let us work out the linearized Ricci tensor, Ricci scalar, and Einstein tensor.

Specifically, show that one contraction of eq. (8.6.29) yields the linearized Ricci tensor:

$$R_{\beta\nu} = \bar{R}_{\beta\nu} + \delta_1 R_{\beta\nu} + O(h^2), \quad (8.6.30)$$

$$\delta_1 R_{\beta\nu} \equiv \frac{1}{2} \left( \nabla^\mu \nabla_{(\beta} h_{\nu)}^\alpha - \nabla_\mu \nabla_{\beta} h_{\nu}^\alpha - \nabla_\nu \nabla_{\beta} h_{\mu}^\alpha + \nabla_\nu \nabla^\alpha h_{\mu\beta} \right). \quad (8.6.31)$$

Contracting this Ricci tensor result with the full inverse metric, verify that the linearized Ricci scalar is

$$\mathcal{R} = \bar{\mathcal{R}} + \delta_1 \mathcal{R} + O(h^2), \quad (8.6.32)$$

$$\delta_1 \mathcal{R} \equiv -h^\beta_{\nu} \bar{R}_{\beta\nu} + \left( \bar{g}^\mu_{\nu} \nabla^{\sigma} \nabla_{\sigma} - g^\mu_{\nu} \nabla^\sigma \nabla_{\sigma} \right) h_{\mu\nu}. \quad (8.6.33)$$

Now, let us define the variable $\bar{h}_{\mu\nu}$ through the relation

$$h_{\mu\nu} \equiv \bar{h}_{\mu\nu} - \frac{\bar{g}_{\mu\nu}}{d-2} \bar{h}, \quad \bar{h} \equiv \bar{h}^\sigma_{\sigma}. \quad (8.6.34)$$

First explain why this is equivalent to

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{\bar{g}_{\mu\nu}}{2} h. \quad (8.6.35)$$

(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

$$G_{\mu\nu} = \bar{G}_{\mu\nu}[\bar{g}] + \delta_1 G_{\mu\nu} + O(\bar{h}^2), \quad (8.6.36)$$

341
where
\[ \delta_1 G_{\mu\nu} \equiv -\frac{1}{2} \left( \Box h_{\mu\nu} + \bar{g}_{\mu\nu} \nabla_{\sigma} \nabla_{\rho} \bar{h}^{\sigma\rho} - \nabla_{(\mu} \nabla_{\nu)} \bar{h}_{\sigma}^{\sigma} \right) + \frac{1}{2} \left( \bar{g}_{\mu\nu} \bar{h}^{\rho\sigma} \bar{R}_{\rho\sigma} + \bar{h}_{(\mu}^{\rho} \bar{R}_{\nu)\rho\sigma} - \bar{h}_{\mu\nu} \bar{R} - 2 \bar{h}^{\rho\sigma} \bar{R}_{\mu\rho\nu\sigma} \right). \] (8.6.37)

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 \[21\].

\[ \delta_2 R_{\mu\nu} = \frac{1}{2} \left\{ \frac{1}{2} \nabla_{\rho} h_{\alpha\beta} \nabla_{\nu} h_{\alpha\beta} + h^{\alpha\beta} \left( \nabla_{\nu} \nabla_{\mu} h_{\alpha\beta} + \nabla_{\beta} \nabla_{\alpha} h_{\mu\nu} - \nabla_{\beta} \nabla_{\nu} h_{\alpha\mu} - \nabla_{\beta} \nabla_{\nu} h_{\mu\alpha} \right) \right\} \] (8.6.38)

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
\[ x^\mu = x'^\mu + \xi^\mu (x') \] (8.6.39)
is infinitesimal, in that \( \xi^\mu \) 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. \[8.6.39\] as referring to the same spacetime point, but expressed within infinitesimally different coordinate systems. Now, transforming from \( x \) to \( x' \),
\[ ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu = (\bar{g}_{\mu\nu}(x) + h_{\mu\nu}(x)) dx^\mu dx^\nu \] (8.6.40)

This teaches us that, the infinitesimal coordinate transformation of eq. \[8.6.39\] amounts to keeping the background metric fixed,
\[ \bar{g}_{\mu\nu}(x) \rightarrow \bar{g}_{\mu\nu}(x), \] (8.6.41)
but shifting
\[ h_{\mu\nu}(x) \rightarrow h_{\mu\nu}(x) + \xi^\sigma(x) \partial_\sigma \bar{g}_{\mu\nu}(x) + \bar{g}_{\sigma\{\mu(x)\partial_{\nu}\} \xi^\sigma(x)}, \quad (8.6.42) \]

followed by replacing
\[ x^\mu \rightarrow x'^\mu \quad \text{and} \quad \partial_\mu \equiv \frac{\partial}{\partial x^\mu} \rightarrow \frac{\partial}{\partial x'^\mu} \equiv \partial_{\mu'}. \quad (8.6.43) \]

However, since \( x \) and \( x' \) refer to the same point in spacetime\footnote{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.}, it is customary within the contemporary physics literature to drop the primes and simply phrase the coordinate transformation as replacement rules:
\[ x^\mu \rightarrow x^\mu + \xi^\mu(x), \quad (8.6.44) \]
\[ \bar{g}_{\mu\nu}(x) \rightarrow \bar{g}_{\mu\nu}(x), \quad (8.6.45) \]
\[ h_{\mu\nu}(x) \rightarrow h_{\mu\nu}(x) + \nabla_{\{\mu \xi_{\nu}\}}(x); \quad (8.6.46) \]

where we have recognized
\[ \xi^\sigma \partial_\sigma \bar{g}_{\mu\nu} + \bar{g}_{\sigma\{\mu \partial_{\nu}\} \xi^\sigma} = \nabla_{\{\mu \xi_{\nu}\}} \equiv (\mathcal{L}_\xi \bar{g})(x). \quad (8.6.47) \]

**Problem 8.58. Lie Derivative of a tensor along generator of coordinate transformation** If \( x \) and \( x' \) are infinitesimally nearby coordinate systems related via eq. (8.6.39), show that \( T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M}(x) \) (the components of a given tensor in the \( x^\mu \) coordinate basis) and \( T^{\mu'_1 \cdots \mu'_N}_{\nu'_1 \cdots \nu'_M}(x') \) (the components of the same tensor but in the \( x'^\mu \) coordinate basis) are in turn related via
\[ T^{\mu'_1 \cdots \mu'_N}_{\nu'_1 \cdots \nu'_M}(x') = T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M}(x \rightarrow x') + (\mathcal{L}_\xi T)^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M}(x \rightarrow x'); \quad (8.6.48) \]

where the Lie derivative of the tensor reads
\[ (\mathcal{L}_\xi T)^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M} = \xi^\sigma \partial_\sigma T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M} \]
\[ - T^{\mu_2 \cdots \mu_N}_{\nu_1 \cdots \nu_M} \partial_\sigma \xi^\mu_1 - \cdots - T^{\mu_1 \cdots \mu_{N-1}\sigma}_{\nu_1 \cdots \nu_M} \partial_\nu \partial_\sigma \xi^\mu_N \]
\[ + T^{\mu_1 \cdots \mu_N}_{\sigma\nu_2 \cdots \nu_M} \partial_\sigma \xi^\nu + \cdots + T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_{M-1}\sigma} \partial_\nu \partial_\sigma \xi^\nu. \quad (8.6.49) \]

The \( x \rightarrow x' \) on the right hand side of eq. (8.6.48) means, the tensor \( T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M} \) and its Lie derivative are to be computed in the \( x^\mu \)-coordinate basis – but \( x^\mu \) is to be replaced with \( x'^\mu \) afterwards.

Explain why the partial derivatives on the right hand side of eq. (8.6.49) may be replaced with covariant ones, namely
\[ (\mathcal{L}_\xi T)^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M} = \xi^\sigma \nabla_\sigma T^{\mu_1 \cdots \mu_N}_{\nu_1 \cdots \nu_M} \]
\[
-T^\sigma_{\mu_2\ldots\mu_N} \nabla_\sigma \xi^{\mu_1} - \ldots - T^\mu_{\mu_1\ldots\mu_N-1\nu_1\ldots\nu_M} \nabla_\mu \xi^{\mu N} \\
+ T^\mu_{\mu_1\ldots\mu_N\nu_2\ldots\nu_M} \nabla_\nu \xi^{\sigma} + \ldots + T^\mu_{\mu_1\ldots\mu_N\nu_1\ldots\nu_{M-1}} \nabla_{\nu M} \xi^{\sigma}.
\]

(8.6.50)

(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 that the former is a tensor.

We defined the Lie derivative of the metric \( \overline{g}_{\mu\nu} \) with respect to \( \xi^\alpha \) in eq. (8.6.47). Is it consistent with equations (8.6.49) and (8.6.50)?

**Lie Derivative of Vector**

Note that the Lie derivative of some vector field \( U^\mu \) with respect to \( \xi^\mu \) is, according to eq. (8.6.50),

\[
\mathcal{L}_\xi U^\mu = \xi^\sigma \nabla_\sigma U^\mu - U^\sigma \nabla_\sigma \xi^\mu
\]

(8.6.51)

\[
= \xi^\sigma \partial_\sigma U^\mu - U^\sigma \partial_\sigma \xi^\mu = [\xi, U]^\mu.
\]

(8.6.52)

We have already encountered the Lie bracket/commutator of vector fields, in eq. (8.3.22). There, we learned that \([\xi, U] = 0\) iff \( \xi \) 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 \( \xi \) as an active coordinate transformation induced by the displacement \( x \to x + \xi \). This in fact provides insight into the above mentioned theorem: if \( \mathcal{L}_\xi U^\mu = 0 \) that means \( U \) remains unaltered upon a coordinate transformation induced along the direction of \( \xi \); that in turn indicates, it is possible to move along the integral curve of \( \xi \), bringing us from one integral curve of \( U \) to the next—while consistently maintaining the same coordinate value along the latter. Similarly, since \([\xi, U] = -[U, \xi] = -\mathcal{L}_U \xi = 0\), the vanishing of the Lie bracket also informs us the coordinate value along the integral curve of \( \xi \) 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 8.59. Gauge transformations of a tensor**

Consider perturbing a spacetime tensor

\[
T^{\mu_1\ldots\mu_N\nu_1\ldots\nu_M} \equiv \overline{T}^{\mu_1\ldots\mu_N\nu_1\ldots\nu_M} + \delta T^{\mu_1\ldots\mu_N\nu_1\ldots\nu_M},
\]

(8.6.53)

where \( \delta T^{\mu_1\ldots\mu_N\nu_1\ldots\nu_M} \) is small in the same sense that \( \xi^\alpha \) and \( h_{\mu\nu} \) are small. Perform the infinitesimal coordinate transformation in eq. (8.6.39) on the tensor in eq. (8.6.53) and attribute all the transformations to the \( \delta T^{\mu_1\ldots\mu_N\nu_1\ldots\nu_M} \). Write down the ensuing gauge transformation, in direct analogy to eq. (8.6.46). 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 (8.58).

8.6.1 Perturbed Flat Spacetimes & Gravitational Waves

In this subsection we shall study perturbations about flat spacetimes

\[
g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1.
\]

(8.6.54)
In 4D, this is the context where gravitational waves are usually studied.

Under a Poincaré transformation in eq. (8.1.7), \( x^\mu = \Lambda^\mu_\nu x^\nu + a^\mu \), where \( \Lambda^\mu_\nu \) satisfies (8.1.5), observe that the metric transforms as

\[
g_{\alpha\beta'}(x') = g_{\mu\nu}(x = \Lambda x' + a)\Lambda^\mu_\alpha \Lambda^\nu_\beta = \left( \eta_{\mu\nu} + h_{\mu\nu}(x = \Lambda x' + a) \right) \Lambda^\mu_\alpha \Lambda^\nu_\beta \equiv \eta_{\alpha\beta} + h_{\alpha\beta'}(x').
\]

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:

\[
h_{\alpha\beta'}(x') = h_{\mu\nu}(x(x')) \Lambda^\mu_\alpha \Lambda^\nu_\beta, \quad x^\mu = \Lambda^\mu_\nu x^\nu + a^\mu.
\]

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

\[
x^\mu = x'^\mu + \xi^\mu(x'),
\]

the linearized Riemann tensor written in the \( x \) versus \( x' \) systems are related as

\[
\delta_1 R_{\mu\nu\alpha\beta}(x) = \delta_1 R_{\mu'\nu'\alpha'\beta'}(x') + \mathcal{O}(h^2, \xi \cdot h, \xi^2).
\]

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:

\[
g_{\mu\nu}dx^\mu dx^\nu = \eta_{\mu\nu}dx^\mu dx^\nu + h_{ij}^{(s)} dx^i dx^j \quad (\text{Synchronous gauge}).
\]

The latter is defined by the Lorentz-covariant constraint

\[
\partial^\mu h_{\mu\nu} = \frac{1}{2} \partial_\nu h, \quad h \equiv \eta^{\alpha\beta} h_{\alpha\beta}, \quad (\text{de Donder gauge}).
\]

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. (8.6.46), the perturbation in a Minkowski background transforms as

\[
h_{\mu\nu} \rightarrow h_{\mu\nu} + \partial_\nu \xi_\mu + \partial_\mu \xi_\nu.
\]
\[ h \to h + 2 \partial_\sigma \xi^\sigma. \] (8.6.65)

Hence, if \( h_{00} \) were not zero, we may render it so by choosing \( \xi_0 = -(1/2) \int^t h_{00} dt \); since

\[ h_{00} \to h_{00} + 2 \partial_0 \xi_0 \] (8.6.66)
\[ = h_{00} + 2 \partial_0 \left( -\frac{1}{2} \int^t h_{00} dt \right) = 0. \] (8.6.67)

Moreover, if \( h_{0i} \) were not zero, an infinitesimal coordinate transformation would yield

\[ h_{0i} \to h_{0i} + \partial_i \xi_0 + \partial_0 \xi_i \] (8.6.68)
\[ = h_{0i} - \frac{1}{2} \int^t \partial_i h_{00} dt + \partial_0 \xi_i. \] (8.6.69)

The right hand side is zero if we choose

\[ \xi_i = - \int^t \left( h_{0i} - \frac{1}{2} \int^t \partial_i h_{00} dt' \right) dt'. \] (8.6.70)

That is, by choosing \( \xi_\mu \) appropriately, \( h_{0\mu} = h_{\mu0} \) may always be set to zero.

As for the de Donder gauge condition in eq. (8.6.63), we first re-write it using eq. (8.6.35)

\[ \bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h. \] (8.6.71)

Namely,

\[ \partial^\mu \bar{h}_{\mu\nu} = \partial^\mu \left( h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h \right) = 0. \] (8.6.72)

Utilizing eq. (8.6.64), we may deduce the gauge-transformation of \( \bar{h}_{\mu\nu} \) is

\[ \bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial \cdot \xi, \quad \partial \cdot \xi \equiv \partial^\sigma \xi_\sigma. \] (8.6.73)

Now, if eq. (8.6.72) were not obeyed, a gauge transformation would produce

\[ \partial^\mu \bar{h}_{\mu\nu} \to \partial^\mu \bar{h}_{\mu\nu} + \partial^\mu (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) - \eta_{\mu\nu} \partial^\mu \partial \cdot \xi \] (8.6.74)
\[ = \partial^\mu \bar{h}_{\mu\nu} + \partial^2 \xi_\nu. \] (8.6.75)

Therefore, by choosing \( \xi_\nu \) 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. (8.6.72). We also note, suppose \( h_{\mu\nu} \) 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 8.60.** Are the synchronous and de Donder gauges “infinitesimally nearby” coordinate systems?
Problem 8.61. Co-moving geodesics in synchronous gauge

Prove that

\[ Z^\mu(t) = (t, \vec{Z}_0), \quad (8.6.76) \]

where \( \vec{Z}_0 \) is time-independent, satisfies the geodesic equation in the spacetime

\[ g_{\mu\nu}dx^\mu dx^\nu = dt^2 + g_{ij}(t, \vec{x})dx^i dx^j. \quad (8.6.77) \]

This result translates to the following interpretation: each \( \vec{x} \) in eq. (8.6.77) 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. (8.6.62) is a special case of eq. (8.6.77).

Linearized Synge’s World Function

In the weak field metric of eq. (8.6.54), according to eq. (8.4.55), half the square of the geodesic distance between \( x \) and \( x' \) is

\[ \sigma(x, x') = \frac{1}{2} \int_0^1 d\lambda (\eta_{\mu\nu} + h_{\mu\nu}(\vec{Z})) \frac{dZ^\mu}{d\lambda} \frac{dZ^\nu}{d\lambda}; \quad (8.6.78) \]

where the trajectories obey geodesic equation (8.4.57)

\[ \frac{d^2Z^\mu}{d\lambda^2} + \Gamma^\mu_{\alpha\beta} \frac{dZ^\alpha}{d\lambda} \frac{dZ^\beta}{d\lambda} = 0 \quad (8.6.79) \]

subject to the boundary conditions

\[ Z^\mu(\lambda = 0) = x'^\mu \quad \text{and} \quad Z^\mu(\lambda = 1) = x^\mu. \quad (8.6.80) \]

If the perturbations were not present, \( h_{\mu\nu} = 0 \), the geodesic equation is

\[ \frac{d^2\vec{Z}^\mu}{d\lambda^2} = 0; \quad (8.6.81) \]

whose solution, in turn, is

\[ \vec{Z}^\mu(\lambda) = x'^\mu + \lambda(x - x')^\mu, \quad (8.6.82) \]

\[ \dot{\vec{Z}}^\mu(\lambda) = (x - x')^\mu. \quad (8.6.83) \]

When the perturbations are non-trivial, \( h_{\mu\nu} \neq 0 \), the full solution \( Z^\mu = \vec{Z}^\mu + \delta Z^\mu \) should deviate from the zeroth order solution \( \vec{Z}^\mu \) at linear order in the perturbations: \( \delta Z^\mu \sim \mathcal{O}(h_{\mu\nu}) \). One may see this from eq. (8.4.71). Hence, if we insert \( Z^\mu = \vec{Z}^\mu + \delta Z^\mu \) into Synge’s world function in eq. (8.6.78),

\[ \sigma(x, x') = \frac{1}{2} \int_0^1 d\lambda (\eta_{\mu\nu} + h_{\mu\nu}(\vec{Z})) (x - x')^\mu (x - x')^\nu \]

\[ - \int_0^1 \delta Z^\mu(\lambda) (\eta_{\mu\nu} + h_{\mu\nu}(\vec{Z})) \frac{D^2\vec{Z}^\nu}{d\lambda^2} d\lambda + \mathcal{O}((\delta Z)^2); \quad (8.6.84) \]
because the zeroth order geodesic equation is satisfied, namely \( d^2 \tilde{Z} / d\lambda^2 = 0 \), \( D^2 \tilde{Z}^\nu / d\lambda^2 = \Gamma^\nu_{\alpha\beta} \tilde{Z}^\alpha \tilde{Z}^\beta \sim \mathcal{O}(h_{\mu\nu}) \) and therefore the second line scales as \( \mathcal{O}(h^2_{\mu\nu}) \) 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. \([8.6.82]\) \[106\]

\[
\sigma(x, x') = \frac{1}{2} (x - x')^2 + \frac{1}{2} (x - x')^\mu(x - x')^\nu \int_0^1 h_{\mu\nu} \left( \tilde{Z}(\lambda) \right) d\lambda + \mathcal{O}(h^2) \quad (8.6.85)
\]

**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. \([8.6.62]\), where \( h_{\mu0} = h_{0\mu} = 0 \), the square of their geodesic spatial separation at a fixed time \( t = t' \) is gotten from eq. \([8.6.85]\) through

\[
\ell^2 = -2\sigma(t = t'; \vec{y}, \vec{y}') = (\vec{y} - \vec{y}')^2 - (y - y')^i(y - y')^j \int_0^1 h^{(s)}_{ij} (t, \vec{y} + \lambda(\vec{y} - \vec{y}')) d\lambda + \mathcal{O}(h^2) \quad (8.6.86)
\]

\[
\ell(t; \vec{y} \leftrightarrow \vec{y}') = |\vec{y} - \vec{y}'| \left( 1 - \frac{1}{2} \hat{R}^i \hat{R}^j \int_0^1 h^{(s)}_{ij} (t, \tilde{Z}(\lambda)) d\lambda + \mathcal{O}(h^2) \right), \quad (8.6.88)
\]

\[
\hat{R} \equiv \frac{\vec{y} - \vec{y}'}{|\vec{y} - \vec{y}'|}. \quad (8.6.89)
\]

(Remember \( \tilde{Z} \) in eq. \([8.6.82]\).)

**Gravitational Wave Polarization & Oscillation Patterns** We may re-phrase eq. \([8.6.88]\) 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 \( h^{(s)}_{ij} \),

\[
\left( \frac{\delta \ell}{\ell_0} \right) (t; \vec{y} \leftrightarrow \vec{y}') = -\frac{1}{2} \hat{R}^i \hat{R}^j \int_0^1 h^{(s)}_{ij} (t, \tilde{Z}(\lambda)) d\lambda + \mathcal{O}(h^2). \quad (8.6.90)
\]

If we define gravitational waves to be simply the finite frequency portion of the tidal signal in eq. \([8.6.90]\), 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

\[
h^{(s)}_{ij} (t, \tilde{Z}(\lambda)) = \int_\mathbb{R} \tilde{h}^{(s)}_{ij} (\omega, \tilde{Z}(\lambda)) e^{-i\omega t} \frac{d\omega}{2\pi}, \quad (8.6.91)
\]

\[106\]This sort of “first-order-variation-vanishes” argument occurs frequently in field theory as well.
then eq. (8.6.90) reads
\[ \left( \frac{\delta \ell}{\ell_0} \right) (\omega; \vec{y} \leftrightarrow \vec{y}') = -\frac{1}{2} \vec{R}^i \vec{R}^j \int_0^1 \hat{h}^{(s)}_{ij}(\omega, \vec{y} + \lambda(\vec{y}' - \vec{y})) d\lambda + \mathcal{O}(h^2). \] (8.6.92)

Now, a direct calculation will reveal
\[ \delta_1 R_{0i0j}(t, \vec{x}) = -\frac{1}{2} \partial_0^2 h^{(s)}_{ij}(t, \vec{x}), \] (Synchronous gauge). (8.6.93)

To translate this statement to frequency space, we replace \( \partial_0 = \partial_t \rightarrow -i\omega \),
\[ \delta_1 \tilde{R}_{0i0j}(\omega, \vec{x}) = \frac{\omega^2}{2} \tilde{h}^{(s)}_{ij}(\omega, \vec{x}), \] (Synchronous gauge). (8.6.94)

Gravitational waves are associated with time dependent radiative processes, capable of performing dissipative work through their oscillatory tidal forces. To this end, eq. (8.6.94) 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. (8.6.94) into eq. (8.6.92), we see that the finite frequency gravitational-wave-driven fractional distortion of space – namely,
\[ \left( \frac{\delta \ell}{\ell_0} \right) (\omega \neq 0; \vec{y} \leftrightarrow \vec{y}') \equiv \frac{\vec{R}^i \vec{R}^j}{\omega^2} \int_0^1 \delta_1 \tilde{R}_{0i0j}(\omega, \vec{y} + \lambda(\vec{y}' - \vec{y})) d\lambda + \mathcal{O}(h^2) \] (8.6.95)
– 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 \( \delta_1 \tilde{R}_{0i0j} \) is a symmetric rank–2 spatial tensor in the \( ij \) indices.

**Problem 8.62.** Verify eq. (8.6.93).

**Problem 8.63.** 4D Gravitational Wave Polarizations

In 3+1 dimensional spacetime, choose the unit vector along the 3–axis \( \hat{e}_3 \) to be the direction of propagation of the finite frequency \( \tilde{h}^{(s)}_{ij} \) in eq. (8.6.92). Then proceed to build upon Problem (4.88) to decompose the fractional distortion of space in eq. (8.6.92) into its irreducible constituents – i.e., the spin–0, spin–1 and spin–2 finite-frequency waves.

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 8.64.** 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,
\[ h^{(s)}_{ij}(t, \vec{x}) = \int \frac{d\omega}{2\pi} \int_{\mathbb{R}^D} \frac{d^D \vec{k}}{(2\pi)^D} \tilde{h}^{(s)}_{ij}(\omega, \vec{k}) e^{-i\omega t} e^{i\vec{k} \cdot \vec{x}}; \] (8.6.96)
so that $\partial_\mu \leftrightarrow -i(\omega, k_\mu)$. The associated synchronous gauge Riemann tensor components then read

$$\delta_1 \tilde{R}_{0ij}(\omega, \vec{k}) = +\frac{\omega^2}{2} \tilde{h}_{ij}^{(s)}(\omega, \vec{k}), \quad \text{(Synchronous gauge).} \quad (8.6.97)$$

Up to this point, we have not assumed a dispersion relation between $\omega$ and $\vec{k}$. Suppose we impose the null condition

$$\omega^2 = \vec{k}^2 \quad (8.6.98)$$

on both the synchronous and de Donder gauge perturbations, so they are both superpositions of traveling waves propagating at unit speed –

$$h_{ij}(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{1}{2} \left\{ \tilde{h}_{ij}(k) e^{-i|\vec{k}| t} + \tilde{h}_{ij}^{(s)}(k)^* e^{+i|\vec{k}| t} \right\} e^{i\vec{k} \cdot \vec{x}} \frac{d^D \vec{k}}{(2\pi)^D}, \quad k^\mu \equiv (|\vec{k}|, \vec{k}) \quad (8.6.99)$$

– now, verify directly that the corresponding Riemann components are

$$\delta_1 \tilde{R}_{0ij}(\omega, \vec{k}) = \frac{\omega^2}{2} \left( \tilde{h}_{ij} + \tilde{k}_i \tilde{k}_j \tilde{k}^k + \tilde{k}_i \tilde{k}_j \tilde{h}_{mn} \tilde{k}^m \tilde{k}^n \right), \quad \text{(de Donder);} \quad (8.6.100)$$

$$\hat{k}^i \equiv k^i / |\vec{k}|, \quad \omega^2 = \vec{k}^2. \quad (8.6.101)$$

Next, verify $\delta_1 \tilde{R}_{0ij}$ in eq. (8.6.100) is transverse and traceless:

$$\delta^{ij} \delta_1 \tilde{R}_{0ij} = 0 = \hat{k}^i \delta_1 \tilde{R}_{0ij}. \quad (8.6.102)$$

Finally, demonstrate that such a traveling-wave $\delta_1 \tilde{R}_{0ij}$ in de Donder gauge is simply the transverse-traceless (TT) portion of the metric perturbation itself:

$$\delta_1 \tilde{R}_{0ij}(\omega, \vec{k}) = \frac{\omega^2}{2} \tilde{P}_{ijab} \tilde{h}_{ab}(\omega, \vec{k}), \quad (8.6.103)$$

where the TT projector is

$$\tilde{P}_{ijab} = \frac{1}{2} \tilde{P}_{i(a} \tilde{P}_{b)j} - \frac{1}{D - 1} \tilde{P}_{ij} \tilde{P}_{ab}, \quad (8.6.104)$$

$$\tilde{P}_{ij} = \delta_{ij} - \hat{k}_i \hat{k}_j. \quad (8.6.105)$$

It enjoys the following properties:

$$\tilde{P}_{ijab} = \tilde{P}_{abij}, \quad \tilde{P}_{jiab} = \tilde{P}_{ijab}, \quad \delta^{ij} \tilde{P}_{ijab} = 0 = \hat{k}^i \tilde{P}_{ijab}. \quad (8.6.106)$$

**Helicity–2 modes**

Finally, by choosing $\hat{k} \equiv \hat{e}_3$, the unit vector along the 3–axis, verify the claim in the previous problem, that the null traveling waves described by these linearized $\delta_1 \tilde{R}_{0ij}$ are purely helicity–2 modes only.

**Hint:** Throughout these calculations, you would need to repeatedly employ the de Donder gauge condition (eq. (8.6.63)) in Fourier spacetime: $k^\mu \tilde{h}_{\mu\nu} = (1/2)k_\nu \tilde{h}$, with $k^\mu \equiv (\omega, \vec{k})$.  

350
From our previous discussion, since the linearized Riemann tensor is gauge-invariant, we may immediately equate the $0i0j$ components in the synchronous (eq. (8.6.97)) and de Donder (eq. (8.6.100)) 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.

$$\tilde{h}^{(s)}_{ij}[\text{Synchronous}] = \tilde{P}_{ijab}\tilde{h}_{ab}[\text{de Donder}] \tag{8.6.107}$$

That this holds only for finite frequencies – the formulas in equations (8.6.97) and (8.6.100) do not contain $\delta(\omega)$ or $\delta'(\omega)$ terms – because $\omega^2\delta(\omega) = 0 = \omega^2\delta'(\omega)$. More specifically, since eq. (8.6.93) involved a second time derivative on $h^{(s)}_{ij}$, by equating it to the (position-spacetime version of) eq. (8.6.100), we may solve the synchronous gauge metric perturbation only up to its initial value and time derivative:

$$h^{(s)}_{ij}(t, \vec{x}) = -2\int_{t_0}^{t} \int_{t_0}^{\tau_2} \delta_1 R_{0i0j}(\tau_1, \vec{x}) d\tau_1 d\tau_2 + (t - t_0)\dot{h}^{(s)}_{ij}(t_0, \vec{x}) + h^{(s)}_{ij}(t_0, \vec{x}). \tag{8.6.108}$$

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. (8.6.107), eq. (8.6.108) does not depend on specializing to traveling waves obeying the null dispersion relation $k^2 \equiv k_\mu k^\mu = 0$. Moreover, eq. (8.6.108) suggests, up to the two initial conditions, $h^{(s)}_{ij}$ itself is almost gauge-invariant – after all it measures something geometrical, eq. (8.6.88), 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. (8.6.64)

$$0 = h^{(s)}_{00} \rightarrow 2\partial_0 \xi_0 = 0. \tag{8.6.109}$$

That is, $\xi_0$ needs to be time-independent. For the $0i$ component to remain zero,

$$0 = h^{(s)}_{0i} \rightarrow \partial_0 \xi_i + \partial_i \xi_0 = 0. \tag{8.6.110}$$

This allows us to assert

$$\xi_i(t, \vec{x}) = -(t - t_0)\partial_i \xi_0(\vec{x}) + \xi_i(t_0, \vec{x}). \tag{8.6.111}$$

Under such a coordinate transformation, $x \rightarrow x + \xi$,

$$h^{(s)}_{ij} \rightarrow h^{(s)}_{ij} + \partial_i \xi_j + \partial_j \xi_i = h^{(s)}_{ij}(t, \vec{x}) - 2(t - t_0)\partial_i \partial_j \xi_0(\vec{x}) + \partial_i \xi_j(t_0, \vec{x}). \tag{8.6.112}$$

Comparison with eq. (8.6.108) 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.

---

107 More specifically, eq. (8.6.107) holds whenever the linearized vacuum Einstein’s equations hold; whereas eq. (8.6.108) is true regardless of the underlying dynamics of the metric perturbations.
8.7 Special Topic: Conformal/Weyl Transformations

In this section, we collect for the reader’s reference, 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,

\[ g_{\mu\nu}(x) \equiv \Omega^2(x) \bar{g}_{\mu\nu}(x). \]  

(8.7.1)

The inverse metric is

\[ g^{\mu\nu}(x) = \Omega(x)^{-2} \bar{g}^{\mu\nu}(x), \quad \bar{g}^{\mu\sigma} \bar{g}_{\sigma\nu} \equiv \delta^\mu_\nu. \]  

(8.7.2)

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 \( \nabla \) is with respect to \( g^{\mu\nu} \) while the \( \bar{\nabla} \) is with respect to \( \bar{g}_{\mu\nu} \).

**Metric Determinant**

Since

\[ \det g_{\mu\nu} = \det (\Omega^2 \bar{g}_{\mu\nu}) = \Omega^{2d} \det \bar{g}_{\mu\nu}. \]  

(8.7.3)

we must also have

\[ |g|^{1/2} = \Omega^d |\bar{g}|^{1/2}. \]  

(8.7.4)

**Scalar Gradients**

The scalar gradient with a lower index is just a partial derivative. Therefore

\[ \nabla_\mu \varphi = \bar{\nabla}_\mu \varphi = \partial_\mu \varphi. \]  

(8.7.5)

while \( \nabla^\mu \varphi = g^{\mu\nu} \nabla_\nu \varphi = \Omega^{-2} \bar{g}^{\mu\nu} \bar{\nabla}_\nu \varphi \), so

\[ \nabla^\mu \varphi = \Omega^{-2} \bar{\nabla}^\mu \varphi. \]  

(8.7.6)

**Scalar Wave Operator**

The wave operator \( \Box \) in the geometry \( g_{\mu\nu} \) is defined as

\[ \Box \equiv g^{\mu\nu} \nabla_\mu \nabla_\nu = \nabla_\mu \nabla^\mu. \]  

(8.7.7)

By a direct calculation, the wave operator \( \Box \) with respect to \( g_{\mu\nu} \) acting on a scalar \( \psi \) is

\[ \Box \varphi = \frac{1}{\Omega^2} \left( \frac{d-2}{\Omega} \nabla_\mu \Omega \cdot \nabla^\mu \varphi + \Box \varphi \right), \]  

(8.7.8)

where \( \Box \) is the wave operator with respect to \( \bar{g}_{\mu\nu} \). We also have

\[ \Box (\Omega^s \psi) = \frac{1}{\Omega^2} \left( \left( s \Omega^{s-2} \nabla_\mu \Omega \cdot \nabla^\mu \Omega \psi + (2s + d - 2) \Omega^{s-2} \nabla_\mu \Omega \bar{\nabla}^\mu \psi + \Box \psi \right) \right). \]  

(8.7.9)
Christoffel Symbols

A direct calculation shows:

\[
\Gamma^\mu_{\alpha\beta}[g] = \Gamma^\mu_{\alpha\beta}[\bar{g}] + (\partial_{\{\alpha} \ln \Omega) \delta^\mu_{\beta\}} - \bar{g}_{\alpha\beta} \bar{g}^{\mu\nu} (\partial^\nu \ln \Omega) \\
= \Gamma^\mu_{\alpha\beta}[\bar{g}] + (\nabla_{\{\alpha} \Omega) \delta^\mu_{\beta\}} - \bar{g}_{\alpha\beta} \nabla^\mu \ln \Omega. 
\] (8.7.10)

Riemann Tensor

By viewing the difference between \(g_{\mu\nu}\) and \(\bar{g}_{\mu\nu}\) as a ‘perturbation’,

\[
g_{\mu\nu} - \bar{g}_{\mu\nu} = (\Omega^2 - 1) \bar{g}_{\mu\nu} \equiv h_{\mu\nu}, 
\] (8.7.12)

we may employ the results in §(8.6). In particular, eq. (8.6.24) may be used to infer that the Riemann tensor is

\[
R^\alpha_{\beta\mu\nu}[g] = \bar{R}^\alpha_{\beta\mu\nu}[\bar{g}] + (d - 2d) \nabla_\beta \nabla_\mu \ln \Omega - \bar{g}_{\beta\nu} \Box \ln \Omega \\
+ \delta_{\mu}^{\alpha} \nabla_\nu \ln \Omega + \bar{g}_{\mu}^{\alpha} \nabla_\nu \ln \Omega \bar{g}_{\nu}^{\beta} + (\nabla \ln \Omega)^2 \bar{g}_{\beta\nu} \delta^\alpha_{\mu}. 
\] (8.7.13)

Ricci Tensor

In turn, the Ricci tensor is

\[
R_{\beta\nu}[g] = \bar{R}_{\beta\nu}[\bar{g}] + (d - 1d) \nabla_\beta \nabla_\nu \ln \Omega - \bar{g}_{\beta\nu} \Box \ln \Omega \\
+ (d - 2d) \left( \nabla_\beta \ln \Omega \nabla_\nu \ln \Omega - \bar{g}_{\beta\nu} \left( \nabla \ln \Omega \right)^2 \right). 
\] (8.7.14)

Ricci Scalar

Contracting the Ricci tensor with \(g^\beta_{\nu} = \Omega^{-2} \bar{g}^\beta_{\nu}\), we conclude

\[
\mathcal{R}[g] = \Omega^{-2} \left( \mathcal{R}[\bar{g}] + 2(1 - 1d) \Box \ln \Omega - (d - 1d)(1 - d) \left( \nabla \ln \Omega \right)^2 \right). 
\] (8.7.16)

Weyl Tensor

The Weyl tensor, for spacetime dimensions greater than two \((d > 2)\), is defined to be the completely trace-free portion of the Riemann tensor:

\[
C_{\mu\nu\alpha\beta} \equiv R_{\mu\nu\alpha\beta} - \frac{1}{d - 2d} (R_{\alpha[\mu} g_{\nu]\beta} - R_{\beta[\mu} g_{\nu]\alpha}) + \frac{g_{\mu[\alpha} g_{\beta\nu]} (d - 2d)(d - 1)}{d - 2d} \mathcal{R}[g]. 
\] (8.7.17)

By a direct calculation, one may verify \(C_{\mu\nu\alpha\beta}\) has the same index-symmetries as \(R_{\mu\nu\alpha\beta}\) and is indeed completely traceless: \(g^{\mu\alpha} C_{\mu\nu\alpha\beta} = 0\). Using equations (8.7.1), (8.7.13), (8.7.14), and (8.7.16), one may then deduce the Weyl tensor with one upper index is \textit{invariant} under conformal transformations:

\[
C^\mu_{\nu\alpha\beta}[g] = C^\mu_{\nu\alpha\beta}[\bar{g}]. 
\] (8.7.18)

If we lower the index \(\mu\) on both sides,

\[
C_{\mu\nu\alpha\beta}[g] = \Omega^2 C_{\mu\nu\alpha\beta}[\bar{g}]. 
\] (8.7.19)

Let us also record that:

In spacetime dimensions greater than 3, 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\[108\]

\[108\text{In } d = 3 \text{ dimensions, a spacetime is locally conformally flat iff its Cotton tensor vanishes.}\]
Problem 8.65. Cosmological Perturbation Theory & Gauge-Invariance

Consider a perturbed metric of the form

\[ g_{\mu\nu} = \Omega^2 (\eta_{\mu\nu} + \chi_{\mu\nu}), \quad |\chi_{\mu\nu}| \ll 1. \]  

(Cosmological perturbation theory is a special case; where \( \Omega \) 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 (8.59).

Einstein Tensor

From equations (8.7.1), (8.7.14) and (8.7.16), we may also compute the transformation of the Einstein tensor

\[ G_{\beta\nu}[g] \equiv R_{\beta\nu} - (g_{\beta\nu}/2)R. \]

\[ G_{\beta\nu}[g] = G_{\beta\nu}[\bar{g}] + (2 - d) \left( \nabla_\beta \nabla_\nu \ln \Omega - \bar{g}_{\beta\nu} \Box \ln \Omega \right) \]

\[ + (d - 2) \left( \nabla_\beta \ln \Omega \nabla_\nu \ln \Omega - \bar{g}_{\beta\nu} \frac{3 - d}{2} (\nabla \ln \Omega)^2 \right) \]  

(8.7.21)

Notice the Einstein tensor is invariant under constant conformal transformations: \( G_{\beta\nu}[g] = G_{\beta\nu}[\bar{g}] \) whenever \( \partial_\mu \Omega = 0. \)

Problem 8.66. 2D Einstein is Zero

Explain why the Einstein tensor is zero in 2D. This implies the 2D Ricci tensor is proportional to the Ricci scalar:

\[ R_{\alpha\beta} = \frac{1}{2} g_{\alpha\beta} R. \]  

(8.7.22)

Hint: Refer to Problem (7.38).

Scalar Field Action

In \( d \) dimensional spacetime, the following action involving the scalar \( \varphi \) and Ricci scalar \( R[g] \),

\[ S[\varphi] \equiv \int d^dx \sqrt{|g|} \frac{1}{2} \left( g^{\alpha\beta} \nabla_\alpha \varphi \nabla_\beta \varphi + \frac{d - 2}{4(d - 1)} R \varphi^2 \right), \]  

is invariant – up to surface terms – under the simultaneous replacements

\[ g_{\alpha\beta} \to \Omega^2 g_{\alpha\beta}, \quad g^{\alpha\beta} \to \Omega^{-2} g^{\alpha\beta}, \quad \sqrt{|g|} \to \Omega^d \sqrt{|g|}, \quad \varphi \to \Omega^{1 - \frac{d}{2}} \varphi. \]  

(8.7.24)

(8.7.25)

The jargon here is that \( \varphi \) 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, \( (\nabla A)^2 \equiv \bar{g}^{\alpha\lambda} \nabla_\alpha A \nabla_\lambda A \) and \( \bar{\nabla}^\alpha A \equiv \bar{g}^{\alpha\lambda} \bar{\nabla}_\lambda A \). 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 8.67. de Sitter as a Maximally Symmetric Spacetime

Verify that de Sitter spacetime, with coordinates \( x^\mu \equiv (\eta, x^i) \),

\[
g_{\mu\nu}(\eta, \vec{x}) = \Omega(\eta)^2 \eta_{\mu\nu}, \quad \Omega(\eta) \equiv -\frac{1}{H\eta}
\]

has the following Riemann tensor:

\[
R_{\mu\nu\alpha\beta} = \frac{\mathcal{R}}{d(d-1)}g_{\mu[\alpha}g_{\beta]\nu}. \tag{8.7.27}
\]

Also verify that the Ricci tensor and scalar are

\[
R_{\mu\nu} = \frac{\mathcal{R}}{d}g_{\mu\nu} \quad \text{and} \quad \mathcal{R} = -\frac{2d\Lambda}{d-2}. \tag{8.7.28}
\]

de Sitter spacetime is a maximally symmetric spacetime, with \( d(d+1)/2 \) Killing vectors. Verify that the following are Killing vector of eq. (8.7.26):

\[
T^\mu \partial_\mu \equiv -Hx^\mu \partial_\mu \tag{8.7.29}
\]

and

\[
K_{(i)}^\mu \partial_\mu \equiv x^i T^\mu \partial_\mu - H \bar{\sigma} \partial_{x^i}, \tag{8.7.30}
\]

\[
\bar{\sigma} \equiv \frac{1}{2} \left( \eta^2 - \vec{x}^2 \right) = \frac{1}{2} \eta_{\mu\nu} x^\mu x^\nu. \tag{8.7.31}
\]

(Hint: It is easier to use the right hand side of eq. (8.5.45) in eq. (8.5.46).) Can you write down the remaining Killing vectors? (Hint: Think about the symmetries on a constant-\( \eta \) surface.) Using (some of) these \( d(d+1)/2 \) Killing vectors and eq. (8.5.74), explain why the Ricci scalar of the de Sitter geometry is a spacetime constant.

Observer time: de Sitter spacetime may also be written as

\[
ds^2 = dt^2 - e^{2Ht} \, d\vec{x} \cdot d\vec{x}. \tag{8.7.32}
\]

Can you describe the relation between \( \eta \) 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. (8.7.29) may also be expressed as

\[
T^\mu \partial_\mu = \frac{1}{\Omega(\eta)} \partial_\eta - Hx^i \partial_i = \partial_t - Hx^i \partial_i. \tag{8.7.33}
\]

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.

\[109\text{As Weinberg} \ [20] \text{explains, maximally symmetric spacetimes are essentially unique, in that they are characterized by a single dimension-ful scale. We see that this scale is nothing but the cosmological constant } \Lambda.\]
9 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.

9.1 Laplacians and Poisson’s Equation

9.1.1 Poisson’s equation, uniqueness of solutions

Poisson’s equation in \( D \)-space is defined to be

\[
-\vec{\nabla}^2 \psi(\vec{x}) = J(\vec{x}),
\]

where \( J \) is to be interpreted as some given mass/charge density that sources the Newtonian/electric potential \( \psi \). The most physically relevant case is in 3D; if we use Cartesian coordinates, Poisson’s equation reads

\[
-\vec{\nabla}^2 \psi(\vec{x}) = -\left( \frac{\partial^2 \psi}{\partial (x^1)^2} + \frac{\partial^2 \psi}{\partial (x^2)^2} + \frac{\partial^2 \psi}{\partial (x^3)^2} \right) = J(\vec{x}).
\]

We will soon see how to solve eq. (9.1.1) 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. (9.1.1)) in some domain \( \mathcal{D} \) is unique once \( \psi \) is specified on the boundary of the domain \( \partial \mathcal{D} \). 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 \( \partial \mathcal{D} \), then \( \psi \) is unique up to an additive constant.

The proof goes by contradiction. Suppose there were two distinct solutions, \( \psi_1 \) and \( \psi_2 \). Let us define their difference as

\[
\Psi \equiv \psi_1 - \psi_2
\]

and start with the integral

\[
I \equiv \int_{\mathcal{D}} d^D \vec{x} \sqrt{|g|} \nabla_i \Psi^i \nabla^i \Psi \geq 0.
\]

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. (7.2.57)), where \( g^{ij} = \varepsilon^i_a \varepsilon^j_b \delta_{ab} \). The \( \sqrt{|g|} \) is always positive, since it describes volume, whereas \( \nabla_i \Psi \nabla^i \Psi \) is really a sum of squares.

\[
\sqrt{|g|} \delta^{ab} \nabla_a \Psi^i \nabla^b \Psi = \sqrt{|g|} \sum_a |\nabla_a \Psi|^2 \geq 0.
\]

\[\text{110Expressing the gradients in an orthonormal frame is, in fact, the primary additional ingredient to this proof, when compared to the flat space case. Moreover, notice this proof relies on the Euclidean (positive definite) nature of the metric.}\]
We may now integrate-by-parts eq. (9.1.4) and use the curved space Gauss' theorem in eq. (7.4.38).

\[ I = \int_{\partial \mathcal{D}} \partial_i \Psi^i \nabla^i \Psi - \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \cdot \Psi^i \nabla^i \nabla^j \Psi. \] (9.1.6)

Remember from eq. (7.4.29) that \( d_{\mathcal{D}} \xi = d_{\mathcal{D}} \xi \sqrt{H(\xi)} n^i \nabla^i \Psi \), where \( \xi(\vec{x}) \) parametrizes the boundary \( \partial \mathcal{D} \); \( H(\xi) \) is the determinant of the induced metric on \( \partial \mathcal{D} \) so that \( d_{\mathcal{D}} \xi \sqrt{H(\xi)} n^i \nabla^i \Psi \) is its infinitesimal area element and \( n^i(\partial \mathcal{D}) \) its unit outward normal. If either \( \psi(\partial \mathcal{D}) \) or \( n^i \partial_i \psi(\partial \mathcal{D}) \) is specified, therefore, the first term on the right hand side of eq. (9.1.6) is zero – since \( \Psi(\partial \mathcal{D}) = \psi_1(\partial \mathcal{D}) - \psi_2(\partial \mathcal{D}) \) and \( n^i \partial_i \psi(\partial \mathcal{D}) = n^i \partial_i \psi_1(\partial \mathcal{D}) - n^i \partial_i \psi_2(\partial \mathcal{D}) \). The second term is zero too, since

\[-\nabla_i \nabla^i \Psi = -\nabla_i \nabla^i (\psi_1 - \psi_2) = J - J = 0. \] (9.1.7)

But we have just witnessed how \( I \) is itself the integral, over the domain, of the sum of squares of \( |\nabla \hat{a} \Psi| \). The only way summing squares of something is zero is that something is identically zero.

\[ \nabla \hat{a} \Psi = \varepsilon \hat{a} \partial^i \Psi = 0, \] (everywhere in \( \mathcal{D} \)). (9.1.8)

Viewing the \( \varepsilon \hat{a} \) as a vector field, so \( \nabla \hat{a} \Psi \) is the derivative of \( \Psi \) in the \( \hat{a} \)th 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., \( \psi_1 \) and \( \psi_2 \) can at most differ by an additive constant. If the normal derivative \( n^i \nabla^i \psi(\partial \mathcal{D}) \) 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 \( \psi(\partial \mathcal{D}) \) were specified on the boundary, \( \Psi(\partial \mathcal{D}) = 0 \) there, and must therefore be zero everywhere in \( \mathcal{D} \). In other words \( \psi_1 = \psi_2 \), and there cannot be more than 1 distinct solution. This completes the proof.

9.1.2 (Negative) Laplacian as a Hermitian operator

We will now demonstrate that the negative Laplacian in some domain \( \mathcal{D} \) can be viewed as a Hermitian operator, if its eigenfunctions obey

\[ \{ \psi_\lambda(\partial \mathcal{D}) = 0 \} \] (Dirichlet) (9.1.9)

or

\[ \{ n^i \nabla^i \psi_\lambda(\partial \mathcal{D}) = 0 \} \] (Neumann), (9.1.10)

or if there are no boundaries\(^{111}\) 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

\[ I \equiv \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \psi_1^i(\vec{x}) \left( -\nabla^2 \psi_2(\vec{x}) \right) = \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \left( -\nabla^2 \psi_1^i(\vec{x}) \right) \psi_2(\vec{x}), \] (9.1.11)

\(^{111}\)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.
for any functions \( \psi_{1,2}(\vec{x}) \) spanned by the eigenfunctions of \( -\nabla^2 \), and therefore satisfy the same boundary conditions. We begin on the left hand side and again employ the curved space Gauss' theorem in eq. (7.4.38).

\[
I = \int_{\partial D} d^{D-1} \Sigma_i \psi_1^\dagger (\nabla_i \psi_2) + \int_D d^D \vec{x} \sqrt{|g|} \nabla_i \psi_1^\dagger \nabla_i \psi_2,
\]

\[
= \int_{\partial D} d^{D-1} \Sigma_i \left\{ \psi_1^\dagger (\nabla_i \psi_2) + \left( \nabla_i \psi_1^\dagger \right) \psi_2 \right\} + \int_D d^D \vec{x} \sqrt{|g|} \left( -\nabla_i \nabla_i \psi_1^\dagger \right) \psi_2, \tag{9.1.12}
\]

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.

**Non-negative eigenvalues**

Let us understand the bounds on the spectrum of the negative Laplacian subject to the Dirichlet (eq. (9.1.9)) or Neumann boundary (eq. (9.1.10)) conditions, or when there are no boundaries. Let \( \psi_\lambda \) be an eigenfunction obeying

\[
-\nabla^2 \psi_\lambda = \lambda \psi_\lambda. \tag{9.1.13}
\]

We have previously argued that

\[
I' = \int_D d^D \vec{x} \sqrt{|g|} \nabla_i \psi_\lambda^\dagger \nabla_i \psi_\lambda
\]

is strictly non-negative. If we integrate-by-parts,

\[
I' = \int_D d^{D-1} \Sigma_i \psi_\lambda^\dagger \nabla_i \psi_\lambda + \int_D d^D \vec{x} \sqrt{|g|} \psi_\lambda^\dagger \left( -\nabla_i \nabla_i \psi_\lambda \right) \geq 0. \tag{9.1.15}
\]

If there are no boundaries – for example, if \( \mathfrak{D} \) is a \( (n \geq 2) \)-sphere (usually denoted as \( S^n \)) – there will be no surface terms; if there are boundaries but the eigenfunctions obey either Dirichlet conditions in eq. (9.1.9) or Neumann conditions in eq. (9.1.10), the surface terms will vanish. In all three cases, we see that the corresponding eigenvalues \{\( \lambda \)\} are strictly non-negative, since \( \int_D d^D \vec{x} \sqrt{|g|} |\psi_\lambda|^2 \geq 0 \):

\[
I' = \lambda \int_D d^D \vec{x} \sqrt{|g|} |\psi_\lambda|^2 \geq 0. \tag{9.1.16}
\]

**Problem 9.1.** Instead of Dirichlet or Neumann boundary conditions, let us allow for mixed (aka Robin) boundary conditions, namely

\[
\alpha \cdot \psi + \beta \cdot n^i \nabla_i \psi = 0 \tag{9.1.17}
\]

on the boundary \( \partial \mathfrak{D} \). Show that the negative Laplacian is Hermitian if we impose

\[
\frac{\alpha}{\alpha^*} = \frac{\beta}{\beta^*}. \tag{9.1.18}
\]

In particular, if \( \alpha \) and \( \beta \) are both real, imposing eq. (9.1.17) automatically yields a Hermitian Laplacian.
9.1.3 Inverse of the negative Laplacian: Green’s function and reciprocity

Given the Dirichlet boundary condition in eq. (9.1.9), i.e., \( \{\psi_\lambda(\partial\mathcal{D}) = 0\} \), we will now understand how to solve Poisson’s equation, through the inverse of the negative Laplacian. Roughly speaking,

\[
-\vec{\nabla}^2 \psi = J \implies \psi = \left(-\vec{\nabla}^2\right)^{-1} J.
\]

(9.1.19)

(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

\[
H = \sum_\lambda \lambda |\lambda\rangle \langle \lambda|, \quad \{\lambda \in \mathbb{R}\},
\]

(9.1.20)

its inverse is

\[
H^{-1} = \sum_\lambda \frac{|\lambda\rangle \langle \lambda|}{\lambda}, \quad \{\lambda \in \mathbb{R}\};
\]

(9.1.21)

we see that the inverse of the negative Laplacian in the position space representation is the following mode expansion involving its eigenfunctions \( \{\psi_\lambda\} \).

\[
G(\vec{x}, \vec{x}') = \frac{1}{-\vec{\nabla}^2} = \sum_\lambda \frac{\psi_\lambda(\vec{x})^{\dagger} \psi_\lambda(\vec{x}')}{\lambda},
\]

(9.1.22)

\[
-\vec{\nabla}^2 \psi_\lambda = \lambda \psi_\lambda, \quad \psi_\lambda(\vec{x}) \equiv \langle \vec{x} | \lambda \rangle.
\]

(9.1.23)

(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 \( \{\psi_\lambda(\partial\mathcal{D}) = 0\} \), the Green’s function itself also obeys Dirichlet boundary conditions:

\[
G(\vec{x} \in \mathcal{D}, \vec{x}') = G(\vec{x}, \vec{x}' \in \mathcal{D}) = 0.
\]

(9.1.24)

The Green’s function \( G \) satisfies the PDE

\[
-\vec{\nabla}^2 G(\vec{x}, \vec{x}') = -\vec{\nabla}^2 G(\vec{x}, \vec{x}') = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{|g(\vec{x})g(\vec{x}')|}},
\]

(9.1.25)

because the negative Laplacian is Hermitian and thus its eigenfunctions obey the following completeness relation (cf. (4.3.24))

\[
\sum_\lambda \psi_\lambda(\vec{x}')^{\dagger} \psi_\lambda(\vec{x}) = \langle \vec{x}' | \left( \sum_\lambda |\lambda\rangle \langle \lambda| \right) |\vec{x}\rangle
\]

\[
= \langle \vec{x}' | \vec{x} \rangle = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{|g(\vec{x})g(\vec{x}')|}}.
\]

(9.1.26)

Eq. (9.1.25) follows from \(-\vec{\nabla}^2 \psi_\lambda = \lambda \psi_\lambda \) and

\[
-\vec{\nabla}^2 G(\vec{x}, \vec{x}') = \sum_\lambda \frac{-\vec{\nabla}^2 \psi_\lambda(\vec{x})^{\dagger} \psi_\lambda(\vec{x}')}{\lambda} = \sum_\lambda \psi_\lambda(\vec{x})^{\dagger} \psi_\lambda(\vec{x}')^{\dagger},
\]

(9.1.27)
\[-\vec{\nabla}_x^2 G(\vec{x}, \vec{x'}) = \sum_\lambda \psi_\lambda(\vec{x}) \left( -\vec{\nabla}_{\vec{x'}}^2 \psi_\lambda(\vec{x'}) \right) \frac{1}{\lambda} = \sum_\lambda \psi_\lambda(\vec{x}) \psi_\lambda(\vec{x'})^\dagger. \] (9.1.28)

Because the $\delta^{(D)}$-functions on the right hand side of eq. (9.1.25) is the (position representation) of the identity operator, the Green’s function itself is really the inverse of the negative Laplacian.

**Physically speaking** these $\delta$-functions also lend eq. (9.1.25) to 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.

\[ \psi_{\vec{k}}(\vec{x}) = \frac{e^{i\vec{k} \cdot \vec{x}}}{(2\pi)^{D/2}}, \quad -\vec{\nabla}_x^2 \psi_{\vec{k}}(\vec{x}) = \vec{k}^2 \psi_{\vec{k}}(\vec{x}). \] (9.1.29)

Because we know the integral representation of the $\delta$-function, eq. (9.1.26) now reads

\[ \int_{\mathbb{R}^D} \frac{d^D\vec{k}}{(2\pi)^D} e^{i\vec{k} \cdot (\vec{x} - \vec{x'})} = \delta^{(D)}(\vec{x} - \vec{x'}). \] (9.1.30)

Through eq. (9.1.22), we may write down the integral representation of the inverse of the negative Laplacian in Euclidean $D$-space.

\[ G(\vec{x}, \vec{x'}) = \int_{\mathbb{R}^D} \frac{d^D\vec{k}}{(2\pi)^D} e^{i\vec{k} \cdot (\vec{x} - \vec{x'})} = \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2} |\vec{x} - \vec{x'}|^{D-2}}. \] (9.1.31)

In 3D, this result simplifies to the (hopefully familiar) result

\[ G_3(\vec{x}, \vec{x'}) = \frac{1}{4\pi |\vec{x} - \vec{x'}|}. \] (9.1.32)

**Boundaries & Method of Images** Suppose we now wish to solve the Green’s function $G_D(\Omega)$ of the negative Laplacian in a finite domain of flat space, $\Omega \subset \mathbb{R}^D$. One may view $G_D(\Omega)$ as the sum of its counterpart in infinite $\mathbb{R}^D$ plus a term that is a homogeneous solution $H_D(\Omega)$ in the finite domain $\Omega$, such that the desired boundary conditions are achieved on $\partial \Omega$. Namely,

\[ G_D(\vec{x}, \vec{x'}; \Omega) = \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2} |\vec{x} - \vec{x'}|^{D-2}} + H(\vec{x}, \vec{x'}; \Omega), \]

\[ -\vec{\nabla}_x^2 G_D(\vec{x}, \vec{x'}; \Omega) = -\vec{\nabla}_{\vec{x'}}^2 G_D(\vec{x}, \vec{x'}; \Omega) = \delta^{(D)}(\vec{x} - \vec{x'}), \quad \text{(Cartesian coordinates)} \]

\[ -\vec{\nabla}_x^2 H_D(\vec{x}, \vec{x'}; \Omega) = -\vec{\nabla}_{\vec{x'}}^2 H_D(\vec{x}, \vec{x'}; \Omega) = 0, \quad \vec{x}, \vec{x'} \in \Omega. \] (9.1.33)

If Dirichlet boundary conditions are desired, we would demand

\[ \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2} |\vec{x} - \vec{x'}|^{D-2}} + H(\vec{x}, \vec{x'}; \Omega) = 0 \] (9.1.34)

whenever $\vec{x} \in \partial \Omega$ or $\vec{x'} \in \partial \Omega$. 

360
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 \( \{ x^1, \ldots, x^{D-1} \} \) run over the real line. We further want the boundary condition

\[
G_D(x^D = 0) = G_D(x'^D = 0) = 0. \tag{9.1.35}
\]

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

\[
G_D(\vec{x}, \vec{x}' ; \mathcal{D}) = \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2}|\vec{x} - \vec{x}'|^{D-2}} - \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2}|\vec{\xi}|^{D-2}}, \tag{9.1.36}
\]

\[|\vec{\xi}| \equiv \sqrt{\sum_{j=1}^{D-1} (x^j - x'^j)^2 + (x^D + x'^D)^2}, \quad x^D, x'^D \geq 0.\]

Mathematically speaking, when the negative Laplacian is applied to the second term in eq. (9.1.36), 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 \( \delta \)-function can be set to zero. Hence, the second term is a homogeneous solution when attention is restricted to \( x^D \geq 0 \).

**Reciprocity** We will also now show that the Green’s function itself is a Hermitian object, in that

\[
G(\vec{x}, \vec{x}')^\dagger = G(\vec{x}', \vec{x}) = G(\vec{x}, \vec{x}'). \tag{9.1.37}
\]

The first equality follows from the real positive nature of the eigenvalues, as well as the mode expansion in eq. (9.1.22)

\[
G(\vec{x}, \vec{x}')^* = \sum_\lambda \frac{\psi_\lambda(\vec{x}')\psi_\lambda(\vec{x})^\dagger}{\lambda} = G(\vec{x}', \vec{x}). \tag{9.1.38}
\]

The second requires considering the sort of integrals we have been examining in this section.

\[
I(x, x') \equiv \int_\mathcal{D} d^D \vec{x}'' \sqrt{|g(\vec{x}'')|} \left\{ G(\vec{x}, \vec{x}'') \left( -\nabla_{2'2'}^2 G(\vec{x}', \vec{x}'') - G(\vec{x}', \vec{x}'') \left( -\nabla_{2'2'}^2 G(\vec{x}, \vec{x}'') \right) \right) \right\}. \tag{9.1.39}
\]

Using the PDE obeyed by \( G \),

\[
I(x, x') = G(\vec{x}, \vec{x}') - G(\vec{x}', \vec{x}). \tag{9.1.40}
\]

We may integrate-by-parts too.

\[
I(x, x') = \int_{\partial\mathcal{D}} d^{D-1} \Sigma_{\nu} \left\{ G(\vec{x}, \vec{x}'')(-\nabla_{\nu}^2)G(\vec{x}', \vec{x}'') - G(\vec{x}', \vec{x}'')(-\nabla_{\nu}^2)G(\vec{x}, \vec{x}'') \right\}
\]

361
\[ + \int d^D \vec{x}' \sqrt{|g(\vec{x}')}} \left\{ \nabla^\nu G(\vec{x}, \vec{x}') \nabla^\nu G(\vec{x}, \vec{x}''') - \nabla^\nu G(\vec{x}', \vec{x}''') \nabla^\nu G(\vec{x}, \vec{x}'') \right\}. \quad (9.1.41) \]

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 9.2.** Verify directly that the Green’s function solution in eq. (9.1.36) obeys reciprocity.

We close this subsection with the following remarks.

**Isolated zero eigenvalue implies non-existence of inverse** Within a finite domain \( \Omega \), we see that the Neumann boundary conditions \( \{ n^i \nabla_i \psi_\lambda(\partial \Omega) = 0 \} \) imply there must be a zero eigenvalue; for, the \( \psi_0 = \text{constant} \) is the corresponding eigenvector, whose normal derivative on the boundary is zero:

\[ -\vec{\nabla}^2 \psi_0 = -\frac{\partial_i \left( \sqrt{|g|} g^{ij} \partial_j \psi_0 \right)}{\sqrt{|g|}} = 0 \cdot \psi_0. \quad (9.1.42) \]

As long as this is an isolated zero – i.e., there are no eigenvalues continuously connected to \( \lambda = 0 \) – this mode will contribute a discrete term in the mode sum of eq. (9.1.22) that yields a \( 1/0 \) infinity. That is, the inverse of the Laplacian does not make sense if there is an isolated zero mode.\footnote{In the infinite flat \( \mathbb{R}^D \) case above, we have seen the \( \{ \exp(ik \cdot \vec{x}) \} \) are the eigenfunctions and hence there is also a zero mode, gotten by setting \( \vec{k} \to 0 \). However the inverse does exist because the mode sum of eq. (9.1.22) is really an integral, and the integration measure \( d^D \vec{k} \) ensures convergence of the integral.}

**Discontinuous first derivatives** Because it may not be apparent from the mode expansion in eq. (9.1.22), it is worth 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, \( \delta \)-functions on the right hand side of eq. (9.1.25). For Green’s functions in a finite domain \( \Omega \), there are potentially additional discontinuities when both \( \vec{x} \) and \( \vec{x}' \) are near the boundary of the domain \( \partial \Omega \).

### 9.1.4 Kirchhoff integral theorem and Dirichlet boundary conditions

Within a finite domain \( \Omega \) 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. (9.1.1).

Suppose we have specified the field on the boundary \( \psi(\partial \Omega) \). 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., \( \{ \psi_\lambda(\partial \Omega) = 0 \} \), so that the resulting Green’s function obey eq. (9.1.24). The solution to Poisson’s equation within the domain \( \Omega \) can now be solved in terms of \( G \), the source \( J \), and its boundary values \( \psi(\partial \Omega) \) through the following Kirchhoff integral representation:

\[
\psi(\vec{x}) = \int_D d^D \vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}') - \int_{\partial \Omega} d^{D-1} \Sigma_i \nabla^i G(\vec{x}, \vec{x}') \psi(\vec{x}'). \quad (9.1.43)
\]
If there are no boundaries, then the boundary integral terms in eq. (9.1.43) 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 $\psi$ 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. (9.1.43)** Let us now consider the following integral

$$I(\vec{x} \in \mathcal{D}) \equiv \int_{\mathcal{D}} d^D\vec{x}' \sqrt{|g(\vec{x}')|} \left\{ G(\vec{x}, \vec{x}') \left( -\nabla^2 \psi(\vec{x}') \right) - \left( -\nabla^2 \psi(\vec{x}') \right) \right\} \psi(\vec{x}')$$

(9.1.44)

If we use the equations (9.1.1) and (9.1.25) obeyed by $\psi$ and $G$ respectively, we obtain immediately

$$I(\vec{x}) = \int_{\mathcal{D}} d^D\vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}') - \psi(\vec{x}).$$

(9.1.45)

On the other hand, we may integrate-by-parts,

$$I(\vec{x}) = \int_{\partial\mathcal{D}} d^{D-1}\Sigma_v \left\{ G(\vec{x}, \vec{x}') \left( -\nabla^i \psi(\vec{x}') \right) - \left( -\nabla^i \psi(\vec{x}') \right) \right\}$$

$$+ \int_{\mathcal{D}} d^D\vec{x}' \sqrt{|g(\vec{x}')|} \left\{ \nabla^i G(\vec{x}, \vec{x}') \nabla^i \psi(\vec{x}') - \nabla^i G(\vec{x}, \vec{x}') \nabla^i \psi(\vec{x}') \right\}.$$  

(9.1.46)

The second line cancels. Combining equations (9.1.45) and (9.1.46) then hands us the following Kirchhoff representation:

$$\psi(\vec{x} \in \mathcal{D}) = \int_{\partial\mathcal{D}} d^{D-1}\Sigma_v \left\{ G(\vec{x}, \vec{x}') \left( \nabla^i \psi(\vec{x}') \right) - \left( \nabla^i \psi(\vec{x}') \right) \right\}$$

$$+ \int_{\mathcal{D}} d^D\vec{x}' \sqrt{|g(\vec{x}')|} G(\vec{x}, \vec{x}') J(\vec{x}').$$

(9.1.47)

(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. (9.1.24)), the first term on the right hand side of the first line drops out and we obtain eq. (9.1.43).

**Problem 9.3. Dirichlet B.C. Variation Principle** In a finite domain (where $\int_{\mathcal{D}} d^D\vec{x} \sqrt{|g|} < \infty$), let all fields vanish on the boundary $\partial\mathcal{D}$ and denote the smallest non-zero eigenvalue of the negative Laplacian $-\nabla^2$ as $\lambda_0$. Let $\psi$ be an arbitrary function obeying the same boundary conditions as the eigenfunctions of $-\nabla^2$. For this problem, assume that the spectrum of the negative Laplacian is discrete. Prove that

$$\frac{\int_{\mathcal{D}} d^D\vec{x} \sqrt{|g|} \nabla_i \psi \nabla^i \psi}{\int_{\mathcal{D}} d^D\vec{x} \sqrt{|g|} |\psi|^2} \geq \lambda_0.$$  

(9.1.48)

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 $\psi$ that vanishes on $\mathcal{D}$ to derive an upper bound for the lowest eigenvalue of the negative Laplacian.

**Hint:** Expand $\psi$ as a superposition of the eigenfunctions of $-\nabla^2$. Then integrate-by-parts one of the $\nabla^i$ in the integrand. □
Example Suppose, within a finite 1D box, \( x \in [0, L] \) we are provided a real field \( \psi \) obeying
\[
\psi(x = 0) = \alpha, \quad \psi(x = L) = \beta
\]
without any external sources. You can probably solve this 1D Poisson’s equation \((-\partial^2_x \psi = 0)\) right away; it is a straight line:
\[
\psi(0 \leq x \leq L) = \alpha + \frac{\beta - \alpha}{L} x.
\]
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,
\[
\langle x | n \rangle = \sqrt{\frac{2}{L}} \sin \left( \frac{n\pi}{L} x \right), \quad n \in \{1, 2, 3, \ldots \}, \quad \sum_{n=1}^{\infty} \langle x | n \rangle \langle n | x' \rangle = \delta(x - x'),
\]
\[
-\partial^2_x \langle x | n \rangle = \left( \frac{n\pi}{L} \right)^2 \langle x | n \rangle.
\]
The mode sum expansion of the Green’s function in eq. \(9.1.22\) is
\[
G(x, x') = \frac{2}{L} \sum_{n=1}^{\infty} \left( \frac{n\pi}{L} \right)^{-2} \sin \left( \frac{n\pi}{L} x \right) \sin \left( \frac{n\pi}{L} x' \right).
\]
The \( J \) term in eq. \(9.1.43\) 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.
\[
\psi(x') = \partial_x G(x, x' = 0)\psi(x' = 0) - \partial_x G(x, x' = L)\psi(x' = L)
\]
\[
= -\frac{2}{L} \sum_{n=1}^{\infty} \frac{L}{n\pi} \sin \left( \frac{n\pi}{L} x \right) \left[ \cos \left( \frac{n\pi}{L} x' \right) \psi(x') \right]_{x'=L}
\]
\[
= -\sum_{n=1}^{\infty} \frac{2}{n\pi} \sin \left( \frac{n\pi}{L} x \right) (\frac{\pi}{L} \cdot \beta - \alpha)
\]
We may check this answer in the following way. Because the solution in eq. \(9.1.53\) is odd under \( x \rightarrow -x \), let us we extend the solution in the following way:
\[
\psi_\infty(-L \leq x \leq L) = \alpha + \frac{\beta - \alpha}{L} x, \quad 0 \leq x \leq L,
\]
\[
= -\left( \alpha + \frac{\beta - \alpha}{L} x \right), \quad -L \leq x < 0.
\]
We will then extend the definition of \( \psi_\infty \) by imposing periodic boundary conditions, \( \psi_\infty(x+2L) = \psi_\infty(x) \). This yields the Fourier series
\[
\psi_\infty(x) = \sum_{\ell=-\infty}^{+\infty} C_\ell e^{i\frac{2\pi \ell}{L} x}.
\]
Multiplying both sides by \(\exp(-i(\pi n/L)x)\) and integrating over \(x \in [-L, L]\),

\[
C_n = \int_{-L}^{L} \psi_\infty(x) e^{-i\pi n x} \frac{dx}{2L} = \int_{-L}^{L} \psi_\infty(x) \left( \cos\left(\frac{\pi n}{L} x\right) - i \sin\left(\frac{\pi n}{L} x\right) \right) \frac{dx}{2L} = -i \int_0^L \left( \alpha + \beta - \frac{\alpha}{L} x \right) \sin\left(\frac{\pi n}{L} x\right) \frac{dx}{L} = \frac{i}{\pi n} ((-)^n \beta - \alpha). \tag{9.1.56}
\]

Putting this back to into the Fourier series,

\[
\psi_\infty(x) = i \sum_{n=1}^{\infty} \frac{1}{\pi n} \left\{ ((-)^n \beta - \alpha) e^{i\pi n x} - ((-)^n \beta - \alpha) e^{-i\pi n x} \right\} = -\sum_{n=1}^{\infty} \frac{2}{\pi n} ((-)^n \beta - \alpha) \sin\left(\frac{\pi n}{L} x\right). \tag{9.1.57}
\]

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 9.4.** 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

\[
\psi(0, y) = \varphi_1(y), \quad \psi(L_1, y) = \varphi_2(y), \quad \psi(x, 0) = \rho_1(x), \quad \psi(x, L_2) = \rho_2(x). \tag{9.1.58}
\]

Write the solution as a mode sum, using the eigenfunctions

\[
\psi_{m,n}(x, y) \equiv \langle x, y | m, n \rangle = \frac{2}{\sqrt{L_1 L_2}} \sin\left(\frac{\pi m}{L_1} x\right) \sin\left(\frac{\pi n}{L_2} y\right). \tag{9.1.59}
\]

Hint: your answer will involve 1D integrals on the 4 boundaries of the rectangle.

\[
\square
\]

### 9.2 Laplacians and their spectra

Let us recall our discussions from both linear algebra and differential geometry. Given a (Euclidean signature) metric

\[
d\ell^2 = g_{ij}(\vec{x})dx^i dx^j, \tag{9.2.1}
\]

the Laplacian acting on a scalar \(\psi\) can be written as

\[
\nabla^2 \psi \equiv \nabla_i \nabla^i \psi = \frac{\partial_i \left( \sqrt{|g|} g^{ij} \partial_j \psi \right)}{\sqrt{|g|}}, \tag{9.2.2}
\]

where \(\sqrt{|g|}\) is the square root of the determinant of the metric.
Now we turn to the primary goal of this section, to study the eigenvector/value problem
\[ -\nabla^2 \psi_{\lambda}(\vec{x}) = -\nabla^2 \langle \vec{x} | \lambda \rangle = \lambda \langle \vec{x} | \lambda \rangle. \] (9.2.3)

If these eigenfunctions are normalized to unit length, namely
\[ \int_{\mathbb{D}} d^D \vec{x} \langle \lambda | \vec{x} \rangle \langle \vec{x} | \lambda' \rangle = \delta_{\lambda \lambda'}, \] (9.2.4)

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
\[ \sum_{\lambda} \langle \vec{x} | \lambda \rangle \langle \lambda | \vec{x}' \rangle = \delta^{(D)}(\vec{x} - \vec{x}'). \] (9.2.5)

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.

### 9.2.1 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 \(-\nabla^2\) with eigenvalues \( \{ k^2 | -\infty < k < \infty \} \). This is a coordinate invariant statement, since the \( \psi \) and Laplacian in eq. (9.2.3) are coordinate scalars. Also notice that the eigenvalue/vector equation (9.2.3) 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,
\[ \psi_{\vec{k}}(\vec{x}) = \langle \vec{x} | \vec{k} \rangle = e^{i\vec{k} \cdot \vec{x}} = \prod_{j=1}^{D} e^{ik_j x_j}, \quad \vec{k}^2 = \delta^{ij} k_i k_j = \sum_{i=1}^{D} (k_i)^2 \equiv \vec{k}^2, \] (9.2.6)

with completeness relations (cf. eq. (9.1.26)) given by
\[ \int_{\mathbb{R}^D} d^D \vec{k} \langle \vec{k} | \vec{k} \rangle \langle \vec{k} | \vec{k}' \rangle = (2\pi)^D \delta^{(D)} \left( \vec{k} - \vec{k}' \right), \] (9.2.7)
\[ \int_{\mathbb{R}^D} d^D \vec{k} \langle \vec{x} | \vec{k} \rangle \langle \vec{k} | \vec{x} \rangle = \delta^{(D)}(\vec{x} - \vec{x}'). \] (9.2.8)

**Translation symmetry and degeneracy** For a fixed \( 1 \leq j \leq D \), notice the translation operator in the \( j \)th Cartesian direction, namely \(-i\partial_j \equiv -i\partial / \partial x^j \) commutes with \(-\nabla^2\). The translation operators commute amongst themselves too. This is why one can simultaneously diagonalize the Laplacian, and all the \( D \) translation operators.
\[ -i\partial_j \langle \vec{x} | k^2 \rangle = k_j \langle \vec{x} | k^2 \rangle \] (9.2.9)
In fact, we see that the eigenvector of the Laplacian $|k^2\rangle$ can be viewed as a tensor product of the eigenstates of $P_j$.

\[
|k^2\rangle = |k_1\rangle \otimes |k_2\rangle \otimes \cdots \otimes |k_D\rangle \quad (9.2.10)
\]

\[
\langle \vec{x} | k^2 \rangle = \left( \langle x^1 | \otimes \cdots \otimes \langle x^D | \right) (|k_1\rangle \otimes \cdots \otimes |k_D\rangle) = \langle x^1 | k_1 \rangle \langle x^2 | k_2 \rangle \cdots \langle x^D | k_D \rangle = \prod_{j=1}^D e^{ik_j x_j} \quad (9.2.11)
\]

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 $\mathbb{R}^D$.

### 9.2.2 1 Dimension

**Infinite Flat Space**

In one dimension, the metric is

\[
d\ell^2 = dz^2, \quad (9.2.12)
\]

for $z \in \mathbb{R}$, and eq. (9.2.6) reduces to

\[
-\vec{\nabla}_1^2 \psi_k(z) = -\partial_z^2 \psi_k(z) = k^2 \psi_k(z), \quad \langle z | k \rangle \equiv \psi_k(z) = e^{ikz}; \quad (9.2.13)
\]

and their completeness relation (cf. eq. (9.1.26)) is

\[
\int_{-\infty}^{\infty} \frac{dk}{2\pi} \langle z | k \rangle \langle k | z' \rangle = \delta(z - z'). \quad (9.2.14)
\]

**Periodic infinite space**

If the 1D space obeys periodic boundary conditions, with period $L$, we have instead

\[
-\vec{\nabla}_1^2 \psi_m(z) = -\partial_z^2 \psi_m(z) = \left( \frac{2\pi m}{L} \right)^2 \psi_m(z),
\]

\[
\langle z | m \rangle \equiv \psi_m(z) = L^{-1/2} e^{i \frac{2\pi m}{L} z}, \quad m = 0, \pm 1, \pm 2, \ldots \quad (9.2.15)
\]

The orthonormal eigenvectors obey

\[
\int_0^L dz \langle m | z \rangle \langle z | m' \rangle = \delta_{m'm}, \quad \langle z | m \rangle = L^{-1/2} e^{i \frac{2\pi m}{L} z}; \quad (9.2.16)
\]

while their completeness relation (eq. (9.1.26)) reads, for $0 \leq z, z' \leq L$,

\[
\sum_{m=-\infty}^{\infty} \langle z | m \rangle \langle m | z' \rangle = \frac{1}{L} \sum_{m=-\infty}^{\infty} e^{i \frac{2\pi m}{L} (z-z')} = \delta(z - z'). \quad (9.2.17)
\]

---

\(^{113}\)One dimensional space(time)s are always flat – the Riemann tensor is identically zero.
**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
\[
d\ell^2 = (d\phi)^2, \quad \phi \in [0, 2\pi).
\] (9.2.18)

We may then bring over the results from the previous discussion.
\[
-\nabla^2_{\phi} \psi_m(\phi) = -\partial_{\phi}^2 \psi_m(\phi) = m^2 \psi_m(\phi), \quad \langle \phi | m \rangle \equiv \psi_m(\phi) = (2\pi)^{-1/2} e^{im\phi}, \quad m = 0, \pm 1, \pm 2, \ldots.
\] (9.2.19)

The orthonormal eigenvectors obey
\[
\int_0^{2\pi} d\phi \langle m | \phi \rangle \langle \phi | m' \rangle = \delta_m^{m'}, \quad \langle \phi | m \rangle = (2\pi)^{-1/2} e^{im\phi}.
\] (9.2.20)

while their completeness relation reads, for \( 0 \leq z, z' \leq L \),
\[
\sum_{m=-\infty}^{\infty} \langle \phi | m \rangle \langle m | \phi' \rangle = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im(\phi - \phi')} = \delta(\phi - \phi').
\] (9.2.21)

**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
\[
\langle \phi | P_{\phi} | \psi \rangle = -i\partial_{\phi} \langle \phi | \psi \rangle
\] (9.2.22)

for any state \(|\psi\rangle\), we shall see it is Hermitian with discrete spectra:
\[
P_{\phi} |m\rangle = m |m\rangle, \quad m = 0, \pm 1, \pm 2, \pm 3, \ldots,
\] (9.2.23)
\[
\langle \phi | m \rangle = e^{im\phi}/\sqrt{2\pi}.
\] (9.2.24)

Given arbitrary states \(|\psi_1, 2\rangle\),
\[
\langle \psi_1 | P_{\phi} | \psi_2 \rangle = \int_0^{2\pi} d\phi \langle \psi_1 | \phi \rangle (-i\partial_{\phi} \langle \phi | \psi_2 \rangle)
\] (9.2.25)
\[
= [-i \langle \psi_1 | \phi \rangle \langle \phi | \psi_2 \rangle]_{\phi=2\pi}^{\phi=0} + \int_0^{2\pi} d\phi \langle i\partial_{\phi} \langle \psi_1 | \phi \rangle \rangle \langle \phi | \psi_2 \rangle.
\]

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,
\[
\langle \psi_1 | P_{\phi} | \psi_2 \rangle = \int_0^{2\pi} d\phi (-i\partial_{\phi} \langle \phi | \psi_1 \rangle^* \langle \phi | \psi_2 \rangle = \int_0^{2\pi} d\phi \langle \phi | P_{\phi} \psi_1 \rangle \langle \phi | \psi_2 \rangle
\]
\[
= \int_0^{2\pi} d\phi \langle \psi_1 | P_{\phi}^\dagger \phi \rangle \langle \phi | \psi_2 \rangle = \langle \psi_1 | P_{\phi}^\dagger | \psi_2 \rangle.
\] (9.2.26)
We must therefore have

\[ \langle \phi | e^{-i\theta P_\phi} | \psi \rangle = e^{-i\theta \partial_\phi} \langle \phi | \psi \rangle = \langle \phi - \theta | \psi \rangle. \quad (9.2.27) \]

Any function on a circle can be expanded in the eigenstates of \( P_\phi \), which in turn can be expressed through its position representation.

\[
|\psi\rangle = \sum_{m=-\infty}^{+\infty} |m\rangle \langle m| \psi \rangle = \sum_{m=-\infty}^{+\infty} \int_0^{2\pi} d\phi \langle m| \phi \rangle \langle \phi | m \rangle \langle m| \psi \rangle = \sum_{m=-\infty}^{+\infty} \int_0^{2\pi} \frac{d\phi}{\sqrt{2\pi}} \langle \phi | m \rangle \langle m| \psi \rangle e^{im\phi},
\]

\[
\langle m| \psi \rangle = \int_0^{2\pi} d\phi' \langle m| \phi' \rangle \langle \phi' | \psi \rangle = \int_0^{2\pi} \frac{d\phi'}{\sqrt{2\pi}} e^{-im\phi'} \psi(\phi').
\]

This is nothing but the Fourier series expansion of \( \psi(\phi) \).

### 9.2.3 2 Dimensions ⊙ Separation-of-Variables for PDEs

**Flat Space, Cylindrical Coordinates**

The 2D flat metric in cylindrical coordinates reads

\[ dl^2 = dr^2 + r^2 d\phi^2, \quad r \geq 0, \quad \phi \in [0, 2\pi), \quad \sqrt{|g|} = r. \quad (9.2.29) \]

The negative Laplacian is therefore

\[
-\vec{\nabla}_2^2 \varphi_k(r, \phi) = -\frac{1}{r} \left( \partial_r (r \partial_r \varphi_k) + \frac{1}{r} \partial_\phi^2 \varphi_k \right) = -\left\{ \frac{1}{r} \partial_r (r \partial_r \varphi_k) + \frac{1}{r^2} \partial_\phi^2 \varphi_k \right\}. \quad (9.2.30)
\]

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

\[ \langle \vec{x} | \vec{k} \rangle = \exp(i\vec{k} \cdot \vec{x}) = \exp(ikr \cos(\phi - \phi_k)), \quad k \equiv |\vec{k}|, \quad r \equiv |\vec{x}|; \quad (9.2.32) \]

because the Cartesian components of \( \vec{k} \) and \( \vec{x} \) are

\[ k_i = k (\cos \phi_k, \sin \phi_k) \quad x^i = r (\cos \phi, \sin \phi). \quad (9.2.33) \]

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 (4.5.165) and (4.5.166),

\[ \langle \vec{x} | \vec{k} \rangle = \sum_{m=-\infty}^{+\infty} \chi_m(kr) e^{im(\phi - \phi_k)} \frac{e^{i kr \cos \phi_k}}{\sqrt{2\pi}}. \quad (9.2.34) \]

Setting \( \phi - \phi_k \to \phi'' \), multiplying both sides with \( \exp(-im\phi'')/\sqrt{2\pi} \), followed by integrating \( \phi'' \) over the unit circle,

\[ \chi_m(kr) = \int_0^{2\pi} \frac{d\phi''}{\sqrt{2\pi}} e^{ikr \cos \phi''} e^{-im\phi''} \quad (9.2.35) \]
(In the last line, we have used the fact that the integrand is itself a periodic function of $\phi'$ with period $2\pi$ 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])

$$J_m(z) = \int_{-\pi}^{\pi} \frac{d\phi'}{2\pi} e^{iz\sin \phi'} e^{-im\phi'}, \quad m \in \{0, \pm1, \pm2, \ldots \},$$ \hspace{1cm} (9.2.37)

$$J_{-m}(z) = (-)^m J_m(z).$$ \hspace{1cm} (9.2.38)

As an aside, let us record that $J_\nu(z)$ also has a series representation

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-)^k (z/2)^{2k}}{k!\Gamma(\nu+k+1)};$$ \hspace{1cm} (9.2.39)

and the large argument asymptotic expansion

$$J_{\nu}(z \gg \nu) \sim \sqrt{\frac{2}{\pi z}} \cos \left( z \mp \frac{\pi}{2} \nu - \frac{\pi}{4} \right).$$ \hspace{1cm} (9.2.40)

Utilizing eq. (9.2.37) in eq. (9.2.36), we see the plane wave in eq. (9.2.34) admits the cylindrical coordinate expansion:

$$\langle \vec{x} | \vec{k} \rangle = \exp(i \vec{k} \cdot \vec{x}) = \exp(ikr \cos(\phi - \phi_k)), \quad k \equiv |\vec{k}|, \quad r \equiv |\vec{x}|$$

$$= \sum_{m=-\infty}^{\infty} i^m J_m(kr)e^{im(\phi - \phi_k)}.$$ \hspace{1cm} (9.2.41)

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 $-\nabla^2$. 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,

$$\frac{(2\pi)^2 \delta(k-k')\delta(\phi_k - \phi_{k'})}{\sqrt{k k'}} = \int_{\mathbb{R}^2} d^2 x \exp(i(\vec{k} - \vec{k}') \cdot \vec{x})$$ \hspace{1cm} (9.2.42)

$$= \sum_{m,m'=\infty}^{+\infty} \int_0^\infty dr \cdot \int_0^{2\pi} d\phi \cdot (-i)^{m'} J_m(kr)J_{m'}(k'r)e^{im(\phi - \phi_k)} e^{-im'(\phi - \phi_{k'})}$$

$$= (2\pi) \sum_{m=-\infty}^{+\infty} \int_0^\infty dr \cdot r J_m(kr)J_m(k'r)e^{im(\phi_{k'} - \phi_k)}.$$ 

**Problem 9.5.** The left hand side of eq. (9.2.42) is $(2\pi)^2 \delta^{(2)}(\vec{k} - \vec{k}')$ if we used Cartesian coordinates in $\vec{k}$–space – see eq. (9.2.7). Can you explain why it takes the form it does? Hint: Use cylindrical coordinates in $k$–space and refer to eq. (9.1.26). \hfill \square
We now replace the $\delta(\phi - \phi_k)$ on the left hand side of eq. (9.2.42) with the completeness relation in eq. (9.2.17), 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

$$\sum_{m=-\infty}^{\infty} \frac{\delta(k - k')}{\sqrt{kk'}} e^{im(\phi_k - \phi_{k'})} = \sum_{m=-\infty}^{\infty} \int_0^\infty dr \cdot r J_m(kr) J_m(k'r) e^{im(\phi_{k'} - \phi_k)}. \quad (9.2.43)$$

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:

$$\int_0^\infty dr \cdot r J_m(kr) J_m(k'r) = \frac{\delta(k - k')}{\sqrt{kk'}}, \quad (9.2.44)$$

$$\int_0^\infty dk \cdot k J_m(kr) J_m(k'r) = \frac{\delta(r - r')}{\sqrt{rr'}}. \quad (9.2.45)$$

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

$$-\nabla^2 \langle r, \phi| k, m \rangle = k^2 \langle r, \phi| k, m \rangle, \quad \langle r, \phi| k, m \rangle \equiv J_m(kr) \frac{\exp(im\phi)}{\sqrt{2\pi}},$$

$$m = 0, \pm 1, \pm 2, \pm 3, \ldots \quad (9.2.46)$$

The eigenvectors are normalized as

$$\int_0^\infty dr \cdot r \int_0^{2\pi} d\phi \langle k, m| r, \phi \rangle \langle r, \phi| k', m' \rangle = \delta_m^m \frac{\delta(k - k')}{\sqrt{kk'}}. \quad (9.2.47)$$

Rotational symmetry and degeneracy Note that $-i\partial_\phi$ is the translation operator in the azimuthal direction (≡ 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 (9.2.46), followed by using the explicit expression in eq. (9.2.30) and the angular eigenvalue/vector equation $\partial^2_\phi \exp(im\phi) = -m^2 \exp(im\phi)$, this hands us

$$k^2 J_m(kr) = -\left\{ \frac{1}{r} \partial_r (r \partial_r J_m(kr)) - \frac{m^2}{r^2} J_m(kr) \right\}. \quad (9.2.48)$$

Let us then re-scale $\rho \equiv kr$, where $k \equiv |\vec{k}|$, so that $\partial_r = k \partial_\rho$.

$$\rho^2 \cdot J''(\rho) + \rho \cdot J'(\rho) + (\rho^2 - m^2)J(\rho) = 0 \quad (9.2.49)$$

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 9.6.** 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** \( S^2 \), **Separation-Of-Variables**, and the Spherical Harmonics

The 2-sphere of radius \( R \) can be viewed as a curved surface embedded in 3D flat space parametrized as

\[
\vec{x}(\vec{\xi} = (\theta, \phi)) = R (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \quad \vec{x}^2 = R^2. \tag{9.2.50}
\]

For concreteness we will consider the case where \( R = 1 \). Its metric is therefore given by

\[
H_{ij}d\xi^i d\xi^j |_{R=1} = \delta_{ij} \partial_i x^i \partial_j x^j d\xi^1 d\xi^2, \quad \sqrt{|H|} = \sin \theta. \tag{9.2.51}
\]

(Or, simply take the 3D flat space metric in spherical coordinates, and set \( dr \to 0 \) and \( r \to 1 \).)

We wish to diagonalize the negative Laplacian on this unit radius 2-sphere. The relevant eigenvector/value equation is

\[
-\vec{\nabla}^2_{S^2} Y(\theta, \phi) = \nu(\nu + 1) Y(\theta, \phi), \tag{9.2.53}
\]

where for now \( \nu \) 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, \( \nu \) 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 \( \tau_i \) depends solely on the \( i \)th variable of the system. Once this has been done – and since no other term now depends on the \( i \)th coordinate so we may vary it without varying others – we may then conclude that each \( \tau_i \) has to be a constant because upon varying this \( i \)th term the entire sum must still remain zero. This in turn leads us to one ODE for every \( \tau_i \). If solutions can be found, we are assured that such an ansatz works.

For the unit radius 2-sphere, we postulate

\[
Y(\theta, \phi) = \Lambda(\theta)\Phi(\phi). \tag{9.2.54}
\]

\[114\]In these notes we focus solely on the spherical harmonics on \( S^2 \); for spherical harmonics in arbitrary dimensions, see [arXiv:1205.3548](arXiv:1205.3548).
First work out the Laplacian explicitly, with $s \equiv \sin \theta$,

$$
- \left\{ \frac{1}{s} \partial_{\theta} (s \partial_{\theta} Y) + \frac{1}{s^2} \partial_{\theta}^2 Y \right\} = - \left\{ \frac{1}{s} \partial_{\theta} (s \partial_{\theta} Y) + \frac{1}{s^2} \nabla_{S^2}^2 Y \right\} = \nu (\nu + 1) Y(\theta, \phi). \quad (9.2.55)
$$

We have identified $\nabla_{S^2}^2 = \partial_{\phi}^2$ to be the Laplacian on the circle, from eq. (9.2.19). 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 multiply the above equation throughout by $s^2$, use the ansatz in eq. (9.2.54), and re-arrange it

$$
\left\{ s \partial_{\theta} (s \partial_{\theta} \Lambda) + s^2 \nu (\nu + 1) \Lambda \right\} + \partial_{\phi}^2 \Phi = 0.
$$

Notice the first term involving the $\{ \ldots \}$ depends only on $\theta$ and not on $\phi$. Whereas the second term $(\partial^2 \Phi)/\Phi$ only depends on $\phi$ and not on $\theta$. This immediately implies both terms must be a constant. For, we may first differentiate both sides with respect to $\theta$,

$$
\partial_{\theta} \left\{ \frac{1}{\Lambda} \left( s \partial_{\theta} (s \partial_{\theta} \Lambda) + s^2 \nu (\nu + 1) \Lambda \right) \right\} = 0 \quad (9.2.58)
$$

and conclude the terms in the curly brackets must be independent of $\theta$. And since they are already independent of $\phi$ by assumption, these terms must be a constant. Similarly, differentiating eq. (9.2.57) with respect to $\phi$,

$$
\partial_{\phi} \left\{ \frac{\partial^2 \Phi}{\Phi} \right\} = 0. \quad (9.2.59)
$$

At this point, we deduce

$$
\frac{1}{\Lambda} \left( s \partial_{\theta} (s \partial_{\theta} \Lambda) + s^2 \nu (\nu + 1) \Lambda \right) = m^2, \quad (9.2.60)
$$

$$
\frac{\partial^2 \Phi}{\Phi} = -m^2. \quad (9.2.61)
$$

Note the relative $-$ sign on the right hand sides of equations (9.2.60) and (9.2.61): this ensures their sum in eq. (9.2.57) is zero. At this point, $m^2$ is an arbitrary constant, but we may see that eq. (9.2.61) is nothing but the simple harmonic oscillator equation: $\partial^2 \Phi + m^2 \Phi = 0$, whose solutions are $\Phi \propto \exp(i m \phi)$. Demanding that $\Phi(\phi + 2\pi) = \Phi(\phi)$ we obtain

$$
\Phi(\phi) \propto \exp(i m \phi), \quad m = 0, \pm 1, \pm 2, \ldots \quad (9.2.62)
$$

Notice this amounts to setting $\Phi$ to be the eigenvector of $\nabla_{S^2}^2$, which we could have guessed from the outset, since the only occurrence of $\partial_{\phi}$ 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. (9.2.60) then reads

$$
\partial_c \left( (1 - c^2) \partial_c \Lambda \right) + \left( \nu (\nu + 1) - \frac{m^2}{1 - c^2} \right) \Lambda = 0. \quad (9.2.63)
$$
This is solved – see eq. 14.2.2 of the NIST page [here] – by the two associated Legendre functions $P^m_\nu(c)$ and $Q^m_\nu(c)$. 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 $P^m_\nu(c)$, set $\nu \equiv \ell$ to be 0 or a positive integer, and have $m$ run from $-\ell$ to $\ell$.

$$\Lambda \propto P^m_\ell(\cos \theta), \quad \ell \in \{0, 1, 2, \ldots \}, \quad m \in \{-\ell, -\ell + 1, \ldots, \ell - 1, \ell\}. \quad (9.2.64)$$

Note that

$$P^0_\ell(x) = P_\ell(x), \quad (9.2.65)$$

where $P_\ell(x)$ is the $\ell$th 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

$$-\hat{\nabla}_S^2 Y^m_\ell(\theta, \phi) = \ell(\ell + 1)Y^m_\ell(\theta, \phi),$$

$$\langle \theta, \phi | \ell, m \rangle = Y^m_\ell(\theta, \phi) = \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} P^m_\ell(\cos \theta) e^{im\phi},$$

$$\ell \in \{0, 1, 2, 3, \ldots \}, \quad m \in \{-\ell, -\ell + 1, \ldots, \ell - 1, \ell\}. \quad (9.2.66)$$

Spherical harmonics should be viewed as “waves” on the 2–sphere, with larger $\ell$ 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.

$Y^0_0 = \frac{1}{\sqrt{4\pi}} \quad (9.2.67)$

The $\ell = 1$ spherical harmonics are:

$$Y^{-1}_1 = \frac{1}{2} \sqrt{\frac{3}{2\pi}} e^{-i\phi} \sin(\theta), \quad Y^0_1 = \frac{1}{2} \sqrt{\frac{3}{\pi}} \cos(\theta), \quad Y^1_1 = -\frac{1}{2} \sqrt{\frac{3}{2\pi}} e^{i\phi} \sin(\theta) \quad (9.2.68)$$

The $\ell = 2$ spherical harmonics are:

$$Y^{-2}_2 = \frac{1}{4} \sqrt{\frac{15}{2\pi}} e^{-2i\phi} \sin^2(\theta), \quad Y^{-1}_2 = \frac{1}{2} \sqrt{\frac{15}{2\pi}} e^{-i\phi} \sin(\theta) \cos(\theta), \quad Y^0_2 = \frac{1}{4} \sqrt{\frac{5}{\pi}} (3 \cos^2(\theta) - 1),$$

$$Y^1_2 = -\frac{1}{2} \sqrt{\frac{15}{2\pi}} e^{i\phi} \sin(\theta) \cos(\theta), \quad Y^2_2 = \frac{1}{4} \sqrt{\frac{15}{2\pi}} e^{2i\phi} \sin^2(\theta) \quad (9.2.69)$$

The $\ell = 3$ spherical harmonics are:

$$Y^{-3}_3 = \frac{1}{8} \sqrt{\frac{35}{\pi}} e^{-3i\phi} \sin^3(\theta), \quad Y^{-2}_3 = \frac{1}{4} \sqrt{\frac{105}{2\pi}} e^{-2i\phi} \sin^2(\theta) \cos(\theta),$$

$$Y^{-1}_3 = \frac{1}{8} \sqrt{\frac{21}{\pi}} e^{-i\phi} \sin(\theta) (5 \cos^2(\theta) - 1), \quad Y^0_3 = \frac{1}{4} \sqrt{\frac{7}{\pi}} (5 \cos^3(\theta) - 3 \cos(\theta)),$$
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.

\[ k_i = k (\sin \theta_k \cdot \cos \phi_k, \sin \theta_k \cdot \sin \phi_k, \cos \theta_k) \equiv \hat{k} \vec{k} \]
\[ x^i = r (\sin \theta \cdot \cos \phi, \sin \theta \cdot \sin \phi, \cos \theta) \equiv r \vec{x} \]

**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.

\[
P_\ell (\hat{k} \cdot \hat{x}) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta, \phi) Y_\ell^m(\theta_k, \phi_k) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta, \phi) Y_\ell^m(\theta_k, \phi_k),
\]

where \( \hat{k} \equiv \vec{k}/k \) and \( \vec{x} \equiv \vec{x}/r \). The second equality follows from the first because the Legendre polynomial is real.

For a fixed direction \( \hat{k} \), note that \( P_\ell (\hat{k} \cdot \hat{x}) \) in eq. (9.2.76) is an eigenvector of the negative Laplacian on the 2-sphere. For, as we have already noted, the eigenvalue equation \(-\nabla_x^2 \psi = \lambda \psi\) is a coordinate scalar. In particular, we may choose coordinates such that \( \hat{k} \) is pointing ‘North’, so that \( \hat{k} \cdot \hat{x} = \cos \theta \), where \( \theta \) is the usual altitude angle. By recalling eq. (9.2.71), we see therefore,

\[
-\nabla_x^2 P_\ell (\hat{k} \cdot \hat{x}) = \ell (\ell + 1) P_\ell (\hat{k} \cdot \hat{x}).
\]
Since $P_\ell (\hat{k} \cdot \hat{x})$ is symmetric under the swap $k \leftrightarrow x$, it must also be an eigenvector of the Laplacian with respect to $\vec{k}$,

$$-\nabla^2 \vec{k} S_2 P_\ell \left( \hat{k} \cdot \hat{x} \right) = \ell (\ell + 1) P_\ell \left( \hat{k} \cdot \hat{x} \right). \quad (9.2.78)$$

**Complex conjugation** Under complex conjugation, the spherical harmonics obey

$$\overline{Y^m_\ell (\theta, \phi)} = (-)^m Y^{-m}_\ell (\theta, \phi). \quad (9.2.79)$$

**Parity** Under a parity flip, meaning if you compare $Y^m_\ell$ evaluated at the point $(\theta, \phi)$ to the point on the opposite side of the sphere $(\pi - \theta, \phi + \pi)$, we have the relation

$$Y^m_\ell (\pi - \theta, \phi + \pi) = (-)^\ell Y^m_\ell (\theta, \phi). \quad (9.2.80)$$

The odd $\ell$ spherical harmonics are thus odd under parity; whereas the even $\ell$ ones are invariant (i.e., even) under parity. That the Laplacian on the sphere $\vec{\nabla}^2 S_2$ and the parity operator $P$ share a common set of eigenvectors is because they commute: $[P, \vec{\nabla}^2 S_2] = 0$.

**Poisson Equation on the $2$-sphere** Having acquired some familiarity of the spherical harmonics, we can now tackle Poisson’s equation

$$-\vec{\nabla}^2 S_2 \psi (\theta, \phi) = J (\theta, \phi) \quad (9.2.81)$$

on the $2$–sphere. Because the spherical harmonics are complete on the sphere, we may expand both $\psi$ and $J$ in terms of them.

$$\psi = \sum_{\ell,m} A^m_\ell Y^m_\ell, \quad J = \sum_{\ell,m} B^m_\ell Y^m_\ell. \quad (9.2.82)$$

(This means, if $J$ is a given function, then we may calculate $B^m_\ell = \int_{S^2} d^2 \Omega Y^m_\ell (\hat{\theta}, \hat{\phi}) J (\theta, \phi)$.) Inserting these expansions into eq. (9.2.81), and recalling the eigenvalue equation $-\vec{\nabla}^2 S_2 Y^m_\ell = \ell (\ell + 1) Y^m_\ell$,

$$\sum_{\ell \neq 0, m} \ell (\ell + 1) A^m_\ell Y^m_\ell = \sum_{\ell, m} B^m_\ell Y^m_\ell. \quad (9.2.83)$$

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\footnote{For, suppose there is a solution to $-\vec{\nabla}^2 \psi = \chi/(4\pi)$, where $\chi$ 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 $\chi$. This inconsistency means no such solution exist.} At this point, we may then equate the $\ell > 0$ coefficients of the spherical harmonics on both sides, and deduce

$$A^m_\ell = \frac{B^m_\ell}{\ell (\ell + 1)}, \quad \ell > 0. \quad (9.2.84)$$

115For, suppose there is a solution to $-\vec{\nabla}^2 \psi = \chi/(4\pi)$, where $\chi$ 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 $\chi$. This inconsistency means no such solution exist.
To summarize, given a \( J(\theta, \phi) \) that has no “zero mode,” such that it can be decomposed as

\[
J(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} B_{\ell}^m Y_{\ell}^m(\theta, \phi) \quad \Leftrightarrow \quad B_{\ell}^m = \int_{-1}^{+1} d(\cos \theta) \int_0^{2\pi} d\phi Y_{\ell}^m(\theta, \phi) J(\theta, \phi),
\]

(9.2.85)

the solution to (9.2.81) is

\[
\psi(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} B_{\ell}^m \ell(\ell + 1) Y_{\ell}^m(\theta, \phi).
\]

(9.2.86)

**Problem 9.7.** Diagonalize the Laplacian in 2D flat space in cylindrical coordinates – i.e., obtain the results in eq. (9.2.46) – using the separation-of-variables technique. Hints: What is the boundary condition in the \( \phi \) direction? For the radial function, consider the appropriate boundary conditions at \( r = 0 \); you may need to refer to here, here, and here.

### 9.2.4 3 Dimensions

**Infinite Flat Space, Cylindrical Coordinates**

We now turn to 3D flat space, written in cylindrical coordinates,

\[
dt^2 = dr^2 + r^2 d\phi^2 + dz^2, \quad r \geq 0, \quad \phi \in [0, 2\pi), \quad z \in \mathbb{R}, \quad \sqrt{|g|} = r.
\]

(9.2.87)

Because the negative Laplacian on a scalar is the sum of the 1D and the 2D cylindrical case,

\[
-\vec{\nabla}_2^2 \psi = -\vec{\nabla}_2^2 \psi - \partial_z^2 \psi,
\]

(9.2.88)

we may try the separation-of-variables ansatz involving the product of the eigenvectors of the respective Laplacians.

\[
\psi(r, \phi, z) = \psi_2(r, \phi) \psi_1(z), \quad \psi_2(r, \phi) \equiv J_m(kr) \frac{e^{im\phi}}{\sqrt{2\pi}}, \quad \psi_1(z) \equiv e^{ikz}.
\]

(9.2.89)

This yields

\[
-\vec{\nabla}^2 \psi = -\psi_1 \vec{\nabla}_2^2 \psi_2 - \psi_2 \partial_z^2 \psi_1 = (k^2 + (k_z)^2) \psi,
\]

(9.2.90)

To sum, the orthonormal eigenfunctions are

\[
\langle r, \phi, z | k, m, k_z \rangle = J_m(kr) \frac{e^{im\phi}}{\sqrt{2\pi}} e^{ikz} \quad \text{(9.2.91)}
\]

\[
\int_0^{2\pi} d\phi \int_0^\infty dr \int_{-\infty}^{+\infty} dz \langle k', m', k'_z | r, \phi, z \rangle \langle r, \phi, z | k, m, k_z \rangle = \delta_{m'}^m \delta(k - k') \sqrt{kk'} \cdot (2\pi) \delta(k'_z - k_z).
\]

(9.2.92)

Since we already figured out the 2D plane wave expansion in cylindrical coordinates in eq. (9.2.41), and since the 3D plane wave is simply the 2D one multiplied by the plane wave in the \( z \) direction, we...
direction, i.e., $\exp(i\vec{k} \cdot \vec{x}) = \exp(ikr \cos(\phi - \phi_k)) \exp(ikz)$, we may write down the 3D expansion immediately

$$\langle \vec{x} | \vec{k} \rangle = \exp(i\vec{k} \cdot \vec{x}) = \sum_{\ell=\infty}^\infty i^\ell J_\ell(kr)e^{i(\phi - \phi_k)}e^{ikz}, \quad (9.2.93)$$

where

$$k_i = (k \cos \phi_k, k \sin \phi_k, k_z), \quad x^i = (r \cos \phi, r \sin \phi, z). \quad (9.2.94)$$

**Infinite Flat Space, Spherical Coordinates**

We now turn to 3D flat space written in spherical coordinates,

$$d\ell^2 = dr^2 + r^2 d\Omega_2^2, \quad d\Omega_2^2 \equiv d\theta^2 + (\sin \theta)^2 d\phi^2,$n

$$r \geq 0, \ \phi \in [0, 2\pi), \ \theta \in [0, \pi], \ \sqrt{|g|} = r^2 \sin \theta. \quad (9.2.95)$$

The Laplacian on a scalar is

$$\vec{\nabla}^2 \psi = \frac{1}{r^2} \partial_r \left( r^2 \partial_r \psi \right) + \frac{1}{r^2} \vec{\nabla}^2_{S^2} \psi. \quad (9.2.96)$$

where $\vec{\nabla}^2_{S^2}$ is the Laplacian on a 2-sphere.

*Plane wave* \hspace{1cm} With

$$k_i = k (\sin(\theta_k) \cos(\phi_k), \sin(\theta_k) \sin(\phi_k), \cos(\theta_k)) \equiv \hat{k}k, \quad (9.2.97)$$

$$x^i = r (\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta)) \equiv r\hat{x}, \quad (9.2.98)$$

we have

$$\langle \vec{x} | \hat{k} \rangle = \exp(i\hat{k} \cdot \vec{x}) = \exp \left( ikr \hat{k} \cdot \hat{x} \right). \quad (9.2.99)$$

If we view $\hat{k}$ 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 $\langle \vec{x} | \hat{k} \rangle$ using $P_\ell(\hat{k} \cdot \hat{x})$.

$$\exp \left( ikr \hat{k} \cdot \hat{x} \right) = \sum_{\ell=0}^\infty \chi_\ell(kr) \sqrt{\frac{2\ell + 1}{4\pi}} P_\ell \left( \hat{k} \cdot \hat{x} \right). \quad (9.2.100)$$

For convenience we have used the $Y_\ell^0$ in eq. (9.2.71) as our basis. Exploiting the orthonormality of the spherical harmonics to solve for the expansion coefficients:

$$\chi_\ell(kr) = 2\pi \int_{-1}^{+1} d\phi' e^{ikr\phi'} Y_\ell^0(\theta', \phi') = \sqrt{(4\pi)(2\ell + 1)\frac{1}{2}} \int_{-1}^{+1} d\phi' e^{ikr\phi'} P_\ell(c). \quad (9.2.101)$$

(Even though the integral is over the entire solid angle, the azimuthal integral is trivial and yields $2\pi$ 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,

$$i^\ell j_\ell(z) = \frac{1}{2} \int_{-1}^{+1} d\phi' e^{iz\phi'} P_\ell(c), \quad \ell = 0, 1, 2, \ldots \quad (9.2.102)$$
(The spherical Bessel function $j_\ell(z)$ is real when $z$ is positive.) We have arrived at

$$
\langle \vec{x} | \vec{k} \rangle = \exp(i \vec{k} \cdot \vec{x}) = \sum_{\ell=0}^{\infty} (2\ell + 1)i^\ell j_\ell(kr) P_\ell(\vec{k} \cdot \vec{x}), \quad k \equiv |\vec{k}|
$$

(9.2.103)

$$
= 4\pi \sum_{\ell=0}^{\infty} i^\ell j_\ell(kr) \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta, \phi) Y_{\ell'}^m(\theta_k, \phi_k),
$$

(9.2.104)

where, for the second equality, we have employed the additional formula in eq. (9.2.76).

**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.

$$
\int_{\mathbb{R}^3} d^3x \exp(i(\vec{k} - \vec{k}') \cdot \vec{x}) = (2\pi)^3 \frac{\delta(k - k')}{k k'} \delta(\cos(\theta_{k'} - \cos(\theta_k)) \delta(\phi_k - \phi_{k'})
$$

(9.2.105)

Switching to spherical coordinates within the integral on the left-hand-side, namely $d^3x = d\cos(\theta)d\phi dr^2 \equiv d\Omega dr^2$; re-expressing $\exp(i \vec{k} \cdot \vec{x})$ and $\exp(-i \vec{k}' \cdot \vec{x})$ using eq. (9.2.103) and its complex conjugate; followed by using eq. (9.2.72) to integrate over the solid angle,

$$
(4\pi)^2 \int_{\mathbb{S}^2} d^2\Omega \int_0^\infty drr^2 \sum_{\ell,\ell'=0}^{+\ell} (-i)^\ell j_\ell(kr) j_{\ell'}(k'r')
$$

$$
\times \sum_{m=-\ell}^{+\ell} \sum_{m'=-\ell'}^{+\ell'} Y_\ell^m(\theta, \phi) Y_{\ell'}^{m'}(\theta_k, \phi_k) Y_{\ell'}^{m'}(\theta_{k'}, \phi_{k'})
$$

$$
= (4\pi)^2 \int_0^\infty drr^2 \sum_{\ell=0}^{\infty} j_\ell(kr) j_{\ell}(k'r') \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k, \phi_k) Y_{\ell'}^m(\theta_{k'}, \phi_{k'}).
$$

(9.2.106)

Let us compare the right hand sides of the two preceding equations, and utilize the completeness relation obeyed by the spherical harmonics (cf. eq. (9.2.73)):

$$
4(2\pi)^2 \int_0^\infty drr^2 \sum_{\ell=0}^{\infty} j_\ell(kr) j_{\ell}(k'r') \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k, \phi_k) Y_{\ell'}^m(\theta_k, \phi_k)
$$

$$
= (2\pi)^3 \frac{\delta(k - k')}{k k'} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} Y_\ell^m(\theta_k, \phi_k) Y_{\ell'}^m(\theta_k, \phi_k).
$$

(9.2.107)

Therefore it must be that

$$
\int_0^\infty drr^2 \sum_{\ell=0}^{\infty} j_\ell(kr) j_{\ell}(k'r') = \frac{\pi \delta(k - k')}{2}. 
$$

(9.2.108)

Referring to eq. (10.47.3) of the NIST page here, we have

$$
\frac{4}{(2\pi)^{1/2}} \exp\left(-\frac{1}{4}z^2\right) = J_{\ell + \frac{1}{2}}(z)
$$

(9.2.109)
we see this is in fact the same result as in eq. (9.2.44).

To sum, we have diagonalized the 3D flat space negative Laplacian in spherical coordinates as follows.

\[-\vec{\nabla}^2 \langle r, \theta, \phi | k, \ell, m \rangle = k^2 \langle r, \theta, \phi | k, \ell, m \rangle ,\]

\[\langle r, \theta, \phi | k, \ell, m \rangle = \sqrt{\frac{2}{\pi}} j_\ell(kr) Y^m_\ell(\theta, \phi),\]  

(9.2.110)

\[\langle k', \ell', m' | k, \ell, m \rangle = \int_{S^2} d^2\Omega \int_0^\infty dr r^2 \langle k', \ell', m' | r, \theta, \phi \rangle \langle r, \theta, \phi | k, \ell, m \rangle ,\]

\[= \frac{\delta(k-k')}{kk'} \delta_\ell^{\ell'} \delta_m^{m'} .\]

**Problem 9.8. 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; \( \rho \) 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,

\[\vec{x} = \left( \sqrt{\rho^2 - R^2 \sin^2 \phi}, \sqrt{\rho^2 - R^2 \sin^2 \phi}, \rho \cos \theta \right) ;\]  

(9.2.111)

\[\rho \geq R, \quad (\theta, \phi) \in S^2 .\]  

(9.2.112)

Explain the geometric meaning of the constant \( R \). Work out the 3D flat metric in prolate ellipsoidal coordinates \((\rho, \theta, \phi)\) and proceed to diagonalize the associated scalar Laplacian \( \vec{\nabla}^2 \equiv g^{ij} \nabla_i \nabla_j \). Hint: Work out the appropriate eigenvector equation and multiply throughout by \( \rho^2 - R^2 \cos^2 \theta \). You should find the \( \phi \)-dependent portions separating after re-writing \( \rho^2 - R^2 \cos^2 \theta = (\rho^2 - R^2) + R^2 \sin^2 \theta \). Also, you may wish to look [here].

### 9.3 Heat/Diffusion Equation

#### 9.3.1 Definition, uniqueness of solutions

We will define the heat or diffusion equation to be the PDE

\[\partial_t \psi (t, \vec{x}) = \sigma \vec{\nabla}^2 \psi (t, \vec{x}) = \frac{\sigma}{\sqrt{|g|}} \partial_i \left( \sqrt{|g|} g^{ij} \partial_j \psi \right) , \quad \sigma > 0 ,\]  

(9.3.1)

where \( \vec{\nabla}^2 \) is the Laplacian with respect to some metric \( g_{ij}(\vec{x}) \), which we will assume *does not* depend on the time \( t \). We will also 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 \( \sigma \) 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 \( \sigma \). As the heat equation, this PDE 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. As we shall witness, the solution of eq. (9.3.1) is aided by the knowledge of the eigenfunctions/values of the Laplacian in question.
Uniqueness of solution

Suppose the following initial conditions are given

\[ \psi(t = t_0, \vec{x}) = \varphi_0(\vec{x}), \quad (9.3.2) \]

and suppose the field \( \psi \) or its normal derivative is specified on the boundaries \( \partial \mathcal{D} \),

\[ \psi(t, \vec{x} \in \partial \mathcal{D}) = \varphi_3(\partial \mathcal{D}), \quad \text{(Dirichlet),} \quad (9.3.3) \]

or \( n^i \nabla_i \psi(t, \vec{x} \in \partial \mathcal{D}) = \varphi_4(\partial \mathcal{D}), \quad \text{(Neumann),} \quad (9.3.4) \]

where \( n^i(\partial \mathcal{D}) \) is the unit outward normal vector. Then, the solution to the heat/diffusion equation in eq. (9.3.1) 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 \( \psi_1 \) and \( \psi_2 \); the proof is established if we can show, in fact, that \( \psi_1 \) has to be equal to \( \psi_2 \). Note that the difference \( \Psi \equiv \psi_1 - \psi_2 \) is subject to the initial conditions

\[ \Psi(t = t_0, \vec{x}) = 0, \quad (9.3.5) \]

and the spatial boundary conditions

\[ \Psi(t, \vec{x} \in \partial \mathcal{D}) = 0 \quad \text{or} \quad n^i \nabla_i \Psi(t, \vec{x} \in \partial \mathcal{D}) = 0. \quad (9.3.6) \]

Let us then consider the following (non-negative) integral

\[ \rho(t) \equiv \frac{1}{2} \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \Psi(t, \vec{x})^2 \geq 0, \quad (9.3.7) \]

as well as its time derivative

\[ \partial_t \rho(t) = \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \Psi \hat{\nabla}^2 \Psi. \quad (9.3.8) \]

We may use the heat/diffusion equation on the \( \hat{\nabla} \) term, and integrate-by-parts one of the gradients on the second term,

\[ \partial_t \rho(t) = \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \Psi \hat{\nabla}^2 \Psi \]

\[ = \int_{\partial \mathcal{D}} d^{D-1} \vec{\xi} \sqrt{|H(\vec{\xi})|} \Psi n^i \nabla_i \Psi - \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \nabla_i \Psi \nabla^i \Psi. \quad (9.3.9) \]

By assumption either \( \Psi \) 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

\[ \nabla_i \Psi \nabla^i \Psi = \sum_i (\nabla_i \Psi)^2 \geq 0. \quad (9.3.10) \]

This implies

\[ \partial_t \rho(t) = - \int_\mathcal{D} d^D \vec{x} \sqrt{|g(\vec{x})|} \nabla_i \Psi \nabla^i \Psi \leq 0. \quad (9.3.11) \]

However, the initial conditions \( \Psi(t = t_0, \vec{x}) = 0 \) indicate \( \rho(t = t_0) = 0 \) (cf. eq. (9.3.7)). 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 \( \Psi \), the only way it can be zero is \( \Psi = 0 \Rightarrow \psi_1 = \psi_2 \). This establishes the theorem. \( \square \)
9.3.2 Heat Kernel, Solutions with $\psi(\partial \mathcal{D}) = 0$

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

$$K(\vec{x}, \vec{r}; s \geq 0) \equiv \langle \vec{x} \mid e^{s \nabla^2} \mid \vec{r} \rangle. \quad (9.3.12)$$

It obeys the heat/diffusion equation

$$\partial_s K(\vec{x}, \vec{r}'; s) = \langle \vec{x} \mid \nabla^2 e^{s \nabla^2} \mid \vec{r}' \rangle = \langle \vec{x} \mid e^{s \nabla^2} \nabla^2 \mid \vec{r}' \rangle = \nabla^2_{\vec{x}} K(\vec{x}, \vec{r}'; s) = \nabla^2_{\vec{r}'} K(\vec{x}, \vec{r}'; s), \quad (9.3.13)$$

where we have assumed $\nabla^2$ is Hermitian. $K$ also obeys the initial condition

$$K(\vec{x}, \vec{r}'; s = 0) = \langle \vec{x} \mid \vec{r}' \rangle = \delta^{(D)}(\vec{x} - \vec{r}'), \quad (9.3.14)$$

If we demand the eigenfunctions of $\nabla^2$ obey Dirichlet boundary conditions,

$$\left\{ \psi_\lambda(\partial \mathcal{D}) = 0 \mid -\nabla^2 \psi_\lambda = \lambda \psi_\lambda \right\}, \quad (9.3.15)$$

then the heat kernel obeys the same boundary conditions.

$$K(\vec{x} \in \partial \mathcal{D}, \vec{r}'; s) = K(\vec{x}, \vec{r}' \in \partial \mathcal{D}; s) = 0. \quad (9.3.16)$$

To see this we need to perform a mode expansion. By inserting in eq. (9.3.14) a complete set of the eigenstates of $\nabla^2$, the heat kernel has an explicit solution

$$K(\vec{x}, \vec{r}; s \geq 0) = \langle \vec{x} \mid e^{s \nabla^2} \mid \vec{r} \rangle = \sum_\lambda e^{-s\lambda} \langle \vec{x} \mid \lambda \rangle \langle \lambda \mid \vec{r} \rangle, \quad (9.3.17)$$

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 $\vec{r}$ lies on the boundary $\partial \mathcal{D}$.

**Initial value problem** In this section we will focus on solving the initial value problem when the field itself is zero on the boundary $\partial \mathcal{D}$ 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:

$$\psi(t = t', \vec{x}) \equiv \langle \vec{x} \mid \psi(t') \rangle \quad \text{(given)}, \quad \psi(t \geq t', \vec{x} \in \mathcal{D}) = 0. \quad (9.3.18)$$

Then $\psi(t, \vec{x})$ at any later time $t > t'$ is given by

$$\psi(t \geq t', \vec{x}) = \langle \vec{x} \mid e^{(t-t')\nabla^2} \mid \psi(t') \rangle = \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \vec{x} \mid e^{(t-t')\nabla^2} \mid \vec{x}' \rangle \langle \vec{x}' \mid \psi(t') \rangle = \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} K(\vec{x}, \vec{x}'; t-t') \psi(t', \vec{x}'). \quad (9.3.19)$$
That this is the correct solution is because the right hand side obeys the heat/diffusion equation through eq. (9.3.13). As \( t \to t' \), we also see from eq. (9.3.14) that the initial condition is recovered.

\[
\psi(t = t', \vec{x}) = \langle \vec{x} | \psi(t') \rangle = \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{|g(\vec{x})|g(\vec{x})|}} \psi(t', \vec{x}') = \psi(t', \vec{x}). \tag{9.3.20}
\]

Moreover, since the heat kernel obeys eq. (9.3.16), the solution automatically maintains the \( \psi(t \geq t', \vec{x} \in \mathcal{D}) = 0 \) 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 smallest eigenvalue that dominates the late time temperature distribution. From eq. (9.3.19) and (9.3.17),

\[
T(t \geq t', \vec{x}) = \sum_\lambda \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \vec{x}' | e^{(t-t')\nabla^2} | \lambda \rangle \langle \lambda | \vec{x}' \rangle \langle \vec{x} | T(t') \rangle
= \sum_\lambda e^{-(t-t')\lambda} \langle \vec{x} | \lambda \rangle \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \lambda | \vec{x}' \rangle \langle \vec{x}' | T(t') \rangle
= \sum_\lambda e^{-(t-t')\lambda} \langle \vec{x} | \lambda \rangle \langle \lambda | T(t') \rangle. \tag{9.3.21}
\]

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

\[
T(t - t' \to \infty, \vec{x}) \approx e^{-(t-t')\lambda_{\text{min}}} \langle \vec{x} | \lambda_{\text{min}} \rangle \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \lambda_{\text{min}} | \vec{x}' \rangle \langle \vec{x}' | T(t') \rangle, \tag{9.3.22}
\]

because all the \( \lambda > \lambda_{\text{min}} \) become exponentially suppressed (relative to the \( \lambda_{\text{min}} \) state) due to the presence of \( e^{-(t-t')\lambda} \). As long as the minimum eigenvalue \( \lambda_{\text{min}} \) is strictly positive, we see the final temperature is zero.

\[
T(t - t' \to \infty, \vec{x}) = 0, \quad \text{if } \lambda_{\text{min}} > 0. \tag{9.3.23}
\]

When the minimum eigenvalue is zero, we have

\[
T(t - t' \to \infty, \vec{x}) \to \langle \vec{x} | \lambda = 0 \rangle \int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \lambda = 0 | \vec{x}' \rangle \langle \vec{x}' | T(t') \rangle, \quad \text{if } \lambda_{\text{min}} = 0. \tag{9.3.24}
\]

The exception to the dominant behavior in eq. (9.3.22) is when there is zero overlap between the initial distribution and that eigenfunction with the smallest eigenvalue, i.e., if

\[
\int d^D \vec{x}' \sqrt{|g(\vec{x}')|} \langle \lambda_{\text{min}} | \vec{x}' \rangle \langle \vec{x}' | T(t') \rangle = 0. \tag{9.3.25}
\]

Generically, we may say that, with the passage of time, the component of the initial distribution along the eigenfunction corresponding to the eigenvalue \( \lambda \) decays as \( 1/\lambda \); i.e., when \( t - t' = 1/\lambda \), its amplitude falls by \( 1/e \).
Static limit  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. (9.3.1) is zero and we obtain Laplace’s equation
\[
0 = \nabla^2 \psi(t \to \infty, \vec{x}).
\] (9.3.26)

Probability interpretation in flat infinite space  In the context of the diffusion equation in flat space, because of the \(\delta\)-functions on the right hand side of eq. (9.3.14), the propagator \(K(\vec{x}, \vec{x'}; t - t')\) itself can be viewed as the probability density (≡ 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
\[
\int_{\mathbb{R}^D} d^D \vec{x} K(\vec{x}, \vec{x'}; t - t') = 1.
\] (9.3.27)

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.

\[
\int_{\mathbb{R}^D} d^D \vec{x} \left\langle \vec{x} \right| e^{(t-t')\nabla^2} \left| \vec{x'} \right\rangle = \int_{\mathbb{R}^D} d^D \vec{k} \int_{\mathbb{R}^D} d^D \vec{x} \left\langle \vec{x} \right| e^{(t-t')\nabla^2} \left| \vec{k} \right\rangle \left\langle \vec{k} \right| \vec{x'} \rangle 
\]
\[
= \int_{\mathbb{R}^D} d^D \vec{k} \int_{\mathbb{R}^D} d^D \vec{x} e^{-(t-t')\vec{k}^2} \left\langle \vec{r} | \vec{k} \right\rangle \left\langle \vec{k} \right| \vec{x'} \rangle 
\]
\[
= \int_{\mathbb{R}^D} d^D \vec{k} \int_{\mathbb{R}^D} d^D \vec{x} e^{-(t-t')\vec{k}^2} e^{i\hat{k} \cdot (\vec{x} - \vec{x'})} \left(\frac{2\pi}{D}\right)^D = \prod_{j=1}^{D} \int_{-\infty}^{+\infty} \frac{dk_j}{2\pi} e^{-(t-t')k_j^2} e^{ik_j(x^j - x'^j)}. \] (9.3.28)

Heat Kernel in flat space  In fact, the same technique allow us to obtain the heat kernel in flat \(\mathbb{R}^D\).

\[
\left\langle \vec{x} \right| e^{(t-t')\nabla^2} \left| \vec{x'} \right\rangle = \int_{\mathbb{R}^D} d^D k \left\langle \vec{x} \right| e^{(t-t')\nabla^2} \left| \vec{k} \right\rangle \left\langle \vec{k} \right| \vec{x'} \rangle 
\]
\[
= \int_{\mathbb{R}^D} \frac{d^D k}{(2\pi)^D} e^{-(t-t')k^2} e^{i\hat{k} \cdot (\vec{x} - \vec{x'})} = \prod_{j=1}^{D} \int_{-\infty}^{+\infty} \frac{dk_j}{2\pi} e^{-(t-t')(k_j^2)} e^{i\hat{k}_j(x^j - x'^j)}. \] (9.3.29)

We may “complete the square” in the exponent by considering
\[
-(t - t') \left(k_j - \frac{x^j - x'^j}{2(t - t')}\right)^2 = -(t - t') \left((k_j)^2 - ik_j \frac{x^j - x'^j}{t - t'} - \left(\frac{x^j - x'^j}{2(t - t')}\right)^2\right). \] (9.3.30)

The heat kernel in flat \(\mathbb{R}^D\) is therefore
\[
\left\langle \vec{x} \right| e^{(t-t')\sigma\nabla^2} \left| \vec{x'} \right\rangle = (4\pi \sigma (t - t'))^{-D/2} \exp \left(-\frac{(\vec{x} - \vec{x'})^2}{4\sigma(t - t')}\right), \quad t > t', \] (9.3.31)

where we have put back the diffusion constant \(\sigma\). If you have taken quantum mechanics, you may recognize this result to be very similar to the path integral \(H \left\langle \vec{x}, t \right| \vec{x'}, t' \right\rangle_H \) of a free particle.
9.3.3 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

\[ G(t - t'; \bar{x}, \bar{x}') = \Theta(t - t')K(\bar{x}, \bar{x}'; t - t'), \tag{9.3.32} \]

where we define \( \Theta(s) = 1 \) for \( s \geq 0 \) and \( \Theta(s) = 0 \) for \( s < 0 \). This Green’s function \( G \) obeys

\[ \left( \partial_t - \nabla^2 \right) G(t - t'; \bar{x}, \bar{x}') = \left( \partial_t - \nabla^2 \right) G(t - t'; \bar{x}, \bar{x}') = \delta(t - t') \frac{\delta(D)(\bar{x} - \bar{x}')}{\sqrt{g(\bar{x})g(\bar{x}')}}, \tag{9.3.33} \]

with the boundary condition

\[ G(\tau; \bar{x} \in \partial \mathcal{D}, \bar{x}') = G(\tau; \bar{x}, \bar{x}' \in \partial \mathcal{D}) = 0, \tag{9.3.34} \]

as well as the causality condition

\[ G(\tau; \bar{x}, \bar{x}') = 0 \quad \text{when} \quad \tau < 0. \tag{9.3.35} \]

The boundary condition in eq. \(9.3.34\) follows directly from eq. \(9.3.16\); whereas eq. \(9.3.33\) follows from equations \(9.3.13\) and \(9.3.14\):

\[ \left( \partial_t - \nabla^2 \right) G(t - t'; \bar{x}, \bar{x}') = \delta(t - t')K(\bar{x}, \bar{x}'; t - t') + \Theta(t - t') \left( \partial_t - \nabla^2 \right) K(\bar{x}, \bar{x}'; t - t') \]
\[ = \delta(t - t') \frac{\delta(D)(\bar{x} - \bar{x}')}{\sqrt{g(\bar{x})g(\bar{x}')}}. \tag{9.3.36} \]

Initial value problem 
Within a spatial domain \( \mathcal{D} \), suppose the initial field configuration \( \psi(t', \bar{x} \in \mathcal{D}) \) is given and suppose its value on the spatial boundary \( \partial \mathcal{D} \) is also provided (i.e., Dirichlet B.C.’s \( \psi(t \geq t', \bar{x} \in \partial \mathcal{D}) \) are specified). The unique solution \( \psi(t \geq t', \bar{x} \in \mathcal{D}) \) to the heat/diffusion equation \(9.3.1\) is

\[ \psi(t \geq t', \bar{x}) = \int_{\mathcal{D}} d^D \bar{x}' \sqrt{|g(\bar{x}')|} G(t - t'; \bar{x}, \bar{x}') \psi(t', \bar{x}') \tag{9.3.37} \]
\[ - \int_{t'}^t dt'' \int_{\partial \mathcal{D}} d^{D-1} \bar{\xi} \sqrt{|H(\bar{\xi})|} n'' \nabla_v G(t - t''; \bar{x}, \bar{x}'(\bar{\xi})) \psi(t'', \bar{x}'(\bar{\xi})), \]

where the Green’s function \( G \) obeys the PDE in eq. \(9.3.33\) and the boundary conditions in equations \(9.3.34\) and \(9.3.35\).

Derivation of eq. \(9.3.37\) 
We begin by multiplying both sides of eq. \(9.3.33\) by \( \psi(t'', \bar{x}') \) and integrating over both space and time (from \( t' \) to infinity).

\[ \psi(t \geq t', \bar{x}) = \int_{t'}^\infty dt'' \int_{\mathcal{D}} d^D \bar{x}' \sqrt{|g(\bar{x}')|} \left( \partial_t - \nabla^2 \right) G(t - t''; \bar{x}, \bar{x}') \psi(t'', \bar{x}') \tag{9.3.38} \]
\[ = \int_{t'}^\infty dt'' \int_{\mathcal{D}} d^D \bar{x}' \sqrt{|g(\bar{x}')|} \left( -\partial_{vv} G \psi + \nabla_v G \nabla_v \psi \right) \]

385
\[
- \int_0^\infty dt'' \int_{\partial D} d^{D-1}\xi' \sqrt{|H(\xi')}| n^i \nabla_i G \psi
\]

\[
= \int_\mathbb{D} d^D x' \sqrt{|g(x')}| \left\{ \left[ -G \psi \right]_{t''=\infty}^{t''=t'} + \int_{t'}^{\infty} dt'' G \left( \partial_{t''} - \tilde{\nabla}^2 \right) \psi \right\}
\]

\[
+ \int_{t'}^{\infty} dt'' \int_{\partial D} d^{D-1}\xi' \sqrt{|H(\xi')}| \left( G \cdot n^i \nabla_i \psi - n^i \nabla_i G \cdot \psi \right).
\]

If we impose the boundary condition in eq. (9.3.35), we see that \([-G \psi]_{t''=\infty}^{t''=t'} = G(t - t')\psi(t')\) because the upper limit contains \(G(t - \infty) \equiv \lim_{t' \to -\infty} \Theta(t - t')K(\vec{x}, \vec{x}'; t - t') = 0\). The heat/diffusion eq. (9.3.1) removes the time-integral term on the first line of the last equality. If Dirichlet boundary conditions were chosen, we may choose \(G(t - t''; \vec{x}, \vec{x}' \in \partial \mathbb{D}) = 0\) (i.e., eq. (9.3.34)) and obtain eq. (9.3.37). Note that the upper limit of integration in the last line is really \(t\), because eq. (9.3.35) tells us the Green’s function vanishes for \(t'' > t\). Finally, recall we have already in §(9.3.1) proven the uniqueness of the solution to the heat equation obeying Dirichlet or Neumann boundary conditions.

### 9.3.4 Problems

**Problem 9.9.** In infinite flat \(\mathbb{R}^D\), suppose we have some initial probability distribution of finding a Brownian particle, expressed in Cartesian coordinates as

\[
\psi(t = t_0, \vec{x}) = \left( \frac{\omega}{\pi} \right)^{D/2} \exp \left( -\omega (\vec{x} - \vec{x}_0)^2 \right), \quad \omega > 0.
\]  

(9.3.39)

Solve the diffusion equation for \(t \geq t_0\).

**Problem 9.10.** 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:

\[
T_0(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a^m_\ell Y^m_\ell(\theta, \phi), \quad a^m_\ell \in \mathbb{C}.
\]  

(9.3.40)

The temperature as a function of time obeys the heat/diffusion equation

\[
\partial_t T(t, \theta, \phi) = \sigma \tilde{\nabla}^2 T(t, \theta, \phi), \quad \sigma > 0,
\]

(9.3.41)

where \(\tilde{\nabla}^2\) is now the Laplacian on the 2–sphere. Since \(\tilde{\nabla}^2\) is dimensionless here, \(\sigma\) has units of \(1/\text{[Time]}\).

1. Solve the propagator \(K\) for the heat/diffusion equation on the 2–sphere, in terms of a spherical harmonic \(\{Y^m_\ell(\theta, \phi)\}\) expansion.

2. Find the solution for \(T(t > t_0, \theta, \phi)\).

3. What is the decay rate of the \(\ell\)th multipole, i.e., how much time does the \(\ell\)th term in the multipole sum take to decay in amplitude by \(1/e\)? Does it depend on both \(\ell\) and \(m\)? And, what is the final equilibrium temperature distribution?
Problem 9.11. 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:

$$-\vec{\nabla}^2 G(\vec{x}, \vec{x}') = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{g(\vec{x})g(\vec{x}')}}.$$  \hfill (9.3.42)

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 $\hat{G} \equiv 1/(-\vec{\nabla}^2)$, namely

$$G(\vec{x}, \vec{x}') = \langle \vec{x} | \hat{G} | \vec{x}' \rangle \equiv \langle \vec{x} | \frac{1}{-\vec{\nabla}^2} | \vec{x}' \rangle.$$ \hfill (9.3.43)

The $\vec{\nabla}^2$ is now an abstract operator acting on the Hilbert space spanned by the position eigenkets $\{|\vec{x}\rangle\}$. Because it is Hermitian, we have

$$-\vec{\nabla}^2 \langle \vec{x} | \frac{1}{-\vec{\nabla}^2} | \vec{x}' \rangle = \langle \vec{x} | -\vec{\nabla}^2 | \vec{x}' \rangle = \langle \vec{x} | \vec{x}' \rangle = \delta^{(D)}(\vec{x} - \vec{x'}).$$ \hfill (9.3.44)

Now use the Gamma function identity, for $\text{Re}(z), \text{Re}(b) > 0$,

$$\frac{1}{b^z} = \frac{1}{\Gamma(z)} \int_0^\infty t^{z-1}e^{-bt} dt,$$ \hfill (9.3.45)

where $\Gamma(z)$ is the Gamma function – to justify

$$G(\vec{x}, \vec{x}') = \int_0^\infty dt K_G (\vec{x}, \vec{x}; t),$$ \hfill (9.3.46)

$$K_G (\vec{x}, \vec{x}; t) \equiv \langle \vec{x} | e^{t\vec{\nabla}^2} | \vec{x}' \rangle.$$ \hfill (9.3.47)

Notice how the integrand itself is the propagator (eq. (9.3.12)) of the heat/diffusion equation.

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. (9.3.45) to deduce

$$G(\vec{x}, \vec{x}') = \int_0^\infty dt \int \frac{d^D\vec{k}}{(2\pi)^D} e^{-t\vec{k}^2} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')},$$ \hfill (9.3.47)

(Going to momentum space allows you to also justify in what sense the restriction $\text{Re}(b) > 0$ of the formula in eq. (9.3.45) was satisfied.) By appropriately “completing the square” in the exponent, followed by an application of eq. (9.3.45), evaluate this integral to arrive at the Green’s function of the Laplacian in $D$ spatial dimensions:

$$G(\vec{x}, \vec{x}') = \langle \vec{x} | \frac{1}{-\vec{\nabla}^2} | \vec{x}' \rangle = \frac{\Gamma \left( \frac{D}{2} - 1 \right)}{4\pi^{D/2} |\vec{x} - \vec{x}'|^{D-2}},$$ \hfill (9.3.48)

\footnote{The perspective that the Green’s function be viewed as an operator acting on some Hilbert space was advocated by theoretical physicist Julian Schwinger.}

387
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 \(r_\equiv \max(r,r')\), \(r_\equiv \min(r,r')\) and \(\hat{n} \cdot \hat{n}'\), where \(r \equiv |\vec{x}|\), \(r' \equiv |\vec{x}'|\), \(\hat{n} \equiv \vec{x}/r\), and \(\hat{n}' \equiv \vec{x}'/r'\)? The \(D = 3\) case reads

\[
\frac{1}{4\pi |\vec{x} - \vec{x}'|} = (4\pi r_\equiv)^{-1} \sum_{\ell=0}^{\infty} P_\ell (\hat{n} \cdot \hat{n}') \left( \frac{r_<}{r_>} \right) \ell = \frac{1}{r_\equiv} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{Y^m_\ell (\hat{n}) Y^m_\ell (\hat{n}')} {2\ell + 1} \left( \frac{r_<}{r_>} \right) ^\ell ,
\]

(9.3.49)

where the \(P_\ell\) are Legendre polynomials and in the second line the addition formula of eq. (9.2.76) was invoked.

Note that while it is not easy to verify by direct differentiation that eq. (9.3.48) is indeed the Green’s function \(1/(-\nabla^2)\), one can do so by performing the integral over \(t\) in eq. (9.3.47), to obtain

\[
G(\vec{x},\vec{x}') = \int \frac{d^Dk}{(2\pi)^D} \frac{e^{i\vec{k} \cdot (\vec{x} - \vec{x}')}} {\vec{k}^2} .
\]

(9.3.50)

We have already seen this in eq. (9.1.31).

Finally, can you use the relationship between the heat kernel and the Green’s function of the Laplacian in eq. (9.3.46), to show how in a finite domain, eq. (9.3.37) leads to eq. (9.1.43) in the late time \(t \to \infty\) limit? (You may assume the smallest eigenvalue of the negative Laplacian is strictly positive; recall eq. (9.1.48).)

\(\square\)

**Problem 9.12.** Is it possible to solve for the Green’s function of the Laplacian on the 2-sphere? Use the methods of the last two problems, or simply try to write down the mode sum expansion in eq. (9.1.22), to show that you would obtain a 1/0 infinity. What is the reason for this apparent pathology? Suppose we could solve

\[
-\nabla^2 G(\vec{x}, \vec{x}') = \frac{\delta^{(2)}(\vec{x} - \vec{x}')}{\sqrt{g(\vec{x})g(\vec{x}')}} .
\]

(9.3.51)

Perform a volume integral of both sides over the 2−sphere – explain the contradiction you get. (Recall the discussion in the differential geometry section.) Hint: Apply the curved space Gauss’ law in eq. (7.4.38) and remember the 2-sphere is a closed surface.

### 9.4 Massless Scalar Wave Equation (Mostly) In Flat Spacetime \(\mathbb{R}^{D,1}\)

#### 9.4.1 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

\[
ds^2 = c^2 dt^2 - d\vec{x} \cdot d\vec{x} \equiv \eta_{\mu\nu} dx^\mu dx^\nu, \quad x^\mu \equiv (ct, x^i) ,
\]

(9.4.1)
where \( c \) is the speed of light in vacuum; \( \mu \in \{0, 1, 2, \ldots, D\} \); and \( D \) is still the dimension of space\(^\text{117}\). We also have defined the flat (Minkowski) spacetime metric
\[
\eta_{\mu\nu} \equiv \text{diag} (1, -1, -1, \ldots, -1).
\] (9.4.2)

The generalization of eq. (9.4.1) to curved spacetime is
\[
\mathrm{d}s^2 = g_{\mu\nu}(t, \vec{x}) \mathrm{d}x^\mu \mathrm{d}x^\nu, \quad x^\mu = (ct, x^i).
\] (9.4.3)

It is common to use the symbol \( \Box \), especially in curved spacetime, to denote the spacetime-Laplacian:
\[
\Box \psi \equiv \nabla_\mu \nabla^\mu \psi = \frac{1}{\sqrt{|g|}} \partial_\mu \left( \sqrt{|g|} g^{\mu\nu} \partial_\nu \psi \right),
\] (9.4.4)

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. (9.4.1), we have \( \sqrt{|g|} = 1 \), \( \eta^{\mu\nu} = \eta_{\mu\nu} \), and
\[
\Box \psi = \eta^{\mu\nu} \partial_\mu \partial_\nu \psi \equiv \delta^2 \psi = (c^{-2} \partial_t^2 - \delta^{ij} \partial_i \partial_j) \psi;
\] (9.4.5)

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
\[
u \cdot v \equiv \eta_{\mu\nu} u^\mu v^\nu = u^0 v^0 - \vec{u} \cdot \vec{v}, \quad u^2 \equiv (u^0)^2 - \vec{u}^2, \quad \text{etc.}
\] (9.4.6)

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,
\[
u^\mu = g^{\mu\nu} u_\nu, \quad u_\mu = g_{\mu\nu} u^\nu.
\] (9.4.7)

In the flat spacetime geometry of eq. (9.4.1), written in Cartesian coordinates,
\[
u^0 = u_0, \quad u^i = -u_i.
\] (9.4.8)

**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. (9.4.1) 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:
\[
0 = (x - x')^2 = \eta_{\mu\nu}(x^\mu - x'^\mu)(x^\nu - x'^\nu) \quad \Rightarrow \quad (t - t')^2 = (\vec{x} - \vec{x}')^2.
\] (9.4.9)

To understand this statement more systematically, let us work out the geodesic distance between any pair of spacetime points in flat spacetime.

\(^{117}\)In this section it is important to distinguish Greek \( \{\mu, \nu, \ldots\} \) and Latin/English alphabets \( \{a, b, i, j, \ldots\} \). The former run over 0 through \( D \), where the 0th index refers to time and the 1st through \( D \)th 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, [29].
Problem 9.13. In Minkowski spacetime expressed in Cartesian coordinates, the Christoffel symbols are zero. Therefore the geodesic equation in (7.3.35) returns the following "acceleration-is-zero" ODE:

\[ 0 = \frac{d^2 Z^\mu(\lambda)}{d\lambda^2}. \]  

(9.4.10)

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

\[ Z^\mu(0 \leq \lambda \leq 1) = x^\mu(0) + \lambda (x^\mu - x^\mu). \]  

(9.4.11)

Use eq. (7.1.25) to show that half the square of the geodesic distance between \( x' \) and \( x \) is

\[ \bar{\sigma}(x, x') = \frac{1}{2} (x - x')^2. \]  

(9.4.12)

\( \bar{\sigma} \) is commonly called Synge’s world function in the gravitation literature.

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 \( d\vec{x} = 0 \) in the Minkowski metric: \( ds^2 \rightarrow dt^2 > 0 \).

- 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 \( dt = 0 \), the (square of the) spatial distance between the both of you is now given by integrating \( ds^2 \rightarrow -d\vec{x}^2 < 0 \) 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 \( ds^2 = 0 \). 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 \( \psi \), subject to eq. (9.4.16), 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 \( \delta_{ij} \) invariant,

\[ x^i \rightarrow R^i_j x^j + a^i \quad \Rightarrow \quad \delta_{ij} dx^i dx^j \rightarrow \delta_{ij} dx^i dx^j. \]  

(9.4.13)

(The \( R^i_j \) and \( a^i \) are constants.) Lorentz transformations \( \{ A^\alpha_\mu \eta_{\alpha\beta} A^\beta_\mu = \eta_{\mu\nu} \} \) and spacetime translations \( \{ a^\mu \} \) are ones that leave the flat Minkowski metric \( \eta_{\mu\nu} \) invariant.

\[ x^\alpha \rightarrow A^\alpha_\mu x^\mu + a^\alpha \quad \Rightarrow \quad \eta_{\mu\nu} dx^\mu dx^\nu \rightarrow \eta_{\mu\nu} dx^\mu dx^\nu. \]  

(9.4.14)
(The $A_\mu$ and $a_\alpha$ are constants.) This in turn leaves the light cone condition $ds^2 = 0$ invariant – the speed of light is unity, $|d\vec{x}|/dt = 1$, in all inertial frames related via eq. (9.4.14).

**Wave Equation In Curved Spacetime** The wave equation (for a minimally coupled massless scalar) in some spacetime geometry $g_{\mu\nu}dx^\mu dx^\nu$ is a 2nd order in time PDE that takes the following form:

$$
\nabla_\mu \nabla^\mu \psi = \frac{1}{\sqrt{|g|}} \partial_\mu \left( \sqrt{|g|} g^{\mu\nu} \partial_\nu \psi \right) = J(x),
$$

(9.4.15)

where $J$ is some specified external source of $\psi$.

**Minkowski** We will mainly deal with the case of infinite flat (aka “Minkowski”) spacetime in eq. (9.4.1), where in Cartesian coordinates $x^\mu = (ct, \vec{x})$. This leads us to the wave equation

$$
\left( \partial_t^2 - c^2 \nabla^2 \right) \psi(t, \vec{x}) = c^2 J(t, \vec{x}),
$$

(9.4.16)

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, $c = 1$.\footnote{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 $(v/c)^2 = 0.76$.}

We will work mainly in flat infinite spacetime, which means the $\nabla^2$ 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 $(t, z)$,

$$
\left( \partial_t^2 - \partial_z^2 \right) \psi(t, z) = 0.
$$

(9.4.17)

We see that the solutions are a superposition of either left-moving $\psi(z + t)$ or right-moving waves $\psi(z - t)$, where $\psi$ can be any arbitrary function,

$$
\left( \partial_t^2 - \partial_z^2 \right) \psi(z \pm t) = (\pm)^2 \psi''(z \pm t) - \psi''(z \pm t) = 0.
$$

(9.4.18)

**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 $[c] = [\text{Length}/\text{Time}]$, so $x \pm ct$ would yield the correct dimensions.

$$
\psi(t, x) = \psi_L(x + ct) + \psi_R(x - ct).
$$

(9.4.19)

These waves move strictly at speed $c$. \footnote{\textit{Minkowski}}
Problem 9.14. Let us define light cone coordinates as $x^\pm \equiv t \pm z$. Write down the Minkowski metric in eq. (9.4.1)

$$ds^2 = dt^2 - dz^2 \quad \text{(9.4.20)}$$
in terms of $x^\pm$ and show by direct integration of eq. (9.4.17) that the most general homogeneous wave solution in 2D is the superposition of left- and right-moving (otherwise arbitrary) profiles.

Uniqueness of Minkowski solutions

Suppose the following initial conditions are given

$$\psi(t = t_0, \vec{x}) = \varphi_0(\vec{x}), \quad \partial_t \psi(t = t_0, \vec{x}) = \varphi_1(\vec{x}); \quad \text{(9.4.21)}$$

and suppose the scalar field $\psi$ or its normal derivative is specified on the spatial boundaries $\partial \mathcal{D}$,

$$\psi(t, \vec{x} \in \partial \mathcal{D}) = \varphi_3(\partial \mathcal{D}), \quad \text{(Dirichlet)}, \quad \text{(9.4.22)}$$

or

$$n^i \nabla_i \psi(t, \vec{x} \in \partial \mathcal{D}) = \varphi_4(\partial \mathcal{D}), \quad \text{(Neumann)}, \quad \text{(9.4.23)}$$

where $n^i(\partial \mathcal{D})$ is the unit outward normal vector. Then, the solution to the wave equation in eq. (9.4.16) 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 $\psi_1$ and $\psi_2$ obeying the same initial and boundary conditions. The proof is established if we can show, in fact, that $\psi_1$ has to be equal to $\psi_2$. Note that the difference $\Psi \equiv \psi_1 - \psi_2$ is subject to the homogeneous wave equation

$$\partial^2 \Psi = \ddot{\Psi} - \vec{\nabla}^2 \Psi = 0 \quad \text{(9.4.24)}$$

since the $J$ cancels out when we subtract the wave equations of $\psi_{1,2}$. For similar reasons the $\Psi$ obeys the initial conditions

$$\Psi(t = t_0, \vec{x}) = 0 \quad \text{and} \quad \partial_t \Psi(t = t_0, \vec{x}) = 0, \quad \text{(9.4.25)}$$

and the spatial boundary conditions

$$\Psi(t, \vec{x} \in \partial \mathcal{D}) = 0 \quad \text{or} \quad n^i \nabla_i \Psi(t, \vec{x} \in \partial \mathcal{D}) = 0. \quad \text{(9.4.26)}$$

Let us then consider the following integral

$$T^{00}(t) \equiv \frac{1}{2} \int_{\mathcal{D}} d^D \vec{x} \left( \dot{\Psi}^2(t, \vec{x}) + \vec{\nabla} \Psi(t, \vec{x}) \cdot \vec{\nabla} \Psi(t, \vec{x}) \right) \quad \text{(9.4.27)}$$

as well as its time derivative

$$\partial_t T^{00}(t) = \int_{\mathcal{D}} d^D \vec{x} \left( \ddot{\Psi} \dot{\Psi} + \vec{\nabla} \dot{\Psi} \cdot \vec{\nabla} \Psi \right). \quad \text{(9.4.28)}$$

\[119\] The integrand, for $\Psi$ obeying the homogeneous wave equation, is in fact its energy density. Therefore $T^{00}(t)$ is the total energy stored in $\Psi$ at a given time $t$. 

392
We may use the homogeneous wave equation on the \( \dot{\psi} \) term, and integrate-by-parts one of the gradients on the second term,

\[
\partial_t T^{00}(t) = \int_{\partial \Omega} d^{D-1} \xi \sqrt{|H(\xi)|} \dot{\psi} n^\alpha \nabla_\alpha \psi + \int_{\Omega} d^D x \left( \dot{\psi} \nabla^2 \psi - \dot{\psi} \nabla^2 \psi \right).
\]

(9.4.29)

By assumption either \( \psi \) or \( n^\alpha \nabla_\alpha \psi \) is zero on the spatial boundary; if it were the former, then \( \dot{\psi}(\partial \Omega) = 0 \) too. Either way, the surface integral is zero. Therefore the right hand side vanishes and we conclude that \( T^{00} \) is actually a constant in time. Together with the initial conditions \( \dot{\psi}(t = t_0, \vec{x}) = 0 \) and \( \psi(t = t_0, \vec{x}) = 0 \) (which implies \( (\nabla \psi(t = t_0, \vec{x}))^2 = 0 \)), we see that \( T^{00}(t = t_0) = 0 \), and therefore has to remain zero for all subsequent time \( t \geq t_0 \). 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.

### 9.4.2 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.

\[
f(t, \vec{x}) = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} \tilde{f}(\omega, \vec{k}) e^{-i\omega t} e^{i\vec{k} \cdot \vec{x}} \quad \text{ (Not quite . . . )},
\]

(9.4.30)

where

\[
k^\mu \equiv (\omega, k^i).
\]

(9.4.31)

Remember the first component is now the 0th one; so

\[
\exp(-ik_\mu x^\mu) = \exp(-i\eta_{\mu\nu} k^\mu x^\nu) = \exp(-i\omega t) \exp(i\vec{k} \cdot \vec{x}).
\]

(9.4.32)

Furthermore, these plane waves in eq. (9.4.32) obey

\[
\partial^2 \exp(-ik_\mu x^\mu) = -k^2 \exp(-ik_\mu x^\mu), \quad k^2 \equiv k_\mu k^\mu.
\]

(9.4.33)

This comes from a direct calculation; note that \( \partial_\mu (ik_\alpha x^\alpha) = ik_\alpha \delta_\mu^\alpha = ik_\mu \) and similarly \( \partial^\mu (ik_\alpha x^\alpha) = ik^\mu \).

\[
\partial^2 \exp(-ik_\mu x^\mu) = \partial_\mu \partial^\mu \exp(-ik_\mu x^\mu) = (ik_\mu)(ik^\mu) \exp(-ik_\mu x^\mu).
\]

(9.4.34)

Therefore, a particular mode \( \tilde{f} e^{-ik_\alpha x^\alpha} \) satisfies the homogeneous scalar wave equation in eq. (9.4.16) with \( J = 0 \) – provided that

\[
0 = \partial^2 \left( \tilde{f} e^{-ik_\alpha x^\alpha} \right) = -k^2 \tilde{f} e^{-ik_\alpha x^\alpha} \quad \Rightarrow \quad k^2 = 0 \quad \Rightarrow \quad \omega^2 = \vec{k}^2.
\]

(9.4.35)
In other words, the two solutions are

$$\tilde{\psi}(\vec{k}) \exp \left( \pm i|\vec{k}| \left\{ t \pm \frac{\vec{k} \cdot \vec{x}}{\vec{k}} \right\} \right), \quad \vec{k} \equiv \frac{\vec{k}}{|\vec{k}|}.$$  \hspace{1cm} (9.4.36)

The $e^{+i\omega t}$ waves propagate along the $\hat{k}$ direction; while the $e^{-i\omega t}$ ones along $-\hat{k}$.

This relationship between the zeroth component of the momentum and its spatial ones, is often known as the dispersion relation. Moreover, the positive root

$$\omega = |\vec{k}|$$  \hspace{1cm} (9.4.37)

can be interpreted as saying the energy $\omega$ of the photon – or, the massless particle associated with $\psi$ obeying eq. (9.4.16) – is equal to the magnitude of its momentum $\vec{k}$.

Therefore, if $\psi$ satisfies the homogeneous wave equation, the Fourier expansion is actually $D$-dimensional not $(D + 1)$ dimensional:

$$\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{d^D \vec{k}}{(2\pi)^D} \left( \tilde{A}(\vec{k}) e^{-i|\vec{k}|t} + \tilde{B}(\vec{k}) e^{i|\vec{k}|t} \right) e^{i\vec{k} \cdot \vec{x}}.$$  \hspace{1cm} (9.4.38)

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 $\psi$, the $\tilde{A}$ and $\tilde{B}$ are related.

$$\psi(t, \vec{x})^* = \psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{d^D \vec{k}}{(2\pi)^D} \left( \tilde{A}(\vec{k})^* e^{i|\vec{k}|t} + \tilde{B}(\vec{k})^* e^{-i|\vec{k}|t} \right) e^{-i\vec{k} \cdot \vec{x}}$$

$$= \int_{\mathbb{R}^D} \frac{d^D \vec{k}}{(2\pi)^D} \left( \tilde{B}(-\vec{k})^* e^{-i|\vec{k}|t} + \tilde{A}(-\vec{k})^* e^{i|\vec{k}|t} \right) e^{i\vec{k} \cdot \vec{x}}.$$  \hspace{1cm} (9.4.39)

Comparing equations (9.4.38) and (9.4.39) indicate $\tilde{A}(-\vec{k})^* = \tilde{B}(\vec{k}) \leftrightarrow \tilde{A}(\vec{k}) = \tilde{B}(-\vec{k})^*$. Therefore,

$$\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{d^D \vec{k}}{(2\pi)^D} \left( \tilde{A}(\vec{k}) e^{-i|\vec{k}|t} + \tilde{A}(-\vec{k})^* e^{i|\vec{k}|t} \right) e^{i\vec{k} \cdot \vec{x}}.$$  \hspace{1cm} (9.4.40)

Note that $\tilde{A}(\vec{k})$ itself, for a fixed $\vec{k}$, has two independent parts – its real and imaginary portions.\(^{120}\)

**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 $\psi = \text{constant}$ 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. (9.4.40) – the need for two independent coefficients $\tilde{A}$ and $\tilde{A}^*$ to describe the homogeneous solution is intimately tied to the need for two independent initial conditions.

---

\(^{120}\)In quantum field theory, the coefficients $\tilde{A}(\vec{k})$ and $\tilde{A}(\vec{k})^*$ of the Fourier expansion in (9.4.40) will become operators obeying appropriate commutation relations.
Suppose
\[
\psi(t = 0, \vec{x}) = \psi_0(\vec{x}) \quad \text{and} \quad \partial_t \psi(t = 0, \vec{x}) = \dot{\psi}_0(\vec{x}),
\] (9.4.41)
where the right hand sides are given functions of space. Then, from eq. (9.4.40),
\[
\psi_0(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \tilde{\psi}_0(\vec{k}) e^{i\vec{k} \cdot \vec{x}} = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \left( \tilde{A}(\vec{k}) + \tilde{A}(-\vec{k})^* \right) e^{i\vec{k} \cdot \vec{x}}
\]
\[
\dot{\psi}_0(\vec{x}) = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \tilde{\psi}_0(\vec{k}) e^{i\vec{k} \cdot \vec{x}} = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \left( -i|\vec{k}| \left( \tilde{A}(\vec{k}) - \tilde{A}(-\vec{k})^* \right) \right) e^{i\vec{k} \cdot \vec{x}}.
\] (9.4.42)

We have also assumed that the initial field and its time derivative admits a Fourier expansion. By equating the coefficients of the plane waves,
\[
\tilde{\psi}_0(\vec{k}) = \tilde{A}(\vec{k}) + \tilde{A}(-\vec{k})^*,
\]
\[
\frac{i}{|\vec{k}|} \tilde{\psi}_0(\vec{k}) = \tilde{A}(\vec{k}) - \tilde{A}(-\vec{k})^*.
\] (9.4.43)

Inverting this relationship tells us the \(\tilde{A}(\vec{k})\) and \(\tilde{A}(\vec{k})^*\) are indeed determined by (the Fourier transforms) of the initial conditions:
\[
\tilde{A}(\vec{k}) = \frac{1}{2} \left( \tilde{\psi}_0(\vec{k}) + \frac{i}{|\vec{k}|} \tilde{\psi}_0(\vec{k}) \right)
\]
\[
\tilde{A}(-\vec{k})^* = \frac{1}{2} \left( \tilde{\psi}_0(\vec{k}) - \frac{i}{|\vec{k}|} \tilde{\psi}_0(\vec{k}) \right)
\] (9.4.44)

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:
\[
\psi(t, \vec{x}) = \frac{1}{2} \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \left\{ \left( \tilde{\psi}_0(\vec{k}) + \frac{i}{|\vec{k}|} \tilde{\psi}_0(\vec{k}) \right) e^{-i|\vec{k}|t} + \left( \tilde{\psi}_0(\vec{k}) - \frac{i}{|\vec{k}|} \tilde{\psi}_0(\vec{k}) \right) e^{i|\vec{k}|t} \right\} e^{i\vec{k} \cdot \vec{x}}
\]
\[
= \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \left( \tilde{\psi}_0(\vec{k}) \cos(|\vec{k}|t) + \frac{\tilde{\psi}_0(\vec{k}) \sin(|\vec{k}|t)}{|\vec{k}|} \right) e^{i\vec{k} \cdot \vec{x}}.
\] (9.4.45)

We see that the initial profile contributes to the part of the field even under time reversal \(t \rightarrow -t\); 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),
\[
\psi(t = t_0, \vec{x}) = \psi_0(\vec{x}), \quad \partial_t \psi(t = t_0, \vec{x}) = \dot{\psi}_0(\vec{x}).
\] (9.4.46)

Because of time-translation symmetry, eq. (9.4.45) becomes
\[
\psi(t, \vec{x}) = \int_{\mathbb{R}^D} \frac{d^Dk}{(2\pi)^D} \left( \tilde{\psi}_0(\vec{k}) \cos(|\vec{k}|(t - t_0)) + \frac{\tilde{\psi}_0(\vec{k}) \sin(|\vec{k}|(t - t_0))}{|\vec{k}|} \right) e^{i\vec{k} \cdot \vec{x}}.
\] (9.4.47)
Problem 9.15. Let’s consider an initial Gaussian wave profile with zero time derivative,
\[
\psi(t = 0, \vec{x}) = \exp\left(-\frac{\vec{x}^2}{\sigma^2}\right), \quad \partial_t \psi(t = 0, \vec{x}) = 0.
\] (9.4.48)

If \( \psi \) 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,
\[
\partial_x^2 \psi(x) = J(x),
\] (9.4.49)
\[
\partial_x^2 \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} \tilde{\psi}(k)e^{-ik\cdot x} = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} \tilde{J}(k)e^{-ik\cdot x},
\] (9.4.50)
\[
\int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} (-k^2) \tilde{\psi}(k)e^{-ik\cdot x} = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} \tilde{J}(k)e^{-ik\cdot x}, \quad k^2 \equiv k_\mu k^\mu. \tag{9.4.51}
\]

Because the plane waves \( \{\exp(-ik\cdot x)\} \) are basis vectors, their coefficients on both sides of the equation must be equal.
\[
\tilde{\psi}(k) = -\frac{\tilde{J}(k)}{k^2}.
\] (9.4.52)

The advantage of solving the wave equation in Fourier space is, we see that this is the particular solution for \( \psi \) – the portion that is sourced by \( J \). Turn off \( J \) and you’d turn off (the inhomogeneous part of) \( \psi \).

**Inhomogeneous solution via Green’s function** We next proceed to transform eq. (9.4.52) back to spacetime.
\[
\psi(x) = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} \tilde{G}_{D+1}(x - x') \tilde{J}(x')e^{-ik\cdot x} = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} J(x')e^{ik\cdot x'}
\] (9.4.53)
\[
= \int_{\mathbb{R}^{D+1}} d^{D+1} x'' \left( \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} e^{-ik(x - x'')} \right) J(x'').
\]

That is, if we define the Green’s function of the wave operator as
\[
G_{D+1}(x - x') = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1} k}{(2\pi)^{D+1}} e^{-ik\cdot (x - x')} = -\int \frac{d\omega}{2\pi} \int \frac{d^{D} k}{(2\pi)^{D}} e^{-i\omega(t-t')} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')},
\] (9.4.54)
eq (9.4.53) translates to
\[
\psi(x) = \int_{\mathbb{R}^{D+1}} d^{D+1} x'' G_{D+1}(x - x'')J(x''). \tag{9.4.55}
\]
The Green’s function $G_{D+1}(x,x')$ itself satisfies the following PDE:

$$\partial^2_x G_{D+1}(x,x') = \delta^{(D+1)}(x-x') = \delta(t-t')\delta^{(D)}(\vec{x} - \vec{x}') .$$  (9.4.56)

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. (9.4.54) will verify eq. (9.4.56). If $\partial^2$ denotes the wave operator with respect to either $x$ or $x'$, and if we recall the eigenvalue equation (9.4.33) as well as the integral representation of the $\delta$-function,

$$\partial^2 G_{D+1}(x - x') = \int_{\mathbb{R}^{D+1}} \frac{d^{D+1}k}{(2\pi)^{D+1}} \frac{\partial^2 e^{-ik\mu(x-x')}\mu}{-k^2} = \delta^{(D+1)}(x-x') .$$  (9.4.57)

**Relation to the Driven Simple Harmonic Oscillator**

If we had performed a Fourier transform only in space, notice that eq. (9.4.49) would read

$$\ddot{\tilde{\psi}}(t, \vec{k}) + \vec{k}^2 \tilde{\psi}(t, \vec{k}) = \tilde{J}(t, \vec{k}) .$$  (9.4.58)

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, and the source $\tilde{J}$ as the external force; even though the wave equation is relativistic while the SHO is non-relativistic.

**Problem 9.16.** Employing the frictionless limit of eq. (5.5.23), explain why the retarded version of the Green’s function in eq. (9.4.54) is

$$G^+(t - t', \vec{x} - \vec{x}') = \Theta(t - t') \int_{\mathbb{R}^{d-1}} \frac{d^{d-1}\vec{k}}{(2\pi)^{d-1}} \frac{\sin (k(t-t'))}{k} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} ,$$  (9.4.59)

where $k \equiv |\vec{k}|$. For each $\vec{k}$ mode, explain why this implies

$$\tilde{\psi}(t, \vec{k}) = \int_{-\infty}^{t} dt' \frac{\sin (k(t-t'))}{k} \tilde{J}(t', \vec{k}) .$$  (9.4.60)

We will shortly recover these results below.

**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. 9.4.56) with that of an external source $J$ in the wave equation for $\psi$ (eq. (9.4.49)), 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 $x'$. According to eq. (9.4.55), 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 $\partial^2$,
obeying eq. (9.4.56), admits the mode sum expansion in eq. (9.4.54). However, you will soon run into a stumbling block if you begin with the \( k_0 = \omega \) integral, because the denominator of the second line of eq. (9.4.54) gives rise to two singularities on the real line at \( \omega = \pm |\vec{k}| \). To ensure the mode expansion in eq. (9.4.54) is well defined, we would need to append to it an appropriate contour prescription for the \( \omega \)-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 \( \omega \)-plane, we can choose to avoid the singularities at \( \omega = \pm |\vec{k}| \) by

1. Making a tiny semi-circular clockwise contour around each of them. 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.

2. Making a tiny semi-circular counterclockwise contour around each of them. 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.

3. Making a tiny semi-circular counterclockwise contour around \( \omega = -|\vec{k}| \) and a clockwise one at \( \omega = +|\vec{k}| \). This will yield the Feynman Green’s function \( G_{D+1,F} \), named after the theoretical physicist Richard P. Feynman. The Feynman Green’s function is used heavily in Minkowski spacetime perturbative Quantum Field Theory. Unlike its retarded and advanced cousins – which are purely real – the Feynman Green’s function is complex. The real part is equal to half the advanced plus half the retarded Green’s functions. The imaginary part, in the quantum field theory context, describes particle creation by an external source.

These are just 3 of the most commonly used contour prescriptions – there are an infinity of others, of course. You may also wonder if there is a heat kernel representation of the Green’s function of the Minkowski spacetime wave operator, i.e., the generalization of eq. (9.3.46) to “spacetime Laplacians”. The subtlety here is that the eigenvalues of \( \partial^2 \), the \( \{-k^2\} \), are not positive definite; to ensure convergence of the proper time \( t \)-integral in eq. (9.3.46) 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 \( D \geq 1 \).

To put the issue of contours on concrete terms, let us tackle the 2 dimensional case. Because the Green’s function enjoys the spacetime translation symmetry of the Minkowski spacetime it resides in – namely, under the simultaneous replacements \( x^\mu \to x^\mu + a^\mu \) and \( x'^\mu \to x'^\mu + a^\mu \), the Green’s function remains the same object – without loss of generality we may set \( x' = 0 \) in eq. (9.4.54).

\[
G_2 (x^\mu = (t, z)) = -\int \frac{d\omega}{2\pi} \int \frac{dk}{2\pi} \frac{e^{-i\omega t} e^{ikz}}{\omega^2 - k^2} \tag{9.4.61}
\]

If we make the retarded contour choice, which we will denote as \( G^+_2 \), then if \( t < 0 \) we would close it in the upper half plane (recall \( e^{-i(\infty)(-|t|)} = 0 \)). Because there are no poles for \( \text{Im}(\omega) > 0 \), we’d get zero. If \( t > 0 \), on the other hand, we will form the closed (clockwise) contour \( C \) via
the lower half plane, and pick up the residues at both poles. We begin with a partial fractions
decomposition of $1/k^2$, followed by applying the residue theorem:

$$
G^+_{2}(t, z) = -i\Theta(t) \int \frac{d\omega}{C} \int_{\mathbb{R}} \frac{dk}{2\pi} e^{-i\omega t} \frac{e^{ikz}}{2k} \left( \frac{1}{\omega - k} - \frac{1}{\omega + k} \right)
$$

(9.4.62)

$$
= +i\Theta(t) \int \frac{dk}{2\pi} \frac{e^{ikz}}{2k} (e^{-ikt} - e^{ikt})
$$

$$
= -i\Theta(t) \int \frac{dk}{2\pi} \frac{e^{ikz}}{2k} \cdot 2i \sin(kt) = \Theta(t) \int \frac{dk}{2\pi} \frac{e^{ikz}}{k} \sin(kt)
$$

(9.4.63)

Let’s now observe that this integral is invariant under the replacement $z \rightarrow -z$. In fact,

$$
G^+_{2}(t, -z) = \Theta(t) \int \frac{dk}{2\pi} \frac{e^{-ikz}}{k} \sin(kt) = G^+_{2}(t, z)^*
$$

(9.4.64)

$$
= \Theta(t) \int \frac{dk}{2\pi} \frac{e^{ikz}}{2\pi - k} \sin(-kt) = G^+_{2}(t, z).
$$

(9.4.65)

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.

$$
G^+_{2}(t, z) = \Theta(t) \int \frac{dk}{2\pi} \frac{e^{ik|z|}}{k} \sin(kt)
$$

(9.4.66)

Note that the integrand $\exp(i k|z|) \sin(kt)/k$ 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:

$$
G^+_{2}(t, z) = \frac{\Theta(t)}{2} \int_{-\infty+io^+}^{+\infty+io^+} \frac{dk}{2\pi} \frac{e^{ik|z|}}{k} (e^{ikt} - e^{-ikt})
$$

$$
= \frac{\Theta(t)}{2} (-) (\Theta(-t - |z|) - \Theta(t - |z|))
$$

(9.4.68)

$$
= \frac{1}{2} \Theta(t - |z|).
$$

(9.4.69)

**Problem 9.17.** Can you explain why

$$
\Theta(t)\Theta(t^2 - z^2) = \Theta(t - |z|)?
$$

(9.4.70)

Re-write $\Theta(-t)\Theta(t^2 - z^2)$ as a single step function.

We have arrived at the solution

$$
G^+_{2}(x - x') = \frac{1}{2} \Theta(t - t')\Theta(\bar{\sigma}) = \frac{1}{2} \Theta(t - t' - |z - z'|),
$$

(9.4.71)

$$
\bar{\sigma} \equiv \frac{(t - t')^2 - (z - z')^2}{2} = \frac{1}{2} (x - x')^2.
$$

(9.4.72)

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

\[
G^-_2(x - x') = \frac{1}{2} \Theta(t' - t) \Theta(\bar{\sigma}).
\]  

(9.4.73)

**Problem 9.18.** From its Fourier representation, calculate \( G^\pm_3(x - x') \), the retarded and advanced Green’s function of the wave operator in 3 dimensional Minkowski spacetime. You should find

\[
G^\pm_3(x - x') = \frac{\Theta(\pm(t - t')) \Theta(\bar{\sigma})}{\sqrt{2(2\pi)}} \sqrt{\bar{\sigma}}.
\]  

(9.4.74)

*Bonus problem:* Can you perform the Fourier integral in eq. (9.4.54) for all \( G_{D+1} \)?

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.\(^{121}\)

For this purpose let’s set the notation. In \( (D + 1) \) dimensional flat spacetime, let the spatial coordinates be denoted as \( x^i = (\vec{x}_L, w^1, w^2) \); and in \( (D - 1) \) dimensions let the spatial coordinates be the \( \vec{x}_L \). Then \( |\vec{x} - \vec{x}'| \) is a \( D \) dimensional Euclidean distance between the observer and source in the former, whereas \( |\vec{x}_L - \vec{x}'_L| \) is the \( D - 1 \) counterpart in the latter.

Starting from the integral representation for \( G_{D+1} \) in eq. (9.4.54), we may integrate with respect to the \( D \)th spatial coordinate \( w^2 \):

\[
\int_{-\infty}^{+\infty} dw^2 G_{D+1}(t - t', \vec{x}_L - \vec{x}'_L, \vec{w} - \vec{w}') = \int_{-\infty}^{+\infty} dw^2 \int_{R^{D+1}} \frac{d\omega d^D k_L \delta^2 k ||}{(2\pi)^{D+1}} \frac{e^{-i\omega(t-t')}e^{i\vec{k}_L \cdot (\vec{x}_L - \vec{x}'_L)} e^{i\vec{k} || (\vec{w} - \vec{w}')}}{-\omega^2 + \vec{k}^2_L + \vec{k}^2 ||} = \\
\int_{R^{D+1}} \frac{d\omega d^D k_L \delta^2 k ||}{(2\pi)^{D+1}} \frac{e^{-i\omega(t-t')}e^{i\vec{k}_L \cdot (\vec{x}_L - \vec{x}'_L)} e^{i\vec{k} || (w^1 - w')}}{-\omega^2 + \vec{k}^2_L + \vec{k}^2 ||} = \\
\int_{R^D} \frac{d\omega d^D k_L \delta^2 k ||}{(2\pi)^{D}} \frac{e^{-i\omega(t-t')}e^{i\vec{k}_L \cdot (\vec{x}_L - \vec{x}'_L)} e^{i\vec{k} || (w^1 - w')}}{-\omega^2 + \vec{k}^2_L + (k^2 ||)^2} = G_D(t - t', \vec{x}_L - \vec{x}'_L, w^1 - w').
\]  

(9.4.75)

The notation is cumbersome, but the math can be summarized as follows. Integrating \( G_{D+1} \) over the \( D \)th 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

\(^{121}\)I will make this statement precise very soon, by you are encouraged to read H. Soodak and M. S. Tiersten, *Wakes and waves in N dimensions*, Am. J. Phys. 61 (395), May 1993, for a pedagogical treatment.
integrated with respect to either \( w^2 \) or \( w^2 \). If we compare our integral here with eq. [9.4.55], we may identify \( J(x^\nu) = \delta(t^\nu - t') \delta^{(D-2)}(\vec{x}_{1} - \vec{x}'_{1}) \delta(w^1 - w'^1) \), an instantaneous line source of unit strength lying parallel to the \( D \)th axis, piercing the \((D - 1)\) space at \((\vec{x}'_{1}, w'^1)\).

We may iterate this integral recursion relation once more,

\[
\int_{\mathbb{R}^2} d^2w G_{D+1}(t - t', \vec{x}_{1} - \vec{x}'_{1}, \vec{w} - \vec{w}') = G_{D-1}(t - t', \vec{x}_{1} - \vec{x}'_{1}). \quad (9.4.76)
\]

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

\[
2\pi \int_{0}^{\infty} d\rho \rho G_{D+1}(t - t', \sqrt{(\vec{x}_{1} - \vec{x}'_{1})^2 + \rho^2}) = G_{D-1}(t - t', |\vec{x}_{1} - \vec{x}'_{1}|), \quad (9.4.77)
\]

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

\[
|\vec{x} - \vec{x}'| = \sqrt{(\vec{x}_{1} - \vec{x}'_{1})^2 + (\vec{w} - \vec{w}')^2}. \quad (9.4.78)
\]

A further change of variables

\[
R' = \sqrt{(\vec{x}_{1} - \vec{x}'_{1})^2 + \rho^2} \quad \Rightarrow \quad dR' = \frac{\rho d\rho}{R'}. \quad (9.4.79)
\]

This brings us to

\[
2\pi \int_{R}^{\infty} dR' R' G_{D+1}(t - t', R') = G_{D-1}(t - t', R), \quad R \equiv |\vec{x}_{1} - \vec{x}'_{1}|. \quad (9.4.80)
\]

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.

\[
G_{D+1}(t - t', R) = -\frac{1}{2\pi R} \frac{\partial}{\partial R} G_{D-1}(t - t', R). \quad (9.4.81)
\]

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 \(|\vec{x}_{1} - \vec{x}'_{1}|\).

**Green’s Function From Extra Dimensional Line Source**

There is an alternate means of obtaining the integral relation in eq. [9.4.76], which was key to deriving eq. [9.4.81]. 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 \( \mathbb{R}^{D,1} \).

\[
G_D(t - t', \vec{x}_{1} - \vec{x}'_{1}, w^1 - w'^1) = \int_{-\infty}^{+\infty} dw^2 G_{D+1}(t - t', \vec{x}_{1} - \vec{x}'_{1}, \vec{w} - \vec{w}')J(w^2) \quad (9.4.82)
\]
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,

\[
\frac{\partial^2}{\partial w^2} \int_{-\infty}^{+\infty} dw^2 G_{D+1} \cdot J \left( \text{where } \frac{\partial^2}{\partial w^2} \equiv \partial^2 - \sum_{i=1}^{D-1} \partial^2_i \right)
\]

\[
= \int_{-\infty}^{+\infty} dw^2 J(w^2) \left( \partial^2_D - \left( \frac{\partial}{\partial w^2} \right)^2 + \left( \frac{\partial}{\partial w^2} \right)^2 \right) G_{D+1}(w^2 - w'^2)
\]

\[
= \int_{-\infty}^{+\infty} dw^2 J(w^2) \left( \partial^2_{D+1} + \left( \frac{\partial}{\partial w^2} \right)^2 \right) G_{D+1}(w^2 - w'^2)
\]

\[
= \int_{-\infty}^{+\infty} dw^2 J(w^2) \left( \delta(t - t')\delta^{(D-2)}(\vec{x}_1 - \vec{x}'_1)\delta^{(2)}(\vec{w} - \vec{w}') + \left( \frac{\partial}{\partial w^2} \right)^2 G_{D+1}(w^2 - w'^2) \right)
\]

\[
= \delta^{(D-1)}(x - x')\delta(w^1 - w'^1) J(w^2)
\]

\[
+ \left[ J(w^2) \frac{\partial G_{D+1}(w^2 - w'^2)}{\partial w^2} \right]_{w^2 = -\infty}^{w^2 = +\infty} - \left[ \frac{\partial J(w^2)}{\partial w^2} G_{D+1}(w^2 - w'^2) \right]_{w^2 = -\infty}^{w^2 = +\infty}
\]

\[
+ \int_{-\infty}^{+\infty} dw^2 J''(w^2) G_{D+1}(w^2 - w'^2).
\]

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 \(\delta\)-function measure, namely \(\delta^{(D-1)}(x - x')\delta(w^1 - w'^1)\), translates to the boundary condition on \(J\):

\[
J(w^2) = 1.
\]

That the second and third terms of the final equality of eq. (9.4.83) 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 \((\vec{x}_1, w^1, w^2)\), they must lie outside the observer’s light cone, where the Green’s function is identically zero. The final term of eq. (9.4.83) is zero if the source obeys the ODE

\[
0 = J''(w^2).
\]

The solution to eq. (9.4.85), subject to eq. (9.4.84), is

\[
J(w^2) = \cos^2 \vartheta + \frac{w^2}{w^2} \sin^2 \vartheta.
\]

Choosing \(\vartheta = 0\) and \(\vartheta = \pi/2\) would return, respectively,

\[
J(w^2) = 1 \quad \text{and} \quad J(w^2) = \frac{w'^2}{w^2}.
\]

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 \(\mathbb{R}^{D+1,1}\):

\[
G_D(t - t', \vec{x}_1 - \vec{x}'_1, w^1 - w'^1) = \int_{-\infty}^{+\infty} dw^2 \left( \cos^2 \vartheta + \frac{w'^2}{w^2} \sin^2 \vartheta \right)
\]
\[ \times G_{D+1}(t-t', \bar{x}_\perp - \bar{x}'_\perp, \bar{w} - \bar{w}'). \tag{9.4.89} \]

Using the simpler expressions in eq. (9.4.87), we obtain

\[ G_D(t-t', \bar{x}_\perp - \bar{x}'_\perp, w^1 - w'^1) = \int_{-\infty}^{+\infty} dw^2 G_{D+1}(t-t', \bar{x}_\perp - \bar{x}'_\perp, \bar{w} - \bar{w}') \tag{9.4.90} \]

\[ = \int_{-\infty}^{+\infty} dw^2 w^2 G_{D+1}(t-t', \bar{x}_\perp - \bar{x}'_\perp, \bar{w} - \bar{w}') \tag{9.4.91} \]

As a reminder, \( \bar{x}_\perp \) and \( \bar{x}'_\perp \) are \( D-1 \) dimensional spatial coordinates; whereas \( \bar{w} \) and \( \bar{w}' \) are two dimensional ones.

\( G^\pm_{D+1} \) in all dimensions, Causal structure of physical signals

At this point we may gather \( G^\pm_{2,3} \) in equations (9.4.71), (9.4.73), and (9.4.74) and apply to them the recursion relation in eq. (9.4.81) to record the explicit expressions of the retarded \( G^+_D \) and advanced \( G^-_{D+1} \) Green’s functions in all \( (D \geq 2) \) dimensions\footnote{When eq. (9.4.81) applied to \( G^\pm_{2,3} \) in equations (9.4.71), (9.4.73), and (9.4.74), 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 even dimensional spacetimes, \( D + 1 = 2 + 2n \) and \( n = 0, 1, 2, 3, 4, \ldots, \)

\[ G^\pm_{2+2n}(x-x') = \Theta(\pm(t-t')) \left( \frac{1}{2\pi} \frac{\partial}{\partial \bar{\sigma}} \right)^n \frac{\Theta(\bar{\sigma})}{2}. \tag{9.4.92} \]

- In odd dimensional spacetime, \( D + 1 = 3 + 2n \) and \( n = 0, 1, 2, 3, 4, \ldots, \)

\[ G^\pm_{3+2n}(x-x') = \Theta(\pm(t-t')) \left( \frac{1}{2\pi} \frac{\partial}{\partial \bar{\sigma}} \right)^n \left( \frac{\Theta(\bar{\sigma})}{2\pi \sqrt{2\bar{\sigma}}} \right). \tag{9.4.93} \]

Recall that \( \bar{\sigma}(x, x') \) is half the square of the geodesic distance between the observer at \( x \) and point source at \( x' \),

\[ \bar{\sigma} \equiv \frac{1}{2}(x-x')^2. \tag{9.4.94} \]

Hence, \( \Theta(\bar{\sigma}) \) is unity inside the light cone and zero outside; whereas \( \delta(\bar{\sigma}) \) and its derivatives are non-zero strictly on the light cone. Note that the inside-the-light-cone portion of a signal – for e.g., 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 \( (x \leftrightarrow x') \).

For a fixed source at \( x' \), we may interpret these coefficients of \( \Theta(\pm(t-t')) \) as the symmetric Green’s function: 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 portion of this symmetric signal; while the advanced \( \Theta(-(t-t')) \) (observer time earlier than emission time) selects the backward light cone part of it.

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 \( \psi \) 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, \( \psi \) 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 \( \partial_\nu J \) occurring within eq. (9.4.93) can be converted into \( \partial_\nu J \) – integrated-by-parts within the integral in eq. (9.4.55) 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.

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: 

\[
\partial^2 \psi = (\partial^2_x - \nabla^2) \psi = 0,
\]

This can be contrasted against Laplace’s equation \( \nabla^2 \psi = \partial_\mu \partial^\mu \psi = 0 \), 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).

\[
\partial^2 \left( \tilde{\psi}(k) e^{-i k \cdot x} \right) = 0 = - \left( k_0^2 - \vec{k} \cdot \vec{k} \right) \tilde{\psi} e^{-i k \cdot x}, \quad \text{(Wave Equation)} \tag{9.4.95}
\]

\[
\nabla^2 \left( \tilde{\psi}(k) e^{i \vec{k} \cdot \vec{x}} \right) = 0 = - \left( \vec{k} \cdot \vec{k} \right) \tilde{\psi} e^{i \vec{k} \cdot \vec{x}}, \quad \text{(Laplace Equation)} \tag{9.4.96}
\]

We see that, for the wave equation, either \( \tilde{\psi} = 0 \) or \( k^2 = k_0^2 - \vec{k}^2 = 0 \). But \( \tilde{\psi} = 0 \) would render the whole solution trivial. Hence, for non-singular \( \tilde{\psi} \) this means \( k_0 = \pm |\vec{k}| \) and we have

\[
\psi = \tilde{\psi}(k) \exp \left( i |\vec{k}| (\vec{k} \cdot \vec{x} \mp t) \right), \quad \vec{k} = \tilde{k} / |\tilde{k}|. \tag{9.4.97}
\]

(We have already encountered this result in eq. (9.4.38).) Whereas, for the Laplace equation either \( \tilde{\psi} = 0 \) 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.\(^{123}\) 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

\(^{123}\)Explicit formulas for the electromagnetic and linear gravitational case can be found in appendices A and B of arXiv: 1611.00018.\(^{36}\)

\(^{124}\)One could allow for singular solutions proportional to the \( \vec{k} \)-space \( \delta^{(d-1)} \)-function and its derivatives, such as \( \tilde{\psi}_0 = \delta^{(3)}(\vec{k}) \exp(i \vec{k} \cdot \vec{x}) \) and \( \tilde{\psi}_1 = \partial_\mu \delta^{(3)}(\vec{k}) \exp(i \vec{k} \cdot \vec{x}) \) (for fixed \( i \)), so that \( \nabla^2 \tilde{\psi} = 0 \) because \( \vec{k}^2 \delta^{(3)}(\vec{k}) = 0 = \vec{k}^2 \partial_\mu \delta^{(3)}(\vec{k}) \). However, the \( \psi_1 \) in position space is simply a spatial constant; while the \( \psi_1 \) 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 \( \{ r^d Y^m_r (\theta, \phi) | | \ell \| = 0, 1, 2, 3, \ldots; m = -\ell, -\ell + 1, \ldots, +\ell - 1, +\ell \} \), but for \( \ell > 0 \) they all blow up at spatial infinity.

404
to the heat kernel in eq. (9.3.31), 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. (9.3.1) 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. (9.3.1) itself – one time derivative versus two derivatives per spatial coordinate – as well as in the heat kernel solution of eq. (9.3.31). On the other hand, the symmetric portion of the spacetime Green’s functions in equations (9.4.92) and (9.4.93) depend on spacetime solely through \( 2\sigma \equiv (t - t')^2 - (\vec{x} - \vec{x}')^2 \), which is invariant under global Poincaré transformations (cf. eq. (9.4.14)).

4 dimensions: Massless Scalar Field We highlight the 4 dimensional retarded case, because it is most relevant to the real world. Using eq. (9.4.92) after we recognize \( \Theta'(\sigma) = \delta(\sigma) \) and \( \Theta(t' - \sigma) = \delta(t' - \sigma) \)|\(^{25} \) of eq. (9.4.98).

\[
G^+_4(x - x') = \frac{\delta(t - t' - |\vec{x} - \vec{x}'|)}{4\pi|\vec{x} - \vec{x}'|}.
\] (9.4.98)

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 \( 1/(\text{observer-source spatial distance}) \) = \( 1/|\vec{x} - \vec{x}'| \).

Problem 9.19. 3D Green’s function from 4D Can you use eq. (9.4.90) to compute the (2+1)D massless scalar retarded Green’s function in eq. (9.4.74) from its (3+1)D counterpart in eq. (9.4.98)?

The solution to \( \psi \) from eq. (9.4.55) is now

\[
\psi(t, \vec{x}) = \int_{-\infty}^{+\infty} dt' \int_{\mathbb{R}^3} d^3 \vec{x}' G^+_4(t - t', \vec{x} - \vec{x}') J(t', \vec{x}')
= \int_{-\infty}^{+\infty} dt' \int_{\mathbb{R}^3} d^3 \vec{x}' \frac{\delta(t - t' - |\vec{x} - \vec{x}'|)}{4\pi|\vec{x} - \vec{x}'|} J(t', \vec{x}')
= \int_{\mathbb{R}^3} d^3 \vec{x}' \frac{J(t_r, \vec{x}')}{4\pi|\vec{x} - \vec{x}'|}, \quad t_r \equiv t - |\vec{x} - \vec{x}'|.
\] (9.4.99)

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 \( \vec{x} = \vec{x}' = 0 \) 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

\[
|\vec{x} - \vec{x}'| = \exp(-x^j \partial_j) r, \quad r \equiv |\vec{x}|
= r - \vec{x}' \cdot \hat{r} + rO\left(\left(\frac{r'}{r}\right)^2\right), \quad \hat{r} \equiv \frac{\vec{x}^j}{r}, \quad r' \equiv |\vec{x}'|.
\] (9.4.100)

\(^{25}\)As for \( \Theta(t - t')\delta(\sigma) = \delta(t - t' - |\vec{x} - \vec{x}'|) \), there is another term involving \( \delta(t - t' + |\vec{x} - \vec{x}'|) \), but for this to be non-zero \( t - t' = -|\vec{x} - \vec{x}'| < 0 \); this is not allowed by the \( \Theta(t - t') \).
\[ r - \vec{x} \cdot \hat{r} + r' \mathcal{O}\left(\frac{r'}{r}\right) = r \left(1 - \frac{\vec{x}}{r} \cdot \hat{r} + \mathcal{O}\left(\frac{(r')^2}{r^2}\right)\right). \] (9.4.102)

(By dimensional analysis, you should be able to deduce this is, schematically, a power series in \(r'/r\).) This leads us from eq. (9.4.100) to the following far zone scalar solution

\[
\psi(t, \vec{x}) = \frac{1}{4\pi r} \int_{\mathbb{R}^3} d^3\vec{x}' \left\{ 1 + \frac{\vec{x}'}{r} \cdot \hat{r} + \mathcal{O}\left(\frac{(r')^2}{r^2}\right) \right\} 
\times J(t - r + \vec{x}' \cdot \hat{r} + r'\mathcal{O}(r'/r), \vec{x}).
\] (9.4.103)

The term in curly brackets arises from the \(1/|\vec{x} - \vec{x}'|\) portion of the 4D Green's function in eq. (9.4.98). In turn, this far zone leading order \(1/r\) behavior teaches us, the field due to some field always fall off as \(1/(\text{observer-source distance})\). On the other hand, by recognizing \(\vec{x} \cdot \hat{r} + r'\mathcal{O}(r'/r) = \vec{x}' \cdot \hat{r}(1 + \mathcal{O}(r'/r))\), followed by Taylor expanding the time argument of the source,

\[
J(t - |\vec{x} - \vec{x}'|, \vec{x}') = J(t - r, \vec{x}') + \sum_{\ell=1}^{+\infty} \frac{(\vec{x}' \cdot \hat{r})^\ell}{\ell!} (1 + \mathcal{O}(r'/r)) \partial_\ell^t J(t - r, \vec{x}').
\] (9.4.104)

If we associate each time derivative acting on \(J\) to scale as

\[
\partial_\ell^t J \sim \frac{J}{(\text{characteristic timescale of source})^{\ell}},
\] (9.4.105)

then the Taylor expansion in eq. (9.4.104) becomes one in powers of the ratio \(r'/\tau_J \equiv (\text{characteristic size of the source})/(\text{characteristic timescale of source})\). (Recall from eq. (9.4.99) 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 \(r'/r \sim (\text{characteristic size of source})/(\text{observer-source distance})\).

**Relativistic Far Zone**

To sum, if we take the far zone limit – i.e., neglect all \((\text{characteristic size of source})/(\text{observer-source distance}) \ll 1\) corrections – but allow for a fully relativistic source, eq. (9.4.103) now reads

\[
\psi(t, \vec{x}) \approx \frac{1}{4\pi r} \int_{\mathbb{R}^3} d^3\vec{x}' J(t - r + \vec{x}' \cdot \hat{r}, \vec{x}'), \quad r \equiv |\vec{x}|.
\] (9.4.106)

**Non-relativistic Far Zone**

If we further assume the source is non-relativistic, namely \((\text{characteristic size of source})/(\text{timescale of source}) \ll 1\),

\[
\psi(t, \vec{x}) \approx \frac{A(t - r)}{4\pi r},
\]

\[
A(t - r) \equiv \int_{\mathbb{R}^3} d^3\vec{x}' J(t - r, \vec{x}').
\] (9.4.108)

---

\(^{126}\)In Quantum Field Theory, this 1/r is attributed to the massless-ness of the \(\psi\)-particles.
In the far zone the amplitude of the wave falls off with increasing distance as \(1/\text{(observer-source spatial distance)}\); and the time-dependent portion of the wave \(A(t - r)\) is consistent with that of an outgoing wave, one emanating from the source \(J\).

**Problem 9.20. Spherical s-Waves**  
The \(A(t - r)/r\) in eq. (9.4.107) turns out to be an exact solution, despite our arrival at it via a non-relativistic and far zone approximation. Referring to eq. (9.4.99), identify the form of \(J(t', \vec{x}')\) that would yield the following exact solution to \(\partial^2 \psi = J\):

\[
\psi(t, \vec{x}) = \frac{A(t - r)}{4\pi r}, \quad r \equiv |\vec{x}|. 
\]  
(9.4.109)

Hint: \(J\) describes a point charge sitting at the spatial origin, but with a time dependent strength. \(\square\)

**Problem 9.21. Spherical s-Waves vs Plane Waves**  
In this problem, we will compare the homogeneous plane wave solutions in eq. (9.4.36) with the spherical wave in eq. (9.4.109). We will assume the amplitude \(A\) in eq. (9.4.109) admits a Fourier transform:

\[
A(\omega) = \int_{\mathbb{R}} \frac{d\omega}{2\pi} e^{-i\omega \xi} \tilde{A}(\omega). 
\]  
(9.4.110)

Then each frequency mode must itself be a solution to the wave equation:

\[
\psi = \text{Re} \left( \tilde{A}(\omega) e^{-i\omega (t-r)} \right) \frac{e^{-i\omega (t-r)}}{4\pi r}. 
\]  
(9.4.111)

Throughout this analysis, we shall assume the high frequency and far zone limit to hold:

\[
\omega r \gg 1. 
\]  
(9.4.112)

First show that the Minkowski metric in spherical coordinates is

\[
g_{\mu\nu} dx^\mu dx^\nu = dt^2 - dr^2 - r^2 \left( d\theta^2 + \sin^2 \theta d\phi^2 \right). 
\]  
(9.4.113)

Then verify that

\[
g^{\mu\nu} \nabla_\mu \left( \frac{e^{-i\omega (t-r)}}{4\pi r} \right) = 0, 
\]  
(9.4.114)

as long as \(r \neq 0\); as well as the null character of the constant-phase surfaces, in that

\[
g^{\mu\nu} \nabla_\mu \left( \omega(t-r) \right) \nabla_\nu \left( \omega(t-r) \right) = 0. 
\]  
(9.4.115)

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

\[127\text{Even for the relativistic case in eq. (9.4.106), we see from eq. (9.4.104) that it consists of an infinite series of various rank amplitudes that are functions of retarded time } t - r: \psi(t, \vec{x}) = (4\pi r)^{-1} \sum_{l=0}^{\infty} t^{2l} \cdots \vec{r}^2 A_1 \cdots (t - r).\]
exp(−iω(t − r)) in eq. [9.4.111] may be identified with exp(−ik(t − ˆk · x)) in eq. (9.4.36) if we identify the propagation direction ˆr in the former with the propagation direction ˆk in the latter:

\[ \hat{r} \leftrightarrow \hat{k} \quad \text{and} \quad \omega \leftrightarrow k. \]  

(9.4.116)

After all, 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 \(\omega \approx 1/(\text{period of wave})\). 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. (9.4.115),

\[ \Box e^{-i\omega(t-r)} = -2\omega \frac{i}{r} e^{-i\omega(t-r)} \]  

(9.4.117)

instead. Thus, relative to the expectation that \(\Box \sim \omega^2\), the actual result scales as \(1/(\omega r)\) relative to it.

\[ \Box \ll 1/(\omega r) \]  

(9.4.118)

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.

4D photons  

In 4 dimensional flat spacetime, the vector potential of electromagnetism, in the Lorenz gauge

\[ \partial_\mu A^\mu = 0 \quad \text{(Cartesian coordinates)}, \]  

(9.4.119)

obeys the wave equation

\[ \partial^2 A^\mu = J^\mu. \]  

(9.4.120)

Here, \(\partial^2\) is the scalar wave operator, and \(J^\mu\) is a conserved electromagnetic current describing the motion of some charge density

\[ \partial_\mu J^\mu = \partial_t J^t + \partial_i J^i = 0. \]  

(9.4.121)

The electromagnetic fields are the “curl” of the vector potential

\[ F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \]  

(9.4.122)
In particular, for a given inertial frame, the electric $E$ and magnetic $B$ fields are, with $i, j, k \in \{1, 2, 3\}$,

$$
E^i = F^{i0} = \partial^i A^0 - \partial^0 A^i = -\partial_i A_0 + \partial_0 A_i = -F_{i0}, \tag{9.4.123}
$$

$$
B^k = -\epsilon^{ijk} \partial_i A_j = -\frac{1}{2} \epsilon^{ijk} F_{ij}, \quad \epsilon^{123} \equiv 1. \tag{9.4.124}
$$

**Problem 9.22. Lorenz Gauge, Relativity & Current Conservation**

Comparison of equations (9.4.106) and (9.4.120) indicates, in the far zone,

$$
A^\mu(t, \vec{x}) \approx \frac{1}{4\pi r} \int d^3 \vec{x}' J^\mu(t - r + \vec{x}' \cdot \hat{r}, \vec{x}'). \tag{9.4.125}
$$

If one takes the non-relativistic limit too (cf. eq. (9.4.107)),

$$
A^\mu(t, \vec{x}) \approx \frac{1}{4\pi r} \int d^3 \vec{x}' J^\mu(t - r, \vec{x}'). \tag{9.4.126}
$$

Compute $\partial_\mu A^\mu$ using equations (9.4.125) and (9.4.126) to leading order in $1/r$. Hint: a key step is to recognize, for a conserved current obeying eq. (9.4.121),

$$
\partial_\tau J^0(\tau, \vec{x}') = -\left( \partial_{\vec{r}} J^i(\tau, \vec{x}') \right)_i + \hat{r}^i \partial_\tau J^i(\tau, \vec{x}'); \tag{9.4.127}
$$

$$
\tau \equiv t - r + \vec{x}' \cdot \hat{r}, \quad \partial_{\vec{r}} \equiv \frac{\partial}{\partial \vec{x}^i}; \tag{9.4.128}
$$

where the subscript $t$ on the first term on the right-hand-side of eq. (9.4.127) means the spatial derivatives are carried out with the observation time $t$ held fixed – which is to be distinguished from doing so but with $\tau$ held fixed.

You should find that the Lorenz gauge in eq. (9.4.119) is respected only by the relativistic solution in eq. (9.4.125), and not by the non-relativistic one in eq. (9.4.126). This is an important point because, even though the Lorenz gauge in eq. (9.4.119) 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. (9.4.120) to obtain $\partial^2(\partial_\mu A^\mu) = \partial_\mu J^\mu$. \hfill \Box

**Problem 9.23. Electromagnetic radiation zone**

Using $G^μ_μ$ in eq. (9.4.98), write down the solution of $A^\mu$ in terms of $J^\mu$. Like the scalar case, take the far zone limit. In this problem we wish to study some basic properties of $A^\mu$ 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 $\partial_\tau J^0 = -\partial_i J^i$, the conservation of the current, show that $A^0$ is independent of time in the far zone limit.

2. Now define the dipole moment as

$$
I^i(t) \equiv \int_{\mathbb{R}^3} d^3 \vec{x}' x'^i J^0(t, \vec{x}'). \tag{9.4.129}
$$
Can you show its first time derivative is
\[ \dot{I}^i(t) \equiv \frac{dI^i(t)}{dt} = \int_{\mathbb{R}^3} d^3\vec{x}' J^i(t, \vec{x}')? \] (9.4.130)

(Hint: Use current conservation and integration-by-parts.)

3. From this, we shall infer it is the ‘transverse’ part of $A^i$ that contains radiative effects. First show that in the far zone, i.e., to leading order in $1/r$,
\[ E^i(t, \vec{x}) \to -\frac{1}{4\pi r} \frac{d^2 I^i(t - r, \hat{r})}{dt^2}, \]
\[ B^i(t, \vec{x}) \to -\frac{1}{4\pi r} \epsilon^{ijk} \frac{d^2 I^j(t - r, \hat{r})}{dt^2} = \epsilon^{ijk} E^k, \]
\[ \vec{I}^i(s, \hat{r}) \equiv P^{ij}(\hat{r}) I^j(s), \quad P^{ij} \equiv \delta^{ij} - \hat{r}^i \hat{r}^j, \quad \hat{r}^i \equiv \frac{x^i}{|\vec{x}|}. \] (9.4.133)

The subscript ‘t’ stands for ‘transverse’; the projector, which obeys $\hat{r}^i P^{ij} = 0$, ensures the $I^i_t$ now consists only of the ‘transverse’ portion of the dipole moment: $\hat{r}^i I^i_t = 0$. Notice, this result indicates not only the electric and magnetic fields are mutually perpendicular, they are also perpendicular to the radial direction – i.e., they are transverse to the propagation direction of (far zone) electromagnetic radiation.

Can you explain whether the results in equations (9.4.131) and (9.4.132) 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. (9.4.129)? That is, what becomes of $\vec{E}$ and $\vec{B}$ if instead of eq. (9.4.129), we defined
\[ I^i(t) \equiv \int_{\mathbb{R}^3} d^3\vec{x}' (x^i - b^i) J^0(t, \vec{x}')? \] (9.4.134)

4. Use the above results in equations (9.4.131) and (9.4.132) 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 $(\vec{E}^2 + \vec{B}^2)/2$. You should find that, with a non-relativistic current, both $\vec{E}$ and $\vec{B}$ in the far zone can be expressed solely in terms of the transverse dipole moment $I^i_t$.

5. Verify the following projector property of $P^{ij}$ in eq. (9.4.133):
\[ P^{ia} P^{ib} = P^{ab}. \] (9.4.135)

Show that the dot product of the Poynting vector with the unit radial vector is
\[ \vec{S} \cdot \hat{r} = \frac{1}{(4\pi)^2 r^2} \left( \vec{P}^2 - \left( \vec{r} \cdot \vec{P} \right)^2 \right). \] (9.4.136)

For an arbitrary unit vector $\hat{n}$, the dot product $\vec{S}(t, \vec{x}) \cdot \hat{n}(t, \vec{x})$ is the energy per unit time per unit area passing through the infinitesimal plane orthogonal to the vector $\hat{n}$ based at
The quantity in eq. \((9.4.136)\), if integrated over the 2-sphere, therefore describes the loss of energy to infinity as \(r \to \infty\). Physically speaking, the non-zero acceleration of the dipole moment responsible for electromagnetic radiation indicates work needs to be done pushing around electric charges, i.e., forces are needed to give rise to acceleration.

**4D gravitational waves**

In a 4D weakly curved spacetime, the metric can be written as one deviating slightly from Minkowski,

\[
g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \quad \text{(Cartesian coordinates),} \tag{9.4.137}
\]

where the dimensionless components of \(h_{\mu\nu}\) are assumed to be much smaller than unity.

The (trace-reversed) graviton

\[
\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \eta^{\alpha\beta} h_{\alpha\beta}, \tag{9.4.138}
\]

in the de Donder gauge

\[
\partial^\mu \bar{h}_{\mu\nu} = \partial_t \bar{h}_{t\nu} - \delta^{ij} \partial_i \bar{h}_{j\nu} = 0, \tag{9.4.139}
\]

obeys the wave equation \(128\)

\[
\partial^2 \bar{h}_{\mu\nu} = -16\pi G_N T_{\mu\nu} \quad \text{(Cartesian coordinates).} \tag{9.4.140}
\]

(The \(G_N\) is the same Newton’s constant you see in Newtonian gravity \(\sim G_N M_1 M_2 / r^2\); both \(\bar{h}_{\mu\nu}\) and \(T_{\mu\nu}\) are symmetric.) The \(T_{\mu\nu}\) is a 4 \(\times\) 4 matrix describing the energy-momentum-shear-stress of matter, and has zero divergence (i.e., it is conserved) whenever the matter is held together primarily by non-gravitational forces \(129\)

\[
\partial_\mu T^{\mu\nu} = \partial_t T^{t\nu} + \partial_i T^{i\nu} = 0. \tag{9.4.141}
\]

**Problem 9.24. de Donder Gauge, Relativity & Energy-Momentum Conservation**

Comparison of equations \(9.4.106\) and \(9.4.120\) indicates, in the far zone,

\[
\bar{h}^{\mu\nu}(t, \vec{x}) \approx -\frac{4 G_N}{r} \int d^3 \vec{x}\, J^\mu (t - r + \vec{x} \cdot \hat{r}, \vec{x}'). \tag{9.4.142}
\]

If one takes the non-relativistic limit too (cf. eq. \(9.4.107\)),

\[
\bar{h}^{\mu\nu}(t, \vec{x}) \approx -\frac{4 G_N}{r} \int d^3 \vec{x}\, J^\mu (t - r, \vec{x}'). \tag{9.4.143}
\]

Compute \(\partial_\mu \bar{h}^{\mu\nu}\) using equations \(9.4.142\) and \(9.4.143\) to leading order in \(1/r\). Hint: a key step is to recognize, for a conserved energy-momentum-stress tensor obeying eq. \(9.4.141\),

\[
\partial_t T^{0\mu}(\tau, \vec{x}') = - \left( \partial_t T^{i\mu}(\tau, \vec{x}') \right)_t + \hat{r}^i \partial_i T^{i\mu}(\tau, \vec{x}'); \tag{9.4.144}
\]

\(128\)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.

\(129\)For systems held together primarily by gravity, such as the Solar System or compact binary black hole(s)/neutron star(s) emitting gravitational radiation, their stress tensor will not be divergence-less.
\[ \tau \equiv t - r + \vec{x}' \cdot \hat{r}, \quad \partial_{\nu} \equiv \frac{\partial}{\partial x'_{\nu}}; \]  

(9.4.145)

where the subscript \( t \) on the first term on the right-hand-side of eq. (9.4.144) means the spatial derivatives are carried out with the observation time \( t \) held fixed – which is to be distinguished from doing so but with \( \tau \) held fixed.

You should find that the de Donder gauge in eq. (9.4.139) is respected only by the relativistic solution in eq. (9.4.142), and not by the non-relativistic one in eq. (9.4.143). This is an important point because, even though the de Donder gauge in eq. (9.4.139) 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. (9.4.140) to obtain \( \partial^{2}(\partial^{\mu}\bar{h}_{\mu\nu}) = -16\pi G_{N}\partial^{\mu}T_{\mu\nu}. \) 

\[ \Box \]

**Problem 9.25. Gravitational radiation zone**

Can you carry out a similar analysis in Problem (9.23), but for gravitational radiation? Using \( G_{4}^{+} \) in eq. (9.4.98), 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 \( \partial_{t}T^{\nu} = -\partial_{i}T^{i\nu} \), 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.

2. Now define the quadrupole moment as

\[ I^{ij}(t) \equiv \int_{\mathbb{R}^{3}} d^{3}\vec{x}' x'^{i} x'^{j} T^{00}(t, \vec{x}'). \]  

(9.4.146)

Can you show its second time derivative is

\[ \dddot{I}^{ij}(t) \equiv \frac{d^{2}I^{ij}(t)}{dt^{2}} = 2 \int_{\mathbb{R}^{3}} d^{3}\vec{x}' T^{ij}(t, \vec{x}'). \]  

(9.4.147)

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:

\[ \bar{h}^{ij}(t, \vec{x}) \rightarrow -\frac{2G_{N}}{r} \frac{d^{2}I^{ij}(t - r)}{dt^{2}}, \quad r \equiv |\vec{x}|. \]  

(9.4.148)

Can you explain what would become of this result if, instead of the quadrupole moment defined in eq. (9.4.146), we had shifted its integration origin by a constant vector \( \vec{b} \), namely

\[ I^{ij}(t) \equiv \int_{\mathbb{R}^{3}} d^{3}\vec{x}'(x'^{i} - b^{i})(x'^{j} - b^{j}) T^{00}(t, \vec{x}'). \]  

(9.4.149)

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.

\[ h_{ij}^{tt} = P_{ijab}h_{ab}, \]  

(9.4.150)

412
\[ P_{ijab} \equiv \frac{1}{2} (P_{ia} P_{jb} + P_{ib} P_{ja} - P_{ij} P_{ab}) , \quad \text{(cf. eq. (9.4.133))}, \quad (9.4.151) \]
\[ \tilde{r}^j h^{tt}_{ij} = 0 \quad \text{(Transverse)} \quad \delta^{ij} h^{tt}_{ij} = 0 \quad \text{(Traceless)}. \quad (9.4.152) \]

Averaged over multiple wavelengths, the energy-momentum-stress tensor of gravitational waves takes the form
\[ \langle t_{\mu \nu} [h^{tt}] \rangle = \frac{1}{32\pi G_N} \langle \partial_{\mu} h^{tt}_{ij} \partial_{\nu} h^{tt}_{ij} \rangle. \quad (9.4.153) \]

Can you work out the energy density \( E \equiv \langle t_{00} \rangle \) and the momentum flux \( P_i \equiv \langle t_{0i} \rangle = -\langle t_{0i} \rangle \) (the gravitational analog to the electromagnetic Poynting vector) in terms of the quadrupole moment?

**Problem 9.26. Waves Around Schwarzschild Black Hole.** The geometry of a non-rotating black hole is described by
\[ ds^2 = \left( 1 - \frac{r_s}{r} \right) dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}} - r^2 (d\theta^2 + \sin(\theta)^2 d\phi^2) , \quad (9.4.154) \]
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
\[ \square \psi(t, r, \theta, \phi) = \nabla_\mu \nabla^\mu \psi = 0. \quad (9.4.155) \]

Consider the following separation-of-variables ansatz
\[ \psi(t, r, \theta, \phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-i\omega t} \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{+\ell} \frac{R_{\ell}(\omega r_s)}{r} Y^m_{\ell}(\theta, \phi) , \quad (9.4.156) \]
where \( \{Y^m_{\ell}\} \) are the spherical harmonics on the 2-sphere and the “tortoise coordinate” is
\[ r_* \equiv r + r_s \ln \left( \frac{r}{r_s} - 1 \right). \quad (9.4.157) \]

Show that the wave equation is reduced to an ordinary differential equation for the \( \ell \)th radial mode function
\[ R''_{\ell}(\xi_* \equiv \omega r) + \left( \frac{\xi_*^2}{\xi^4} + \frac{(\ell(\ell+1) - 1) \xi_*}{\xi^3} - \frac{\ell(\ell+1)}{\xi^2} + 1 \right) R_{\ell}(\xi_*) = 0, \quad (9.4.158) \]
where \( \xi \equiv \omega r, \xi_* \equiv \omega r_s \) and \( \xi_* \equiv \omega r_* \).

An alternative route is to first perform the change-of-variables
\[ x \equiv 1 - \frac{\xi}{\xi_*} , \quad (9.4.159) \]
and the change of radial mode function

\[
\frac{R_\ell(\xi_*)}{r} \equiv \frac{Z_\ell(x)}{\sqrt{x(1-x)}}. \tag{9.4.160}
\]

Show that this returns the ODE

\[
Z''_\ell(x) + \left(\frac{1}{4(1-x)^2} + \frac{1 + 4\xi_*^2}{4x^2} + \xi_*^2 + \frac{2\ell(\ell + 1) + 1 - 4\xi_*^2}{2x} - \frac{2\ell(\ell + 1) + 1}{2(x-1)}\right)Z_\ell(x) = 0.
\]

\[
(9.4.161)
\]

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. (9.4.161) 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.

9.4.3 4D frequency space, Static limit, Discontinuous first derivatives

Wave Equation in Frequency Space

We begin with eq. (9.4.53), and translate it to frequency space.

\[
\psi(t, \vec{x}) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \tilde{\psi}(\omega, \vec{x}) e^{-i\omega t} = \int_{-\infty}^{+\infty} dt'' \int_{\mathbb{R}^D} d^{D+1} \vec{x}'' G_{D+1}(t-t'', \vec{x} - \vec{x}'') \tilde{J}(\omega, \vec{x}'')e^{-i\omega t''}
\]

\[
= \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \int_{-\infty}^{+\infty} d(t-t') e^{i\omega(t-t')} e^{-i\omega t} \int_{\mathbb{R}^D} d^{D} \vec{x}'' G_{D+1}(t-t'', \vec{x} - \vec{x}'') \tilde{J}(\omega, \vec{x}'')
\]

\[
= \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \int_{\mathbb{R}^D} d^{D} \vec{x}'' \tilde{G}_{D+1}(\omega, \vec{x} - \vec{x}'') \tilde{J}(\omega, \vec{x}''). \tag{9.4.162}
\]

Equating the coefficients of \(e^{-i\omega t}\) on both sides,

\[
\tilde{\psi}(\omega, \vec{x}) = \int_{\mathbb{R}^D} d^{D} \vec{x}'' \tilde{G}_{D+1}(\omega, \vec{x} - \vec{x}'') \tilde{J}(\omega, \vec{x}''); \tag{9.4.163}
\]

\[
\tilde{G}_{D+1}(\omega, \vec{x} - \vec{x}'') \equiv \int_{-\infty}^{+\infty} d\tau e^{i\omega \tau} G_{D+1}(\tau, \vec{x} - \vec{x}''). \tag{9.4.164}
\]

Equation (9.4.163) tells us that the \(\omega\)-mode of the source is directly responsible for that of the field \(\tilde{\psi}(\omega, \vec{x})\). 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. (9.4.98), but replace the \(\delta\)-function with its integral representation. This leads us to \(\tilde{G}_{D+1}^+(\omega, \vec{x} - \vec{x}'')\), the frequency space representation of the retarded Green’s function of the wave operator.

\[
G_{D+1}^+(x - x') = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \exp\left(-i\omega(t - t' - |\vec{x} - \vec{x}'|)\right)
\]

\[
\frac{1}{4\pi |\vec{x} - \vec{x}'|}.
\]
\[
\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \tilde{G}^+_4(\omega, \vec{x} - \vec{x}'),
\]
(9.4.165)

where

\[
\tilde{G}^+_4(\omega, \vec{x} - \vec{x}') \equiv \frac{\exp (i\omega|\vec{x} - \vec{x}'|)}{4\pi|\vec{x} - \vec{x}'|}.
\]
(9.4.166)

As we will see, \(\omega\) 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. (9.4.56) in frequency space reads

\[
\left(\partial_0^2 - \vec{\nabla}^2\right) \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \exp \left(-i\omega(t-t' - |\vec{x} - \vec{x}'|)\right) \delta(t-t')\delta^{(3)}(\vec{x} - \vec{x}') ,
\]
(9.4.167)

\[
\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \left(-\omega^2 - \vec{\nabla}^2\right) \frac{\exp (i\omega|\vec{x} - \vec{x}'|)}{4\pi|\vec{x} - \vec{x}'|} = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \delta^{(3)}(\vec{x} - \vec{x}') ,
\]
(9.4.168)

– where \(\partial_0^2\) can be either \(\partial_t^2\) or \(\partial_{\vec{x}}^2\); \(\vec{\nabla}^2\) 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 \(\omega\), the frequency space Green’s function of eq. (9.4.166) obeys the PDE

\[
\left(-\omega^2 - \vec{\nabla}^2\right) \tilde{G}^+_4(\omega, \vec{x} - \vec{x}') = \delta^{(3)}(\vec{x} - \vec{x}').
\]
(9.4.169)

**Problem 9.27. Far Zone In Frequency Space**

Show that the frequency transform of the far zone wave in eq. (9.4.106) is

\[
\tilde{\psi}(\omega, \vec{x}) \approx \frac{e^{i\omega r}}{4\pi r} \tilde{J}(\omega, \omega\vec{r}) ,
\]
(9.4.170)

where

\[
\tilde{J}(\omega, \vec{k}) \equiv \int_{\mathbb{R}} dt \int_{\mathbb{R}^3} d^3\vec{x} e^{+i\omega t} e^{-i\vec{k}\cdot\vec{x}} J(t, \vec{x}) .
\]
(9.4.171)

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. (9.4.15) now reads

\[
\nabla_\mu \nabla^\mu G = g^{tt} \partial_t \partial_t G + g^{ti} \partial_t \partial_i G + \frac{\partial_t \left(\sqrt{|g|}g^{ti} \partial_t G\right)}{\sqrt{|g|}} + \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|}g^{ij} \partial_j G\right)
\]
In the zero frequency limit (\(\omega\) zero frequency limit, \(\omega\) \(\rightarrow\) \(-\infty\)) an outward external force, for otherwise the mass will move towards the center of the star.

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 charge/mass held still at \(\vec{x}'\) from past infinity to future infinity.

### 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. (9.4.98), we may apply the infinite time integral in eq. (9.4.176).

\[
G^{(\text{static})}(\vec{x}, \vec{x}') = \int_{\tau = 0}^{\infty} d\tau e^{i\omega\tau} G^+(\tau; \vec{x}, \vec{x}') \left(\frac{\delta(t - t')}{\sqrt{|g(\vec{x})|}}\right),
\]

(9.4.173)

and note that solving the static equation

\[
\nabla_\nu \nabla^\mu G^{(\text{static})}(\vec{x}, \vec{x}') = \frac{\partial_i \left(\sqrt{|g(\vec{x})|} g^{ij}(\vec{x}) \partial_j G^{(\text{static})}(\vec{x}, \vec{x}')\right)}{\sqrt{|g(\vec{x})|}} = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{|g(\vec{x})|}}
\]

(9.4.174)

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 \(\delta\)-functions on the right hand side is \(D\)-dimensional.

The reason is the frequency transform of eq. (9.4.172) replaces \(\partial_i \rightarrow -i\omega\) and the \(\delta(t - t')\) on the right hand side with unity.

\[
g^{tt}(-i\omega)^2 \tilde{G} + g^{tt}(-i\omega) \partial_t \tilde{G} + \frac{\partial_i \left(\sqrt{|g(\vec{x})|} g^{ij}(-i\omega)G\right)}{\sqrt{|g|}} + \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|} g^{ij} J^j \tilde{G}\right) = \frac{\delta^{(D)}(\vec{x} - \vec{x}')}{\sqrt{|g(\vec{x})|}}
\]

(9.4.175)

In the zero frequency limit \((\omega \rightarrow 0)\) we obtain eq. (9.4.174). And since the static limit is the zero frequency limit,

\[
G^{(\text{static})}(\vec{x}, \vec{x}') = \lim_{\omega \rightarrow 0} \int_{\tau = 0}^{\infty} d\tau e^{i\omega\tau} G^+(\tau; \vec{x}, \vec{x}')
\]

(9.4.176)

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.\[130\]

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.

130Note, 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.
\[-\delta^i \partial_i \partial_j G^{(\text{static})}(\vec{x}, \vec{x}') = -\vec{\nabla}^2 G^{(\text{static})}(\vec{x}, \vec{x}') = \delta^{(3)}(\vec{x} - \vec{x}'). \tag{9.4.178}\]

On the other hand, we may also take the zero frequency limit of eq. \[9.4.166\] to arrive at the same answer.

\[
\lim_{\omega \to 0} \frac{\exp(i\omega |\vec{x} - \vec{x}'|)}{4\pi |\vec{x} - \vec{x}'|} = \frac{1}{4\pi |\vec{x} - \vec{x}'|}. \tag{9.4.179}\]

**Problem 9.28. Discontinuous first derivatives of the radial Green's function**  In this problem we will understand the discontinuity in the radial Green’s function of the frequency space retarded Green’s function in 4D Minkowski spacetime. We begin by switching to spherical coordinates and utilizing the following ansatz

\[
\tilde{G}_4^+(\omega, \vec{x} - \vec{x}') = \sum_{\ell=0}^{\infty} \tilde{g}_\ell(r, r') \sum_{m=-\ell}^{\ell} Y_\ell^m(\theta, \phi) Y_\ell^m(\theta', \phi'),
\]

\[
\vec{x} = r(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \quad \vec{x}' = r'(\sin \theta' \cos \phi', \sin \theta' \sin \phi', \cos \theta'). \tag{9.4.180}\]

Show that this leads to the following ODE(s) for the \(\ell\)th radial Green’s function \(\tilde{g}_\ell\):

\[
\frac{1}{r^2} \partial_r \left( r^2 \partial_r \tilde{g}_\ell \right) + \left( \omega^2 - \frac{\ell(\ell + 1)}{r^2} \right) \tilde{g}_\ell = -\frac{\delta(r - r')}{rr'}, \tag{9.4.181}\]

\[
\frac{1}{r^2} \partial_r' \left( r'^2 \partial_r \tilde{g}_\ell \right) + \left( \omega^2 - \frac{\ell(\ell + 1)}{r'^2} \right) \tilde{g}_\ell = -\frac{\delta(r - r')}{rr'}. \tag{9.4.182}\]

Because \(\tilde{G}_4^+(\omega, \vec{x} - \vec{x}') = \tilde{G}_4^+(\omega, \vec{x}' - \vec{x})\), 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: \(\tilde{g}(r, r') = \tilde{g}(r', r)\). That means the results in \(§9.6\) may be applied here. Show that

\[
\tilde{g}_\ell(r, r') = i\omega j_\ell(\omega r_<) h_{\ell}^{(1)}(\omega r_>), \tag{9.4.183}\]

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. \[9.6.52\] is recovered, by taking the limits \(\omega r, \omega r' \to 0\).

Some useful formulas include

\[
j_\ell(x) = (-x)^\ell \left( \frac{1}{x} \frac{d}{dx} \right)^\ell \frac{\sin x}{x}, \quad h_{\ell}^{(1)}(x) = -i(-x)^\ell \left( \frac{1}{x} \frac{d}{dx} \right)^\ell \frac{\exp(ix)}{x}, \tag{9.4.184}\]

their small argument limits

\[
j_\ell(x \ll 1) \to \frac{x^\ell}{(2\ell + 1)!!} \left( 1 + \mathcal{O}(x^2) \right), \quad h_{\ell}^{(1)}(x \ll 1) \to -\frac{i(2\ell - 1)!!}{x^{\ell + 1}} \left( 1 + \mathcal{O}(x) \right), \tag{9.4.185}\]

as well as their large argument limits

\[
j_\ell(x \gg 1) \to \frac{1}{x} \sin \left( x - \frac{\pi \ell}{2} \right), \quad h_{\ell}^{(1)}(x \gg 1) \to (-i)^{\ell + 1} e^{ix} x. \tag{9.4.186}\]
Their Wronskian is
\[
\text{Wr}_z \left( j_\ell(z), h^{(1)}_\ell(z) \right) = \frac{i}{z^2}.
\] (9.4.187)

Hints: First explain why
\[
\bar{g}_\ell(r, r') = A_1^\ell j_\ell(\omega r) j_\ell(\omega r') + A_2^\ell h^{(1)}_\ell(\omega r) h^{(1)}_\ell(\omega r') + \mathcal{G}_\ell(r, r'),
\] (9.4.188)
\[
\mathcal{G}_\ell(r, r') \equiv F \left\{ (\chi_\ell - 1) j_\ell(\omega r) h^{(1)}_\ell(\omega r') + \chi_\ell \cdot j_\ell(\omega r) h^{(1)}_\ell(\omega r') \right\},
\] (9.4.189)

where \( A_1^\ell, A_2^\ell, F \) and \( \chi_\ell \) are constants. Fix \( F \) by ensuring the “jump” in the first \( r \)-derivative at \( r = r' \) yields the correct \( \delta \)-function measure. Then consider the limits \( r \to 0 \) and \( r \gg r' \). For the latter, note that
\[
|\vec{x} - \vec{x}'| = e^{-\vec{x}' \cdot \vec{n}} |\vec{x}| = |\vec{x}| \left( 1 - (r'/r) \hat{n} \cdot \hat{n}' + \mathcal{O}((r'/r)^2) \right),
\] (9.4.190)
where \( \hat{n} \equiv \vec{x}/r \) and \( \hat{n}' \equiv \vec{x}'/r' \).

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
\[
-\vec{\nabla}^2 \psi(\vec{x}) = J(\vec{x})
\] (9.4.191)
is solved via the static limit of the 4D retarded Green’s function we have been discussing. This static limit is given in eq. (9.6.52) 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., \( J(\vec{r} > R) = 0 \). Therefore we may replace \( r < \to r' \) and \( r > \to r \) in eq. (9.6.52), and the exact solution to \( \psi \) now reads
\[
\psi(\vec{x}; r > R) = \int_{\mathbb{R}^3} d^3\vec{x}' G(\vec{x} - \vec{x}') J(\vec{x}') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} \frac{\rho_{\ell m}^m}{2\ell + 1} \frac{Y_{\ell m}^m(\theta, \phi)}{r^{\ell+1}},
\] (9.4.192)
where the multipole moments \( \{\rho_{\ell m}^m\} \) are defined
\[
\rho_{\ell m}^m \equiv \int_{S^2} d(\cos \theta') d\phi' \int_0^\infty dr' r'^{\ell+2} Y_{\ell m}^m(\theta', \phi') J(r', \theta', \phi').
\] (9.4.193)

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 \( \rho_{\ell m}^m \) in eq. (9.4.193) with larger \( \ell \) 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 $Y_{\ell m}(\vartheta, \varphi)/r^{\ell+1}$ multiplying the $(\ell, m)$-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 $\psi$ 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. (9.6.52) that allowed us to cleanly separate the contribution from the source (the multipoles $\{\rho_{\ell m}^m\}$) 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 $\tilde{J}$, i.e., mathematically, $\omega r \gg 1$. We begin with eq. (9.4.183) inserted into eq. (9.4.180).

$$\tilde{G}_4^+ (\omega, \vec{x} - \vec{x}') = \frac{\exp (i\omega|\vec{x} - \vec{x}'|)}{4\pi|\vec{x} - \vec{x}'|}$$
$$= i\omega \sum_{\ell=0}^{\infty} j_\ell(\omega r') h^{(1)}_\ell(\omega r) \sum_{m=-\ell}^{\ell} Y_{\ell m}(\vartheta, \varphi) Y_{\ell m}(\vartheta', \varphi')^*$$

(9.4.194)
(9.4.195)

Our far zone assumptions means we may replace the Hankel function in eq. (9.4.183) with its large argument limit in eq. (9.4.186).

$$\tilde{G}_4^+ (\omega r \gg 1) = \frac{e^{i\omega r}}{r} \left(1 + O\left((\omega r)^{-1}\right)\right) \sum_{\ell=0}^{\infty} (-i)^\ell j_\ell(\omega r') \sum_{m=-\ell}^{\ell} Y_{\ell m}(\vartheta, \varphi) Y_{\ell m}(\vartheta', \varphi')^*.$$  
(9.4.196)

Applying this limit to the general wave solution in eq. (9.4.163),

$$\tilde{\psi}(\omega, \vec{x}) = \int_{\mathbb{R}^3} d^3 \vec{x}'' \tilde{G}_4^+ (\omega, \vec{x} - \vec{x}'') \tilde{J}(\omega, \vec{x}''),$$

(9.4.197)

$$\tilde{\psi}(\omega r \gg 1) \approx \frac{e^{i\omega r}}{r} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{Y_{\ell m}(\vartheta, \varphi)}{2\ell + 1} \tilde{\Omega}_\ell^m(\omega),$$

(9.4.198)

where now the frequency dependent multipole moments are defined as

$$\tilde{\Omega}_\ell^m(\omega) \equiv (2\ell + 1)(-i)^\ell \int_{S^2} d(\cos \theta') d\varphi' \int_0^{\infty} dr' r'^2 j_\ell(\omega r') Y_{\ell m}(\vartheta', \varphi') \tilde{J}(\omega, r', \theta', \varphi').$$  
(9.4.199)

Problem 9.29. Far zone in position/real space Use the plane wave expansion in eq. (9.2.104) to show that eq. (9.4.198) is equivalent to eq. (9.4.170).

Low frequency limit equals slow motion limit How are the multipole moments $\{\rho_{\ell m}^m\}$ in eq. (9.4.193) (which are pure numbers) related to the frequency dependent ones $\{\tilde{\Omega}_\ell^m(\omega)\}$ in eq. (9.4.199)? 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. (9.4.185).

$$\tilde{\Omega}_m^m(\omega R \ll 1) \approx \frac{(-i\omega)^\ell}{(2\ell - 1)!!} (1 + \mathcal{O}(\omega^2)) \int_{\mathbb{S}^2} d(\cos \theta') d\phi' \int_0^\infty dr' r'^{2+\ell} Y^m_\ell(\theta', \phi') \tilde{J}(\omega, r', \theta', \phi')$$

(9.4.200)

Another way to see this “small $\omega$ equals slow motion limit” is to ask: what is the real time representation of these $\{\tilde{\Omega}_m^m(\omega R \ll 1)\}$? By recognizing every $-i\omega$ as a $t$-derivative,

$$\Omega^m_\ell(t) \equiv \int_{\mathbb{R}} \frac{d\omega}{2\pi} e^{-i\omega t} \tilde{\Omega}_m^m(\omega)$$

$$\approx \frac{\partial^\ell_t}{(2\ell - 1)!!} \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \int_{\mathbb{S}^2} d(\cos \theta') d\phi' \int_0^\infty dr' r'^{2+\ell} Y^m_\ell(\theta', \phi') \tilde{J}(\omega, r', \theta', \phi'),$$

$$\equiv \frac{\partial^\ell_t \rho^m_\ell(t)}{(2\ell - 1)!!}. \quad (9.4.201)$$

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. (9.4.192) comes from the monopole.

**Spherical waves in small $\omega$ 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 $t - r$.

$$\psi(t, \vec{x}) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \tilde{\psi}(\omega, \vec{x})$$

$$\approx \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{i\omega(r - t)} \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \frac{Y^m_\ell(\theta, \phi)}{2\ell + 1} \tilde{\Omega}_m^m(\omega), \quad \text{(Far zone spherical wave expansion)}$$

$$\approx \frac{1}{r} \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \frac{Y^m_\ell(\theta, \phi)}{(2\ell + 1)!!} \frac{d^\ell \rho^m_\ell(t - r)}{dt^\ell}, \quad \text{(Slow motion limit).} \quad (9.4.202)$$

### 9.4.4 Initial value problem via Kirchhoff representation

**Massless scalar fields** Previously we showed how, if we specified the initial conditions for the scalar field $\psi$ – then via their Fourier transforms – eq. (9.4.45) tells us how they will evolve forward in time. Now we will derive an analogous expression that is valid in curved spacetime, using the retarded Green’s function $G^+_{D+1}$. To begin, the appropriate generalization of equations (9.4.16) and (9.4.56) are

$$\Box_x \psi(x) = J(x),$$

420
\[
\Box^+_s G^+_{D+1}(x, x') = \Box^+_s G^+_{D+1}(x, x') = \frac{\delta^{(D+1)}(x-x')}{\sqrt{|g(x)g(x')|}}. \tag{9.4.203}
\]

The derivation is actually very similar in spirit to the one starting in eq. \(9.1.44\). Let us consider some ‘cylindrical’ domain of spacetime \(D\) with spatial boundaries \(\partial D_s\) that can be assumed to be infinitely far away, and ‘constant time’ hypersurfaces \(\partial D(t)\) \((\text{final time } t_f)\) and \(\partial D(t_0)\) \((\text{initial time } t_0)\). (These constant time hypersurfaces need not correspond to the same time coordinate used in the integration.) We will consider an observer residing \((\text{at } x)\) within this domain \(D\).

\[
I(x \in D) \equiv \int_\partial D d^{D+1}x' \sqrt{|g(x')|} \left\{ G^+_{D+1}(x, x') \Box^+_s \psi(x') - \Box^+_s G^+_{D+1}(x, x') \cdot \psi(x') \right\} = \int_\partial D d^{D}\alpha' \left\{ G^+_{D+1}(x, x') \nabla^\alpha' \psi(x') - \nabla^\alpha' G^+_{D+1}(x, x') \cdot \psi(x') \right\} \tag{9.4.204}
\]

The terms in the very last line cancel. What remains in the second equality is the surface integrals over the spatial boundaries \(\partial D_s\), and constant time hypersurfaces \(\partial D(t_f)\) and \(\partial D(t_0)\) – where we have used the Gauss’ theorem in eq. \(7.4.38\). 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 \(\partial D(t_f)\) cannot contribute to the integral because it lies to the future of \(x\). Also, if we assume that \(G^+_{D+1}(x, x')\), like its Minkowski counterpart, vanishes outside the past light cone of \(x\), then the spatial boundaries at infinity also cannot contribute.\(^{131}\) (Drawing a spacetime diagram here helps.)

Within eq. \(9.4.204\), if we now proceed to invoke the equations obeyed by \(\psi\) and \(G^+_{D+1}\) in eq. \(9.4.203\), what remains is

\[
- \psi(x) + \int_\partial D d^{D+1}x' \sqrt{|g(x')|} G^+_{D+1}(x, x') J(x')
\]

\[
= - \int_{\partial D(t_0)} d^D\xi \sqrt{|H(\tilde{\xi})|} \left\{ G^+_{D+1} \left( x, x'(\tilde{\xi}) \right) n'^\nu \nabla_{\alpha'} \psi \left( x'(\tilde{\xi}) \right) - n'^\nu \nabla_{\alpha'} G^+_{D+1} \left( x, x'(\tilde{\xi}) \right) \cdot \psi \left( x'(\tilde{\xi}) \right) \right\}.
\]

Here, we have assumed there are \(D\) coordinates \(\tilde{\xi}\) such that \(x'^\mu(\tilde{\xi})\) parametrizes our initial time hypersurface \(\partial D(t_0)\). The \(\sqrt{|H|}\) is the square root of the determinant of its induced metric. More specifically,

\[
H_{ij}(\tilde{\xi})d\xi^i d\xi^j = \left( g_{\mu\nu}(x(\tilde{\xi})) \frac{\partial x^\mu}{\partial \xi^i} \frac{\partial x^\nu}{\partial \xi^j} \right) d\xi^i d\xi^j. \tag{9.4.206}
\]

Also, remember in Gauss’ theorem (eq. \(7.4.38\)), the unit normal vector dotted into the gradient \(\nabla_{\alpha'}\) is the \textit{outward} one (see equations \(7.4.29\) and \(7.4.30\)), which in our case is therefore pointing \textit{backward} in time: this is our \(-n'^\alpha\), we have inserted a negative sign in front so that \(n'^\alpha\) itself is the unit timelike vector pointing \textit{towards} the future:

\[
d^D\alpha' = d^D\xi \sqrt{|H(\tilde{\xi})|} \left( -n'^\alpha(\tilde{\xi}) \right). \tag{9.4.207}
\]

\(^{131}\)In curved spacetimes where any pair of points \(x\) and \(x'\) can be linked by a unique geodesic, this causal structure of \(G^+_{D+1}\) can be readily proved for the 4 dimensional case.
With all these clarifications in mind, we gather from eq. (9.4.205):

\[
\psi(x; x^0 > t_0) = \int_{\mathcal{D}} d^{D+1}x' \sqrt{|g(x')|} G_{D+1}(x, x') J(x')
\]

\[
+ \int_{\partial \mathcal{D}(t_0)} d^D \xi \sqrt{|H(\xi)|} \left\{ G_{D+1} \left( x, x'(\xi) \right) n^\alpha \nabla_\alpha \psi \left( x'(\xi) \right) - n^\alpha \nabla_\alpha G_{D+1} \left( x, x'(\xi) \right) \cdot \psi \left( x'(\xi) \right) \right\}.
\]

In Minkowski spacetime, we may choose \( t_0 \) to be the constant \( t \) surface of \( ds^2 = dt^2 - d\vec{x}^2 \). Then, expressed in these Cartesian coordinates,

\[
\psi(t > t_0, \vec{x}) = \int_{t' \geq t_0} dt' \int_{\mathbb{R}^D} d^D \vec{x}' G_{D+1}(t - t', \vec{x} - \vec{x}') J(t', \vec{x}')
\]

\[
+ \int_{\mathbb{R}^D} d^D \vec{x}' \{ G_{D+1}(t - t_0, \vec{x} - \vec{x}') \partial_{t_0} \psi(t_0, \vec{x}') - \partial_{t_0} G_{D+1}(t - t_0, \vec{x} - \vec{x}') \cdot \psi(t_0, \vec{x}') \}.
\]

We see in both equations (9.4.208) and (9.4.209), that the time evolution of the field \( \psi(x) \) can be solved once the retarded Green’s function \( G_{D+1} \), as well as \( \psi \)'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. (9.4.55) is in fact a special case of eq. (9.4.209): the initial time surface is the infinite past \( t_0 \rightarrow -\infty \), 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 9.30.** In 4 dimensional infinite flat spacetime, let the initial conditions for the scalar field be given by

\[
\psi(t = 0, \vec{x}) = e^{i \vec{k} \cdot \vec{x}}, \quad \partial_t \psi(t = 0, \vec{x}) = -i |\vec{k}| \psi(0, \vec{x}).
\]

Use the Kirchhoff representation in eq. (9.4.209) 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 9.31. Two Dimensions** In 1+1 dimensional flat spacetime, suppose \( \partial^2 \psi = (\partial_t^2 - \partial_z^2) \psi = 0 \) and

\[
\psi(t = 0, x) = Q(x) \quad \text{and} \quad \partial_t \psi(t = 0, x) = P(x).
\]

Explain why

\[
\psi(t > 0, x) = \frac{1}{2} Q(x + t) + \frac{1}{2} Q(x - t) + \frac{1}{2} \int_{x-t}^{x+t} P(x') dx'.
\]

Hint: Remember eq. (9.4.71). Note that, if \( t > 0 \), the \( \delta(t - |z|) \) implies \( z = t \) and \( z = -t \).
9.4.5 JWKB Approximation for Wave Equations

9.5 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 $\psi$ sourced by some externally prescribed $J(\vec{x})$.

$$S_{\text{Poisson}}[\psi] \equiv \int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \left( \frac{1}{2} \nabla_i \psi(\vec{x}) \nabla^i \psi(\vec{x}) - \psi(\vec{x}) J(\vec{x}) \right)$$

We claim that the action $S_{\text{Poisson}}$ is extremized iff $\psi$ is a solution to Poisson’s equation (eq. [9.1.1]), provided the field at the boundary $\partial \mathcal{D}$ 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,

$$\psi(\vec{x}) = \bar{\psi}(\vec{x}) + \delta \psi(\vec{x}).$$

($\delta \psi$ is one field; the $\delta$ is pre-pended as a reminder this is a deviation from $\bar{\psi}$.) A direct calculation yields

$$S_{\text{Poisson}}[\bar{\psi} + \delta \psi] = \int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \left( \frac{1}{2} \nabla_i \bar{\psi} \nabla^i \bar{\psi} - \bar{\psi} J \right)$$

$$+ \int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \left( \nabla_i \bar{\psi} \nabla^i \delta \psi - J \delta \psi \right)$$

$$+ \int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \left( \frac{1}{2} \nabla_i \delta \psi \nabla^i \delta \psi \right).$$

We may integrate-by-parts, in the second line, the gradient acting on $\delta \psi$.

$$S_{\text{Poisson}}[\bar{\psi} + \delta \psi] = \int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \left( \frac{1}{2} \nabla_i \bar{\psi} \nabla^i \bar{\psi} - \bar{\psi} J + \frac{1}{2} \nabla_i \delta \psi \nabla^i \delta \psi + \delta \psi \left\{ -\nabla^2 \bar{\psi} - J \right\} \right)$$

$$+ \int_{\partial \mathcal{D}} d^{D-1} \vec{\xi} \sqrt{|H(\vec{\xi})|} \delta \psi n^i \nabla_i \bar{\psi}$$

Provided Dirichlet boundary conditions are specified and not varied, i.e., $\psi(\partial \mathcal{D})$ is given, then by definition $\delta \psi(\partial \mathcal{D}) = 0$ and the surface term on the second line is zero. Now, suppose Poisson’s equation is satisfied by $\bar{\psi}$, then $-\nabla^2 \bar{\psi} - J = 0$ and because the remaining quadratic-in-$\delta \psi$ is strictly positive (as argued earlier) we see that any deviation increases the value of $S_{\text{Poisson}}$ 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 $S_{\text{Poisson}}$ is extremized by $\bar{\psi}$ if the first-order-in-$\delta \psi$ term

$$\int_D d^D \vec{x} \sqrt{|g(\vec{x})|} \delta \psi \left\{ -\nabla^2 \bar{\psi} - J \right\}$$

(9.5.5)
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 $\psi(t_i, \vec{x})$ and final $\psi(t_f, \vec{x})$ configurations are specified and fixed, we now discuss why the action

$$S_{\text{Wave}} \equiv \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left\{ \frac{1}{2} \nabla_\mu \psi(t', \vec{x}) \nabla^\mu \psi(t', \vec{x}) + J(t', \vec{x}) \psi(t', \vec{x}) \right\}$$

(9.5.6)

(where $x \equiv (t', \vec{x})$) is extremized iff the wave equation in eq. [9.4.15] is satisfied.

Just as we did for $S_{\text{Poisson}}$, let us consider adding to some given field $\bar{\psi}$, a deviation $\delta \psi$. That is, we will consider

$$\psi(x) = \bar{\psi}(x) + \delta \psi(x),$$

(9.5.7)

without first assuming $\bar{\psi}$ solves the wave equation. A direct calculation yields

$$S_{\text{Wave}}[\bar{\psi} + \delta \psi] = \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left\{ \frac{1}{2} \nabla_\mu \bar{\psi} \nabla^\mu \bar{\psi} + \bar{\psi} J \right\}
+ \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left( \nabla_\mu \bar{\psi} \nabla^\mu \delta \psi + J \delta \psi \right)
+ \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left( \frac{1}{2} \nabla_\mu \delta \psi \nabla^\mu \delta \psi \right).$$

(9.5.8)

We may integrate-by-parts, in the second line, the gradient acting on $\delta \psi$. By assuming that the fields fall off sufficiently quickly at spatial infinity, the remaining surface terms involve the fields at the initial and final time hypersurfaces.

$$S_{\text{Wave}}[\bar{\psi} + \delta \psi] = \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left\{ \frac{1}{2} \nabla_\mu \bar{\psi} \nabla^\mu \bar{\psi} + \bar{\psi} J + \frac{1}{2} \nabla_\mu \delta \psi \nabla^\mu \delta \psi + \delta \psi \left\{ -\nabla_\mu \nabla^\mu \bar{\psi} + J \right\} \right\}
+ \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \delta \psi(t = t_i, \vec{x}) \nabla_\mu \bar{\psi}(t = t_i, \vec{x})
- \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \delta \psi(t = t_f, \vec{x}) \nabla_\mu \bar{\psi}(t = t_f, \vec{x})
+ \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^{D-1}} d^{D-1} \xi \sqrt{|H(\xi)|} \delta \psi \nabla^\mu \bar{\psi}
= \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \left\{ \frac{1}{2} \nabla_\mu \bar{\psi} \nabla^\mu \bar{\psi} + \bar{\psi} J + \frac{1}{2} \nabla_\mu \delta \psi \nabla^\mu \delta \psi + \delta \psi \left\{ -\nabla_\mu \nabla^\mu \bar{\psi} + J \right\} \right\}
+ \left[ \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{|g(x)|} \delta \psi(t, \vec{x}) \nabla_\mu \bar{\psi}(t, \vec{x}) \right]_{t = t_i}^{t = t_f}.$$

(9.5.9)

The second and third lines of the first equality (and the second line of the second equality) come from the time derivative part of

$$\int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \sqrt{g(x)} \nabla_\mu (\delta \psi \nabla^\mu \bar{\psi}) = \int_{t_i}^{t_f} dt' \int_{\mathbb{R}^D} d^D \vec{x} \partial_\mu \left( \sqrt{g(x)} \delta \psi g^{\mu\nu} \nabla_\nu \bar{\psi} \right).$$
\begin{align*}
&= \left[ \int_{\mathbb{R}^D} d^D \mathcal{E} \sqrt{g(x)} \delta \psi g^{0\nu} \partial_\nu \bar{\psi} \right]_{t'' = t_f}^{t'' = t_i} + \ldots \quad (9.5.10)
\end{align*}

Provided the initial and final field values are specified and not varied, then \( \delta \psi(t'' = t_i,f) = 0 \) and the surface terms are zero. In eq. (9.5.9), we see that the action is extremized, i.e., when the term

\[ \int_{t_i}^{t_f} dt'' \int_{\mathbb{R}^D} d^D \mathcal{E} \sqrt{|g(x)|} \left( \delta \psi \left\{ -\nabla_\mu \nabla^\mu \bar{\psi} + J \right\} \right) \quad (9.5.11) \]

is zero for all deviations \( \delta \psi \), if the terms in the curly brackets vanish, and the wave equation eq. (9.4.15) is satisfied. Note that, unlike the case for \( S_{\text{Poisson}} \), 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 \( S_{\text{Wave}} \).

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. (9.3.17), is in fact an example of Norbert Wiener’s version of the path integral, which was the precursor of Feynman’s.

**Problem 9.32. 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 \( x^\mu \equiv (t'', \mathcal{E}) \).

\[ S[\psi] \equiv \int_{t_i}^{t_f} dt'' \int_{\mathbb{R}^D} d^D x \sqrt{|g|} \mathcal{L}(\psi, \nabla \psi) \quad (9.5.12) \]

Show that, demanding the action to be extremized leads to the Euler-Lagrange equations

\[ \frac{\partial \mathcal{L}}{\partial \psi} = \nabla_\mu \frac{\partial \mathcal{L}}{\partial \nabla_\mu \psi} . \quad (9.5.13) \]

What sort of boundary conditions are sufficient to make the variational principle well-defined? What happens when \( \mathbb{D} \) 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.

\[ \square \]

**9.6 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
\[ D_z f(z) \equiv p_2(z) \frac{d^2 f(z)}{dz^2} + p_1(z) \frac{df(z)}{dz} + p_0(z) f(z), \quad a \leq z \leq b, \]  
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
\[ D_z f_{1,2}(z) = 0. \]  
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
\[ D_z G(z, z') = \lambda(z) \delta(z - z'), \quad \text{and} \quad D_{z'} G(z, z') = \lambda(z') \delta(z - z'), \]  
where \( \delta(z - z') \) is the Dirac \( \delta \)-function and \( \lambda \) 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 \( J \),
\[ D_z f_p(z) = J(z) \quad \Rightarrow \quad f_p(z) = \int_a^b \frac{dz'}{\lambda(z')} G(z, z') J(z'). \]  
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. (9.6.3) in terms of \( f_{1,2} \)?

**Wronskian**  
The Wronskian of the two linearly independent solutions, defined to be
\[ \text{Wr}_z(f_1, f_2) \equiv f_1(z) f_2'(z) - f_1'(z) f_2(z), \quad a \leq z \leq b, \]  
will be an important object in what is to follow. We record the following facts.

- If \( \text{Wr}_z(f_1, f_2) \neq 0 \), then \( f_{1,2}(z) \) are linearly independent.
- The Wronskian itself obeys the 1st order ODE
\[ \frac{d}{dz} \text{Wr}_z(f_1, f_2) = -\frac{p_1(z)}{p_2(z)} \text{Wr}_z(f_1, f_2), \]  
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 \( f_{1,2} \),
\[ \text{Wr}_z(f_1, f_2) = W_0 \exp \left( -\int_a^z \frac{p_1(z'')}{p_2(z'')}dz'' \right), \quad W_0 = \text{constant}. \]  
- 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 \( M_1^J \),
\[ f_1(z) = M_1^J g_1(z), \quad I, J \in \{1, 2\}, \ \det M_1^J \neq 0; \]  
then
\[ \text{Wr}_z(f_1, f_2) = (\det M_1^J) \text{Wr}_z(g_1, g_2). \]

\[^{132}\text{This can be readily proven using eq. (9.6.2).}\]
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. (9.6.3). Therefore \( G(z, z') \) has to obey the homogeneous equation

\[
D_z G(z, z') = D_{z'} G(z, z') = 0, \quad z \neq z'.
\]  
(9.6.10)

For \( z > z' \), if we solve \( D_z G = 0 \) first,

\[
G(z, z') = \alpha^1(z') f_1(z),
\]

i.e., it must be a superposition of the linearly independent solutions \( \{f_1(z)\} \) (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

\[
D_{z'} G(z, z') = D_{z'} \alpha^1(z') f_1(z) = 0,
\]

(9.6.12)

(for \( z \neq z' \)), we see that the \( \{\alpha^1(z')\} \) must in turn each be a superposition of the linearly independent solutions in the variable \( z' \).

\[
\alpha^1(z') = A^1_{z'} f_1(z').
\]

(9.6.13)

(The \( \{A^1_{z'}\} \) are now constants, because \( \alpha^1(z') \) 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:

\[
G(z > z') = A^1_{z'} f_1(z) f_3(z'), \quad A^1_{z'} = \text{constant}.
\]

(9.6.14)

Similar arguments will tell us,

\[
G(z < z') = A^1_{z} f_1(z) f_3(z'), \quad A^1_{z} = \text{constant}.
\]

(9.6.15)

This may be summarized as

\[
G(z, z') = \Theta(z - z') A^1_{z'} f_1(z) f_3(z') + \Theta(z' - z) A^1_{z} f_1(z) f_3(z').
\]

(9.6.16)

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. (9.6.3). Therefore we may impose the continuity conditions

\[
A^1_{z'} f_1(z) f_3(z) = A^1_{z} f_1(z) f_3(z),
\]

(9.6.17)

\[
A^1_{z'} f_1(z)^2 + A^2_{z'} f_2(z)^2 + (A^1_{z} + A^2_{z}) f_1(z) f_2(z) = A^1_{z'} f_1(z)^2 + A^2_{z'} f_2(z)^2 + (A^1_{z} + A^2_{z}) f_1(z) f_2(z).
\]

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,

\[
A^1_{z'} = A^1_{z} \equiv A^1, \quad A^2_{z'} = A^2_{z} \equiv A^2, \quad A^1_{z} + A^2_{z} = A^1_{z'} + A^2_{z'}.
\]

(9.6.18)

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 \( \partial_z \),

\[
\int_{z' - \epsilon}^{z' + \epsilon} dz \lambda(z) \delta(z - z') = \int_{z' - \epsilon}^{z' + \epsilon} dz \{p_2 G'' + p_1 G' + p_0 G\}
\]

427
\[
\lambda(z') = \left[ p_2 G' + p_1 G \right]_{z' - \epsilon} + \int_{z' - \epsilon}^{z' + \epsilon} dz \left\{ -p_2' G' - p_1' G + p_0 G \right\} \\
= \left[ (p_1(z) - \partial_z p_2(z)) G(z, z') + p_2(z) \partial_z G(z, z') \right]_{z' = z' - \epsilon}^{z' = z' + \epsilon} \\
+ \int_{z' - \epsilon}^{z' + \epsilon} dz \left\{ p_2''(z) G(z, z') - p_1'(z) G(z, z') + p_0(z) G(z, z') \right\}.
\]

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.

\[
\lim_{\epsilon \to 0} \left\{ p_2(z' + \epsilon) \frac{\partial G(z = z' + \epsilon, z')}{\partial z} - p_2(z' - \epsilon) \frac{\partial G(z = z' - \epsilon, z')}{\partial z} \right\} = \lambda(z')
\]

We can set \( z' + \epsilon \to z' \) in the \( p_2 \) because it is smooth; the error incurred would go as \( O(\epsilon) \). 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. (9.6.3).

\[
p_2(z') \left\{ \frac{\partial G(z = z'^+, z')}{\partial z} - \frac{\partial G(z = z'^-, z')}{\partial z} \right\} = \lambda(z')
\]

This translates to

\[
p_2(z') \left( A_{21}^{12} f_1^1(z') f_3(z') - A_{21}^{12} f_1^1(z') f_3(z') \right) = \lambda(z').
\]

By taking into account eq. (9.6.18),

\[
p_2(z') \left( (A_{21}^{12} - A_{21}^{12}) f_1^1(z') f_2(z') + (A_{21}^{21} - A_{21}^{21}) f_1^1(z') f_2(z') \right) = \lambda(z'),
\]

Since \( A_{21}^{12} + A_{21}^{21} = A_{21}^{12} + A_{21}^{21} \leftrightarrow A_{21}^{12} - A_{21}^{21} = -(A_{21}^{21} - A_{21}^{21}) \),

\[
p_2(z')(A_{21}^{21} - A_{21}^{21}) \text{Wr}_{z'}(f_1, f_2) = \lambda(z'),
\]

\[
p_2(z')(A_{21}^{21} - A_{21}^{21}) W_0 \exp \left(-\int_b^{z'} \frac{p_1(z''')}{p_2(z''')} dz''' \right) = \lambda(z'),
\]

where eq. (9.6.7) was employed in the second line. We see that, given a differential operator \( D \) of the form in eq. (9.6.1), this amounts to solving for the measure \( \lambda(z') \): it is fixed, up to an overall multiplicative constant \( (A_{21}^{21} - A_{21}^{21}) W_0 \), by the \( p_{1,2} \). (Remember the Wronskian itself is fixed up to an overall constant by \( p_{1,2} \); cf. eq. (9.6.7).) Furthermore, note that \( A_{21}^{21} - A_{21}^{21} \) can be absorbed into the functions \( f_{1,2} \), since the latter’s normalization has remained arbitrary till now. Thus, we may choose \( A_{21}^{21} - A_{21}^{21} = 1 = -(A_{21}^{21} - A_{21}^{21}) \). At this point,

\[
G(z, z') = A^1 f_1(z) f_1(z') + A^2 f_2(z) f_2(z') \\
+ \Theta(z - z')((A_{21}^{12} - 1) f_1(z) f_2(z') + A_{21}^{21} f_2(z) f_1(z')) \\
+ \Theta(z' - z)((A_{21}^{12} - 1) f_1(z) f_2(z') + (A_{21}^{21} - 1) f_2(z) f_1(z')).
\]

Because we are seeking a symmetric Green’s function, let us also consider

\[
G(z', z) = A^1 f_1(z') f_1(z) + A^2 f_2(z') f_2(z)
\]

428
We also reiterate, up to the overall multiplicative constant $W$ is fixed once the differential operator $D$ together with the third line of eq. (9.6.25) versus second line of eq. (9.6.26), says the terms 

tically symmetric; whereas the second line of eq. (9.6.25) versus the third line of eq. (9.6.26), says the terms involving $A_{z}^{1,2}$ are symmetric iff $A_{z}^{1,2} = A_{z}^{21} \equiv \chi$. We gather, therefore,

\[
G(z, z') = A^{1}f_{1}(z)f_{1}(z') + A^{2}f_{2}(z)f_{2}(z') + G(z, z'; \chi),
\]

\[
G(z, z'; \chi) \equiv (\chi - 1) \left\{ \Theta(z - z')f_{1}(z)f_{2}(z') + \Theta(z' - z)f_{1}(z')f_{2}(z) \right\} + \chi \left\{ \Theta(z - z')f_{2}(z)f_{1}(z') + \Theta(z' - z)f_{2}(z')f_{1}(z) \right\}.
\]

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 contribute 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

\[
D_{z}f_{1,2}(z) \equiv p_{2}(z) \frac{d^{2}f_{1,2}(z)}{dz^{2}} + p_{1}(z) \frac{df_{1,2}(z)}{dz} + p_{0}(z)f_{1,2}(z) = 0, \quad a \leq z \leq b,
\]

we may solve the symmetric Green’s function equation(s)

\[
D_{z}G(z, z') = p_{2}(z)W_{0} \exp \left( - \int_{b}^{z} \frac{p_{1}(z'')}{p_{2}(z'')} dz'' \right) \delta(z - z'),
\]

\[
D_{z'}G(z, z') = p_{2}(z')W_{0} \exp \left( - \int_{b}^{z'} \frac{p_{1}(z'')}{p_{2}(z'')} dz'' \right) \delta(z - z'),
\]

\[
G(z, z') = G(z', z),
\]

by using the general ansatz

\[
G(z, z') = G(z', z) = A^{1}f_{1}(z)f_{1}(z') + A^{2}f_{2}(z)f_{2}(z') + G(z, z'; \chi),
\]

\[
G(z, z'; \chi) \equiv (\chi - 1)f_{1}(z_{>}f_{2}(z_{<}) + \chi \cdot f_{1}(z_{<})f_{2}(z_{>}),
\]

\[
z_{>} \equiv \max(z, z'), \quad z_{<} \equiv \min(z, z').
\]

Here $W_{0}$, $A^{1,2}$, and $\chi$ 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,

\[
\text{Wr}_{z}(f_{1}, f_{2})(z) = f_{1}(z)f_{2}'(z) - f_{1}'(z)f_{2}(z) = \left. \left( \frac{\partial G(z = z'^{+}, z')}{\partial z} - \frac{\partial G(z = z'^{-}, z')}{\partial z} \right) \right|_{z' \rightarrow z} = W_{0} \exp \left( - \int_{b}^{z} \frac{p_{1}(z'')}{p_{2}(z'')} dz'' \right).
\]

We also reiterate, up to the overall multiplicative constant $W_{0}$, the right hand side of eq. (9.6.30) is fixed once the differential operator $D$ (in eq. (9.6.29)) is specified; in particular, one may not always be able to set the right hand side of eq. (9.6.30) to $\delta(z - z')$.  

429
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

\[ x^i = r(s_\theta c_\phi, s_\theta s_\phi, c_\theta) \],
\[ x'^i = r'(s_\theta c_{\phi'}, s_\theta s_{\phi'}, c_{\theta'}) \];

and try to solve

\[ -\nabla^2 x G(\vec{x} - \vec{x}') = -\nabla^2 x' G(\vec{x} - \vec{x}') = \delta(r - r') \delta(c_\theta - c_{\theta'}) \delta(\phi - \phi'). \] (9.6.39)

Because of the rotation symmetry of the problem – we know, in fact,

\[ G(\vec{x} - \vec{x}') = \frac{1}{4\pi |\vec{x} - \vec{x}'|} = (4\pi)^{-1} (r^2 + r'^2 - 2rr' \cos \gamma)^{-1/2} \] (9.6.40)

depends on the angular coordinates through the dot product \( \cos \gamma \equiv \vec{x} \cdot \vec{x}' / (rr') = \hat{x} \cdot \hat{x}' \). This allows us to postulate the ansatz

\[ G(\vec{x} - \vec{x}') = \sum_{\ell=0}^{\infty} \frac{\tilde{g}_{\ell}(r, r')}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y^m_{\ell}(\theta, \phi) Y^m_{\ell}(\theta', \phi'). \] (9.6.41)

By applying the Laplacian in spherical coordinates (cf. eq. (9.2.96)) and using the completeness relation for spherical harmonics in eq. (9.2.73), eq. (9.6.39) becomes

\[ \sum_{\ell=0}^{\infty} \frac{\tilde{g}_{\ell}(r, r')}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y^m_{\ell}(\theta, \phi) Y^m_{\ell}(\theta', \phi') \]

\[ = -\frac{\delta(r - r')}{rr'} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y^m_{\ell}(\theta, \phi) Y^m_{\ell}(\theta', \phi'), \] (9.6.42)

with each prime representing \( \partial_r \). Equating the \((\ell, m)\) term on each side,

\[ D_r \tilde{g}_{\ell} \equiv \tilde{g}'_{\ell} + \frac{2}{r} \tilde{g}_{\ell} - \frac{\ell(\ell + 1)}{r^2} \tilde{g}_{\ell} = -(2\ell + 1) \frac{\delta(r - r')}{rr'}. \] (9.6.43)

We already have the \( \delta \)-function measure – it is \(- (2\ell + 1)/r^2\) – but it is instructive to check its consistency with the right hand side of (9.6.30); here, \( p_1(r) = 2/r \) and \( p_2(r) = 1 \), and

\[ W_0 \exp \left(-2 \int^r dr''/r''\right) = W_0 e^{-2\ln r} = W_0 r^{-2}. \] (9.6.44)

Now, the two linearly independent solutions to \( D_r f_{1,2}(r) = 0 \) are

\[ f_1(r) = \frac{F_1}{r^{\ell+1}}, \quad f_2(r) = F_2 r^\ell, \quad F_{1,2} = \text{constant}. \] (9.6.45)
The radial Green’s function must, according to eq. (9.6.33), take the form

$$\tilde{g}_\ell(r, r') = \frac{A_1^\ell}{(rr')^{\ell+1}} + A_2^\ell (rr')^\ell + G_\ell(r, r'),$$  \hspace{1cm} (9.6.46)

where $A_1^\ell$, $F$, and $\lambda_\ell$ are constants. (What happened to $F_{1,2}$? Strictly speaking $F_1 F_2$ should multiply $A_1^\ell$ but since the latter is arbitrary their product(s) may be assimilated into one constant(s); similarly, in $G_\ell(r, r')$, $F = F_1 F_2$ but since $F_{1,2}$ occurs as a product, we may as well call it a single constant.) To fix $F$, we employ eq. (9.6.36).

$$-\frac{2\ell + 1}{r^2} = F \nabla_r (r^{1-\ell}, r^\ell) = \frac{\partial G(r = r^{\ell+})}{\partial r} - \frac{\partial G(r = r^{\ell-})}{\partial r}. \hspace{1cm} (9.6.49)$$

Carrying out the derivatives explicitly,

$$-\frac{2\ell + 1}{r^2} = F \left\{ \frac{\partial}{\partial r} \left( \frac{1}{r} \left( \frac{r}{r'} \right)^{\ell} \right)_{r=r^{\ell+}} - \frac{\partial}{\partial r} \left( \frac{1}{r} \left( \frac{r'}{r} \right)^{\ell} \right)_{r=r^{\ell-}} \right\}
= F \left\{ \frac{\ell}{r^{\ell+1}} + \frac{(\ell + 1) r^\ell}{r^{\ell+2}} \right\} = F \frac{2\ell + 1}{r^2}. \hspace{1cm} (9.6.50)$$

Thus, $F = -1$. We may take the limit $r \to 0$ or $r' \to 0$ and see that the terms involving $A_1^\ell$ and $(\lambda_\ell/r^\ell)(r^{\ell+}/r^{\ell-})$ in eq. (9.6.46) will blow up for any $\ell$; while $1/(4\pi|x - \vec{x}'|)$ will $1/(4\pi r')$ or $\to 1/(4\pi r)$ does not. This implies $A_1^\ell = 0$ and $\lambda_\ell = 0$. Next, by considering the limits $r \to \infty$ or $r' \to \infty$, we see that the $A_2^\ell$ term will blow up for $\ell > 0$, whereas, in fact, $1/(4\pi|x - \vec{x}'|) \to 0$. Hence $A_2^{\ell > 0} = 0$. 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 (9.6.46) now stands as

$$\tilde{g}_\ell(r, r') = \frac{1}{r^\ell} \left( \frac{r}{r'} \right)^{\ell}, \hspace{1cm} (9.6.51)$$

which in turn means eq. (9.6.41) is

$$G(\vec{x} - \vec{x}') = \frac{1}{4\pi|\vec{x} - \vec{x}'|} = \frac{1}{r^\ell} \sum_{\ell=0}^{\infty} \frac{1}{2\ell + 1} \left( \frac{r}{r'} \right)^{\ell} \sum_{m=-\ell}^{\ell} Y_{\ell}^m(\theta, \phi) Y_{\ell}^{m'}(\theta', \phi'). \hspace{1cm} (9.6.52)$$

If we use the addition formula in eq. (9.2.76), we then recover eq. (9.3.49).

**Problem 9.33.** 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. (2.0.39). Hint: Start by justifying the ansatz

$$G_2(\vec{x} - \vec{x}') = \sum_{\ell=-\infty}^{+\infty} \tilde{g}_\ell(r, r') e^{i\ell(\phi - \phi')}, \hspace{1cm} (9.6.53)$$
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 $\tilde{g}_e$ is a symmetric one. You should be able to appeal to the solution in eq. (7.4.54) to argue there are no homogeneous contributions to this 2D Green’s function; i.e., the $A_1 = A_2 = 0$ in eq. (9.6.33) are zero in this case.
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! I do have solutions to some of the problems. If you are using these notes for self-study, write to me and I will e-mail them to you.

B  Group Theory

What is a group? A group is a collection of objects \{a, b, c, \ldots\} 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. 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 \(\{X_i | i = 1, \ldots, N\}\) is a collection of such invertible linear operators that are closed under multiplication, namely

\[ X_i X_j = c_{ijk} X_k \]  

(B.0.1)

for any \(i, j, k \in \{1, 2, \ldots, N\}\); 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 \(\{g_1, g_2, \ldots\}\) to a set of invertible linear operators \(\{D(g_1), D(g_2), \ldots\}\) 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

\[ D(g_1)D(g_2) = D(g_1 g_2). \]  

(B.0.2)

The identity maps to the identity

\[ D(e) = \mathbb{I}. \]  

(B.0.3)

because \(D(e)D(g) = D(e \cdot g) = D(g) = \mathbb{I} \cdot D(g)\) for all \(g\). Also,

\[ D(g^{-1}) = D(g)^{-1} \]  

(B.0.4)
because \( D(g^{-1})D(g) = D(g^{-1}g) = I = D(g)^{-1}D(g) \).

**Example** A simple example is that of the rotation group in 2D. We may take as definition of the group the 1-parameter matrix:

\[
R(\phi) = \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}.
\]  

(B.0.5)

The group multiplication rule is, via a direct computation,

\[
R(\phi)R(\phi') = R(\phi + \phi').
\]  

(B.0.6)

On the other hand, we may also map these rotations to the 2D complex plane, via

\[
D(R(\phi)) \equiv e^{i\phi}.
\]  

(B.0.7)

Witness the preservation of the group multiplication rule \( D(R(\phi))D(R(\phi')) = D(R(\phi + \phi')) \) because the left hand side is

\[
D(R(\phi))D(R(\phi')) = e^{i\phi}e^{i\phi'}
\]  

(B.0.8)

whereas the right hand side is simply

\[
D(R(\phi + \phi')) = e^{i(\phi + \phi')}.
\]  

(B.0.9)

### 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:

\[
\xi^i p_i \equiv \sum_i \xi^i p_i.
\]  

(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. \( \xi^i \) is the \( i \)th component of \( (\xi^1, \xi^2, \xi^3, \ldots) \), not some variable \( \xi \) raised to the \( i \)th 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

$$\eta_{\mu\nu} = \text{diag}[1, -1, \ldots, -1],$$  \hfill (C.0.2)

where “$\text{diag}[a_1, \ldots, a_N]$” refers to the diagonal matrix, whose diagonal elements (from the top left to the bottom right) are respectively $a_1, a_2, \ldots, a_N$. 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

$$\partial^2 = \Box = \eta^{\mu\nu}\partial_\mu\partial_\nu.$$  \hfill (C.0.3)

The Laplacian in flat space, in Cartesian coordinates, read instead

$$\vec{\nabla}^2 = \delta_{ij}\partial_i\partial_j,$$  \hfill (C.0.4)

where $\delta_{ij}$ is the Kronecker delta, the unit $D \times D$ matrix $\mathbb{I}$:

$$\delta_{ij} = 1, \quad i = j$$

$$= 0, \quad i \neq j.$$  \hfill (C.0.5)

**Index (anti-)symmetrization**  The symbols $[\ldots]$ and $\{\ldots\}$ denote anti-symmetrization and symmetrization respectively. In particular,

$$T_{[i_1\ldots i_N]} = \sum_{\text{even permutations } \Pi} T_{\Pi[i_1\ldots i_N]} - \sum_{\text{odd permutations } \Pi} T_{\Pi[i_1\ldots i_N]},$$  \hfill (C.0.6)

$$T_{\{i_1\ldots i_N\}} = \sum_{\text{all permutations } \Pi} T_{\Pi[i_1\ldots i_N]}.$$  \hfill (C.0.7)

For example,

$$T_{[ijk]} = T_{ijk} - T_{ikj} - T_{jik} + T_{kji} + T_{kij}$$  \hfill (C.0.8)

$$T_{\{ijk\}} = T_{ijk} + T_{ikj} + T_{jik} + T_{kji} + T_{kij}.$$  \hfill (C.0.9)

**Caution**  Beware that many relativity texts define their (anti-)symmetrization with a division by a factorial; namely,

$$T_{[i_1\ldots i_N]} = \frac{1}{N!} \left( \sum_{\text{even permutations } \Pi} T_{\Pi[i_1\ldots i_N]} - \sum_{\text{odd permutations } \Pi} T_{\Pi[i_1\ldots i_N]} \right),$$  \hfill (C.0.10)

$$T_{\{i_1\ldots i_N\}} = \frac{1}{N!} \sum_{\text{all permutations } \Pi} T_{\Pi[i_1\ldots i_N]}.$$  \hfill (C.0.11)

I prefer not to do so, because of the additional baggage incurred by these numerical factors when performing concrete computations.
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 \( G_N \) 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

\[
[\text{Length}] = [\text{Time}] .
\]

In particular, since in SI units \( c = 299,792,458 \) meters/second, we have

\[
1 \text{ second} = 299,792,458 \text{ meters}, \quad (c = 1). \tag{D.0.1}
\]

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,

\[
1 \text{ second} \approx 1/(1.0545718 \times 10^{-34} \text{ Joules}) \quad (\hbar = 1). \tag{D.0.2}
\]

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 \( \text{fm} = \text{femtometer} = 10^{-15} \text{ meters} \). Hence,

\[
10^{-15} \text{ meters} \approx 1/(197.326,98 \text{ MeV}), \quad (\hbar c = 1). \tag{D.0.4}
\]

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 \( 1/(\exp(E/(k_B T)) + 1) \) – where \( E, k_B \) and \( T \) are respectively the energy of the particle(s), \( k_B \) is the Boltzmann constant, and \( T \) is 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.

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 $G_N M/r$ through the virial theorem (kinetic energy $\sim$ potential energy) $v \sim \sqrt{G_N M/r}$.

Newton’s gravitational constant takes the form

$$G_N \approx 6.7086 \times 10^{-39} \hbar c (\text{GeV}/c^2)^{-2}. \quad (D.0.5)$$

Just from this dimensional analysis alone, when $\hbar = c = 1$, one may form a mass-energy scale (‘Planck mass’)

$$M_{\text{pl}} \equiv \frac{1}{\sqrt{32\pi G_N}}. \quad (D.0.6)$$

(The $32\pi$ is for technical convenience.) This suggests – since $M_{\text{pl}}$ appears to involve relativity ($c$), quantum mechanics ($\hbar$) and gravitation ($G_N$) – that the energy scale required to probe quantum aspects of gravity is roughly $M_{\text{pl}}$. Therefore, it may be useful to set $M_{\text{pl}} = 1$ 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.:


**Problem D.1.** Let $\hbar = c = 1$.

- If angular momentum is 3.34, 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?
- Work out the Planck mass in eq. (D.0.6) in seconds, meters, and GeV.
Problem D.2. 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 \(V(r) \sim e^{-mr}/r\) 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].

Problem D.3. Consider the following wave operator for a particle of mass \(m > 0\),

\[ \mathcal{W} \equiv \partial_\mu \partial^\mu + m^2, \quad x^\mu \equiv (t, \vec{x}). \]

\[ (D.0.7) \]

• In \(\mathcal{W}\), put back the \(\hbar\)s only.

• In \(\mathcal{W}\), put back the \(c\)s only.

Assume that \(\mathcal{W}\) has dimensions of \(1/[\text{Length}^2]\).
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.: Wei Hao Chen, Afif Ismail, Jake Leistico, Kuan-Nan Lin, Alec Lovlein, Chern Xi Pang, Hadi Papei, Jason Payne, Evan Severson, Leon Tsai, Chun-Yu Wei (Viona), Cheng Han Wu, and Afidah Zuroida.

I thank Tolga Birkandan for pointing me to Sages Math (linked below; [38]); and Leo Stein for clarifying how xAct [39] 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: May 27, 2020

References


[3] Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables (AMS55)


[5] Philip M. Morse, Herman Feshbach, “Methods of Theoretical Physics”


Matthias Blau, “Plane Waves and Penrose Limits,” Available here


[100 Years Of Relativity : space-time structure: Einstein and beyond, 257 (2005)].


doi:10.1103/PhysRevLett.13.789

[34] I. I. Shapiro, G. H. Pettengill, M. E. Ash, M. L. Stone, W. B. Smith, R. P. Ingalls and


arXiv:1611.00018 [gr-qc].

[37] Numerical Receipes (Older versions available online.)

[38] Sages Math (“. . . free open-source mathematics software system. . .”)