PY501 - Mathematical Physics

Hongwan Liu

Table of Contents

1 Calculus of Variations

References: Stone & Goldbart (SG) Chapter 1; Byron & Fuller (BF) Chapter 2; Arfken, Weber & Harris (AWH) Chapter 22.

Finding where the minimum or maximum value of some quantity occurs is an extremely common task. For example, you might want to know where the highest point on a map is, or when you had the highest heartrate throughout the day. Mathematically, we have a function $f(x)$, and we want to find the value of x which maximizes or minimizes $f(x)$. To do so, for a differentiable function f, we simply take the derivative and set it to zero

$$f'(x) = 0 \tag{1.1}$$

and solve for x to find the stationary points of the function.

Often though, we run into problems where we want to find the *function* at which the minimum or maximum value of some function occurs. Some examples include:

1. What is the shortest path to take between points A and B?
2. What closed curve of fixed length encloses the maximum possible area?
3. What form does a hanging heavy chain of fixed length take, so as to minimize its potential energy?

To answer these questions mathematically, we need an object called a **functional** $J[y]$, which maps smooth[1] functions y (e.g. a path, a curve) to a real number (e.g. a distance, an area). This is just another map, like a function is. But now, we want to develop the tools required to define a **functional derivative** such that setting

$$\frac{\delta J}{\delta y(x)} = 0 \tag{1.2}$$

will allow us to find a function $y(x)$ (e.g. a path, a curve) that maximizes $J[y]$ (e.g. a distance, an area).

[1] "Smooth" means that all derivatives of the function exists. We won't ever be interested in subtleties involving continuity and differentiability in this course.

1.1 Functionals

What does a functional look like? For our purposes, we will be dealing with functionals that have the following form:

$$J[y] = \int_{x_1}^{x_2} dx\, f(x, y, y', y'', \cdots, y^{(n)})\,, \tag{1.3}$$

where f is a function of the real numbers x, y, $y' \cdots$, *independently*.[2] We call these kinds of functionals **local** in x. As you can see, J takes in a function y, performs an integral, and returns a real number, which is exactly what a functional should do.

[2] This is a cause of endless confusion, so pay attention! From the perspective of f, y and y' are *independent variables*.

1.1.1 The functional derivative

Let us work out the functional derivative for the case where

$$J[y] = \int_{x_1}^{x_2} dx\, f(x, y, y')\,. \tag{1.4}$$

To do this, suppose we make an infinitesimal shift $y(x) \to y(x) + \varepsilon\eta(x)$, where ε is an infinitesimally small constant,[3] and $\eta(x)$ is some arbitrary function. Then

[3] i.e. with ε^2 and higher powers of ε all being zero.

the shift in J is

$$\delta J = J[y+\varepsilon\eta] - J[y] = \int_{x_1}^{x_2} dx\ \left[f(x, y+\varepsilon\eta, y'+\varepsilon\eta') - f(x,y,y')\right]\ . \quad (1.5)$$

Since ε is an infinitesimal quantity, we can perform a Taylor expansion up to first order about y and y' to find

$$\delta J = \int_{x_1}^{x_2} dx\ \left[\varepsilon\eta\frac{\partial f}{\partial y} + \varepsilon\eta'\frac{\partial f}{\partial y'}\right]\ . \quad (1.6)$$

To make further progress, we integrate the second term by parts, giving

$$\delta J = \left[\varepsilon\eta\frac{\partial f}{\partial y'}\right]_{x_1}^{x_2} + \int_{x_1}^{x_2} dx\ \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]\varepsilon\eta\,. \quad (1.7)$$

We are frequently—but not always!—concerned with finding functions of y with fixed endpoints;[4] in that case, $\eta(x_1) = \eta(x_2) = 0$, and the boundary terms in the first term on the right vanishes, leaving

$$\delta J = \int_{x_1}^{x_2} dx\ \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]\varepsilon\eta\,. \quad (1.8)$$

This can be written suggestively as

$$\delta J = \int_{x_1}^{x_2} dx\,\delta y(x)\left(\frac{\delta J}{\delta y(x)}\right)\ , \quad (1.9)$$

where $\delta y(x) \equiv \varepsilon\eta(x)$, and

$$\frac{\delta J}{\delta y(x)} \equiv \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \quad (1.10)$$

is the **functional derivative** of J with respect to $y(x)$.

To aid our understanding, it can be helpful to think discretely. We can discretize x between x_1 and x_2 into N discrete steps, so that the function y takes up values $y_i = y(x_i)$, where $i = 1, 2\cdots, N$. A choice of the function y corresponds in this discrete picture to a choice of $\{y_i\}$, which is a single point in an N-dimensional space. Fig. 1 has a visualization of this. At every point in this N-dimensional space, we can assign a value to J. The small variation $\varepsilon\eta$ can likewise be discretized, so that $\delta y_i = \varepsilon\eta_i$, which can be thought of as a step in a particular direction in the same N-dimensional space. In this discrete picture,

$$\delta J = \sum_{i=1}^{N}\frac{\partial J}{\partial y_i}\delta y_i\,, \quad (1.11)$$

just as one might expect for a function J defined in the N-dimensional space indexed by y_i. In the continuous limit, we need to trade the summation over discrete i to an integral over the continuous label x, leading to Eq. (1.9).[5]

1.1.2 The Euler-Lagrange equation

Now, to find the **stationary points**—maxima, minima or saddle points—of J, we want to set $\delta J = 0$ for any arbitrary variation $\varepsilon\eta$, just like for a function g

[4] For example, if we are interested in finding the path with the short distance between two fixed points.

Figure 1: A discretized visualization of varying over functions. (**HL**: To be completed, but not difficult to imagine!)

[5] In this picture, a choice of the function y is a point in an uncountably infinite dimensional space, and $\epsilon\eta$ is a step in some arbitrary direction, and J is function that returns a real number at every point in this space.

on $\mathbb{R}^n$, we want δg to be zero for a step in any direction at a stationary point. Referring to Eq. (1.8), we require

$$\int_{x_1}^{x_2} dx \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right] \varepsilon\eta = 0\,. \tag{1.12}$$

Since this applies for *any* $\eta(x)$, the term in the square brackets $[\cdots]$ must vanish.[6] we must have

$$\boxed{\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0} \tag{1.13}$$

for a stationary point for J. This is the famous **Euler-Lagrange equation**.

Through derivations similar to what we saw above, we can get generalized Euler-Lagrange equations for more complicated versions of J. If J depends on more than one function y_i, for example, the stationary points are given by

$$\frac{\partial f}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial f}{\partial y_i'}\right) = 0\,, \tag{1.14}$$

which is one equation for each variable y_i. If on the other hand, f depends on higher derivatives y'', y''' and so on, then the generalized Euler-Lagrange equation we get is

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) - \frac{d^3}{dx^3}\left(\frac{\partial f}{\partial y'''}\right) + \cdots = 0\,. \tag{1.15}$$

[6] We can prove that this is true rigorously, given various conditions on η and f. This is often known as the **fundamental lemma of the calculus of variations**. For further discussion, see SG 1.2.2 and Wikipedia.

1.1.3 Applications

Time to apply what we've learnt! We'll apply the Euler-Lagrange equations to two examples.

1.1.3.1 Soap film supported by a pair of coaxial rings. Consider Fig. 2, where a pair of co-axial rings support a soap film. The energy associated with the configuration is directly proportional to the area, and hence the soap film tends to minimize this energy by minimizing its area, subject to the constraint that the soap film has to end on the rings at either end. The area of associated with a segment of the film of width dx is

$$dA = 2\pi y(x)\sqrt{dx^2 + dy^2} = 2\pi y(x)\sqrt{1 + y'^2}\,dx\,, \tag{1.16}$$

and so the functional that we want to minimize is

$$J[y] = \int_{x_1}^{x_2} dx\, f(y, y')\,, \quad f(y, y') \equiv y\sqrt{1 + y'^2}\,, \tag{1.17}$$

with the endpoint values fixed at $y(x_1)$ and $y(x_2)$. The minimum for this functional can therefore be found by applying the Euler-Lagrange equations. The partial derivatives that we need for the Euler-Lagrange equation are

$$\frac{\partial f}{\partial y} = \sqrt{1 + y'^2}\,, \quad \frac{\partial f}{\partial y'} = \frac{yy'}{\sqrt{1 + y'^2}}\,, \tag{1.18}$$

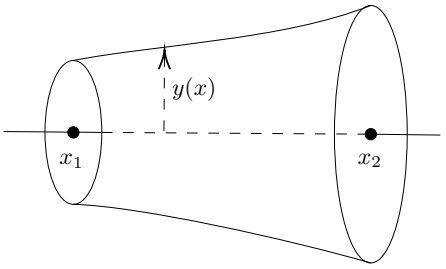

Figure 2: Soap film between two rings, centered at x_1 and x_2, with radii $y(x_1)$ and $y(x_2)$.

and so the Euler-Lagrange equation says that the minimal surface area profile $y(x)$ must satisfy

$$\sqrt{1+y'^2} - \frac{d}{dx}\left(\frac{yy'}{\sqrt{1+y'^2}}\right) = 0$$
$$\implies \sqrt{1+y'^2} - \frac{y'^2}{\sqrt{1+y'^2}} - \frac{yy''}{\sqrt{1+y'^2}} + \frac{yy'^2y''}{(1+y'^2)^{3/2}} = 0$$
$$\implies \frac{1}{\sqrt{1+y'^2}} - \frac{yy''}{(1+y'^2)^{3/2}} = 0\,. \qquad (1.19)$$

This differential equation looks difficult to solve, but fortunately there's a neat little trick to do so. Multiplying by y' on both sides gives

$$0 = \frac{y'}{\sqrt{1+y'^2}} - \frac{yy'y''}{(1+y'^2)^{3/2}} = \frac{d}{dx}\left(\frac{y}{\sqrt{1+y'^2}}\right)\,. \qquad (1.20)$$

We'll return to how we knew this trick would work later on. In the mean time, the solution is

$$\frac{y}{\sqrt{1+y'^2}} = \kappa \qquad (1.21)$$

for some constant κ. Rewriting this as

$$\frac{dy}{dx} = \sqrt{\frac{y^2}{\kappa^2} - 1} \implies \int \frac{dy}{\sqrt{y^2/\kappa^2 - 1}} = \int dx\,, \qquad (1.22)$$

we can integrate this first-order ordinary differential equation by substituting $y = \kappa\cosh t$ and $dy = \kappa\sinh t$ to find

$$\kappa\int dt = \int dx \implies \kappa t = x + C \implies y = \kappa\cosh\left(\frac{x+C}{\kappa}\right) \qquad (1.23)$$

for some constants κ and C. These can be determined by enforcing the two boundary conditions—the radii of the two rings, $y(x_1)$ and $y(x_2)$.

1.1.3.2 The brachistochrone The next problem we will consider is a famous one, posed by Johann Bernoulli in 1696. What shape should a wire with endpoints $(0,0)$ and (a,b) take, in order that a frictionless bead will slide from rest down the wire in the shortest possible time?

First, the total time T taken down a given path can be written as

$$T = \int_0^T dt = \int_0^L \frac{ds}{v}\,, \qquad (1.24)$$

where v is the speed of the bead, and s is the distance along the path, with a total length L. However, we can once again write $ds^2 = dx^2 + dy^2$ so that $ds = \sqrt{1+y'^2}\,dx$, and apply conservation of energy to find $v = \sqrt{2gy}$. Thus, we can define a functional T

$$T[y] = \int_0^a dx\,\sqrt{\frac{1+y'^2}{2gy}} \qquad (1.25)$$

that we want to minimize with respect to y, again with fixed end points. We can therefore apply the Euler-Lagrange equation, which gives after some algebra

$$yy'' + \frac{1}{2}(1+y'^2) = 0\,. \qquad (1.26)$$

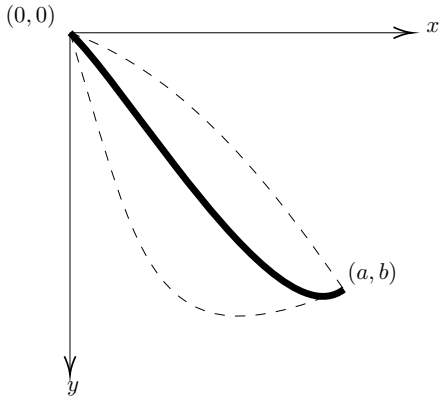

Figure 3: Possible shapes of a wire for a frictionless bead to travel from the origin $(0,0)$ to a point (a,b).

Once again, we can use the trick of multiplying by y' to find that

$$y'\left(yy'' + \frac{1}{2}(1+y'^2)\right) = \frac{1}{2}\frac{d}{dx}\left(y(1+y'^2)\right) = 0\,, \tag{1.27}$$

or

$$y(1+y'^2) = 2C \tag{1.28}$$

for some constant C. From this point, one can check that the following parametrization $(x(t), y(t))$ is indeed a solution to the differential equation above:

$$\begin{aligned} x &= C(\theta - \sin\theta) \\ y &= C(1-\cos\theta)\,, \end{aligned} \tag{1.29}$$

although it is surprisingly hard to pin down the details regarding this solution.[7] This parametric curve is known as the **cycloid**, which is the curve traced out by a fixed point on the rim of a wheel that is rolling without slipping along a flat surface.

[7] For example, is there a unique solution, and does the solution always occur with $\theta \in [0, 2\pi)$ for every point (a, b)? See for example Ref. [1] for more details.

1.1.4 First integral

In both applications discussed in Sec. 1.1.3, we were able to rephrase the differential equation to be solved as $dI/dx = 0$, implying that I is some constant associated with the problem. This quirk, which somewhat resembles the conservation of quantities like energy, is something we will revisit in much greater depth later: it is far deeper than just a mathematical coincidence. For now, let's just take a quick look at where it comes from. In both cases, the function inside the integral was $f(y, y')$, with no explicit dependence on x, implying that

$$\frac{df}{dx} = \cancelto{0}{\frac{\partial f}{\partial x}} + y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'}\,. \tag{1.30}$$

We define the **first integral** of the Euler-Lagrange equation as

$$I \equiv f - y'\frac{\partial f}{\partial y'}\,, \tag{1.31}$$

from which we can check that

$$\begin{aligned} \frac{dI}{dx} &= y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'} - y''\frac{\partial f}{\partial y'} - y'\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) \\ &= y'\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right]\,. \end{aligned} \tag{1.32}$$

Thus, $dI/dx = 0$ if the Euler-Lagrange equation is satisfied.

You can show that if f depends on more than one variable, so that we have a functional of the form

$$J[y_1, y_2, \cdots, y_n] = \int dx\, f(y_1, y_2, \cdots, y_n; y_1', y_2', \cdots, y_n')\,, \tag{1.33}$$

the first integral is of the form

$$I = f - \sum_i y_i'\frac{\partial f}{\partial y_i'}\,. \tag{1.34}$$

Note that there is only one first integral, even when there are multiple dependent variables y_i.

1.2 Lagrangian Mechanics

It turns out that classical mechanics can be reformulated as a problem of finding the stationary function of some functional. Given some system, we first define the **Lagrangian** function $L = T - V$, where T and V are the kinetic and potential energy functions of the system. We can make any choice of coordinates we would like to describe T and V; let's say we choose some set of **generalized coordinates** q with components q^i and time derivatives $\dot{q}^i$.[8]

The equations of motion governing the system between times t_i and t_f can then be obtained by finding the stationary function $q(t)$ of the **action functional**,

$$S[q] = \int_{t_i}^{t_f} dt\, L(t; q^i, \dot{q}^i)\,. \tag{1.35}$$

This is known as the **principle of least action**. It is no exaggeration to say that a lot of theoretical physics basically involves finding the appropriate action that describes the system of interest, once the principle of least action is applied.

(**End of Lecture**: Wednesday Sep 4 2024)

1.2.1 The central force problem

We'll now turn our attention to an important problem in mechanics—a particle of mass m moving in a potential $V(r)$ which depends only on the distance r from the origin, with the radial component of the force being $F_r = -\partial_r V$. We'll compute this two ways: the first using Newtonian mechanics, and the second using Lagrangian mechanics. The coordinate system we'll use is shown in Fig. 4.

In Newtonian mechanics, we would write down $-\partial_r V = m a_r$, and $-\partial_\theta V/r = 0 = m a_\theta$, where a_r and a_θ are the radial and tangential accelerations respectively. But to solve this equation, we need to find the acceleration in polar coordinates. To do so, we'll use a neat trick: take the particle to be traveling in the complex plane, with coordinate given by the complex number $z = re^{i\theta}$. Then

$$\begin{aligned} \dot{z} &= \dot{r}e^{i\theta} + i\dot{\theta}re^{i\theta} \\ \ddot{z} &= (\ddot{r} - r\dot{\theta}^2)e^{i\theta} + (2\dot{r}\dot{\theta} + r\ddot{\theta})(ie^{i\theta})\,. \end{aligned} \tag{1.36}$$

On the complex plane, the first term is a complex number that is represented by a vector that is parallel to $re^{i\theta}$, while the second term is represented by a vector that is perpendicular to $re^{i\theta}$. We can therefore conclude that the acceleration in polar coordinates is

$$a_r = \ddot{r} - r\dot{\theta}^2\,, \qquad a_\theta = 2\dot{r}\dot{\theta} + r\ddot{\theta}\,. \tag{1.37}$$

Newton's laws therefore read

$$m(\ddot{r} - r\dot{\theta}^2) = -\partial_r V \tag{1.38}$$

$$m(2\dot{r}\dot{\theta} + r\ddot{\theta}) = 0 \implies \frac{d}{dt}(mr^2\dot{\theta}) = 0\,. \tag{1.39}$$

You should recognize the second equation as expressing the conservation of angular momentum, with $l = mr^2\dot{\theta}$. Subtituting this expression into Eq. (1.38), we find

$$m\ddot{r} - \frac{l^2}{mr^3} = -\frac{\partial V}{\partial r}\,, \tag{1.40}$$

[8] This can be the usual x, y and z in 3D space, r, θ and ϕ in 3D spherical coordinates, or something even more abstract than these choices, it doesn't matter.

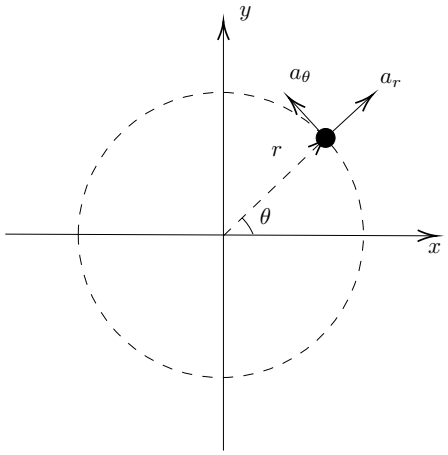

Figure 4: Coordinate system for the central force problem.

the equation of motion governing all central force problems.

Now that was a bit of hike, and again, a lot of issue was that we had to deal with vectors. Let's see how the Lagrangian approach works out. First, we write down the Lagrangian for the system,

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r)\,. \tag{1.41}$$

This depends on two coordinates, but is independent of t, and therefore we can write down two Euler-Lagrange equations—one for each coordinate. These are:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) - \frac{\partial L}{\partial r} = 0 \implies m\ddot{r} - mr\dot{\theta}^2 + \frac{\partial V}{\partial r} = 0 \tag{1.42}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0 \implies \frac{d}{dt}(mr^2\dot{\theta}) = 0\,. \tag{1.43}$$

Notice that we obtain the conservation of angular momentum immediately from the second equation involving θ. Substituting once again $l = mr^2\dot{\theta}$, we get the same expression as before in Eq. (1.40).[9]

Let's also compute the first integral, defined for this problem as

$$I = L - \dot{r}\frac{\partial L}{\partial \dot{r}} - \dot{\theta}\frac{\partial L}{\partial \dot{\theta}} = -\frac{1}{2}m\dot{r}^2 - \frac{1}{2}mr^2\dot{\theta}^2 - V(r) = -(T+V)\,, \tag{1.44}$$

which is simply (the negative of) the total energy. Thus, the fact that the first integral is constant is just a statement about conservation of energy in this problem.

[9] You may be tempted to plug l directly into the Lagrangian, which looks like it would remove one coordinate from the Lagrangian, simplifying matters significantly. *But this is pure folly*, and gets you the wrong answer. The coordinates and their derivatives are—each and every one of them—*independent variables of the Lagrangian*. The Lagrangian's only aim in life is to take in these coordinates, their derivatives, and the time coordinate—not knowing that they're related in any way—and spit out a number. The Lagrangian *knows absolutely nothing about trajectories involving these coordinates as a function of time*.

1.3 Variable Endpoints

So far, we've been dealing with situations where we want the endpoints to be fixed. Now let's see what happens when we relax that assumption. Consider the problem of constructing a railway between two ports, located across a strip of land with straight, parallel sides, illustrated in Fig. 5. Suppose that the cost of construction is proportional to the length of the track, but the cost of sea transport is negligible, and so that the ports can be wherever we want. We therefore want to minimize the total length given once again by

$$L[y] = \int_{x_1}^{x_2} dx\, \sqrt{1 + y'^2}\,. \tag{1.45}$$

This time however, we want to allow variations at the endpoints too. Considering a small perturbation, we see that

$$\delta L = \int_{x_1}^{x_2} dx\, \frac{y'}{\sqrt{1+y'^2}}\delta y' \tag{1.46}$$

By now you should be familiar with what to do next: integrate by parts! Don't forget, however, the boundary terms. This gives

$$\delta L = \frac{y'(x_2)}{\sqrt{1+y'(x_2)^2}}\delta y(x_2) - \frac{y'(x_1)}{\sqrt{1+y'(x_1)^2}}\delta y(x_1) - \int_{x_1}^{x_2} dx \left[\frac{d}{dx}\left(\frac{y'}{\sqrt{1+y'^2}}\right)\right]\delta y(x)\,. \tag{1.47}$$

The extremum for L is achieved when $\delta L = 0$. But since we have complete freedom to choose $y(x)$, including $y(x_2)$ and $y(x_1)$, the coefficients to each of

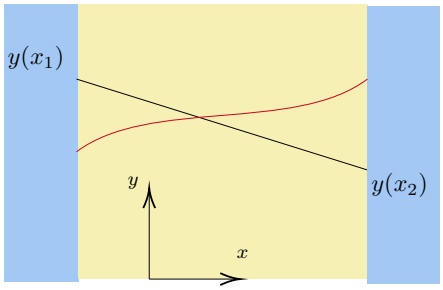

Figure 5: Railway across a strip of land with straight, parallel sides.

$\delta y(x_2)$, $\delta y(x_1)$ and δy inside the integral must be zero. Let's first consider the coefficient for $\delta y(x)$. This just says that

$$\frac{y'}{\sqrt{1+y'^2}} = C \implies y' = \text{constant}\,, \tag{1.48}$$

and so we'd better be building a straight track, as expected. In addition however, the other two terms now enforce

$$\frac{y'(x_2)}{\sqrt{1+y'(x_2)^2}} = \frac{y'(x_1)}{\sqrt{1+y'(x_1)^2}} = 0 \implies y'(x_1) = y'(x_2) = 0\,. \tag{1.49}$$

We therefore want to build the railway exactly perpendicular to the sea. All very reasonable!

What you might have noticed is that by allowing the endpoints to float, you ended up getting a boundary term that enforces some boundary conditions at the two endpoints. Boundary conditions obtained through variation are known as **natural boundary conditions**. We'll come back to more examples with natural boundary conditions later on.

1.4 Noether's Theorem

We now come to one of the most profound results in mathematical physics. We've seen in a few of the previous examples above that when we apply the principle of least action, we can sometimes obtain equations of the form $d(\cdots)/dt = 0$ when the Euler-Lagrange equation is satisfied. For example, for a Lagrangian of the form $L = L(q, \dot{q})$ with no explicit dependence on t, we know from Sec. 1.1.4 that the time derivative of the first integral is conserved:

$$\frac{d}{dt}\left(\dot{q}\frac{\partial L}{\partial \dot{q}} - L\right) = 0\,. \tag{1.50}$$

We also saw that in the central force problem in Sec. 1.2.1, we can show that the angular momentum is conserved, i.e.

$$\frac{d}{dt}\left(mr^2\dot{\theta}\right) = 0\,, \tag{1.51}$$

and in that case, the Lagrangian had not explicit dependence on θ. Equations of this form are known as **conservation laws**, and you have, up to this point in your physics career, already encountered many of them. But you may now be seeing a pattern emerge: whenever there is no explicit dependence of L on some quantity, you get conservation laws. And indeed this is true! These are all manifestations of **Noether's theorem**,[10] which says that

> Any continuous symmetry of the action corresponds to a conservation law.

[10] This is actually Noether's *first* theorem; the second theorem is much more obscure, and we won't discuss it in this course.

Now, your first thought might be, it's a little strange that we have a theorem that's just written out in words. That's really because you can prove many different mathematical statements that are described by those words, with various levels of generality and formality. Because Noether's theorem is so important, I want to walk you through a proof of it at a relatively general level. Your second thought might be that there are a lot of terms in there that I haven't carefully defined. But I think in this case, it's best just to define things as we go along. The discussion that follows below is mostly based on Ref. [2].

So let's start with the classical action, written as

$$S[q^i] = \int_{t_i}^{t_f} dt\, L(t; q^i, \dot{q}^i)\,, \tag{1.52}$$

where I remind you that q is some generalized coordinate with components $q^i(t)$ and time derivatives $\dot{q}^i(t)$. We say that the action is **invariant up to a boundary term** under a transformation $q^i(t) \to q^i(t) + \varepsilon\eta^i(t)$, where ε is taken to be a small, time-independent quantity, if

$$\delta S \equiv S[q^i + \varepsilon\eta^i] - S[q^i] = \varepsilon \int_{t_i}^{t_f} dt\, \frac{dK}{dt}\,, \tag{1.53}$$

for all $q^i(t)$, and for some **boundary term** K.

If this is the first time you're seeing this, you might be wondering why even bother with K at all—after all, that's what the word *invariant* should mean. But in fact, a lot of extremely interesting models (in both particle physics and condensed matter physics) have actions which are invariant up to boundary terms, and we would like to apply Noether's theorem to them. An equivalent way to say the same thing is that the action is invariant up to a boundary term if the Lagrangian transforms by a **total derivative** dK/dt.

Any transformation which leaves the action invariant up to a boundary term, or transforms the Lagrangian by a total derivative, is called a **symmetry**.[11] The first question you should ask when you see a Lagrangian is always, "what are its symmetries?" This is the organizing principle behind all of the models that we study in physics. So let's check out some examples. Consider once again the central force problem,

$$S = \int_{t_i}^{t_f} dt\, \left[\frac{1}{2}m\dot{\vec{r}}^2 - V(r)\right]\,, \tag{1.54}$$

The action is invariant under rotations, i.e. perform the transformation $\vec{r} \to \mathsf{R}\vec{r}$, where R is constant, orthogonal matrix with $\mathsf{R}^\intercal = \mathsf{R}^{-1}$, even for very large rotations, since $\dot{\vec{r}}^2 \to (\mathsf{R}\dot{\vec{r}})^\intercal \mathsf{R}\dot{\vec{r}} = \dot{\vec{r}}^\intercal \mathsf{R}^\intercal \mathsf{R}\dot{\vec{r}} = \dot{\vec{r}}^2$, in addition to the fact that rotations preserve length, i.e. $|\mathsf{R}\vec{r}| = r$.

Let's consider another transformation to the action

$$\vec{r}(t) \to \vec{r}(t) - \varepsilon\dot{\vec{r}}(t)\,, \tag{1.55}$$

again with ε being small. Under this transformation, we find[12]

$$\begin{aligned} S[r^i] \to S[r^i - \varepsilon\dot{r}^i] &= \int dt\, \left[\frac{1}{2}m(\dot{\vec{r}} - \varepsilon\ddot{\vec{r}})^2 - V(\vec{r} - \varepsilon\dot{\vec{r}})\right] \\ &= \int dt\, \left[\frac{1}{2}m\dot{\vec{r}}^2 - \varepsilon m\dot{\vec{r}}\ddot{\vec{r}} - V(\vec{r}) + \varepsilon\dot{\vec{r}}^i\partial_i V\right] \\ &= S[r^i] - \varepsilon\int dt\, \frac{d}{dt}\left[\frac{1}{2}m\dot{\vec{r}}^2 - V(r)\right]\,. \end{aligned} \tag{1.56}$$

We can therefore conclude that the action is invariant under this transformation up to a boundary term, $K = L$. What is this strange transformation? Well, it really isn't that mysterious if you just think of it as relabeling time by the coordinate τ, which is just a constant shift from t, i.e. $\tau = t + \varepsilon$. With this transformation, $\vec{r}(t) = \vec{r}(\tau) - \varepsilon\dot{\vec{r}}(\tau)$, and it is easy to see that you get the same transformation as in Eq. (1.56).

Let's now consider an *arbitrary* transformation, one that isn't necessarily a symmetry. It also doesn't necessarily have to leave the endpoints fixed. Denoting this transformation as $q^i \to q^i + \varepsilon\zeta^i(t)$ and apply this to the classical action

[11] Sometimes you will hear people say that a transformation is only a symmetry if $K = 0$, and should be called a *quasisymmetry* otherwise. We won't make this distinction.

[12] For notation convenience, I'll drop the limits of the integral whenever nothing much ever happens to it.

Eq. (1.52). This gives (using index notation)

$$\begin{aligned}\delta S &= \int_{t_i}^{t_f} dt\, \frac{\partial L}{\partial q^i}\varepsilon\zeta^i(t) + \frac{\partial L}{\partial \dot{q}^i}\varepsilon\dot{\zeta}^i(t) \\ &= \int_{t_i}^{t_f} dt\, \frac{\partial L}{\partial q^i} + \frac{\partial L}{\partial \dot{q}^i}\varepsilon\zeta^i\Big|_{t_i}^{t_f} - \int_{t_i}^{t_f} dt\, \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\varepsilon\zeta^i \\ &= \int_{t_i}^{t_f} dt\, \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\right]\varepsilon\zeta^i + \int_{t_i}^{t_f} dt\, \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\varepsilon\zeta^i\right), \end{aligned} \tag{1.57}$$

where we have kept the boundary term after performing integration by parts.

Now, let's choose the arbitrary transformation $\zeta^i(t)$ to be one that leaves the action invariant. Then we have the following relation, after dividing throughout by ε,

$$\int_{t_i}^{t_f} dt\, \frac{d}{dt}\left(K - \frac{\partial L}{\partial \dot{q}^i}\zeta^i\right) = \int_{t_i}^{t_f} dt\, \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\right]\zeta^i\,. \tag{1.58}$$

In classical mechanics, we're often interested in the behavior of a system at arbitrary times, and indeed, it doesn't really matter what the precise values of t_i and t_f are. Because of that, we need the integrand themselves to be equal, i.e.

$$\frac{d}{dt}\left(K - \frac{\partial L}{\partial \dot{q}^i}\zeta^i\right) = \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\right]\zeta^i\,. \tag{1.59}$$

This relation is known as the **Noether identity**, and is true for any ζ^i that is a symmetry. Note that at this point, we *have not even mentioned the Euler-Lagrange equation*. This relation is true for *arbitrary* $q^i(t)$, whether or not it corresponds to a the physical solution. When an equation is true before having enforced the Euler-Lagrange equation, we call in an **off-shell** relation, a term that you're bound to run into again.

Now let's take the relation **on-shell**, so that we are only considering $q^i(t)$ that satisfies the Euler-Lagrange equation. For such $q^i(t)$, the right-hand side of the previous equation vanishes, and we find

$$\frac{dQ}{dt} \overset{\text{E-L}}{=} 0\,, \qquad Q = K - \frac{\partial L}{\partial \dot{q}^i}\zeta^i\,. \tag{1.60}$$

This is our big result, **Noether's theorem**: given a symmetry of the action ζ^i, we can define a quantity called a **Noether charge** (or just "charge") that is conserved on-shell, i.e. for physical trajectories.

That was a lot of abstract discussion; let's now take a look at some examples.

(**End of Lecture**: Monday Sep 9 2024)

1.4.1 The central force problem revisited

Let's return once again to the to the central force problem, with

$$S[\vec{r}] = \int_{t_i}^{t_f} dt\, \left[\frac{1}{2}m\dot{\vec{r}}^2 - V(r)\right]. \tag{1.61}$$

We saw earlier that the action was invariant under rotations. You can convince yourself (see Fig. 6 that an infinitesimal rotation about some arbitrary axis with

unit vector $\hat{\alpha}$ can be written as $\vec{r} \to \vec{r} + \varepsilon\hat{\alpha} \times \vec{r}$. We also checked that the rotation leaves the action invariant with $K = 0$. Putting this altogether, we find that the Noether charge associated with rotational symmetry is (using index notation)

$$Q = -\frac{\partial L}{\partial \dot{r}^i}(\hat{\alpha} \times \vec{r})^i = -(m\dot{\vec{r}}_i)(\hat{\alpha} \times \vec{r})^i = \hat{\alpha} \cdot (\vec{r} \times \vec{p}) \,, \tag{1.62}$$

where $\vec{p} \equiv m\dot{\vec{r}}$ is the momentum. You should recognize immediately that $\vec{L} = \vec{r} \times \vec{p}$ is the **angular momentum**, the Noether charge corresponding to rotational symmetry, and since $\hat{\alpha}$ is arbitrary, $\vec{L}$ is conserved.

We also considered another transformation $\vec{r} \to \vec{r} - \varepsilon\dot{\vec{r}}$, which left the action invariant up to a boundary term, $K = L$, the Lagrangian. We also saw that this transformation was equivalent to time translation, $t \to t + \varepsilon$. Once again, the Noether charge associated with this symmetry is

$$Q = L - \frac{\partial L}{\partial \dot{r}^i}\dot{\vec{r}}^i = \frac{1}{2}m\dot{\vec{r}}^2 - V(r) - m\dot{\vec{r}}^2 = -\left[\frac{1}{2}m\dot{\vec{r}}^2 + V(r)\right] \,, \tag{1.63}$$

which is simply the statement that the Noether charge is the **total energy**, and is conserved. More generally, for any action that has time-translation as a symmetry, e.g. $L = L(q^i, \dot{q}^i)$, the total energy is

$$E = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L \,, \tag{1.64}$$

and is conserved. You should recognize that E is nothing but the first integral that we described in Eq. (1.31).

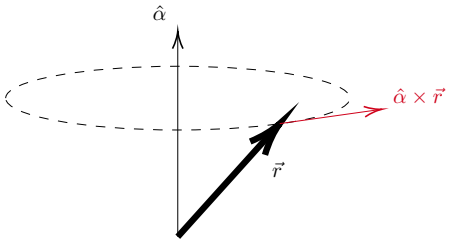

Figure 6: Infinitesimal rotations of $\vec{r}$ about axis $\hat{\alpha}$ are in the direction $\hat{\alpha} \times \vec{r}$.

1.4.2 Common symmetries

Before we move on from Noether's theorem, let's examine some common symmetries that you've probably already encountered in classical mechanics, and how they usually arise from Noether's theorem.

1.4.2.1 Spatial translation invariance implies momentum conservation.

If a Lagrangian is invariant under a small, constant, spatial translation ε of some coordinate q, i.e. $L(t; q + \varepsilon, \dot{q}) = L(t; q, \dot{q})$.[13] Noether's theorem (Eq. (1.60)) tells us that the associated charge is

$$Q = p \equiv \frac{\partial L}{\partial \dot{q}} \text{ and } \frac{dp}{dt} = 0 \,, \tag{1.65}$$

where p is the momentum conjugate to q. One simple scenario where momentum conservation holds is when the Lagrangian simply doesn't depend explicitly on q at all; in this case, q is known as a **cyclic coordinate**. This is what happens in systems with cylindrical symmetry, for example, and therefore have Lagrangians that do not depend on the azimuthal angle, leading to an associated conserved angular momentum.

[13] Since ε is constant, $\dot{q}$ remains unchanged.

1.4.2.2 Time translation invariance implies energy conservation.

Suppose we performed the translation $q(t) \to q(t + \varepsilon)$, where ε is once again small and constant. You can imagine this as taking a trajectory that previously started

at time t, and now starting it at $t-\varepsilon$ instead. Then the action transforms as

$$\begin{aligned} S[q] &\to \int_{t_i}^{t_f} dt\, L(t; q(t+\varepsilon), \dot{q}(t+\varepsilon)) \\ &= \int_{t_i+\varepsilon}^{t_f+\varepsilon} d\tau\, L(\tau-\varepsilon; q(\tau), \dot{q}(\tau)) \\ &= S[q] + \varepsilon\left[L(t_f, q(t_f), \dot{q}(t_f)) - L(t_i, q(t_i), \dot{q}(t_i))\right] - \varepsilon \int_{t_i}^{t_f} d\tau\, \frac{\partial L}{\partial \tau}\,. \end{aligned} \tag{1.66}$$

Here, the second term is a boundary term, but the third term might not be: we can therefore see that time translation is a symmetry of the action if the Lagrangian does not depend explicitly on t, with the conserved charge being energy, i.e.

$$\frac{\partial L}{\partial t} = 0 \implies \text{Energy is conserved.} \tag{1.67}$$

We have derived the expression for energy several times already, for example in Eq. (1.64): we do this by noting that $q(t+\varepsilon) = q(t) + \varepsilon\dot{q}(t)$, and then applying Noether's theorem.

1.5 Continuous Systems

So far, we have been looking at Lagrangians that have only a single particle. For multiple particles, we can similarly write down Lagrangians involving $\{q_i, \dot{q}_i\}$ for each particle. Frequently, however, we're interested in *fields*, like the electromagnetic field, and we want to be able to write down *field theories* by similarly writing down an action, extremizing it, and finding the equations of motion. Suppose we have a field with value $\varphi(x)$, where $x \equiv x^\mu$, $\mu = 0, 1, \cdots, d$ is the coordinate of a $(d+1)$-dimensional space (or spacetime). We can write down actions that look like

$$S[\varphi] = \int dt\, L = \int d^{d+1}x\, \mathcal{L}(x, \varphi, \partial_\mu \varphi)\,, \tag{1.68}$$

where $\mathcal{L}$ is the **Lagrangian density**,

$$L \equiv \int d^d x\, \mathcal{L}\,, \tag{1.69}$$

and

$$\partial_\mu \varphi \equiv \frac{\partial \varphi}{\partial x^\mu}\,. \tag{1.70}$$

We're going to follow a similar path to what we discussed in Sec. 1.1.1. We'll take a variation over the action, $\varphi(x) \to \varphi(x) + \varepsilon\eta(x)$, and so correspondingly $\partial_\mu\varphi(x) \to \partial_\mu\varphi(x) + \varepsilon\partial_\mu\eta(x)$, with no variation on the d-dimensional boundary of our $(d+1)$-dimensional space. We get

$$\delta S = \int d^{d+1}x \left[\frac{\partial \mathcal{L}}{\partial \varphi}\varepsilon\eta + \frac{\partial \mathcal{L}}{\partial(\partial_\mu\varphi)}\varepsilon\partial_\mu\eta\right]\,. \tag{1.71}$$

At this point, we want to integrate by parts, which we can do using the **divergence theorem**,

$$\int_\Omega d^{n+1}x\,\partial_\mu f^\mu = \int_{\partial\Omega} dS\, n_\mu f^\mu\,, \tag{1.72}$$

where $\partial\Omega$ is the boundary of Ω, dS is an element of area on the boundary, and n_μ is the outward-pointing normal vector. Much more on this later on in the course, but for now, we see that

$$\begin{aligned}\int_\Omega d^{n+1}x\,\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\epsilon\eta\right) &= \int d^{n+1}x\,\left(\partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\varepsilon\eta + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\varepsilon\partial_\mu\eta\right)\\ &= \int_{\partial\Omega} dS\, n^\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\right)\epsilon\eta\\ &= 0\,,\end{aligned} \tag{1.73}$$

since η does not vary on the boundary, allowing us to write

$$\delta S = \int d^{d+1}x\,\left[\frac{\partial\mathcal{L}}{\partial\varphi} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\right]\varepsilon\eta\,. \tag{1.74}$$

Effectively, if we are only allowing for variations that are zero on the boundary, we can simply integrate by parts by switch the position of ∂_μ. Extremizing the action therefore means that

$$\frac{\delta S}{\delta\varphi} \equiv \frac{\partial\mathcal{L}}{\partial\varphi} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)} = 0\,, \tag{1.75}$$

which is the Euler-Lagrange equation in continuous form.

1.5.1 The vibrating string

Our first example of a continuous system is a string, shown in Fig. 7. The string has fixed ends, a mass per unit length of ρ, and is under tension T. If we assume only small displacements from equilibrium, the Lagrangian is

$$L = \int_0^L dx\,\left(\frac{1}{2}\rho\dot{y}^2 - \frac{1}{2}Ty'^2\right)\,, \tag{1.76}$$

where the dot denotes a partial derivative with respect to t, and the prime a partial derivative with respect to x.

The variation of the action is

$$\begin{aligned}\delta S &= \int dt\int_0^L dx\,\left(\rho\dot{y}\delta\dot{y} - Ty'\delta y'\right)\\ &= \int dt\int_0^L dx\,\left(-\rho\ddot{y} + Ty''\right)\delta y\,,\end{aligned} \tag{1.77}$$

where to reach the second line, we perform integration by parts, and because the endpoints are fixed, $\delta y = 0$ when $x = 0$ and $x = L$. Extremizing the action $\delta S = 0$ gives the equation of motion

$$\rho\ddot{y} - Ty'' = 0\,, \tag{1.78}$$

which is the wave equation with transverse waves propagating with speed $c = \sqrt{T/\rho}$.

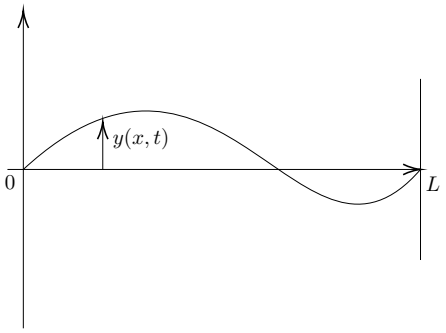

Figure 7: Set up for the vibrating string.

(**End of Lecture**: Wednesday Sep 11 2024)

1.5.2 The canonical energy-momentum tensor

Just as in the case of a single-particle Lagrangian, with conservation of momentum and energy, if the Lagrangian density $\mathcal{L} \equiv \mathcal{L}(\varphi, \partial_\mu \varphi)$ does not depend explicitly on x^μ, then we should expect a conservation law as well.

We'll go through the derivation of the Noether charge here again, as the result is presented in often confusing ways in many textbooks. Suppose our action is of the form

$$S = \int d^{d+1}x \, \mathcal{L}(\varphi, \partial_\mu \varphi) \,. \tag{1.79}$$

Then, we know that under the transformation $x^\mu \to x^\mu + \varepsilon^\mu$, where ε^μ is constant, we can expand $\mathcal{L}$ to give

$$\delta S = \int d^{d+1}x \, \varepsilon^\mu \partial_\mu \mathcal{L} \,, \tag{1.80}$$

where

$$\partial_\mu \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \varphi} \partial_\mu \varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_\mu (\partial_\nu \varphi) \,. \tag{1.81}$$

Note that in this transformation is a symmetry of the action, since $\varepsilon^\mu \partial_\mu \mathcal{L}$ is a total derivative. However, if we promote $\varepsilon^\mu \to \varepsilon^\mu(x^\mu)$ from a *global* to a local transformation, this stops being true.

Now, let's consider the *arbitrary* transformation, $x^\mu \to x^\mu + \zeta^\mu(x^\mu)$ (not necessarily constant, and not necessarily with no variation on the boundary!). This gives

$$\varphi \to \varphi + \zeta^\mu \partial_\mu \varphi \,, \quad \partial_\mu \varphi \to \partial_\mu \varphi + \partial_\mu (\zeta^\nu \partial_\nu \varphi) \,. \tag{1.82}$$

$$\begin{aligned} \delta S &= \int d^{d+1}x \left[\frac{\partial \mathcal{L}}{\partial \varphi} \zeta^\mu \partial_\mu \varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_\nu (\zeta^\mu \partial_\mu \varphi) \right] \\ &= \int d^{d+1}x \left[\frac{\partial \mathcal{L}}{\partial \varphi} - \partial_\nu \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \right] \zeta^\mu \partial_\mu \varphi + \int d^{d+1}x \, \partial_\nu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \zeta^\mu \partial_\mu \varphi \right) . \end{aligned} \tag{1.83}$$

Now, setting $\zeta^\mu = \varepsilon^\mu$, and going on-shell, the first term goes to zero as the Euler-Lagrange equations are satisfied, giving

$$\begin{aligned} \int d^{d+1}x \, \varepsilon^\mu \partial_\mu \mathcal{L} &= \int d^{d+1}x \, \varepsilon^\mu \partial_\nu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_\mu \varphi \right) \\ &\implies \int d^{d+1}x \, \varepsilon^\mu \partial_\nu \left(\mathcal{L} \delta^\nu_\mu - \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_\mu \varphi \right) = 0 \,. \end{aligned} \tag{1.84}$$

We call the Noether charge here the **canonical energy-momentum tensor**,[14]

$$T^\nu{}_\mu \equiv \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_\mu \varphi - \delta^\nu_\mu \mathcal{L} \,, \tag{1.85}$$

with the statement

$$\partial_\nu T^\nu{}_\mu = 0 \tag{1.86}$$

denoting *local* energy-momentum conservation. This can be extended to the case where there are multiple fields, in which case the first term on the right-hand side of Eq. (1.84) picks up a sum over all fields.

[14] The reason its called "canonical" and not just "the" energy-momentum tensor is because it's often the case that this isn't quite what you want to work with. Instead, you often want to work with $T^{\mu\nu}$ shifted by some other object whose total derivative is zero. In electromagnetism, this procedure of "fixing" $T^{\mu\nu}$ is often called the *Belinfante improvement procedure*.

1.5.3 The vibrating string revisited

We now return to the vibrating string, with Lagrangian given in Eq. (1.76). We can now compute the canonical energy-momentum tensor for this system. For $\mu = 0$,

$$T^{\nu}{}_{0} = \frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_0 \varphi - \delta^\nu_0 \mathcal{L} \,, \tag{1.87}$$

and so

$$T^{0}{}_{0} = \rho \dot{y}^2 - \left(\frac{1}{2}\rho \dot{y}^2 - \frac{1}{2} T y'^2 \right) = \frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \,, \tag{1.88}$$

$$T^{1}{}_{0} = -T y' \dot{y} \,. \tag{1.89}$$

In addition, the conservation laws read

$$\begin{aligned} 0 = \partial_\nu T^{\nu}{}_{0} &= \partial_\nu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_0 \varphi \right] - \partial_\nu \delta^\nu_0 \mathcal{L} \\ &= \partial_t (\rho \dot{y}^2 - \mathcal{L}) + \partial_x (-T y' \dot{y}) \\ &= \partial_t \left(\frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \right) - \partial_x (T y' \dot{y}) \,, \end{aligned} \tag{1.90}$$

$$\begin{aligned} 0 = \partial_\nu T^{\nu}{}_{1} &= \partial_\nu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\nu \varphi)} \partial_1 \varphi \right] - \partial_\nu \delta^\nu_1 \mathcal{L} \\ &= \partial_t (\rho \dot{y} y') + \partial_x (-T y'^2 - \mathcal{L}) \\ &= \partial_t (\rho \dot{y} y') - \partial_x \left(\frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \right) \,. \end{aligned} \tag{1.91}$$

These are local conservation laws of the same form as the *continuity equation*,

$$\frac{\partial q}{\partial t} + \nabla \cdot \vec{J} = 0 \,, \tag{1.92}$$

where q should be thought of as a local density (e.g. charge density of fluid density) with a globally conserved quantity $Q = \int d^d x \, q$, and $\vec{J}$ is a flux. For $\partial_\nu T^{\nu}{}_{0} = 0$, the local density is

$$T^{0}{}_{0} = \frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \,, \tag{1.93}$$

which is the energy density. The energy flux is $T^{1}{}_{0} = -T \dot{y} y'$, which is the rate that a segment of string is doing work on its neighbor to the right. Let's check that the total energy is globally conserved:

$$\begin{aligned} \frac{d}{dt} \int_0^L dx \left(\frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \right) &= \int_0^L dx \, \frac{\partial}{\partial t} \left(\frac{1}{2}\rho \dot{y}^2 + \frac{1}{2} T y'^2 \right) \\ &= \int_0^L dx \, \frac{\partial}{\partial x} (T y' \dot{y}) \\ &= T y' \dot{y} \big|_0^L = 0 \,, \end{aligned} \tag{1.94}$$

since the string is clamped at the ends, as expected.

That's really the last we'll say about continuous systems, but of course, there is so much more to say. You'll learn more about these systems in classes like quantum field theory and statistical physics.

1.6 Constraints

When we want to find the extremum of some function or functional, we often want to impose some constraints on that extremum. Here are some classic examples of problems that involve finding extrema under some constraint:

1. Given a curve a fixed length on a plane, what is the maximum area that it can enclose? What is the shape of the curve in that case?

2. Consider a chain with fixed length suspended between two points, both a fixed height above the ground. What is the shape of the chain that minimizes the potential energy?

The strategy that we are going to adopt here is identical to what you may have seen in vector calculus, with the use of **Lagrange multipliers**. Suppose you have a functional $J[y]$ that you want to extremize, subject to some constraint given by $K[y] = 0$. Consider the modified functional $\tilde{J}[y, \lambda] = J[y] - \lambda K[y]$, for some new parameter λ. First, we note that if vary the new parameter λ, we find

$$\tilde{J}[y, \lambda + \delta\lambda] - \tilde{J}[y, \lambda] = -\delta\lambda \cdot K[y]\,, \tag{1.95}$$

and so when we extremize $\tilde{J}$, including over possible variations in λ, we require $K[y] = 0$, the precise constraint that we wanted to impose!

Under a variation of y, on the other hand, we obtain

$$\tilde{J}[y + \delta y, \lambda] - \tilde{J}[y, \lambda] = \delta J[y, \delta y] - \lambda \cdot \delta K[y, \delta y]\,, \tag{1.96}$$

where $\delta J[y, \delta y] = \delta J[y + \delta y] - \delta J[y]$, and likewise for δK. At the extremum $\delta\tilde{J} = 0$ for some $y = y^*$ then, we must necessarily have

$$\delta J[y^*, \delta y] = \lambda \cdot \delta K[y^*, \delta y] \tag{1.97}$$

for all variations δy away from y^*. What's going on here? First of all, $K[y^*] = 0$, and so y^* corresponds to a point where the constraint is satisfied. Now, if the variation δy preserves this constraint, then $\delta K[y^*, y + \delta y] = 0$; in that case, the previous equation says that $\delta J[y^*, y + \delta y] = 0$ as well. In other words, J *itself is also extremized, as long as we limit ourselves to* δy *that preserves the constraint*.

Fig. 8 shows an illustration of how this method works. The equation of motion for λ enforces the constraint, and fixes the trajectory to lie along the blue line, with $K[y] = 0$. We can see that if we choose variations along the blue trajectory so that the constraint is always satisfied, $\delta K = 0$, but not necessarily δJ, *except* at the extremum of J along the trajectory (marked in red). At this point, in the direction $\delta K = 0$, $\delta J = 0$ as well, since it is a local extremum. Our clever choice of $\tilde{J}$ gave us *1)* an equation of motion for λ that constrained us to the blue trajectory, and *2)* an equation of motion for y that is satisfied for points where J is extremized along the trajectory.

Another common situation where Lagrange multipliers can be used is in classical mechanics, when you want to extremize an action given by $S = \int dt\, L(t; q, \dot{q})$, subject to some constraint on the *coordinate* q, e.g. $g(t; q, \dot{q}) = 0$. We can again consider the modified action $\tilde{S}$, where

$$\tilde{S}[q, \lambda] = \int dt\, [L - \lambda(t) g(t; q, \dot{q})]\,, \tag{1.98}$$

where once again, extremizing $\tilde{S}$ over all q and λ simultaneously enforces the desired constraint, and extremizes L when subject to the constraint.[15]

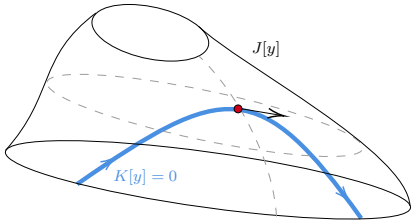

Figure 8: An illustration of the Lagrange multiplier method.

[15] One has to be careful when applying this to Newtonian mechanics, where the constraints should be applied *to the Euler-Lagrange equation themselves* in order to get the right dynamics. Equivalently, the extremization of $\tilde{S}$ cannot be over arbitrary q, but must also be over only those δq where the constraints are satisfied. Using our procedure for constraints of the form $g(t; q)$, with no dependence on $\dot{q}$, works; these are called *holonomic constraints*. However, what we have here does not work for general constraints. For a much more detailed discussion of holonomic and non-holonomic constraints, and when the Lagrange multiplier method fails, see Ref. [3] for an excellent discussion of this topic.

1.6.1 Maximum entropy distribution

Our first example of how to use this comes from statistical mechanics. Let Γ denote the classical phase space of a mechanical system of N particles governed by a Hamiltonian $H(p_n^i, q_n^i)$, with $n = 1, \cdots, N$, $i = 1, 2, 3$ and $d\Gamma \equiv \prod_{n=1}^{N} (d^3\vec{p}_n \, d^3\vec{q}_n)$. We define the phase space density $\rho(p_n^i, q_n^i)$ such that $\rho(p_n^i, q_n^i) d\Gamma$ is the probability of the system being in some state given by p_n^i and q_n^i in the small region $d\Gamma$.

The entropy related to this probability distribution can be defined as the functional

$$S[\rho] = -\int_\Gamma d\Gamma \, \rho \log \rho \, . \tag{1.99}$$

We now want to find ρ that maximizes the entropy for a given mean energy,

$$\langle E \rangle = \int_\Gamma d\Gamma \, \rho H \, . \tag{1.100}$$

However, ρ is subject to the constraint that it is a probability density function, and so we must enforce the constraint

$$\int_\Gamma d\Gamma \, \rho = 1 \, . \tag{1.101}$$

We can solve this problem by defining a new functional $\tilde{S}[\rho, \alpha, \beta]$, where

$$\tilde{S}[\rho, \alpha, \beta] = S[\rho] + \alpha \left(\int_\Gamma d\Gamma \, \rho - 1 \right) + \beta \left(\int_\Gamma d\Gamma \, \rho H - \langle E \rangle \right) . \tag{1.102}$$

The equations of motion for the real-number[16] Lagrange multipliers α and β enforce the two constraints that we mentioned above. Now for ρ, we have

$$\rho \log \rho \to (\rho + \delta\rho) \log(\rho + \delta\rho) = \rho \log \rho + \delta\rho \log \rho + \rho \cdot \frac{\delta\rho}{\rho} \tag{1.103}$$

and so

$$\delta\tilde{S} = \int_\Gamma d\Gamma \, (1 + \log \rho + \alpha + \beta H) \, \delta\rho \, , \tag{1.104}$$

and setting the term in parentheses to zero to extremize $\delta\tilde{S}$ gives

$$\rho(p_n^i, q_n^i) = e^{-1-\alpha-\beta H(p_n^i, q_n^i)} \, . \tag{1.105}$$

At this point, α and β can be determined from the normalization and energy constraints detailed above, with the exact answer being determined by the expect form of H. This probability density is the usual *canonical distribution*. Note that you can see from this procedure that the temperature $T \equiv 1/\beta$ appears as a Lagrange multiplier in this procedure.

[16] They're not functions! In particular, they can live *outside* the integral over Γ. There's nothing wrong with having a functional depend on real numbers. You can also use Lagrange multipliers that are functions, if for example the constraint you are imposing is on ρ itself, and not the integral of ρ, as we are in this case.

(**End of Lecture**: Monday Sep 16 2024)

1.6.2 The catenary

We are now ready to tackle another classic calculus of variations problem: given a chain of fixed length and constant linear density suspended between two poles situated at $x = -R/2$ and $x = R/2$ (the height of these poles are fixed as well), what is the shape of the chain that minimizes the potential energy?

The total potential energy of the chain is the functional

$$E[y] = \int_{-R/2}^{R/2} d\ell\, y = \int_{-L}^{L} dx\, y\sqrt{1+y'^2}\,, \tag{1.106}$$

where I've dropped the density and the acceleration due to gravity for simplicity: they enter only as an overall multiplicative factor to the energy. I would like to extremize $E[y]$ subject to the constraint that the length of the chain is some constant L, i.e.

$$\int_{-R/2}^{R/2} dx\, \sqrt{1+y'^2} = L\,. \tag{1.107}$$

As before, we define a new functional $\tilde{E}[y,\lambda]$, where λ is a real number Lagrange multiplier, and

$$\begin{aligned}\tilde{E}[y,\lambda] &= E[y] - \lambda\left(\int_{-R/2}^{R/2} dx\, \sqrt{1+y'^2} - L\right) \\ &= \int_{-R/2}^{R/2} dx\, \left[(y-\lambda)\sqrt{1+y'^2} + \frac{\lambda L}{R}\right]. \end{aligned}\tag{1.108}$$

The equation of motion for y is then given by extremizing the first term in the equation above. We can just focus on the part of the action given by the integral since the other term doesn't depend on y, and use the Euler-Lagrange equations. Or, if we are a little bit cleverer, we can use the fact that the first integral is constant, since there is no explicit x dependence, i.e.

$$\begin{aligned}(y-\lambda)\sqrt{1+y'^2} - \frac{y'^2(y-\lambda)}{\sqrt{1+y'^2}} = C \implies 1+y'^2 - y'^2 &= \frac{C}{y-\lambda}\sqrt{1+y'^2} \\ \implies \frac{(y-\lambda)^2}{C^2} - y'^2 &= 1\,. \end{aligned}\tag{1.109}$$

This equation should remind you of $\cosh^2 t - \sinh^2 t = 1$, and so the general solution is

$$y - \lambda = C\cosh\left(\frac{x+D}{C}\right), \tag{1.110}$$

i.e. the minimum energy curve should look like a hyperbola. This shape is also known as the **catenary**, and the constants are fixed by the heights of the two poles, and the requirement that the total length of the curve is L.

2 Calculus on Manifolds

References: Stone & Goldbart (SG) Chapter 10, Appendix A; Carroll Chapter 2

We're all very familiar by now with how to do calculus in $\mathbb{R}^n$. But often in physics, we want to be able to do calculus in spaces that don't quite look like $\mathbb{R}^n$. One example you've probably already seen is doing calculus in *Minkowski space* in special relativity. Another perhaps more mundane example is doing calculus on a sphere. Most of our intuition, however, is grounded in the logic and structures of $\mathbb{R}^n$, which we sometimes take for granted. And so it will help for us to think very carefully about how $\mathbb{R}^n$ works in a more formal language, in order for us to see how to generalize to other spaces.

2.1 Some Facts from Linear Algebra

The study of spaces like $\mathbb{R}^n$ falls under the subject of **linear algebra**. While a course in mathematical physics might feel a little incomplete without covering this topic in detail, linear algebra is generally well-covered in undergraduate curricula. We'll content ourselves with a lightning review of some key facts.

2.1.1 Vector spaces and inner products

So what are the structures of $\mathbb{R}^n$ that we take for granted? First, the fundamental quantities that we deal with in real space are **vectors**. A collection of objects that live in real space is simply a set, but the interesting thing about real space is that there are relations between the objects in the space. In fact, real space is an example of a **vector space**, a structure which is defined as follows:

A **vector space** V over a field $\mathbb{F}$ is a set equipped with two operations: a binary operation called **vector addition** which assigns to each pair of elements $\vec{x}, \vec{y} \in V$ a third element denoted $\vec{x} + \vec{y}$, and **scalar multiplication** which assigns to an element $\vec{x} \in V$ and $\lambda \in \mathbb{F}$ a new element $\lambda\vec{x} \in V$. There is also a distinguished element $\vec{0} \in V$ such that the follow axioms are obeyed:

1. Vector addition is commutative: $\vec{x} + \vec{y} = \vec{y} + \vec{x}$;
2. Vector addition is associative: $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$;
3. Additive identity: $\vec{0} + \vec{x} = \vec{x}$;
4. Existence of an additive inverse: for any $\vec{x} \in V$, there is an element $-\vec{x} \in V$ such that $\vec{x} + (-\vec{x}) = \vec{0}$;
5. Scalar distributive law: $\lambda(\vec{x}+\vec{y}) = \lambda\vec{x}+\lambda\vec{y}$, as well as $(\lambda+\mu)\vec{x} = \lambda\vec{x} + \mu\vec{x}$;
6. Scalar multiplication is associative: $(\lambda\mu)\vec{x} = \lambda(\mu\vec{x})$, and
7. Multiplicative identity: $1\vec{x} = \vec{x}$.

A lot of that just seems very natural, and so it might seem like a lot of useless abstraction. But the point is to be clear about what a vector in the abstract actually is, so that when we're in much less familiar settings, these formal structures are going to help us cut through the confusion. Furthermore, we can study properties of all vector spaces that would apply equally well to $\mathbb{R}^n$ as it does to any other vector space.

Here, you can see that $\mathbb{R}^n$ is a vector space over the field $\mathbb{R}$. However, you're also familiar with vector spaces over the complex numbers $\mathbb{C}$: one example is the *Hilbert space*, which underpins *quantum mechanics*. The states of a system are described as vectors $|\psi\rangle$ in a Hilbert space. The results of linear algebra apply equally well to both $\mathbb{R}^n$ and Hilbert spaces.

In addition to being able to add vectors, or multiply vectors by real numbers, another important thing you can do in $\mathbb{R}^n$ is talk about distances: you can take the dot product or **inner product** of a vector with itself to talk about length, or take the inner product of two different vectors and talk about angles. Formally, vector spaces with this additional structure are called **inner product spaces**. Inner product spaces are defined as follows:[17]

An **inner product space** is a vector space V over a field $\mathbb{F}$, together with an **inner product**, which is a map

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}\,, \tag{2.1}$$

that satisfies the following properties for all $\vec{x}, \vec{y}, \vec{z} \in V$ and $\lambda, \mu \in \mathbb{F}$:

1. Conjugate symmetry: $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle^*$, where $*$ denotes complex conjugation. If $\mathbb{F}$ is real, then this just means that the inner product should be symmetric;
2. Linearity in the second argument, i.e. $\langle \vec{x}, \lambda\vec{y} + \mu\vec{z} \rangle = \lambda \langle \vec{x}, \vec{y} \rangle + \mu \langle \vec{x}, \vec{z} \rangle$. Note that this together with conjugate symmetry implies that $\langle \lambda\vec{x} + \mu\vec{y}, z \rangle = \lambda^* \langle \vec{x}, \vec{z} \rangle + \mu^* \langle \vec{y}, \vec{z} \rangle$. The inner product is only linear in both arguments when $\mathbb{F} = \mathbb{R}$, and
3. Nondegenerate, i.e. if $\langle \vec{x}, \vec{y} \rangle = 0$ for all $\vec{y}$, then $\vec{x} = 0$.

The inner product on $\mathbb{R}^n$ is the dot product; in Hilbert space, it is denoted $\langle \psi' | \psi \rangle$; in Minkowski space, we have the metric tensor. We'll go into a lot more detail on this in just a bit.

The last thing that we'll talk about are **linear transformations** (also known as linear operators or linear maps), which are functions that take us between vector spaces. Let V and W be vector spaces with dimensions n and m respectively; $A : V \to W$ is a linear transformation if

$$A(\lambda\vec{x} + \mu\vec{y}) = \lambda A(\vec{x}) + \mu A(\vec{y})\,. \tag{2.2}$$

2.1.2 Bases and components

At this point, the vectors on a vector space are still abstract objects. In order to make contact with our usual representation of vectors as a column of numbers, we need to define a **basis** for the vector space. This is something that you've probably seen in linear algebra, but we'll state some facts and definitions that all are somewhat intuitive:

1. A set of vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is **linearly dependent** if there exist $\lambda^1, \cdots, \lambda^n \in \mathbb{F}$, written as λ^μ for $\mu = 1, \cdots, n$, not all zero, such that
$$\lambda^1\vec{e}_1 + \lambda^2\vec{e}_2 + \cdots \lambda^n\vec{e}_n = \vec{0}\,. \tag{2.3}$$
2. If it is not linearly dependent, a set of vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is **linearly independent**. For a linearly independent set, the relation
$$\lambda^1\vec{e}_1 + \lambda^2\vec{e}_2 + \cdots \lambda^n\vec{e}_n = \vec{0} \tag{2.4}$$
holds only if $\lambda^1 = \cdots = \lambda^n = 0$.

[17] There are a few differences here compared to the usual definition in mathematics. First, in mathematics, it is common to have linearity apply to the first argument. This is of course is entirely equivalent. We use this definition to conform with our usual intuition in braket notation. Second, the inner product space is usually defined as having a *positive definite* inner product; but this unfortunately excludes Minkowski space, which is more properly classified as a pseudo-inner product. We don't really care about these differences in physics though.

3. A set of vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is said to **span** V if for any $\vec{x} \in V$, there are numbers x^μ such that $\vec{x}$ can be written (not necessarily uniquely) as

$$\vec{x} = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n \,. \tag{2.5}$$

   A vector space is **finite dimensional** if a finite spanning set exists.

4. A set of vectors $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is a **basis** if it is a *maximally linearly independent set*, i.e. introducing any additional vector makes the set linearly dependent. Equivalently, a basis is a *minimal spanning set*, i.e. deleting any of the $\vec{e}_i$ destroys the spanning property.

5. If $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ is a basis, then any $\vec{x} \in V$ can be written

$$\vec{x} = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n \,, \tag{2.6}$$

   where the x^μ, known as the **components** of the vector with respect to this basis, are unique in that two vectors coincide if and only if they have the same components.

6. If the sets $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ and $\{\vec{f}_1, \vec{f}_2, \cdots, \vec{f}_m\}$ are both bases for the space V, then $m = n$. This invariant integer is the **dimension**, $\dim(V)$, of the space.

At this point, you may be looking at the notation above and wondering about the placement of indices: when are indices placed above, and when are they placed below? This will made clear in the next part of our discussion.

2.2 Change of Bases, Covariant and Contravariant Transformations

Having defined the concepts of a basis, and the components of a vector with respect to a basis, we now want to understand how these components change as we choose different bases, since bases are not unique.

Suppose a vector space V has two different bases given by $\{\vec{e}_1, \cdots, \vec{e}_n\}$ and $\{\vec{e}'_1, \cdots, \vec{e}'_n\}$. Since both sets span V, every vector in $\{\vec{e}_1, \cdots, \vec{e}_n\}$ can be written as a sum of $\{\vec{e}'_1, \cdots, \vec{e}'_n\}$, and we can define a set of n^2 numbers ${a^\mu}_\nu$ which maps

$$\vec{e}_\nu = \sum_{\mu=1}^{n} {a^\mu}_\nu \vec{e}'_\mu \equiv {a^\mu}_\nu \vec{e}'_\mu = \vec{e}'_\mu {a^\mu}_\nu \,. \tag{2.7}$$

At this point, we've introduced the famous **Einstein notation**, which just says that every repeated index should be regarded as being summed over all possible values. Again, you may be worried about the placement of the indices, but all be clear as we go along. The final expression is helpful in helping you visualize the object $\vec{e}'_\mu$ as a row vector, multiplied by the **matrix** ${a^\mu}_\nu$, where μ indexes its rows, and ν indexes its columns.

${a^\mu}_\nu$ is clearly **invertible**: every vector has a unique representation in each basis, and the map takes the coordinates of any vector in one basis to another, and so it is certainly a bijective map. We can therefore define ${(a^{-1})^\mu}_\nu$ as the inverse map,

$$\vec{e}'_\nu = {(a^{-1})^\mu}_\nu \vec{e}_\mu \,, \tag{2.8}$$

with

$${(a^{-1})^\mu}_\nu {a^\nu}_\sigma = \delta^\mu_\sigma \,, \tag{2.9}$$

where δ^μ_σ is the Kronecker delta or the identity matrix.

So far, we have dealt with the transformation of the basis. But how does a general vector transform? Given the transformation between bases above, we see that for any arbitrary vector $\vec{x}$, which can be written as $x^\nu \vec{e}_\nu$ in one basis and $x'^\mu \vec{e}'_\mu$ in the other, are related by

$$\vec{x} = x'^\mu \vec{e}'_\mu = x^\nu \vec{e}_\nu = x^\nu (a^\mu{}_\nu \vec{e}'_\mu) = (a^\mu{}_\nu x^\nu)\vec{e}'_\mu \,, \tag{2.10}$$

or in other words,[18]

$$x'^\mu = a^\mu{}_\nu x^\nu \,. \tag{2.11}$$

One thing you should notice immediately is that the basis and the coordinates transform in the opposite way:

$$x'^\mu = a^\mu{}_\nu x^\nu \,, \qquad \vec{e}'_\mu = (a^{-1})^\sigma{}_\mu \vec{e}_\sigma \,, \tag{2.12}$$

because of course the vector itself, $x'^\mu \vec{e}'_\mu = x^\mu \vec{e}_\mu$, doesn't transform under a coordinate change at all! Any quantity that transforms under a change of basis like the basis itself is said to transform **covariantly**, while any quantity that transforms like the coordinates, i.e. in the opposite manner as the basis, is said to transform **contravariantly**. *We will always use indices on the top to indicate a quantity that transforms contravariantly, and indices on the bottom to indicate a quantity that transform covariantly.*

The best intuition for this comes from imagining a change of basis via rotation in $\mathbb{R}^2$, as shown in Fig. 9. Either we can imagine the basis vectors actually rotating counterclockwise and defining a new set of axes, as we would do in taking $x'^\nu \vec{e}'_\nu = x'^\nu (a^\mu{}_\nu \vec{e}_\mu)$, or equivalently, we can think of the components of the vector themselves rotating clockwise, with the axes just being relabeled, which corresponds to $x'^\nu \vec{e}'_\nu = (a^\mu{}_\nu x'^\nu)\vec{e}_\mu$.

(**End of Lecture**: Wednesday 18 Sep 2024)

2.3 The Dual Space

For every vector space V, we can define a **dual space** V^*, which is a set of linear transformations $f : V \to \mathbb{F}$, each of which takes in a vector and returns a number. The functions f are called **covectors** or **one-forms**, and you can convince yourself that V^* is also a vector space. Since these functions are linear, we have

$$f(\vec{x}) = f(x^\mu \vec{e}_\mu) = x^\mu f(\vec{e}_\mu) \equiv x^\mu f_\mu \,, \tag{2.13}$$

where in the last equality I have defined the set of numbers $f_\mu \equiv f(\vec{e}_\mu)$, which I can construct given the basis $\{\vec{e}_\mu\}$ in V. Under a change of basis in V,

$$f_\mu = f(\vec{e}_\mu) = f(a^\nu{}_\mu \vec{e}'_\nu) = a^\nu{}_\mu f(\vec{e}'_\nu) \equiv a^\nu{}_\mu f'_\nu \,, \tag{2.14}$$

where $f'_\nu \equiv f(\vec{e}'_\nu)$ are again a set of numbers that we can construct given the basis $\{\vec{e}'_\nu\}$ in V. Notice that under a change of basis in V, f_μ transforms *covariantly*, i.e. in the same manner as the change of basis in V.

Given a basis $\vec{e}_\mu$ of V, we can define a **dual basis** for V^*, which is the set of covectors $\vec{e}^{*\mu} \in V^*$ such that

$$\vec{e}^{*\mu}(\vec{e}_\nu) = \delta^\mu_\nu \,. \tag{2.15}$$

This is clearly a basis, since for any $f \in V^*$,

$$f(\vec{x}) = x^\mu f_\mu = x^\mu f_\nu \delta^\nu_\mu = x^\mu f_\nu \vec{e}^{*\nu}(\vec{e}_\mu) = f_\nu \vec{e}^{*\nu}(x^\mu \vec{e}_\mu) \,, \tag{2.16}$$

[18] Very often, you will see the notation $a^\mu{}_\nu \equiv \partial x'^\mu / \partial x^\nu$, which makes total sense if you look at Eq. (2.11). In fact, the advantage of writing it this way tells you how to obtain the matrix $a^\mu{}_\nu$.

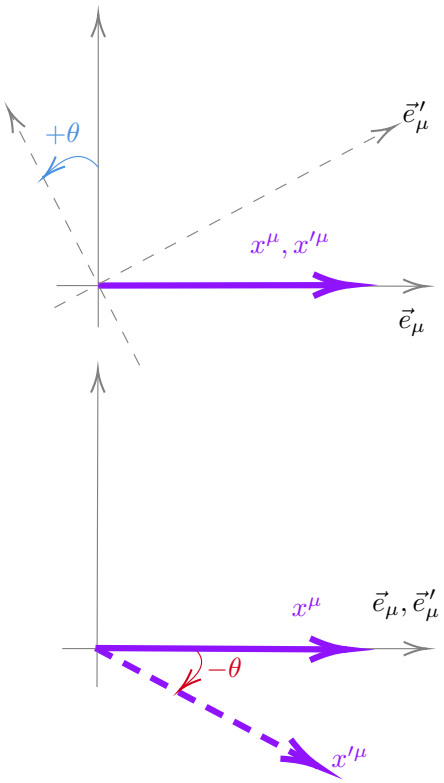

Figure 9: Two equivalent ways to think of a change of basis. (Top) the basis vectors themselves are transformed, or (bottom) the coordinates are transformed in the opposite direction. These pictures are equivalent.

or in other words,

$$f = f_\mu \vec{e}^{*\mu} \,. \tag{2.17}$$

We should therefore view $f_\mu = f(\vec{e}_\mu)$ as the *components* of f under the induced dual basis $\vec{e}^{*\mu}$.

You should already have a sense that V and V^* are very closely related; in fact, the map $\vec{e}_\mu \mapsto \vec{e}^{*\mu}$ is an **isomorphism**, i.e. a map of every element in V to another in V^* that preserves their respective relation to each other under addition and scalar multiplication.

2.4 The Metric

So far, everything we have discussed has been about vector spaces. We are now going to turn our attention to inner product spaces over $\mathbb{R}$, where the additional inner product structure is defined, giving us a way of talking about distances and angles.

As a reminder, the inner product is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that takes two vectors in V, and spits out a number. Having chosen a basis $\{\vec{e}_\mu\}$ for our vector space V, we can now define a quantity known as the **metric** or **metric tensor** $g_{\mu\nu}$,

$$g_{\mu\nu} \equiv \langle \vec{e}_\mu, \vec{e}_\nu \rangle \,. \tag{2.18}$$

For $\mathbb{R}^n$, for example, with the inner product given by the dot product, we have simply $g_{\mu\nu} = \delta_{\mu\nu}$, while for Minkowski space, the metric is given by[19]

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \,. \tag{2.19}$$

[19] I will generally stick with the mostly minus convention for the Minkowski metric. My apologies to mostly plus aficionados, but I'm just slightly more comfortable with the mostly minus convention at this point.

Knowing the metric fully defines the inner product, since

$$\langle \vec{x}, \vec{y} \rangle = \langle x^\mu \vec{e}_\mu, y^\nu \vec{e}_\nu \rangle = g_{\mu\nu} x^\mu y^\nu \,. \tag{2.20}$$

Moreover, the structure of the inner product also guarantees that $g_{\mu\nu} = g_{\nu\mu}$, i.e. $g_{\mu\nu}$ is **symmetric**. Another thing we can note is that as a matrix, $g_{\mu\nu} x^\nu = 0$ only if $x^\nu = 0$ by the definition of the inner product; this means that $g_{\mu\nu}$ is invertible. We can therefore define the **inverse of the metric**, which we denote $g^{\mu\nu}$, with

$$g_{\mu\nu} g^{\nu\sigma} = g^{\mu\nu} g_{\nu\sigma} = \delta^\mu_\sigma \,. \tag{2.21}$$

For now, this is just a relationship between matrices; we'll come back and revisit the metric when we have discussed tensors later on.

2.4.1 Raising and lowering indices

Let's look at the expression in Eq. (2.20) more closely. We can reinterpret $\langle \vec{x}, \vec{y} \rangle = g_{\mu\nu} x^\mu y^\nu$ as $(g_{\mu\nu} x^\mu) y^\nu$. With this rewriting, can think of $(g_{\mu\nu} x^\mu)$ as being the components of the object $\langle \vec{x}, \cdot \rangle$, which takes in a vector $\vec{y}$ and returns $\langle \vec{x}, \vec{y} \rangle$. In fact, $\langle \vec{x}, \cdot \rangle$ is an object in V^*, mapping vectors in V to real numbers, and can be written as $\langle \vec{x}, \cdot \rangle = g_{\mu\nu} x^\mu \vec{e}^{*\nu}$, so that

$$g_{\mu\nu} x^\mu \vec{e}^{*\nu} (y^\sigma \vec{e}_\sigma) = g_{\mu\nu} x^\mu y^\sigma \delta^\nu_\sigma = g_{\mu\nu} x^\mu y^\nu = \langle \vec{x}, \vec{y} \rangle \,. \tag{2.22}$$

Clearly then, $g_{\mu\nu}x^\mu$ are indeed the components of $\langle\vec{x},\cdot\rangle$ in the basis of V^* induced by our chosen basis of V, and therefore transforms *covariantly*.

If all that was a bit dense, the upshot is that, starting from a contravariant quantity x^μ, we can **lower its index** by defining

$$x_\nu \equiv g_{\mu\nu}x^\mu\,, \tag{2.23}$$

which 1) is a quantity that transforms covariantly (and so has a lower index), and 2) can be **contracted** with a contravariant quantity to form a real number, or a **scalar** or an **invariant**. Intuitively, it is one half of the inner product: you need to put together a covariant and contravariant piece to obtain a scalar.

Multiplying Eq. (2.23) by $g^{\sigma\nu}$ on both sides, we also find

$$g^{\sigma\nu}x_\nu \equiv g^{\sigma\nu}g_{\mu\nu}x^\mu = g^{\sigma\nu}g_{\nu\mu}x^\mu = \delta^\sigma_\mu x^\mu = x^\sigma\,, \tag{2.24}$$

which shows that I can also **raise an index** by multiplying by the inverse tensor. Ultimately, all I'm doing is switching between the components of the two objects

$$\langle\vec{x},\cdot\rangle \leftrightarrow \vec{x}\,, \tag{2.25}$$

which are in 1-to-1 correspondence with each other between the isomorphic vector spaces V and V^*.

Finally, notice that every time we perform a contraction, we sum over one upper and one lower index. This is because every contraction represents the pairing of a function in V^*, with a vector in V, and results in a scalar. Another way of understanding this is that you want to pair up a contravariant with a covariant quantity, so that you end up with a quantity that doesn't transform, i.e. a scalar. I have never encountered a situation where you want to sum over the components of two objects which transform in the same way.

2.4.2 Example: Some Common Metrics

Let's pause for a moment and take a look at some important examples.

To digest all of this information, let's revisit $\mathbb{R}^2$ with all of this technology. $\mathbb{R}^2$ is a 2D vector space, with vectors $x^i\vec{e}_i$ that look like, for example, $3\vec{e}_x+2\vec{e}_y$, where 3 and 2 are the components of the vector, and $\{\vec{e}_x,\vec{e}_y\}$ is a chosen basis for the space. $\mathbb{R}^2$ also comes with an inner product, which is the usual dot product. In $\mathbb{R}^2$, we can choose a basis that is **orthonormal**, i.e. with a metric given by

$$g_{ij} = \langle\vec{e}_i,\vec{e}_j\rangle = \vec{e}_i\cdot\vec{e}_j = \delta_{ij}\,. \tag{2.26}$$

I can use this metric to raise and lower indices of covariant or contravariant quantities, so for example

$$g_{ij}x^j = \delta_{ij}x^j = x_i\,, \tag{2.27}$$

but if you explicitly plug in the indices, you can see that $x^0 = x_0$ and $x^1 = x_1$, which shows that *the position of indices doesn't matter in* $\mathbb{R}^n$. Inner products between two vectors can be written in component form as

$$g_{ij}x^iy^j = x_jy^j\,, \tag{2.28}$$

i.e. the sum of the product of individual components, as in the usual dot product. You can think of x_j as the components of $\langle\vec{x},\cdot\rangle$. In $\mathbb{R}^2$, you can also think of

x_j as a row matrix, which maps column vectors (which contain components of a vector) to numbers by matrix multiplication.

Now let's consider a basis change, given by $\vec{e}_i' = a^j{}_i \vec{e}_j$, where

$$a^j{}_i = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{2.29}$$

where j indexes the row and i indexes the column. This is a clockwise rotation of the basis vectors by some constant angle θ. Explicitly,

$$\begin{aligned} \vec{e}_1' &= \cos\theta\, \vec{e}_1 - \sin\theta\, \vec{e}_2\,, \\ \vec{e}_2' &= \sin\theta\, \vec{e}_1 + \cos\theta\, \vec{e}_2\,. \end{aligned} \tag{2.30}$$

At the same time, for any vector $\vec{x} = x^i \vec{e}_i$, the coordinates x^i transforms in the opposite sense, i.e. $x'^i = (a^{-1})^i{}_j x^j$, where

$$(a^{-1})^i{}_j = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{2.31}$$

so that

$$\begin{aligned} x'^1 &= \cos\theta\, x^1 - \sin\theta\, x^2 \\ x'^2 &= \sin\theta\, x^1 + \cos\theta\, x^2\,, \end{aligned} \tag{2.32}$$

which is instead a *counterclockwise* rotation of the components.[20] You can check that the vector itself, $x^i \vec{e}_i$, remains unchanged. This is the same intuition we had from Fig. 9.

[20] What may be confusing is that Eq. (2.30) and Eq. (2.32) look identical! But how we interpret what's happening is different. In the first, each line tells us how each basis vector is separately rotated, so you're looking out for $\vec{e}_1$ transforming into something else. But in the second, it is the transformation of the arrow denoted by (x^1, x^2) going into (x'^1, x'^2) that we are interested in.

We now graduate to something hopefully still familiar, but a little more nontrivial: 4D Minkowski space, where the 0-dimension is time, and dimensions 1,2,3 are spatial dimensions. The vectors that live in this space are called **4-vectors**, and they are of the form $x^\mu \vec{e}_\mu$, where $\{\vec{e}_0, \vec{e}_1, \vec{e}_2, \vec{e}_3\}$ forms a basis. Once again, we have an inner product and an associated metric; we can choose a basis such that the metric is the **Minkowski metric**

$$\eta_{\mu\nu} = \langle \vec{e}_i, \vec{e}_j \rangle\,, \tag{2.33}$$

where

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.34}$$

You can verify for yourself that $\eta^{\mu\nu}$, the inverse of the matrix, has the same entries as $\eta_{\mu\nu}$.

Once again, I can lower indices by hitting a contravariant quantity with the metric, e.g. $x_\mu = \eta_{\mu\nu} x^\nu$, but this time, you can see that $x_0 = x^0$, and $x_i = -x^i$. Therefore, *the position of the indices does matter in Minkowski space*, and we need to be a little more careful. You can still think of x_μ as the components of $\langle \vec{x}, \cdot \rangle$, but x_μ is now no longer just a simple transposition (i.e. a row matrix) relative to x^μ (which we can view as a column matrix); you also need to change the sign of the spatial components. Inner products can be written, as before, as

$$g_{\mu\nu} x^\mu y^\nu = x_\nu y^\nu\,, \tag{2.35}$$

but note that because of the negative signs in the metric, *you are no longer guaranteed that the inner product is positive*.

We can again consider basis changes such as the **Lorentz boost**, given by $\vec{e}'_\nu = \Lambda^\mu{}_\nu \vec{e}_\mu$, where for example

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.36}$$

where $|\beta| < 1$ and $\gamma = (1-\beta^2)^{-1/2}$. Under this transformation,

$$\begin{aligned} \vec{e}'_0 &= \gamma\vec{e}_0 + \beta\gamma\vec{e}_1 \\ \vec{e}'_1 &= \beta\gamma\vec{e}_0 + \gamma\vec{e}_1 \\ \vec{e}'_2 &= \vec{e}_2 \\ \vec{e}'_3 &= \vec{e}_3 \,. \end{aligned} \tag{2.37}$$

On the other hand, the coordinates transform as $x'^\nu = (\Lambda^{-1})^\nu{}_\mu x^\mu$, where

$$(\Lambda^{-1})^\nu{}_\mu = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.38}$$

i.e.

$$\begin{aligned} x'^0 &= \gamma x^0 - \beta\gamma x^1 \\ x'^1 &= -\beta\gamma x^0 + \gamma x^1 \\ x'^2 &= x^2 \\ x'^3 &= x^3 \,. \end{aligned} \tag{2.39}$$

2.5 Tensors

We have now seen vector spaces and their dual spaces. We can now start defining even more general objects by putting vector spaces and dual spaces together!

Consider three vector spaces U, V and W over $\mathbb{F}$. We can define the **tensor product** of spaces such as $V \otimes W$, or even $U \otimes V \otimes W$.

1. It is distributive, i.e. for $\vec{a} \in U$ and $\vec{x} \in V$,

$$\begin{aligned} \vec{a} \otimes (\vec{x} + \vec{y}) &= \vec{a} \otimes \vec{x} + \vec{a} \otimes \vec{y}\,, \\ (\vec{a} + \vec{b}) \otimes \vec{x} &= \vec{a} \otimes \vec{x} + \vec{b} \otimes \vec{x}\,; \end{aligned} \tag{2.40}$$

2. It is associative, so that we can chain together three vector spaces like $U \otimes V \otimes W$ without worrying about whether it's $(U \otimes V) \otimes W$ or $U \otimes (V \otimes W)$;

3. It commutes with $\mathbb{F}$, i.e.

$$\lambda(\vec{a} \otimes \vec{x}) = (\lambda\vec{a}) \otimes \vec{x} = \vec{a} \otimes (\lambda\vec{x})\,, \tag{2.41}$$

but it is *not* commutative over the vectors, i.e. $\vec{a} \otimes \vec{b} \neq \vec{b} \otimes \vec{a}$ in general.

Consider a vector space V with a basis $\{\vec{e}_\mu\}$. This basis induces a basis in tensor products of V and V^* in the natural way; for example, in $V^* \otimes V^*$, this

induces the basis $\vec{e}^{*\mu} \otimes \vec{e}^{*\nu}$. These basis vectors act on pairs of basis vectors, $(\vec{e}_\alpha, \vec{e}_\beta)$, in the expected way, so for our $V^* \otimes V^*$ example, we get

$$\vec{e}^{*\mu} \otimes \vec{e}^{*\nu}(\vec{e}_\alpha, \vec{e}_\beta) = \delta^\mu_\alpha \delta^\nu_\beta \,. \tag{2.42}$$

It can also act on elements of $V \otimes V$ in a manner that you might also expect:

$$\vec{e}^{*\mu} \otimes \vec{e}^{*\nu}(\vec{e}_\alpha \otimes \vec{e}_\beta) = \delta^\mu_\alpha \delta^\nu_\beta \,. \tag{2.43}$$

A good example of what a tensor is and what it does is the metric itself, which we often refer to as **metric tensor**. It is a tensor in $V^* \otimes V^*$,

$$\mathbf{g} = g_{\mu\nu} \vec{e}^{*\mu} \otimes \vec{e}^{*\nu} \,. \tag{2.44}$$

It acts on pairs of vectors $\vec{x}$ and $\vec{y}$ and returns a number in $\mathbb{F}$:

$$\mathbf{g}(\vec{x}, \vec{y}) = g_{\mu\nu} \vec{e}^{*\mu} \otimes \vec{e}^{*\nu}(x^\alpha \vec{e}_\alpha, y^\beta \vec{e}_\beta) = g_{\mu\nu} x^\alpha y^\beta \delta^\mu_\alpha \delta^\nu_\beta = g_{\mu\nu} x^\mu y^\nu = x^\mu y_\mu \,. \tag{2.45}$$

As with vectors and covectors, once we've picked a basis, we will only need to worry about the components, with the understanding that the object itself is specified by both the components and the basis, and that the spaces act in the natural way that you expect.

Under a change of basis $\{\vec{e}_\mu\} \mapsto \{\vec{e}'_\mu\}$, the metric itself undergoes a transformation:

$$g_{\mu\nu} \mapsto g'_{\mu\nu} = \langle \vec{e}'_\mu, \vec{e}'_\nu \rangle = \langle a^\sigma{}_\mu \vec{e}_\sigma, a^\lambda{}_\nu \vec{e}_\lambda \rangle = a^\sigma{}_\mu a^\lambda{}_\nu g_{\sigma\lambda} \,. \tag{2.46}$$

Each lower index is acted on by the change-of-basis transformation, with each transformation given by a covariant transformation. We say therefore that $g_{\mu\nu}$ a doubly covariant **tensor**, which explains why we often also refer to $g_{\mu\nu}$ as the **metric tensor**. You can see now that we can look into objects with more general number of indices, say $Q^{\alpha\beta}{}_{\gamma\delta\epsilon}$, which transforms as

$$Q'^{\alpha\beta}{}_{\gamma\delta\epsilon} = (a^{-1})^\alpha{}_{\alpha'} (a^{-1})^\beta{}_{\beta'} a^{\gamma'}{}_\gamma a^{\delta'}{}_\delta a^{\epsilon'}{}_\epsilon Q^{\alpha'\beta'}{}_{\gamma'\delta'\epsilon'} \,, \tag{2.47}$$

which is a doubly contravariant, triply covariant tensor, or a type $(2,3)$ tensor. The total number of indices is what we call the **rank** of the tensor. Notice how when we were writing down the transformation of $Q^{\alpha\beta}{}_{\gamma\delta\epsilon}$, the indices lined up: we contracted upper indices with lower indices, so that under each application of the change-of-basis transformation, upper indices remain upper indices.

Another thing you will notice is that I have been very careful with the relative positions of the tensors. This is good practice, but very often you'll find people get sloppy and collapse all the indices when they think the notation is obvious. The one tensor where this is always okay is the Kronecker delta δ^μ_ν, since we always have this tensor returning 1 if $\mu = \nu$ and 0 if $\mu \neq \nu$ regardless of the position of the indices.

(**End of Lecture**: Monday Sep 23 2024)

2.5.1 Tensor algebra

So what can we do with tensors? Well, we can add tensors together, but we have to ensure that you're adding things that are transforming with a change of bases in the same way. So for example

$$A^\mu{}_{\nu\lambda} = B^{\mu\tau}{}_{\nu\lambda\tau} + C^\mu{}_{\nu\lambda} \tag{2.48}$$

is legal, but

$$A^{\mu}{}_{\nu\lambda} \overset{\text{wrong}}{=} B^{\nu}{}_{\mu\lambda} + C^{\mu}{}_{\nu\lambda\sigma\sigma} + D^{\mu}{}_{\nu\lambda\tau} \tag{2.49}$$

makes no sense.

We can multiply tensors together. Suppose we have tensors $A^{\mu}{}_{\nu\lambda}$ and $B^{\mu}{}_{\nu\lambda\tau}$, which are tensors of type $(1,2)$ and $(1,3)$ respectively. Then, we can multiply them together to get

$$C^{\alpha\beta}{}_{\nu\lambda\rho\sigma\tau} = A^{\alpha}{}_{\nu\lambda} B^{\beta}{}_{\rho\sigma\tau} \,, \tag{2.50}$$

which is a tensor of type $(2,5)$.

You can also contract indices, or equivalently, multiply tensors by the metric, so that for example

$$C^{\alpha\beta}{}_{\alpha\beta\rho\sigma\tau} = g^{\alpha\lambda} g^{\beta\mu} C_{\lambda\mu\alpha\beta\rho\sigma\tau} \tag{2.51}$$

is now a tensor of type $(0,3)$. Contracting two vectors is a special case that leads to a real number, also called a scalar or sometimes an **invariant**, which does not transform under a change of basis. Let me stress again: *you must contract one upper and one lower index!* You can write down objects like $B_{\alpha\beta\beta}$, but these objects are *not* tensors.

2.5.2 Example: Rotations and Lorentz transformations

Let's study the properties of basis transformations $a^{\mu}{}_{\nu}$ that leave the metric invariant. In $\mathbb{R}^n$, these are the transformations such that angles and lengths are all preserved. Under the transformation $a^{\mu}{}_{\nu}$, the metric tensor transforms as

$$g_{\mu\nu} \mapsto a^{\sigma}{}_{\mu} a^{\lambda}{}_{\nu} g_{\sigma\lambda} \,, \tag{2.52}$$

Transformations O that leave the metric invariant are therefore of the form

$$O^{\sigma}{}_{\mu} g_{\sigma\lambda} O^{\lambda}{}_{\nu} = g_{\mu\nu} \,. \tag{2.53}$$

In $\mathbb{R}^n$, if we start with the canonical metric $g_{\sigma\lambda} = \delta_{\sigma\lambda}$, then these transformations must satisfy

$$O^{\sigma}{}_{\mu} \delta_{\sigma\lambda} O^{\lambda}{}_{\nu} = \delta_{\mu\nu} \implies O^{\lambda}{}_{\mu} O_{\lambda\nu} = \delta_{\mu\nu} \implies (O^{-1})_{\mu}{}^{\lambda} = O^{\lambda}{}_{\mu} \,. \tag{2.54}$$

Therefore, as matrices, we must have $O^{-1} = O^{\mathsf{T}}$. The set of all such matrices is the **orthogonal matrices**, which can be thought of as a **group** called $O(n)$, called the **orthogonal group**.[21] This group is made up of matrices corresponding to *rotations* and *reflections*. The group containing matrices corresponding only to rotations is called the **special orthogonal group**, $SO(n)$.

In Minkowski space, we have instead

$$\Lambda^{\sigma}{}_{\mu} \eta_{\sigma\lambda} \Lambda^{\lambda}{}_{\nu} = \eta_{\mu\nu} \,. \tag{2.55}$$

The Λ matrices also form a group that we call the **Lorentz group**, $O(1,3)$ for spacetime. The set of all Lorentz transformations (both boosts and spatial rotations), as well as time reversal and reflections, make up this group. Again, the group with just boosts and rotations is called $SO(1,3)$.

[21] A group G is a set of elements with an operation $\cdot$, that obeys the following: 1) associativity, i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$; 2) it contains an identity element e in G such that $a \cdot e = e \cdot a = a$, and 3) every element a in G has an inverse b, such that $a \cdot b = b \cdot a = e$.

2.5.3 Example: Linear transformations

In the crash course in linear algebra, I mentioned the concept of linear transformations briefly. Let's linear transformations $M : V \to V$, mapping vectors in a

vector space V over a field $\mathbb{F}$ to other vectors in V. The linear transformation must satisfy the following property:

$$M(\lambda\vec{x}+\mu\vec{y}) = \lambda M(\vec{x}) + \mu M(\vec{y}) \tag{2.56}$$

for all $\vec{x},\vec{y}\in V$, and $\lambda,\mu\in\mathbb{F}$. This object exists independently of any basis, but given a basis, it can be represented by a matrix $M^{\mu}{}_{\nu}$, obtained by examining the action of the transformation on the basis vectors:

$$M(\vec{e}_\mu) = M^{\nu}{}_{\mu}\vec{e}_\nu\,. \tag{2.57}$$

Suppose M acting on some arbitrary vector $\vec{x}=x^\mu\vec{e}_\mu$ gives the result $\vec{y}=y^\nu\vec{e}_\nu$. We can see however that

$$\vec{y} = y^\nu\vec{e}_\nu = M(x^\mu\vec{e}_\mu) = x^\mu M(\vec{e}_\mu) = M^{\nu}{}_{\mu}x^\mu\vec{e}_\nu\,, \tag{2.58}$$

or in other words, the components transform as

$$y^\nu = M^{\nu}{}_{\mu}x^\mu\,. \tag{2.59}$$

We can therefore see that given a basis, M behaves just like matrix multiplication, as we already knew to be true from linear algebra.

In another basis related to the old one via $\vec{e}_\nu = a^{\sigma}{}_{\nu}\vec{e}'_\sigma$, we see that

$$M(\vec{e}_\mu) = M^{\nu}{}_{\mu}\vec{e}_\nu = M^{\nu}{}_{\mu}a^{\sigma}{}_{\nu}\vec{e}'_\sigma\,, \tag{2.60}$$

but also

$$M(\vec{e}_\mu) = M(a^{\lambda}{}_{\mu}\vec{e}'_\lambda) = a^{\lambda}{}_{\mu}M(\vec{e}'_\lambda) \implies M(\vec{e}'_\lambda) = (a^{-1})^{\mu}{}_{\lambda}M(\vec{e}_\mu)\,. \tag{2.61}$$

Therefore, in the new basis,

$$M(\vec{e}'_\lambda) = a^{\sigma}{}_{\nu}M^{\nu}{}_{\mu}(a^{-1})^{\mu}{}_{\lambda}\vec{e}'_\sigma\,, \tag{2.62}$$

i.e. under the change of basis,

$$M^{\sigma}{}_{\lambda} \mapsto a^{\sigma}{}_{\nu}M^{\nu}{}_{\mu}(a^{-1})^{\mu}{}_{\lambda}\,, \tag{2.63}$$

which is the transformation rule for type (1,1) tensors, and also what you may be familiar with in linear algebra about the change of basis of linear transformations.

Compare this with the transformation rule for the metric tensor,

$$g_{\mu\nu} \mapsto a^{\sigma}{}_{\mu}a^{\lambda}{}_{\nu}g_{\sigma\lambda}\,, \tag{2.64}$$

and you'll notice a big difference! And that's because the metric tensor is a type (0,2) tensor. The key lesson here is that *2D tensors are much more than just matrices*; while it can be useful to write out tensor components as matrices, one needs to keep in mind that tensors also come equipped with a transformation rule!

2.5.4 Invariants of 2D tensors

Invariants are extremely useful: they don't transform under a basis change, and so they're very easy to deal with. They're also giving you information about the tensor that is independent of the basis. The first commonly discussed invariant of 2D tensors are the **determinant**, which you can compute from the matrix representation of the tensor; we will discuss this in much greater detail after we've built up some machinery to discuss it. The second invariant is called the **trace**. In matrix language, this is simply the sum of the terms along the diagonal. It's hard to understand how this could be an invariant from the matrix perspective, but in terms of indices, it is simply given by

$$\mathrm{tr}(M) = g^{\mu\nu}M_{\mu\nu} = g_{\mu\nu}M^{\mu\nu} = \delta^{\nu}_{\mu}M^{\mu}{}_{\nu} = M^{\mu}{}_{\mu}\,, \tag{2.65}$$

where I'm showing you how to obtain the trace for 2D tensors of all types.

2.5.5 Symmetric and antisymmetric tensors

A tensor is said to be **symmetric** in some indices if the tensor values are the same when the two indices are swapped. For example, we say that $S^{\mu\nu}$ is a symmetric tensor if $S^{\mu\nu} = S^{\nu\mu}$. Note that we have also

$$S^{\mu}{}_{\nu} = g_{\nu\alpha}S^{\mu\alpha} = g_{\nu\alpha}S^{\alpha\mu} = S_{\nu}{}^{\mu}\,, \tag{2.66}$$

and so the position of the indices doesn't matter in determining whether a tensor is symmetric or not. Moreover, under a change of basis,

$$S'^{\alpha\beta} = a^{\alpha}{}_{\mu}a^{\beta}{}_{\nu}S^{\mu\nu} = a^{\alpha}{}_{\mu}a^{\beta}{}_{\nu}S^{\nu\mu} = S'^{\beta\alpha}\,, \tag{2.67}$$

and so a symmetric matrix stays symmetric under a change of basis.

We can similarly define a tensor to be **antisymmetric** in some indices if, under a swap of the two indices, the tensor picks up a minus sign, e.g. $A^{\mu\nu} = -A^{\nu\mu}$. Again, a tensor that is antisymmetric when indices are in one position are still antisymmetric if the indices are raised or lowered; they also remain antisymmetric under an arbitrary change of basis.

The contraction of a symmetric and an antisymmetric tensor is always zero, since e.g.

$$S_{\mu\nu}A^{\mu\nu} = S_{\nu\mu}A^{\mu\nu} = -S_{\nu\mu}A^{\nu\mu} = -S_{\mu\nu}A^{\mu\nu} = 0\,, \tag{2.68}$$

where in the second last step I have simply relabeled the contracted indices (since they are dummy indices), and noted that $S_{\mu\nu}A^{\mu\nu}$ is equal to its negative, and therefore has to be zero.

Every tensor can always be decomposed into a symmetric piece and an antisymmetric piece. To see this, take an arbitrary tensor $B^{\mu\nu}$. We can always rewrite this as

$$B^{\mu\nu} = \frac{1}{2}(B^{\mu\nu} + B^{\nu\mu}) + \frac{1}{2}(B^{\mu\nu} - B^{\nu\mu}) \equiv S^{\mu\nu} + A^{\mu\nu}\,, \tag{2.69}$$

where $S^{\mu\nu} = (B^{\mu\nu} + B^{\nu\mu})/2$ a symmetric tensor, and $A^{\mu\nu} = (B^{\mu\nu} - B^{\nu\mu})/2$ is antisymmetric.

2.5.6 Kronecker and Levi-Civita tensors

We'll now discuss two special tensors that arise very commonly in tensor algebra. The first of them, the **Kronecker delta** δ^{μ}_{ν}, is defined as a (1,1) tensor that, in some basis, is unity if $\mu = \nu$ and zero otherwise. Let's check what happens under an arbitrary change of basis:

$$\delta^{\mu}_{\nu} \mapsto (a^{-1})^{\lambda}{}_{\nu}a^{\mu}{}_{\sigma}\delta^{\sigma}_{\lambda} = \delta^{\mu}_{\nu}\,. \tag{2.70}$$

In other words, *the Kronecker delta always has the same numerical components in all coordinate systems*.[22] [23]

The **Levi-Civita symbol** $\epsilon_{\mu_1\mu_2\cdots\mu_n}$ is defined as an object with n indices such that $\epsilon_{12\cdots n} = 1$, and $\epsilon_{\cdots i_p\cdots i_q\cdots} = -\epsilon_{\cdots i_q\cdots i_p\cdots}$, i.e. every time two indices are exchanged, the result differs by a minus sign. This definition guarantees that when two indices are equal, the Levi-Civita symbol is zero. One particularly important use-case of the Levi-Civita symbol is in expressing antisymmetric

[22] This explains why the indices are almost never distinguished in the literature: both $\delta^{\mu}{}_{\nu}$ and $\delta_{\nu}{}^{\mu}$ act the same way on any object: just replace μ with ν or vice-versa.

[23] Sometimes, people talk about the (0,2) tensor $\delta_{\mu\nu} = g_{\mu\lambda}\delta^{\lambda}_{\nu} = g_{\mu\nu}$, which as you can see is really the metric tensor (whose coordinates *can* change with a change of basis). I will usually avoid using $\delta_{\mu\nu}$, unless we are talking about $\mathbb{R}^n$, where the index position doesn't matter.

quantities in component form. For example, the **cross product** $\vec{a} \times \vec{b}$ can be written as

$$(\vec{a} \times \vec{b})_k = \epsilon_{ijk} a^i b^j \,. \tag{2.71}$$

The **determinant** of an $n \times n$ matrix M can likewise be written as (assuming summation over repeated indices)

$$\epsilon_{\mu_1 \cdots \mu_n} \mathsf{det}(M) = \epsilon_{\nu_1 \cdots \nu_n} M^{\nu_1}{}_{\mu_1} \cdots M^{\nu_n}{}_{\mu_n} \,. \tag{2.72}$$

Let's suppose there is an n-dimensional tensor $\eta_{\mu_1 \mu_2 \cdots \mu_n}$ whose components coincide with $\epsilon_{\mu_1 \mu_2 \cdots \mu_n}$ in one particular basis. Then under a change of basis,

$$\begin{aligned} \eta_{\mu_1 \cdots \mu_n} &\mapsto a^{\nu_1}{}_{\mu_1} a^{\nu_2}{}_{\mu_2} \cdots a^{\nu_n}{}_{\mu_n} \epsilon_{\nu_1 \cdots \nu_n} \\ &= \epsilon_{\mu_1 \cdots \mu_n} \mathsf{det}(a) = \eta_{\mu_1 \cdots \mu_n} \mathsf{det}(a) \,. \end{aligned} \tag{2.73}$$

We see that the Levi-Civita symbol is *almost* a tensor whose components do not transform, up to a pesky determinant.

At this point, let's examine the determinant of the metric tensor itself, $g \equiv \mathsf{det}(g_{\mu\nu})$. We see that under the same transformation, the determinant after the change of basis is

$$g' \equiv \mathsf{det}(g'_{\mu\nu}) = \mathsf{det}(a^\lambda{}_\mu a^\sigma{}_\nu g_{\lambda\sigma}) = (\mathsf{det}(a))^2 g \,, \tag{2.74}$$

and therefore the quantity $\sqrt{|g'|} = |\mathsf{det}(a)| \sqrt{|g|}$; note that the absolute value is important, since the metric tensor can have negative determinant (e.g. in Minkowski space!). Now let's consider the object

$$\varepsilon_{\mu_1 \mu_2 \cdots \mu_n} = \sqrt{|g|} \epsilon_{\mu_1 \mu_2 \cdots \mu_n} \tag{2.75}$$

Under the same transformation, we now see that provided $\mathsf{det}(a) > 0$, which are referred to as *orientation preserving changes of basis*,[24]

$$\varepsilon_{\mu_1 \cdots \mu_n} \mapsto \sqrt{|g|} \mathsf{det}(a) \epsilon_{\mu_1 \cdots \mu_n} = \sqrt{|g'|} \epsilon_{\mu_1 \cdots \mu_n} \,, \tag{2.76}$$

and so once again, we find a tensor known as the **Levi-Civita tensor** $\varepsilon_{\mu_1 \cdots \mu_n}$ that always has the same form in any basis (although we must evaluate the determinant of the transformed metric, which is in general different for each basis).

If we limit ourselves to rotations in $\mathbb{R}^n$ and boosts (and rotations) in Minkowski space, then $\varepsilon_{\mu_1 \cdots \mu_n} = \epsilon_{\mu_1 \cdots \mu_n}$ is a good old tensor. However, if we include transformations like reflections, for example, then we need a lot more care.[25]

[24] An easy way to see that $\varepsilon_{\mu_1 \cdots \mu_n}$ needs to pick up a sign under non-orientation-preserving transformations is by considering what happens to a cross-product under reflection, where $\vec{x} \mapsto -\vec{x}$. With the sign flip, $\varepsilon^{ijk} x_j y_k \mapsto -\varepsilon^{ijk} x_j y_k$ as needed, since the final vector as a result of the cross product also needs to pick up a minus sign under reflection.

[25] You might have heard a joke about how 'a tensor is something that transforms like a tensor'. We can always exclude some class of basis changes from consideration to make a class of objects well-behaved and tensor-like.

2.5.7 Isotropic Cartesian tensors

Another special tensor that we'll now discuss is the isotropic, Cartesian tensor. Consider a Cartesian coordinate system with orthonormal basis vectors, so that $g_{ij} = \delta_{ij}$, the Kronecker delta function. We looked at the set of orthogonal matrices form what we called $O(n)$, which are matrices with the property that $O^{-1} = O^\mathsf{T}$. When we perform a change of basis under these matrices, we found that this leaves the metric invariant, since

$$g'_{kl} = O^i{}_k O^j{}_l \delta_{ij} = O_{jk} O^j{}_l = O^\mathsf{T}_{kj} O^j{}_l = \delta_{kl} \,. \tag{2.77}$$

You can check that the same thing is true for products of δ_{ij}, for example $T_{ijklmn} = \delta_{ij}\delta_{kl}\delta_{mn}$.

What is the most general form a tensor of rank m that is invariant under an $O(n)$ transformation? These are important questions that have tedious answers to them, and so we won't try to prove these results (see for example Ref. [4], but rather just state them. The most general $O(n)$ invariant tensor of rank 4 I_{ijkl} is

$$I_{ijkl} = \alpha\delta_{ij}\delta_{kl} + \beta\delta_{ik}\delta_{lj} + \gamma\delta_{il}\delta_{jk} \tag{2.78}$$

for some numbers α, β and γ.

How about for $SO(n)$, which is simply all the orthogonal matrices with determinant 1, corresponding to just rotations, and not reflections? We can check, for example, that ϵ_{ijk} is invariant under an $SO(3)$ transformation O, since

$$\begin{aligned}\epsilon'_{lmn} &= O^i{}_l O^j{}_m O^k{}_n \epsilon_{ijk} \\ &= \epsilon_{lmn}\mathrm{det}(O) \\ &= \epsilon_{lmn}\,,\end{aligned} \tag{2.79}$$

since by definition, an $SO(n)$ matrix has determinant 1. With this, it's not surprising that, for example, the most general $SO(4)$-invariant, rank-4 tensor J_{ijkl} is

$$J_{ijkl} = \alpha\delta_{ij}\delta_{kl} + \beta\delta_{il}\delta_{jk} + \gamma\delta_{il}\delta_{jk} + \lambda\epsilon_{ijkl}\,.$$

(**End of Lecture**: Wednesday Sep 25 2024)

2.6 Vectors and Tensors on Manifolds

We are now armed and ready to do calculus on surfaces that aren't just flat spaces like $\mathbb{R}^n$, or even flat spacetimes like Minkowski space. We want to build up the machinery to be able to do calculus on what are called **manifolds**.

So what are manifolds? There's a lot of technical set-up that Carroll goes through in his notes/textbook, while Stone and Goldbart basically condenses it down into a paragraph of dense text. But for our purposes the details don't really matter. A manifold is something that, if you zoom in enough, looks like $\mathbb{R}^n$ or Minkowski space, $\mathbb{R}^{1,3}$ (from here on, I will include Minkowski space when I say $\mathbb{R}^n$), in the sense that you can map patches of a manifold to these spaces in a completely invertible way. Spheres and tori are examples of manifolds; a line ending on a plane is an example that is *not* a manifold. The 2D surface of a cone is also a manifold, but there's something badly behaved about the tip of the cone: curves that go through it have to under a sudden change in direction, and so calculus on the cone sounds problematic near the tip. We will in general only care about **differentiable** manifolds, which do not have such troubling singularities.

Note that we only need to be able to map the manifold into $\mathbb{R}^n$ one patch at a time; manifolds in general do not usually allow for the entire manifold to be mapped at one go into $\mathbb{R}^n$. Take for example the sphere: any patch of the sphere can be mapped without trouble onto $\mathbb{R}^2$, but not the whole sphere at one go. This is why map making is so complicated!

So how do we set ourselves up to do calculus on manifolds? There are a few things to consider:

1. One thing we take for granted in $\mathbb{R}^n$, however, is that vectors can "slide around" on $\mathbb{R}^n$ trivially, allowing us to compare vectors from one point in $\mathbb{R}^n$ with another without much thought. This is very different on arbitrary manifolds however; on a sphere like to Earth, sliding a vector around in a closed loop causes the vector to rotate with respect to itself (see Fig. 10)! We therefore have to be extremely careful about sliding vectors around, or comparing vectors across different points. As such, vectors on general manifolds are therefore objects that belong to each particular *point* on the manifold.

2. We know that any manifold looks locally like $\mathbb{R}^n$. Picture a sphere for example: just like on Earth, every point locally looks like $\mathbb{R}^2$. Instinctively, we understand that at every point, we can set up an n-dimensional vector space with a metric. This is known as the **tangent space**: the space of all vectors at a point on a manifold. With the picture of the Earth in mind, the tangent space is a 2D vector space with vectors tangent to the surface of the Earth. But keep in mind that this picture is misleading, because the "arrows" that we have in our heads point out into outer space, whereas a general manifold doesn't have to embedded into a higher dimensional space at all.

3. Which brings us to our next point. We also want to find a way to describe the n-dimensional vector space that doesn't depend on a choice of basis or an embedding like the Earth into outer space. A vector in the tangent space should be an object that knows only about the manifold itself, and doesn't depend on the choice of basis. Of course, we will often want to choose a basis to calculate things, but the definition of the space should be basis independent.

2.6.1 Tangent spaces

We'll now try to put this intuition on solid mathematical ground. To properly define the tangent space at the point p on a manifold M, let's consider the set of all smooth functions $f : M \to \mathbb{R}$. Take any curve on the manifold passing through p parametrized by λ. Then at point p, each curve defines an operator $d/d\lambda$ that maps $f \mapsto df/d\lambda$, which is simply the **directional derivative** at the point of f along the curve. *The tangent space T_p at the point p is then the space of directional derivative operators $d/d\lambda$ through p.* This intuition is illustrated in Fig. 11.

We can check is that the directional derivatives do form a vector space. A lot of the desired properties in the definition in Sec. 2.1.1 follow from the properties of derivatives; you can check many of them yourself. Most importantly, the linear combination of any two directional derivatives, $a(d/d\lambda) + b(d/d\eta)$, is itself a perfectly good directional derivative, obeying all the expected rules of the derivative, e.g. the product rule, and so the vector space is closed on itself.

Let's move on next to constructing a basis for T_p. By definition of a manifold M, we can zoom in on a small patch around p, which looks like $\mathbb{R}^n$, where n is the **dimension** of M. On this patch, I can define coordinates x^μ on this $\mathbb{R}^n$-like patch, so that for any function $f : M \to \mathbb{R}$ on this patch, the directional derivative is simply

$$\frac{df}{d\lambda} = \frac{\partial f}{\partial x^\mu}\frac{dx^\mu}{d\lambda}\,, \tag{2.80}$$

as expected from the chain rule. This shows that any element of T_p can be written as

$$\frac{d}{d\lambda} = \frac{dx^\mu}{d\lambda}\partial_\mu\,, \tag{2.81}$$

where we've defined $\partial_\mu \equiv \partial/\partial x^\mu$. Notice that this has the form of $y^\mu \vec{e}_\mu$ as we have come to expect from vectors in a vector space; we should therefore interpret $dx^\mu/d\lambda$ as components, and ∂_μ as a basis. The physical intuition is very similar to what is shown in Fig. 11 once again, except now you should imagine throwing on a small $\mathbb{R}^2$ surface with a Cartesian grid, which locally

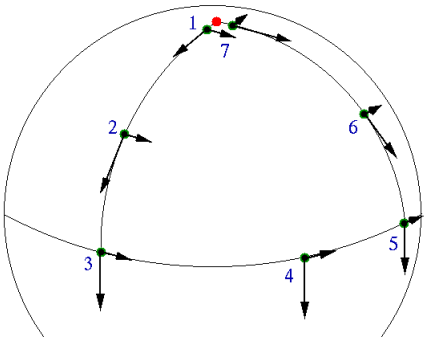

Figure 10: Sliding a vector around on a sphere causes it to rotate with respect to itself!

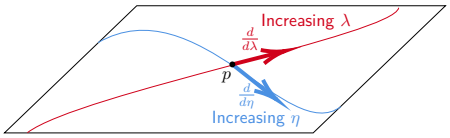

Figure 11: Directional derivatives as elements of the tangent space T_p.

defines a coordinate system x^μ, and which then allows T_p to inherit the basis $\{\partial_\mu\}$. This is known as a **coordinate basis** for T_p. Keep in mind that you are *not limited to choosing a Cartesian basis!* You are free to choose, for example, polar coordinates (r, θ) in $\mathbb{R}^2$, in which case your basis is $\{\partial_r, \partial_\theta\}$; this basis is *neither normalized nor orthogonal*, but it is perfectly acceptable.

What happens under a change of coordinates, say switching between Cartesian and polar coordinates on $\mathbb{R}^2$? That induces also a change of basis. Suppose we perform a change of coordinates from x^μ to x'^μ; all of the derivatives then transform as[26]

$$\partial'_\nu \equiv \frac{\partial}{\partial x'^\nu} = \frac{\partial x^\mu}{\partial x'^\nu}\partial_\mu \,. \tag{2.82}$$

[26] Note that Carroll adopts the notation $x^{\mu'}$ instead, whereas SG avoids using x'^μ, using z^μ to denote the components in the new basis, but then using the notation ∂'_μ to denote differentiation in the new basis. Every notation is nonideal in its own way.

A simple example is the transformation from polar to Cartesian coordinates, given by $x = r\cos\theta$ and $y = r\sin\theta$, so that

$$\partial_r = \frac{\partial x}{\partial r}\partial_x + \frac{\partial y}{\partial r}\partial_y = \cos\theta\cdot\partial_x + \sin\theta\cdot\partial_y \,, \tag{2.83}$$

$$\partial_\theta = \frac{\partial x}{\partial \theta}\partial_x + \frac{\partial y}{\partial \theta}\partial_y = -r\sin\theta\cdot\partial_x + r\cos\theta\cdot\partial_y \,. \tag{2.84}$$

Given this, any arbitrary $V \in T_p$ can be written as $V = V^\mu\partial_\mu$ given any choice of coordinate system. Under a change of coordinates, becomes

$$V^\mu\partial_\mu = V'^\nu\partial'_\nu = V'^\nu\frac{\partial x^\mu}{\partial x'^\nu}\partial_\mu \,, \tag{2.85}$$

and therefore

$$V'^\nu = \frac{\partial x'^\nu}{\partial x^\mu}V^\mu \,. \tag{2.86}$$

Comparing Eqs. (2.82) and (2.86), we can see that the components of the vector transforms **contravariantly**, while the basis vectors transform **covariantly**. As before, we'll move away rapidly from the formal beginnings we are at here, and will almost entirely deal with the components of vector, leaving the basis implicit. It is therefore common to say that "vectors transform contravariantly" according to Eq. (2.86), even though we really just mean the components.

Finally, we call the collection of all the tangent spaces at every point p on M the **tangent bundle**.

(**End of Lecture**: Monday Sep 30 2024)

2.6.2 Cotangent spaces, tensors and the metric

As we did before, we can now proceed to look at the dual of the tangent space T_p^* at the point p on M, which we call the **cotangent space**. This is a vector space of linear maps $\omega : T_p \to \mathbb{R}$, and the elements that live in T_p^* are called **one-forms**. As before, a choice of basis $\{\partial_\mu\}$ in T_p induces a basis on T_p^*, which we write suggestively as $\{\mathrm{d}x^\mu\}$, such that

$$\mathrm{d}x^\mu(\partial_\nu) = \frac{\partial x^\mu}{\partial x^\nu} = \delta^\mu_\nu \,. \tag{2.87}$$

Any arbitrary $\omega \in T_p^*$ can therefore be written as $\omega = \omega_\mu \mathrm{d}x^\mu$. One way to think of elements in T_p^* is again to go back to the space of functions $f : M \to \mathbb{R}$ that

pass through the point p, and to view elements in T_p^* as $\mathrm{d}f$, the gradient of the function, so that given a coordinate system,

$$\mathrm{d}f = \frac{\partial f}{\partial x^\mu}\mathrm{d}x^\mu \,. \tag{2.88}$$

Under a change of basis, we can see that the basis vectors

$$\mathrm{d}x'^\nu = \frac{\partial x'^\nu}{\partial x^\mu}\mathrm{d}x^\mu \,, \tag{2.89}$$

which transform contravariantly, and the components

$$\omega'_\nu = \frac{\partial x^\mu}{\partial x'^\nu}\omega_\mu \,, \tag{2.90}$$

transform covariantly. As usual, we will only usually see the components discussed when talk about one-forms.

Similarly, the collection of all cotangent spaces at every point p on M is called the **cotangent bundle**.

We can also define **tensors** given the tangent space T_p and the cotangent space T_p^* by performing a series of tensor products. Given a coordinate system, we can define objects like

$$T = T^{\mu_1\cdots\mu_j}{}_{\nu_1\cdots\nu_k}\,\partial_{\mu_1}\otimes\cdots\otimes\partial_{\mu_j}\otimes\mathrm{d}x^{\nu_1}\otimes\cdots\otimes\mathrm{d}x^{\nu_k} \,, \tag{2.91}$$

that can act on some combination of vectors, one-forms or tensors in the same manner as before. Once again, the coordinates transform under a change of coordinate system as e.g.

$$T'^{\alpha_1\cdots\alpha_j}{}_{\beta_1\cdots\beta_k} = \frac{\partial x^{\alpha_1}}{\partial x^{\mu_1}}\cdots\frac{\partial x^{\alpha_j}}{\partial x^{\mu_j}}\frac{\partial x^{\nu_1}}{\partial x^{\beta_1}}\cdots\frac{\partial x^{\nu_k}}{\partial x^{\beta_k}}T^{\mu_1\cdots\mu_j}{}_{\nu_1\cdots\nu_k} \,. \tag{2.92}$$

Very often, however, it is better to take ∂_μ and $\mathrm{d}x^\mu$ and apply the change of coordinates we are familiar with from calculus, treating ∂_μ as a partial derivative and $\mathrm{d}x^\mu$ as an infinitesimal quantity respectively.

An extremely important rank-2 tensor that you're already familiar with is the **metric tensor**, which given some choice of coordinate system can be written as $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\otimes\mathrm{d}x^\nu$. The components of the metric must be symmetric (given the properties of the inner product), and so even though the tensor product does not commute, it is very common to see the metric written as $g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$; you should keep in mind that expressions like $2\mathrm{d}x\mathrm{d}y$ really should be thought of as $\mathrm{d}x\otimes\mathrm{d}y + \mathrm{d}y\otimes\mathrm{d}x$. Just like any $(0,2)$ tensor, $\mathrm{d}s^2$ takes in two vectors, and spits out a real number, with

$$\mathrm{d}s^2(V,W) = g_{\mu\nu}V^\mu W^\nu \,, \tag{2.93}$$

which should be familiar to you from our previous discussions about the metric.

A very simple example of a metric is the typical Euclidean $\mathbb{R}^3$ metric,

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \,, \tag{2.94}$$

which under a transformation from Cartesian to spherical coordinates gives

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \,. \tag{2.95}$$

These two metrics are really exactly the same thing, referring to the same inner product space, just written in a different coordinate basis. Another slightly less trivial metric is the metric of a 2-sphere,

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\,, \tag{2.96}$$

which is emphatically *different* from the metric of flat, Euclidean $\mathbb{R}^2$. The metric contains all relevant information about the curvature of the manifold. While we will unfortunately not have time to discuss curvature in any detail, you are now armed with all the tools necessary to talk about curvature and general relativity.

2.7 Differentiation on Manifolds

We're now finally ready to start tackling the problem of doing calculus on manifolds, starting with differentiation. This is a vast area of knowledge that we do not have time to fully explore; instead, we will take a look at two different types of derivatives we can take on a manifold, and leave a more complete treatment of this topic to more advanced courses e.g. in general relativity.

Going forward, we will now turn our attention to tensor *fields* over a manifold M that we can use a coordinate system given by x^μ to describe. This defines a tensor at every point on the manifold, all describable by components and basis vectors given by the coordinate basis induced by x^μ.

We have already seen that given a scalar field f defined everywhere on some manifold M, on which we can assign a coordinate system given by x^μ, we can construct the gradient

$$\mathrm{d}f = \frac{\partial f}{\partial x^\mu}\mathrm{d}x^\mu\,, \tag{2.97}$$

which is an object that lives in T_p^*, and is therefore a well-behaved one-form. However, now suppose instead we have a *one-form field* V_ν defined everywhere on the same manifold M. If we take the gradient of this object, $\partial_\mu V_\nu$, we do *not* in general obtain a tensor. This is because under a change of coordinates, we find

$$\begin{aligned}\partial'_\alpha W'_\beta &= \frac{\partial x^\mu}{\partial x'^\alpha}\partial_\mu\left(\frac{\partial x^\nu}{\partial x'^\beta}W_\nu\right)\\ &= \frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta}\partial_\mu W_\nu + W_\nu\frac{\partial x^\mu}{\partial x'^\alpha}\cdot\frac{\partial}{\partial x^\mu}\frac{\partial x^\nu}{\partial x'^\beta}\,.\end{aligned} \tag{2.98}$$

If the second term wasn't there, we would have obtained an honest-to-goodness $(0,2)$ tensor, but unfortunately that's not what we have. Objects like $\partial_\mu W_\nu$ are therefore not tensors in general, and are coordinate-dependent objects, and not the coordinate-independent objects that we want tensors to be. This shouldn't have come as too much of a surprise; we began our discussion of manifolds by emphasizing that vectors and one-forms are very much objects that are tied to the tangent space at point p, whereas the gradient of a vector or one-form naturally means subtracting vectors between nearby points, which we need to define carefully.

We therefore have to work a bit harder to define operations that look like derivatives on manifolds. It turns out that there are several operations that behave like derivatives on the manifold; in this course, we will only have time to look at two of them: the **Lie derivative**, and the **exterior derivative**. We will leave the **covariant derivative** to courses focused on classical and quantum field theories to explore.

2.7.1 The Lie derivative

We saw earlier that vector fields $X \equiv X^\mu \partial_\mu$ on a manifold M can be thought of as directional derivatives acting on functions on the manifold M, i.e. $X(f) = X^\mu \partial_\mu f$, the directional derivative of f in the direction of X^μ. Consider yet another vector field $Y \equiv Y^\mu \partial_\mu$, and define the **Lie bracket**

$$[X, Y](f) \equiv X(Y(f)) - Y(X(f)) \,. \tag{2.99}$$

In your problem set, you will prove that the following statements about the Lie bracket are true:

1. The Lie bracket is linear, i.e. for f, g functions on M and a, b real numbers,

$$[X, Y](af + bg) = a[X, Y](f) + b[X, Y](g) \,. \tag{2.100}$$

2. The Lie bracket obeys Leibniz's rule, i.e.

$$[X, Y](fg) = f[X, Y](g) + g[X, Y](f) \,. \tag{2.101}$$

3. In terms of the components of X and Y, we can write

$$[X, Y]^\mu = X^\nu \partial_\nu Y^\mu - Y^\nu \partial_\nu X^\mu \,. \tag{2.102}$$

4. $[X, Y]^\mu$ transforms under a change of coordinates as a vector.

What does the Lie bracket represent? For a given vector field Y, we can define the **flow** associated with Y as the map that takes a point x_0^μ and maps it to $x(t)$ by solving

$$\frac{dx^\mu}{dt} = Y^\mu \tag{2.103}$$

with initial condition $x^\mu(0) = x_0^\mu$. In other words, the vector field Y can be thought of as a velocity field of a flowing fluid, with x riding along with the fluid. Now consider two vector fields X and Y. Starting at a point, flow along X for some small time t, and then along Y for some small time s. Now, try to undo the flow by now flowing along $-X$ for some time t again, and along $-Y$ for some time s. The Lie bracket tells you how far away you are after this procedure from the starting point: you will miss by $\delta x^\mu = st[X, Y]^\mu$ as s and t go to zero.

To get a better sense of what the Lie bracket is doing, let's consider $\mathbb{R}^2$ with polar coordinates (r, ϕ), with the vector fields

$$A = \frac{1}{r}\partial_\phi \,, \qquad B = \partial_r \,. \tag{2.104}$$

The flow lines on the 2D-plane corresponds to circular paths around the origin for A, and radial lines outward for B. The Lie bracket for these vector fields is

$$[A, B] = \frac{1}{r}\partial_\phi \partial_r - \partial_r \left(\frac{1}{r}\partial_\phi\right) = \frac{1}{r^2}\partial_\phi \,. \tag{2.105}$$

We can see that if we take a little circuit where we flow along A over some time t, followed by B over some time s, and then back along $-A$ over time t and then along $-B$ over time $-s$, we find that

$$\delta r = 0 \,, \qquad \delta\phi = \frac{t}{r} - \frac{t}{r+s} = \frac{ts}{r^2} = st[A, B]^\theta \,. \tag{2.106}$$

We are now ready to define an operation called the **Lie derivative** $\mathcal{L}_X$ along some vector field X. It acts on a scalar function f through X itself, i.e.

$$\mathcal{L}_X f \equiv X(f) = X^\mu \partial_\mu f \,, \tag{2.107}$$

which is, once again, the directional derivative of f along X. Next, on a vector field Y, it acts via the Lie bracket, i.e.

$$\mathcal{L}_X Y \equiv [X, Y] \,. \tag{2.108}$$

We can now deduce the action of the Lie derivative on more complicated objects by ensuring that the Lie derivative respects *1) linearity*, with $\mathcal{L}_X(aT + bS) = a\mathcal{L}_X T + b\mathcal{L}_X S$, and *2) Leibniz's rule*, i.e. the product rule, $\mathcal{L}_X(T \otimes S) = (\mathcal{L}_X T) \otimes S + T \otimes (\mathcal{L}_X S)$.

For example, to see how the Lie derivative acts on a one-form ω_μ, we see that given a vector field V^μ, the Lie derivative should act on the object $\omega_\mu V^\mu$ in the following manner:

$$\mathcal{L}_X(\omega_\mu V^\mu) = X_\nu \partial^\nu (\omega_\mu V^\mu) = X_\nu (V^\mu \partial^\nu \omega_\mu + \omega_\mu \partial^\nu V^\mu) \,. \tag{2.109}$$

On the other hand, Leibniz's rule requires

$$\begin{aligned} \mathcal{L}_X(\omega_\mu V^\mu) = V^\mu \mathcal{L}_X \omega_\mu + \omega_\mu \mathcal{L}_X V^\mu &= V^\mu \mathcal{L}_X \omega_\mu + \omega_\mu [X, V]^\mu \\ &= V^\mu \mathcal{L}_X \omega_\mu + \omega_\mu (X^\nu \partial_\nu V^\mu - V^\nu \partial_\nu X^\mu) \,. \end{aligned} \tag{2.110}$$

Comparing both expressions, we find

$$V^\mu \mathcal{L}_X \omega_\mu = X_\nu V^\mu \partial^\nu \omega_\mu + \omega_\mu V^\nu \partial_\nu X^\mu \implies \mathcal{L}_X \omega_\mu = X_\nu \partial^\nu \omega_\mu + \omega_\nu \partial_\mu X^\nu \,. \tag{2.111}$$

You can check that $\mathcal{L}_X \omega_\mu$ transforms as a one-form; in fact, the Lie derivative of any (k, l)-tensor always transforms as a (k, l)-tensor.

Of particular importance is the Lie derivative of $(0, 2)$-tensors such as the metric, which reads (and you will prove in your problem set)

$$(\mathcal{L}_X g)_{\mu\nu} = X^\alpha \partial_\alpha g_{\mu\nu} + g_{\mu\alpha} \partial_\nu X^\alpha + g_{\alpha\nu} \partial_\mu X^\alpha \,. \tag{2.112}$$

Let's compare this with what happens to the metric $g_{\mu\nu}(x) \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu$ when we make a displacement $x^\alpha \mapsto x^\alpha + \epsilon X^\alpha$ in the direction of the vector field X. Under such a displacement, we have

$$\begin{aligned} \mathrm{d}x^\mu &\mapsto \mathrm{d}x^\mu + \epsilon \partial_\beta X^\mu \cdot \mathrm{d}x^\beta \,, \\ g_{\mu\nu} &\mapsto g_{\mu\nu} + \partial_\beta g_{\mu\nu} \cdot \epsilon X^\beta \,. \end{aligned} \tag{2.113}$$

Therefore,

$$\begin{aligned} g_{\mu\nu}(x)\, \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu \mapsto & g_{\mu\nu}(x)\, \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu + \epsilon X^\beta \partial_\beta g_{\mu\nu}\, \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu \\ & + \epsilon g_{\mu\nu} \left(\partial_\beta X^\mu\, \mathrm{d}x^\beta \otimes \mathrm{d}x^\nu + \partial_\beta X^\nu\, \mathrm{d}x^\mu \otimes \mathrm{d}x^\beta \right) \\ = & \left[g_{\mu\nu} + \epsilon \left(X^\beta \partial_\beta g_{\mu\nu} + g_{\beta\nu} \partial_\mu X^\beta + g_{\mu\beta} \partial_\nu X^\beta \right) \right] \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu \\ = & \left(g_{\mu\nu} + \epsilon (\mathcal{L}_X g)_{\mu\nu} \right) \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu \,. \end{aligned} \tag{2.114}$$

Therefore, for an infinitesimal displacement along a vector field X, the metric changes by $\epsilon \mathcal{L}_X g$. In particular, if $\mathcal{L}_X g = 0$, then the metric is preserved. X

is said to be a **Killing field**, and indicates an existence of an **isometry** of the manifold, a transformation of the manifold that preserves distances. You should have in your mind a field indicating how points transform if a sphere is rotated about any axis, for example, as an example of a Killing field.

(**End of Lecture**: Monday Oct 7 2024)

2.7.2 Differential forms and the wedge product

Before we can move on to the next derivative operation, we need to introduce a special class of tensors known as **differential forms**, or simply **forms**. These are simply $(0,p)$ tensors that are completely antisymmetric. Thus, a scalar is a 0-form, a co-vector is a one-form—a term we have already been consistently using—and, for example, the Levi-Civita tensor $\varepsilon_{\mu\nu\rho\sigma}$ is a 4-form.

To make differential forms starting from one-forms, we can define a new operator called the **wedge product**; we can then take two forms $\omega \in T_p^*$ and $\eta \in T_p^*$ and construct $\omega \wedge \eta$, a higher order form. Suppose in a coordinate basis, $\omega = \omega_{\mu_1\cdots\mu_p} \mathrm{d}x^{\mu_1} \otimes \cdots \otimes \mathrm{d}x^{\mu_p}$ is a p-form, and $\eta = \eta_{\mu_{p+1}\cdots\mu_{p+q}} \mathrm{d}x^{\mu_{p+1}} \otimes \cdots \otimes \mathrm{d}x^{\mu_{p+q}}$ is a q-form. Then the wedge product is a $(p+q)$-form, defined as

$$\begin{aligned}\omega \wedge \eta &= \frac{(p+q)!}{p!q!} \omega_{[\mu_1\cdots\mu_p} \eta_{\mu_{p+1}\cdots\mu_{p+q}]} \mathrm{d}x^{\mu_1} \otimes \cdots \otimes \mathrm{d}x^{\mu_{p+q}} \\ &= \frac{(p+q)!}{p!q!} \omega_{\mu_1\cdots\mu_p} \eta_{\mu_{p+1}\cdots\mu_{p+q}} \mathrm{d}x^{[\mu_1} \otimes \cdots \otimes \mathrm{d}x^{\mu_{p+q}]}\,, \end{aligned} \tag{2.115}$$

where

$$T_{[\mu_1\mu_2\cdots\mu_n]} = \frac{1}{n!}(T_{\mu_1\mu_2\cdots\mu_n} + \text{alternating sum over permutations of indices})\,. \tag{2.116}$$

So, for example, for two one-forms A and B,

$$(A \wedge B)_{\mu\nu} = 2A_{[\mu}B_{\nu]} = A_\mu B_\nu - A_\nu B_\mu\,. \tag{2.117}$$

From the definition of the wedge product, we can see that it inherits the distributivity, associativity, and commutativity with $\mathbb{R}$ from the tensor product; just to be very clear, this means that for any three forms A, B and C, and real number λ,

$$\begin{aligned} A \wedge (B + C) &= A \wedge B + A \wedge C\,, \\ (A \wedge B) \wedge C &= A \wedge (B \wedge C) = A \wedge B \wedge C\,, \\ A \wedge \lambda B &= \lambda A \wedge B = \lambda(A \wedge B)\,. \end{aligned} \tag{2.118}$$

In addition, from the definition, we can also show that for a p-form ω and a q-form η,

$$\omega \wedge \eta = (-1)^{pq} \eta \wedge \omega\,, \tag{2.119}$$

so one has to be extra careful in switching the order of forms in a wedge product! The set of all wedge products of one-forms from T_p^* forms a vector space, $T_p^* \wedge T_p^*$. You can of course perform arbitrarily many wedge products to form higher-order forms in spaces that we denote as $\bigwedge^k(T_p^*)$, which is a vector space of k-forms.

As before, once a coordinate system is chosen on the tangent space, this induces a basis on $\bigwedge^k(T_p^*)$, denoted by $\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}$. Two-forms are objects that act on a pair of vectors, and return a number. From the definition of the wedge product that the action of the basis two-forms is

$$\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu(\partial_\alpha,\partial_\beta)=\delta^\mu_\alpha\delta^\nu_\beta-\delta^\nu_\alpha\delta^\mu_\beta\,. \tag{2.120}$$

The analogous expression for a p-term would have $p!$ terms, corresponding to all the different permutations with the appropriate sign depending on the number of swaps to get to the permutation at hand.

As an example, on a 3-dimensional manifold with some choice of coordinate basis, a one-form looks like

$$A=A_\mu\mathrm{d}x^\mu=A_1\mathrm{d}x^1+A_2\mathrm{d}x^2+A_3\mathrm{d}x^3\,, \tag{2.121}$$

a two-form has the form

$$F=\frac{1}{2}F_{\mu\nu}\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu=F_{12}\mathrm{d}x^1\wedge\mathrm{d}x^2+F_{23}\mathrm{d}x^2\wedge\mathrm{d}x^3+F_{31}\mathrm{d}x^3\wedge\mathrm{d}x^1\,, \tag{2.122}$$

where $\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu=-\mathrm{d}x^\nu\wedge\mathrm{d}x^\mu$. A three-form looks like

$$\Omega=\frac{1}{3!}\Omega_{\mu\nu\sigma}\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu\wedge\mathrm{d}x^\sigma=\Omega_{123}\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3\,. \tag{2.123}$$

All of the components have to be completely antisymmetric. Once again, it is unfortunately extremely common for people to simply drop the wedge product, writing $F_{\mu\nu}\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu$ as $F_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$, leaving the reader to remember that in fact F is a two-form.

In d-dimensions, the space of p-forms is $d!/[p!(d-p)!]$ dimensional, and all p-forms with $p>d$ vanish identically; the basis of the p-form would involve the wedge product of two basis one-forms with the same index, which is necessarily zero by the antisymmetrization of the indices in the definition of the wedge product.

2.7.3 The exterior derivative

The wonderful world of forms and the antisymmetry of swapping the order of the wedge product of two one-forms essentially gives us a nice safe space for us to define yet another useful derivative-like operator. This is known as the **exterior derivative**, which takes a p-form $\omega=\omega_{\mu_1\cdots\mu_p}\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_p}$ in the coordinate basis, and produces a $(p+1)$-form $\mathrm{d}\omega$, given by

$$\begin{aligned}\mathrm{d}\omega&=(p+1)\partial_{[\mu_1}\omega_{\mu_2\cdots\mu_{p+1}]}\mathrm{d}x^{\mu_1}\otimes\cdots\otimes\mathrm{d}x^{\mu_{p+1}}\\&=(p+1)\partial_{\mu_1}\omega_{\mu_2\cdots\mu_{p+1}}\mathrm{d}x^{[\mu_1}\otimes\cdots\otimes\mathrm{d}x^{\mu_{p+1}]}\\&=\frac{p+1}{(p+1)!}p!\partial_{\mu_1}\omega_{\mu_2\cdots\mu_{p+1}}\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_{p+1}}\\&=\partial_{\mu_1}\omega_{\mu_2\cdots\mu_{p+1}}\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_{p+1}}\,.\end{aligned} \tag{2.124}$$

where in the second last step, I used the definition of the wedge product to rewrite tensor products into wedge products. The last line is typically the form that is most useful.

Let's warm up by doing some examples. Taking the exterior derivative of a zero-form, which is just a scalar field f on the manifold, gives

$$\mathrm{d}f=\partial_\mu f\,\mathrm{d}x^\mu\,, \tag{2.125}$$

which once again is the gradient. Comparing this with the Lie derivative, you might start to see a pattern here: all derivatives acting on scalar fields on the manifold must return the same thing.

For a one-form A, define $F = \mathrm{d}A$; we find

$$F = \mathrm{d}A = 2\partial_{[\mu}A_{\nu]}\mathrm{d}x^\mu \otimes \mathrm{d}x^\nu = \partial_\mu A_\nu \, \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu = \frac{1}{2}F_{\mu\nu}\, \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \,, \tag{2.126}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$; you should at this point be noticing a striking resemblance to the electromagnetic field tensor! Let's write this out explicitly in component form, assuming we're on a manifold in 3D:

$$\begin{aligned} F = \mathrm{d}A = &\left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\right)\mathrm{d}x^1 \wedge \mathrm{d}x^2 + \left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\right)\mathrm{d}x^2 \wedge \mathrm{d}x^3 \\ &+ \left(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\right)\mathrm{d}x^3 \wedge \mathrm{d}x^1 \end{aligned} \tag{2.127}$$

What you should notice here is that the exterior of the one-form A gives something that resembles the curl, $\nabla\times$ on $\mathbb{R}^3$!

Finally, suppose we have a two-form $G = G_{12}\mathrm{d}x^1 \wedge \mathrm{d}x^2 + G_{23}\mathrm{d}x^2 \wedge \mathrm{d}x^3 + G_{31}\mathrm{d}x^3 \wedge \mathrm{d}x^1$. Then

$$\mathrm{d}G = \left(\frac{\partial G_{23}}{\partial x^1} + \frac{\partial G_{31}}{\partial x^2} + \frac{\partial G_{12}}{\partial x^3}\right)\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 \,, \tag{2.128}$$

which resembles the divergence, $\nabla\cdot$ in $\mathbb{R}^3$.

How is the exterior derivative a derivative? Remember that a derivative-like operator should do two things: *1)* it should be linear, i.e. $\mathrm{d}(a\omega + b\eta) = a\,\mathrm{d}\omega + b\,\mathrm{d}\eta$, which clearly holds given that the partial derivative itself is linear, and *2)* that the operator should obey Leibniz's rule. In fact, that exterior derivative obeys a slightly modified version of Leibniz's rule: given a p-form ω and a q-form η,

$$\mathrm{d}(\omega \wedge \eta) = (\mathrm{d}\omega) \wedge \eta + (-1)^p \omega \wedge (\mathrm{d}\eta) \,, \tag{2.129}$$

the proof of which I will not show here, but can be straightforwardly derived from all the definitions above.

Let's now consider what happens when you take the exterior derivative twice. We can observe from the definition that

$$\mathrm{d}^2 A = \partial_k \partial_l A_{\mu_1 \cdots \mu_p} \mathrm{d}x^k \wedge \mathrm{d}x^l \wedge \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_k} \,. \tag{2.130}$$

However, we should be able to commute the two partial derivatives, while k and l are supposed to be antisymmetric indices in the wedge product, implying that the expression is zero. This therefore implies the following extremely important result:

$$\mathrm{d}^2 = 0 \,, \tag{2.131}$$

sometimes known as **Poincaré's lemma**. We say that a p-form A is **closed** if $\mathrm{d}A = 0$, and **exact** if $A = \mathrm{d}B$ for some $(p-1)$-form B. All exact forms are closed, but whether or not all closed forms are also exact depends on the topology of the manifold you are in; all closed p-forms where $p \geq 1$ are exact in Minkowski or real space.

What are we to make of all of this? By studying forms and the exterior derivative, we have developed machinery that is the generalization of ∇ to

arbitrary dimensions and manifolds with arbitrary coordinate basis. In fact, Poincaré's lemma itself tells us that if I take the curl of the grad, or the div of a curl, I get zero! The exterior derivative and the wedge product are the much-hoped-for generalization of vector calculus to manifolds, and are extremely powerful tools indeed.

(**End of Lecture**: Wednesday Oct 9 2024)

2.7.4 Electromagnetism

As we already alluded to earlier, the language of differential forms is particularly suited to describing electromagnetism, especially given that Maxwell's equations are basically a bunch of relations involving divs and curls! The manifold that is relevant here is Minkowski space; we'll take the metric to be $(+1,-1,-1,-1)$, and we'll work everything out in natural units, where $c=\varepsilon_0=\mu_0=1$.

The basic object of interest is the four-potential, $A^\mu=(\phi,\vec{A})$, where ϕ is the electric potential, and $\vec{A}$ is the vector potential. The one-form that we are interested in is therefore $A=A_\mu\,\mathrm{d}x^\mu$, where $A_\mu=(\phi,-\vec{A})$. As we saw earlier, I can define the two-form[27]

$$F=-\mathrm{d}A=-\frac{1}{2}F_{\mu\nu}\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu\,, \tag{2.132}$$

where $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$. Using the standard relation that

$$\vec{E}=-\nabla\phi-\frac{\partial\vec{A}}{\partial t}\,,\qquad \vec{B}=\nabla\times\vec{A}\,, \tag{2.133}$$

we can see that in fact the components of F are precisely the electromagnetic fields, e.g.

$$\begin{aligned}F_{01}&=\partial_0A_1-\partial_1A_0=-\partial_t\vec{A}_x-\partial_x\phi=\vec{E}_x\\ F_{12}&=\partial_1A_2-\partial_2A_1=-(\partial_x\vec{A}_y-\partial_y\vec{A}_x)=-\vec{B}_z\end{aligned} \tag{2.134}$$

with

$$F_{\mu\nu}=\begin{pmatrix}0&E_x&E_y&E_z\\-E_x&0&-B_z&B_y\\-E_y&B_z&0&-B_x\\-E_z&-B_y&B_x&0\end{pmatrix}\,. \tag{2.135}$$

Since F is exact, we arrive immediately at our first equation of motion:

$$\mathrm{d}F=0\,, \tag{2.136}$$

with almost no effort. You'll prove in the next problem set that this is equivalent to $\nabla\times\vec{E}=-\partial_t\vec{B}$ and $\nabla\cdot\vec{B}=0$, which are the two source-free Maxwell equations.

The other two equations take a little more work to write down in form language. We first have to define the **Hodge star operator** on a manifold of dimension n, which maps p-forms to $(n-p)$-forms via the following transformation of the coordinate basis:

$$\star\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_{n-p}}=\frac{1}{p!}\varepsilon^{\mu_1\cdots\mu_{n-p}}{}_{\nu_1\cdots\nu_p}\mathrm{d}x^{\nu_1}\wedge\cdots\wedge\mathrm{d}x^{\nu_p}\,. \tag{2.137}$$

[27] Unfortunately, the choice of the mostly minus metric and sign conventions used in the literature have culminated in this horrendous minus sign, which I wouldn't personally introduce here, but will to be consistent with e.g. Ref. [5].

where $\varepsilon_{\mu_1\cdots\mu_{n-p}\nu_1\cdots\nu_p} = \sqrt{|g|}\epsilon_{\mu_1\cdots\mu_{n-p}\nu_1\cdots\nu_p}$ is the Levi-Civita tensor. This looks scary, but in an orthonormal Cartesian or Minkowski basis, this is quite straightforward. For 2-forms on Minkowski space, it is simply the map

$$\begin{aligned}
\star(\mathrm{d}x \wedge \mathrm{d}y) &= \mathrm{d}t \wedge \mathrm{d}z \\
\star(\mathrm{d}y \wedge \mathrm{d}z) &= \mathrm{d}t \wedge \mathrm{d}x \\
\star(\mathrm{d}x \wedge \mathrm{d}z) &= -\mathrm{d}t \wedge \mathrm{d}y \\
\star(\mathrm{d}t \wedge \mathrm{d}x) &= -\mathrm{d}y \wedge \mathrm{d}z \\
\star(\mathrm{d}t \wedge \mathrm{d}y) &= \mathrm{d}x \wedge \mathrm{d}z \\
\star(\mathrm{d}t \wedge \mathrm{d}z) &= -\mathrm{d}x \wedge \mathrm{d}y
\end{aligned}$$

The rule of thumb is that for a given form ω, $\star\omega$ can be obtained by writing the components not involved in ω in an order such that $\omega \wedge (\star\omega) = \mathrm{d}t \wedge \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$, and adding an extra minus sign for every spatial basis component in ω.[28] The two Maxwell's equations with sources, $\nabla \cdot \vec{E} = \rho$ and $\nabla \times \vec{B} = \vec{j} + \partial_t \vec{E}$, can ultimately be written as

$$\mathrm{d}(\star F) = \star J\,, \tag{2.138}$$

where $J = J_\mu \mathrm{d}x^\mu$ is the one-form corresponding to the four-current, $J_\mu = (\rho, -\vec{j})$, with ρ being the charge density, and $\vec{j}$ being the current density.

On your problem set, you will also see that charge conservation and gauge transformations fall out quite naturally from using the language of differential forms.

[28] This is true in the mostly minus metric convention we are using. In the mostly plus convention, you add a minus sign for every $\mathrm{d}t$ in ω. For $\mathbb{R}^n$, there is no need to worry about extra minus signs.

2.8 Integration on Manifolds

Having seen all the added work we needed to do to define operations that look like derivatives on a manifold, you might be worried that integration might be similarly cumbersome. It turns out, however, that integration on manifolds is a much more familiar beast: once we inroduce some formalities, you'll realize soon enough that you've always been comfortable integrating on manifolds!

2.8.1 Volume forms

Integrating a function over a manifold involves breaking up the manifold into little pieces, each with a **volume**[29] and a function value, and adding up all of the contributions from the pieces. To do this, we will have to start by understanding how to define volume locally. We begin by trying to understand how to define the **volume** of k vectors in a tangent space. For two vectors $\vec{v}_1, \vec{v}_2$ in 3D space, for example, we know that we can define a volume by taking the area of the parallelogram spanned by these vectors, i.e. $|\vec{v}_1 \times \vec{v}_2|$. Similarly, in 3D space, we can take three vectors and form them into a parallelopiped, More generally, the volume ω_k is therefore something that takes k vectors and returns a real number, and so ω_k is a $(0,k)$-object that lives in $\bigotimes^k T_p^*$. In addition, we would like the following properties for the volume form:

[29] When speaking generally, area will also be regarded as a volume

1. *Scaling*: If we double the length of one of the vectors, we expect the volume to double, i.e. given k vectors $\vec{v}_1, \cdots, \vec{v}_k \in T_p$, and $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\omega_k(\lambda\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k) = \omega_k(\vec{v}_1, \lambda\vec{v}_2, \cdots, \vec{v}_k) = \cdots = \omega_k(\vec{v}_1, \vec{v}_2, \cdots, \lambda\vec{v}_k) \\
= \lambda\omega_k(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k)\,.
\end{aligned} \tag{2.139}$$

   Note that this allows for negative values for the volume.

2. *Additivity*: For an additional vector $\vec{u} \in T_p$, we also want

$$\omega_k(\vec{v}_1 + \vec{u}, \vec{v}_2, \cdots, \vec{v}_k) = \omega_k(\vec{u}, \vec{v}_2, \cdots, \vec{v}_k) + \omega_k(\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k)\,, \tag{2.140}$$

   and similarly for all the other slots.

3. *Degeneracy*: If two sides coincide, the volume should be zero, i.e. if any two vectors are equal, the form must be zero.

These properties combine to ensure that in fact, ω_k *has to be a* k*-form*. To see this, observe that

$$\begin{aligned} &\omega_k(\cdots, \vec{u} + \vec{v}, \cdots, \vec{u} + \vec{v}, \cdots) = 0 \\ &\implies \omega_k(\cdots, \vec{u}, \cdots, \vec{v}, \cdots) + \omega_k(\cdots, \vec{v}, \cdots, \vec{u}, \cdots) = 0 \\ &\implies \omega_k(\cdots, \vec{u}, \cdots, \vec{v}, \cdots) = -\omega_k(\cdots, \vec{v}, \cdots, \vec{u}, \cdots)\,, \end{aligned} \tag{2.141}$$

i.e. ω_k is *completely antisymmetric with respect any pair of its arguments*, which is exactly what a k-form is.

2.8.2 Defining an integral

So we've got a notion of volume in the tangent space of every point on the manifold, and we now want to be able to add these volumes up, just like how an integral should work. The natural setting we have in mind is to have a p-form defined in $\mathbb{R}^n$, e.g.

$$\omega = \frac{1}{p!}\omega_{\mu_1 \cdots \mu_p} \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_p}\,, \tag{2.142}$$

where x^μ is a choice of coordinate system for $\mathbb{R}^n$ (e.g. Cartesian coordinate system) and to integrate that form over some p-dimensional manifold Ω embedded in $\mathbb{R}^n$, to obtain a real number. Suppose Ω is parametrized by ξ^i, where

$$x^1 = x^1(\xi^1, \cdots, \xi^p)\,, \cdots x^d = x^d(\xi^1, \cdots, \xi^p)\,. \tag{2.143}$$

so that

$$\mathrm{d}x^\mu = \frac{\partial x^\mu}{\partial \xi^i} \mathrm{d}\xi^i \tag{2.144}$$

Then we can write, in this new parametrization,

$$\omega = \omega_{\mu_1 \mu_2 \cdots \mu_p} \frac{\partial x^{\mu_1}}{\partial \xi^1} \cdots \frac{\partial x^{\mu_p}}{\partial \xi^p} \mathrm{d}\xi^1 \wedge \cdots \wedge \mathrm{d}\xi^p\,, \tag{2.145}$$

which is a p-form on Ω.

We now define the integral of ω over the manifold Ω as

$$I = \int_\Omega \omega \equiv \int_\Omega d\xi^1 \cdots d\xi^p\, \omega_{\mu_1 \mu_2 \cdots \mu_p} \frac{\partial x^{\mu_1}}{\partial \xi^1} \cdots \frac{\partial x^{\mu_p}}{\partial \xi^p}\,, \tag{2.146}$$

where $I \in \mathbb{R}$, and the object on the right is just the regular multidimensional integral. Note that if $\xi^i = x^i$, i.e. the parametrization of Ω is the same or some subset of $\mathbb{R}^n$, then this is just a straightforward integral over the component of the form; otherwise, because $\omega_{\mu_1 \mu_2 \cdots \mu_p}$ is antisymmetric in all indices, the contraction against x^μ *is really the Jacobian that we're used to in regular*

integration. The punchline really is that *you can simply treat forms as regular differentials*, remembering that a change of coordinates requires a Jacobian.

A few examples will help. Let's consider a 1-dimensional path Γ starting at some point P_0 and ending at point P_1, and a function f defined across the path. We can integrate the 1-form $\mathrm{d}f$, if we parametrize the path as $x(s)$, for $0 \leq s \leq 1$, so that $x(0) = P_0$ and $x(1) = P_1$, we have

$$I = \int_\Gamma \mathrm{d}f = \int_0^1 ds \frac{df}{ds} = f(P_1) - f(P_0)\,. \tag{2.147}$$

Next, let's integrate the 2-form $\omega = x\,\mathrm{d}y \wedge \mathrm{d}z$ over the surface of a two-dimensional sphere S^2 of radius R. We can parametrize the sphere using polar angles, which are related to the usual $\mathbb{R}^3$ Cartesian coordinates via

$$\begin{aligned} x &= R\cos\phi\sin\theta \\ y &= R\sin\phi\sin\theta \\ z &= R\cos\theta\,. \end{aligned} \tag{2.148}$$

To perform the integral, we need the Jacobian $J \equiv |\partial x^i/\partial\xi^j|$. We have

$$\begin{aligned} \frac{\partial y}{\partial\theta} &= R\sin\phi\cos\theta\,, & \frac{\partial y}{\partial\phi} &= R\cos\phi\sin\theta \\ \frac{\partial z}{\partial\theta} &= -R\sin\theta\,, & \frac{\partial z}{\partial\phi} &= 0\,. \end{aligned} \tag{2.149}$$

Using the standard orientation with $\xi^i = \theta$ and $\xi^2 = \phi$ so that $\hat\theta \times \hat\phi$ points outward, $J = R^2\cos\phi\sin^2\theta$. Hence,

$$\begin{aligned} \int_{S^2} x\,\mathrm{d}y\wedge\mathrm{d}z &= \int_0^{2\pi} d\phi \int_0^\pi d\theta (R\cos\phi\sin\theta)(R^2\cos\phi\sin^2\theta) \\ &= \frac{4}{3}\pi R^3\,. \end{aligned} \tag{2.150}$$

We could also have noted that We also have the relation

$$\begin{aligned} \mathrm{d}y &= R\sin\phi\cos\theta\,\mathrm{d}\theta + R\cos\phi\sin\theta\,\mathrm{d}\phi \\ \mathrm{d}z &= -R\sin\theta\,\mathrm{d}\theta\,, \end{aligned} \tag{2.151}$$

so that the form on the sphere is

$$x\,\mathrm{d}y\wedge\mathrm{d}z = (R\cos\phi\sin\theta)(R^2\cos\phi\sin^2\theta)\,\mathrm{d}\theta\wedge\mathrm{d}\phi\,, \tag{2.152}$$

and then proceeding directly with the integral. This picture is perhaps much more intuitive though: you can think of the sphere as being tiled by little patches formed by intersecting longitudinal and latitudinal lines, each bordered by a vector $\vec v_\theta$ in the $\hat\theta$ direction and $\vec v_\phi$ in the $\hat\phi$ direction. The integral is then just summing up $\omega(\vec v_\theta, \vec v_\phi)$ in all of the little squares (taking the limit as the squares become infinitesimally small).

(**End of Lecture**: Tuesday Oct 15 2024)

2.8.3 Orientability

You might have noticed that we had to make a choice about the **orientation** of the 2-sphere, i.e. pick an ordering to the basis elements $\mathrm{d}\theta\wedge\mathrm{d}\phi$ (as opposed to $\mathrm{d}\phi\wedge\mathrm{d}\theta$), so that the direction of $\hat\theta\times\hat\phi$ always points outward. This is possible on the sphere, making the sphere an **orientable** manifold. I won't go into a detailed discussion of the technical definition of orientable, but the picture you should

have in mind is that the basis vectors are changing as you move from point to point, but the transformation mapping basis vectors from one point to another must always have a positive determinant for an orientable manifold. One simple example of this failing is the Möbius strip, where you can see that one of the two basis vectors transported in a loop ends up in the opposite orientation when it comes back to itself. Integration as defined here really only makes sense on orientable manifolds.

2.8.4 The volume form

Note that so far, in discussing the integration of forms, we have made no mention whatsoever of the metric, and hence we do not need any notion of length to make sense of these integrals. However, when a manifold of dimension $p < d$ is embedded in $\mathbb{R}^d$, it *inherits* a metric from the usual metric of $\mathbb{R}^d$.

For a manifold Ω, given any two vectors $\vec{v}, \vec{w} \in \Omega$; if the manifold is embedded in $\mathbb{R}^d$, then we can also view these vectors as belonging to $\mathbb{R}^d$. The induced metric $g_{\mu\nu}$ on Ω, on which we make a choice of coordinates ξ^μ on the manifold, is simply the metric that, when acting on $\vec{v}$ and $\vec{w}$, gives the same result as the usual metric $\delta_{\alpha\beta}$ on $\mathbb{R}^d$ with coordinates x^α, i.e.

$$g_{\mu\nu} v'^\mu w'^\nu = \delta_{\alpha\beta} v^\alpha w^\beta \,, \tag{2.153}$$

where v^α and w^β are the components of v and w in the usual Cartesian coordinate basis respectively, while v'^μ and v'^ν are the components in coordinates parametrizing the manifold, ξ^μ. These are of course related to v^μ and w^ν through the usual change-of-coordinates formula, i.e.

$$g_{\mu\nu} v'^\mu w'^\nu = \delta_{\alpha\beta} \frac{\partial x^\alpha}{\partial \xi^\mu} \frac{\partial x^\beta}{\partial \xi^\nu} v'^\mu w'^\nu \,, \tag{2.154}$$

i.e. the induced metric on the manifold is

$$g_{\mu\nu} = \delta_{\alpha\beta} \frac{\partial x^\alpha}{\partial \xi^\mu} \frac{\partial x^\beta}{\partial \xi^\nu} = \sum_{i=1}^{d} \frac{\partial x^i}{\partial \xi^\mu} \frac{\partial x^i}{\partial \xi^\nu} \,. \tag{2.155}$$

For a p-dimensional manifold Ω, we can define the **volume form** associated with this induced metric,

$$\mathrm{d}\omega = \sqrt{g}\, \mathrm{d}\xi^1 \wedge \cdots \wedge \mathrm{d}\xi^p \,, \tag{2.156}$$

where you can see that $\sqrt{g}$ is the Jacobian between ξ^μ and x^α, the Cartesian coordinates on $\mathbb{R}^d$. Thus, integrating this form over the Ω gives the total p-dimensional volume of Ω.

Let's look at a simple example, the unit-radius two-sphere embedded in $\mathbb{R}^3$. Choosing polar angle coordinates on the sphere, we have

$$\begin{aligned} x &= \cos\phi \sin\theta \,, \\ y &= \sin\phi \sin\theta \,, \\ z &= \cos\theta \,, \end{aligned} \tag{2.157}$$

from which we see that the induced metric has components

$$\begin{aligned} g_{\phi\phi} &= (-\sin\phi\sin\theta)^2 + (\cos\phi\sin\theta)^2 = \sin^2\theta \,, \\ g_{\phi\theta} &= (-\sin\phi\sin\theta)(\cos\phi\cos\theta) + (\cos\phi\sin\theta)(\sin\phi\cos\theta) = 0 \,, \\ g_{\theta\theta} &= (\cos\phi\cos\theta)^2 + (\sin\phi\cos\theta)^2 + (-\sin\theta)^2 = 1 \,, \end{aligned} \tag{2.158}$$

i.e. the induced metric is

$$g = \mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta\, \mathrm{d}\phi \otimes \mathrm{d}\phi\,. \tag{2.159}$$

The volume form is then

$$\sqrt{g}\, \mathrm{d}\theta \wedge \mathrm{d}\phi = \sin\theta\, \mathrm{d}\theta \wedge \mathrm{d}\phi\,, \tag{2.160}$$

which corresponds to the usual surface element on the two-sphere.

2.8.5 Stokes' Theorem

We finish our discussion of integration—and the calculus of manifolds!—with Stokes' theorem. You should be familiar with this result from 3D vector calculus: given a surface S in $\mathbb{R}^3$, with boundary Γ, for a smooth vector field $\vec{F}$ throughout $\mathbb{R}^3$ we have the relation

$$\int_S d\vec{S} \cdot (\nabla \times \vec{F}) = \oint_\Gamma d\vec{\Gamma} \cdot \vec{F}\,. \tag{2.161}$$

This allows us to move between differential and integral forms of Maxwell's equations, for example

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \implies \int_S d\vec{A} \cdot (\nabla \times \vec{E}) = -\frac{\partial}{\partial t} \int_S d\vec{A} \cdot \vec{B} = \oint_\Gamma d\vec{\ell} \cdot \vec{E}\,. \tag{2.162}$$

The generalized version of Stokes' theorem relates the integral of a exterior derivative $\mathrm{d}\omega$ of a form ω across a manifold Ω with an integral along the boundary $\partial\Omega$ of Ω:

$$\int_\Omega \mathrm{d}\omega = \int_{\partial\Omega} \omega\,. \tag{2.163}$$

This is an incredibly beautiful and useful result. We won't go through the formal proof, but the intuitive explanation for this is pretty straightforward. Notice that for a p-form,

$$\mathrm{d}(\omega_{\mu_1\cdots\mu_p}\, \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_p}) = \partial_\nu \omega_{\mu_1\cdots\mu_p}\, \mathrm{d}x^\nu \wedge \mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_p}\,. \tag{2.164}$$

Divide ω up into tiny little boxes, and perform the integral in each of these boxes. This gives you a bunch of terms that look like the sum of ω evaluated on one side of a p-dimensional box. However, the adjacent box shares one surface with the current box, with the opposite orientation, and therefore when we sum the effect of these two boxes together, the term corresponding to the shared surface cancels out. Therefore, when we sum over all of the small boxes, we are left with ω summed along the boundary of Ω.

Let's see a few simple examples:

1. If Ω is a region in $\mathbb{R}^2$, then since

$$\mathrm{d}\left[\frac{1}{2}(x\,\mathrm{d}y - y\,\mathrm{d}x)\right] = \mathrm{d}x \wedge \mathrm{d}y\,, \tag{2.165}$$

we can write

$$\text{Area}(\Omega) = \int_\Omega \mathrm{d}x \wedge \mathrm{d}y = \frac{1}{2}\int_{\partial\Omega} (x\,\mathrm{d}y - y\,\mathrm{d}x)\,, \tag{2.166}$$

which is just Green's theorem.

2. For Ω a region in $\mathbb{R}^2$, we have

$$\mathrm{d}\left[\frac{1}{2}r^2\,\mathrm{d}\theta\right] = r\,\mathrm{d}r\wedge\mathrm{d}\theta\,, \tag{2.167}$$

and so

$$\text{Area}(\Omega) = \int_\Omega r\,\mathrm{d}r\wedge\mathrm{d}\theta = \frac{1}{2}\int_{\partial\Omega} r^2\,\mathrm{d}\theta\,, \tag{2.168}$$

which are two equivalent ways of evaluating the area in polar coordinates.

3. If Ω is the interior of a sphere of radius R, then

$$\int_\Omega \mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z = \int_{\partial\Omega} x\,\mathrm{d}y\wedge\mathrm{d}z = \frac{4}{3}\pi R^3\,, \tag{2.169}$$

a result which we obtained in Eq. (2.150).

3 Complex Analysis

References: Churchill and Brown (CB) Chapters 1–7

From doing calculus on general spaces but with a focus on real numbers, we'll now turn our attention to doing calculus with functions of complex numbers. From our experience with real analysis, one may have expected the calculus of complex numbers to be as complicated, varied and sometimes pathological as the calculus of real numbers; but we'll see that the structure of the complex plane makes life very simple and pretty, making complex analysis a very powerful tool, even when you're really only interested in the real line.

3.1 Preliminaries

We'll begin with a crash course on complex numbers and how to manipulate them, which should be familiar to you from your undergraduate experience. Complex numbers can be obtained by supplementing the real numbers with an additional element i, satisfying $i^2 = -1$. It turns out that with this new element, we can still sensibly define addition, subtraction, multiplication and division; they all work in a manner very similar to what happens in $\mathbb{R}$ and also for the rational numbers $\mathbb{Q}$. The mathematical jargon here is that the complex numbers form a **field**, with $\mathbb{R}$ as a subfield. With this new element, the bigger field $\mathbb{C}$ actually behaves even more nicely than $\mathbb{R}$; for example, every polynomial of degree n with complex coefficients always has n complex roots, while it certainly isn't the case polynomials with real coefficients of degree n have n real roots. A straightforward example is $x^2 + 1$, which has no real roots, but two complex roots i and $-i$.

A general complex number z can be written as $z = x + iy$, where $x, y \in \mathbb{R}$. We call x and y the real and imaginary part of z respectively, often written as $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$.

In addition to the usual field operations, we have two other important operations acting on a general complex number $z = x + iy \in \mathbb{C}$:

1. **Complex conjugate**, written either as $z*$ or $\overline{z}$, defined as $z^* = x - iy$, and
2. **Modulus** or absolute value, written as $|z|$, and defined as $|z| = \sqrt{x^2 + y^2}$.

One useful result is that $|z|^2 = zz^*$, which is often used to divide complex numbers; for example,

$$\frac{3+4i}{3-4i} = \frac{(3-4i)^2}{(3-4i)(3+4i)} = \frac{9-24i-16}{25} = \frac{-7-24i}{25}\,. \tag{3.1}$$

Aside from writing $z = x + iy$, it is also common to write complex numbers in **polar form**,

$$z = r(\cos\theta + i\sin\theta)\,, \tag{3.2}$$

where $x = r\cos\theta$ and $y = r\sin\theta$, and $r, \theta \in \mathbb{R}$. In fact, it is often helpful to visualize the complex plane, with the real component on the x-axis and the imaginary component on the y-axis, so that complex numbers $z = x + iy$ can be visualized as points (x, y), or in polar coordinates in terms of (r, θ).

You can check that $r = |z|$. We call the value of θ the **argument** of the complex number, written as $\theta = \arg(z)$. You can see, however, that given θ as the argument of z, $\theta + 2n\pi$ for any integer n gives you the same complex number as well. We therefore define the **principal value of the argument**

$\text{Arg}(z)$, which is the unique value of θ such that $-\pi < \theta \leq \pi$. Note that $\text{Arg}(-1) = \pi$.

Euler's formula, which we will study in more detail later, is given by

$$e^{i\theta} = \cos\theta + i\sin\theta\,, \tag{3.3}$$

and so we can also write $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$, which we call the **exponential form**. Switching between forms can help make algebra easier. In particular, for $z = x + iy$,

$$z^* = x - iy = r(\cos\theta - i\sin\theta) = re^{-i\theta}\,. \tag{3.4}$$

Euler's formula also directly leads to **de Moivre's formula**, which reads

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta\,, \tag{3.5}$$

which you can check by writing $(e^{i\theta})^n = e^{in\theta}$.

One particularly important operation is to find the nth roots of any complex number $z_0 = r_0 e^{i\theta_0}$, which by definition are complex numbers $z = re^{i\theta}$ such that $z^n = z_0$. This means that we require $r^n = r_0$, and $n\theta = \theta_0 + 2k\pi$ for some $k \in \mathbb{Z}$, since adding an additional 2π to the argument doesn't change the complex number at all. We can see here that there are n unique values of θ, namely

$$\theta = \frac{\theta_0 + 2k\pi}{n}\,, \quad k = 0, \cdots, n-1\,. \tag{3.6}$$

Therefore the nth roots of z_0 are

$$z = r_0^{1/n} \exp\left[i\left(\frac{\theta_0}{n} + \frac{2k\pi}{n}\right)\right]\,, \quad k = 0, \cdots, n-1\,. \tag{3.7}$$

(**End of Lecture**: Wednesday Oct 16 2024)

3.2 Complex Functions and Derivatives

We next move on to studying functions of one complex variable $f(z)$, which in general maps a complex number to another complex number on some subset of $\mathbb{C}$. In a rigorous study of complex analysis, we would make sure to define things like continuity and taking limits rigorously, but as this is a physics class, we will just rely on the intuitive definitions of continuity and limit; you can read more about robust definitions in CB15–19. At this point, some functions that you can easily define are things like polynomials, e.g. $f(z) = z^2$, as well as familiar functions like exponential and trigonometric functions, which all can be generalized onto the complex plane (more on that later).

3.2.1 Complex derivative

We will however define the **complex derivative** carefully. Let f be a function defined over some small neighborhood around a point z_0. Then the derivative of f and z_0 is the limit

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}, \tag{3.8}$$

and the function f is said to be **differentiable** at z_0 when $f'(z_0)$ exists. This definition should come as no surprise if you're familiar with the derivative over $\mathbb{R}$. However, some thought should now convince you that this definition is quite a bit *stronger* than that in $\mathbb{R}$; the limit that you obtain has to be the same *regardless of which direction you approach z_0 from*, as otherwise the limit cannot be defined.

Take for example the function $f(z) = |z|^2$, which intuitively looks like a nice function over $\mathbb{R}^2$; on the complex plane, we find that $f(z)$ is in fact not differentiable anywhere except at $z = 0$! You can see this by picking a point x_0 on the real axis, and trying to find the limit when 1) you approach along the real axis, where the result of the limit is $2x_0$, and 2) you approach along the purely imaginary direction, where the result of the limit is 0. *Complex differentiability is a much more stringent requirement*.

Assuming that the derivative does exist, all of our intuition from calculus on $\mathbb{R}$ carries over. So, for example, we have:

1. The usual differentiation formula for a polynomial,

$$\frac{d}{dz} z^n = n z^{n-1}. \tag{3.9}$$

2. Complex differentiation is additive, i.e.

$$\frac{d}{dz}[f(z) + g(z)] = f'(z) + g'(z). \tag{3.10}$$

3. We have the product rule, i.e.

$$\frac{d}{dz}[f(z)g(z)] = f(z)g'(z) + f'(z)g(z). \tag{3.11}$$

4. Finally, we have the chain rule: suppose that f has a derivative at z_0, and g has a derivative at the point $f(z_0)$. Then the function $F(z) = g[f(z)]$ has a derivative at z_0, and

$$F'(z_0) = g'[f(z_0)]f'(z_0). \tag{3.12}$$

Thus, everything we know and love about real analysis carries over neatly here!

3.2.2 Holomorphic functions and the Cauchy-Riemann equations

At this point, you might still be really uncomfortable with taking complex derivatives, because you've got to be real careful about whether a given function is differentiable or not. In this part, we're going to develop some nice necessary and sufficient conditions for a complex function to be differentiable. Let me just remind you of what "necessary" and "sufficient" mean:

- The statement "A is necessary for B" means that A has to be true in order for B to be true, or in other words, if B is true, A must also be true. This is sometimes written as $B \implies A$. So, for example, "passing your ACE exams is necessary for obtaining your PhD"; if you successfully obtained your PhD, you must have passed your ACE exams, because it was a necessary condition.

- The statement "A is sufficient for B" means that if A is true, B is true as well. This is sometimes written as $A \implies B$. So, for example, "getting a 70 on this course is sufficient to pass", and so if you do get a 70, that's sufficient for you to pass; but you might also have passed with a 60.

Let's examine more closely how derivatives work on the complex plane, and derive a *necessary* condition for a derivative to exist. Any complex function $f(z)$ can be written as a pair of component functions u and v, such that for any $z = x + iy$,

$$f(z) \equiv u(x,y) + iv(x,y)\,, \tag{3.13}$$

where u, v take in a pair of real numbers, and return a real number. Suppose the derivative $f'(z_0)$ exists at the point $z_0 = x_0 + iy_0$. Then

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}\,. \tag{3.14}$$

This should apply as I approach z_0 from any direction on the complex plane, so let's start by approaching z_0 along a direction parallel to the real axis. We can write $(x, y) = (x_0 + \Delta x, y_0)$, so that going along the real direction, we find

$$\begin{aligned} f'(z_0) &= \lim_{\Delta x \to 0} \frac{u(x_0 + \Delta x, y_0) - u(x_0, y_0) + iv(x_0 + \Delta x, y_0) - iv(x_0, y_0)}{x_0 + \Delta x - x_0} \\ &= \lim_{\Delta x \to 0} \frac{u(x_0 + \Delta x, y_0) - u(x_0, y_0)}{\Delta x} + \lim_{\Delta x \to 0} i\frac{v(x_0 + \Delta x, y_0) - v(x_0, y_0)}{\Delta x} \\ &= \partial_x u + i\partial_x v\,, \end{aligned} \tag{3.15}$$

where I have just used the usual definition of the partial derivative of a function of two real numbers. On the other hand, I can also approach z_0 parallel to the imaginary axis. In this case, we should write $(x, y) = (x_0, y_0 + \Delta y)$, so that going along the imaginary direction, we find

$$\begin{aligned} f'(z_0) &= \lim_{\Delta y \to 0} \frac{u(x_0, y_0 + \Delta y) - u(x_0, y_0) + v(x_0, y_0 + \Delta y) - v(x_0, y_0)}{i(y_0 + \Delta y) - iy_0} \\ &= \lim_{\Delta y \to 0} \frac{u(x_0, y_0 + \Delta y) - u(x_0, y_0)}{i\Delta y} + \lim_{\Delta y \to 0} i\frac{v(x_0, y_0 + \Delta y) - v(x_0, y_0)}{i\Delta y} \\ &= -i\partial_y u + \partial_y v \end{aligned} \tag{3.16}$$

Comparing the two expressions we just obtained, we find

$$\partial_x u = \partial_y v\,, \qquad \partial_y u = -\partial_x v\,, \tag{3.17}$$

where all functions are to be evaluated at (x_0, y_0). These are known as the **Cauchy-Riemann equations**. Once again, these are *necessary* conditions for the existence of a derivative: if they are not satisfied, then the derivative does not exist, but if they are satisfied, it does *not* imply that the derivative must therefore exist. Furthermore, one can see that

$$f'(z_0) = \partial_x u + i\partial_x v\,, \tag{3.18}$$

when u and v are evaluated at (x_0, y_0).

What about *sufficient* conditions? In addition to satisfying the Cauchy-Riemann equations, for $f'(z_0)$ to exist, *1)* the first order partial derivatives must exist everywhere in some neighborhood of z_0, and *2)* these partial derivatives

are continuous at (x_0, y_0). I won't prove this statement, since it is not very essential in physics, but you can read about it in CB22.

Let's try a simple example. Consider the function $f(z) = z^*$. We can write $f(z) = u(x,y) + iv(x,y)$, where $u(x,y) = x$, and $v(x,y) = -y$. From this, we see that $\partial_x u = 1$, but $\partial_y v = -1$, and therefore $\partial_x u \neq \partial_y v$. We can thus conclude that $f(z) = z^*$ is not differentiable anywhere.

Now for some terms that can thrown around a lot: a function f is called **holomorphic** in some region of $\mathbb{C}$ if it has a complex derivative everywhere in the region. Sometimes people use the word **analytic** instead; technically, analytic means something else that we will discuss later on the course, but fortunately *all holomorphic functions are analytic and vice-versa in complex analysis, so to all intents and purposes, they are synonyms for complex functions*, and we won't worry about this distinction any further. A function that is holomorphic on $\mathbb{C}$ is an **entire** function; for example, any polynomial is an entire function.

Frequently, a function f is not holomorphic at a point z_0, but is holomorphic at points arbitrarily close to z_0; for example, $f(z) = 1/z$ is holomorphic everywhere except at $z = 0$. Such points z_0 are called **singularities**. Lots more to come about singularities ahead.

3.2.3 Harmonic functions

A **harmonic function** $H(x,y)$ is a real-valued function of two real variables x and y which satisfies **Laplace's equation**,

$$(\partial_x^2 + \partial_y^2)H = 0\,. \tag{3.19}$$

Laplace's equation comes up frequently in physics; for example, the electrostatic and gravitational potentials in vacuum must satisfy Laplace's equation. Complex analysis turns out to be very useful for solving Laplace's equation, due to the following relation:

> If a function $f(z) = u(x,y) + iv(x,y)$ is analytic in some domain D, then u and v are harmonic in D.

Let's see why this is true. Since f is analytic in D, the Cauchy-Riemann equations must be satisfied, i.e.

$$\partial_x u = \partial_y v\,, \qquad \partial_y u = -\partial_x v\,. \tag{3.20}$$

Let's take the x-derivative of the first equation, and the y-derivative of the second equation. This gives

$$\partial_x^2 u = \partial_x \partial_y v\,, \qquad \partial_y^2 u = -\partial_y \partial_x v\,. \tag{3.21}$$

However, we can commute the partial derivatives,[30] from which we find

$$(\partial_x^2 + \partial_y^2)u = 0\,, \tag{3.22}$$

i.e. u is harmonic. By taking the y-derivative of the first Cauchy-Riemann equation, and taking the x-derivative of the second one, we can also see that $(\partial_x^2 + \partial_y^2)v = 0$, thus proving the relation above.

Suppose we have two functions u and v that are both solutions to Laplace's equation. We say that v is the **harmonic conjugate** of u if they also satisfy the Cauchy-Riemann equations, i.e. $\partial_x u = \partial_y v$ and $\partial_y u = -\partial_x v$. Confusingly, v being the harmonic conjugate of u doesn't mean that u is the harmonic conjugate of v, so be careful! This gives us powerful way of checking if a function is analytic:

[30] This is not just us being sloppy as physicists: since f is analytic, we must have the partial derivatives of u and v being continuous everywhere in D, which implies mathematically that the partial derivatives do indeed commute.

A function $f(z) = u(x,y)+iv(x,y)$ is analytic in a domain D if and only if v is a harmonic conjugate of u, i.e. v being the harmonic conjugate of u is a necessary *and* sufficient condition for $f(z)$ to be analytic.

Let's see why this is so. In one direction, this is easy: f being analytic implies that u and v satisfy the Cauchy-Riemann equations, and are also harmonic. On the other hand, u and v being harmonic conjugates implies that they satisfy the Cauchy-Riemann equations, and also that u and v have well-defined partial derivatives in D that are continuous everywhere. This therefore satisfies the sufficient condition for f to be analytic that we discussed in Sec. 3.2.2.

What is the big-picture takeaway here? For a complex function $f(z) = u(x,y) + iv(x,y)$ to be analytic, the component functions u and v face very strong constraints: they satisfy the Cauchy-Riemann equations, and are harmonic. The result of this is that analytic complex functions are extremely well-behaved, and that is going to lead to all kinds of very pretty results as we go forward.

3.2.4 Elementary functions

Let's take a look at some functions that you're already familiar with from real space, and see how they work on the complex plane. Most of the properties of these functions that you're used to in real space carries over to the complex plane; I'll only highlight things that work differently between $\mathbb{R}$ and $\mathbb{C}$.

The **exponential function** e^z for $z = x + iy$ is simply given by

$$e^z = e^x e^{iy} = e^x(\cos y + i \sin y)\,. \tag{3.23}$$

One important property of e^z that we should note is that $e^z \neq 0$ for any $z \in \mathbb{C}$, since $|e^z| = e^x > 0$ for $z = x + iy$. We say that e^z has no **zeros**, where a zero z_0 of a function $f(z)$ is defined as $f(z_0) = 0$. In addition, unlike the exponential function on the real line, e^z is actually *periodic*, since

$$e^{z+2\pi i} = e^z e^{2\pi i} = e^z\,. \tag{3.24}$$

Furthermore, even though $e^z \neq 0$, we still can have $e^z < 0$, e.g.

$$e^{i\pi} = \cos\pi + i\sin\pi = -1\,. \tag{3.25}$$

Finally, e^z is an *entire function*, with

$$\frac{d}{dz}e^z = e^z \tag{3.26}$$

everywhere on the complex plane. Otherwise, all of our intuition on real space about the exponential carries over to $\mathbb{C}$.

Next, we move on to the **trigonometric functions**. From Euler's formula, we have $e^{i\theta} = \cos\theta + i\sin\theta$, from which we can see that for a real number θ, $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$, and $\sin\theta = (e^{i\theta} - e^{-i\theta})/(2i)$. We can therefore generalize this and define the following complex trigonometric functions:

$$\sin z \equiv \frac{e^{iz} - e^{-iz}}{2i}\,, \qquad \cos z \equiv \frac{e^{iz} + e^{-iz}}{2}\,. \tag{3.27}$$

Once again, all of the usual properties in trigonometry and calculus generalize to the complex plane, which you can check from the definition. One thing to note from the definition of $\sin z$ and $\cos z$ is that for $y \in \mathbb{R}$,

$$\sin(iy) = i\sinh y\,, \qquad \cos(iy) = \cosh y\,. \tag{3.28}$$

From these definitions, we can go on to define

$$\tan z \equiv \frac{\sin z}{\cos z}\,, \qquad \cot z \equiv \frac{\cos z}{\sin z}\,,$$
$$\sec z \equiv \frac{1}{\cos z}\,, \qquad \csc z \equiv \frac{1}{\sin z}\,. \tag{3.29}$$

Similarly, we have the **hyperbolic functions**, defined as

$$\sinh z = \frac{e^z - e^{-z}}{2}\,, \quad \cosh z = \frac{e^z + e^{-z}}{2}\,. \tag{3.30}$$

Nothing really surprising happens in the generalization here.

We next turn our attention to the **logarithmic function**, which requires more careful treatment. Let's write a general complex number z in polar form, i.e. $z = re^{i\theta}$. Then we define

$$\log z \equiv \ln r + i(\theta + 2n\pi)\,, n \in \mathbb{Z}\,, \tag{3.31}$$

where we use $\ln$ to denote the real-valued natural logarithm, to distinguish it from the complex logarithm $\log$. Note that $\log$ defined in this way is *multivalued*. Importantly, since for any $z = x + iy$, $e^z = e^x e^{iy}$, we have

$$\log e^z = x + i(y + 2n\pi) = z + 2in\pi\,, n \in \mathbb{Z}\,. \tag{3.32}$$

With this definition, we can also define a **complex exponent**, with the function z^c for $c \in \mathbb{C}$ being defined as

$$z^c \equiv e^{c \log z}\,. \tag{3.33}$$

Note that this means that the complex exponent is a multi-valued function as well in general, so for example

$$i^{-2i} = e^{-2i \log i} = e^{-2i \cdot i(\pi/2 + 2n\pi)} = e^{(4n+1)\pi}\,, n \in \mathbb{Z}\,. \tag{3.34}$$

Therefore, i^{-2i} is multivalued, and all of these values are actually real numbers!

3.2.5 Branches

Let's consider the multi-valued logarithm function again, which for $z = re^{i\theta}$ is $\log z = \ln r + i\theta$, where $\theta = \Theta + 2n\pi$, with Θ being the principal value of the argument. To make this single-valued, we can choose any real number α and restrict the value of θ so that $\alpha < \theta \leq \alpha + 2\pi$. In this restricted domain, $\log(z)$ becomes *single-valued*, and in fact is analytic for $r > 0$, $\alpha < \theta < \alpha + 2\pi$.[31] Fig. 12 shows an illustration of what we've done, which is to define what's known as a **branch** of the $\log$ function, with the $\theta = \alpha$ line on the complex plane being called a **branch cut**. The origin in this case is known as the **branch point**. The function *can* be defined on the branch cut; however, the domain over which the function is continuous and even analytic must include it.

[31] Note that we cannot include $\theta = \alpha$ here, as the function has a discontinuity at that point; a little bit off the $\theta = \alpha$ line, we have complex numbers with arguments close to α, while on the other side, we have numbers with arguments close to $\alpha + 2\pi$.

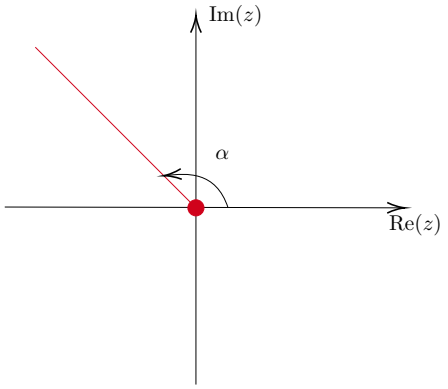

Figure 12: A branch cut on the complex plane for $\log$, restricting the values of θ to $\alpha < \theta < \alpha + 2\pi$.

The special case where we choose $\alpha = -\pi$, so that the branch cut is along the negative real axis, and $-\pi < \theta \leq \pi$ corresponds to the *principal value of the argument*, defines **the principal value of the logarithm**,

$$\mathrm{Log}(z) = \ln r + i\Theta\,, \tag{3.35}$$

written with a capital 'L'; it is single-valued, and well-defined whenever $z \neq 0$. This is also known as the **principal branch**. Thus for example

$$\begin{aligned} &\log(-1) = i(\pi + 2n\pi) = (2n+1)\pi i\,, n \in \mathbb{Z}\,, \\ &\mathrm{Log}(-1) = i\pi\,. \end{aligned} \tag{3.36}$$

Note that while the usual rules of how logarithms work carry over for $\log$, many of these rules don't work for Log. For example, if $z_1 = z_2 = -1$, then

$$\mathrm{Log}(z_1 z_2) = 0\,, \tag{3.37}$$

but

$$\mathrm{Log}(z_1) + \mathrm{Log}(z_2) = i\pi + i\pi = 2\pi i\,, \tag{3.38}$$

i.e. $\mathrm{Log}(z_1 z_2) \neq \mathrm{Log}(z_1) + \mathrm{Log}(z_2)$.

(**End of Lecture**: Monday Oct 21 2024)

3.3 Integration Part I

For the rest of the course, the main focus will be on integration on the complex plane, which is a fascinating and beautiful area of mathematics, and also extremely powerful, mostly due to the restricted structure of differentiable functions on the complex plane.

Let's begin with integration along the real line.

3.3.1 Integration along the real line

Suppose we have some function $w(t)$ which is a complex valued function, but takes as an argument a real variable t; you can imagine that this is just a general complex function restricted to the real axis. Then we can write $w(t) = u(t) + iv(t)$ for two real-valued functions u and v. The definite integral of $w(t)$ can then be straightforwardly defined as

$$\int_a^b dt\, w(t) = \int_a^b dt\, u(t) + i \int_a^b dt\, v(t)\,, \tag{3.39}$$

with the integral behaving in the usual ways that an integral would behave, e.g.

$$\int_a^b dt\, w(t) = \int_a^c dt\, w(t) + \int_c^b dt\, w(t)\,. \tag{3.40}$$

You also get the fundamental theorem of calculus here, which says that

$$\int_a^b dt\, w(t) = W(b) - W(a)\,, \qquad W'(t) = w(t)\,. \tag{3.41}$$

So, for example, we have

$$\int_0^{\pi/2} dt\, e^{it} = -i\, e^{it}\big|_0^{\pi/2} = 1 - (-i) = 1 + i\,. \tag{3.42}$$

3.3.2 Contour integrals

We're looking at functions living on the complex plane though, so we should really be able to define integrals on the complex plane. We can generalize integrating along the real axis to an integral over any **contour** C in the complex plane, written as the **contour integral**

$$\int_C dz\, f(z)\,. \tag{3.43}$$

How would we go about defining this integral? Well, this is really the same as how you would do a line integral on a flat surface, which you already have some experience with. We try to find a parametrization of the complex numbers $z(t)$ along C, so that as t varies in some range $a \le t \le b$, $z(t)$ takes us along the contour. Then, following the usual change of variables formula,

$$\int_C dz\, f(z) = \int_a^b dt \frac{dz}{dt} f[z(t)]\,, \tag{3.44}$$

converting the contour integral to just an integral along one real variable, which we now know how to do. Note that the contour is *oriented*, i.e. there is a direction associated with it. If you choose to perform the integral along the contour $-C$, i.e. C in the opposite sense (starting at $z(b)$ and heading to $z(a)$ in the previous parametrization), then you can see quite easily that

$$\int_{-C} dz\, f(z) = -\int_C dz\, f(z)\,. \tag{3.45}$$

You are also free to split a contour up into parts, and sum the separate contributions. This is all very similar to line integrals or even integrals on the real line so far. Now let's try a few examples!

1. Let's evaluate the integral

$$I = \int_C dz\, z^*\,. \tag{3.46}$$

where C is the contour

$$z = 2e^{i\theta}\,, \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}\,, \tag{3.47}$$

shown in Fig. 13. Along the contour, we have $z^* = 2e^{-i\theta}$. We can transform the integral over the contour to just an integral over θ, which is a real variable, and so

$$\begin{aligned}\int_C dz\, z^* &= \int_{-\pi/2}^{\pi/2} d\theta\, z^*(\theta) \frac{dz}{d\theta} \\ &= \int_{-\pi/2}^{\pi/2} d\theta\, (2e^{-i\theta}) \cdot 2ie^{i\theta} \\ &= 4i\left(\frac{\pi}{2} + \frac{\pi}{2}\right) \\ &= 4\pi i\,.\end{aligned} \tag{3.48}$$

2. Consider an arbitrary smooth contour C, which can be parametrized as $z = z(t)$, where $a \le t \le b$, from a fixed point $z_1 = z(t = a)$ to a fixed point $z_2 = z(t = b)$. Now let's consider the integral

$$I = \int_C dz\, z \tag{3.49}$$

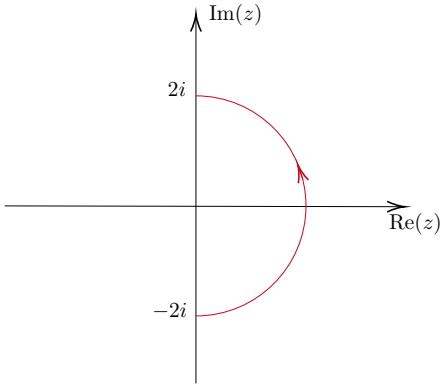

Figure 13: The contour corresponding to $z = 2e^{i\theta}$.

for this arbitrary contour. I can always write

$$\begin{aligned}\int_C dz\, z &= \int_a^b dt\, z(t) z'(t) \\ &= \frac{1}{2}\int_a^b dt\, \frac{d}{dt}(z^2(t)) \\ &= \frac{1}{2}z^2(b) - \frac{1}{2}z^2(a) \\ &= \frac{1}{2}\left(z_2^2 - z_1^2\right) .\end{aligned} \tag{3.50}$$

Notice that this integral is independent of the actual contour, and only depends on the endpoints; and is very reminiscent of line integrals of conservative forces. In particular, if you perform the integral over a closed contour, then the result is simply zero.

3. Consider the semicircular path

$$z = 3e^{i\theta}\,, 0 \le \theta \le \pi \tag{3.51}$$

from the point $z = 3$ to $z = -3$. Along this path, we want to integrate the following branch of $z^{1/2}$, defined as

$$f(z) = z^{1/2} = \exp\left(\frac{1}{2}\log z\right)\,, 0 \le \arg(z) < 2\pi\,. \tag{3.52}$$

This corresponds to a branch cut taken along the positive real axis; the contour together with this branch cut is shown in Fig. 14. There's no issue with the integral starting on the branch cut. First, we note that

$$\begin{aligned} f(3e^{i\theta}) &= \exp\left(\frac{1}{2}\log(3e^{i\theta})\right) \\ &= \exp\left[\frac{1}{2}(\ln 3 + i\theta)\right] \\ &= \sqrt{3}e^{i\theta/2}\,. \end{aligned} \tag{3.53}$$

Therefore, the integral we want to compute is

$$\begin{aligned}\int_C dz\, z^{1/2} &= \int_0^\pi d\theta\, \sqrt{3}e^{i\theta/2}\cdot 3ie^{i\theta} \\ &= 3\sqrt{3}i\int_0^\pi d\theta\, e^{i3\theta/2} \\ &= 3\sqrt{3}i\cdot\left.\frac{e^{i3\theta/2}}{3i/2}\right|_0^\pi \\ &= 2\sqrt{3}\left(e^{i3\pi/2} - 1\right) \\ &= -2\sqrt{3}(1+i)\,.\end{aligned} \tag{3.54}$$

(End of Material for Midterm 2)

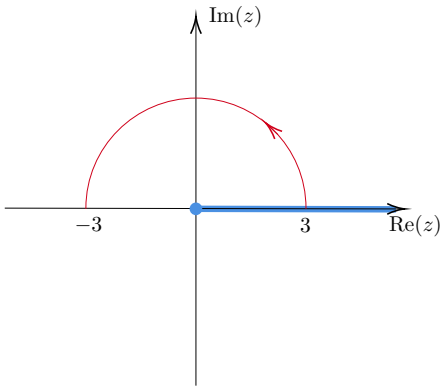

Figure 14: Contour $z = 3e^{i\theta}$ for the integral over $\sqrt{z}$, choosing the branch where $0 < \arg(z) < 2\pi$.

3.3.3 The modulus inequality

At this point, it seems like we know how to evaluate contour integrals, so we should just go out and do a bunch of them, right? Well, *it gets even better than that:* analytic functions on the complex plane are so well-behaved that there are

some surprisingly powerful theorems ahead of us. For now though, we'll take a little detour and prove some important results related to integrals that will come up repeatedly.

The first result relates the modulus of an integral to the integral of a modulus, and will be particularly useful when we solve improper integrals on the real line later on. In particular,

If $w(t)$ is a complex-valued function defined on an interval $a \leq t \leq b$, then

$$\left| \int_a^b dt\, w(t) \right| \leq \int_a^b dt\, |w(t)| \,. \tag{3.55}$$

The proof is rather straightforward, so we'll go through it: let the result of the integral of $w(t)$ over the interval be

$$\int_a^b dt\, w(t) = r_0 e^{i\theta_0} \,, \tag{3.56}$$

where the integral results in some arbitrary complex number determined by r_0 and θ_0. With this definition,

$$\left| \int_a^b dt\, w(t) \right| = r_0 \,. \tag{3.57}$$

On the other hand, we also have

$$r_0 = \int_a^b dt\, e^{-i\theta_0} w(t) \tag{3.58}$$

Since r_0 is just a real number, however, we can write the following:

$$r_0 = \text{Re} \int_a^b dt\, e^{-i\theta_0} w(t) = \int_a^b dt\, \text{Re}[e^{-i\theta_0} w(t)] \,. \tag{3.59}$$

But we also know that

$$\text{Re}[e^{-i\theta_0} w(t)] \leq \left| e^{-i\theta_0} w(t) \right| = \left| e^{-i\theta_0} \right| |w(t)| = |w(t)| \,. \tag{3.60}$$

Putting everything together proves the statement above.

This leads immediately to another important result:

Let C denote a contour of length L, and suppose that a function $f(z)$ is piecewise continuous on C. Then there exists M a nonnegative constant such that

$$|f(z)| \leq M \tag{3.61}$$

for all points z on C, and therefore by the previous result,

$$\left| \int_C dz\, f(z) \right| \leq ML \,. \tag{3.62}$$

This is a really powerful result for taking contours to infinity. Here is an example: consider the semicircular path C given by $z = Re^{i\theta}$, $0 \leq \theta \leq \pi$, and $z^{1/2}$ denotes the branch

$$z^{1/2} = \sqrt{r} e^{i\theta/2} \,,\, r > 0 \,,\, -\frac{\pi}{2} < \theta < \frac{3\pi}{2} \,. \tag{3.63}$$

Here, the branch cut is placed along the negative imaginary axis. Let's consider the integral

$$\lim_{R\to\infty}\int_C dz\,\frac{z^{1/2}}{z^2+1}\,. \tag{3.64}$$

First, we note that

$$\begin{aligned}\left|\frac{z^{1/2}}{z^2+1}\right| &= \frac{|z^{1/2}|}{|z^2+1|}\\ &= \frac{\sqrt{R}}{|R^2e^{i2\theta}+1|}\end{aligned} \tag{3.65}$$

But

$$\begin{aligned}|R^2e^{i2\theta}+1| &= \sqrt{(1+R^2\cos 2\theta)^2+(R^2\sin 2\theta)^2}\\ &= \sqrt{1+2R^2\cos 2\theta+R^4}\\ &\geq R^2-1\,,\end{aligned} \tag{3.66}$$

where we've taken $R>1$ since we are going to take $R\to\infty$. This means that

$$\left|\frac{z^{1/2}}{z^2+1}\right| \leq \frac{\sqrt{R}}{R^2-1}\,, \tag{3.67}$$

and that therefore

$$\left|\int_C dz\,\frac{z^{1/2}}{z^2+1}\right| \leq \pi R\cdot\frac{\sqrt{R}}{R^2-1} \tag{3.68}$$

This shows that

$$\lim_{R\to\infty}\left|\int_C dz\,\frac{z^{1/2}}{z^2+1}\right| = 0\,, \tag{3.69}$$

and therefore the integral itself also becomes zero in the limit.

3.3.4 Antiderivatives

Thinking back to the example of integrating a function over an arbitrary contour in Sec. 3.3.2, we should already be able to see that something like the fundamental theorem of calculus in real analysis should carry over into the complex plane. Here is the theorem:[32]

Suppose that a function $f(z)$ is continuous on a domain D. Then the following statements are all equivalent:

1. $f(z)$ has an antiderivative $F(z)$ throughout D, i.e. $F'(z)=f(z)$;
2. The integrals of $f(z)$ along contours lying entirely in D and extending from any fixed point z_1 to any fixed point z_2 all have the same value, namely
$$\int_{z_1}^{z_2} dz\,f(z) = F(z_2)-F(z_1)\,, \tag{3.70}$$
where $F'(z)=f(z)$;
3. The integrals of $f(z)$ around closed contours lying entirely in D all have value zero.

[32] Whenever the integral only depends on the endpoints and not on the particular contour, we sometimes will write the integral as having complex numbers as the upper and lower limits.

The fact that Statement 1) implies Statement 2) is easy to see, and is very similar to the earlier example: along any contour, we can parametrize the contour between z_1 and z_2 by $a \leq t \leq b$, and see that

$$\begin{aligned}\int_C dz\, f(z) &= \int_a^b dt\, f[z(t)]z'(t) \\ &= \int dt\, \frac{dF}{dz}\frac{dz}{dt} \\ &= F(z(b)) - F(z(a)) \\ &= F(z_2) - F(z_1)\,. \end{aligned} \tag{3.71}$$

Statement 2) implying Statement 3) is easy too: since contour integration only depends on the end points, any closed contour is a sum of two contour integrals in opposite directions, and hence has to be zero. The hard part of the proof is that Statement 3) implies Statement 1) as well; I won't show you the proof here, but you can consult CB45.

Let's do an example where try to perform the integral

$$I = \int_C \frac{dz}{z}\,, \tag{3.72}$$

where C is a circle of radius 2. Let's evaluate this integral using our knowledge of antiderivatives. We know that $d(\log z)/dz = 1/z$ with an appropriate choice of a branch, so that along the contour of interest this relation holds. First, let's start by considering C_1, the semicircle going from $-\pi/2 \leq \theta \leq \pi/2$ in the counterclockwise direction. For this, we choose the principal branch of the $\log$, i.e.

$$\mathrm{Log}\, z = \ln r + i\Theta\,,\; r > 0\,,\; -\pi < \Theta < \pi\,. \tag{3.73}$$

Then, since we know the antiderivative, we can immediately obtain

$$\begin{aligned}\int_{C_1} \frac{dz}{z} = \int_{-2i}^{2i} \frac{dz}{z} &= \mathrm{Log}\,(2i) - \mathrm{Log}\,(-2i) \\ &= \ln 2 + \frac{i\pi}{2} - \ln 2 + \frac{i\pi}{2} \\ &= i\pi\,. \end{aligned} \tag{3.74}$$

Next, we consider C_2, the semicircle going from $\pi/2 \leq \theta \leq 3\pi/2$ in the counterclockwise direction. This time, we'll choose another branch of the $\log$,

$$\log z = \ln r + i\theta\,,\; r > 0\,,\; 0 < \Theta < 2\pi\,, \tag{3.75}$$

so that the branch cut lies on the positive real axis. Doing the integral, we find

$$\begin{aligned}\int_{C_2} \frac{dz}{z} = \int_{2i}^{-2i} \frac{dz}{z} &= \log(-2i) - \log(2i) \\ &= \ln 2 + \frac{3\pi i}{2} - \ln 2 - \frac{i\pi}{2} \\ &= i\pi\,. \end{aligned} \tag{3.76}$$

Thus,

$$\int_C \frac{dz}{z} = \int_{C_1} \frac{dz}{z} + \int_{C_2} \frac{dz}{z} = 2\pi i\,. \tag{3.77}$$

(**End of Lecture**: Wednesday Oct 23 2024)

3.3.5 Cauchy-Goursat theorem

We now come to one of the most remarkable result in complex analysis, known as the **Cauchy-Goursat Theorem**:

If a function f is analytic at all points interior to and on a closed contour C, then

$$\int_C dz\, f(z) = 0\,. \tag{3.78}$$

This is the result of the strong constraints on analytic functions that we have discussed so far, in particular the requirement that the Cauchy-Riemann equations are satisfied. For a full proof of the theorem, I will refer you to CB47; however, it's worth motivating where this result comes from. As we have seen, the contour integral can be written as

$$\int_C dz\, f(z) = \int_a^b dt\, f[z(t)]z'(t)\,, \tag{3.79}$$

where we will now once again write $z = x + iy$, $f(z) = u(x,y) + iv(x,y)$, and $z'(t) = x'(t) + iy'(t)$. With this notation, we can write

$$\int_C dz\, f(z) = \int_a^b dt\,(ux' - vy') + i\int_a^b dt\,(vx' + uy')\,. \tag{3.80}$$

On the complex plane, however, we can view the contour integral as a line integral, with line element given by

$$d\vec{\ell} = (x'\hat{x} + y'\hat{y})dt\,, \tag{3.81}$$

where $\hat{x}, \hat{y}$ are the unit vectors in the real and imaginary directions respectively. Therefore, we can define $\vec{A} = u\hat{x} - v\hat{y}$ and $\vec{B} = v\hat{x} + u\hat{y}$, and rewrite

$$\int_C dz\, f(z) = \oint_C d\vec{\ell}\cdot\vec{A} + i\oint_C d\vec{\ell}\cdot\vec{B}\,, \tag{3.82}$$

At this point, we can think about these integrals as simply being a line integral *in three dimensional real space*. Stokes' theorem now tells us

$$\oint_C d\vec{\ell}\cdot A = \int_S d\vec{S}\cdot(\nabla\times\vec{A})\,, \tag{3.83}$$

where we are now integrating over the area S enclosed by the contour C, and the vector $d\vec{S}$ is the normal vector given by the right-hand rule applied to the contour C. On the other hand,

$$\nabla\times\vec{A} = -\partial_x v - \partial_y u = 0\,, \tag{3.84}$$

where the last equality sign follows from the Cauchy-Riemann equations! Similarly,

$$\nabla\times\vec{B} = \partial_x u - \partial_y v = 0 \tag{3.85}$$

as well. This does *not* constitute a full proof, since we have not been careful about whether the partial derivatives of u and v are continuous, but the details are not pertinent to us here.

The Cauchy-Goursat theorem makes many integrals basically trivial. Consider for example the contour C that is the unit circle $|z| = 1$; we have

$$\int_C dz\,\frac{z^2}{z-3} = \int_C dz\, ze^{-z} = \int_C dz\,\tan z = \int_C dz\,\mathrm{Log}(z+2) = 0\,. \tag{3.86}$$

The only thing you have to be careful about in the above expressions is to locate where the integrated functions are *not* analytic, and check that they

lie outside of the contour. A point z_0 is a **singularity** of a function f if f is not analytic at z_0, but there are points in any neighborhood of z_0 where f is analytic. For example, $z^2/(z-3)$ is not analytic at $z=3$, but arbitrarily close to $z=3$, you're fine. This singularity lies outside of the contour C that we were considering above, and so the Cauchy-Goursat theorem applies. Similarly, $\text{Log}(z+2)$ has a branch cut starting at $z=-2$, and extending to the left of that point to $-\infty$, but that's all outside the contour once again.

3.3.6 Cauchy integral formula

With the Cauchy-Goursat theorem, we can now prove another powerful result, known as the **Cauchy integral formula**:

> Let f be analytic everywhere inside and on a closed contour C, oriented in the positive direction (i.e. in the counterclockwise orientation). If z_0 is any point interior to C, then
>
> $$f(z_0) = \frac{1}{2\pi i}\int_C \frac{dz\, f(z)}{z-z_0}\,. \tag{3.87}$$

It'll be quite instructive to see where this result comes from. Let's recall from the last problem set that we showed that given a contour C_ρ given by $z = z_0 + \rho e^{i\theta}$, $0 \le \theta < 2\pi$ (which is simply a circle of radius ρ centered on the point z_0, going in the counterclockwise direction),

$$\int_{C_\rho} \frac{dz}{z-z_0} = 2\pi i\,. \tag{3.88}$$

Now let's look at the integration contours that we want to consider on the complex plane, shown in Fig. 15. We want to evaluate the integral shown Eq. (3.87) around the contour C; but first, let's take a look at the full contour that's shown in Fig. 15. It consists of *1)* the contour C, *2)* a straight path to the circle of radius ρ, *3)* a *clockwise* path around the circle (corresponding to the contour $-C_\rho$), and finally *4)* the same straight path back to the C, in the opposite direction as *2)*. Since $f(z)/(z-z_0)$ is analytic everywhere in the region enclosed by this full contour, we have that

$$\int_C \frac{dz\, f(z)}{z-z_0} = \int_{C_\rho} \frac{dz\, f(z)}{z-z_0}\,. \tag{3.89}$$

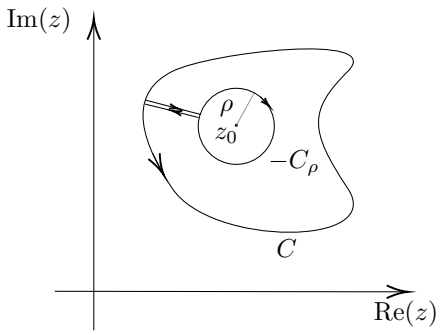

Figure 15: Cauchy integral formula contours. The outer contour is C, while the inner contour, a circular path of radius ρ centered around z_0 in the clockwise direction is $-C_\rho$. Two extra legs connect the C with C_ρ.

This illustrates a very general principle:

> When integrating over some function $f(z)$, you are allowed to deform the contour through any region where $f(z)$ is analytic.

Now, given Eq. (3.89), we can write

$$\int_C \frac{dz\, f(z)}{z-z_0} - f(z_0)\int_{C_\rho}\frac{dz}{z-z_0} = \int_{C_\rho} dz\,\frac{f(z)-f(z_0)}{z-z_0}\,. \tag{3.90}$$

Taking $\rho \to 0$, we see that the integrand on the right-hand side approaches $f'(z_0)$, and so we are left with an integral of a constant over a circle, which just goes to zero, proving (in a way that can be made rigorous) the integral formula.

Let's do a simple example here. Suppose we want to evaluate the integral

$$\int_C dz\,\frac{\cos z}{z(z^2+8)}\,, \tag{3.91}$$

where C is the positively oriented boundary of the square whose sides lie along the lines $x = \pm 2$ and $y = \pm 2$. We can write this integral as

$$\int_C dz \, \frac{\cos z}{z(z^2+8)} = \int_C dz \, \frac{\cos z}{z(z+2\sqrt{2}i)(z-2\sqrt{2}i)} \,. \tag{3.92}$$

One can easily see that two of the singularities of the function are located at $\pm 2\sqrt{2}i$, which both lie outside C. Hence, defining $g(z) \equiv \cos z/((z+2\sqrt{2}i)(z-2\sqrt{2}i))$, we have

$$\int_C dz \, \frac{\cos z}{z(z^2+8)} = \int_C dz \, \frac{g(z)}{z-0} = 2\pi i g(0) = \frac{i\pi}{4} \,, \tag{3.93}$$

where we have used the Cauchy integral formula in the second last equality.

Another intriguing consequence of the Cauchy integral formula can be obtained by taking the derivative of the expression; we find

$$f'(z) = \frac{1}{2\pi i} \int_C \frac{ds \, f(s)}{(s-z)^2} \,. \tag{3.94}$$

We can go on taking derivatives, to find that

for a closed contour C around z_0 with f analytic inside,

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \int_C dz \, \frac{f(z)}{(z-z_0)^{n+1}} \,. \tag{3.95}$$

In fact, given that we can construct an expression for any arbitrary derivative of an analytic function f at z_0, we arrive at the following remarkable theorem:

If a function f is differentiable at a point z_0, it is infinitely differentiable; i.e. if f is analytic at a point z_0, all derivatives of f at z_0 exist.

Let's do another example: let's evaluate

$$\int_C dz \, \frac{\cosh z}{z^4} \,, \tag{3.96}$$

around the contour C which is the positively oriented boundary of the square with sides lying on lines $x = \pm 2$ and $y = \pm 2$. Using the formula above in Eq. (3.95), we have

$$f'''(0) = \frac{3!}{2\pi i} \int_C dz \, \frac{f(z)}{(z-0)^4} \,. \tag{3.97}$$

Thus,

$$\int_C dz \, \frac{\cosh z}{z^4} = \sinh(0) \cdot \frac{2\pi i}{3!} = 0 \,. \tag{3.98}$$

3.4 Series Representations of Complex Functions

So far, the Cauchy-Goursat theorem gives us a powerful tool for integrating functions which are analytic within the contour of integration. But we can do even better! To do this, we need to take a small but important diversion to discuss how to represent complex functions as a series. The basic idea of this is already familiar to you, from your years of experience with **Taylor series**:

Let f be analytic throughout a disk $|z - z_0| < R_0$, centered at z_0 and with radius R_0. Then $f(z)$ has the power series representation

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \,, \tag{3.99}$$

where

$$a_n = \frac{f^{(n)}(z_0)}{n!} \,, n = 0, 1, 2 \cdots \,. \tag{3.100}$$

This series converges to $f(z)$ for any z on the open disk. The largest possible R_0 over which f is analytic is called the **radius of convergence**.

This is just the usual Taylor series, generalized to complex numbers, with

$$f(z) = f(z_0) + \frac{f'(z_0)}{1!}(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \cdots \,, \tag{3.101}$$

when we expand about the complex number z_0, in some disk about z_0 where f is analytic. In particular, if the function is entire, i.e. analytic on the entire complex plane, then this series representation converges everywhere on the complex plane.

Thus, for example, we have

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \,, \tag{3.102}$$

which can be taken to be the definition of e^z. Furthermore, we can take the product and sum of series in the natural way, with e.g.

$$\begin{aligned} \sin z &= \frac{e^{iz} - e^{-iz}}{2i} \\ &= \frac{1}{2i} \left(\sum_{n=0}^{\infty} \frac{(iz)^n}{n!} - \sum_{n=0}^{\infty} \frac{(-iz)^n}{n!} \right) . \end{aligned} \tag{3.103}$$

We note here that the terms with even n cancel, while the odd terms add, and so we can rewrite this as

$$\sin z = \frac{1}{2i} \left(2 \sum_{n=0}^{\infty} \frac{(iz)^{2n+1}}{(2n+1)!} \right) = \frac{1}{i} \sum_{n=0}^{\infty} \frac{i(-1)^n z^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!} \,, \tag{3.104}$$

which you can check is the usual Taylor series for $\sin z$ for the real line, generalized to the complex plane. We can even take the *derivative* of the series, so that for example

$$\frac{d}{dz} \sin z = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!} \,, \tag{3.105}$$

which you can readily check is the Taylor series expansion for $\cos z$. Every function also has a *unique* series expansion in each domain, i.e. you can identify a function by its series expansion and vice-versa.

Another example is the series representation of $f(z) = 1/(1 - z)$ about $z = 0$. First, we note that

$$f^{(n)}(z) = \frac{n!}{(1 - z)^{n+1}} \,, \tag{3.106}$$

and therefore the Taylor series at $z = 0$ (sometimes also known as the **Maclaurin series**) is

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n = \sum_{n=0}^{\infty} z^n \,. \tag{3.107}$$

Note that this series expansion only applies for $|z| < 1$, since f is clearly not analytic at $z = 0$.

Even when a function is not analytic at z_0, it turns out we can *still* obtain a series expansion about the point.

If $f(z)$ is analytic in some annulus $R_1 < |z - z_0| < R_2$ centered at z_0, we can expand f as a **Laurent series**:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n + \sum_{n=1}^{\infty} \frac{b_n}{(z - z_0)^n} \,, (R_1 < |z - z_0| < R_2) \,. \tag{3.108}$$

Given a positively oriented contour C lying in the annulus, we can evaluate

$$\begin{aligned} a_n &= \frac{1}{2\pi i} \int_C \frac{dz\, f(z)}{(z - z_0)^{n+1}} \\ b_n &= \frac{1}{2\pi i} \int_C \frac{dz\, f(z)}{(z - z_0)^{-n+1}} \,. \end{aligned} \tag{3.109}$$

As was the case for the Taylor series, you can add, multiply or differentiate Laurent series, and the Laurent series in any particular domain is unique for each function. I won't prove the expressions for a_n and b_n, but if you look back to Eq. (3.95), you can see that a_n has the right structure for a Taylor series, and b_n is simply a generalization of that. If you're interested in the proof, I refer you to CB61.

We can frequently obtain the Laurent series of a function by starting from its Taylor series, but being careful about the domain of applicability of the series, since the Laurent series is defined in some annulus around a singularity. So for example,

$$z^2 \sin\left(\frac{1}{z^2}\right) = z^2 \sum_{n=0}^{\infty} \frac{(-1)^n (1/z^2)^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \frac{1}{z^{4n}} \,, \, (0 < |z| < \infty) \,, \tag{3.110}$$

where the domain of validity can be deduced from the location of the singularity at $z = 0$. In this case, the inner radius of the annulus can be taken to zero, since the function is analytic arbitrarily close to the singularity.

Another instructive example is the function

$$f(z) = \frac{1}{(z-1)(z-2)} = \frac{1}{z-2} - \frac{1}{z-1} \,, \tag{3.111}$$

which has singularities at $z = 1$ and $z = 2$. We want to obtain three different Laurent series centered about $z_0 = 0$, in the annuli given by *1)* $D_1 = |z| < 1$, *2)* $D_2 = 1 < |z| < 2$, and *3)* $D_3 = 2 < |z| < \infty$ (see Fig. 16). For the first annulus, we can use the usual Taylor series expression that we saw above,

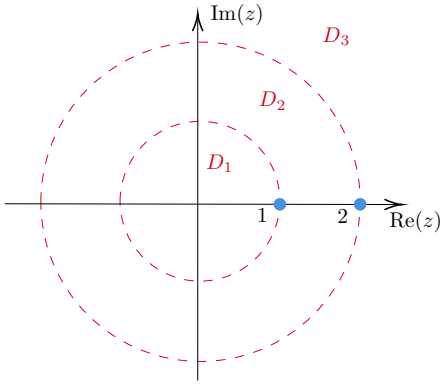

Figure 16: The annuli for computing the Laurent series of $1/((z-1)(z-2))$.

$1/(1-z) = \sum_{n=0}^{\infty} z^n$, to find

$$\begin{aligned} f(z) &= -\left(-\sum_{n=0}^{\infty} z^n\right) + \left(-\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n\right) \\ &= \sum_{n=0}^{\infty}\left(1-2^{-n-1}\right) z^n\,,\ (|z|<1)\,. \end{aligned} \tag{3.112}$$

This is just a Taylor series, since there is no singularity in D_1.

Moving on, in D_2, we know that the Taylor series expression $1/(1-z) = \sum_{n=0}^{\infty} z^n$ fails, and that we must have a true Laurent series in this annulus, with nonzero b_n in Eq. (3.108). However, we know that in D_2, $|1/z| < 1$, and so we can write

$$\frac{1}{z-1} = \frac{1}{z}\left(\frac{1}{1-z^{-1}}\right) = \frac{1}{z}\sum_{n=0}^{\infty}\frac{1}{z^n}\,. \tag{3.113}$$

Therefore,

$$\begin{aligned} f(z) &= -\frac{1}{z}\sum_{n=0}^{\infty}\frac{1}{z^n} + \left(-\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^n\right) \\ &= -\sum_{n=0}^{\infty}\frac{z^n}{2^{n+1}} - \sum_{n=1}^{\infty}\frac{1}{z^n}\,,\ (1<|z|<2)\,. \end{aligned} \tag{3.114}$$

Finally, in D_3, we need to do the same trick with $1/(z-2)$,

$$\frac{1}{z-2} = \frac{1}{z}\left(\frac{1}{1-2/z}\right) = \frac{1}{z}\sum_{n=0}^{\infty}\left(\frac{2}{z}\right)^n\,, \tag{3.115}$$

so that

$$\begin{aligned} f(z) &= -\sum_{n=1}^{\infty}\frac{1}{z^n} + \sum_{n=1}^{\infty}\frac{2^{n-1}}{z^n} \\ &= \sum_{n=1}^{\infty}\frac{2^{n-1}-1}{z^n}\,,\ (|z|>2)\,. \end{aligned} \tag{3.116}$$

3.5 Integration Part II

We are now ready to continue our discussion of complex integration, armed with our ability to write down a series expansion for analytic functions.

3.5.1 Residues and Poles

First, some terminology. We've already defined a *singularity* as a point z_0 where a function f fails to be analytic, but is otherwise analytic at some point in any neighborhood around z_0. If, in addition, we can cut the point out, and find a disk $0 < |z-z_0| < \varepsilon$ where f is analytic, we call z_0 an **isolated singularity**.

Now, we can write down a Laurent series for the function on this disk, expanding about the singularity. In general, this will look like

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n + \sum_{n=1}^{\infty}\frac{b_n}{(z-z_0)^n}\,,\ (0<|z-z_0|<\varepsilon)\,. \tag{3.117}$$

There are three possibilities for the Laurent series here:

1. $b_n = 0$ for all n. This is called a **removable singularity**. For example, consider the function $f(z) = (1-\cos z)/z^2$. This function has a singularity at $z = 0$, but if we look at its Laurent series expanded about 0, we have

$$\begin{aligned} f(z) &= \frac{1}{z^2}\left(-\frac{z^2}{2!} + \frac{z^4}{4!} + \cdots\right) \\ &= \frac{1}{2!} + \frac{z^2}{4!} + \cdots . \end{aligned} \tag{3.118}$$

   We see here that if we simply redefine $f(0) = 1/2$, we get an *entire function*.

2. $b_m \neq 0$ for some m, but $b_{m+1} = b_{m+2} = \cdots = 0$. This is called a **pole of order** m. A pole of order $m = 1$ is usually referred to as a **simple pole**. For example,

$$\frac{z^2 - 2z + 3}{z - 2} = (z-2) + 2 + \frac{3}{z-2}\,, \; (0 < |z-2| < \infty)\,, \tag{3.119}$$

   and so we have a simple pole at $z = 2$.

3. An infinite number of b_n is nonzero. These are called **essential singularities**.

Finally, given the Laurent series, we call b_1 the **residue** of f at the singularity. We often denote the residue of f at the singularity z_0 as

$$b_1 = \operatorname*{Res}_{z=z_0} f(z)\,. \tag{3.120}$$

(**End of Lecture**: Monday Nov 4 2024)

3.5.2 Cauchy's residue theorem

Let's revisit the expression for the Laurent series in Eq. (3.108). Suppose f has an isolated singularity at at z_0. We have

$$b_1 = \operatorname*{Res}_{z=z_0} f(z) = \frac{1}{2\pi i}\int_C dz\, f(z)\,, \tag{3.121}$$

if C is a closed contour around z_0, lying within a disk where f is analytic everywhere except z_0. This generalizes immediately to the immensely powerful **Cauchy residue theorem**:

> Let C be a closed contour oriented positively. If a function f is analytic inside the contour except for a finite number of singular points z_k, $k = 1, \cdots, n$, inside C, then
>
> $$\int_C dz\, f(z) = 2\pi i \sum_{k=1}^{n} \operatorname*{Res}_{z=z_k} f(z)\,. \tag{3.122}$$

The proof of this theorem is pretty straightforward. Given a contour C such as the one shown in Fig. 17, we first consider the deformed path as shown. Along the deformed path, we get the contour integral given by

$$\int_C dz\, f(z) - \sum_k \int_{C_k} dz\, f(z) = 0\,, \tag{3.123}$$

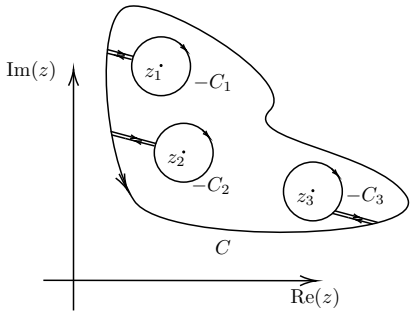

Figure 17: Proving Cauchy's residue theorem.

where we get zero because the deformed path encloses a region where f is analytic. But we can immediately evaluate each term using the Laurent series expression for b_1, thus proving the residue theorem.

Before we dive into numerous examples demonstrating the power of the residue theorem, let's discuss the concept of the **residue at infinity**. Pick a contour C_0 centered at the origin that is large enough that it encompasses all of the singularities of the function, *oriented clockwise*. Then we define the residue at infinity as

$$\int_{C_0} dz\, f(z) = 2\pi i \operatorname*{Res}_{z=\infty} f(z)\,. \tag{3.124}$$

Since we can deform contours through regions that are analytic, for any contour C oriented positively encompassing all singularities of a function f,

$$\int_{C} dz\, f(z) = -2\pi i \operatorname*{Res}_{z=\infty} f(z)\,. \tag{3.125}$$

How do we evaluate this residue? Well, using Eq. (3.108), if we write the Laurent series expanded about the origin, we have

$$f(z) = \sum_{n=0}^{\infty} a_n z^n + \sum_{n=1}^{\infty} \frac{b_n}{z^n} = \sum_{n=-\infty}^{\infty} c_n z^n\,, \tag{3.126}$$

with $c_n = a_n$ for $n \geq 0$, and $c_{-n} = b_n$ for $n < 0$. From Eq. (3.108) that

$$c_n = \frac{1}{2\pi i} \int_{-C_0} dz\, \frac{f(z)}{z^{n+1}}\,, \tag{3.127}$$

since C_0 is big enough that it lies away from all of the isolated singularities, and therefore lies on an annulus where f is analytic. In particular,

Notice the negative sign, defined to be clockwise.

$$c_{-1} = \frac{1}{2\pi i} \int_{-C_0} dz\, f(z) = -\operatorname*{Res}_{z=\infty} f(z)\,. \tag{3.128}$$

So if we have a way of evaluating c_{-1} efficiently, we would have the residue at infinity, and also the contour integrals of interest. Let's note however that

$$\frac{1}{z^2} f\left(\frac{1}{z}\right) = \frac{1}{z^2} \sum_{n=-\infty}^{\infty} \frac{c_n}{z^n} = \sum_{n=-\infty}^{\infty} \frac{c_{n-2}}{z^n}\,. \tag{3.129}$$

Therefore,

$$\operatorname*{Res}_{z=\infty} f(z) = -c_{-1} = -\operatorname{Res}\left[\frac{1}{z^2} f\left(\frac{1}{z}\right)\right]\,. \tag{3.130}$$

This gives the following result:

> If a function f is analytic everywhere in the complex plane except for a finite number of singular points interior to a positively oriented closed contour C, then
>
> $$\int_{C} dz\, f(z) = 2\pi i \operatorname*{Res}_{z=0} \left[\frac{1}{z^2} f\left(\frac{1}{z}\right)\right]\,. \tag{3.131}$$

On to a bunch of examples!

Example 1. Let's evaluate the integral

$$\int_{C} dz\, \frac{\cosh \pi z}{z(z^2+1)}\,, \tag{3.132}$$

where C is the circle $|z| = 2$, oriented positively. There are three poles in this expression: $z = 0$, and $z = \pm i$. Instead of actually finding the Laurent series about each of these poles, let's use the following fact that makes life a lot easier:

An isolated singular point z_0 of a function f is a pole of order m if and only if $f(z)$ can be written in the form

$$f(z) = \frac{\phi(z)}{(z - z_0)^m}\,, \tag{3.133}$$

where $\phi(z)$ is analytic and nonzero at z_0. Moreover,

$$\operatorname*{Res}_{z=z_0} f(z) = \phi(z_0)\,,\ (m = 1)\,, \tag{3.134}$$

and

$$\operatorname*{Res}_{z=z_0} f(z) = \frac{\phi^{(m-1)}(z_0)}{(m-1)!}\,,\ (m \geq 2)\,, \tag{3.135}$$

where $\phi^{(m-1)}(z_0)$ is the $(m-1)$th derivative, evaluated at z_0.

Let's see why the formula for the residue works. Since $\phi(z)$ is analytic at z_0, we can Taylor expand $\phi(z)$ as

$$\phi(z) = \phi(z_0) + \phi'(z_0)(z - z_0) + \frac{1}{2!}\phi''(z_0)(z - z_0)^2 + \cdots\,, \tag{3.136}$$

and therefore

$$f(z) = \frac{\phi(z_0)}{(z - z_0)^m} + \frac{\phi'(z_0)}{(z - z_0)^{m-1}} + \frac{1}{2!}\frac{\phi''(z_0)}{(z - z_0)^{m-2}} + \cdots\,. \tag{3.137}$$

We can see from this that the residue, which corresponds to the term with just $(z - z_0)$ in the denominator, is exactly what is given above.

Returning to our example, we have for $f(z) = \cosh(\pi z)/(z(z^2 + 1))$,

$$\begin{aligned}
&\operatorname*{Res}_{z=0} f(z) = 1\,,\\
&\operatorname*{Res}_{z=i} f(z) = \frac{\cosh i\pi}{i \cdot (i + i)} = \frac{1}{2}\,,\\
&\operatorname*{Res}_{z=-i} f(z) = \frac{\cosh(-i\pi)}{-i \cdot (-i - i)} = \frac{1}{2}\,.
\end{aligned} \tag{3.138}$$

Putting everything together, we find

$$\int_C dz\, \frac{\cosh \pi z}{z(z^2 + 1)} = 2\pi i \left(1 + \frac{1}{2} + \frac{1}{2}\right) = 4\pi i\,. \tag{3.139}$$

Example 2. Let's evaluate the integral of $f(z)$ around the positively oriented circle $|z| = 3$, with

$$f(z) = \frac{z^3 e^{1/z}}{1 + z^3}\,. \tag{3.140}$$

The singularities are located at $z = 0$, as well as at $z = e^{i\pi/3}$, $z = -1$ and $z = e^{i5\pi/3}$. But these singularities are all located inside the contour, and so we can use the residue at infinity to evaluate the integral. We find

$$\frac{1}{z^2} f\left(\frac{1}{z}\right) = \frac{1}{z^2}\frac{e^z}{z^3(1 + z^{-3})} = \frac{1}{z^2}\frac{e^z}{1 + z^3}\,. \tag{3.141}$$

We can expand this as a Laurent series about the origin as

$$\begin{aligned}\frac{1}{z^2}f\left(\frac{1}{z}\right) &= \frac{1}{z^2}\left(\sum_{n=0}^{\infty}\frac{z^n}{n!}\right)\left(\sum_{n=0}^{\infty}(z^3)^n\right)\\ &= \frac{1}{z^2}+\frac{1}{z}+\frac{1}{2}+\cdots\,,\end{aligned} \tag{3.142}$$

and so the residue at $z=0$ is simply 1. We deduce that

$$\int_C dz\, f(z) = 2\pi i \operatorname*{Res}_{z=0}\frac{1}{z^2}f\left(\frac{1}{z}\right) = 2\pi i\,. \tag{3.143}$$

Example 3. Let's integrate $f(z)=\tan z$ along the contour C, the positively oriented circle $|z|=2$. Since $\tan z = \sin z/\cos z$, there are singularities within the contour at $z=\pm\pi/2$. About either of these singularities, we can expand

$$\begin{aligned}\sin z &= \sin z_0 + \cdots\\ \cos z &= \cos(z_0) - \sin z_0\cdot(z-z_0)+\cdots\\ &= -\sin z_0\cdot(z-z_0)+\cdots\,,\end{aligned} \tag{3.144}$$

since $\cos(z_0)=0$. $\cdots$ represents terms that are of order $z-z_0$ or higher. Hence,

$$\begin{aligned}\tan z &= \frac{\sin z_0+\cdots}{-\sin z_0\cdot(z-z_0)+\cdots}\\ &= -\frac{1}{z-z_0}\frac{1+\cdots}{1-\cdots}\\ &= -\frac{1}{z-z_0}\left(1+\cdots\right)\,.\end{aligned} \tag{3.145}$$

Clearly then, the residue at both singularities is -1, and

$$\int_C dz\,\tan z = -4\pi i\,. \tag{3.146}$$

3.6 Improper integrals

By far the most powerful application of the residue theorem is in evaluating improper integrals. These are integrals of the form

$$\int_0^\infty dx\, f(x) = \lim_{R\to\infty}\int_0^R dx\, f(x)\,, \tag{3.147}$$

and also

$$\int_{-\infty}^{\infty} dx\, f(x) = \lim_{R_1\to\infty}\int_{-R_1}^{\infty} dx\, f(x) + \lim_{R_2\to\infty}\int_0^{R_2} dx\, f(x)\,, \tag{3.148}$$

if these limits are well-defined.

In addition, sometimes we are interested in the **Cauchy principal value** of an integral, which is

$$\text{P.V.}\int_{-\infty}^{\infty} dx\, f(x) = \lim_{R\to\infty}\int_{-R}^{R} dx\, f(x)\,. \tag{3.149}$$

So, for example,

$$\text{P.V.}\int_{-\infty}^{\infty} dx\, x = \lim_{R\to\infty}\left[\frac{x^2}{2}\right]_{-R}^{R} = 0\,, \tag{3.150}$$

even though the limit in Eq. (3.148) doesn't exist.

For even functions, i.e. $f(x) = f(-x)$, we also have the relation that

$$\int_0^\infty dx\, f(x) = \frac{1}{2}\int_{-\infty}^\infty dx\, f(x)\,. \tag{3.151}$$

The most basic strategy to evaluate many such integrals is to choose a contour that runs from $-R$ to R along the real axis, and then a semicircular contour connecting these two ends. Finally, we take $R \to \infty$, and check that the semicircular part of the contour goes to zero in that limit. The contour integral can then be evaluated using the residue theorem.

The first example we will examine is

$$\int_0^\infty dx\, f(x) = \int_0^\infty dx\, \frac{x^2}{x^6+1} = \frac{1}{2}\int_{-\infty}^\infty dx\, \frac{x^2}{x^6+1}\,. \tag{3.152}$$

To evaluate this, we consider the contour C described above, which we write as

$$\int_C dz\, f(z) = \int_{-R}^R dx\, \frac{x^2}{x^6+1} + \int_{C_R} dz\, f(z)\,, \tag{3.153}$$

where C_R is the semicircular contour of radius R, centered on the origin. The poles of $f(z)$ are to be found at

$$z = e^{i\pi/6},\ i,\ e^{5\pi i/6},\ e^{7\pi i/6},\ -i,\ e^{11\pi i/6}\,. \tag{3.154}$$

Of these poles, three of them are in the upper half of the complex plane, and will be inside C when we take $R \to \infty$.

Let's first evaluate the residues. Defining

$$\phi_{z_0}(z) = \frac{z^2}{z^6+1}(z - z_0)\,, \tag{3.155}$$

we know from the theorem shown in Eq. (3.134) that

$$\begin{aligned}
\operatorname*{Res}_{z=e^{i\pi/6}} f(z) &= \phi_{e^{i\pi/6}}(e^{i\pi/6}) = -\frac{i}{6}\,,\\
\operatorname*{Res}_{z=i} f(z) &= \phi_i(i) = \frac{i}{6}\,,\\
\operatorname*{Res}_{z=e^{5i\pi/6}} f(z) &= \phi_{5i\pi/6}(e^{5i\pi/6}) = -\frac{i}{6}\,,
\end{aligned} \tag{3.156}$$

with the aid of Mathematica.

Next, let's turn our attention to the integral over the semicircle. The integral can be parametrized by θ where $z = Re^{i\theta}$,

$$\begin{aligned}
\int_{C_R} dz\, f(z) &= \int_0^\pi d\theta\, \frac{dz}{d\theta}\, \frac{R^2 e^{2i\theta}}{R^6 e^{6i\theta}+1}\\
&= iR^3 \int_0^\pi \frac{e^{3i\theta}}{R^6 e^{6i\theta}+1}\,.
\end{aligned} \tag{3.157}$$

We now want to use the modulus inequality, given in Eq. (3.62), to show that this integral goes to zero as $R \to \infty$. We have

$$\left|\frac{e^{3i\theta}}{R^6 e^{6i\theta}+1}\right| = \frac{1}{|R^6 e^{6i\theta}+1|}\,. \tag{3.158}$$

A useful general result is that

$$\begin{aligned}|R^n e^{in\theta} \pm 1| &= \sqrt{(R^n\cos(n\theta)\pm 1)^2 + (R^n\sin(n\theta))^2} \\ &= \sqrt{R^{2n} \pm 2R^n\cos(n\theta) + 1} \\ &\geq R^n - 1 \end{aligned} \tag{3.159}$$

Thus,

$$\left|\frac{e^{3i\theta}}{R^6 e^{6i\theta}+1}\right| \leq \frac{1}{R^6-1}\,, \tag{3.160}$$

From this, we deduce from the modulus inequality in Eq. (3.62) that

$$\lim_{R\to\infty}\left|\int_{C_R} dz\, f(z)\right| \leq \lim_{R\to\infty}\frac{\pi\cdot R^3}{R^6-1} = 0\,, \tag{3.161}$$

i.e. the integral itself must go to zero.

Thus, using the residue theorem, we find

$$\lim_{R\to\infty}\int_C dz\, f(z) = \int_{-\infty}^{\infty} dx\,\frac{x^2}{x^6+1} = 2\pi i\left(-\frac{i}{6}+\frac{i}{6}-\frac{i}{6}\right) = \frac{\pi}{3}\,, \tag{3.162}$$

so that

$$\int_0^{\infty} dx\,\frac{x^2}{x^6+1} = \frac{\pi}{6}\,. \tag{3.163}$$

Let's do another example, this time evaluating the Cauchy principal value of

$$\int_{-\infty}^{\infty} dx\, f(x) = \int_{-\infty}^{\infty} dx\,\frac{x\sin x}{x^2+2x+2} \tag{3.164}$$

Let's begin by locating the poles; we can solve $x^2+2x+2=0$ to find that the poles are at

$$z_0 = -1 \pm \frac{1}{2}\sqrt{4-8} = -1\pm i\,. \tag{3.165}$$

To evaluate the integral, we want to do the following:

$$\int_C dz\, f(z) = \mathrm{Im}\int_C dz\,\frac{ze^{iz}}{z^2+2z+2}\,, \tag{3.166}$$

where if we choose the contour to go from $-R$ to R on the real axis, and then take $R\to\infty$, the imaginary part will pick out $\sin x$ on the real axis, giving us the integral that we want. But how do we close the contour? Well, we will have to close it *upward* in a semicircle, so that when you plug in z with a large positive imaginary component into e^{iz}, you'll get something that is exponentially *suppressed*. So using the same contour as before, let's examine the integral along the semicircle given by $|z|=R$, where $z=Re^{i\theta}$:

$$\int_{C_R} dz\,\frac{ze^{iz}}{z^2+2z+2} = \int_0^{\pi} d\theta\,\frac{iR^2e^{2i\theta}e^{iRe^{i\theta}}}{(Re^{i\theta})^2+2Re^{i\theta}+2}\,. \tag{3.167}$$

Notice however that I can write

$$e^{iRe^{i\theta}} = e^{iR\cos\theta}e^{-R\sin\theta}\,. \tag{3.168}$$

The exponential suppression here, which is not compensated for by any other terms, guarantees that the integrand goes to zero as $R \to \infty$. Thus

$$\lim_{R\to\infty} \int_{C_R} dz\, \frac{z e^{iz}}{z^2+2z+2} = 0\,. \tag{3.169}$$

Let's turn our attention back to the poles again. Only one pole is in the upper half plane, and that is $-1+i$. The residue at this point is

$$\operatorname*{Res}_{z=-1+i} g(z) = \frac{(-1+i)e^{i(-1+i)}}{(-1+i-(-1-i))} = \frac{-1+i}{2i} e^{-1-i} \tag{3.170}$$

Hence,

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, \frac{x\sin x}{x^2+2x+2} &= \operatorname{Im}\left(2\pi i\cdot \frac{-1+i}{2i} e^{-1-i}\right)\\ &= \operatorname{Im}\left(\pi(-1+i)e^{-1}(\cos 1 - i\sin 1)\right)\\ &= \frac{\pi}{e}\left(\cos 1 + \sin 1\right)\,.\end{aligned} \tag{3.171}$$

(**End of Lecture**: Wednesday Nov 6 2024)

The next example is one that comes up very regularly in the study of Fourier transforms:

$$\int_{-\infty}^{\infty} dx\, \frac{\sin x}{x} = \pi\,. \tag{3.172}$$

To evaluate this integral, we'll consider the contour integral $\int_C dz\, f(z)$ where $f(z) = e^{iz}/z$, with the following contour C: *1)* a big semicircular, anticlockwise contour C_R of radius R, starting at $z = R$ and ending at $z = -R$; *2)* a segment along the real line from $z = -R$ to $z = -\rho$; *3)* another smaller semicircle C_ρ, this time clockwise, of radius ρ; *4)* another segment along the real line from $z = \rho$ to $z = R$. The contour is shown in Fig. **??**. We can see that as $\rho \to 0$ and $R \to \infty$, the two line segments will add up to give the desired integral once we take the imaginary part. Furthermore, since the only pole of $f(z)$ occurs at $z = 0$, which lies outside of the contour, $\int_C dz\, f(z) = 0$.

Let's look at the two contour integrals along C_R and C_ρ. For C_R, we have

$$\begin{aligned}\int_{C_R} dz\, f(z) &= \int_0^\pi d\theta\, (iRe^{i\theta}) \frac{e^{iRe^{i\theta}}}{Re^{i\theta}}\\ &= i\int_0^\pi d\theta\, e^{iR\cos\theta} e^{-R\sin\theta}\,,\end{aligned} \tag{3.173}$$

which clearly goes to zero as $R \to \infty$. On the other hand,

$$\begin{aligned}\lim_{\rho\to 0}\int_{C_\rho} dz\, f(z) &= -i\lim_{\rho\to 0}\int_0^\pi d\theta e^{i\rho\cos\theta} e^{-\rho\sin\theta}\\ &= -i\int_0^\pi d\theta\\ &= -i\pi\,.\end{aligned} \tag{3.174}$$

Therefore,

$$\begin{aligned}&\int_{-\infty}^{\infty} dx\, \frac{e^{ix}}{x} + \lim_{\rho\to 0}\int_{C_\rho} dz\, f(z) = 0\\ &\implies \int_{-\infty}^{\infty} dx\, \frac{\sin x}{x} = \operatorname{Im}\int_{-\infty}^{\infty} dx\, \frac{e^{ix}}{x} = \pi\end{aligned} \tag{3.175}$$

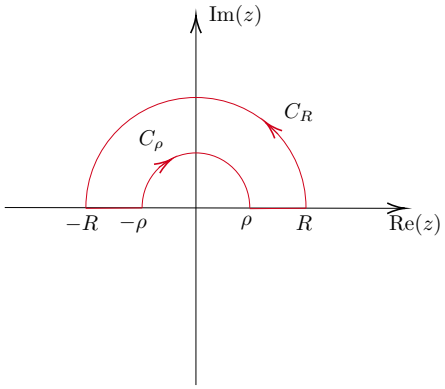

Figure 18: Contour for the integral $\int_{-\infty}^{\infty} dx(\sin x/x)$.

as required.

Our final example involves a branch cut. Consider the integral

$$\int_0^\infty dx\, f(x) = \int_0^\infty dx\, \frac{x^{-a}}{x+1}\,, \tag{3.176}$$

where $0 < a < 1$, and the corresponding complex function

$$f(z) = \frac{z^{-a}}{z+1}\,, \tag{3.177}$$

where $z^{-a} = \exp(-a \log z)$. This is a multivalued function, and requires us to choose a branch in order for it to be single valued, with the branch point occurring at $z = 0$.

The way that we'll evaluate this contour is as follows: consider two concentric circular contours, C_ρ and C_R, with $\rho < R$, both positively oriented. Choose the branch where $0 < \arg(z) < 2\pi$, so that the branch cut lies along the positive real axis. We are going to start at $x = \rho$ on the real axis, integrate slightly above the branch cut until $x = R$. We'll then go in a counterclockwise circle until we hit $x = R$ from just below the branch cut, integrate along R to ρ, and then go around $-C_\rho$ to complete the integral. This is shown in Fig. **??**.

With our choice of branch cut, when we are just above the branch cut, the argument of z along that contour is 0, whereas just below the branch cut, the argument is 2π, so that for $z = re^{i2\pi}$ there, $z^{-a} = \exp(-a \log(re^{i2\pi})) = r^{-a}e^{-2\pi i a}$. We then want to take $\rho \to 0$ and $R \to \infty$.

Let's consider the integral over C_R. This is given by

$$\int_{C_R} dz\, f(z) = \int_0^{2\pi} d\theta\, iRe^{i\theta} \frac{R^{-a}e^{-ia\theta}}{Re^{i\theta}+1}\,, \tag{3.178}$$

so that the integrand has an absolute value given by

$$\left| iRe^{i\theta} \frac{R^{-a}e^{-ia\theta}}{Re^{i\theta}+1} \right| = \frac{R^{1-a}}{|Re^{i\theta}+1|} \le \frac{R^{1-a}}{R-1}\,, \tag{3.179}$$

so that by the modulus inequality in Eq. (3.62),

$$\left| \int_{C_R} dz\, f(z) \right| \le \frac{2\pi \cdot R^{1-a}}{R-1}\,, \tag{3.180}$$

which goes to zero as $R \to \infty$ since $a > 0$.

Similarly, we also have

$$\left| \int_{C_\rho} dz\, f(z) \right| \le \frac{2\pi \cdot \rho^{1-a}}{\rho-1}\,, \tag{3.181}$$

which also goes to zero as $\rho \to 0$! Hence, the integral over the full contour when we take the limits of $\rho \to 0$ and $R \to \infty$ is

$$\begin{aligned} &\lim_{\rho\to 0}\lim_{R\to\infty} \left(\int_\rho^R dx\, \frac{x^{-a}}{x+1} + \int_R^\rho dx\, \frac{x^{-a}e^{-2\pi i a}}{x+1} \right) \\ &= \lim_{\rho\to 0}\lim_{R\to\infty} \left[\left(1 + e^{-2\pi i a}\right) \int_\rho^R dx\, \frac{x^{-a}}{x+1} \right] \\ &= \left(1 + e^{-2\pi i a}\right) \int_0^\infty dx\, \frac{x^{-a}}{x+1}\,. \end{aligned} \tag{3.182}$$

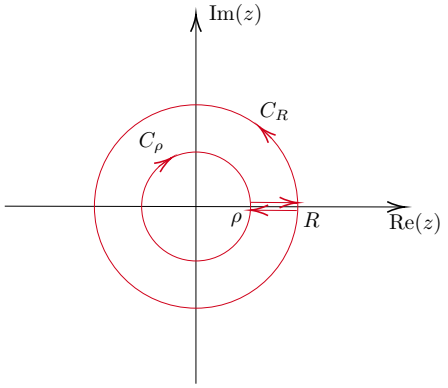

Figure 19: Contour for the integral $\int_{-\infty}^{\infty} x^{-a}/(x+1)$.

It remains to use the Cauchy residue theorem to actually evaluate the integral! There is one pole in the integral, located at $x = -1$, with residue given by $(-1)^{-a} = e^{-ia\pi}$. Thus,

$$\begin{aligned}
\left(1 + e^{-2\pi i a}\right) \int_0^\infty dx \, \frac{x^{-a}}{x+1} &= 2\pi i e^{-i\pi a} \\
\int_0^\infty dx \, \frac{x^{-a}}{x+1} &= \frac{2\pi i e^{-i\pi a}}{1 + e^{-2\pi i a}} \\
&= \frac{2\pi i}{e^{\pi i a} - e^{-\pi i a}} \\
&= \frac{\pi}{\sin a\pi} \, .
\end{aligned} \tag{3.183}$$

4 Fourier Analysis

References: Altland and van Delft, Chapter C6

Given a vector $\vec{v}$ in an arbitrary vector space, we know it is always possible to choose a basis $\{\vec{e}_i\}$ in the vector space, and write the vector as a linear combination $\vec{v} = v^i \vec{e}_i$. In an inner product space, we can furthermore choose an *orthonormal* basis, where the inner product $\langle \vec{e}_i, \vec{e}_j \rangle = \delta_{ij}$.

Functions can be viewed as vectors on an inner product space over $\mathbb{R}$ or $\mathbb{C}$, with the inner product being defined in a form similar to

$$\langle f, g \rangle \sim \int dx\, f^*(x) g(x)\,, \tag{4.1}$$

where * denotes the complex conjugate of the function.[33] You can check that this satisfies the axioms of an inner product discussed in Sec. 2.1.1: $\langle f, g \rangle = \langle g, f \rangle^*$, $\langle f, \lambda g + \mu h \rangle = \lambda \langle f, g \rangle + \mu \langle f, h \rangle$, and $\langle f, g \rangle = 0$ for all g implies $f = 0$. Given this, we can now try to decompose functions into a linear combination of well-behaved, orthogonal basis functions that often have a useful, physical meaning. This idea is ubiquitous in applied mathematics and physics, and depending on the exact definition of the inner product and teh class of functions you are interested in, you may already be familiar with many different choices of basis functions, including Legendre polynomials, Chebyshev polynomials, spherical harmonics, Bessel functions, etc. We're now going to study the most basic of these: the decomposition of a function into sines and cosines, which a technique known as Fourier analysis.

[33] This is fine also if you're just considering real functions; the complex conjugate just doesn't do anything.

4.1 Dirac Delta Function

Before we dive into Fourier analysis, let's spend some time familiarizing ourselves with the **Dirac delta function**, denoted $\delta(x)$, which will feature heavily in Fourier analysis.

It is defined as

$$\int_a^b dx\, f(x)\delta(x) = \begin{cases} f(0)\,, & a < 0 < b\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{4.2}$$

Integrating against $\delta(x)$ across the origin therefore picks out the function value at $x = 0$. More generally, we have

$$\int_a^b dx\, f(x)\delta(x-y) = \begin{cases} f(y)\,, & a < y < b\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{4.3}$$

which can be deduced by performing the substitution $z = x - y$.

What is this peculiar object? We can get a sense of what it is by considering the integral

$$\int_{-\infty}^x dy\, \delta(y) = \begin{cases} 1\,, & x > 0\,, \\ 0 & x < 0 \end{cases} \equiv \Theta(x)\,, \tag{4.4}$$

where $\Theta(x)$ is known as the **Heaviside step function**. We see that the Dirac delta function is therefore the *derivative* of the Heaviside step function, which

has zero gradient at $x \neq 0$, but a jump discontinuity at the origin, i.e. heuristically,

$$\delta(x) = \begin{cases} 0\,, & x \neq 0\,, \\ \infty\,, & x = 0\,. \end{cases} \tag{4.5}$$

This really isn't a function at all, since functions shouldn't be returning infinite values. Nevertheless, it is an extremely useful object that mathematicians call a *distribution*, and that physicists use with abandon. Sometimes, it is useful to think of $\delta(x)$ as being the limit of various functions. Indeed, you can show that e.g.

$$\delta(x) = \lim_{\epsilon\to 0} \frac{1}{\epsilon\sqrt{\pi}} e^{-(x-y)^2/\epsilon^2} = \lim_{\epsilon\to 0} \frac{1}{\pi}\frac{\epsilon}{\epsilon^2+x^2} = \lim_{k\to\infty} \frac{\sin kx}{\pi x}\,, \tag{4.6}$$

in the sense that the limit of these functions satisfy Eq. (4.2). Let's check the last one since it comes up quite often. We see that for any function $f(x)$, for any $a, b \in \mathbb{R}$,

$$\begin{aligned} \lim_{k\to\infty} \int_a^b dx\, f(x) \frac{\sin kx}{\pi x} &= \frac{1}{\pi} \lim_{k\to\infty} \int_{ka}^{kb} \frac{dy}{k}\, f\left(\frac{y}{k}\right) \frac{\sin y}{y/k} \\ &= \begin{cases} \frac{1}{\pi} f(0) \int_{-\infty}^{\infty} dy\, \sin y/y = f(0)\,, & a < 0 < b\,, \\ 0 & \text{otherwise}\,, \end{cases} \end{aligned} \tag{4.7}$$

which is exactly how the delta function should behave. $\int_{-\infty}^{\infty} dy\, \sin y/y = \pi$, which we showed earlier.

A few other properties of the Dirac delta function are worth mentioning:

1. For some real number c, if $a < 0 < b$,

$$\int_a^b dx\, \delta(cx) f(x) = \int_{ac}^{bc} \frac{dy}{c} \delta(y) f(y/c) = \frac{f(0)}{|c|}\,, \tag{4.8}$$

   where the absolute value sign appears because if $c < 0$, the direction of integration is reversed, which picks up an extra minus sign. Therefore,

$$\delta(cx) = \frac{\delta(x)}{|c|}\,. \tag{4.9}$$

2. More generally, for $a < 0 < b$,

$$\int_a^b dx\, \delta(g(x)) f(x) = \int_{g(a)}^{g(b)} dg\, \delta(g)\, |g'(x)|^{-1}\, f(x) = \sum_i \frac{f(x_i)}{|g'(x_i)|}\,, \tag{4.10}$$

   where x_i are all the points $a < x_i < b$ satisfying $g(x_i) = 0$.

4.2 Fourier Series

Let's consider complex functions f that are defined on the interval $-L/2 < x < L/2$, subjected to periodic boundary conditions, i.e. $f(-L/2) = f(L/2)$. We'll define the inner product between two functions f and g as[34]

$$\langle f, g\rangle \equiv \int_{-L/2}^{L/2} dx\, f^*(x) g(x)\,. \tag{4.11}$$

[34] Clearly not all functions lead to a well-defined integral when defined this way. We won't spend any time discussing defining the vector space of functions exactly. For real functions, functions that satisfy what are known as the Dirichlet conditions are well-behaved with respect to what we will discuss next; see AvD C6.2 for more information.

The goal now is to find a nice, orthogonal basis for f, i.e. a set of functions such that any function f that is periodic on $-L/2 < x < L/2$ can be written as a linear combination of the basis functions. The fact that we are interested in periodic functions strongly suggests we should be looking for functions that are well-defined on a circle, such as sines and cosines. In fact, the orthogonal basis $\{\phi_n\}$ we're looking for is

$$\phi_n(x) = \frac{1}{\sqrt{L}} e^{2\pi i n x/L}\,,\; n \in \mathbb{Z}\,. \tag{4.12}$$

Let's prove a few useful statements:

1. **Orthonormality**. First, we note that

$$\begin{aligned}\int_{-L/2}^{L/2} dx\, \phi_m^*(x)\phi_n(x) &= \frac{1}{L}\int_{-L/2}^{L/2} dx\, e^{2\pi i(n-m)x/L} \\ &= \begin{cases} 1\,, & n = m \\ \frac{1}{2\pi i(n-m)} e^{2\pi i(n-m)x/L}\Big|_{x=-L/2}^{x=L/2}\,, & n \neq m \end{cases} \\ &= \delta_{mn}\,, \end{aligned} \tag{4.13}$$

   which already shows that $\phi_n(x)$ is *orthonormal* with respect to the defined inner product. But note that for any arbitrary function $f(x)$,

2. **Completeness**. The completeness condition reads

$$\frac{1}{\sqrt{L}} \sum_{n=-\infty}^{n=\infty} \phi_n(x) = \delta(x)\,, \tag{4.14}$$

   for $-L/2 < x < L/2$. Let's see where this comes from. First, consider

$$\begin{aligned}\frac{1}{\sqrt{L}} \sum_{n=-k}^{n=k} \phi_n(x) &= \frac{1}{L} e^{-2\pi i k x/L}\left(1 + e^{2\pi i x/L} + \cdots + e^{2\pi i (2k)x/L}\right) \\ &= \frac{1}{L} e^{-2\pi i k x/L} \frac{1 - e^{2\pi i(2k+1)x/L}}{1 - e^{2\pi i x/L}} \\ &= \frac{1}{L} e^{-2\pi i k x/L} \frac{e^{\pi i(2k+1)x/L}}{e^{\pi i x/L}} \frac{\sin\left((2k+1)\pi x/L\right)}{\sin(\pi x/L)} \\ &= \frac{1}{L} \frac{\sin\left((2k+1)\pi x/L\right)}{\sin(\pi x/L)}\,. \end{aligned} \tag{4.15}$$

   To see that in the limit as $k \to \infty$ this behaves exactly like the Dirac delta-function, we see that for $-L/2 < x < L/2$,

$$\begin{aligned}&\int_{-L/2}^{L/2} dx \lim_{k\to\infty} f(x) \frac{1}{L} \frac{\sin\left((2k+1)\pi x/L\right)}{\sin(\pi x/L)} \\ &= \lim_{k\to\infty} \frac{1}{L} \int_{-(2k+1)\pi/2}^{(2k+1)\pi/2} dy\, f\left(\frac{yL}{(2k+1)\pi}\right) \frac{\sin y}{\sin(y/(2k+1))} \frac{L}{(2k+1)\pi} \\ &= \int_{-\infty}^{\infty} dy\, f(0) \frac{\sin y}{y} \frac{1}{\pi} \\ &= f(0)\,, \end{aligned} \tag{4.16}$$

   where we have made use of $\int_{-\infty}^{\infty} \sin(x)/x = \pi$. In fact, given that $\phi_n(x) = \phi_n(x + mL)$ for any integer m, we can see that over the entire real line,

$$\frac{1}{L} \sum_{m=-\infty}^{\infty} e^{2\pi i m x/L} = \sum_{m=-\infty}^{\infty} \delta(x - mL)\,. \tag{4.17}$$

We've checked that the functions $\phi_n(x)$ are orthogonal; we'll now show that they do indeed form a basis. First, given any arbitrary, periodic function $f(x)$ on $0 < x < L$, consider the set of coefficients

$$\tilde{f}_n = \int_{-L/2}^{L/2} dx\, \phi_n^*(x) f(x)\,, \tag{4.18}$$

for $n \in \mathbb{Z}$. Then

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} \tilde{f}_n \phi_n(x) &= \sum_{n=-\infty}^{\infty} \phi_n(x) \int_{-L/2}^{L/2} dy\, \phi_n^*(y) f(y) \\
&= \int_{-L/2}^{L/2} dy\, \frac{1}{\sqrt{L}} \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{L}} e^{2\pi i n(x-y)/L} f(y) \\
&= \int_{-L/2}^{L/2} dy\, \delta(x-y) f(y) \\
&= f(x)\,,
\end{aligned} \tag{4.19}$$

for any $-L/2 \leq x \leq L/2$, where we have used the completeness relation to trade the sum for the Dirac delta function. Clearly then, any $f(x)$ can be written as a sum of $\phi_n(x)$, demonstrating that $\{\phi_n(x)\}$ is indeed an orthonormal basis. Notice that $\tilde{f}_n = \langle \phi_n, f \rangle$, and so

$$f = \sum_n \langle \phi_n, f \rangle \phi_n\,, \tag{4.20}$$

which is exactly how you would write f as a linear combination of its basis elements: you 'project' f along each basis vector to find the component. c_n are known as the **Fourier coefficients** of f, and are in general complex.

(**End of Lecture**: Monday Nov 11 2024)

As an example, let's expand the square wave,

$$f(x) = \begin{cases} -1\,, & -\frac{1}{2} < x < 0\,, \\ 1\,, & 0 \leq x < \frac{1}{2}\,, \end{cases} \tag{4.21}$$

with $f(-1/2) = f(1/2) = 0$ to respect periodicity (although there is a discontinuity there). The Fourier coefficients are

$$\begin{aligned}
\tilde{f}_n &= \int_{-1/2}^{1/2} dx\, \phi_n^*(x) f(x) \\
&= -\int_{-1/2}^{0} dx\, e^{-2\pi i n x} + \int_0^{1/2} dx\, e^{-2\pi i n x} \\
&= \int_{1/2}^{0} dx\, e^{2\pi i n x} + \int_0^{1/2} dx\, e^{-2\pi i n x} \\
&= \int_0^{1/2} dx\, \left(e^{-2\pi i n x} - e^{2\pi i n x} \right) \\
&= -2i \int_0^{1/2} dx\, \sin(2\pi n x) \\
&= 2i \left(\frac{\cos 2\pi n x}{2\pi n} \right)_0^{1/2} \\
&= \frac{i}{n\pi} (\cos n\pi - 1) \\
&= -\frac{i}{n\pi} \begin{cases} 0\,, & n \text{ even}\,, \\ 2\,, & n \text{ odd}\,. \end{cases}
\end{aligned} \tag{4.22}$$

Therefore the Fourier series expansion for the square wave is

$$\begin{aligned} f(x) &= \frac{-2i}{\pi}\left(\cdots - \frac{1}{3}e^{-6\pi ix} - e^{-2\pi ix} + e^{2\pi ix} + \frac{1}{3}e^{6\pi ix} + \cdots\right) \\ &= \frac{4}{\pi}\left(\sin(2\pi x) + \frac{1}{3}\sin(6\pi x) + \cdots\right) \\ &= \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin(2\pi(2n+1)x)\,. \end{aligned} \tag{4.23}$$

Notice that the Fourier series contains only sines, since the function is odd about the origin. More generally, the Fourier series is a deconstruction of a function into constituent pure sine and cosine waves. The utility of such a procedure is immense. Here are some possible reasons why you might want to perform a Fourier decomposition, or a more general decomposition in other basis functions:

1. If the function is dominated by only a few terms in the series expansion, then you can *compress* the information stored by the function by just keeping the most relevant Fourier coefficients;

2. Often, you're interested in features on particular scales, e.g. we might only be interested long-distance behavior and not short-distance ones, and so writing the function as a sum of increasingly short wavelength modes allows you to disentangle this (the same is true for angles and spherical harmonics);

3. Certain mathematical operations are simpler when performed on the series expansion. We'll see several examples of this later on;

4. Often, you have detectors that are only sensitive to particular wavelengths, and so reconstructing an image can be done by going back from Fourier coefficients to the function in real space.

4.3 Fourier Transform

The Fourier transform is essentially the generalization of the Fourier decomposition method above as $L \to \infty$. We still have an inner product of functions

$$\langle f, g\rangle \equiv \int_{-\infty}^{\infty} dx\, f^*(x)g(x)\,, \tag{4.24}$$

and again we want to find an orthogonal basis for f. Notice that as L gets bigger, the basis functions from above, $\phi_n(x) = \exp(ikx)/\sqrt{L}$ with $k \equiv 2\pi n/L$, have values of k that become increasingly close together, with separation $\Delta k = 2\pi/L$. This suggests that the right set of basis functions in the limit as $L \to \infty$ is simply e^{ikx} for any $k \in \mathbb{R}$. We are choosing then to decompose functions into a linear combination of *plane waves*.

4.3.1 Definitions and higher dimensions

We therefore define the **Fourier transform** as

$$\tilde{f}(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx} f(x)\,, \tag{4.25}$$

where now $\tilde{f}(k)$ is *function* of k, and is the generalization of Fourier coefficients discussed previously. Sometimes you will also see the notation $\mathcal{F}f$ instead of

$\tilde{f}$. The first thing to note is that there are many conventions for the Fourier transform: you will see different choices for the sign in the exponent, as well as for the overall normalization of the transform. I am adopting what I think is the usual physics convention for *spacelike* Fourier transforms, but be careful![35]

Let's check for orthogonality and completeness of our new basis functions. First, we see that

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, e^{-ikx} &= \lim_{R\to\infty}\int_{-R}^{R} dx\, e^{-ikx} \\ &= \lim_{R\to\infty}\left[-\frac{e^{-ikx}}{ik}\right]_{-R}^{R} \\ &= 2\lim_{R\to\infty}\frac{\sin(kR)}{k} \end{aligned} \tag{4.26}$$

or

$$\int_{-\infty}^{\infty} dx\, e^{-ikx} = 2\pi\delta(k)\,, \tag{4.27}$$

which is the completeness relation for our basis functions or **Fourier modes**. This also gives us a way of writing the Dirac delta-function as an integral, which is a very powerful mathematical identity. From this expression, the orthogonality condition follows easily, since

$$\int_{-\infty}^{\infty} dx\, e^{ik'x} e^{-ikx} = 2\pi\delta(k'-k)\,. \tag{4.28}$$

Notice that instead of the Kronecker delta function, we have the Dirac delta function for orthogonality, which is the correct generalization when the dimension of the basis is uncountably infinite.

Because of these relations, we have

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ikx}\tilde{f}(k) &= \int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ikx}\int_{-\infty}^{\infty}dx' e^{-ikx'}f(x') \\ &= \int_{-\infty}^{\infty}dx'\, f(x')\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-ik(x'-x)} \\ &= \int_{-\infty}^{\infty}dx'\, f(x')\delta(x'-x) \\ &= f(x)\,, \end{aligned} \tag{4.29}$$

and therefore I can define an **inverse Fourier transform**, given by

$$f(x) = \int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ikx}\tilde{f}(k)\,. \tag{4.30}$$

This is sometimes denoted $\mathcal{F}^{-1}\tilde{f}$. The pair f and $\tilde{f}$ are called **Fourier transform pairs**.

Let's do a simple Fourier transform example. We'll find the Fourier transform of the rectangular function,

$$\mathrm{rect}(x) = \begin{cases} 1\,, & |x| < 1/2\,, \\ 0\,, & |x| \geq 1/2\,. \end{cases} \tag{4.31}$$

[35] For *timelike* transforms, the usual convention is $\tilde{f}(\omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t} f(t)$. I know this is confusing, and I'm sorry, but it does make sense: when both time and space are involved, the decomposition is into right-going waves for $\omega > 0$ and $k > 0$.

The Fourier transform is simply

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, e^{-ikx}\mathrm{rect}(x) &= \int_{-1/2}^{1/2} dx\, e^{-ikx} \\ &= \left.\frac{e^{-ikx}}{-ik}\right|_{-1/2}^{1/2} \\ &= \frac{2}{k}\sin\left(\frac{k}{2}\right)\,. \end{aligned} \tag{4.32}$$

We can define Fourier transforms in higher dimensions, which is simply the Fourier transform applied along each dimension, and so

$$\tilde{f}(\vec{k}) = \int_{-\infty}^{\infty} dz\, e^{-ik_n x_n}\cdots\int_{-\infty}^{\infty} dx\, e^{-ik_1 x_1} f(\vec{x}) = \int_{\mathbb{R}^n} d^n\vec{x}\, e^{-i\vec{k}\cdot\vec{x}} f(\vec{x})\,, \tag{4.33}$$

where $\vec{x} \equiv (x_1, \cdots, x_n)$, and $\vec{k} \equiv (k_1, \cdots, k_n)$. Similarly, the inverse Fourier transform is given by

$$f(\vec{x}) = \int_{\mathbb{R}^n} \frac{d^n\vec{k}}{(2\pi)^n}\, e^{i\vec{k}\cdot\vec{x}}\tilde{f}(\vec{k})\,. \tag{4.34}$$

Below, we will simply omit the label of $\mathbb{R}^n$ for the Fourier transform for simplicity.

The Dirac delta function can be generalized to n-dimensions, with

$$\int_{\mathbb{R}^n} d^n\vec{x}\, e^{-i\vec{k}\cdot\vec{x}} = (2\pi)^n\delta^n(\vec{x})\,, \tag{4.35}$$

with the expected property that

$$\int_{\mathbb{R}^n} d^n\vec{x}\, f(\vec{x})\delta^n(\vec{x}-\vec{y}) = f(\vec{y})\,. \tag{4.36}$$

4.3.2 Properties

Let's examine some basic properties of the Fourier transform:

1. It is *linear*, so that $\mathcal{F}(f+g) = \mathcal{F}f + \mathcal{F}g$, which follows immediately from the definition;

2. We can see that

$$\tilde{f}(0) = \int d^n\vec{x}\, f(\vec{x})\,, \tag{4.37}$$

   while

$$f(0) = \int \frac{d^n\vec{k}}{(2\pi)^n}\tilde{f}(\vec{k})\,. \tag{4.38}$$

3. For an even or odd function, where $f(-x) = \pm f(x)$, we see that

$$\begin{aligned}\tilde{f}(-\vec{k}) &= \int d^n\vec{x}\, e^{+i\vec{k}\cdot\vec{x}} f(\vec{x}) \\ &= \int d^n\vec{x}\, e^{-i\vec{k}\cdot\vec{x}} f(-\vec{x}) \\ &= \pm\tilde{f}(\vec{k})\,, \end{aligned} \tag{4.39}$$

where in the second last line we performed the change of variables $x \to -x$. This shows that if $f(x)$ is even/odd, then $\tilde{f}(k)$ is also even/odd respectively.

4. The complex conjugate of the Fourier transform $(\mathcal{F}f)^*$ is given by

$$(\mathcal{F}f)^*(\vec{k}) = \int d^n\vec{x}\, e^{+i\vec{k}\cdot\vec{x}} f^*(\vec{x}) = (\mathcal{F}f^*)(-\vec{k})\,. \tag{4.40}$$

Furthermore, if $f^* = f$ is a real function, then $(\mathcal{F}f)^*(\vec{k}) = \mathcal{F}f(-\vec{k})$.

5. A function that is more narrowly supported on the real axis will have a Fourier transform that has support over a larger range in Fourier space. To make this concrete, consider the function

$$f_s(x) = \frac{1}{\sqrt{s}} f\left(\frac{x}{s}\right)\,, \tag{4.41}$$

where I am free to adjust the value of s. One example could be a Gaussian, with a width given by the value of s; as s increases, the Gaussian widens. Then the Fourier transform of this is

$$\begin{aligned}\tilde{f}_s(k) &= \int dx\, e^{-ikx} \frac{1}{\sqrt{s}} f\left(\frac{x}{s}\right) \\ &= \int dy\, e^{-ikys}\sqrt{s} f(y) \\ &= \sqrt{s}\tilde{f}(sk)\,,\end{aligned} \tag{4.42}$$

where in the second last line we've made the substitution $y = x/s$. We can see here that as s increases, f gets wider, while the Fourier transform $\tilde{f}$ gets *narrower*, going in the opposite direction of f itself. This behavior is known as **Fourier reciprocity**, and is ultimately responsible for the uncertainty principle.

6. The Fourier transform *converts derivatives to multiplication*, making it a powerful tool for solving differential equations. For example, in one dimension,

$$\begin{aligned}\frac{d}{dx} f(x) &= \frac{d}{dx}\int \frac{dk}{2\pi} e^{ikx}\tilde{f}(k) \\ &= \int \frac{dk}{2\pi} e^{ikx}\cdot ik\tilde{f}(k)\,,\end{aligned} \tag{4.43}$$

and so $\mathcal{F}f' = ik(\mathcal{F}f)$. Similarly, for n-dimensions, we can determine that $\mathcal{F}(\nabla f) = i\vec{k}(\mathcal{F}f)$.

This last property is particularly useful for differential equation solving. As a simple example, consider the partial differential equation known as the *diffusion equation*

$$\frac{\partial^2\varphi}{\partial x^2} = \frac{1}{D}\frac{\partial\varphi}{\partial t}\,, \tag{4.44}$$

where we want to solve for $\varphi(x,t)$ for some initial conditions, with D being a constant. Let's perform a Fourier transform in the spatial coordinate to obtain $\tilde{\varphi}(k,t)$. Using the fact that derivatives become multiplication after the Fourier transform, we find

$$-k^2\tilde{\varphi}(k,t) = \frac{1}{D}\frac{\partial\tilde{\varphi}}{\partial t}\,. \tag{4.45}$$

For fixed k, we see clearly that

$$\log \tilde{\varphi} = -Dk^2 t + C \tag{4.46}$$

for some C an arbitrary constant, i.e.

$$\tilde{\varphi}(k,t) = A \exp\left(-Dk^2 t\right) \tag{4.47}$$

for some other arbitrary constant A set by initial conditions. We can now take the inverse Fourier transform to obtain $\varphi(x,t)$:

$$\begin{aligned} \varphi(x,t) &= \int \frac{dk}{2\pi} e^{ikx}\, A e^{-Dk^2 t} \\ &= \frac{A}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right), \end{aligned} \tag{4.48}$$

a result that you will prove in your problem set.

4.4 Plancherel's (or Parseval's) Theorem

If one views the the function f like a wave as a function of x, then $|f|^2$ is proportional to the energy density of the wave, just like e.g. $|E|^2$ in an electromagnetic wave is proportional to the energy density. In the Fourier mode picture, we would like to have a similar quantity that expresses the total energy density calculated in real space in terms of the Fourier modes. Let's try to see how this works by starting from the inner product of two functions f and g, and going to the Fourier domain:

$$\begin{aligned} \int_{-\infty}^{\infty} dx\, f^*(x) g(x) &= \int_{-\infty}^{\infty} dx \left[\int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \tilde{f}(k)\right]^* \int_{-\infty}^{\infty} \frac{dk'}{2\pi} e^{ik'x} \tilde{g}(k') \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{-ikx} \tilde{f}^*(k) \int_{-\infty}^{\infty} \frac{dk'}{2\pi} e^{ik'x} \tilde{g}(k') \\ &= \int_{-\infty}^{\infty} \frac{dk}{2\pi} \tilde{f}^*(k) \int_{-\infty}^{\infty} \frac{dk'}{2\pi} \tilde{g}(k') \int_{-\infty}^{\infty} dx\, e^{-i(k-k')x}. \end{aligned} \tag{4.49}$$

The innermost integral over x gives simply the Dirac delta function, and so

$$\int_{-\infty}^{\infty} dx\, f^*(x) g(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \tilde{f}^*(k) \int_{-\infty}^{\infty} \frac{dk'}{2\pi} \tilde{g}(k') 2\pi \delta(k-k'), \tag{4.50}$$

and integrating over k' gives us **Plancherel's Theorem** or **Parseval's Theorem**:[36]

$$\int_{-\infty}^{\infty} dx\, f^*(x) g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, \tilde{f}^*(k) \tilde{g}(k). \tag{4.51}$$

In particular, if we set $f = g$, then we find

$$\int_{-\infty}^{\infty} dx\, |f(x)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, |\tilde{f}(k)|^2. \tag{4.52}$$

Thus, if f like the energy density, the integral on the left is the total energy; on the right, we therefore have $\tilde{f}$ representing the energy per Fourier mode, with the integral also giving the total energy. Plancherel's or Parseval's theorem

[36] I could not figure out which name is actually the right one to use. Opinions differ.

[37] You might be worried about the $1/2\pi$, but that just is due to the convention that we have chosen for the Fourier transform. There are conventions where the factor is distributed as $1/\sqrt{2\pi}$ on both sides, which is more aesthetically pleasing, but doesn't jive with our usual picture of $2\pi/k$ being the inverse of the wavelength.

[38] Again, up to the factor of 2π.

therefore expresses some form of *conservation of energy*.[37] It also shows that the Fourier transform preserves the inner product, i.e. is *unitary*,[38] with

$$\langle f, g\rangle = \frac{1}{2\pi}\langle \tilde{f}, \tilde{g}\rangle \,. \tag{4.53}$$

(**End of Lecture**: Wednesday Nov 13 2024)

4.5 Convolution Theorem

The **convolution** of two functions f and g is defined as

$$(f * g)(x) \equiv \int_{-\infty}^{\infty} dy f(y) g(x-y) \,. \tag{4.54}$$

Intuitively, if you imagine f and g to be concentrated in a particular region the real line, the convolution involves reflecting the function g about the origin, and then "sliding" the function g over the function f from left to right as x increases. The function g acts as an "window" that performs an average over f, and as you slide that window around, you're averaging over different points on the real line.[39] Averaging an array over its nearest neighbors super common mathematical operation in physics and engineering, making the convolution worth studying in detail. The convolution is easily generalized into $\mathbb{R}^n$, where you should imagine the procedure as taking an average over some n-dimensional region around each point.

[39] See the Wikipedia article on convolution for a nice illustration of this.

Some properties of the convolution are as follows:

1. It is commutative, i.e. $f * g = g * f$, since

$$\begin{aligned}(g * f)(x) &= \int_{-\infty}^{\infty} dy\, g(y) f(x-y) \\ &= \int_{-\infty}^{\infty} d\xi\, g(x-\xi) f(\xi)\,, \end{aligned} \tag{4.55}$$

   where I've simply made a change of variables of $\xi = x - y$ in the last line.

2. It is associative, i.e. $(f*g)*h = f*(g*h)$, and so we don't have to worry about how to put parentheses around convolutions. This comes from the fact that you can always flip the order of integrals.

3. $f*\delta = f$, where δ represents the Dirac delta function. This is the definition of how the Dirac delta function behaves.

As an example, let's consider the convolution of the rectangular function,

$$\mathrm{rect}(x) = \begin{cases} 1\,, |x| < 1/2\,, \\ 0\,, |x| \geq 1/2\,, \end{cases} \tag{4.56}$$

with itself. This can be written as

$$\begin{aligned}(\mathrm{rect} * \mathrm{rect})(x) &= \int_{-\infty}^{\infty} dy\, \mathrm{rect}(y)\mathrm{rect}(x-y) \\ &= \int_{-1/2}^{1/2} dy\, \mathrm{rect}(x-y)\,. \end{aligned} \tag{4.57}$$

The integrand is only nonzero when

$$-\frac{1}{2} \leq x - y \leq \frac{1}{2} \implies x - \frac{1}{2} \leq y \leq x + \frac{1}{2}\,. \tag{4.58}$$

Therefore, there are five scenarios:

1. if $x + 1/2 < -1/2$ or $x < -1$, then the integral is zero;
2. if $x - 1/2 < -1/2 < x + 1/2 < 1/2$, i.e. $-1 < x < 0$, then the integral is $(x + 1/2) - (-1/2) = x + 1$;
3. if $x = 0$, then the integral is 1;
4. if $-1/2 < x - 1/2 < 1/2 < x + 1/2$, i.e. $0 < x < 1$, then the integral is $1/2 - (x - 1/2) = 1 - x$;
5. and finally if $1/2 < x - 1/2$, we have zero for the integral again.

We can see clearly that $\mathrm{rect} * \mathrm{rect} = \mathrm{tri}$, the triangular function, compactly defined as

$$\mathrm{tri}(x) = \begin{cases} 1 - |x|\,, & |x| < 1\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{4.59}$$

As you can see, convolutions are annoying to compute; they are also annoying numerically, since it involves sliding a window and taking averages across some space. But let's see what happens when we take the Fourier transform of a convolution:

$$\begin{aligned} \int_{-\infty}^{\infty} dx\, e^{-ikx} (f * g)(x) &= \int_{-\infty}^{\infty} dx\, e^{-ikx} \int_{-\infty}^{\infty} dy\, f(y) g(x - y) \\ &= \int_{-\infty}^{\infty} dy\, f(y) \int_{-\infty}^{\infty} d\xi\, e^{-ik(\xi + y)} g(\xi) \\ &= \int_{-\infty}^{\infty} dy\, e^{-iky} f(y) \int_{-\infty}^{\infty} d\xi\, e^{-ik\xi} g(\xi)\,. \end{aligned} \tag{4.60}$$

This gives the **convolution theorem**, and its converse, which can proven by similar means:

$$\begin{aligned} \mathcal{F}(f * g)(k) &= (\mathcal{F} f(k))(\mathcal{F} g(k))\,, \\ \mathcal{F}(fg)(k) &= \frac{1}{2\pi} (\mathcal{F} f * \mathcal{F} g)(k)\,. \end{aligned} \tag{4.61}$$

With this theorem, to perform a convolution between f and g, you can first perform a Fourier transform on each to get $\tilde{f}$ and $\tilde{g}$, multiply them together, and then perform the inverse Fourier transform; by the convolution theorem, we see that this is simply $f * g = \mathcal{F}^{-1}(\mathcal{F} f \cdot \mathcal{F} g)$. This is often numerically a much better option, since numerical Fourier transforms are blazingly fast (and so are products).

Let's use this theorem to check that $\mathrm{rect} * \mathrm{rect} = \mathrm{tri}$, which will also serve as a useful example of how to calculate Fourier transforms. We saw earlier in Eq. (4.32) that $\mathcal{F}(\mathrm{rect}) = (2/k)\sin(k/2)$. Thus, we expect the inverse Fourier transform of $(2/k)^2 \sin^2(k/2)$ to be given by tri. Let's show this. We have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \frac{4}{k^2} \sin^2\left(\frac{k}{2}\right) &= \frac{2}{\pi} \int_{-\infty}^{\infty} dk\, \frac{e^{ikx}}{k^2} \left(\frac{1 - \cos k}{2}\right) \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} dk\, \frac{e^{ikx}}{k^2} - \frac{1}{\pi} \int_{-\infty}^{\infty} dk\, \frac{e^{ikx}}{k^2} \cos k\,. \end{aligned} \tag{4.62}$$

First, let's evaluate the first integral. To do this, let's first consider the integral

$$\int_{-\infty}^{\infty} dk\, \frac{e^{ikx}}{k}\,, \tag{4.63}$$

and exploit the fact that the Fourier transform allows us to write derivatives as products:

$$\frac{d}{dx}\int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k^2} = i\int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k} \tag{4.64}$$

But from the work we had done to get the results hown in Eq. (3.175), i.e.

$$\int_{-\infty}^{\infty} dk\,\frac{e^{ik}}{k} = i\pi\,, \tag{4.65}$$

we can deduce that

$$\int_{-\infty}^{\infty} dk\frac{e^{ikx}}{k} = i\pi\,\mathrm{sgn}(x)\,, \tag{4.66}$$

which you can check by making the substitution $y = kx$.

And so, we find

$$\frac{d}{dx}\int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k^2} = -\pi\,\mathrm{sgn}(x) \implies \int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k^2} = -\pi x\,\mathrm{sgn}(x) = -\pi|x|\,. \tag{4.67}$$

Notice that there is no constant of integration, which you can check is correct by setting $x = 0$ of both sides: on the left-hand side, the principal value of the integral is indeed equal to zero.

The next step is to note that

$$\begin{aligned}\int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k^2}\cos k &= \frac{1}{2}\int_{-\infty}^{\infty} dk\,\frac{e^{ikx}}{k^2}\left(e^{ik}+e^{-ik}\right)\\ &= \frac{1}{2}\int_{-\infty}^{\infty} dk\,\frac{e^{ik(x+1)}}{k^2} + \frac{1}{2}\int_{-\infty}^{\infty} dk\,\frac{e^{ik(x-1)}}{k^2}\\ &= -\frac{1}{2}\pi|x+1| - \frac{1}{2}\pi|x-1|\,.\end{aligned} \tag{4.68}$$

Putting everything together, we find

$$\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{ikx}\frac{4}{k^2}\sin^2\left(\frac{k}{2}\right) = -|x| + \frac{1}{2}|x+1| + \frac{1}{2}|x-1|\,, \tag{4.69}$$

which you can check is simply $\mathrm{tri}(x)$, as required.

4.6 The Radon Transform

As a parting example of the power of the Fourier transform, let's discuss an idea that is used very frequently in tomographic imaging: the **radon transform**.

Imagine you want to make a 2D map of the density $\rho(\vec{r})$ of some disc-like object, say a cross section through the human body, but you can only do so by shining light along the plane of the disc. You can move your light source perpendicular to the direction the light is shining, ξ, and you can rotate the light source about the disc by some angle, ϕ. The absorption of the light passing through the disc is proportional to $\rho(\vec{r})$. Can we reconstruct the density map by imaging this way?

The total loss of light at some ξ and some ϕ can be written as (see Fig. 20 for the set-up)

$$a(\xi,\phi) = c\int d^2\vec{r}\,\rho(\vec{r})\delta^2(\vec{r}\cdot\hat{e}_r - \xi)\,, \tag{4.70}$$

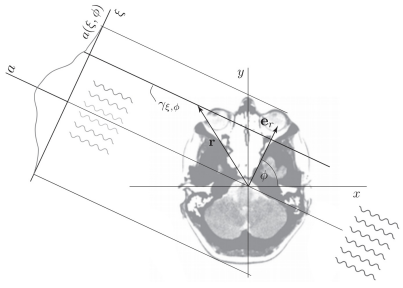

Figure 20: Set-up for tomography, and the coordinate system being used. Figure taken from Altland and von Delft.

where $\hat{e}_r$ is the direction along which the light source can slide. c is some proportionality constant relating the light lost to the density. Let's perform a Fourier transform of the ξ coordinate now. We obtain

$$\begin{aligned}\tilde{a}(k,\phi) &= c\int d^2\vec{r}\,\rho(\vec{r})\int dx\, e^{-ik\xi}\delta^2(\vec{r}\cdot\hat{e}_r-\xi)\\ &= c\int d^2\vec{r}\,\rho(\vec{r})e^{-ik(\vec{r}\cdot\hat{e}_r)}\,. \end{aligned} \tag{4.71}$$

From the last expression, we see that

$$\tilde{a}(k,\phi) = c\tilde{\rho}(k\hat{e}_r) \equiv c\tilde{\rho}(\vec{k})\,, \tag{4.72}$$

where we have defined $\vec{k} \equiv k\hat{e}_r$, with $\hat{e}_r = (\cos\phi, \sin\phi)$, and therefore $\vec{k} = (k\cos\phi)\hat{x} + (k\sin\phi)\hat{y}$. And so we find that

$$\begin{aligned}\rho(\vec{r}) &= \int \frac{d^2\vec{k}}{(2\pi)^2}\, e^{i\vec{k}\cdot\vec{r}}\tilde{\rho}(\vec{k})\\ &= \frac{1}{c}\int \frac{d^2\vec{k}}{(2\pi)^2}\, e^{i\vec{k}\cdot\vec{r}}\tilde{a}(k,\phi)\\ &= \frac{1}{(2\pi)^2 c}\int_0^\infty dk\, k\int_0^{2\pi} d\phi\, e^{i\vec{k}\cdot\vec{r}}\tilde{a}(k,\phi)\,. \end{aligned} \tag{4.73}$$

Somewhat miraculously (at least to me), we can recover the full, 2D density map by performing side-on imaging! This of course underpins essential imaging technologies today like medical tomography and electron microscopy.

5 Statistics

In classical physics, we're used to trying to model everything as *deterministic*; for example, we derive the trajectory of a projectile, or the electric field due to a point charge, and these things are known exactly with infinite precision. However, *even in theoretical physics*, this is not the right approach to a significant portion of problems:

1. You often want to understand how to characterize systems about which you have *incomplete information*, e.g. a box containing a mole of gas. This system contains $\sim 10^{23}$ particles, and it would be hopeless to track what each individual particle (known generically as *microstates*) are doing. But given some information about the whole ensemble (or *macrostate*), we hope to be able to say something about the system. This is the idea behind *statistical mechanics*;

2. Even more fundamental than that is the fact that nature itself appears to be *inherently random*. Performing a measurement of some physical property of a system appears to give nondeterministic outcomes in *quantum mechanics*.

So as you can see, randomness is fully baked into modern physics! Of course, once you include *experimental uncertainties*, things get even fuzzier. When you perform a measurement in a laboratory, while the measurement itself is obviously important, what is equally important is quantifying how well you have done that measurement, and in fact in many fields, that's the part that takes lots of work. In this case, *data* has some intrinsic randomness that we must account for when testing our models.

5.1 Introduction to Probability

We are now going to study how to quantify randomness. We won't spend any time building the formal framework of probability, but instead trust our intuition on this subject. Consider a range of outcomes that span all the possibilities of random variation we call S, the sample space. For a six-sided die, for example, $S = \{1, 2, 3, 4, 5, 6\}$. To every subset $A \subset S$, we can assign what we call a **probability** $P(A)$, which satisfies the following axioms:

1. For every subset $A \subset S$, $P(A) \geq 0$, i.e. probability cannot be negative.

2. For any two subsets that are disjoint, i.e. $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

3. The probability assigned to the entire sample space is $P(S) = 1$.

From these axioms, one can prove several straightforward, intuitive results, such as $P(\overline{A}) = 1 - P(A)$, where $\overline{A}$ is the complement of A, and $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, which is obvious when you draw a Venn diagram. Thus, for a fair die, one has $P(\{1, 2, 3\}) = 1 - P(\{4, 5, 6\}) = 0.5$, $P(\{1, 2, 3\}) = P(\{1, 2\}) + P(\{2, 3\}) - P(\{2\})$, and so on.

5.1.1 Conditional probability, independence and Bayes' theorem

Let's build on the basics to explore some fundamentally important (and at times confusing) concepts in probability.

Setting aside quantum mechanics, the probability that you assign to something is dependent on what information you have. This seems like an almost obvious statement, but is fundamental to understanding probability. For example, the probability that the temperature outside is zero degrees fahrenheit

should generally be assigned a low probability, but the probability that the temperature is zero degrees fahrenheit *given that you're at the South Pole in June* would raise that probability significantly. In the context of our more formal description of probability above,

The **conditional probability** $P(A|B)$ is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,. \tag{5.1}$$

For example, going back to our example of the die,

$$P(\{1\}|\{1,2\}) = \frac{P(\{1\} \cap \{1,2\})}{P(\{1,2\})} = \frac{P(\{1\})}{P(\{1,2\})} = 0.5\,. \tag{5.2}$$

In words, even though the probability of rolling a 1 is $1/6$, the probability of having rolled a 1 *given that we've rolled a 1 or a 2* goes up to 0.5. One way to understand this formula is that by taking B to be given, the sample space shrinks to B; A happening in this more restricted sample space is given by $A \cap B$, and the probability of A given B is therefore $\propto P(A \cap B)$, with the right normalization being $P(B)$.

Suppose we can split up the sample space S into disjoint parts B_i, so that $\cup_i B_i = S$. Then from the definition of conditional probability, we obtain the **law of total probability**, given by

$$\sum_i P(A|B_i)P(B_i) = \sum_i P(A \cap B_i) = P(A)\,. \tag{5.3}$$

The last equality follows because all of the B_i form the full sample space and are disjoint, and so adding up all the disjoint intersections $A \cap B_i$ gives A.

Another important concept is **independence**. Let's consider two rolls of the die, with the first roll being a 6. What is the probability of the second roll being a 6 again, given that the first roll was a 6? Well, in a fair die, that's just the same probability as any roll being a 6; the fact that the first roll was a 6 *gave you no new information*. More formally,

A and B are **independent** if

$$P(A|B) = P(A)\,, \tag{5.4}$$

or equivalently,

$$P(A \cap B) = P(A)P(B)\,. \tag{5.5}$$

You can check that $P(A|B) = P(A)$ also implies that $P(B|A) = P(B)$, so that independence is always mutual.

Although easy to define mathematically, trying to understanding if two possible outcomes of a single random variable are independent or not can be the cause of a lot of statistical heartache. We won't have much to say about this in this class unfortunately.

Notice from the definition of conditional probability, we have both

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,, \tag{5.6}$$

as well as

$$P(B|A) = \frac{P(B \cap A)}{P(A)}\,. \tag{5.7}$$

Putting these results together, we have at once a simple but also one of the most profound statements in probability and statistics, known as **Bayes' theorem**:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}\,. \tag{5.8}$$

There's a lot to be said about this result, and we'll come back to it several times in this course. First, let's look at two very famous examples. Both of these examples really show how counterintuitive probability can be!

Imagine you've developed a test for a disease which afflicts 0.1% of the population, so that for every person,

$$\begin{aligned} P(\text{disease}) &= 0.001\,, \\ P(\text{no disease}) &= 0.999\,. \end{aligned} \tag{5.9}$$

Given that a patient has the disease, the test gives a positive result 98% of the time, i.e. the false negative rate is 2%, or

$$\begin{aligned} P(+|\text{disease}) &= 0.98\,, \\ P(-|\text{disease}) &= 0.02\,. \end{aligned} \tag{5.10}$$

On the other hand, given that a patient does *not* have the disease, the test returns a positive result 3% of the time, i.e. the false positive rate is 3%:

$$\begin{aligned} P(-|\text{no disease}) &= 0.97\,, \\ P(+|\text{no disease}) &= 0.03\,. \end{aligned} \tag{5.11}$$

Now suppose a patient gets a positive result. What is the probability that this patient actually has the disease? This is the quantity $P(\text{disease}|+)$. Instinctively, you might think that the probability is quite high. But a direct use of Bayes' theorem gives

$$\begin{aligned} P(\text{disease}|+) &= \frac{P(+|\text{disease})P(\text{disease})}{P(+)} \\ &= \frac{P(+|\text{disease})P(\text{disease})}{P(+|\text{disease})P(\text{disease}) + P(+|\text{no disease})P(\text{no disease})} \\ &= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} \\ &= 3.2\%\,, \end{aligned} \tag{5.12}$$

where in the second line, I've made use of the law of the total probability. So, even though the false positive and false negative rates are so low, the probability of the patient actually having the disease is still pretty low, at 3.2%! The ultimate reason for this is that the *prevalence of the disease is thought to be really low in the first place*, and so a positive test result isn't strong enough evidence to offset your *prior* belief that the patient is just unlikely to have the disease. More on this later.

The next example we'll review is the famous **Monty Hall problem**, which stumped even the great Paul Erdős. You're a contestant in a game show, standing in front of three closed doors. Behind two of the doors are goats, and

[40] Presumably you want the car; no judgment if you'd prefer a goat.

behind one of them is a car; you're asked to pick a door, and you get to bring whatever is behind the door home.[40] Without loss of generality, let's say you picked door 1. At this point, you have a $2/3$ chance of getting a goat, and $1/3$ chance of getting a car. But now comes the twist: the host of the show opens door 3 (again without loss of generality), and offers you one last chance to switch doors. Should you?

This problem is really fascinating, because for some reason, for most people, the overwhelming intuition is to say that it doesn't matter. *But switching is actually the better strategy*, with a probability of winning given by $2/3$, as compared to not switching, which is $1/3$. The most straightforward way to understand this is to see that if you had picked a goat initially (a possibility with probability $2/3$), then switching wins; if you had picked the car (a possibility with probability $1/3$), then switching loses. But let's be a little more careful, and calculate the probability $P(\text{switching wins}|\text{host opens door 3})$. This is, from the definition of conditional probability,

$$P(\text{switching wins}|\text{host opens door 3}) = \frac{P(\text{switching wins} \cap \text{host opens door 3})}{P(\text{host opens door 3})}\,. \quad (5.13)$$

There are three possibilities: CGG, GCG and GGC, where 'C' stands for car, and 'G' stands for goat, each possibility with a probability of $1/3$. The scenario where switching wins *and* the host opens door 3 corresponds only to GCG, which occurs with probability $1/3$. On the other hand, the probability that the host opens door 3 corresponds to GCG, *and* CGG, followed by the host choosing door 3 over door 2. If the host is as likely to choose door 2 as door 3, then $P(\text{host opens door 3}) = 1/3 + 1/6 = 1/2$. Therefore,

$$P(\text{switching wins}|\text{host opens door 3}) = \frac{1/3}{1/3 + 1/6} = \frac{2}{3}\,, \quad (5.14)$$

as before. Note however that if the host has a preference for either door 2 or door 3, things change. If the host always chooses door 2 over door 3 when possible, for example, then switching always wins; the fact that the host chose to open door 3 means that door 2 was unavailable, because it was the car! This is a perfect example of how *probability is always about how much information you have, and having more information allows you to update your prior information*.

5.1.2 Frequentist vs. Bayesian statistics

We'll take a brief pause now to dive into how probabilities are used in physics and the two main ways of thinking about data and models which you might have already heard about at some point. We'll revisit this discussion frequently in this topic, but for now, I want to introduce some ideas and terminology early on so that we get used to them.

One of the goals of experimental physics is to collect data to better understand or test models of a system. But the data that one collects is necessarily probabilistic, due to measurement uncertainty; **statistics** is concerned with the analysis and interpretation of data in light of our models. We use the data—which is probabilistic in nature—to perform **statistical inference** on the model, i.e. *determine the values of particular parameters in the model with some degree of confidence from the data*.

There are two broad approaches that one can take to perform statistical inference: the **frequentist** or **classical** approach, or the **Bayesian** approach. The debate about the whats and the whys of this great divide is massive, and much cleverer people than me have made their much better informed opinions

known. I can, however, only give you my not super expert interpretation of what the difference between these two approaches, and why you might choose one over the other.

Let me start with the *frequentist* or *classical* point of view, which is the more traditional approach to statistics, and which will be the main focus of this chapter.

1. In the frequentist approach, the model is thought of as an immutable thing, with absolutely true values of the parameter that are fixed and infinitely precise. You may not know what they are, but they are certainly real and in existence.

2. The only quantity that you are ever concerned with in life as a frequentist is the **likelihood**, which is $P(\text{data}|\text{model})$. With the likelihood you can answer questions such as, "how consistent is the data with respect to the model?" or "given the model, how likely was it to have obtained the data that you did?".

3. The way you should view the data that you have collected is that it is one of many possible outcomes in N possible universes. You can test a model (also known as **hypothesis testing**) by saying, "given the model is true, there was an x% chance of getting data at the level of inconsistency with the model that I did", and then choosing to reject the model if x is less than some number (a common but arbitrary choice is 5%).

4. If you're trying to infer the value of a parameter θ in the model, in each of the N universes, you would construct what is called a **confidence interval** for θ, in such a way that in some percentage of these N universes (a common choice is 90% or 95%), the true value of θ lies in the interval constructed.

Again, I want to emphasize that in the frequentist point of view, the model is the model is the model. If you ever say something like 'the probability of the model...' or 'the probability of the model parameter being this value is...' or 'the parameter lies inside the confidence interval 95% of the time' or something like that, where the model is mutable and parameter values can move around, you're saying something completely blasphemous and nonsensical to a frequentist.

The big advantage of the frequentist approach is that it is rigorous and sounds quite objective: you pick a model, and you check how consistent the data is with the model, or you construct the confidence interval which should contain the true model parameter some percent of the time. The problem with the frequentist approach is that it's not *quite* answering the questions that you're interested in. Let's think back to the example we went through earlier with the medical test for a disease. A frequentist approach would be as follows: model the patient as having no disease, and run the test to obtain the data. If the test returns a positive result, since the *likelihood* of the data $(+)$ given the model ('no disease') is only 3%, we might reject the hypothesis in favor of believing the patient has a disease. But, as we saw in the same example, if the probability of having the disease is just very low in general, $P(\text{disease}|+)$ is also going to be very low, and we would very likely have been led to a false positive had we followed the frequentist approach.

The main problem with the frequentist approach is that it only deals with the likelihood, $P(\text{data}|\text{model})$ (e.g. $P(+|\text{no disease})$), when really scientists are often more interested in $P(\text{model}|\text{data})$ (e.g. $P(\text{disease}|+)$), a quantity we call the **posterior probability**. The Bayesian approach is to note that Bayes' theorem

says that

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}\,, \tag{5.15}$$

i.e. that the posterior probability $P(\text{model}|\text{data})$ proportional to the likelihood $P(\text{data}|\text{model})$ times $P(\text{model})$, a quantity we call the **prior**.[41] The Bayesian approach then tells us to calculate the likelihood as in the frequentist case, but then multiply that probability by *the probability of the model being right*, to get the quantity that you truly care about: a probability of the model parameters having a certain value, given the data observed.

This is complete anathema to a frequentist, of course, but in some cases, it is a very natural thing to do. In the medical test example, for example, if you had some *prior information or belief* about the prevalence of the disease $P(\text{disease})$, then you can use it as a prior to get the posterior. The Bayesian approach is particularly useful here, when you have strong prior information or beliefs that you want to include in the analysis. The problem, however, is that you're forced to add a huge dose of *subjectivity* to your analysis—the prior can literally be anything that you want. There are ways to try to be as prior agnostic as possible, usually by using priors that are flat over very large ranges of different model parameters; but you always have to be careful that your choice of priors isn't dominating the inferred posterior in ways that you didn't intend, which can be tricky to determine.

How do we choose which approach to use? The answer in my experience is usually one of practicality: Bayesian methods are usually computationally simpler, explaining its popularity in e.g. cosmology where the number of parameters in the model is generally very large. However, there are notable examples of studies with posteriors that are unwittingly dominated by a choice of priors, and so the tide appears to be turning a little toward frequentist methods, at least in the field of cosmology. I generally adopt frequentist approaches when they are computationally tractable, and Bayesian approaches when they're not. Much more on this as we go along in the course, but I'll leave you with a nice quote that explains the frequentist vs. Bayesian divide:

> *"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentists use impeccable logic to deal with an issue that is of no interest to anyone."* – Louis Lyons

(**End of Lecture**: Monday Nov 18 2024)

[41] $P(\text{data})$ is known as the **evidence**, which is interesting in its own right, but in practice we don't need it to compute the posterior; the proportionality is enough, since we can just normalize everything at the end.

5.1.3 Probability density functions

So far we have been paying attention to the sample space, and the probability of various subsets of the sample space. In practice, we are usually concerned with the value of a certain *random variable* X, and the probability that X takes on certain values. For the six-sided die, for example, X would be the value of the die after a roll, and we may be interested in the probability of $P(X = x)$, for example. This is an example of a *discrete* random variable. Much of the previous discussion centered on this case. Sometimes, however, the random variable is *continuous*, e.g. the height of a randomly selected person. In that case, we define the following:

The **probability density function** $f(x)$ (or PDF for short) is defined via

$$P(x \leq X \leq x + dx) = f(x)\,dx\,. \tag{5.16}$$

The probability to find $a \leq X \leq b$ is therefore given by

$$P(a \leq X \leq b) = \int_a^b dx\, f(x)\,. \tag{5.17}$$

By definition, if we integrate over the entire sample space, then

$$\int_S dx\, f(x) = 1\,, \tag{5.18}$$

i.e. $f(x)$ must be properly **normalized**.

We are often also interested in e.g. $P(x \leq b)$, with no interest in the lower limit. To do this, we define the following:

The **cumulative distribution function** $F(x)$ (or CDF for short) is defined as

$$F(x) = \int_{-\infty}^{x} dx'\, f(x')\,, \tag{5.19}$$

with $P(X < a) = F(a)$.

You're probably already familiar with the **quantile** (or **percentile**, if we express it as a percentage): the quantile of order α is simply the value x_α at which $F(x_\alpha) = \alpha$, or $x_\alpha = F^{-1}(\alpha)$.[42] One of the most common quantiles that we use is the **median**, which is defined as $x_{1/2} = F^{-1}(1/2)$.

[42] Note that $F(x)$ is always invertible, since it must be a monotonically increasing function, since probabilities are always positive.

We can of course have multidimensional random variables, e.g. the distribution of momenta in an ideal gas, which has a sample space that is 3D. The generalization is pretty straightforward: in 2D, for example, we can define $f(x, y)$ the **joint PDF** for two random variables X and Y as

$$f(x,y)\,dx\,dy = P(x \leq X \leq x + dx \text{ and } y \leq Y \leq y + dy)\,. \tag{5.20}$$

This should be viewed as the continuous random variables version of $P(A \cap B)$. As before, the joint PDF should be normalized:

$$\int_{S_x} dx \int_{S_y} dy\, f(x,y) = 1\,, \tag{5.21}$$

where S_x and S_y are the allowed values of x and y respectively. From here on out, we'll drop the notation S_x and S_y when it is obvious from context. Note that it is often the case that X and Y are *not independent*, so that certain values of X are more favored given certain values of Y, and vice versa.

With multiple variables, we are frequently interested in the distribution of *just one variable*. In a 2D PDF, you can imagine drawing narrow, vertical strips between x and $x + dx$, and just summing up the probability in the strip, leaving just a function of a single variable x. Mathematically, we call this the **marginal PDF** for x, $f_X(x)$, and is defined by

$$f_X(x) = \int dy\, f(x,y)\,. \tag{5.22}$$

We often call this *marginalizing over* Y, or *integrating out* Y. We can of course marginalize over x instead, to obtain

$$f_Y(y) = \int dx\, f(x,y)\,. \tag{5.23}$$

We can also discuss the conditional probability of continuous variables. We know from our previous discussion of conditional probability that

$$\begin{aligned} P(Y = y|X = x_0) &= \frac{P(Y = y \text{ and } X = x_0)}{P(X = x_0)} \\ &= \frac{f(x_0, y)\,dx\,dy}{f_X(x_0)\,dx}\,, \end{aligned} \tag{5.24}$$

where I have used x_0 to emphasize that we are thinking of fixing X to that particular value, and viewing y as a variable. We can therefore define the **conditional PDF** $h(y|X = x_0)$ as

$$h(y|X = x_0) = \frac{f(x_0, y)}{f_X(x_0)} = \frac{f(x_0, y)}{\int dy\, f(x_0, y)}\,. \tag{5.25}$$

The last equality gives a nice interpretation: to find the conditional PDF of y when $X = x_0$, just take $f(x_0, y)$, but then normalize it correctly over all values of y. The intuitive picture on the 2D distribution is to draw a line cutting through the distribution corresponding to $x = x_0$, and normalizing the resulting distribution so that the total integral is 1.

From the definition of the conditional PDF, we can also find the following relation between the marginal PDF $f_Y(y)$ and the conditional PDF $h(y|X = x)$:

$$f_Y(y) = \int dx\, f(x, y) = \int dx\, h(y|X = x) f_X(x)\,. \tag{5.26}$$

This is the equivalent of the law of total probability: to get the marginal probability, we integrate the conditional probability over all possible conditions, weighted by the probability of the condition itself.

Finally, we note that two continuous random variables are *independent* if

$$f(x, y) = f_X(x) f_Y(y)\,, \tag{5.27}$$

which is the equivalent of $P(A \cap B) = P(A)P(B)$. Intuitively, the probability near the point (x, y) is determined separately by the probability of $X = x$ multiplied by the probability of $Y = y$ in a little patch around the point, with no crosstalk between the two variables.

5.1.4 Functions of random variables

So far we have considered random variables and their probability density functions. Another common thing that we do is to construct *functions* of random variables; for example, given random variables X, Y and Z corresponding to the x, y and z coordinates of a particle in a gas, we might be interested in the random variable $R = \sqrt{X^2 + Y^2 + Z^2}$, which is the distance of a particle from the origin. We're in particular interested in getting the PDF of functions of random variables.

Suppose we have a random variable X, and a function of the random variable $a(X)$, and suppose a is invertible. Then consider the PDF of X, $f(x)$, which is given by $f(x)\,dx = P(x < X < x + dx)$. Then the PDF of $a(X)$, $g(a)$, should satisfy $f(x)\,dx = g(a)|da|$, where $da = a'(x)\,dx$. The absolute value is so that the signs work out, since $dx > 0$. In other words,

> The PDF of a function $a(X)$ of a random variable X is given by
>
> $$g(a) = f(x)\left|\frac{dx}{da}\right|\,. \tag{5.28}$$

In higher dimensions, the derivative is simply replaced by a Jacobian, using the same argument. If $a(X)$ is not invertible, then you have to be a bit more careful, but the idea is to split up x into several domains in which you can actually invert $a(X)$, and add up the contributions. For example, if $a(X) = X^2$, we have $|dx/da| = 1/2\sqrt{a}$, and

$$g(a) = \frac{1}{2\sqrt{a}} f(\sqrt{a}) + \frac{1}{2\sqrt{a}} f(-\sqrt{a}) \,. \tag{5.29}$$

5.2 Expectation Values

5.2.1 Definitions

The **expectation value** of a random variable X, variously written as $E[X]$, $\mathbb{E}(X)$ or $\langle x \rangle$, is defined for continuous random variable as

$$\mathbb{E}(X) = \int dx \, x f(x) \,. \tag{5.30}$$

The expectation value is often denoted by μ, and is also known as the **mean**. For a discrete random variable, the expectation value is simply

$$\mathbb{E}(X) = \sum_x x P(X = x) \,, \tag{5.31}$$

which is an obvious parallel to continuous variables, and so we'll only discuss continuous variables in this section.

Besides the mean, we are often interested in

$$\mathbb{E}(X^n) = \int dx \, x^n f(x) \,, \tag{5.32}$$

which is known as the algebraic moment, raw moment, or simply the **moment**. Another quantity that is often discussed is the nth **central moment**, which is defined as

$$\mathbb{E}[(X - \mathbb{E}(X))^n] = \int dx \, (x - \mu)^n f(x) \,. \tag{5.33}$$

Of particular importance is the second central moment;

This is known as the **variance** of the distribution, defined as

$$\mathrm{Var}(X) = \mathbb{E}[(x - \mathbb{E}(x))^2] = \int dx \, (x - \mathbb{E}(x))^2 f(x) \,, \tag{5.34}$$

and is often also denoted as σ^2.

In addition, the square root of the variance is known as the **standard deviation** of X, denoted by σ. Note that (switching up notations just so that you're familiar with all the ways people write these things)

$$\begin{aligned} \mathrm{Var}(X) = \langle (x - \langle x \rangle)^2 \rangle &= \langle x^2 - 2x\langle x \rangle \rangle + \langle x \rangle^2 \\ &= \langle x^2 \rangle - 2\langle x \rangle^2 + \langle x \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2 \,, \end{aligned} \tag{5.35}$$

which is a useful formula for computing the variance. More generally,

$$\mathbb{E}(a(X)) = \int dx \, a(x) f(x) \,. \tag{5.36}$$

For n random variables $X_1, \cdots, X_n$, we can define the **covariance matrix**

$$V_{ij} \equiv \text{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]\,, \tag{5.37}$$

which is an $n \times n$ matrix, with $\text{Var}(X_i)$ on the diagonals.

The intuitive picture is that V_{ij} is positive if a higher value of X_i tends to produce a higher value of X_j (we say X_i and X_j are **positively correlated**), while V_{ij} is negative if a high value of X_i tends to produce a lower value of X_j (which we call **negative correlation**). If X has dimensions, then V_ij has dimensions of X^2; to create a dimensionless measure of correlation,

We define the **correlation coefficient** between two random variables as

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}\,, \tag{5.38}$$

where σ_X and σ_Y are the standard deviations of X and Y respectively.

From this definition, we can show that $-1 \leq \rho_{XY} \leq 1$, with 1 showing complete positive correlation, -1 being complete anticorrelation.

5.2.2 Error Propagation

One of the most basic procedures in the lab is the measurement of certain physical quantities with an associated uncertainty, and then using these measurements to calculate other derived quantities, and you would like to know how to *propagate the uncertainties* associated with your measurements to these derived quantities. The right way to think about your measurements is that they are random variables, with mean given by the measured value, and the variance being determined by the measurement uncertainties.

Consider making measurements of n different physical quantities, described by random variables $\vec{X} \equiv (X_1, \cdots, X_n)$, with means $\vec{\mu} = (\mu_1, \cdots, \mu_n)$, and an $n \times n$ covariance matrix V_{ij}. Suppose in addition that $V_{ij} \ll \mu_i \mu_j$, so that the variances are all small. Given another random variable Y, which is a function of $\vec{X}$, you would now like to understand what you should expect for the mean and standard deviation of Y. For example, you could be making two measurements X_1 and X_2, and are interested in the derived parameter $Y = X_1 X_2$, or anything in general.

First, given that the variances are small, given any measurement $\vec{x}$, $y(\vec{x})$ can be expanded as

$$y(\vec{x}) = y(\vec{\mu}) + \partial_i y|_{\vec{x}=\vec{\mu}}\,(x_i - \mu_i) + \frac{1}{2}\partial_i \partial_j y|_{\vec{x}=\vec{\mu}}(x_i - \mu_i)(x_j - \mu_j) + \cdots\,, \tag{5.39}$$

where $\partial_i \equiv \partial/\partial x_i$, we are using Einstein summation notation here with i's being summed over. $\cdots$ represents higher order terms. We see from this expression that

$$\begin{aligned}\langle y(\vec{x})\rangle &= y(\vec{\mu}) + \partial_i|_{\vec{x}=\vec{\mu}}\langle x_i - \mu_i\rangle + \frac{1}{2}\partial_i\partial_j y|_{\vec{x}=\vec{\mu}}\langle (x_i - \mu_i)(x_j - \mu_j)\rangle + \cdots \\ &= y(\vec{\mu}) + \frac{1}{2}\partial_i \partial_j y|_{\vec{x}=\vec{\mu}} V_{ij} + \cdots\,.\end{aligned} \tag{5.40}$$

Therefore, the mean value of the derived parameter $y(\vec{x})$, to leading order, is just what you expect: in the example where $Y = X_1 X_2$, $\langle Y\rangle \approx \mu_1 \mu_2$, i.e. to

get a derived parameter, just plug in the central values in the expression of the derived parameter.

Next, we see

$$y^2(\vec{x}) = y^2(\vec{\mu}) + \partial_i y^2|_{\vec{x}=\vec{\mu}}(x_i - \mu_i) + \frac{1}{2}[\partial_i\partial_j y^2]_{\vec{x}=\vec{\mu}}(x_i - \mu_i)(x_j - \mu_j) + \cdots , \tag{5.41}$$

and

$$\langle y^2(x)\rangle = y^2(\vec{\mu}) + \frac{1}{2}[\partial_i\partial_j y^2]_{\vec{x}=\vec{\mu}}V_{ij} + \cdots \tag{5.42}$$

At the same time,

$$\langle y(\vec{x})\rangle^2 = y^2(\vec{\mu}) + [y\partial_i\partial_j y]_{\vec{x}=\vec{\mu}}V_{ij} + \cdots \tag{5.43}$$

Noting that $\partial_i\partial_j y^2 = 2\partial_i y\partial_j y + 2y\partial_i\partial_j y$, we finally see that

$$\mathsf{Var}(y(x)) = [\partial_i y]_{\vec{x}=\vec{\mu}}[\partial_j y]_{\vec{x}=\vec{\mu}}V_{ij} \, . \tag{5.44}$$

If there is no correlation between X_i, then V_{ij} is diagonal, and

$$\mathsf{Var}(y) = \sigma_Y^2 = \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma_{X_i}^2 \, , \tag{5.45}$$

a standard formula for error propagation you may be familiar with.

5.3 Common Distributions

Having discussed PDFs and the moments of the PDFs, let's take a look at some common distributions that you are likely to encounter in research life.

5.3.1 Binomial distribution

The **binomial distribution** is a distribution that is used to model the outcome of N independent experiments with two possible outcomes, e.g. "success" and "failure", with successes happening with probability p. The number of successes is then a random variable X. Suppose the number of successes is n; within N independent experiments, there are $N!/(n!(N-n)!)$ combinations of n successful experiments. The probability of each combination happening is $p^n(1-p)^{N-n}$ and we can directly compute the probability $P(X = n)$ as

$$P(X = n) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} \, . \tag{5.46}$$

As a warm up, let's check that the total probability is 1. To see this, note that the binomial expansion is given by

$$(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!}p^n q^{N-n} \, , \tag{5.47}$$

and therefore

$$\begin{aligned} \sum_{n=0}^{\infty} P(X = n) &= \sum_{n=0}^{\infty} \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} \\ &= (p + 1 - p)^N \\ &= 1 \, . \end{aligned} \tag{5.48}$$

Let's investigate the mean and variance. First,

$$\mathbb{E}(X) = \sum_{n=0}^{N} nP(X=n) = \sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \,. \tag{5.49}$$

To evaluate this, we'll borrow a trick that you can use also with integrals, that is due to Feynman. Using the binomial expansion above,

$$\frac{d}{dp}(p+q)^N = N(p+q)^{n-1} = \sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^{n-1} q^{N-n} \,. \tag{5.50}$$

If I now set $q = 1-p$ on both sides, I can conclude that

$$N = \frac{1}{p} \sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \,, \tag{5.51}$$

or in other words,

$$\mathbb{E}(X) = Np \,. \tag{5.52}$$

The mean of a binomial random variable is therefore just the total number of trials times the probability of success of each trial, which makes intuitive sense.

A similar argument involving taking two derivatives of p for the polynomial $(p+q)^N$ leads to

$$\mathrm{Var}(X) = Np(1-p) \,. \tag{5.53}$$

Suppose you were running an election poll between two candidates, with candidate A having support given by p, and candidate B given by $1-p$. If you take a survey of N people, the random variable corresponding to the number of people supporting candidate A will have a mean of Np, and a variance of $Np(1-p)$. Taking the ratio of the standard deviation to the mean, we find

$$\frac{\sigma}{\mu} = \frac{\sqrt{Np(1-p)}}{Np} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{p} - 1} \,. \tag{5.54}$$

So, unless one candidate is winning by a whopping landslide, you should expect that the typical uncertainty in your poll scales as $1/\sqrt{N}$. This is a very powerful rule of thumb: if you did a poll of about 1,000 people, you should expect uncertainties at the 3% level or so.

(**End of Lecture**: Wednesday Nov 20 2024)

5.3.2 Poisson distribution

Consider a time interval T over which we expect some event to happen, with some fixed probability per unit time λ/T. Events occur entirely independently, and never simultaneously. This could be, for example, the number of muons received every minute from cosmic rays, the number of photons received at a pixel from a far away source, the number of Prussian soldiers killed by horse kick per year, and more; basically any thing that happens sporadically, independently, and with fixed probability over time. You can take the time interval T and break it up into say 10 intervals, each with length $T/10$, so that the probability in each interval is now $\lambda/10$. If you keep doing this, you'll get N intervals, each with probability λ/N, but these intervals are so small that there is basically only going to be 1 event in them, if ever. We can then model the probability distribution of X, the number of events in time interval T, as N 'success/failure' experiments,

each with a probability of λ/N, and take $N \to \infty$. This is, of course, just a binomial distribution, but taking $p \to 0$, $N \to \infty$ but with $Np = \lambda$.

Let's see how this works. The probability distribution for a binomial random variable can be written as

$$\begin{aligned} P(X = n) &= \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ &= \frac{(Np)^n}{n!} \frac{N!}{(N-n)!} \frac{(1 - Np/N)^N}{N^n (1 - Np/N)^n} \end{aligned} \tag{5.55}$$

Taking the limit as $N \to \infty$, $p \to 0$ with $Np = \lambda$, and noting that $\lim_{N\to\infty}(1 - \lambda/N)^N = e^{-\lambda}$, we find

$$P(X = n) \to \frac{\lambda^n e^{-\lambda}}{n!} \frac{N!}{(N-n)! N^n} \,. \tag{5.56}$$

Using Stirling's approximation $N! \to \sqrt{2\pi N}(N/e)^N$,

$$\frac{N!}{(N-n)! N^n} = \frac{\sqrt{2\pi N}}{\sqrt{2\pi(N-n)}} \frac{N^N}{(N-n)^{N-n} N^n} \to 1 \tag{5.57}$$

as $N \to \infty$ once again. We therefore arrive at the **Poisson probability distribution function** for the number of events in a time interval T,

$$P(X = n) = \frac{\lambda^n e^{-\lambda}}{n!} \,, \tag{5.58}$$

where λ/T is the probability of an event happening per unit time.

Once again, let's begin by checking that this distribution is normalized. We have

$$\begin{aligned} \sum_{n=0}^{\infty} P(X = n) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} e^{\lambda} \\ &= 1 \,, \end{aligned} \tag{5.59}$$

where we have just recognized the Taylor expansion of the exponential function in the second last line. To get the mean,

$$\begin{aligned} \mathbb{E}(X) = \sum_{n=0}^{\infty} n P(X = n) &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\ &= \lambda \,. \end{aligned} \tag{5.60}$$

Finally, a very similar set of tricks also ends up showing that

$$\mathrm{Var}(X) = \lambda \,. \tag{5.61}$$

A few comments are in order here. First, let's consider the sum of two independent Poisson processes, $Y = X_1 + X_2$, with $\mathbb{E}(X_1) = \lambda_1$ and $\mathbb{E}(X_2) = \lambda_2$. Then the probability distribution function is given by

$$\begin{aligned} P(Y=y) &= P(X_1=0)P(X_2=y) + P(X_1=1)P(X_2=y-1) + \cdots \\ &= \sum_{n=0}^{y} \frac{\lambda_1^n}{n!} e^{-\lambda_1} \frac{\lambda_2^{y-n}}{(y-n)!} e^{-\lambda_2} \\ &= \frac{1}{y!} e^{-(\lambda_1+\lambda_2)} \sum_{n=0}^{y} \frac{y!}{n!(y-n)!} \lambda_1^n \lambda_2^{y-n} \\ &= \frac{1}{y!} e^{-(\lambda_1+\lambda_2)} (\lambda_1+\lambda_2)^y \,, \end{aligned} \tag{5.62}$$

and so Y is itself a Poisson process, with mean given by $\lambda_1 + \lambda_2$. That's useful: for example, if you're doing a particle physics experiment involving counting up e.g. the number of dimuon pairs with a certain invariant mass, you can model both the signal and the background as a Poisson process with some mean, and just add up the two contributions into a single Poisson process.

Second, thinking about the example of having a Poisson-like signal and a Poisson-like background, how does the signal-to-statistical noise ratio scale, i.e. just fluctuations in the number of events due to pure chance? If the mean number of signal events is S and the mean number of background events is B, then the total fluctuations is a Poisson process with mean $S + B$, and standard deviation $\sqrt{S+B}$. You therefore would like to make sure that $S/\sqrt{S+B} \gtrsim 1$ at least in order to detect your signal.

5.3.3 Uniform distribution

The **uniform distribution** is our first example of a continuous random variable, and it's also really simple: it is a random variable with uniform probability between values a and b, and so has PDF

$$\begin{cases} \frac{1}{b-a} \,, & a < x < b \,, \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{5.63}$$

We can quickly evaluate

$$\begin{aligned} \langle x \rangle &= \int_a^b dx \, \frac{x}{b-a} \\ &= \frac{1}{2} \frac{b^2 - a^2}{b-a} \\ &= \frac{1}{2}(a+b) \,, \end{aligned} \tag{5.64}$$

i.e. the midway point between a and b, and

$$\begin{aligned} \mathsf{Var}(X) = \langle (x - \langle x \rangle)^2 \rangle &= \int_a^b dx \, \frac{(x - (a+b)/2)^2}{b-a} \\ &= \frac{1}{3} \frac{1}{b-a} \left[\left(b - \frac{a+b}{2} \right)^3 - \left(a - \frac{a+b}{2} \right)^3 \right] \\ &= \frac{1}{12} (a-b)^2 \,. \end{aligned} \tag{5.65}$$

5.3.4 Exponential distribution

The **exponential distribution** models the distance between Poisson processes, i.e. processes that occur independently and with a constant average rate. This

is for example the distribution of the decay lifetime of an ensemble of particles with some constant halflife, or equivalently the distance traveled by particles with some constant halflife. Another example would be time between production errors, or the time interval between consecutive buses, etc.

It has a PDF given by

$$f(x) = \frac{1}{\xi} e^{-x/\xi}\,, \tag{5.66}$$

or sometimes it is parametrized using $\lambda \equiv 1/\xi$ instead. Elementary integration tells you that

$$\mathbb{E}(X) = \sqrt{\mathsf{Var}(X)} = \xi\,. \tag{5.67}$$

In addition, the probability that $P(x > s)$ is simply

$$P(x > s) = \int_s^\infty dx\, f(x) = e^{-s/\xi} \tag{5.68}$$

The exponential distribution satisfies the **memorylessness property**, i.e. that

$$\begin{aligned} P(x > s + t | x > s) &= \frac{P(x > s + t \cap x > s}{P(s)} \\ &= \frac{P(x > s + t)}{P(s)} \\ &= \frac{e^{-(s+t)/\xi}}{e^{-s/\xi}} \\ &= e^{-t/\xi} \\ &= P(x > t)\,, \end{aligned} \tag{5.69}$$

for any s and t. In other words, given that the mean time between events is 10 seconds, but one event hasn't happened in a 1,000 seconds, the probability that an event will happen in the next 10 seconds is just the unconditional probability; the fact that 1,000 seconds has already elapsed doesn't make it more likely than an event will happen in the next 10 seconds.

5.3.5 Gaussian distribution

The **Gaussian distribution** or **normal distribution** is probably already pretty well-known to you at this point. The one-dimensional Gaussian random variable X has the PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,, \tag{5.70}$$

where $\mathbb{E}(X) = \mu$ and $\mathsf{Var}(X) = \sigma^2$. We very frequently write $X \sim N(\mu, \sigma^2)$. The particular choice of $\mu = 0$ and $\sigma = 1$ defines the **standard Gaussian**,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,, \tag{5.71}$$

and you can see pretty easily that given a Gaussian random variable Y with mean μ and variance σ^2, $(Y - \mu)/\sigma$ is the standard Gaussian.

The n-dimensional generalization is the **multivariate Gaussian**, which can be thought of as a set of random variables $X_1, \cdots, X_n$ with the joint PDF

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}_i - \vec{\mu}_i)V_{ij}^{-1}(\vec{x}_j - \vec{\mu}_j)\right)\,, \tag{5.72}$$

where $\mathbb{E}(\vec{x}) = \vec{\mu}$, and V_{ij} is the $n \times n$ covariance matrix between the random variables $X_1, \cdots, X_n$. $|V|$ is the determinant of the covariance matrix.

Like the Poisson distribution, if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $X+Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$, and so you can add Gaussian random variables together easily.

Much more to be said about the Gaussian later.

5.4 Statistics: The Main Idea

Having discussed probabilities and common probability density functions, we now have the tools to understand how to do science with uncertain data. This is a vast field of knowledge that we can only get a flavor of in this course, but we should be able to cover the main ideas and a lot of the vocabulary.

How do we do science? From the statistical perspective, we perform an experiment and make some measurements with appropriate uncertainties; these measurements form what we call **data**. With the data, we want to perform **statistical inference**, a procedure to gain some information about our description of nature in some way, generically what we call the **model**. The main point-of-view we will take in this course is the *frequentist* perspective. Here, the model is a fixed, immutable thing, with parameters that have true, unique values, which you may or may not know; *it makes no sense to discuss the probability distribution of model parameters, or allowing model parameters to vary, and so on*. Understanding this is crucial in correctly interpreting conclusions drawn from frequentist procedures!

The sole quantity of interest in frequentist statistics is the **likelihood**, which is $P(\text{data}|\text{model})$. More accurately, let's represent the data by a vector $\boldsymbol{x}$, and the parameters of the model by a vector $\boldsymbol{\theta}$; the likelihood is then the PDF $f(\boldsymbol{x}; \boldsymbol{\theta})$. This notation emphasizes that $\boldsymbol{x}$ is the value of the random variable, and $\boldsymbol{\theta}$ is the parameters of the model assumed. There are two main types of statistical inference we perform:

1. If all model parameters $\boldsymbol{\theta}$ are specified, we can perform **hypothesis testing**, which asks whether the data $\boldsymbol{x}$ is consistent with the model $\boldsymbol{\theta}$. This is often what we do when we set *experimental limits* on parameters that, if they exist, would produce "new physics", e.g. the mass of the photon.

2. If some model parameters are unknown, we can perform **parameter estimation** given $f(\boldsymbol{x}, \boldsymbol{\theta})$, mostly usually in the form of a **confidence interval** that attempts to bracket the true value of $\boldsymbol{\theta}$.

Great! All we need now is the likelihood, which is easier said than done.

The brute-force, but actually modern way of getting the likelihood is to perform *simulations*: given a model $\boldsymbol{\theta}$, you run many numerical experiments, and try to figure what is the distribution of all possible outcomes. A silly example could be simulating flipping a fair coin 20 times and trying to understand the probability of getting n heads out of 20; a more sophisticated (and real!) example would be to take a cosmological model, and run it through a simulation of galaxy formation, and try to understand how the probability of observing a certain distribution of galaxies depends on model parameters.

But fortunately, we don't always have to go to such extremes. Two mathematical facts help make life significantly easier. Suppose we have a sequence of n *independent and identically distributed* (often written as *i.i.d*) random variables, $X_1, \cdots, X_n$, often called **samples** in statistics, but in physics is often as repeat measurements of some quantity. These could, for example, be n independent measurements of the same quantity like the Hubble constant, or heads/tails of a coin flip, etc. Consider a new random variable $\overline{X} = (X_1 + \cdots + X_n)/n$.

The first result is the **law of large numbers**, which says that the pdf of X will become arbitrarily concentrated around μ as we increase the value of n to an arbitrarily large value.

This fact guarantees that, *as long as* $X_1, \cdots, X_n$ *really are i.i.d*, that you will always get the right answer if you have enough statistics. To see why this works, we can simply compute

$$\mathbb{E}(\overline{X}) = \frac{1}{n}\mathbb{E}(X_1 + \cdots + X_n) = \frac{1}{n} \cdot n\mathbb{E}(X_1) = \mu\,, \tag{5.73}$$

and

$$\begin{aligned}
\mathsf{Var}(\overline{X}) &= \mathbb{E}\left[\left(\frac{X_1 + \cdots + X_n}{n}\right)^2\right] - \mathbb{E}\left(\overline{X}\right)^2 \\
&= \frac{1}{n^2}\mathbb{E}\left[(X_1^2 + \cdots + X_n^2) + (X_1X_2 + X_1X_3 + \cdots + X_{n-1}X_n)\right] - \mu^2 \\
&= \frac{1}{n}\mathbb{E}(X_1^2) + \frac{n^2 - n}{n^2}\mathbb{E}(X_1)^2 - \mu^2 \\
&= \frac{1}{n}\mathsf{Var}(X_i)\,,
\end{aligned} \tag{5.74}$$

where in the second last line I used the fact that $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)^2$ since X_1 and X_2 are not correlated, etc. You can see that $\mathsf{Var}(\overline{X})$ therefore shrinks as $1/n$, giving us the desired result, at least on a non-rigorous level.

Note that the key assumption is that $X_1, \cdots, X_n$ *really are i.i.d*; statistics cannot save you if you're averaging over mismeasured data, or if you're polling a group which is not representative of the full population, and so on. This is the key difference between **statistical uncertainty** and **systematic uncertainty**: the statistical uncertainty of an experiment is due to random fluctuations in a finite sample size, and can be averaged down with a large enough sample/number of repetitions, whereas a systematic uncertainty basically would throw the measured mean value off from the true value, and cannot be averaged away. This could be things like the fact that your detector misses 5% of all events due to incomplete coverage, or an instrument not being calibrated correctly, etc.

The second result is almost certainly the jewel in the crown of statistics, and really goes a long way toward making everything tractable:

The **central limit theorem** states that the sum n independent, continuous random variables X_i with means μ_i and variances σ_i^2 becomes a Gaussian random variable with mean $\mu = \sum_{i=1}^n \mu_i$ and variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ in the limit of $n \to \infty$ under fairly generic assumptions about X_i.

We won't go into the fine details of what the assumptions are, nor will we give a proof, although the proof is well within your abilities now that you have sat through most of mathematical physics (see Cowan 10.3). Concretely, the central limit theorem tells us that as long as we are constructing a random variable X by summing up a bunch of other random variables, the resulting distribution will be Gaussian. But if you think about it, a large chunk of measurements corresponds to summing things up/averaging: adding up all the photons received in different parts of the sky, integrating and then averaging a signal over some small time window (or equivalently binning your data), etc. This theorem *justifies the assumption that random error of a measurement can usually be modeled as a Gaussian random variable with some mean and variance*, and allows us in many instances to write down a likelihood.

5.5 Hypothesis Testing

We're now ready to discuss an important class of frequentist methods called **hypothesis testing**. As with all frequentist methods, we start by assuming a fixed model, which is frequently also referred to as a **hypothesis**. Hypothesis testing involves evaluating how consistent the data is with a given hypothesis. This can come in several flavors. The first main variant, called a *goodness-of-fit test* by Cowan, and also variously known as a *significance test* or *Fisher's significance testing*, involves only one hypothesis, called the **null hypothesis**, and simply provides a statement of how good the agreement is between data and the null hypothesis. The second variant, called simply *hypothesis testing* or *Neyman-Pearson hypothesis testing*, involves two hypotheses, a null and an alternate hypothesis; the end result of the test is a recommendation to accept one hypothesis and reject the other. The relation between the two has been the source of great controversy in statistics (and of great bitterness between Fisher, Neyman and Pearson!), a fascinating discussion that we will not engage in. *Both* types of hypothesis testing are used in physics, but we'll focus on the first case here.

5.5.1 An example: testing the fairness of a coin

The best way to explain the procedure is to just start off with an example. Suppose we have a coin, and we would like to establish whether the coin is a fair coin. The *null hypothesis* here is that the coin is fair. Next, we design our statistical test. We first fix the number of coin tosses to $N = 20$; we know from the binomial distribution that the number of heads n_h we should observe is $\langle n_h \rangle = 10$. If we obtain n_h, then the likelihood is

$$P(n_h|H_0) = \frac{20!}{n_h!(20-n_h)!}(0.5)^{n_h}(0.5)^{N-n_h}\,. \tag{5.75}$$

We also have to decide on what we mean by the data being consistent with the model. Clearly, $n_h = 10$ is more consistent than $n_h = 17$; but what about $n_h = 3$? If we are only interested in whether the coin is fair or not, which we take to be the case here, then we would say $n_h = 3$ is also inconsistent, and as inconsistent as $n_h = 17$ is; this is known as a **two-tailed test**. In other situations, say where our null hypothesis is "background only", and we're interested in seeing whether there is a signal on top of our background, then we would prefer a **one-tailed test**, where we don't consider downward fluctuations from the mean as being discrepant.

Next, we run an experiment, where we flip the coin 20 times. Say for example that we obtain 17 heads, which intuitively seems inconsistent with the null hypothesis. To quantify this, we compute the **significance** or the p-**value**, which is the probability of obtaining this level of discrepancy *or higher*. This means summing over the probability for $n_h = 0, 1, 2, 3, 17, 18, 19, 20$:

$$\begin{aligned} P(n_h \leq 3|H_0) + P(n_h \geq 17|H_0) &= 2P(n_h \leq 3|H_0) \\ &= 2\sum_{n_h=0}^{3} \frac{20!}{(20-n_h)!n_h!}(0.5)^{20} \\ &= 0.0026\,. \end{aligned} \tag{5.76}$$

Now comes the question: what do we *do* with the p-value? A couple of points:

1. First, let's make clear what the p-value is *not*. It is *emphatically not* the probability that the hypothesis is true: it makes no sense in the frequentist perspective to talk about the probability with regard to models, even

though this is often what you want in science! Rather, the p-value tells you the probability of getting a result as extreme as or more extreme than $n_h = 17$, *assuming the hypothesis*.

2. One common way in which we use hypothesis testing as a community is by deciding beforehand what p-value α is low enough to *reject the null hypothesis*. α is also called the *significance level* of the test. In the social sciences, $\alpha = 0.05$ is a common benchmark. This translates to a probability of being 2 standard deviations from the mean of a Gaussian distribution. Once again, let me emphasize what α means: choosing $\alpha = 0.05$ does not mean that we reject the hypothesis if the there is a less than 5% chance that the hypothesis is true, since this way of thinking is illogical in the frequentist perspective. Rather, $\alpha = 0.05$ gives us a 5% chance of mistakenly rejecting H_0 when it is true; this is known as making a **type-I error**, which is an unfortunately uninformative name. *The choice of α is a completely arbitrary choice, and is basically saying, "Nah, I don't think I'm unlucky at the level of α in being misled by the data that I collected."* In particle physics, we adopt the convention of setting $\alpha = 5.7 \times 10^{-7}$ (or a 5σ discrepancy if the likelihood is Gaussian) for any statistical test used to claim a discovery. 2σ or 5σ, there is simply no basis at all in this choice, and is symptomatic of the bigger problem of frequentist statistics: the question that we are interested in, $P(\text{model}|\text{data})$, lies outside the scope of frequentist statistics.

3. Another way in which hypothesis testing seems unsatisfactory is the following. Consider running twenty different tests of twenty different null hypotheses, all with $\alpha = 0.05$.[43] We don't even need to know what these hypotheses are: we're very likely going to reject one of them, just due to the choice of significance level, and the fact that we're doing so many tests. This problem goes another the name p-**hacking** in the social sciences, and the **look-elsewhere effect** in particle physics.

[43] See https://xkcd.com/882/ for an illustration.

(**End of Lecture**: Monday Nov 25 2024)

5.5.2 Asymptotic distributions and Pearson's chi-squared test

One of the main difficulties with performing a hypothesis test is the construction of a suitable test statistic $t(\boldsymbol{x})$, with a known likelihood function $P(t(\boldsymbol{x})|H_0)$. But as we alluded to earlier, the *central limit theorem* goes a long way toward the construction of statistics with known likelihoods.

Consider a random variable X which has some distribution. We observe several values of X that we represent as a histogram over N bins. Let the number of entries in bin i be n_i. Suppose we have a model H_0 that predicts the expected number of entries in bin i, denoted ν_i. The goal is to construct a test statistic that reflects the level of agreement between observed and expected histograms.

One choice is Pearson's χ^2 statistic, given by

$$t(\boldsymbol{n}) = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i} \,. \tag{5.77}$$

It can be shown that if the data $\boldsymbol{n} = (n_1, \cdots, n_N)$ are Poisson distributed with mean values $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_N)$, and if the number of entries are not too small, then $t(\boldsymbol{n})$ follows the χ^2 distribution with N degrees of freedom, i.e. $f(t(\boldsymbol{n})|H_0)$ has the PDF of a χ^2 distribution with N degrees of freedom. Tests using this

statistic are also known as χ^2 goodness-of-fit tests. The associated p-value is then given by

$$p = \int_{\chi^2}^{\infty} dz\, f_{\chi^2}(z; N)\,, \tag{5.78}$$

where $f_{\chi^2}(z; N)$ is the χ^2 distribution with N degrees of freedom.

Another common scenario is when the data $\boldsymbol{n} = (n_1, \cdots, n_N)$ are constrained to $n_1 + \cdots + n_n = n_{\text{tot}}$, which is what you might expect when the data is proportional to some probability density function. In that case, if we again know the expected mean values $\boldsymbol{\nu}$ given the null hypothesis H_0, the χ^2 statistic can be shown to follow a χ^2 distribution with $N - 1$ degrees of freedom; you can think of the one fewer degree of freedom as the fact that the normalization condition must be enforced.

While we won't prove this particular result here, these kinds of asymptotic guarantees—that a particular statistic follows some distribution in the limit of large enough sample sizes—are ultimately due to the central limit theorem. There are many other useful results that I won't have time to talk about in this course, such as *Wilk's theorem* and the *profile likelihood*, but the idea is very similar to this simpler example: perform a hypothesis test by constructing a test statistic, which has a likelihood that is guaranteed to have a known distribution under some mild conditions.[44] I consider this one of the best reasons to use frequentist statistics: you have full analytic control of the likelihood, which can be difficult to arrive at otherwise, and can get a statistical test done without much fuss.

[44] Cowan et al. is the standard reference on this in particle physics.

5.6 Parameter Estimation

The other common reason to perform a statistical analysis is to obtain some kind of estimate for unknown model parameters. In physics, we also want to give an estimate for how *confident* we are in this estimation. The basic idea of doing this as follows: given some data $\boldsymbol{x}$, and unknown model parameters $\boldsymbol{\theta}$, determine the likelihood $f(x; \theta)$. We can estimate the parameters as $\hat{\boldsymbol{\theta}}$, the value of the parameters that *maximizes* the likelihood. How quickly the likelihood falls off from this maximum tells us roughly how confident we should be of the estimate.

5.6.1 Estimators

Consider n samples $X_1, \cdots, X_n$ are n i.i.d. samples, e.g. n height measurements in some population of people, or the number of photons received in n pixels, all expected to follow the same distribution of photons. Suppose that the true, underlying probability distribution has mean μ and variance σ^2. We would now like to construct a function of the sample values $x_1, \cdots, x_n$ that would be an **estimate** of μ. Intuitively, the function would simply be to take the mean of all n samples, i.e.

$$\overline{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i\,, \tag{5.79}$$

which we call an **estimator** for μ. This particular estimator is called the **sample mean estimator**. Note that the estimator is a random variable, and its true value depends on what samples you get. The **estimate** given the values $x_1, \cdots, x_n$ would then be

$$\hat{\mu} \equiv \frac{1}{n} \sum_{i=1}^{n} x_i\,, \tag{5.80}$$

where we use the hat to denote quantities that are estimates, and *not* the true value. Again, let me stress that $\hat{\mu}$ is drawn from the distribution of $\overline{X}$, and will vary depending on what samples got drawn originally.

Let's check a few properties of $\overline{X}$. It's expectation value is given by

$$\begin{aligned}\mathbb{E}(\overline{X}) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) \\ &= \frac{1}{n}\cdot n\mu \\ &= \mu\,. \end{aligned} \tag{5.81}$$

Well, that's a good sign: the expectation value of the estimator is equal to the true quantity. An estimator that satisfies this property is called an **unbiased estimator**.

Let's now check the *variance* of $\overline{X}$, which tells us how large of a spread the estimates are going to be about the true mean. This was already calculated in Eq. (5.74),

$$\begin{aligned}\mathsf{Var}(\overline{X}) &= \mathbb{E}(\overline{X}^2) - \mathbb{E}(\overline{X})^2 \\ &= \frac{\sigma^2}{n}\,. \end{aligned} \tag{5.82}$$

This is another restatement of the *law of large numbers*: as the sample size gets larger, the variance of the unbiased estimator of the mean scales as $1/n$.

Next, let's construct an estimator for σ^2 itself, the variance of the true probability distribution. You might naively think that the right estimator is

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\,, \tag{5.83}$$

but let's check its expectation value. We have

$$\begin{aligned}\mathbb{E}(S_n^2) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i - \overline{X}\right)^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}(X_i^2) - 2\mathbb{E}(X_i\overline{X}) + \mathbb{E}(\overline{X}^2)\right]\,. \end{aligned} \tag{5.84}$$

Let's go through each term carefully. First,

$$\mathbb{E}(X_i^2) = \mathsf{Var}(X_i) + \mathbb{E}(X_i)^2 = \sigma^2 + \mu^2\,. \tag{5.85}$$

Next,

$$\begin{aligned}\mathbb{E}(X_i\overline{X}) &= \frac{1}{n}\mathbb{E}\left(X_i\sum_{j=1}^{n} X_j\right) \\ &= \frac{1}{n}\left((n-1)\mathbb{E}(X_i)^2 + \mathbb{E}(X_i^2)\right) \\ &= \frac{1}{n}\left((n-1)\mu^2 + \sigma^2 + \mu^2\right) \\ &= \mu^2 + \frac{\sigma^2}{n}\,. \end{aligned} \tag{5.86}$$

Finally,

$$\mathbb{E}(\overline{X}^2) = \mathsf{Var}(\overline{X}) + \mathbb{E}(\overline{X})^2 = \frac{\sigma^2}{n} + \mu^2 \,. \tag{5.87}$$

Putting everything together,

$$\begin{aligned}
\mathbb{E}(S_n^2) &= \frac{1}{n}\sum_{i=1}^{n}\left[\sigma^2 + \mu^2 - 2\mu^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} + \mu^2\right] \\
&= \frac{1}{n}\cdot n \cdot \frac{n-1}{n}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2 \,.
\end{aligned} \tag{5.88}$$

As it turns out, S_n^2 is a *biased estimator*! The *unbiased estimator* (known as the **sample variance estimator**) is then

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \,. \tag{5.89}$$

More generally, for an estimator $\hat{\theta}(\boldsymbol{X})$ for the parameter θ, where $\boldsymbol{X}$ is a vector of random samples, with values given by $\boldsymbol{x}$, we define

$$b \equiv \mathbb{E}(\hat{\theta}(\boldsymbol{X})) - \theta \tag{5.90}$$

to be the **bias** of the estimator. The variance of the estimator is defined to be

$$\mathsf{Var}(\hat{\theta}(\boldsymbol{X})) = \mathbb{E}\left[\left(\hat{\theta}(\boldsymbol{X}) - \mathbb{E}(\hat{\theta}(\boldsymbol{X}))\right)^2\right] \,. \tag{5.91}$$

Finally, we define the **mean square error** of the estimator,

$$\begin{aligned}
\mathsf{MSE}(\hat{\theta}(\boldsymbol{X})) &= \mathbb{E}\left[\left(\hat{\theta}(\boldsymbol{X}) - \theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta}(\boldsymbol{X}) - \mathbb{E}(\hat{\theta}(\boldsymbol{X})) + b\right)^2\right] \\
&= \mathsf{Var}(\hat{\theta}(\boldsymbol{X})) + b^2 + 2b\mathbb{E}\left[\left(\hat{\theta}(\boldsymbol{X}) - \mathbb{E}(\hat{\theta}(\boldsymbol{X}))\right)\right] \\
&= \mathsf{Var}(\hat{\theta}(\boldsymbol{X})) + b^2 \,.
\end{aligned} \tag{5.92}$$

In the context of measurement, $\hat{\theta}(\mathbf{X})$ is the random variable representing the outcome of a measurement, with $\mathsf{Var}(\hat{\theta}(\boldsymbol{X}))$ representing statistical uncertainty, and b representing systematic uncertainty.

5.6.2 Maximum likelihood estimator

Consider a random variable $\boldsymbol{X}$ distributed according to a likelihood $f(\boldsymbol{x};\boldsymbol{\theta})$, where $\boldsymbol{x}$ in your experiment, and $\boldsymbol{\theta}$ are the parameters in your model. Suppose you don't know the values of $\boldsymbol{\theta}$, and would like to estimate their values, based on your data.

One intuitive way to perform this estimate is to find the value of $\boldsymbol{\theta}$ that maximizes the likelihood. Putting it more formally, let's define the **likelihood function**

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x};\boldsymbol{\theta}) \,, \tag{5.93}$$

where we view the likelihood as being a function of the *parameters*, instead of the data. Note that from this viewpoint, the likelihood function $L(\boldsymbol{\theta})$ is *not a*

[45] I am using the unfortunately usual convention of the estimator (a random variable) and the estimate (a real number) being denoted by the same symbol here.

PDF over $\boldsymbol{\theta}$, and is not properly normalized; after all, in the frequentist point of view, we don't assign probabilities to $\boldsymbol{\theta}$! They are fixed to values that we don't know, and we're just trying to find an estimate for them. The **maximum likelihood estimator** $\hat{\boldsymbol{\theta}}$ (ML estimator for short)[45] is obtained by finding the value of $\boldsymbol{\theta}$ such that

$$\left.\frac{\partial L}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0\,, \tag{5.94}$$

where $\partial L/\partial \boldsymbol{\theta}$ is a shorthand for taking the partial derivatives with respect to each of the components of $\boldsymbol{\theta}$.

Let's try an example, involving n draws from a Gaussian distribution, $\boldsymbol{x} = (x_1, \cdots, x_n)$, with unknown μ and σ^2. The likelihood function is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\,. \tag{5.95}$$

It's often more convenient to think about the log-likelihood function instead; maximizing the log-likelihood function is then equivalent to maximizing the likelihood. For the 1D Gaussian distribution, this is simply

$$\begin{aligned}\log L(\mu, \sigma^2) &= \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right] \\ &= -n\log\sqrt{2\pi\sigma^2} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\,.\end{aligned} \tag{5.96}$$

From here, we have

$$\frac{\partial}{\partial\mu}\log L = \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2}\,. \tag{5.97}$$

The maximum likelihood estimate for μ is therefore

$$\sum_{i=1}^{n}\frac{x_i-\hat{\mu}}{\sigma^2} = 0 \implies n\hat{\mu} = \sum_{i=1}^{n} x_i \implies \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i\,. \tag{5.98}$$

This is just the sample mean estimator, which we know is an unbiased estimator.

Let's see what the ML estimator for the variance looks like. We have

$$\frac{\partial}{\partial\sigma^2}\log L = -\frac{n\sqrt{2\pi}}{\sqrt{2\pi\sigma^2}}\frac{1}{2\sqrt{\sigma^2}} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2\,. \tag{5.99}$$

Setting this to zero, we find

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\,. \tag{5.100}$$

As we discussed earlier, this is *not* unbiased, and should therefore disabuse you of the fact that the ML estimator is always well-behaved. However, in the limit of infinitely many samples, the central limit theorem saves the day again, and ML estimators are guaranteed to have zero bias under very generous conditions.

5.6.3 Rao-Cramér-Frechet Bound and the Fisher Information Matrix

Let's try to tie everything we've learnt so far about estimators into how we would normally deploy this in science. We deploy estimators like the ML estimator to

experiments designed to measure certain parameters in our model. We take a bunch of measurements, and construct the likelihood of the data given the model parameters. Based on this, we can construct the ML estimate for the model parameters of interest. But keep in mind that the estimator $\hat{\boldsymbol{\theta}}$ is itself a random variable, and so our job as scientists is not done when we obtain the ML estimate; we still need to determine the *uncertainty* associated with our estimate.

If the exact form of the likelihood function is known, this can be done analytically. We can always, of course, revert to our usual backup plan of running simulations of our experiments, and determining the the likelihood function through Monte Carlo methods. However, it turns out that a *lower bound* on the variance of any estimator can actually be derived. Given a likelihood function $L(\theta)$ for an unknown parameter θ, for any *unbiased* estimator of a single model parameter $\hat{\theta}$,

The **Rao-Cramér-Frechet** bound says that the variance of the unbiased estimator satisfies

$$\mathrm{Var}(\hat{\theta}) \geq \left[\mathbb{E}\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right)\right]^{-1}, \tag{5.101}$$

under very general conditions,

where as a reminder,

$$\mathbb{E}\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) = -\int d^n\boldsymbol{x}\, f(\boldsymbol{x};\theta)\frac{\partial^2}{\partial \theta^2}\log L(\boldsymbol{x};\theta)\,. \tag{5.102}$$

An estimator is said to be **efficient** when the variance is equal to its lower bound. In practice, ML estimators are almost always efficient in the limit of large sample sizes, again mostly thanks to the central limit theorem, and so it is very common to assume that ML estimators are efficient.

If there are multiple unknown parameters, given an unbiased and efficient estimator $\hat{\boldsymbol{\theta}}$, we can construct the covariance matrix $V_{ij} = \mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j)$.

For an unbiased, efficient estimator, the covariance matrix is given by

$$(V^{-1})_{ij} = \mathbb{E}\left(-\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right), \tag{5.103}$$

where the quantity on the right is known as the **Fisher information matrix**.

Often, the expectation value itself is difficult to calculate; however, once again in the limit of large samples, if we have the ML estimates $\hat{\boldsymbol{\theta}}$, one can instead evaluate the estimate

$$\widehat{V^{-1}}_{ij} = -\left.\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{5.104}$$

in order to obtain an estimate for the variance and uncertainty of the ML estimate. This procedure often goes under the name of a **Fisher forecast**, and is used extensively in fields such as cosmology to forecast the reach of future experiments. The idea is as follows:

1. Start with a model with parameters $\boldsymbol{\theta}$.

2. Establish a simplified model of the instrumental noise or the backgrounds that would interfere with the signal that you're trying to measure with **nuisance parameters** $\boldsymbol{\nu}$.

3. Either write down analytically the complete likelihood function $L(\boldsymbol{\theta}, \boldsymbol{\nu})$, or perform a simulation/numerical computation to obtain the likelihood function.

4. Any ML estimator of the data $\hat{\boldsymbol{\theta}}$ will then have an expectation value given by $\boldsymbol{\theta}$, and a variance that can be calculated using Eq. (5.104). This is often accomplished by taking numerical derivatives of the likelihood as each parameter is varied.

At the end of the day, this procedure gives $\widehat{V^{-1}}$, which one can then invert to provide an estimate for the *sensitivity* of your experiment to $\boldsymbol{\theta}$ (i.e. the covariance of all of the parameters, which represent the statistical uncertainty you would obtain from your experiment), taking into account the estimated background or noise $\boldsymbol{\nu}$.

5.6.4 Confidence intervals

Suppose you have an estimator $\hat{\theta}$ for a single parameter θ in your model; this can be the ML estimator, or something else. On top of that, you also have the PDF of $\hat{\theta}$ *given the value of the true parameter* θ, denoted as $g(\hat{\theta}; \theta)$. For any real number $0 \leq \alpha \leq 1$, I can determine a function $u_\alpha(\theta)$ such that

$$\alpha = P(\hat{\theta} \geq u_\alpha(\theta)) = \int_{u_\alpha(\theta)}^{\infty} d\hat{\theta}\, g(\hat{\theta}; \theta)\,. \tag{5.105}$$

$u_\alpha(\theta)$ here is the value of $\hat{\theta}$ such that the total probability of getting $\hat{\theta} \geq u_\alpha(\theta)$ is α. Similarly, for some other real number $0 \leq \beta \leq 1$, we can determine the function $v_\beta(\theta)$ such that

$$\beta = P(\hat{\theta} \leq v_\beta(\theta)) = \int_{-\infty}^{v_\beta(\theta)} d\hat{\theta}\, g(\hat{\theta}; \theta)\,. \tag{5.106}$$

But now I can invert the functions $u_\alpha(\theta)$ and $v_\beta(\theta)$, and obtain the inverse functions $u_\alpha^{-1}(\hat{\theta})$ and $v_\beta^{-1}(\hat{\theta})$. Furthermore,

$$P(\hat{\theta} \geq u_\alpha(\theta)) = P(u_\alpha^{-1}(\hat{\theta}) \geq \theta)\,, \tag{5.107}$$

where I've just applied u_α^{-1} to both sides of the inequality; and likewise

$$P(\hat{\theta} \leq v_\beta(\theta)) = P(v_\beta^{-1}(\hat{\theta}) \leq \theta)\,. \tag{5.108}$$

Therefore, if we have an estimate $\hat{\theta}$, we can construct an interval for θ such that

$$P(v_\beta^{-1}(\hat{\theta}) \leq \theta \leq u_\alpha^{-1}(\hat{\theta})) = 1 - \alpha - \beta\,, \tag{5.109}$$

where the interval $(v_\beta^{-1}(\hat{\theta}), u_\alpha^{-1}(\hat{\theta}))$ is known as the **confidence interval** with **confidence level** of $1 - \alpha - \beta$.

We can construct **one-sided confidence intervals** or **limits**, such that α or β is zero. Another common choice is to choose $\alpha = \beta = \gamma/2$ to get a **central confidence interval** with confidence level $1 - \gamma$.

Let's remind ourselves one more time: *we currently have a frequentist hat on, and therefore it makes no sense to talk about the probability of parameters taking on values*. And therefore it should be crystal clear to you what the confidence interval is *not*: a 95% confidence interval for a parameter θ is *most definitely not* the interval which captures with 95% probability the value of θ. Rather, if you conduct n experiments an obtain n different estimates $\hat{\theta}$, the confidence interval that you would construct from these estimates would contain the true value θ 95% of the time, as $n \to \infty$. Again, this is not *quite* what you want, but unless you happen to be unlucky at the 5% level, you'd be okay, maybe.

Our previous discussion makes the construction of confidence intervals sound pretty cumbersome, since it first relies on knowing $g(\hat{\theta};\theta)$, which sounds like a difficult problem. Once again, however, the central limit theorem comes to the rescue. The ML estimator $\hat{\theta}$ for θ in the limit of a large sample size has a pdf $g(\hat{\theta};\theta)$ that approaches a Gaussian:

$$g(\hat{\theta};\theta) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(-\frac{(\hat{\theta}-\theta)^2}{2\sigma_{\hat{\theta}}^2}\right), \tag{5.110}$$

where we can once again use the Cramér-Rao-Frechet bound to obtain an estimate $\hat{\sigma}_{\hat{\theta}}^2$ of the variance (unbiased in the large sample limit),

$$\hat{\sigma}_{\hat{\theta}}^2 = \left[\left.\frac{\partial^2 \log L}{\partial\theta^2}\right|_{\theta=\hat{\theta}}\right]^{-1}. \tag{5.111}$$

This can once again be evaluated numerically.

Furthermore, it is also true that in the large sample limit, the likelihood function about the ML estimate itself approaches a Gaussian, i.e.

$$\begin{aligned} L(\theta) &= L_{\max} \exp\left(-\frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}\right), \\ \log L(\theta) &= \log L_{\max} - \frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}. \end{aligned} \tag{5.112}$$

We can therefore also just make a plot of $\log L(\theta)$ as a function of θ and use that to numerically construct the confidence interval, since when $|\theta-\hat{\theta}| = N\sigma_{\hat{\theta}}$, i.e. $\hat{\theta}$ and θ are different by N standard deviations, $\log L(\theta) = \log L_{\max} - N^2/2$. We can set for example $N = 1, 2, 3$ and get the confidence interval $(\hat{\theta} - N\sigma_{\hat{\theta}}, \hat{\theta} + N\sigma_{\hat{\theta}})$, which we call the $N\sigma$ confidence interval, or the 68.3%, 95.4% and 99.7% confidence intervals respectively.

5.7 Bayesian Statistics

Before we end this discussion about statistics, we must discuss the Bayesian point-of-view, which is increasingly the default point-of-view in performing statistical analyses. We've seen that the frequentist approach centers on the *likelihood*, which is roughly given by $P(\text{data}|\text{model})$, or more precisely the pdf $f(\boldsymbol{x};\boldsymbol{\theta})$. But as we've seen time and again, from hypothesis testing to constructing confidence intervals, this isn't quite what you want. Instead, you want to be making statements about which models are more probable, and which are not, given the data, i.e. $P(\text{model}|\text{data})$, a quantity which is known as the **posterior**. More precisely, you want the pdf $p(\boldsymbol{\theta};\boldsymbol{x})$.

The Bayesian framework is as follows. Suppose I now think of probabilities as a *statement of beliefs*, so that when I say something like "the probability of rolling a 6 in a single die roll", it really is a statement about what I *believe* about

the die being fair etc. Under this point of view, I can now define something that looks like $P(\text{model})$, called the **prior**, which allows me to assign probabilities to the model acquiring certain values. With this, I now can use Bayes' Theorem to find that

$$P(\text{model}|\text{data}) \propto P(\text{data}|\text{model})P(\text{model})\,, \tag{5.113}$$

and voilà we have the posterior! More precisely, one would specify the prior distribution $\pi(\boldsymbol{\theta})$, and find that $p(\boldsymbol{\theta};\boldsymbol{x}) \propto f(\boldsymbol{x};\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. There are many advantages to this point-of-view, including the following:

1. A lot of the most powerful results in frequentist statistics rely on the central limit theorem and various "niceness" conditions which I've swept under the rug for this course. These are of course not guaranteed to hold.

2. Sometimes priors are super useful. Going back to our medical test example, the probability that you have a disease given a positive test can be dominated by the prior, since knowing that a disease is just extremely, extremely rare would lead you to believe that a positive test is likely to be a false positive. A pure frequentist has no way of incorporating such beliefs.

3. In many experiments, you have lots of terms in your model that account for noise or unwanted backgrounds that you really want to marginalize over. This is very easy to do in the Bayesian viewpoint: just take the final posterior pdf and marginalize over parameters that you're not interested in. From the frequentist viewpoint, this requires more thought; it can be done with *profile likelihoods* and *Wilks' theorem*, but things get computationally expensive.

4. **Markov Chain Monte Carlo** (MCMC) methods are a powerful class of algorithms that can be deployed effectively to obtain the posterior. Take a numerical methods class to learn more about them!

So why aren't we all Bayesians? Well, it all boils down to the pesky prior and how you decide on $\pi(\boldsymbol{\theta})$. It's completely arbitrary, for starters. Since the posterior is the likelihood times the prior, if the likelihood dominates what the inferred posterior is, that's fine; but when it's the other way around, things can get hairy. First, different groups may disagree on what is the right choice of a prior; what range of $\boldsymbol{\theta}$ to include, for example, in inferring the posterior. Second, the prior is sensitive to how you parametrize the problem. You might think that if you set extremely broad, uniformly distributed priors, you will be as agnostic as possible, and hope that this gives you something "objective". However, consider a parameter θ that you expect to lie between 0.1 and 1. Should give a flat prior on θ itself, or on $\log(\theta)$? Choosing a flat prior in $\log(\theta)$ gives you a prior that is definitely *not* flat in θ, but which to choose is completely arbitrary. Finally, priors can unwittingly come to dominate the posterior due to the **prior volume effect**. This is a situation where the true parameters correspond to one particular point in parameter space, but the likelihood is very similar across large parts of parameter space, and also somewhat large, thus biasing the posterior away from the true values. This happens very often in cosmological searches for new physics (where setting more than 1 parameter to zero recovers standard cosmology, for example), and is precipitating a bit of shift back toward frequentist analyses in cosmology.

All in all, if the signal is large (in Bayesian terms, your likelihood dominates the posterior), then both methods should yield similar results; in which case, my personal preference between the two approaches is whichever happens to be easier. For less clear cut situations, one should always inject a healthy dose of

scepticism to any statistical analysis. With frequentist analyses, you should ask if there are reasonable priors to adopt, and if the asymptotic conditions typically required to obtain sensible results hold. For Bayesian analyses, be aware of the priors, the assumptions that go into them, and be wary about the possible dominance of the priors on the posteriors.

A Midterm 1 Review

A.1 Disk on an Inclined Plane

Consider a disk of mass m and radius a rolling without slipping down a plane inclined at an angle θ with respect to the horizontal. Denote the distance of the center of mass of the disk from the top of the plane as x, and denote the angle of rotation of the disk as ϕ. Obtain the equations of motion. Is the first integral conserved?

SOLUTION:
Rolling without slipping implies a constraint given by $x - a\phi = 0$. Therefore, the Lagrangian for this constrained system can be written as

$$L(x, \dot{x}, \phi, \dot{\phi}) = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}I\dot{\phi}^2 + mgx\sin\theta - \lambda(x - a\phi)\,, \tag{A.1}$$

where $I = ma^2/2$ is the moment of inertia of the disk, and we have inserted the Lagrange multiplier λ. Note that for the purposes of classical mechanics, *only so-called holonomic constraints* of the form $f(x, \phi) = 0$ can be treated by adding the constraint to the Lagrangian through a Lagrange multiplier. Other types of constraints require more sophisticated treatments.
Let's write down the equations of motion for this system. The Euler-Lagrange equation with respect to x is

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0 \implies \ddot{x} = g\sin\theta - \frac{\lambda}{m} \tag{A.2}$$

$$\frac{\partial L}{\partial \phi} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\phi}} = 0 \implies I\ddot{\phi} = \lambda a\,, \tag{A.3}$$

with $x = a\phi$ imposed by the constraint. We thus find

$$\ddot{\phi} = \frac{g\sin\theta - \lambda/m}{a} \implies I\frac{g\sin\theta - \lambda/m}{a} = \lambda a \implies \lambda = \frac{mg\sin\theta}{1 + ma^2/I}\,. \tag{A.4}$$

Note that since $\ddot{x} = g\sin\theta - \lambda/m$, we can see that λ is the force of friction acting on the disk as it rolls down without slipping, which is the force responsible for enforcing the constraint $x - a\phi = 0$.
What are the conserved quantities? Since the Lagrangian depends on both x and ϕ coordinates directly, the conjugate momenta to these coordinates are not conserved. However, L does not explicitly depend on time, and therefore the first integral is conserved. This is

$$\frac{\partial L}{\partial \dot{x}}\dot{x} + \frac{\partial L}{\partial \dot{\phi}}\dot{\phi} - L = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}I\dot{\phi}^2 - mgx\sin\theta + \lambda(x - a\phi) \tag{A.5}$$

$$= \frac{1}{2}m\dot{x}^2 + \frac{1}{2}I\dot{\phi}^2 - mgx\sin\theta\,, \tag{A.6}$$

where the last term drops out since the constraint is enforced on-shell. This is the translational kinetic energy plus the rotational kinetic energy plus the total gravitational potential energy, and is conserved.

A.2 (SG 10.13) Fluid Equations

Write out Euler's equation for fluid motion

$$\dot{\vec{v}} + (\vec{v} \cdot \nabla)\vec{v} = -\nabla h \tag{A.7}$$

in Cartesian tensor notation. Transform it into

$$\dot{\vec{v}} - \vec{v} \times \vec{\omega} = -\nabla \left(\frac{1}{2}\vec{v}^2 + h \right) \tag{A.8}$$

where $\vec{\omega} = \nabla \times \vec{v}$ is the vorticity. Deduce Bernoulli's theorem, that for steady $\dot{\vec{v}} = 0$ flow the quantity $\vec{v}^2/2 + h$ is constant along streamlines.

SOLUTION:
In index notation, Euler's equation reads

$$\partial_t v_i + v_j \partial_j v_i = -\partial_i h \,. \tag{A.9}$$

In order to perform the transformation, first let's observe that

$$\begin{aligned} (\vec{v} \times \vec{\omega})_i &= \epsilon_{ijk} v_j \omega_k \\ &= \epsilon_{ijk} v_j (\epsilon_{kmn} \partial_m v_n) \\ &= \epsilon_{kij}\epsilon_{kmn} v_j \partial_m v_n \\ &= (\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}) v_j \partial_m v_n \\ &= v_j \partial_i v_j - v_j \partial_j v_i \,. \end{aligned} \tag{A.10}$$

Therefore, we can write Euler's equation as

$$\partial_t v_i + v_j \partial_i v_j - \epsilon_{ijk} v_j \omega_k = -\partial_i h \,. \tag{A.11}$$

Note however that $v_j \partial_i v_j = \partial_i (v_j v_j)/2 = \partial_i \vec{v}^2/2$, and so grouping terms appropriately,

$$\partial_t v_i - \epsilon_{ijk} v_j \omega_k = -\partial_i \left(\frac{1}{2}\vec{v}^2 + h \right) \implies \dot{\vec{v}} - \vec{v} \times \vec{\omega} = -\nabla \left(\frac{1}{2}\vec{v}^2 + h \right) \tag{A.12}$$

as required.
To deduce Bernoulli's theorem, we set $\partial_t v_i = 0$, and contract both sides with v_i. This gives

$$-\epsilon_{ijk} v_i v_j \omega_k = -v_i \partial_i \left(\frac{1}{2}\vec{v}^2 + h \right) . \tag{A.13}$$

But contracting the Levi-Civita symbol with the same vector twice is zero. Therefore,

$$v_i \partial_i \left(\frac{1}{2}\vec{v}^2 + h \right) , \tag{A.14}$$

which is the statement that the quantity $\vec{v}^2/2 + h$ is constant along streamlines.

A.3 Maxwell's Equations from the Lagrangian

The Lagrangian density for the electromagnetic field is given by

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,, \tag{A.15}$$

in Minkowski space, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, and A^μ is the four-potential. The metric that we use is

$$\eta_{\mu\nu} \equiv \text{diag}(+1, -1, -1, -1)\,. \tag{A.16}$$

Obtain the equations of motion.

From the properties of the electromagnetic tensor, Prove the Bianchi identity

$$\partial^\mu F^{\alpha\beta} + \partial^\alpha F^{\beta\mu} + \partial^\beta F^{\mu\alpha} = 0\,. \tag{A.17}$$

Find the canonical energy-momentum tensor, and show explicitly that it is conserved.

SOLUTION:

Let's write out the Lagrangian density carefully as

$$\mathcal{L} = -\frac{1}{4}\eta^{\mu\alpha}\eta^{\nu\beta}\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)\left(\partial_\alpha A_\beta - \partial_\beta A_\alpha\right)\,. \tag{A.18}$$

Now let's take the derivative. We have

$$\frac{\partial \mathcal{L}}{\partial A_\sigma} = 0 \tag{A.19}$$

and

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial(\partial_\rho A_\sigma)} &= -\frac{1}{4}\eta^{\mu\alpha}\eta^{\nu\beta}\left(\delta^\rho_\mu\delta^\sigma_\nu - \delta^\rho_\nu\delta^\sigma_\mu\right)F_{\alpha\beta} - \frac{1}{4}\eta^{\mu\alpha}\eta^{\nu\beta}F_{\mu\nu}\left(\delta^\rho_\alpha\delta^\sigma_\beta - \delta^\rho_\beta\delta^\sigma_\alpha\right)\\ &= -\frac{1}{4}(F^{\rho\sigma} - F^{\sigma\rho}) - \frac{1}{4}(F^{\rho\sigma} - F^{\sigma\rho})\\ &= -F^{\rho\sigma}\,,\end{aligned} \tag{A.20}$$

where in the last line I used the fact that $F^{\rho\sigma}$ is antisymmetric.
Thus the equation of motion is

$$\partial_\rho F^{\rho\sigma} = 0\,. \tag{A.21}$$

For the Bianchi identity, we note that

$$\begin{aligned}\partial^\mu F^{\alpha\beta} &= \partial^\mu(\partial^\alpha A^\beta - \partial^\beta A^\alpha)\\ &= \partial^\alpha\partial^\mu A^\beta - \partial^\alpha\partial^\beta A^\mu + \partial^\alpha\partial^\beta A^\mu - \partial^\beta\partial^\mu A^\alpha\\ &= \partial^\alpha F^{\mu\beta} + \partial^\beta F^{\alpha\mu}\\ &= -\partial^\alpha F^{\beta\mu} - \partial^\beta F^{\mu\alpha}\,,\end{aligned} \tag{A.22}$$

and so $\partial^\mu F^{\alpha\beta} + \partial^\alpha F^{\beta\mu} + \partial^\beta F^{\mu\alpha} = 0$ as required.
The canonical energy-momentum tensor is

$$\begin{aligned}T^\mu{}_\nu &= \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\sigma)}\partial_\nu A_\sigma - \delta^\mu_\nu\mathcal{L}\\ &= -F^{\mu\sigma}\partial_\nu A_\sigma + \frac{1}{4}\delta^\mu_\nu F_{\alpha\beta}F^{\alpha\beta}\end{aligned} \tag{A.23}$$

Taking the derivative, we find

$$\partial_\mu T^\mu{}_\nu = -\partial_\mu F^{\mu\sigma}\partial_\nu A_\sigma - F^{\mu\sigma}\partial_\mu\partial_\nu A_\sigma + \frac{1}{4}\partial_\nu(F_{\alpha\beta}F^{\alpha\beta})\,. \tag{A.24}$$

The first term vanishes by the equation of motion, and we can rewrite the last term as $(1/2)\partial_\nu F_{\mu\sigma}F^{\mu\sigma}$. However, by relabeling indices, commuting derivatives, and exploiting the antisymmetry of $F^{\mu\sigma}$,

$$\begin{aligned}\frac{1}{2}\partial_\nu F_{\mu\sigma}F^{\mu\sigma} &= \frac{1}{2}\left(\partial_\nu\partial_\mu A_\sigma F^{\mu\sigma} - \partial_\nu\partial_\sigma A_\mu F^{\mu\sigma}\right) \\ &= \frac{1}{2}\left(\partial_\mu\partial_\nu A_\sigma F^{\mu\sigma} - \partial_\nu\partial_\mu A_\sigma F^{\sigma\mu}\right) \\ &= \partial_\mu\partial_\nu A_\sigma F^{\mu\sigma}\,, \end{aligned} \tag{A.25}$$

where in the last line I used $F^{\sigma\mu} = -F^{\mu\sigma}$. Therefore,

$$\partial_\mu T^\mu{}_\nu = -F^{\mu\sigma}\partial_\mu\partial_\nu A_\sigma + \partial_\mu\partial_\nu A_\sigma F^{\mu\sigma} = 0\,, \tag{A.26}$$

i.e. the canonical energy-momentum tensor is conserved.

B Midterm 2 Review

B.1 Matrix Derivatives

Consider an $n \times n$ matrix X, with each entry X_{ij} being an independent variable that can be varied. In many machine learning applications, it is useful to compute the derivative of various quantities involving the matrix X with respect to its entries. A giant compendium of such results can be found here. We now want to use index notation to demystify several of these results.

Show the following results for an invertible $n \times n$ matrix X, which you can think of representing some $(1,1)$-tensor in $\mathbb{R}^n$:

1.

$$\frac{\partial X^{-1}_{kl}}{\partial X_{ij}} = -X^{-1}_{ki} X^{-1}_{jl} \,. \tag{B.1}$$

Let me review the golden rules of index notation:

(a) Free indices (indices that are not summed over) must agree on both the left-hand side and the right-hand side of any equation.

(b) Indices that are not free should be contracted in pairs, with one upper and one lower index (unless in $\mathbb{R}^n$, where the position doesn't matter).

(c) Contracted indices can always be relabeled, since they are dummy indices.

(d) There should never be more than two of the same index appearing in a term formed by the product of a bunch of tensors, vectors etc. If this happens to you, it is likely because sum of the indices are supposed to be contracted in a sum. Relabel your contracted indices!

First, let's note that the multiplication of two matrices AB is written in index notation as $A_{ij}B_{jk}$, and we don't have to worry about index position here. We know that $X^{-1}_{ij} X_{jk} = \delta_{ik} = I$, where I is the $n \times n$ identity matrix, and therefore

$$\frac{\partial}{\partial X_{ij}} (X^{-1}_{kl} X_{lm}) = \frac{\partial}{\partial X_{ij}} \delta_{km} = 0 \,. \tag{B.2}$$

On the other hand, using Leibniz's rule gives

$$\begin{aligned} \frac{\partial}{\partial X_{ij}} (X^{-1}_{kl} X_{lm}) &= \frac{\partial X^{-1}_{kl}}{\partial X_{ij}} X_{lm} + X^{-1}_{kl} \frac{\partial X_{lm}}{\partial X_{ij}} \\ &= \frac{\partial X^{-1}_{kl}}{\partial X_{ij}} X_{lm} + X^{-1}_{kl} \delta_{il} \delta_{jm} \end{aligned} \tag{B.3}$$

From here, we see that

$$\begin{aligned}\frac{\partial X^{-1}_{kl}}{\partial X_{ij}} X_{lm} &= -X^{-1}_{kl}\delta_{il}\delta_{jm} \\ \frac{\partial X^{-1}_{kl}}{\partial X_{ij}} X_{lm} X^{-1}_{mp} &= -X^{-1}_{kl}\delta_{il}\delta_{jm}X^{-1}_{mp} \\ \frac{\partial X^{-1}_{kl}}{\partial X_{ij}} \delta_{lp} &= -X^{-1}_{ki}X^{-1}_{jp} \\ \frac{\partial X^{-1}_{kp}}{\partial X_{ij}} &= -X^{-1}_{ki}X^{-1}_{jp}\,, \end{aligned} \tag{B.4}$$

which you can check is equivalent to the required expression by just a relabeling of the indices.

2. For a constant matrix A, show explicitly that $\mathrm{Tr}(XA)$ is a scalar if these matrices are taken to represent tensors, and that

$$\frac{\partial}{\partial X_{ij}}\mathrm{Tr}(XA) = (A^T)_{ij} = A_{ji}\,, \tag{B.5}$$

where ^{T} denotes the transpose.

First, let's write the trace properly in index notation, paying some attention to the placement of indices: $\mathrm{Tr}(XA) = X^{ij}A_{ji}$. Under a coordinate transformation,

$$\begin{aligned} X^{ij} &\mapsto X'^{ij} = a^i{}_k a^j{}_l X^{kl} \\ A_{ji} &\mapsto A'_{ji} = (a^{-1})^k{}_j (a^{-1})^l{}_i A_{kl} \end{aligned} \tag{B.6}$$

Therefore,

$$\begin{aligned} X^{ij}A_{ji} &\mapsto (a^{-1})^p{}_i a^i{}_k (a^{-1})^m{}_j a^j{}_l X^{kl} A_{mp} \\ &= \delta^p_k \delta^m_l X^{kl} A_{mp} \\ &= X^{kl}A_{lk}\,, \end{aligned} \tag{B.7}$$

i.e. $X^{ij}A_{ji}$ does not transform under a coordinate transformation and is a scalar. Notice how I relabeled the contracted indices on A' to avoid having more than two of the same index.
At this point, we can see immediately that

$$\frac{\partial}{\partial X_{ij}} X^{kl}A_{lk} = \delta^k_i \delta^l_j A_{lk} = A_{ji}\,, \tag{B.8}$$

proving the relation. Notice again that I was careful to label XA using indices k and l, which again avoids having more than two of the same index.

3. The following relation is true for any $n \times n$ matrix X:

$$X_{jk}\frac{\partial}{\partial X_{ik}}\det(X) = \delta_{ij}\det(X)\,. \tag{B.9}$$

Show that it is true for the case where $n = 3$.

The determinant in the 3×3 case can be written as

$$\det(X)\epsilon_{lmn} = \epsilon^{ijk} X_{il} X_{jm} X_{kn} \,. \tag{B.10}$$

Taking the derivative, we find

$$\begin{aligned}
\frac{\partial}{\partial X_{pq}} \det(X)\epsilon_{lmn} =& \frac{\partial}{\partial X_{pq}} \epsilon^{ijk} X_{il} X_{jm} X_{kn} \\
=& \epsilon^{ijk} \left(\delta_{ip}\delta_{lq} X_{jm} X_{kn} + X_{il}\delta_{jp}\delta_{mq} X_{kn} \right. \\
& \left. + X_{il} X_{jm} \delta_{kp}\delta_{nq} \right) \\
X_{rq} \frac{\partial}{\partial X_{pq}} \det(X)\epsilon_{lmn} =& \epsilon^{ijk} \left(\delta_{ip}\delta_{lq} X_{jm} X_{kn} + X_{il}\delta_{jp}\delta_{mq} X_{kn} \right. \\
& \left. + X_{il} X_{jm} \delta_{kp}\delta_{nq} \right) X_{rq}
\end{aligned} \tag{B.11}$$

Multiplying both sides by ϵ^{lmn}, and noting that $\epsilon_{lmn}\epsilon^{lmn} = 3!$, we find

$$\begin{aligned}
3! X_{rq} \frac{\partial}{\partial X_{pq}} \det(X) =& \epsilon^{ijk}\epsilon^{lmn} \left(\delta_{ip}\delta_{lq} X_{jm} X_{kn} + X_{il}\delta_{jp}\delta_{mq} X_{kn} \right. \\
& \left. + X_{il} X_{jm} \delta_{kp}\delta_{nq} \right) X_{rq} \\
=& \epsilon^{pjk}\epsilon^{lmn} X_{rl} X_{jm} X_{kn} + \epsilon^{ipk}\epsilon^{lmn} X_{il} X_{rm} X_{kn} \\
& + \epsilon^{ijp}\epsilon^{lmn} X_{il} X_{jm} X_{rn} \,.
\end{aligned} \tag{B.12}$$

You can show that for example $\epsilon^{lmn} X_{rl} X_{jm} X_{kn} = \det(X)\epsilon_{rjk}$ from Eq. (B.10), by noting that $\det(A) = \det(A^T)$. Thus, we find

$$\begin{aligned}
3! X_{rq} \frac{\partial}{\partial X_{pq}} \det(X) &= (\epsilon^{pjk}\epsilon_{rjk} + \epsilon^{ipk}\epsilon_{irk} + \epsilon^{ijp}\epsilon_{ijr}) \det(X) \\
&= 3\epsilon^{pjk}\epsilon_{rjk} \det(X) \\
&= 3 \times 2\delta_{rp} \det(X) \,,
\end{aligned} \tag{B.13}$$

giving finally

$$X_{rq} \frac{\partial}{\partial X_{pq}} \det(X) = \delta_{pr} \det(X) \,, \tag{B.14}$$

as required.

B.2 2+1D Electrodynamics

Consider electrodynamics with 1 time-like dimension, and 2 space-like dimensions. The electromagnetic field strength two form is still defined as

$$F = -\frac{1}{2} F_{\mu\nu} \, \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \,, \tag{B.15}$$

with coordinates $(x^0, x^1, x^2) \equiv (t, x, y)$, and metric $g = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2$, adopting the mostly minus convention. In terms of the components of the usual electric and magnetic fields,

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y \\ -E_x & 0 & -B \\ -E_y & B & 0 \end{pmatrix} \,. \tag{B.16}$$

Furthermore, let the one-form current $J = \rho \, \mathrm{d}t - j_x \, \mathrm{d}x - j_y \, \mathrm{d}y$, where $J^\mu = (\rho, \vec{j})$ are the usual charge density and current density vector respectively. Note that

we can write

$$F = -E_x \, \mathrm{d}t \wedge \mathrm{d}x - E_y \, \mathrm{d}t \wedge \mathrm{d}y + B \, \mathrm{d}x \wedge \mathrm{d}y \, . \tag{B.17}$$

1. Write down Maxwell's equations, $\mathrm{d}F = 0$ and $\mathrm{d} \star F = \star J$, in component form for 2+1D electrodynamics.

   First, we see that

   $$\mathrm{d}F = (-\partial_y E_x + \partial_x E_y + \partial_t B)\mathrm{d}t \wedge \mathrm{d}x \wedge \mathrm{d}y = 0 \, , \tag{B.18}$$

   which in index notation reads

   $$\epsilon^{ij} \partial_i E_j = -\partial_t B \, , \tag{B.19}$$

   where ϵ^{ij} is the two-dimensional Levi-Civita tensor. This is the equivalent of Faraday's law. Next, we find

   $$\star F = E_x \, \mathrm{d}y - E_y \, \mathrm{d}x + B \, \mathrm{d}t \, , \tag{B.20}$$

   and therefore

   $$\begin{aligned} \mathrm{d} \star F =& \partial_x E_x \, \mathrm{d}x \wedge \mathrm{d}y + \partial_t E_x \, \mathrm{d}t \wedge \mathrm{d}y - \partial_t E_y \, \mathrm{d}t \wedge \mathrm{d}x \\ & - \partial_y E_y \, \mathrm{d}y \wedge \mathrm{d}x + \partial_x B \, \mathrm{d}x \wedge \mathrm{d}t + \partial_y B \, \mathrm{d}y \wedge \mathrm{d}t \\ =& - (\partial_t E_y + \partial_x B)\mathrm{d}t \wedge \mathrm{d}x + (\partial_t E_x - \partial_y B)\mathrm{d}t \wedge \mathrm{d}y \\ & + (\partial_x E_x + \partial_y E_y)\mathrm{d}x \wedge \mathrm{d}y \end{aligned} \tag{B.21}$$

   On the other hand,

   $$\star J = \rho \, \mathrm{d}x \wedge \mathrm{d}y - j_x \mathrm{d}t \wedge \mathrm{d}y + j_y \, \mathrm{d}t \wedge \mathrm{d}x \, . \tag{B.22}$$

   Comparing both expressions, we find the equivalent of Gauss's law,

   $$\partial_x E_x + \partial_y E_y = \rho \, , \tag{B.23}$$

   and the equivalent of Ampere's law,

   $$\begin{aligned} \partial_x B &= -j_y - \partial_t E_y \, , \\ \partial_y B &= j_x + \partial_t E_x \, . \end{aligned} \tag{B.24}$$

2. Derive the continuity equation.

   The continuity equation can be derived from the fact that $\mathrm{d} \star J = \mathrm{d}^2 \star F = 0$. But

   $$\mathrm{d} \star J = (\partial_t \rho + \partial_x j_x + \partial_y j_y)\mathrm{d}t \wedge \mathrm{d}x \wedge \mathrm{d}y = 0 \, , \tag{B.25}$$

   and we obtain the continuity equation

   $$\partial_t \rho + \partial_x j_x + \partial_y j_y = 0 \, . \tag{B.26}$$

B.3 Complex Numbers, Cauchy-Riemann Equations, Harmonic Functions

1. Show that for $z = x + iy$,

   $$\mathrm{Re}[\log(z-1)] = \frac{1}{2} \ln[(x-1)^2 + y^2] \, . \tag{B.27}$$

Why must this function satisfy Laplace's equation when $z \neq 1$?

First,

$$\begin{aligned}\log(z-1) &= \log((x-1)+iy) \\ &= \ln R + i(\Theta + 2k\pi)\,, k \in \mathbb{Z}\,,\end{aligned} \tag{B.28}$$

where $R = \sqrt{(x-1)^2+y^2}$, and $\Theta = \tan^{-1}((y-1)/x)$. Therefore,

$$\mathrm{Re}[\log(z-1)] = \ln\sqrt{(x-1)^2+y^2} = \frac{1}{2}\ln[(x-1)^2+y^2]\,, \tag{B.29}$$

as required. It must be satisfy Laplace's equation since it is the real part of the function $\log(z-1)$, which is analytic except at $z=1$.

2. Find the values of $(1+i)^i$, as well as the principal value.

$$\begin{aligned}(1+i)^i &= \exp\left(i\log(1+i)\right) \\ &= \exp\left[i\log\left(\sqrt{2}e^{i\pi/4}\right)\right] \\ &= \exp\left[i\left(\ln\sqrt{2} + i\left(\frac{\pi}{4}+2k\pi\right)\right)\right]\,, k\in\mathbb{Z} \\ &= \exp\left(-\frac{\pi}{4}-2k\pi\right)e^{i\ln\sqrt{2}}\,, k\in\mathbb{Z}\,.\end{aligned} \tag{B.30}$$

The principal value is obtained by using Log instead of $\log$, giving

$$\mathrm{P.V.}(1+i)^i = e^{-\pi/4}e^{i\ln\sqrt{2}}\,. \tag{B.31}$$

3. Let $u(x,y) = \ln(x^2+y^2)$, $x^2+y^2 > 0$. Show that u is harmonic, and find its harmonic conjugate.

$$\partial_x u = \frac{2x}{x^2+y^2} \implies \partial_x^2 u = \frac{2}{x^2+y^2} - \frac{4x^2}{(x^2+y^2)^2}\,, \tag{B.32}$$

and similarly

$$\partial_y^2 u = \frac{2}{x^2+y^2} - \frac{4y^2}{(x^2+y^2)^2} \tag{B.33}$$

by symmetry. We now see that

$$(\partial_x^2+\partial_y^2)u = \frac{4}{x^2+y^2} - \frac{4(x^2+y^2)}{(x^2+y^2)^2} = 0\,, \tag{B.34}$$

and so $(\partial_x^2+\partial_y^2)u = 0$, and u is harmonic. To find the harmonic conjugate, we need to find $v(x,y)$ such that $\partial_x u = \partial_y v$, and $\partial_y u = -\partial_x v$. We have

$$\partial_x v = -\frac{2y}{x^2+y^2} \implies v = -2\tan^{-1}\left(\frac{x}{y}\right) + f(y)\,, \tag{B.35}$$

where $f(y)$ is an arbitrary function of y (including possible constants). Similarly,

$$\partial_y v = \frac{2x}{x^2+y^2} \implies v = 2\tan^{-1}\left(\frac{y}{x}\right) + g(x) \tag{B.36}$$

with $g(x)$ being another arbitrary function, this time of x (including possible constants). Recall from trigonometry that

$$\tan^{-1}(1/x) + \tan^{-1}(x) = \frac{\pi}{2}\,, \tag{B.37}$$

and so we see that the harmonic conjugate to $u(x,y)$ is

$$v(x,y) = 2\tan^{-1}\left(\frac{y}{x}\right) + C \tag{B.38}$$

for some arbitrary constant C.

B.4 Complex Integration I

Let C be the contour, consisting of the following four pieces:

1. From -2 to -1 along a straight line.
2. From -1 to $+1$ along the semicircular arc (radius 1, centered at 0) lying above the real axis.
3. From 1 to 2 along a straight line.
4. From 2 to -2 along the semicircular arc (radius 2, centered at 0) lying above the real axis.

Compute

$$\int_C dz \frac{z}{z^*}\,. \tag{B.39}$$

Along leg 1, we have

$$\int_{-2}^{-1} dx\,\frac{x}{x} = 1\,. \tag{B.40}$$

Along leg 2, we have $z = e^{i\theta}$, and

$$\int_{\pi}^{0} d\theta\,\frac{dz}{d\theta}\frac{e^{i\theta}}{e^{-i\theta}} = i\int_{\pi}^{0} d\theta\, e^{i3\theta} = \frac{1}{3}(1-e^{i3\pi}) = \frac{2}{3}\,. \tag{B.41}$$

Along leg 3, we have once again

$$\int_{1}^{2} dx\,\frac{x}{x} = 1\,. \tag{B.42}$$

Finally, for leg 4, we have $z = 2e^{i\theta}$, and

$$\int_{0}^{\pi} d\theta\,\frac{dz}{d\theta}\frac{2e^{i\theta}}{2e^{-i\theta}} = 2i\int_{0}^{\pi} e^{i3\theta} = \frac{2}{3}(e^{i3\pi}-1) = -\frac{4}{3}\,, \tag{B.43}$$

with the result being very similar to above. Therefore,

$$\int_C dz\frac{z}{z^*} = \frac{4}{3}\,. \tag{B.44}$$

B.5 Complex Integration II

Let's evaluate

$$\int_C dz\, z^{a-1} \tag{B.45}$$

for some nonzero real number a, where C is the positively oriented circle $z = Re^{i\theta}$, with $-\pi \leq \theta \leq \pi$. We choose the principal branch with $-\pi < \arg(z) < \pi$.

The integral is simply

$$\begin{aligned}
\int_{-\pi}^{\pi} d\theta\, \frac{dz}{d\theta}\, z^{a-1} &= \int_{-\pi}^{\pi} d\theta\, iRe^{i\theta} \cdot R^{a-1} e^{i(a-1)\theta} \\
&= iR^a \int_{-\pi}^{\pi} d\theta\, e^{ia\theta} \\
&= iR^a \left[\frac{e^{ia\theta}}{ia}\right]_{-\pi}^{\pi} \\
&= \frac{R^a}{a}\left(e^{ia\pi} - e^{-ia\pi}\right) \\
&= \frac{R^a}{a}\left(2i\sin(a\pi)\right) \\
&= \frac{2i}{a} R^a \sin(a\pi)\,.
\end{aligned} \tag{B.46}$$

References

[1] Philippe G. Ciarlet and Cristinel Mardare. "On the Brachistochrone Problem". In: *Communications in Mathematical Analysis and Applications* 1.1 (2022), pp. 213–240. ISSN: 2790-1939. DOI: https://doi.org/10.4208/cmaa.2021-0005. URL: http://global-sci.org/intro/article_detail/cmaa/20161.html.

[2] Máximo Bañados and Ignacio A. Reyes. "A short review on Noether's theorems, gauge symmetries and boundary terms". In: *Int. J. Mod. Phys. D* 25.10 (2016), p. 1630021. DOI: 10.1142/S0218271816300214. arXiv: 1601.03616 [hep-th].

[3] M. R. Flannery. "The enigma of nonholonomic constraints". In: *American Journal of Physics* 73.3 (Mar. 2005), pp. 265–272. ISSN: 0002-9505. DOI: 10.1119/1.1830501. eprint: https://pubs.aip.org/aapt/ajp/article-pdf/73/3/265/7531446/265_1_online.pdf. URL: https://doi.org/10.1119/1.1830501.

[4] July 2016. URL: https://mathematica.stackexchange.com/questions/77855/finding-basis-of-isotropic-tensors-of-rank-n.

[5] L.H. Ryder. *Quantum Field Theory*. Cambridge University Press, 1996. ISBN: 9780521478144. URL: https://books.google.com/books?id=nnuW_kVJ500C.