Applied Mathematics for Chemists II

Vector fields, partial differentiation, cylindrical and spherical coordinates, multiple integrals, line integrals, the wave and the Schrödinger equations, separation of variables method. Inner Product Spaces. Fourier Series.

By

Colin Roberts

Colorado State University

Created in Fall 2019

Contents

Part

Part V

Further Topics in Linear Algebra

1

Complex Functions and Transformations

1 Introduction

When we ended the prequel with linear algebra, we found the complex number system to be highly useful in many ways. We'll want to keep this in mind as we progress into further topics in the field of linear algebra. Instead of dealing with transformations of finite dimensional vector spaces like $\mathbb{R}^n$ and $\mathbb{C}^n$, we will care about the spaces of functions on these spaces. So we find ourselves studying a bit of a nested structure.

Spaces of functions are of great importance. In studying these spaces, we find ways to solve problems we will approach in the future (e.g., partial differential equations). These spaces are, in some sense, infinite dimensional which means we can no longer draw pictures that accurately describe what is occuring. Luckily enough, the intuition gained from the finite dimensional case will work just fine.

We begin with complex functions as they are immensely fundamental in the study of the physical world and our mathematical development. Once we have covered this area, we can adjust our view to the relevant spaces of functions that arise in areas such as quantum mechanics and partial differential equations in general. As we did in the finite dimensional case, we can consider how these linear spaces transform under linear operators. Finally, we make a nudge towards the spectral theory (eigenvalues and eigenvectors) via Fourier theory.

2 Complex Functions

In the prequel, we studied in depth single variable real valued functions $f\colon \mathbb{R} \to \mathbb{R}$. That is, functions with a single real variable as an input that outputs a single real number. Analogously, a ***complex function*** is a function, $f\colon \mathbb{C} \to \mathbb{C}$, with a complex number given as input and a complex number output as well. The interesting quality to note is

that we specified a complex number $z \in \mathbb{C}$ by putting

$$z = x + iy,$$

which means that single complex number is defined by two real numbers. Recall as well that we could write a complex number in polar form

$$z = re^{i\theta},$$

which again requires the specification of two real numbers. All of this is to say that we are allowed to (when it is helpful) think of complex functions as functions that input two real numbers $x, y \in \mathbb{R}$ and outputs two real numbers. Hence we would write $f\colon \mathbb{R}^2 \to \mathbb{R}^2$. The additional structure with complex numbers (in how we multiply them) forces us to think of $f\colon \mathbb{C} \to \mathbb{C}$ in a slightly different manner than their real valued counterparts which is why we cannot always make this identification!

2.1 Cartesian and Polar Representations

Consider a complex function $f\colon \mathbb{C} \to \mathbb{C}$. Then, as always, we define this function by providing an output for each input and specify this by

$$f(z) = w,$$

where both $w \in \mathbb{C}$ and $z \in \mathbb{C}$ are complex numbers. Hence, we can further decompose this function by writing

$$f(z) = u(z) + iv(z),$$

where $u(z)$ and $v(z)$ are real valued functions $u, v\colon \mathbb{C} \to \mathbb{R}$. This decomposition is rather helpful in providing us a way to visualize the complex function f. In this case, we are seeing what happens to the real $u(z)$ and imaginary part $v(z)$ of the output as we vary the complex input.

Of course, we can also write

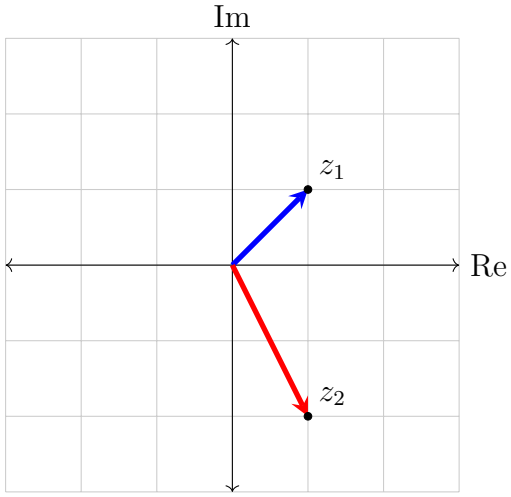
$$f(z) = r(z)e^{i\theta(z)},$$

where again $r, \theta\colon \mathbb{C} \to \mathbb{R}$. In this perspective, we are seeing what happens to the argument $\theta(z)$ and modulus $r(z)$ as we vary the complex input. Which way of decomposing f we choose is typically decided on the situation at hand. It has more to do with the symmetry of the function than anything else! In this polar representation of the function, we refer to $\theta(z)$ as a ***phase***.

When we try to plot a complex function $f\colon \mathbb{C} \to \mathbb{C}$, we run into a bit of an issue. When we plot real functions $g\colon \mathbb{R} \to \mathbb{R}$, we can draw this in a 2-dimensional plane. However, if we were to draw a complex function using the same idea, it would be in a 4-dimensional space. This cannot work. But that doesn't mean we are out of options!

Remember, we can think of a point z in the complex plane as a 2-dimensional vector.

So, for example, if we take $z_1 = 1 + i$ and $z_2 = 1 - 2i$, we can plot this like



Using this idea, let us see how we can plot a complex function in a different way.

Example 2.1: Plotting a Complex Function

Let's consider a complex function $f\colon \mathbb{C} \to \mathbb{C}$ given by the function

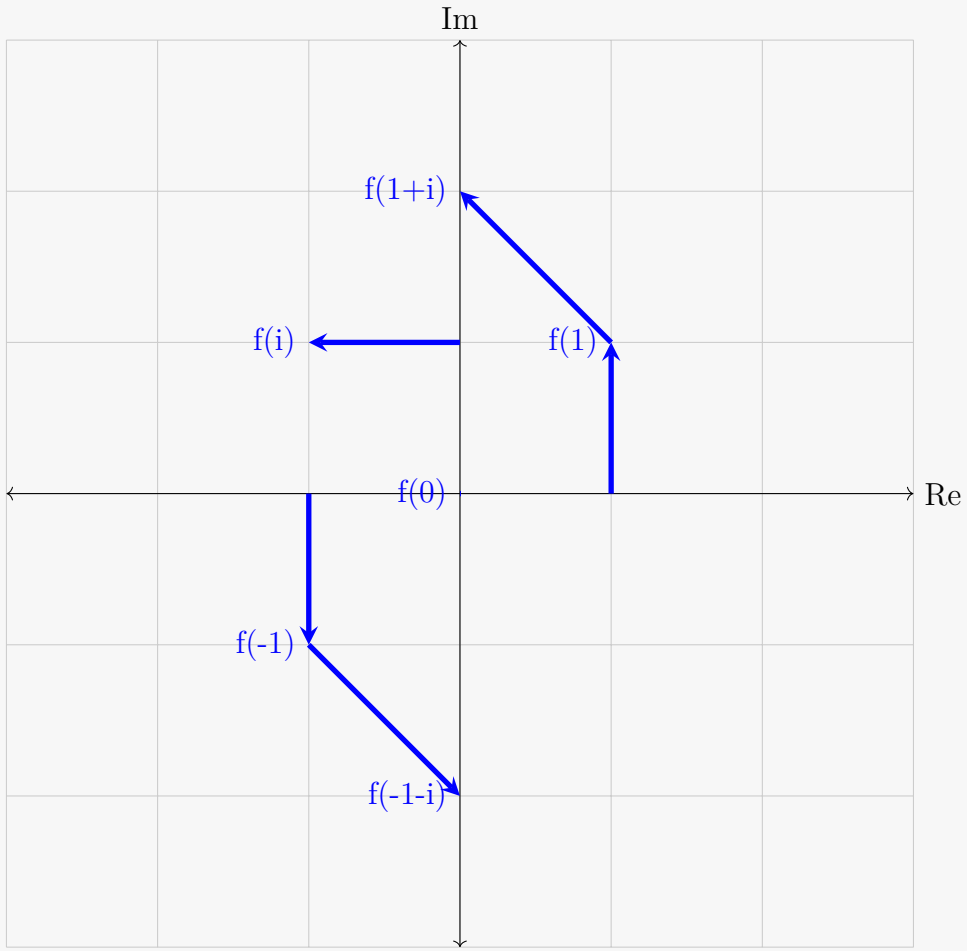
$$f(z) = iz = -\mathrm{Im}(z) + i\mathrm{Re}(z).$$

Recall that multiplication by i rotates a complex number by an angle of $\pi/2$ in the counterclockwise direction. Thinking this way will help us understand what this function is doing. For some more concrete results, we should compute a few values for this function

$$\begin{array}{ccc} f(0) = 0 & f(1) = i & f(i) = -1 \\ f(-1) = -i & f(1+i) = -1 + i & f(-1-i) = 1 - i. \end{array}$$

We can plot these values as vectors emanating from the input point z. That is, we

can place the arrow given by the output $f(z)$ at the point z.



We can then do this for many more input points to get a picture like the following.

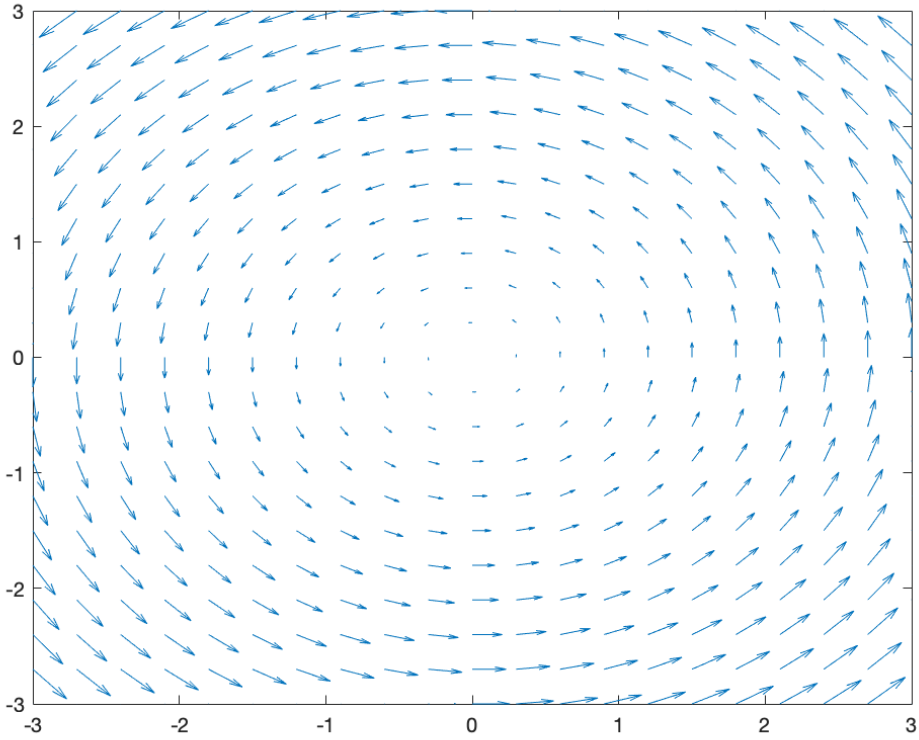


Figure 1.1: A plot of various different outputs for their corresponding inputs.

What we see here is typically referred to as a vector field. We will get to this notion later on in the text.

2.2 Complex Valued Functions

A major focus in this course is understanding the mathematics behind quantum mechanics. For a chemist, this knowledge is rather important since modern theory is mostly quantum in nature. What isn't quantum is likely thermodynamical or electrodynamical in nature and we will get to these topics a bit later on.

Recall that wavefunctions are solutions to Schrödinger's equation. In the broadest generality, wavefunctions are functions that are complex valued and whose domain of definition is on some region Ω in space $\mathbb{R}^3$. More generally, we can allow for Ω to a be a region in other spaces as well. To restate this, we are considering a function of the form $\Psi\colon \Omega \to \mathbb{C}$ where we will specify what the domain Ω is. Previously, we looked at models in lower dimensions (e.g., the free particle in the 1-dimensional box) since we have yet to properly discuss multivariate functions.

For now, consider a complex function $\Psi\colon [a,b] \to \mathbb{C}$ that has a single real variable as an input. Thus, we define this function by $\Psi(x) = z$, where $z \in \mathbb{C}$. Of course, we get the Cartesian decomposition

$$\Psi(x) = u(x) + iv(x),$$

or the polar decomposition

$$\Psi(x) = r(x)e^{i\theta(x)}.$$

The great thing in this case is that we can differentiate and integrate wavefunctions in a way that's no different than single variable real functions! Fundamentally, this is due to the fact that our understanding of the derivative has only been defined for a single real value input. We will deepen our understanding later. So, for a wavefunction we have that

$$\Psi'(x) = u'(x) + iv'(x),$$

and in the polar case we have

$$\Psi'(x) = r'(x)e^{i\theta(x)} + r(x)e^{i\theta(x)}\theta'(x),$$

which follows from the chain rule.

Exercise 2.1. Verify the polar derivative above is correct.

Integration follows the fundamental theorem of calculus and hence we have

$$\int_a^b \Psi'(x)dx = \Psi(b) - \Psi(a).$$

So, for example, in the cartesian representation we have

$$\int_a^b \Psi'(x)dx = \int_a^b u'(x)dx + i\int_a^b v'(x)dx = [u(b) - u(a)] + i[v(b) - v(a)].$$

Remark 2.1. Complex functions (i.e., functions with complex valued inputs) have different behavior with integration and differentiation which we will not discuss at all. The closest we will get to this structure is calculus in $\mathbb{R}^2$.

Example 2.2: A Line in $\mathbb{C}$

Let's consider the complex function $f : \mathbb{R} \to \mathbb{C}$ given by the Cartesian representation

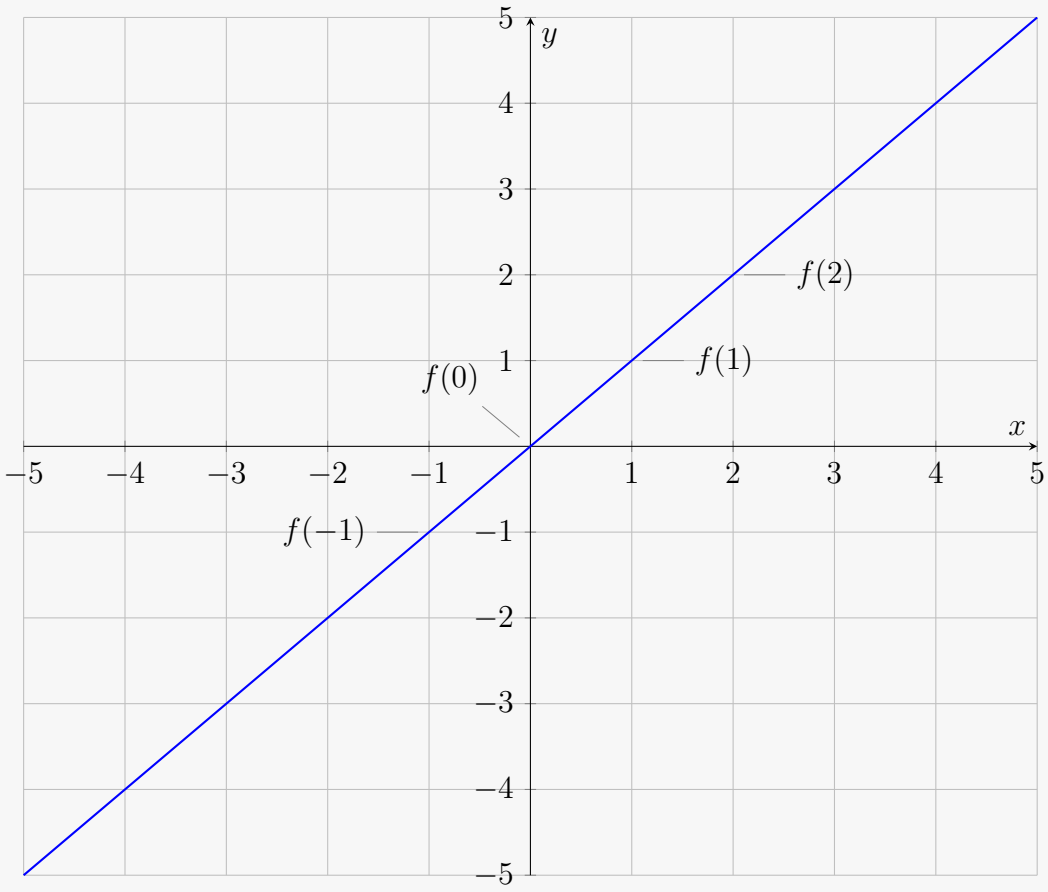
$$f(x) = x + ix.$$

How should we think of this function? For one, we can see how the output changes as the input changes by testing a few values

$$f(0) = 0 \qquad f(1) = 1 + i$$
$$f(-1) = -1 - i \qquad f(2) = 2 + 2i.$$

We can also visualize this function in the following way. For every input, we will just place the complex output into the complex plane. Unlike plotting real functions, we

will have to pay a bit more attention to what the input value is.



We often refer to this type of function as a curve or, in the complex case specifically, a contour. Again, we will revisit curves later on in this text.

Example 2.3: Wavefunctions in the Box

Let $\Omega = [0, L]$ and recall that the normalized states of the particle in the 1-dimensional box were given by

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right).$$

Recall as well that we could write a wavefunction as a superposition of states by

$$\Psi(x) = \sum_{n=0}^{\infty} a_n \psi_n(x).$$

Though these states are real valued, there is no physics that requires this. Similarly, the coefficients a_n are also not constrained to be real valued constants either. In the broadest generality, Ψ can be a complex valued function and the coefficients a_n can be complex as well.
Fundamentally, this is due to the physical understanding of the solutions to Schrödinger's equation. When we are looking for physically meaningful interpre-

tations of a wavefunction, we must evaluate an integral. We can think of this act of integration as performing a measurement. For example, let $[a, b]$ be a subinterval of $[0, L]$, then we can compute the probability of the particle with wavefunction $\Psi(x)$ to be in the region $[a, b]$ by

$$P_{[a,b]}(\Psi) = \int_a^b \|\Psi(x)\|^2 dx,$$

where we have the pointwise modulus of the complex valued function

$$\|\Psi(x)\|^2 = \Psi^*(x)\Psi(x),$$

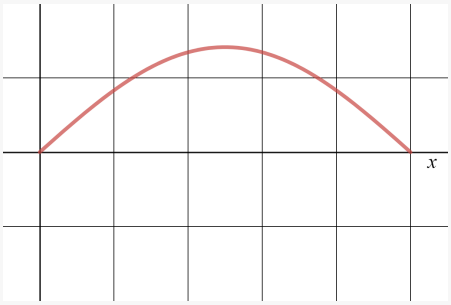
where * indicates the complex conjugate. Say we take the cartesian representation for $\Psi(x)$ by $\Psi(x) = u(x) + iv(x)$, then

$$\|\Psi(x)\|^2 = u^2(x) + v^2(x).$$

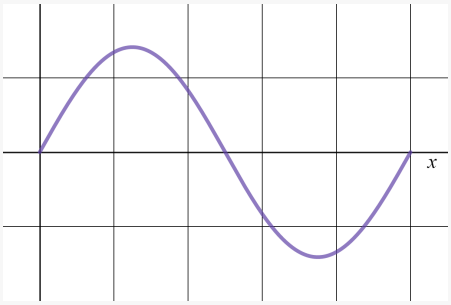
Let $\Psi(x) = \frac{1}{\sqrt{2}}\psi_1(x) + \frac{1}{\sqrt{2}}\psi_2(x)$ be a superposition state. We can compute the probability that the particle is in the first half of the region $[0, L]$ by computing

$$\begin{aligned}
P_{[0,L/2]}(\Psi) &= \int_0^{L/2} \|\Psi(x)\|^2 dx \\
&= \int_0^{L/2} \frac{1}{2}\psi_1^2(x) + \psi_1(x)\psi_2(x) + \frac{1}{2}\psi_2^2(x) dx \\
&= \int_0^{L/2} \frac{1}{2}\left(\sqrt{\frac{2}{L}}\sin\left(\frac{\pi x}{L}\right)\right)^2 + \sqrt{\frac{2}{L}}\sin\left(\frac{\pi x}{L}\right)\sin\left(\frac{2\pi x}{L}\right) + \frac{1}{2}\left(\sqrt{\frac{2}{L}}\sin\left(\frac{2\pi x}{L}\right)\right)^2 dx \\
&= \frac{1}{2} + \frac{4}{3\pi} \\
&\approx .924.
\end{aligned}$$

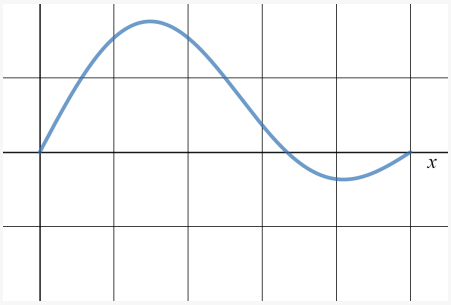
Through this calculation we have found that the probability that the particle is the first half of box is about 92.5%. Since the particle must be in the box, it follows that the probability of the particle being in $[L/2, L]$ must be $1 - \frac{1}{2} - \frac{4}{3\pi}$ or roughly 7.5%. This is quite different than we would expect classically!
We can plot the functions used above to see why this is the case.



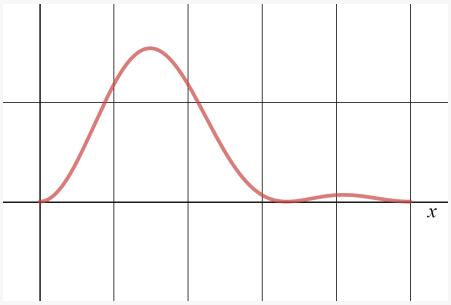
(a) Normalized state $\psi_1(x)$.



(b) Normalized state $\psi_2(x)$.



(c) The wavefunction $\Psi(x)$.



(d) The probability function $\|\Psi(x)\|^2$.

The plots above of course show us that the integration makes sense. We can see in (d) that the function that describes the probability is heavily weighted towards the first half of the interval $[0, L]$. One may then wonder if this is always true? That is, if I were to check back later in time, is the probability still distributed in the same way? The answer is no. Later, we will introduce the time dependent version of the Schrödinger equation where we will see that these wavefunctions also evolve over time. To some extent, we can see a bit of this behavior now.

If we instead change our wavefunction by introducing a phase difference for each of the components. What will happen in this case? If you have seen the double slit

experiment, you may guess that introducing a phase difference can change the result (as phase difference causes interference). Instead of the $\Psi(x)$ above, take

$$\tilde{\Psi}(x) = \frac{e^{i\theta}}{\sqrt{2}}\psi_1(x) + \frac{e^{i\phi}}{\sqrt{2}}\psi_2(x).$$

In this case, all we have done is made the wavefunction complex. If, however, we consider the probability distribution given by this new wave function, we find

$$\|\tilde{\Psi}(x)\|^2 = \tilde{\Psi}^*(x)\tilde{\Psi}(x) = \frac{1}{2}\psi_1^2(x) + \frac{e^{i(\theta-\phi)}}{2}\psi_1(x)\psi_2(x) + \frac{e^{i(\phi-\theta)}}{2}\psi_1(x)\psi_2(x) + \frac{1}{2}\psi_2^2(x).$$

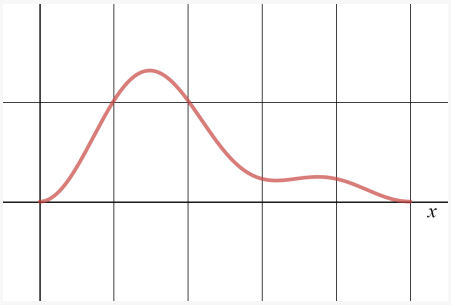
This is now slightly different! However, we can note that

$$\frac{e^{i(\theta-\phi)} + e^{i(\phi-\theta)}}{2} = \cos(\theta - \phi),$$

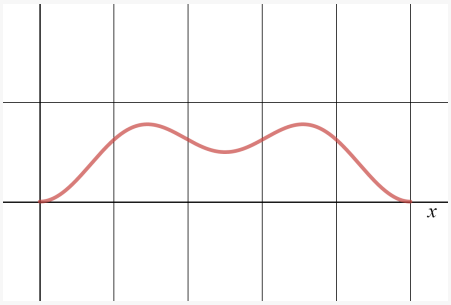
and hence we have

$$\|\tilde{\Psi}(x)\|^2 = \tilde{\Psi}^*(x)\tilde{\Psi}(x) = \frac{1}{2}\psi_1^2(x) + \cos(\theta-\phi)\psi_1(x)\psi_2(x) + \frac{1}{2}\psi_2^2(x).$$

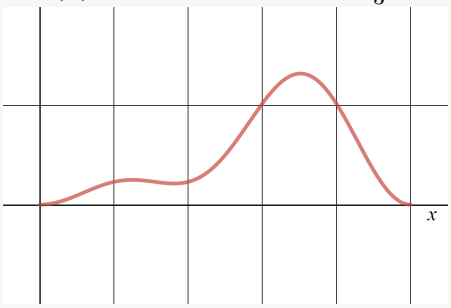
So the phase difference $\delta = |\theta - \phi|$ between the two states causes the wavefunction to change. In the first example with $\Psi(x)$, the phase difference $\delta = 0$ and we observed the probability function $\|\Psi(x)\|^2$. However, let us see what happens as we change the phase.



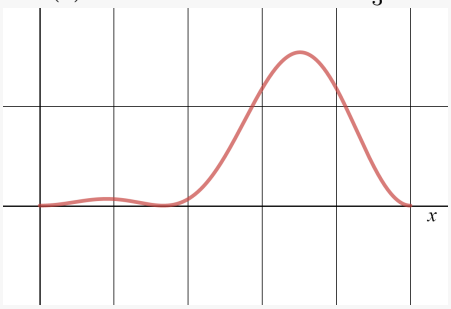
(a) Phase difference $\delta = 0$.



(b) Phase difference $\delta = \frac{\pi}{3}$.



(c) Phase difference $\delta = \frac{2\pi}{3}$.



(d) Phase difference $\delta = \pi$.

Interestingly enough, it seems that the phase difference "moves" the particle around in the box. Of course, the particle itself is not moving, but the function that represents the likelihood of its position changes as the phase changes. The largest difference in phase is π, and when we see this, we find that the distribution given by $\|\tilde{\Psi}(x)\|^2$ is the mirror image of the original $\|\Psi(x)\|^2$.
There are two important remarks to note here.

1. We can change the global phase of the system without changing the probability of measurement. That is, $e^{i\theta}\Psi(x)$ has no discernable difference from $\Psi(x)$ (you can verify this from the work above).

2. This difference in phase seems to drive some form of motion for a particle. It is with this insight that we will later revisit the time dependent version of the Schrödinger equation and see how the time component relates to phase.

Remark 2.2. In a sense, the integral defined above $P_{[a,b]}(\Psi)$ is a real valued function with a function as an input. Though we have not noted this until now, it becomes important in the future.

All of this is to say that we must be able to work with complex valued functions. They show up in physics and help us describe what we observe through nature. It's important to remember that all measurements we make in a lab must be real valued, and so our mathematical models for these measurements must take that into account as well.

2 Hilbert Spaces

1 Introduction

Recall the importance of the dot product in space. Given two vectors $\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \in \mathbb{R}^3$, we defined the dot product by

$$\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = u_1 v_1 + u_2 v_2 + u_3 v_3,$$

and we also referred to this as an inner product. The dot product allowed us to project a vector onto its components by, for example,

$$\vec{\boldsymbol{u}} \cdot \hat{\boldsymbol{x}} = u_1.$$

This was extremely useful for us. On top of that, the dot product provided us a means of computing the length of a vector by putting

$$\|\vec{\boldsymbol{u}}\| = \sqrt{\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{u}}}.$$

Underlying much of the theory of space was this structure.

Later, we introduced the Hermitian inner product on complex vectors. As it turns out, this inner product is strictly more general than the dot product. If we had two vectors $\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \in \mathbb{C}^n$ (i.e., vectors with n complex number entries) then we defined the inner product by

$$\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \rangle = \sum_{j=1}^{n} a_j b_j^*.$$

Note that if $\vec{\boldsymbol{a}}$ and $\vec{\boldsymbol{b}}$ only have real entries, then the complex conjugate $b_j^* = b_j$ and we are left with the typical dot product for $\mathbb{R}^n$. It suffices to say, that we need only care about

this Hermitian inner product. In the same vein, we receive all the wonderful benefits of the dot product. For example, we can project a vector by taking

$$\hat{\boldsymbol{x}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and computing

$$\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{x}}_1 \rangle = a_1.$$

Likewise, the length of a complex vector is given by

$$\|\vec{\boldsymbol{a}}\| = \sqrt{\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{a}} \rangle}.$$

Nothing is lost from this more general approach, and this more general approach extends far beyond finite dimensional complex vectors!

1.1 Infinite Dimensions

The dimension of a vector is the number of entries needed to fully describe the vector. From the examples before, we can say that the vectors $\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \in \mathbb{R}^3$ are 3-dimensional real vectors and the vectors $\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \in \mathbb{C}^n$ are n-dimensional complex vectors. There is no restriction on the size of n, and n can in fact be infinite!

This section of the text is primarily concerned with extending our linear algebra techniques to the infinite dimensional case. Though this may sound ominous, it simply builds upon what we already know. In essence, we will combine our knowledge of functions, infinite series, integrals, and linear algebra to complete the theory for infinite dimensions. Put simply, functions will play the role of vectors while series and integrals will play the role of inner products. This viewpoint places us viewing mathematics from the top, where we can always reduce the general story to something more specific when need be. Ultimately, this allows one to understand one general structure instead of many individual ones.

2 Function Spaces

Rather than leave this section as an abstract approach to more advanced linear algebra, we shall take the viewpoint of working with some specific ideas in mind. The canonical infinite dimensional vector spaces are formed as spaces of functions. That is, for example, the set of all solutions to a homogeneous linear ODE of the form

$$x''(t) + f(t)x'(t) + g(t)x(t) = 0.$$

In the prequel, we proved that any linear combination of solutions to the above equation is also a solution which, along with the fact that $x(t) = 0$ is a solution, proves that this set of functions forms a vector space.

We will not concern ourselves with names for these spaces of functions or their formal definitions. But, let's list a few more other than the one I mentioned above. Let $\Omega = [0, L]$, then we can consider a few different sets of functions $f : \Omega \to \mathbb{R}$ that are vector spaces.

- The set of continuous functions.
- The set of differentiable functions (whose derivatives are also continuous).
- The set of analytic (functions with convergent Taylor series) functions.
- The set of functions whose square is integrable. That is, all functions f so that

$$\int_0^L \|f(x)\|^2 dx < \infty.$$

- The set of solutions to the Schrödinger equation for a free particle in a 1-dimensional box.
- The set of solutions to the Legendre equation for any nonnegative integer choice of m.

We will revisit the same spaces again and again as these are prototypes for other problems you will encounter. If you grasp the content for the above examples, seeing it with new examples will not be too challenging.

3 Inner Products and Norms

Before we define general inner products, let us recall the definition of a vector space. In the prequel, we had that a vector space V over some field $\mathbb{F}$ (the numbers we choose as entries) is a set containing vectors that satisfy eight different properties.

Exercise 3.1. Find the definition in the previous text and review it.

Definition 3.1: Inner Product

An ***inner product*** on a vector space V over a field $\mathbb{F}$ is a bilinear (sometimes sesquilinear) function

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

that satisfies

i. (Nondegenerate) For a $\vec{\boldsymbol{a}} \in V$ we have that $\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{a}} \rangle = 0$ if and only if $\vec{\boldsymbol{a}} = \vec{\boldsymbol{0}}$;

ii. (Positive definite) For any nonzero $\vec{\boldsymbol{a}} \in V$ we have that $\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{a}} \rangle > 0$;

iii. (Symmetric) For any $\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \in V$ we have that $\left\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \right\rangle = \left\langle \vec{\boldsymbol{b}}, \vec{\boldsymbol{a}} \right\rangle$. If the vector space is complex, then we have conjugate symmetry $\left\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{b}} \right\rangle = \left\langle \vec{\boldsymbol{b}}, \vec{\boldsymbol{a}} \right\rangle^*$.

What we are denoting is a function $\langle \cdot, \cdot \rangle$ that has two vectors ($V \times V$) as inputs where see $\cdot$ and outputs some number in the designated field $\mathbb{F}$. When we say bilinear, we mean that the function is linear in each input. For example, we have for vectors $\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}}, \vec{\boldsymbol{c}} \in V$ and a scalar $\alpha \in \mathbb{F}$ that

$$\left\langle \alpha \vec{\boldsymbol{a}} + \vec{\boldsymbol{b}}, \vec{\boldsymbol{c}} \right\rangle = \alpha \left\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{c}} \right\rangle + \left\langle \vec{\boldsymbol{b}}, \vec{\boldsymbol{c}} \right\rangle,$$

which shows the linearity in the first input. The second input is linear as well.

Similarly, if the field $\mathbb{F} = \mathbb{C}$, then the inner product need be sesquilinear in that we instead have the addition of a complex conjugate in the second position. That is, let $\alpha, \beta \in \mathbb{C}$ and we have

$$\left\langle \alpha\vec{\boldsymbol{a}} + \vec{\boldsymbol{b}}, \beta\vec{\boldsymbol{c}} \right\rangle = \alpha\beta^* \left\langle \vec{\boldsymbol{a}}, \vec{\boldsymbol{c}} \right\rangle + \beta^* \left\langle \vec{\boldsymbol{b}}, \vec{\boldsymbol{c}} \right\rangle.$$

The first position is simply linear.

Exercise 3.2. Verify that the dot product for $\mathbb{R}^n$ and the Hermitian inner product for $\mathbb{C}^n$ are indeed inner products.

Since we have previously covered two different inner products for the finite dimensional vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$, we can use our intuition from these spaces with their inner product structure to define other important inner products. We have in fact come across another example while studying the particle in the 1-dimensional box. Recall that the problem we solved was the equation

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi(x)}{dx^2} = E\Psi(x),$$

on the region $[0, L]$, where $\Psi(x)$ is the wavefunction. We also imposed the boundary conditions that $\Psi(0) = \Psi(L) = 0$ since the particle cannot be found on the boundary of this domain.

We found that the solutions to this equation were the normalized states

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right),$$

with corresponding energies $E_n = \frac{n^2h^2}{8mL^2}$. Then, a wavefunction could be written as a linear combination of these states by

$$\Psi(x) = \sum_{n=1}^{\infty} a_n\psi_n(x),$$

where $a_n \in \mathbb{C}$. In order for the wavefunction $\Psi(x)$ to be normalized, we required that

$$\sum_{n=1}^{\infty} \|a_n\|^2 = 1.$$

Now, we can consider a set V of all the possible wavefunctions for the above problem as well as the zero function (which is indeed a solution to the problem, but it is not physically meaningful).

Exercise 3.3. Show that V is a vector space.

We can add an inner product to the vector space V by defining the inner product on two wavefunctions Ψ and Φ by

$$\langle \Psi, \Phi \rangle := \int_0^L \Psi(x)\Phi^*(x)dx.$$

To see that this is an inner product, we need to show that the above function is sesquilinear and satisfies the three conditions for an inner product (nondegeneracy, positive definite, and symmetric). Sesquilinearity follows from the linearity of the integral in that we have

$$\begin{aligned}\langle \Psi, \Phi + \alpha\Theta \rangle &= \int_0^L \Psi(x)(\Phi(x) + \alpha\Theta(x))^* dx \\ &= \int_0^L \Psi(x)\Phi(x)\Phi^*(x)dx + \alpha^* \int_0^L \Psi(x)\Theta^*(x)dx \\ &= \langle \Psi, \Phi \rangle + \alpha^* \langle \Psi, \Theta \rangle .\end{aligned}$$

Showing the linearity in the first argument is analogous but there will not be a complex conjugate.

Next, we can see that the inner product is nondegenerate by noting that if we take the zero function 0, we have

$$\langle 0, 0 \rangle = \int_0^L 0 dx = 0,$$

and if we have that

$$0 = \langle \Psi, \Psi \rangle = \int_0^L \Psi(x)\Psi^*(x)dx = \int_0^L \|\Psi(x)\|^2 dx,$$

it must be that $\|\Psi(x)\| = 0$ since this integral cannot be zero otherwise. Hence, $\Psi(x)$ is the zero function and we have that the inner product is indeed nondegenerate.

By the above work, if $\Psi(x)$ is not the zero function, then $\|\Psi(x)\|^2 > 0$ and thus we have

$$\langle \Psi, \Psi \rangle > 0.$$

Hence, the inner product is positive definite.

Lastly, we can see that the inner product is symmetric by taking the Cartesian representation for $\Psi(x)$ by $\Psi(x) = a(x) + ib(x)$ and for $\Phi(x) = c(x) + id(x)$ and noting

$$\begin{aligned}\Psi(x)\Phi^*(x) &= (a(x) + ib(x))(c(x) - id(x)) \\ &= (a(x)c(x) + b(x)d(x)) + i(b(x)c(x) - a(x)d(x)),\end{aligned}$$

and

$$\begin{aligned}\Phi(x)\Psi^*(x) &= (c(x) + id(x))(a(x) - ib(x)) \\ &= (a(x)c(x) + b(x)d(x)) + i(a(x)d(x) - b(x)c(x),\end{aligned}$$

which means that we have

$$\langle \Psi, \Phi \rangle = \langle \Phi, \Psi \rangle^* .$$

Thus we have shown that this is indeed an inner product.

3.1 Norms

We have used the notation $\|\cdot\|$ throughout the prequel and in this text as well. At a first glance, it seems as if it is notationally similar to the absolute value of a real number $|x|$. The reason why we differentiate the notation slightly is that $\|\cdot\|$ means something a bit more general than the absolute value. We call $\|\cdot\|$ a ***norm***. For example, the modulus of a complex number z is a norm where we define

$$\|z\| = \sqrt{zz^*}.$$

We can also compute the norm of a complex (or real) vector $\vec{\boldsymbol{u}}$ in $\mathbb{C}^m$ (resp. $\mathbb{R}^m$) by computing

$$\|\vec{\boldsymbol{v}}\| = \sqrt{\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{u}} \rangle}.$$

In fact, the modulus of a complex number is the special case of $m = 1$. But what is the norm?

The norm of a vector is traditionally referred to as the length of the vector when this vector is something we can picture. For example, the norm of a vector $\vec{\boldsymbol{v}} \in \mathbb{R}^3$ is given by

$$\|\vec{\boldsymbol{v}} = \sqrt{\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}}}.$$

Notice how we are always using an inner product to define a norm! This is a key fact about inner products – they always induce a norm (or length).

Remark 3.1. From here on out we shall use the notation $\langle \cdot, \cdot \rangle$ to refer to the inner product and $\| \cdot \|$ to refer to the induced norm from that inner product. The choice of inner product should be noted or clear from the choice of vector space.

4 Inner Product Spaces

Given a vector space V with an inner product, we refer to the vector space as an ***inner product space***. In fact, all the vector spaces we have dealt with are inner product spaces! We tend to prefer working with these spaces as they allow us to nicely compare vectors (like we can with the dot product) and we can also compute lengths and distances. Needless to say, inner product spaces are immensely important in the physical world.

However, when the vector space is not finite dimensional (such as the space of solutions to the 1-dimensional box with the added zero function), we must be a bit more careful. Without going into far too much detail, we must add one other attribute to these spaces to make them work as we need. In this case, we must require that the inner product space is also ***complete***. A space is complete if and only if all Cauchy sequences in the space converge. We call a complete inner product space a ***Hilbert space***.

Exercise 4.1. We defined a Cauchy sequence in the prequel. Find the definition.

This extra requirement rules out some oddities and makes the infinite dimensional space much more like the finite dimensional spaces such as $\mathbb{R}^n$ and $\mathbb{C}^n$. We showed in the prequel that in $\mathbb{R}$ a convergent sequence is also Cauchy. That is, the definitions are analogous. The same happens to be true in $\mathbb{R}^n$ and $\mathbb{C}^n$ (you can picture taking a sequence of vectors instead of a sequence of real numbers). Thus, in a Hilbert space, Cauchy and convergent are again equal. Let us see why one should believe this.

Example 4.1: A Cauchy Sequence of Functions

Before, we studied power series that define functions. We would write

$$f(x) = \sum_{n=0}^{\infty} a_n x^n,$$

where x is in the domain of convergence for the series. As we worked through what it meant for a series to converge, we found that we could view a series as a sequence

of partial sums. That is, for each value of x we can create a sequence $\{A_n(x)\}_{n=0}^{\infty}$ by letting

$$A_N(x) = \sum_{n=0}^{N} a_n x^n.$$

We noted that as we increased N, the function $A_N(x)$ became closer and closer to the function $f(x)$. This was entirely reasonable as if the contrary were true, at some point a large N would provide us a worse approximation to $f(x)$.
The completeness assumption for a Hilbert space will give us this ability. It will allow one to properly approximate quantities such as infinite sums of functions in a way that makes intuitive sense.

No more detail is needed on the notion of completeness. We will completely avoid spaces that are not complete as they behave badly. Take the completeness of any space as given unless it is mentioned otherwise.

5 Symmetries

As previously discussed, symmetry is an important aspect of problem solving that is present in most physical systems. The prior example is no exception. We discussed the phase of a complex function and viewed this in an example from quantum mechanics. There, we found that when a wavefunction is altered by adding a global phase, the probability of making a measurement is not changed. This is in fact a specific example of something far more general. But in this case for the particle in the 1-dimensional box, we can see that if alter two wave functions by the same phase and take the inner product

$$\begin{aligned}\left\langle e^{i\theta}\Psi, e^{i\theta}\Phi \right\rangle &= \int_0^L e^{i\theta}\Psi(x)(e^{i\theta}\Phi(x))^* dx \\ &= \int_0^L e^{i\theta}\Psi(x)e^{-i\theta}\Phi^*(x)dx \\ &= \int_0^L \Psi(x)\Phi^*(x)dx \\ &= \langle \Psi, \Phi \rangle ,\end{aligned}$$

then the inner product is not changed. This is an example of a ***unitary operator***. A unitary operator preserves the inner product between two vectors. That is, if we have vectors Ψ and Φ from an Hilbert product space H, then U is a unitary operator if $U\colon H \to H$ is onto and

$$\langle U\Psi, U\Phi \rangle = \langle \Psi, \Phi \rangle .$$

This is of course not special for just the particle in the 1-dimensional box either. Take the space $\mathbb{R}^2$ with two vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$. Then consider a matrix $[A]$ that is in the group $\mathrm{O}(2)$ (which means that $[A]$ is a matrix that solely rotates and or reflects vectors).

Exercise 5.1. Recall the definition of the matrix group $\mathrm{O}(2)$.

Now, we can actually realize any matrix in $\mathrm{O}(2)$ as a reflection matrix, or a product

of two reflection matrices. For the sake of example, take a reflection matrix

$$[\text{Ref}]_\theta = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix},$$

which reflects a vector about the line passing through the origin with angle θ measured from the x-axis. Then letting

$$\vec{\boldsymbol{u}} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \qquad \text{and} \qquad \vec{\boldsymbol{v}} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

we have

$$[\text{Ref}]_\theta \vec{\boldsymbol{u}} = \begin{pmatrix} \cos(2\theta)u_1 + \sin(2\theta)u_2 \\ \sin(2\theta)u_1 - \cos(2\theta)u_2 \end{pmatrix} \qquad \text{and} \qquad [\text{Ref}]_\theta \vec{\boldsymbol{v}} = \begin{pmatrix} \cos(2\theta)v_1 + \sin(2\theta)v_2 \\ \sin(2\theta)v_1 - \cos(2\theta)v_2 \end{pmatrix}.$$

Put a picture here Then we can compute the inner product of the reflected vectors

$$\begin{aligned} \langle [\text{Ref}]_\theta \vec{\boldsymbol{u}}, [\text{Ref}]_\theta \vec{\boldsymbol{v}} \rangle &= (\cos(2\theta)u_1 + \sin(2\theta)u_2)(\cos(2\theta)v_1 + \sin(2\theta)v_2) \\ &\quad + (\sin(2\theta)u_1 - \cos(2\theta)u_2)(\sin(2\theta)v_1 - \cos(2\theta)v_2) \\ &\vdots \\ &= u_1 v_1 + u_2 v_2. \end{aligned}$$

Exercise 5.2. Show the remaining steps in the above computation.

Since this is true for any reflection matrix, it will be true for any product of reflection matrices and hence the group of O(2) matrices are unitary transformations on the inner product space $\mathbb{R}^2$ (where the inner product is the standard dot product).

Every Hilbert space will have a group of unitary symmetries that preserve the inner product. Hence, we tend to have different names for these groups to reflect what the underlying Hilbert product space is. In the particle in a box case, the underlying symmetry is the (slightly inaptly named) unitary group $U(1)$ whereas the case for $\mathbb{R}^2$ it is the orthogonal group O(2).

6 Bases

In the case of finite dimensions, we found that we could construct a minimal set of vectors in a space V such that any vector in V can be written as a linear combination of those vectors. We called such a set a basis. Take for example, the space $\mathbb{R}^3$, where we had the standard orthonormal basis given by $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$. We put

$$\hat{\boldsymbol{x}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \hat{\boldsymbol{y}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \hat{\boldsymbol{z}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

so that any vector

$$\vec{\boldsymbol{v}} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

could be written as

$$\vec{\boldsymbol{v}} = v_1\hat{\boldsymbol{x}} + v_2\hat{\boldsymbol{y}} + v_3\hat{\boldsymbol{z}}.$$

The latter notation proves to be convenient in the infinite dimensional case.

We have already discovered one way in which we can form a basis for certain infinite dimensional spaces. This came in the form of power series. The major difference is that we have infinite sums to build functions instead of just finite sums to build vectors that live in spaces like $\mathbb{R}^n$ and $\mathbb{C}^n$.

Let us continue on with the power series example. For the sake of simplicity, we can consider the space of analytic functions on the region $[0, L]$ (recall that analytic meant the function has a power series). Then, by definition, every function $f(x)$ in this space can be written as an infinite sum

$$f(x) = \sum_{n=0}^{\infty} a_n x^n,$$

so long as for every $x \in [0, L]$ we have that the above series converges. Thus we have really investigated infinite dimensional spaces and their bases a bit. In this case, our basis is the set of all powers of x. That is to say, we use $\{x^0, x^1, x^2, x^3, \dots\}$ as our basis vectors and the coefficients are the a_n.

In general, the definition of a basis that we have is already correct so long as we understand what it means in general to take a sum. In the previous example with power series, a sum could potentially be infinite. In some cases (e.g., the Fourier transform) we will see that we may need to take an integral to be our method of summation.

6.1 Orthonormal Bases

Once again, the finite dimensional case led us to discovering the usefulness of an inner product as a method for determining bases that are more useful. Above, we mentioned the standard orthonormal basis for $\mathbb{R}^3$ which gives us the most natural decomposition of a vector. Each of the vectors in this basis is of length one and is mutually orthogonal to one another. Intuitively, this lets you describe a point in space by how much you must walk back or forth, left or right, and up or down to reach your desired location. The orthonormal basis gave us a way to naturally decompose a vector into separate components through the dot product. That is, we could find the x, y, or z-component of a vector and none of these components depend on the others. If you were to take a basis

$$\vec{\boldsymbol{u}} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \vec{\boldsymbol{v}} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad \vec{\boldsymbol{w}} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

then if you were to break a vector up into these components, you would find that there is some overlap between what each component describes.

Exercise 6.1. Show the vectors $\vec{\boldsymbol{u}}$, $\vec{\boldsymbol{v}}$, and $\vec{\boldsymbol{w}}$ above form a basis. Then, to see that there is some overlap, you can find the components of a vector (of your choice) in the directions $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{w}}$. *Hint: to see the overlap, determine the length of the vector you chose, and compare it to the length computed from components un the $\hat{\boldsymbol{u}}$,$\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{w}}$.*

When a basis is not orthogonal, then to find independent components of a vector, you must somehow subtract off the overlap. This process is a bit tedious, so instead of giving an example, we can just assert the fact that orthogonal bases are indeed helpful.

Remark 6.1. Orthogonal bases in finite dimensions also appear when we find eigenvalues and eigenvectors for Hermitian matrices. Specifically, the set of eigenvectors corresponding to different eigenvalues for a Hermitian matrix are always orthogonal!

One may wish that the Hilbert spaces would have some sense of a natural decomposition into different orthogonal components. It turns out that in the realm of spaces we care about we can do this, and that some Hilbert spaces have particularly nice bases to work with. In the prequel, we studied series solutions to differential equations as a means of solving a large class of equations. During that time, we solved Legendre's equation and found the solutions were particularly nice.

Recall the equation we wished to solve was

$$(1-x^2)f''(x) - 2xf'(x) + m(m+1)f(x) = 0$$

with our domain $\Omega = [-1, 1]$ and m a non-negative integer ($m = 0, 1, 2, \dots$). Ignoring the details, we found that there were solutions for each m (which in some sense are like the eigenvalues for the 1-dimensional box). The normalized solutions were known as the Legendre polynomials and the first few were

$$f_0(x) = \sqrt{\frac{1}{2}} \qquad f_1(x) = \sqrt{\frac{3}{2}}x$$
$$f_2(x) = \sqrt{\frac{5}{8}}(1-3x^2) \qquad f_3(x) = \sqrt{\frac{63}{8}}\left(x - \frac{5x^3}{3}\right).$$

One may be wondering what we mean by normalized here. The solutions to the Legendre equation live in a Hilbert space whose inner product is given by

$$\langle \Psi, \Phi \rangle = \int_{-1}^{1} \Psi(x)\Phi(x)dx,$$

in which these functions are normalized. That is, we have

$$\|f_i\| = \sqrt{\langle f_i, f_i \rangle} = 1,$$

for every i. Note we do not need a complex conjugate in this inner product since these functions are all real valued. Moreover, it turns out that this set of polynomials is orthonormal since we also have

$$\langle f_i, f_j \rangle = 0$$

if $i \neq j$. Finally, the Legendre polynomials form a basis for the solution to the Legendre equation. Specifically, we can write any other solution as a series of Legendre polynomials by

$$f(x) = \sum_{n=1}^{\infty} a_n f_n(x),$$

so long as the series converges for all values of x in the domain Ω.

In the realm of quantum mechanics, the orthonormal basis of eigenfunctions are extremely important. Previously, we even donated these functions with the term ***states***. Though we never defined this term in the prequel, we now have a rough working definition. When observing a quantum system, it is only possible to observe a particle in an

eigenstate of the operator that is mathematical representation for the observation you are wishing to make. We will define operators, eigenfunctions, and observables later.

Another example of a basis arises from the solutions to the free particle in the 1-dimensional box. If we let the Hilbert space be any solution to the boundary value problem along with the zero solution which we ruled out as not being physically meaningful (it is a solution, just not physically), then we found the orthonormal basis was given by all the states

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right).$$

Hence, any solution to that problem can be written as

$$\Psi(x) = \sum_{n=1}^{\infty} a_n \psi_n(x),$$

so long as this sum converges for x values in the domain $[0, L]$.

6.2 Series and Integrals as Linear Combinations

The two main points of a basis is really to give us a way to extract interesting components of a vector and to be able to uniquely write any vector as a linear combination of the fewest possible elements of a vector space. As discussed above, we can make our lives easiest by choosing nicer bases such as bases that are orthonormal. Let us assume that we will do this.

For certain Hilbert spaces $\mathcal{H}$, we can write any vector in the space as a sum or series. First, take for example, $\mathbb{R}^n$. Let us choose the orthonormal basis

$$\hat{\boldsymbol{x}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \hat{\boldsymbol{x}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \hat{\boldsymbol{x}}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then we can write any vector in the space $\mathbb{R}^n$ as a finite sum of the basis elements above by

$$\vec{\boldsymbol{v}} = \sum_{i=1}^{n} v_i \hat{\boldsymbol{x}}_i = v_1 \hat{\boldsymbol{x}}_1 + v_2 \hat{\boldsymbol{x}}_2 + \cdots + v_n \hat{\boldsymbol{x}}_n = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix}.$$

As always, the $\hat{\boldsymbol{x}}_i$ form the basis for $\mathbb{R}^n$ and the v_i are the coefficients for the vector $\vec{\boldsymbol{v}}$.

When $\mathcal{H}$ is infinite dimensional, we can no longer easily manipulate vectors as column vectors (it just doesn't always make sense)! So, we our hand is forced and we must write our vectors as some form of an infinite sum. Of course, this is just an infinite version of a linear combination! For now, let us assume H is sufficiently nice[1] so that we can write any vector $\Psi \in \mathcal{H}$ by

$$\Psi = \sum_{i=1}^{\infty} a_n \psi_n$$

[1] Sufficiently nice here means ***separable***. Do not worry about this definition!

where the collection of all ψ_n is a basis for H and the a_n are the coefficients for Ψ. Not much has changed here, but the sum has now become infinite so we must handle convergence with care! One may also find that the infinite sum may not converge to the intended function at *every* point in the domain (but this is a non-issue). We will see this when we cover the Fourier series.

Finally, for the Hilbert spaces that are not quite as nice, we may in fact need to use integration as our form of linear combination. We will see this in the case of the Fourier transform where we must put

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\omega)\chi_x(\omega)d\omega.$$

Here, the basis elements are the $\chi_x(\omega)$ which for the Fourier transform have the form

$$\chi_x(\omega) = e^{i\omega x}.$$

The coefficients are then given by the function $\hat{f}(\omega)$ where, for example, $\omega \in \mathbb{R}$. This is analogous the the above cases, its just that the list of coefficients is not finite, or indexed by integers, but rather it is indexed by the whole (larger) set of real numbers $\mathbb{R}$.

Remark 6.2. There is much going on here. Much if it is coming down to the sizes of infinities and underlying spaces we are working with. The point is to see that we are solely generalizing the notion of writing a vector in a basis. Some spaces just have a very large amount of basis elements!

6.3 Projection

When we are handed a vector $\vec{\boldsymbol{v}} \in \mathbb{R}^3$ we can recover the components of $\vec{\boldsymbol{v}}$ by projecting onto our orthonormal basis. As I've said before, we do this by

$$\vec{\boldsymbol{v}} \cdot \hat{\boldsymbol{x}} = v_1.$$

Nothing changes when we work with an infinite orthonormal basis. Say we have the basis ψ_n for some Hilbert space, then for a vector Ψ, we can recover the component a_n by

$$\langle \Psi, \psi_n \rangle = a_n.$$

This is a surprisingly helpful fact.

Sometimes, we want to see what the fundamental components of a function are in order to approximate a function. We did this with power series, but this is not always the best choice. We also have developed machines that work via extracting, for example, frequency components of functions (see Fourier transform spectroscopy). If one is able to find that a complicated function is well approximated by the sum of a few far less complicated functions, this will simplify the analysis we do when we work with the function. Let us work with an example.

Example 6.1: Triangle Wave

Let's consider the function $f\colon [0, L] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x & 0 \leq x \leq L/2 \\ L - x & L/2 < x \leq L \end{cases}$$

We wish to find a suitable orthonormal basis of functions with domain $[0, L]$. In this case, let us define the inner product by

$$\langle g, h \rangle = \frac{1}{L} \int_0^L g(x)h(x)dx,$$

which is a scaled version of the Hermitian inner product (with omitted conjugate since the functions are real). We will revisit this inner product when we arrive at Fourier series.

Now, since this is a scaled version of the Hermitian inner product and since $f(x)$ satisfies the zero boundary conditions requirement, we can conjecture that a suitable orthogonal basis is given by

$$\sin\left(\frac{n\pi x}{L}\right)$$

where $n = 1, 2, 3, \ldots$. We know that these elements are orthogonal in that

$$\left\langle \sin\left(\frac{n\pi x}{L}\right), \sin\left(\frac{m\pi x}{L}\right) \right\rangle = 0$$

when $n \neq m$. However, we must normalize these vectors as well since our inner product is not the same. We take

$$\begin{aligned} 1 &= \left\langle c\sin\left(\frac{n\pi x}{L}\right), c\sin\left(\frac{n\pi x}{L}\right) \right\rangle \\ &= \frac{1}{L}\int_0^L c^2 \sin^2\left(\frac{n\pi x}{L}\right) dx \\ &= \frac{c^2}{2}. \end{aligned}$$

We can then solve for the normalization constant c to find that we get $c = \sqrt{2}$. Hence we have that our orthonormal basis for this inner product is

$$\psi_n(x) = \sqrt{2}\sin\left(\frac{n\pi x}{L}\right),$$

for $n = 1, 2, 3, \ldots$. Now, we have our $f(x)$, and we want to extract the components of $f(x)$. To do this, we use the inner product and project by

$$\begin{aligned} a_n &= \langle f, \psi_n \rangle \\ &= \frac{1}{L}\int_0^L f(x)\sqrt{2}\sin\left(\frac{n\pi x}{L}\right) dx \\ &= \sqrt{2}\left(\frac{L\sin\left(\frac{\pi n}{2}\right) - \frac{1}{2}\pi L n\cos\left(\frac{\pi n}{2}\right)}{\pi^2 n^2} + \frac{L\left(2\sin\left(\frac{\pi n}{2}\right) + \pi n\cos\left(\frac{\pi n}{2}\right)\right)}{2\pi^2 n^2}\right). \end{aligned}$$

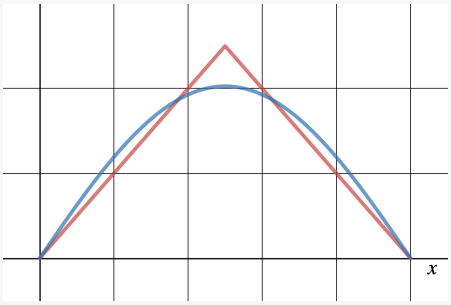
You should verify the integral above.
What this means is that we can now put

$$f(x) = \sum_{n=1}^{\infty} \sqrt{2} \left(\frac{L \sin\left(\frac{\pi n}{2}\right) - \frac{1}{2}\pi L n \cos\left(\frac{\pi n}{2}\right)}{\pi^2 n^2} + \frac{L\left(2\sin\left(\frac{\pi n}{2}\right) + \pi n \cos\left(\frac{\pi n}{2}\right)\right)}{2\pi^2 n^2} \right) \sqrt{2} \sin\left(\frac{n\pi x}{L}\right).$$

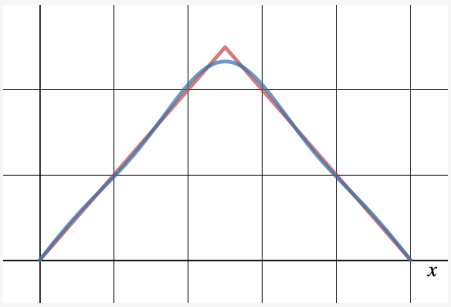
The question now is whether this is believable at all. Because, let's face it, this seems pretty odd. So, let's examine what these functions look like.
There is no nice way of evaluating the above infinite sum, but we can take finite sum approximations from this. For example, we can choose some positive integer N and take

$$f(x) \approx \sum_{n=1}^{N} a_n \psi_n(x).$$

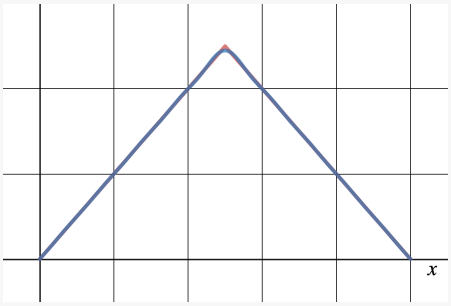
Let us plot various degrees of approximations with the original function $f(x)$.



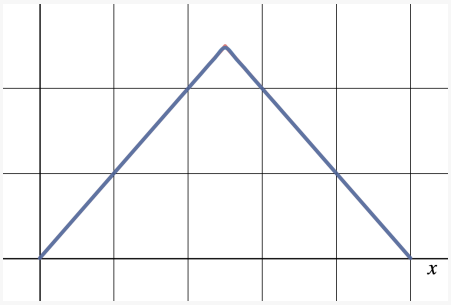
(a) Approximation of $f(x)$ with $n = 1$ basis elements..



(b) Approximation of $f(x)$ with $n = 5$ basis elements.



(c) Approximation of $f(x)$ with $n = 20$ basis elements.



(d) Phase difference $\delta = \pi$.

Figure 2.1: Approximation of $f(x)$ with $n = 50$ basis elements.

The above example is of dire importance in the realm of experimentation. If we were not able to approximate functions using a methodology like this, all computation would be far more difficult and possible infeasible. One can see from the above example that an approximation with 20 or more terms is fantastic. What we have

done here is also important for modelling differential equations. For example, in some equations, one can take the first 20 basis functions and see how the coefficients a_n change over time as opposed to explicitly solving the differential equation. This is the essence of finite element approaches.

3

Linear Operators

1 Matrices

In the spirit of making analogies with the finite dimensional case, we will briefly revisit the idea of a linear transformations and matrices. We previously studied linear transformations from one vector space to another. Most often we considered transformations of the form

$$T\colon \mathbb{R}^3 \to \mathbb{R}^3$$

which transformed space. We wrote

$$T(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{w}}.$$

Recall that a linear transformation is one that satisfies

$$T(\vec{\boldsymbol{u}} + \alpha\vec{\boldsymbol{v}}) = T(\vec{\boldsymbol{u}}) + \alpha T(\vec{\boldsymbol{v}})$$

for any two vectors $\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \in \mathbb{R}^3$ and scalar $\alpha \in \mathbb{R}$. Since this transformation is linear, we found that we could represent this transformation as a matrix $[A]$ of nine numbers and multiply the vector by this matrix. So we wrote

$$[A]\vec{\boldsymbol{v}} = \vec{\boldsymbol{w}}.$$

Part of the reason we introduced the notion of linear transformations and not just matrices is that we will now need to ditch the notion of a matrix (unless you wish to write an infinite matrix, which, in some cases, you can do). Linear transformations are far more general and we can study their structure in a similar way that we did with matrices.

Linear Equations

Matrices give rise to linear equations. For example, we may be given a vector $\vec{\boldsymbol{b}} \in \mathbb{R}^3$ and a 3×3-matrix $[A]$ and will be asked to solve the equation

$$[A]\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}, \tag{Eq. 3.1.1}$$

for the vector $\vec{\boldsymbol{x}}$. This problem was solved by row reduction. However, we also found that we could (in general) invert the matrix $[A]$ to produce $[A]^{-1}$. This method was far more powerful since we could quickly solve the Eq. 3.1.1 with any given vector. Specifically, we have that

$$\vec{\boldsymbol{x}} = [A]^{-1}\vec{\boldsymbol{b}}.$$

In fact, the eigenvalue problem is extremely related to this idea as well. Given the problem in Eq. 3.1.1, we can isolate the matrix $[A]$ and decompose the action of the matrix into scaling on individual vectors. That is, we solve the eigen-equation

$$[A]\vec{\boldsymbol{e}} = \lambda\vec{\boldsymbol{e}}.$$

Scaling is an easy process to invert since we have

$$[A]^{-1}\vec{\boldsymbol{e}} = \frac{1}{\lambda}\vec{\boldsymbol{e}}.$$

Exercise 1.1. Can you prove the above statement?

The moral is, if we can find a full set of eigenvalues and eigenvectors, then we can diagonalize our matrix by

$$[\Lambda] = [P]^{-1}[A][P].$$

The diagonal matrix $[\Lambda]$ is very easy to invert and so we can find

$$[A]^{-1} = [P]^{-1}[\Lambda]^{-1}[P].$$

This may seem like a bit more work than just inverting the matrix $[A]$, but this method allows us to see how the method is really working. To some extent, you are performing this process when you compute $[A]^{-1}$ without directly mentioning so.

Exercise 1.2. Can you show the work between the above steps?

We will return to this idea of inversion as we investigate more general operators.

Adjoints

When we were given a matrix $[A]$ (whether real or complex) we could compute its adjoint by

$$[A]^\dagger = \left([A]^T\right)^*.$$

That is, we can take the complex conjugate transpose of the matrix $[A]$. If $[A]$ is purely real, then this amounts to just taking the transpose. The notion of the adjoint was an important one in studying the eigenvalue problem. For example, we stated that if a matrix is Hermitian (or self adjoint) then

$$[A]^\dagger = [A].$$

In this case, we know that all the eigenvalues of $[A]$ are real, and that the eigenvectors corresponding to different eigenvectors are orthogonal. Recall, the eigenvalue problem

$$[A]\vec{e} = \lambda\vec{e}.$$

What the above says is that if $[A]$ is Hermitian, then λ must be real. In other words, one may say that the ***spectrum*** of a Hermitian matrix is always real. The notion of a spectrum is of core importance in the study of quantum mechanics. Physicist Paul Dirac had developed a theory of quantum mechanics after Werner Heisenberg and Erwin Schrödinger where one computes the spectrum of the hamiltonian operator to find solutions.

2 Linear Operators

Just as we generalized the notion of vectors from the finite dimensional spaces into the infinite dimensional case, we will repeat with linear transformations. The definitions of a vector space and linear transformation were properly general enough, but we will provide a new name for the linear transformations. Let H be a Hilbert space. Then we have that a ***linear operator*** is a linear transformation $\mathcal{L}\colon \mathcal{H} \to \mathcal{H}$. Some may relax this definition a bit to allow for the input and output spaces to be different but we should not be concerned by this in any way.

What is an example of a linear operator? Let us first choose a Hilbert space. We can, for example, choose the Hilbert space $\mathcal{H}$ of analytic functions on a region $[0, L]$. Then consider the linear operator $x\colon \mathcal{H} \to \mathcal{H}$ which multiplies a given function by the variable x. Is this indeed linear? Let us check by taking two function $f, g \in \mathcal{H}$ and a constant $\alpha \in \mathbb{R}$ and note

$$x(f(x) + \alpha g(x)) = xf(x) + \alpha x g(x),$$

which is indeed linear. One should also check that, for example, $xf(x)$ is still an analytic function on $[0, L]$. Note that

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

which converges on $[0, L]$. Then

$$xf(x) = \sum_{n=0}^{\infty} a_n x^{n+1}$$

also converges on $[0, L]$.

Exercise 2.1. Can you argue why this must be true? *Hint: use the ratio test.*

Another example of a linear operator is taking the differential of a function. That is, the derivative $\mathcal{L} = \frac{d}{dx}$ is an operator $\frac{d}{dx}\colon \mathcal{H} \to \mathcal{H}$. Keeping $\mathcal{H}$ as the analytic functions on $[0, L]$ then we can see that the derivative is linear by

$$\frac{d}{dx}(f(x) + \alpha g(x)) = \frac{df}{dx} + \alpha\frac{dg}{dx}.$$

Previously one refers to these rules of the derivative as the sum rule and constant multiple rule. However, we should now just refer to this quality as the ***linearity*** of the derivative operator.

These operators show up in the study of (ordinary) differential equations as we have seen before. Take for example the Hamiltonian operator

$$\hat{H} = \frac{-\hbar^2}{2m}\frac{d^2}{dx^2} + V(x).$$

This is a linear operator as well. In fact, this operator $\hat{H}$ is even Hermitian which means its spectrum is real valued! This is a key component of the quantum theory as we only have the ability to measure real numbers. We in fact require all observable operators to be Hermitian. More on this in a bit.

2.1 Adjoints

Though we had defined adjoints of matrices through an operation, we need to instead provide a more general definition. Let $\mathcal{H}$ be a Hilbert space and $\mathcal{L}$ a linear operator. Then we define the ***adjoint*** $\mathcal{L}^\dagger$ to be the unique operator satisfying

$$\langle \mathcal{L}\Psi, \Phi\rangle = \left\langle \Psi, \mathcal{L}^\dagger\Phi\right\rangle.$$

Let us see why this makes sense for the matrix case with an explicit example.

Example 2.1: Matrix Adjoint

Let

$$[A] = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad \vec{\boldsymbol{u}} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \qquad \vec{\boldsymbol{v}} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Then let us take the inner (dot) product

$$\langle [A]\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}\rangle.$$

First, we compute

$$[A]\vec{\boldsymbol{u}} = \begin{pmatrix} u_1 \\ u_1 + u_2 \end{pmatrix},$$

and then

$$\langle [A]\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}\rangle = \begin{pmatrix} u_1 \\ u_1 + u_2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = u_1v_1 + (u_1 + u_2)v_2.$$

Now let us solve for $[A]^\dagger$. We require

$$\left\langle \vec{\boldsymbol{u}}, [A]^\dagger\vec{\boldsymbol{v}}\right\rangle = u_1v_1 + (u_1 + u_2)v_2.$$

Let us put

$$[A]^\dagger = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$[A]^\dagger\vec{\boldsymbol{v}} = \begin{pmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{pmatrix}$$

Then we also require that

$$u_1v_1 + (u_1 + u_2)v_2 = \left\langle \vec{\boldsymbol{u}}, [A]^\dagger \vec{\boldsymbol{v}} \right\rangle = u_1(av_1 + bv_2) + u_2(cv_1 + dv_2).$$

Thus we can solve for a, b, c, and d to find that $a = b = d = 1$ and $c = 0$. That is,

$$[A]^\dagger = [A]^T,$$

is just the transpose of $[A]$.

Exercise 2.2. Can you prove that $[A]^\dagger = ([A]^*)^T$ when $[A]$ is an arbitrary complex 2×2 matrix? *Hint: use the same steps as above but start with an arbitrary matrix and use the Hermitian inner product.*

In the case for functions and function spaces, the adjoint may not be as easy to compute but it is still well defined. For many of the cases we care about, we will not need to compute the adjoint since it will be equal to the original operator (i.e., it is Hermitian).

We previously covered the idea of phase for complex valued functions and specifically looked at how phase can effect the inner product for solutions to the particle in the 1-dimensional box. We can see this in a different light by taking

$$\Psi(x) = \frac{1}{\sqrt{2}}\psi_1(x) + \frac{1}{\sqrt{2}}\psi_2(x)$$

and letting $\mathcal{U}$ be an operator defined by

$$\mathcal{U} = e^{i\theta},$$

which changes the phase of a wavefunction.

Example 2.2: Phase Operator

Taking the notions from above, we can take two wave function $\Psi(x)$ and $\Phi(x)$ and compute

$$\langle \mathcal{U}\Psi, \Phi \rangle = \int_0^L e^{i\theta}\Psi(x)\Phi^*(x)dx.$$

In this case, we can find the adjoint $\mathcal{U}^\dagger$ to U by requiring

$$\int_0^L e^{i\theta}\Psi(x)\Phi^*(x)dx = \left\langle \Psi, \mathcal{U}^\dagger\Phi \right\rangle = \int_0^L \Psi(x)(\mathcal{U}^\dagger\Phi(x))^*dx.$$

This leads us to the equation

$$\int_0^L e^{i\theta}\Psi(x)\Phi^*(x)dx = \int_0^L \Psi(x)(\mathcal{U}^\dagger)^*\Phi^*(x).$$

Thus, it must be that

$$(\mathcal{U}^\dagger)^* = e^{i\theta}.$$

Then, taking the complex conjugate of both sides we have

$$\mathcal{U}^\dagger = e^{-i\theta},$$

which means that † is acting as the complex conjugate itself in this example. Note that † is not always just the complex conjugate!

In the example, we took $\mathcal{U} = e^{i\theta}$ and found $\mathcal{U}^\dagger = e^{-i\theta}$ and one can note that $\mathcal{U}^\dagger\mathcal{U} = \mathcal{U}\mathcal{U}^\dagger = 1$. This is exactly the requirement we put on, for example, matrices in the group of spatial rotation matrices SO(3). In that case, we said that a matrix $[A] \in \text{SO}(3)$ satisfies $[A]^T[A] = [A][A]^T = I$.

What we have seen above is an example of a ***unitary operator***. A unitary operator is an operator $\mathcal{U}\colon \mathcal{H} \to \mathcal{H}$ that is onto (every possible output value is achieved) and satisfies $\langle \mathcal{U}\Psi, \mathcal{U}\Phi\rangle = \langle \Psi, \Phi\rangle$. Unitary operators are the symmetry operators for a given Hilbert space as they do not affect the inner product measurement we perform on that space. For example, when $\mathcal{U} = e^{i\theta}$, we can see that

$$\langle \mathcal{U}\Psi, \mathcal{U}\Phi\rangle = \int_0^L e^{i\theta}\Psi(x)\left(e^{i\theta}\Phi(x)\right)^* dx = \int_0^L \Psi(x)\Phi^*(x)dx = \langle \Psi, \Phi\rangle .$$

In other words, the multiplication by the same phase on both functions does not change the inner product between them. If we let $\Phi(x) = \Psi(x)$, this means that the probability of observing a particle at some point $x \in [0, L]$ does not change if we rotate our measurement device.

In the case for SO(3) where we rotate vectors we do not see the inner product between vectors change either. Matrices in SO(3) are also unitary matrices for the dot product on $\mathbb{R}^3$ and they can be realized as rotations of the whole space. Clearly, rotating the whole space won't change the angle between two vectors!

2.2 Hermitian Operators

In finite dimensions, we came across the notion of matrices that were ***self-adjoint*** or ***Hermitian*** (both mean the same thing). That is, a matrix $[A]$ is Hermitian if

$$[A]^\dagger = [A].$$

In other words, we have

$$\langle [A]\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}\rangle = \left\langle \vec{\boldsymbol{u}}, [A]^\dagger\vec{\boldsymbol{v}}\right\rangle = \langle \vec{\boldsymbol{u}}, [A]\vec{\boldsymbol{v}}\rangle .$$

These matrices had real eigenvalues and moreover each eigenvector corresponding to a different eigenvalue are orthogonal to one another. We never proved this fact, but with the updated notion of the adjoint, we can prove this rather easily.

Theorem 2.1: Hermitian Matrix Eigenvalues and Eigenvectors

Let $[A]$ be a Hermitian matrix with complex entries and $\langle \cdot, \cdot\rangle$ be the Hermitian inner product. Then $[A]$ has all real eigenvalues and if λ_j and λ_k are distinct eigenvalues, then the corresponding eigenvectors $\vec{\boldsymbol{e}}_j$ and $\vec{\boldsymbol{e}}_k$ are orthogonal.

Proof. To prove the first part, let λ and $\vec{\boldsymbol{e}}$ be an eigenvalue and eigenvector pair.

Then we have

$$\langle [A]\vec{\boldsymbol{e}}, \vec{\boldsymbol{e}} \rangle = \langle \lambda \vec{\boldsymbol{e}}, \vec{\boldsymbol{e}} \rangle = \lambda \langle \vec{\boldsymbol{e}}, \vec{\boldsymbol{e}} \rangle ,$$

but also

$$\langle [A]\vec{\boldsymbol{e}}, \vec{\boldsymbol{e}} \rangle = \langle \vec{\boldsymbol{e}}, [A]\vec{\boldsymbol{e}} \rangle = \lambda^* \langle \vec{\boldsymbol{e}}, \vec{\boldsymbol{e}} \rangle .$$

Hence it must be that $\lambda = \lambda^*$ and thus λ must be real.
To prove the second statement, we take two eigenvectors $\vec{\boldsymbol{e}}_j$ and $\vec{\boldsymbol{e}}_k$ corresponding to distinct eigenvalues λ_j and λ_k and note

$$\langle [A]\vec{\boldsymbol{e}}_j, \vec{\boldsymbol{e}}_k \rangle = \lambda_j \langle \vec{\boldsymbol{e}}_j, \vec{\boldsymbol{e}}_k \rangle ,$$

as well as

$$\langle [A]\vec{\boldsymbol{e}}_j, \vec{\boldsymbol{e}}_k \rangle = \langle \vec{\boldsymbol{e}}_j, [A]\vec{\boldsymbol{e}}_k \rangle = \lambda_k^* \langle \vec{\boldsymbol{e}}_j, \vec{\boldsymbol{e}}_k \rangle .$$

Then, note that we know λ_k must be real by the first part and hence we must have

$$\lambda_j = \lambda_k$$

which contradicts our original supposition or that

$$\langle \vec{\boldsymbol{e}}_j, \vec{\boldsymbol{e}}_k \rangle = 0.$$

Since we cannot have a contradiction, we then have that the eigenvectors are orthogonal. □

We don't want to solely review the finite dimensional examples, but move onto more general ones. The reason why we review this is we have the same result for more general operators. Our intuition thus carries over from our knowledge about matrices. When dealing with a linear operator, we define Hermitian (or self-adjointness) in the exact same way we do for matrices. Namely, if we have an inner product for functions $\langle \cdot, \cdot \rangle$ and a linear operator L, then

$$\langle L\Psi, \Phi \rangle = \langle \Psi, L\Phi \rangle ,$$

means that L is Hermitian.

Theorem 2.2: Hermitian Operators Eigenvalues and Eigenfunctions

Let L be a Hermitian linear operator and $\langle \cdot, \cdot \rangle$ be the inner product for the function space. Then L has real eigenvalues and if λ_j and λ_k are distinct eigenvalues, then the corresponding eigenfunctions $\psi_j(x)$ and $\psi_k(x)$ are orthogonal.

Proof. The proof is analogous to the previous. □

Let us see this by reviewing one of our go-to examples.

Example 2.3: Laplace Operator is Self Adjoint

Consider the free particle in the 1-dimensional box $[0, L]$. There, we solved the for the eigenfunctions of the Laplace operator $-\frac{d^2}{dx^2}$ by finding all $\Psi(x)$ such that

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi(x) = E\Psi(x).$$

Dividing both sides by the constants on both sides yields

$$-\frac{d^2}{dx^2}\Psi(x) = \omega^2\Psi(x),$$

where $\omega^2 = \frac{2mE}{\hbar^2}$. Hence, we are indeed finding eigenfunctions of the Laplace operator. Recall also that we require $\Psi(0) = \Psi(L) = 0$, and thus we can see this operator is Hermitian by performing integration by parts twice. That is, we take

$$\begin{aligned}\left\langle -\frac{d^2}{dx^2}\Psi, \Phi \right\rangle &= \int_0^L \left(-\frac{d^2\Psi}{dx^2}\right)\Phi^*(x)dx \\ &= \int_0^L \left(\frac{d\Psi}{dx}\right)\left(\frac{d\Phi^*}{dx}\right)dx \\ &= \int_0^L \Psi(x)\left(-\frac{d^2\Phi^*}{dx^2}\right)dx \\ &= \left\langle \Psi, -\frac{d^2}{dx^2}\Phi \right\rangle.\end{aligned}$$

Thus, we know that the eigenvalues E must be real. Indeed, we found that the eigenvalues were

$$E_n = \frac{n^2\pi^2\hbar^2}{2mL^2},$$

which correspond to the normalized eigenfunctions

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right).$$

Note that for each ψ_n, we have a unique eigenvalue and thus we expect that the eigenfunctions ψ_n are orthogonal. Again, we found this to be true since we can show

$$\int_0^L \psi_n(x)\psi_m(x)dx = 0$$

when $n \neq m$.

Exercise 2.3. Fill in the missing steps in the integration by parts above.

Now, we can say something more meaningful about linear operators. For one, we have seen that an equation with a linear operator can be a differential equation. Also, couple that with our results about the eigenvalue problem and inversion for matrices, and we should be able to use this theory to solve said differential equations.

For example, one may wonder if this allows us to consider a more general equation. Take, for example, the ***Poisson equation***

$$-\frac{d^2}{dx^2}f(x) = g(x).$$

In this problem, we are given a domain, say $[0, L]$, and boundary values, say $f(0) = f(L) = 0$, and the function $g(x)$ and we are asked to solve for the function $f(x)$. In this case, the domain $[0, L]$ represents an elastic rod attached at endpoints and $g(x)$ is an external force pushing down on the rod. Our solution $f(x)$ describes the configuration of the rod once the external force is allowed to act. Let $L = -\frac{d^2}{dx^2}$ and we can put

$$Lf = g,$$

which exactly parallels the equation

$$[A]\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}.$$

So, if we have some way of inverting L, we could put

$$f = L^{-1}g.$$

Our means of doing this will come with first solving the eigenvalue problem as we did in the previous example. This gives us a new way to write our functions $f(x)$ and $g(x)$ and solve our differential equation using algebraic methods.

3 Differential and Integral Operators

Sometimes operators (like matrices) are not "square." When a matrix was not square, we end up with a transformation like $[A]\colon \mathbb{R}^m \to \mathbb{R}^n$. So, the output space looks different than the input space. We avoided studying these types of operators for the most part, but we did consider the inner product which is indeed an example of this in a more general case.

Derivatives or more generally, ***differential operators***, are examples of operators that, in general, do not act like square matrices. At the very least, they act like matrices that have a non-trivial nullspace. For example, if we take a constant function c, then

$$\frac{d}{dx}c = 0.$$

Thus, all constant functions are in the nullspace of the derivative operator.

Now, our interest in studying differential operators is to provide us a new way of interpreting a differential equation. Namely, we will think of a differential equation as we have previously mentioned. If L is a linear operator, then we can write a linear differential equation as

$$Lf = g,$$

for some given function g and either initial or boundary values for our input domain. Let us see a few examples.

Before, we let $L = -\frac{d^2}{dx^2}$ which gives us Poisson's equation

$$Lf = g \qquad \Longleftrightarrow \qquad -\frac{d^2}{dx^2}f(x) = g(x).$$

This is a second order linear inhomogeneous ODE. We could take another operator $D = -\frac{d}{dx}$ and put

$$Df = g \qquad \Longleftrightarrow \qquad \frac{d}{dx}f(x) = g(x),$$

which is a first order linear equation. Specifically, this equation is even separable in the way we have written it.

Moreover, any second order linear equation could be written by taking

$$L = p(x)\frac{d^2}{dx^2} + q(x)\frac{d}{dx} + r(x)$$

so that

$$Lf = p(x)f''(x) + q(x)f'(x) + r(x)f(x).$$

Aside from a few nonlinear ODEs that we solved, we primarily dealt with the linear equations. It's just that the linear theory is quite a bit easier than the nonlinear version!

Later on in this course we visit new forms of differential operators as we increase the spatial dimension. For example, in 3-dimensional space, we have a notion of rotation of a vector field. The example in 2.1 shows us a field that rotates. To compute how much rotation the field has, we need the differential operator

$$\nabla\times,$$

known as the curl operator. If a vector field has a source, we can see this using the divergence operator

$$\nabla\cdot.$$

Some of the notation should seem familiar as we will simply combine differential operators ∇ with vector operators $\times$ and $\cdot$.

These new operators will give us new and important problems to consider. For example, one can find the streamlines of a vector field by solving a higher dimensional ODE. One could also study material properties by seeing how heat passes through a material over time, or how a material is deformed under a force. Most every physical problem out there falls into this general category of partial differential equations.

3.1 Integral Operators

Now, let us consider one other type of non-square operator $J\colon \mathcal{H} \to \mathbb{R}$. This operator will be that of integration. Specifically, we have already seen that derivatives act as operators, and one may be interested in if an integral operator can act as an inverse. It turns out that integral operators can in fact help us with that, but they also show up with their own interpretation.

For one example, we can take the norm of a function as an integral operator. Say that our function f is defined on $[0, L]$, then this is because we have

$$\|f\| = \sqrt{\langle f, f\rangle} = \int_0^L f(x)f^*(x)dx.$$

Therefore, finding the length of a function is simply an operator.

We also saw for the particle in the 1-dimensional box that we could compute the probability of a particle being in a region $[a, b]$. We found this by computing

$$P_{[a,b]}(f) = \int_a^b f(x)f^*(x)dx.$$

Note that here $P_{[a,b]}$ is an integral operator that inputs a function, and outputs a probability between 0 and 1.

We will not spend extra time studying integral operators, but it is worth noting that operators encompass nearly all that we have studied throughout calculus. This is why we express the importance of studying linear algebra! Though it, we can study much more general problems using similar ideas.

4 Spectra

All of this has culminated to investigation of the spectrum of an operator. We have found that the differential operators allow us to rephrase differential equations in a new light. We now seek to break down these differential operators so that we may hope to find new (and possibly easier) ways to approach differential equations.

Earlier in this chapter, we compared the equations

$$[A]\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}} \qquad \text{and} \qquad Lf = g.$$

For the matrix problem, we discussed the inversion process in finding the eigenvectors for $[A]$ and mentioned we could extend this notion to a linear operator L. In general, we refer to the set of all eigenvalues of a matrix as the ***spectrum***. For a linear operator L, we shall take the same definition though it is not completely general.

The idea is as follows: Find the spectrum and eigenfunctions for a linear operator L, and use these to invert the operator in order to solve the equation $Lf = g$. We have worked through two problems where we have done this.

Exercise 4.1. Revisit the solution for the free particle in the 1-dimensional box. The equation given was

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi(x) = E\Psi(x).$$

In this case, our linear operator is

$$\mathcal{L} = -\frac{\hbar^2}{2m}\frac{d^2}{dx},$$

and the spectrum are the possible E values one can attain.

Example 4.1: Spectrum for the Derivative on $\mathbb{R}$

In the prequel, we began studying ODEs with the equation

$$\frac{d}{dx}f(x) = kf(x).$$

Notice, that we can take

$$\mathcal{L} = \frac{d}{dx},$$

and that k is an eigenvalue with $f(x)$ the corresponding eigenfunction. Of course, we did not state the problem this way. Let us also add in the constraint of an initial condition that $f(0) = 1$. Then, we can search for the spectrum of $\mathcal{L}$.
We can note that this equation is separable, and the particular solution is

$$f(x) = e^{kx}.$$

Here, k is not constrained in any way! Thus, k can be any complex number. So, the spectrum for the first derivative operator (given this initial data) is all of $\mathbb{C}$. For any given $k \in \mathbb{C}$, the corresponding eigenfunction is then just e^{kx}.
One may ask how this helps us solve the equation

$$\mathcal{L}f = g,$$

for some given function g. There, the answer lies in the Fourier transform, which we will soon analyze.

Exercise 4.2. Review how to solve the above separable equation.

Remark 4.1. The story of this methodology for solving differential equations is a long one. There is a great history of this method beginning in the early 1800's with Joseph Fourier. As of now, there is still active research in differential equations where these types of methods are generalized and applied to more problems. We will seek to cover the linear theory with a few specific examples. Just know that many details are being glossed over for the sake of promoting understanding as opposed to rigor.

4

Fourier Theory

1 Fourier Series

With our newfound toolbox, we will investigate solving types of differential equations with an underlying periodic structure. Specifically, we will consider now linear differential equations on the region $\Omega = [0, L]$ with periodic boundary conditions. That is, we will be given some differential operator $\mathcal{L}$, a function g, and we will be asked to solve for an f such that

$$\mathcal{L}f = g, \tag{Eq. 4.1.1}$$

with f subject to the boundary conditions $f(0) = f(L)$. We refer to these boundary conditions as ***periodic boundary conditions***. Another way to view this problem is as defining the function f to be continuous on a circle domain. This point of view shows its head in the solution for the hydrogen atom.

For the sake of example, we will take the operator $\mathcal{L} = -\frac{d^2}{dx^2}$ which is referred to as the 1-dimensional ***Laplace operator*** or ***Laplacian***. To follow the methodology outlined in the previous chapter, before we approach Eq. 4.1.1, we must find the spectrum of $\mathcal{L}$. Hence, we must solve the eigenvalue equation

$$\mathcal{L}f = \omega^2 f,$$

where ω^2 is some constant that was chosen to simplify the later notation. Thus, we arrive at the differential equation

$$-\frac{d^2}{dx^2}f(x) = \omega^2 f(x) \qquad \text{with } f(0) = f(L). \tag{Eq. 4.1.2}$$

We have in fact solved this equation before! This is the harmonic oscillator equation with angular frequency ω. However, there is a difference. In the previous case, we had a distinct

and predetermined value for ω that was given by the system we were investigating (e.g., the stiffness of a spring). Now, we allow ω to be a parameter for the problem as well.

When we solved the matrix eigenvalue problem

$$[A]\vec{\boldsymbol{e}} = \lambda\vec{\boldsymbol{e}}$$

we wanted to take the determinant of $[A] - \lambda[I]$ to find what the eigenvalues were and then use this information to determine the eigenvectors. Differential operators work a bit differently. Instead, we would find a general solution to the differential equation and continue on from there to refine our answer.

In our equation Eq. 4.1.2, we know the general solution is

$$f(x) = C_1 \sin(\omega x) + C_2 \cos(\omega x).$$

Exercise 1.1. Determine the above general solution using either the characteristic polynomial (which in reality comes from a determinant) or by using a power series.

With our general solution in hand, we can determine what eigenvalues for ω are reasonable. Specifically, we use our boundary conditions and we force

$$f(0) = C_1 \sin(\omega \cdot 0) + C_2 \cos(\omega \cdot 0) = C_2$$

and

$$f(L) = C_1 \sin(\omega L) + C_2 \cos(\omega L).$$

In order for the sin term above to disappear, we must have

$$\omega L = n\pi,$$

for any integer value n. But we also require that $\cos(\omega L) = 1$ which means that we restrict further to

$$\omega L = 2n\pi.$$

We can then solve for ω to find

$$\boxed{\omega = \frac{2n\pi}{L}.}$$

Exercise 1.2. Review the analysis above for determining ω.

Then, what we have found is the spectrum is discrete, and there is an eigenvalue

$$\omega^2 = \frac{4n^2\pi^2}{L^2}$$

for each value of n. But, note that since n is an integer, n^2 is always positive, and so we require $n \geq 0$ since we wish to remove the redundant results. We can then list off the eigenfunctions

$$1,\ \sin\left(\frac{2\pi x}{L}\right),\ \cos\left(\frac{2\pi x}{L}\right),\ \sin\left(\frac{4\pi x}{L}\right),\ \cos\left(\frac{4\pi x}{L}\right),\ \ldots.$$

1.1 Solving Equations

The equation Eq. 4.1.1 before is linear, and hence any sum of solutions is also a solution. Thus, we can write a solution $f(x)$ as a series by putting

$$f(x) = \sum_{n=1}^{\infty} a_n \sin\left(\frac{2n\pi x}{L}\right) + \sum_{n=0}^{\infty} b_n \cos\left(\frac{2n\pi x}{L}\right).$$

However, we can then consider if we can take any g and write this as a series as well.

Let us define the following inner product

$$\langle F, G\rangle = \frac{1}{L}\int_0^L F(x)G*(x)dx,$$

which we will refer to as the ***Fourier inner product***. Notice, this inner product is very similar to the inner products we have used for functions previously (we just divide by the length of the interval).

Though we won't prove it here, the eigenfunctions found above serve as a basis for the solutions for the equation Eq. 4.1.1. To simplify future work, we will normalize our eigenfunctions. Thus, we must solve

$$\langle c1, c1\rangle = 1, \quad \left\langle c_o \sin\left(\frac{2n\pi x}{L}\right), c_o \sin\left(\frac{2n\pi x}{L}\right)\right\rangle = 1, \quad \left\langle c_e \cos\left(\frac{2n\pi x}{L}\right), c_e \cos\left(\frac{2n\pi x}{L}\right)\right\rangle = 1$$

Solving for the constants, we get

$$c = 1, \quad c_o = \sqrt{2}, \quad c_e = \sqrt{2}.$$

Thus, our orthonormal basis for this inner product is given by

$$\boxed{1,\ \sqrt{2}\sin\left(\frac{2\pi x}{L}\right),\ \sqrt{2}\cos\left(\frac{2\pi x}{L}\right),\ \sqrt{2}\sin\left(\frac{4\pi x}{L}\right),\ \sqrt{2}\cos\left(\frac{4\pi x}{L}\right),\ \ldots.}$$

Thus, using this basis, we can decompose functions into their ***Fourier series***. That is, given a function f, we can compute coefficients a_n, and b_n so that

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n\sqrt{2}\cos\left(\frac{2\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n\sqrt{2}\sin\left(\frac{2\pi x}{L}\right).$$

Since we have constructed these functions to be an orthonormal basis, we can compute the coefficients by projection. For example, we can compute the constant term a_0 by

$$a_0 = \langle f, 1\rangle,$$

as well as the terms a_n and b_n by

$$a_n = \left\langle f, \sqrt{2}\cos\left(\frac{2n\pi x}{L}\right)\right\rangle \quad \text{and} \quad b_n = \left\langle f, \sqrt{2}\sin\left(\frac{2n\pi x}{L}\right)\right\rangle.$$

Example 1.1: Square Wave

The classic example one computes for a Fourier series is for the square wave

$$f(x) = \begin{cases} 0 & x < L/2 \\ 1 & x \geq L/2, \end{cases}$$

for $x \in [0, L]$. Now, this function is piecewise constant and we will use the fact that we can split up an integral by

$$\int_0^L f(x)\psi_n(x)dx = \int_0^{L/2} f(x)\psi_n(x)dx + \int_{L/2}^L f(x)\psi_n(x)dx.$$

Hence, we can compute the Fourier coefficients for this function f. First, we compute the constant term

$$\begin{aligned} a_0 = \langle f, 1\rangle &= \frac{1}{L}\int_0^L f(x)\cdot 1 dx \\ &= \frac{1}{L}\int_0^{L/2} f(x)dx + \frac{1}{L}\int_{L/2}^L f(x)dx \\ &= \frac{1}{L}\int_{L/2}^L 1 dx \\ &= \frac{1}{2}. \end{aligned}$$

This value for a_0 is the average value of the function f. Next, we can compute the constants a_n by

$$\begin{aligned} a_0 &= \left\langle f, \sqrt{2}\cos\left(\frac{2n\pi x}{L}\right)\right\rangle \\ &= \frac{1}{L}\int_0^L f(x)\cdot 1 dx \\ &= \frac{1}{L}\int_0^{L/2} f(x)\sqrt{2}\cos\left(\frac{2n\pi x}{L}\right)dx + \frac{1}{L}\int_{L/2}^L f(x)\sqrt{2}\cos\left(\frac{2n\pi x}{L}\right)dx \\ &= \frac{1}{L}\int_{L/2}^L \sqrt{2}\cos\left(\frac{2n\pi x}{L}\right)dx \\ &= \frac{\sin(2\pi n) - \sin(\pi n)}{\sqrt{2}\pi n} \\ &= 0, \end{aligned}$$

since n is an integer and $\sin(\pi n) = \sin(2\pi n) = 0$ for every integer. Next, we can

compute the constants b_n by

$$\begin{aligned} a_0 &= \left\langle f, \sqrt{2}\sin\left(\frac{2n\pi x}{L}\right)\right\rangle \\ &= \frac{1}{L}\int_0^L f(x)\cdot 1 dx \\ &= \frac{1}{L}\int_0^{L/2} f(x)\sqrt{2}\sin\left(\frac{2n\pi x}{L}\right)dx + \frac{1}{L}\int_{L/2}^{L} f(x)\sqrt{2}\sin\left(\frac{2n\pi x}{L}\right)dx \\ &= \frac{1}{L}\int_{L/2}^{L} \sqrt{2}\sin\left(\frac{2n\pi x}{L}\right)dx \\ &= \frac{\cos(\pi n)-\cos(2\pi n)}{\sqrt{2}\pi n}. \end{aligned}$$

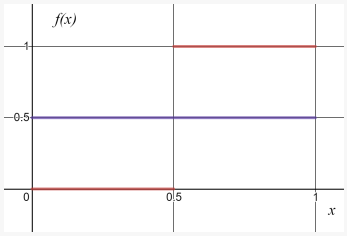
Thus, we have that

$$f(x) = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{\cos(\pi n)-\cos(2\pi n)}{\pi n}\sin\left(\frac{2n\pi x}{L}\right).$$

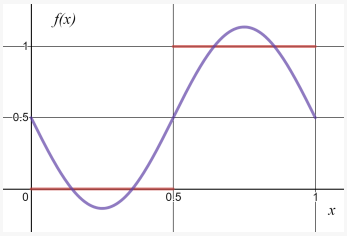
We can then use a finite sum to approximate our function by

$$f(x) \approx \frac{1}{2} + \sum_{n=1}^{N} \frac{\cos(\pi n)-\cos(2\pi n)}{\pi n}\sin\left(\frac{2n\pi x}{L}\right),$$

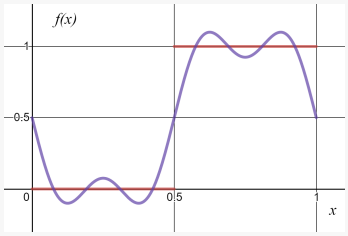
for various values of N



(a) $N = 0$ approximation.



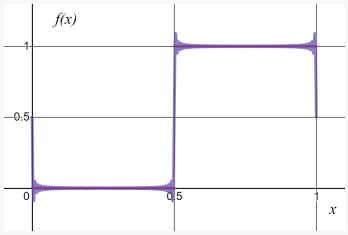
(b) $N = 1$ approximation.



(c) $N = 3$ approximation.

(d) $N = 10$ approximation.

(e) $N = 50$ approximation.



(f) $N = 100$ approximation.

Figure 4.1: True $f(x)$ in red. Approximations in purple.

2 Fourier Transform

Let us transition into a more general point of view of the Fourier series. Specifically, we wish to utilize the Fourier transform. The idea behind the Fourier transform is to

convert a given function f into a new one that we denote by $\hat{f}$ that contains the frequency information of f. This new function will look different, but it will be (in some way) equivalent to our original function. However, we do gain some ability to solve differential equations with the transformed function.

First, we can rewrite the Fourier series by using complex functions. Again, we consider functions defined on the interval $[0, L]$ and take the functions

$$e^{\frac{i2n\pi x}{L}} \qquad n = \ldots, -2, -1, 0, 1, 2, \ldots,$$

as our orthonormal basis functions. Note that these are indeed normalized when we take the inner product

$$\begin{aligned}\left\langle e^{\frac{i2n\pi x}{L}}, e^{\frac{i2n\pi x}{L}} \right\rangle &= \frac{1}{L}\int_0^L e^{\frac{i2n\pi x}{L}} e^{\frac{-i2n\pi x}{L}} dx \\ &= \frac{1}{L}\int_0^L 1 dx \\ &= 1.\end{aligned}$$

In some way we can now see that the functions are a bit more natural to use. Recall of course that Euler's formula gives us the connection from this representation to the one we previously discussed. That is,

$$e^{\frac{i2n\pi x}{L}} = \cos\left(\frac{2n\pi x}{L}\right) + i\sin\left(\frac{2n\pi x}{L}\right).$$

Thus, we can write functions by taking the following linear combination

$$f(x) = \sum_{-\infty}^{\infty} c_n e^{\frac{i2n\pi x}{L}},$$

where the constants c_n are allowed to be complex valued. Then, the ***Fourier transform*** of the function f, denoted by $\hat{f}$, is given by

$$\hat{f}(k) = c_k,$$

is the coefficients of from the Fourier series. Thus, what we find is that the Fourier transform describes "how much" of the frequency $\frac{2k\pi}{L}$ makes up the function f.

Example 2.1: Square Wave Fourier Transform

If we consider the square wave function given by

$$f(x) = \begin{cases} 0 & x < L/2 \\ 1 & x \geq L/2, \end{cases},$$

then we can compute the Fourier components with the complex basis functions

$$e^{\frac{i2n\pi x}{L}}.$$

Thus, we compute the coefficients in the same by

$$\begin{aligned} c_k &= \left\langle f, e^{\frac{i2n\pi x}{L}} \right\rangle \\ &= \frac{ie^{i\pi k}(-1-e^{i\pi k})}{2\pi k} \\ &= \frac{i(-1)^k(-1-(-1)^k)}{2\pi k} \end{aligned}$$

Hence, the Fourier transform of f is given by

$$\hat{f}(k) = \frac{i(-1)^k(-1-(-1)^k)}{2\pi k}.$$

We'll arrive at the usefulness of the Fourier transform in a bit, but we will first want to consider a yet more general version of the Fourier transform. Instead of taking functions defined solely on the interval $[0, L]$, we can create a transform that works for functions defined on all of $\mathbb{R}$ or any subset. The difference is this: we now take a basis of functions to be

$$e^{i2\pi kx} \qquad k \in \mathbb{R}.$$

That is, instead of taking a discrete set of basis functions, we will now need a continuous set of basis functions. The reason for this, heuristically, is due to the fact that there can be a full continuum of frequencies for functions defined on $\mathbb{R}$. The details are suitably more involved than we can get into for a course at this level!

One may now see why we have built up this framework in the way that we did. It was a bit of a drawn out process, but it allows us to compute the Fourier transform in a consistent way. Given a function $f(x)$, we can define the Fourier transform by taking the Hermitian inner product for $\mathbb{R}$ and projecting our function onto the basis elements. For example, we have

$$\hat{f}(k) = \left\langle f, e^{i2\pi kx} \right\rangle = \int_{-\infty}^{\infty} f(x)e^{-i2\pi kx}dx.$$

And that's really all! However, one must mention an immensely important function before we compute some examples.

2.1 Fourier Transform Operator

Often time, it will be useful to talk of the Fourier transform as an operator. For this, we will use the notation

$$\mathcal{F}[f(x)] = \hat{f}(k).$$

Later, we will introduce the inverse operation in that

$$\mathcal{F}^{-1}\left[\hat{f}(k)\right] = f(x).$$

These operators are fundamental in solving differential equations!

One should also note that the Fourier transform is a unitary linear operator. Specifically, if we have two functions f and g and a constant $\alpha \in \mathbb{C}$, then

$$\mathcal{F}[f(x) + \alpha g(x)] = \hat{f}(k) + \alpha\hat{g}(k).$$

Now, the operator is unitary since we also have that

$$\langle f, g\rangle = \left\langle \hat{f}, \hat{g}\right\rangle,$$

where the integrals used to evaluate the inner products are taken over the variables x and k respectively.

The fact that the Fourier transform is a unitary operator is rather important. It means that the transformation does not disturb the measurements one could wish to make. In the same vein, it allows one to work with the transformed function $\hat{f}(k)$ without losing any information about the system. We'll find that working with $\hat{f}(k)$ is often an easier task.

3 The Dirac Delta Function

Studying the Fourier theory naturally brings about a very special (generalized) function known as the ***Dirac delta function*** which we denote by $\delta(x)$. Quite simply, we define this function via an integral. Specifically, we have for any function $f(x)$ that

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0).$$

Moreover, if the interval $[a, b]$ contains 0, we have

$$\int_{a}^{b} f(x)\delta(x)dx = f(0).$$

We can change the input value for $\delta(x)$ by taking $\delta(x - x_0)$ and we have

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)dx = f(x_0).$$

Put in more simple terms, δ is the function that, when integrated with another function f, evaluates that function at a given input value (so long as we integrate over that input value).

Remark 3.1. $\delta(x)$ is in fact not a function at all. It is something more general known as a distribution. But this fact is not entirely relevant. We will continue to refer to $\delta(x)$ as a function despite this slight misnomer.

Why should one even consider such a function? For one, it will show up quite readily for us when using the Fourier transform. But, moreover, it is a physically meaningful function in that describes a concentration of mass at a single point. For example, when studying electromagnetism, one will talk of charged particles. Often, those charged particles are thought of as single points with a charge q. To determine the total charge in a region, one would perform an integral like

$$\int_{-\infty}^{\infty} q\delta(x)dx = q$$

which shows the total charge q. The fact that we have $q\delta(x)$ tells us that all of the charge is concentrated at $x = 0$.

One can define the delta function in another convenient manner. Specifically, via the Fourier transform. One can also show that

$$\delta(x) = \int_{-\infty}^{\infty} 1 e^{-i2\pi kx} dx,$$

which means that the δ function is the Fourier transform of the constant function. Moreover,

$$c\delta(x) = \int_{-\infty}^{\infty} c e^{-i2\pi kx} dx.$$

One should believe this on an intuitive level since the Fourier transform of a function returns the function's frequency components. A constant function has only a zero frequency component and hence the transform must reflect this.

Remark 3.2. Computing these integrals requires tools from complex analysis that we have not seen. Thus, we will have to blackbox how these integrals are computed in exchange for their usefulness.

4 Computing Fourier Transforms

To see what we are working with, we should compute a few examples. Once again, the techniques to compute these integrals are beyond the scope of this text, but we can show a few results anyways. Our first goal is to see what is meant by frequency components of a function.

If we start with functions with a constant period of oscillation, we can try to digest what the output of the Fourier transform is telling us. We will consider three functions, each with the same period, and compare their transforms.

Example 4.1: Fourier Transform of Cosine

Let us consider the Fourier transform of the function $f(x) = \cos(x)$. We can compute this transformation by

$$\begin{aligned}\mathcal{F}[f(x)] &= \int_{-\infty}^{\infty} \cos(x) e^{i2\pi kx} dx \\ &= \frac{\delta\left(k - \frac{1}{2\pi}\right) - \delta\left((k + \frac{1}{2\pi}\right)}{2}.\end{aligned}$$

Here, we can see that the recovered frequency components are where the input to the delta functions are zero. That is,

$$-2\pi k - 1 = 0 \implies k = -\frac{1}{2\pi},$$

and

$$1 - 2\pi k = 0 \implies k = \frac{1}{2\pi}.$$

Recall that the period of $\cos(x)$ is 2π and hence the frequency $\nu = \frac{1}{2\pi}$. Here, we can see that there is both a $\pm\frac{1}{2\pi}$.

Part V

In the previous example, one can see how the Fourier transform breaks the function down into frequency components. However, we also know of another function whose frequency is 2π. Namely, $\sin(x)$. Note that these functions are different in that $\cos(x)$ is an even function and $\sin(x)$ is an odd function, so we should expect their Fourier transforms to differ as well (or else this is not an invertible process).

Example 4.2: Fourier Transform of Sine

Similarly, we could consider the Fourier transform of the function $f(x) = \sin(x)$. We'll take

$$\begin{aligned}\mathcal{F}[f(x)] &= \int_{-\infty}^{\infty} \sin(x) e^{i2\pi kx} dx \\ &= \frac{\delta\left(k - \frac{1}{2\pi}\right) - \delta\left(k + \frac{1}{2\pi}\right)}{2i}\end{aligned}$$

What we see is that the constants in front of the delta functions are different for $\sin(x)$. Specifically, we see the inclusion of the imaginary unit i.

The differences above display how the Fourier transform is capturing more information about a function than just the frequency information. In some sense, the Fourier transform can determine how even or odd a function is as well. To finalize this section, let us consider the Fourier transform of another function with frequency $\frac{1}{2\pi}$.

Example 4.3: Fourier Transform of Complex Exponential

Let $f(x) = e^{ix}$, and note that this function has a period of 2π and hence a frequency of $\frac{1}{2\pi}$. We can also recall that by Euler's formula, we have

$$e^{ix} = \cos(x) + i\sin(x),$$

and thus we can compute the Fourier transform by

$$\begin{aligned}\mathcal{F}\left[e^{ix}\right] &= \mathcal{F}[\cos(x) + i\sin(x)] \\ &= \mathcal{F}[\cos(x)] + i\mathcal{F}[\sin(x)] \\ &= \delta\left(k - \frac{1}{2\pi}\right)\end{aligned}$$

Again, we find that the Fourier transform captures enough information about our functions to properly differentiate between them.

This section would not be complete without some reference of common Fourier transforms. We'll place a few here, but there are many other references to compute more examples (see for example Wikipedia's page on *Tables of Important Fourier Transforms*.

$f(x)$	$\hat{f}(k)$
$\delta(x)$	1
1	$\delta(k)$
e^{iax}	$\delta\left(k-\frac{a}{2\pi}\right)$
$\cos(ax)$	$\frac{\delta\left(k-\frac{a}{2\pi}\right)+\delta\left(k+\frac{a}{2\pi}\right)}{2}$
$\sin(ax)$	$\frac{\delta\left(k-\frac{a}{2\pi}\right)-\delta\left(k+\frac{a}{2\pi}\right)}{2i}$
$e^{-\alpha x^2}$	$\sqrt{\frac{\pi}{\alpha}}e^{\frac{(\pi k)^2}{\alpha}}$

5 The Inverse Fourier Transform

The most striking property of the Fourier transform is that it is invertible. Coupled with its linear and unitary nature, the Fourier transform has many nice properties that allow one to make great use of it. Up until this point, we have only dealt with the forward direction of the Fourier transform. If we are given a function $f(x)$, we can convert this function into $\hat{f}(k)$ which we refer to as the transformed function in the ***frequency domain***. We put

$$\mathcal{F}\left[f(x)\right] = \hat{f}(k).$$

Now, the claim is that this process is invertible in that if we are given a function $\hat{f}(k)$ in the frequency domain, we can convert it back to a function $f(x)$. That is, we want

$$\mathcal{F}^{-1}\left[\hat{f}(k)\right] = f(x).$$

Now, since the Fourier transform is a unitary operator, we need only find the adjoint to the Fourier transform. Recall that, given a unitary operator $\mathcal{U}$ we can find $\mathcal{U}^{-1}$ by solving

$$\mathcal{U}\mathcal{U}^{-1} = I,$$

where I is the identity. However, the defining property of a unitary operator is that we have

$$\mathcal{U}^{-1} = \mathcal{U}^{\dagger}.$$

That is, the inverse of a unitary operator is its adjoint!

Hence, it follows that the ***inverse Fourier transform*** of a function $\hat{f}(k)$ is given by

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(k)e^{i2\pi kx}dk.$$

If one revisits example 2.2, then we can see that the adjoint to multiplication by a phase is to multiply by the negative phase. That is, the adjoint to $e^{i\theta}$ is $e^{-i\theta}$. That is the fact that lets us define the inverse Fourier transform above.

In **??**, one can find the inverse Fourier transform of a given function by starting with a $\hat{f}(k)$ and seeing what the corresponding $f(x)$ is. We'll not worry so much about inverting other functions for now.

6 Solving Differential Equations

Perhaps the most unique feature of the Fourier transform is the transformation of derivatives of a function. One could in fact argue that this was the true nature of the transformation from the beginning. Consider a function $f(x)$, then we can differentiate $f(x)$ to get the function $f'(x)$. Now, what is the Fourier transform

$$\mathcal{F}[f'(x)]?$$

First, (though we have yet to mention this) let us note that any $f(x)$ must be a function that decays rapidly as $x \to \pm\infty$. Let's write this out. We have

$$\begin{aligned}\mathcal{F}[f'(x)] &= \int_{-\infty}^{\infty} f'(x)e^{-i2\pi kx}dx \\ &= i2\pi k\int_{-\infty}^{\infty} f(x)e^{i2\pi kx}dx \\ &= i2\pi k\hat{f}(k).\end{aligned}$$

Exercise 6.1. Using integration by parts, show that the work above is correct.

What we have found is that the Fourier transform turns derivatives into a multiplication operator! This can be summed up in the following way:

$$\boxed{\mathcal{F}\left[\frac{d^n}{dx^n}f(x)\right] = (i2\pi k)^n\hat{f}(k).}$$

We also have the following as well

$$\boxed{\mathcal{F}[x^n f(x)] = \left(\frac{i}{2\pi}\right)^n \frac{d^n}{dk^n}\hat{f}(k).}$$

This means that the Fourier transform turns multiplication into differentiation.

What follows is an algebraic method for solving differential equations. Specifically, any linear differential equation can be quickly transformed into a new (possibly easier) differential equation. The outline can be summarized in the following steps.

1. Take a Fourier transform of both sides of your differential equation.
2. Solve the new equation for $\hat{f}(k)$ in terms of the frequency variable k.
3. Use the inverse Fourier transform to convert $\hat{f}(k)$ back to $f(x)$.

We'll first start with an example that uses the Fourier series as opposed to the Fourier transform. Underlying the solution is the same principal, but it gives us a better means for understanding the methodology.

Example 6.1: Fourier Series Solution of a Second Order Equation

Consider an elastic band suspended from atop two poles subject to an external force.

We can model this equation by

$$-\mu \frac{d^2}{dx^2} f(x) = g(x),$$

with the boundary conditions $f(0) = f(L) = 0$. Here, μ is an elastic constant that describes the stiffness of the elastic. Then, we let

$$g(x) = \begin{cases} 0 & x < L/4 \\ 1 & L/4 \leq x < 3L/4 \\ 0 & 3L/4 < x \end{cases},$$

which has the following graph.

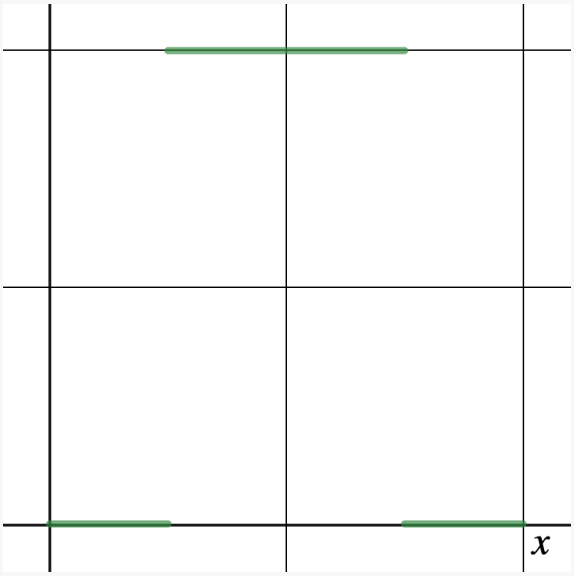


Figure 4.2: A plot of the external force $g(x)$.

We can also write $g(x)$ as a Fourier series by

$$g(x) = \frac{1}{2} + \sum_{n=1}^{N} \frac{\sin\left(\frac{3\pi n}{2}\right) - \sin\left(\frac{\pi n}{2}\right)}{\pi n} \cos\left(\frac{2n\pi x}{L}\right).$$

Let us also suppose that $f(x)$ is written as a Fourier series,

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \sqrt{2} \cos\left(\frac{2n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sqrt{2} \sin\left(\frac{2n\pi x}{L}\right).$$

Then, we can plug both Fourier series into the differential equation to get

$$\frac{4n^2\pi^2\mu}{L^2} \left(\sum_{n=1}^{\infty} a_n \sqrt{2} \cos\left(\frac{2n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sqrt{2} \sin\left(\frac{2n\pi x}{L}\right) \right)$$
$$= \frac{1}{2} + \sum_{n=1}^{N} \frac{\sin\left(\frac{3\pi n}{2}\right) - \sin\left(\frac{\pi n}{2}\right)}{\sqrt{2}\pi n} \sqrt{2} \cos\left(\frac{2n\pi x}{L}\right).$$

Our task now is to solve for the coefficients a_n and b_n and later determine the constant a_0 from boundary data. In the above, it's clear that $b_n = 0$ and we also have

$$a_n = \frac{L^2\left(\sin\left(\frac{3n\pi}{2}\right) - \sin\left(\frac{n\pi}{2}\right)\right)}{4\sqrt{2}\pi^3 n^3 \mu}.$$

Thus we have the general solution

$$f(x) = a_0 + \frac{1}{\mu}\sum_{n=1}^{\infty} \frac{L^2\left(\sin\left(\frac{3n\pi}{2}\right) - \sin\left(\frac{n\pi}{2}\right)\right)}{4\pi^3 n^3} \cos\left(\frac{2n\pi x}{L}\right).$$

Since we require that $f(0) = f(L) = 0$ we take

$$0 = f(0) = a_0 + \frac{1}{\mu}\sum_{n=1}^{\infty} \frac{L^2\left(\sin\left(\frac{3n\pi}{2}\right) - \sin\left(\frac{n\pi}{2}\right)\right)}{4\pi^3 n^3}.$$

Evaluating the above infinite series gives us

$$0 = a_0 - \frac{1}{64\mu},$$

and so $a_0 = \frac{1}{64\mu}$. Finally, we have our solution

$$\boxed{f(x) = \frac{1}{64\mu} + \frac{1}{\mu}\sum_{n=1}^{\infty} \frac{L^2\left(\sin\left(\frac{3n\pi}{2}\right) - \sin\left(\frac{n\pi}{2}\right)\right)}{4\pi^3 n^3} \cos\left(\frac{2n\pi x}{L}\right).}$$

Letting $\mu = \frac{1}{10}$, we can plot our solution.

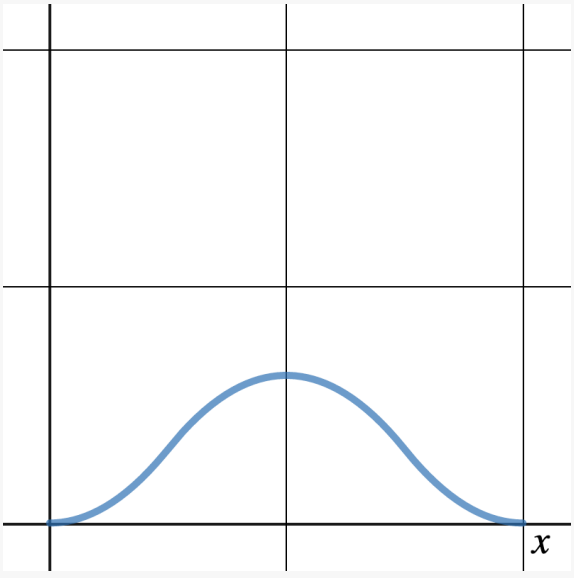


Figure 4.3: An approximation to the solution $f(x)$ with $N = 100$.

Exercise 6.2. Verify that the Fourier series for $g(x)$ above is correct.

Example 6.2: Fourier Transform Solution of a Second Order Equation

Let us consider the same problem as the previous example but with a slightly different approach. We wish to solve the boundary value problem

$$-\mu \frac{d^2}{dx^2} f(x) = g(x)$$

with $f(0) = f(L) = 0$ and with

$$g(x) = \begin{cases} 0 & x < L/4 \\ 1 & L/4 \leq x < 3L/4 \\ 0 & 3L/4 < x \end{cases}.$$

Then, we can compute the Fourier transform of $g(x)$ by

$$\hat{g}(k) = \frac{1}{L} \int_0^L g(x) e^{\frac{i2\pi kx}{L}} dx = \frac{e^{ik\pi} \sin\left(\frac{k\pi}{2}\right)}{k\pi}.$$

Similarly, we can compute the Fourier transform of the left hand side of our differential equation to get

$$\mathcal{F}\left[-\mu \frac{d^2}{dx^2} f(x)\right] = \mu \frac{4\pi^2 k^2}{L^2} \hat{f}(k).$$

Hence, the Fourier transform of the whole differential equation reads

$$\mu \frac{4\pi^2 k^2}{L^2} \hat{f}(k) = \frac{e^{ik\pi} \sin\left(\frac{k\pi}{2}\right)}{k\pi}.$$

Here, we can solve for $\hat{f}(k)$ to get

$$\hat{f}(k) = \frac{1}{\mu} \frac{L^2 e^{ik\pi} \sin\left(\frac{k\pi}{2}\right)}{4\pi^3 k^3}.$$

Inverting the Fourier transform on $[0, L]$ amounts to writing $f(x)$ as the complex Fourier series

$$\boxed{f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{\frac{i2k\pi x}{L}}.}$$

Remark 6.1. Though arguably less intuitive, the Fourier transform approach is a bit cleaner to work with. If one were to plot approximations for our function $f(x)$, we would receive the same answer as in the prior example.

In the previous examples, we were able to solve differential equations on the interval $[0, L]$ whose right hand sides were discontinuous functions. Prior to this technique, we would have had no way of solving these problems. The further usefulness of the Fourier transform comes into play as we investigate differential equations defined on unbounded sets of $\mathbb{R}$. For example, we can consider an equation defined on $\mathbb{R}$ with a discontinuous

(technically, not even function) right hand side.

Example 6.3: Fourier Transform for a Second Order Equation on $\mathbb{R}$

Consider the following equation

$$-\frac{d^2}{dx^2}f(x) = \delta(x),$$

defined on all of $\mathbb{R}$. We wish to find a general solution for this problem. To solve this, we take a Fourier transform of both sides of the equation to get

$$4\pi^2 k^2 \hat{f}(k) = 1.$$

We can then solve for the function $\hat{f}(k)$ and get

$$\hat{f}(k) = \frac{1}{4\pi^2 k^2}.$$

Thus, we can then find

$$f(x) = \mathcal{F}^{-1}\left[\frac{1}{4\pi^2 k^2}\right].$$

Computing this inverse Fourier transform yields

$$\boxed{f(x) = -\frac{|x|}{2}.}$$

The above example computes what we refer to as the ***fundamental solution*** for a differential operator. Specifically, this is when we have a differential operator $\mathcal{L}$ and we solve

$$\mathcal{L}f(x) = \delta(x).$$

In our case above, we found the fundamental solution for the 1-dimensional Laplace operator. If one generalizes this to into 3-dimensions, we will have the equation

$$\nabla \cdot (\nabla V(x, y, z)) = \delta(x, y, z),$$

which garners the solution

$$V(x, y, z) = \frac{1}{4\pi\sqrt{x^2 + y^2 + z^2}}.$$

This is how one derives the form of the electrostatic potential mathematically!

For now, we will move on to developing calculus in higher dimensions so that we can approach physical problems in space.

Part VI

Calculus in Higher Dimensions

5

Curves and Scalar Fields

1 Overview of multivariate functions

Aside from the study of linear transformations on $\mathbb{R}^n$ or $\mathbb{C}^n$, we have limited ourselves to functions of a single variable. Given that we have performed a deep analysis for functions of one variable, we can take what we have learned and apply it to new types of functions. Specifically, we will concentrate on $\mathbb{R}^3$ (or $\mathbb{R}^2$) and functions of the form:

$$\vec{\boldsymbol{\gamma}}\colon \mathbb{R} \to \mathbb{R}^3$$
$$f\colon \mathbb{R}^3 \to \mathbb{R}$$
$$\vec{\boldsymbol{F}}\colon \mathbb{R}^3 \to \mathbb{R}^3.$$

Of these three functions, the latter two are often referred to as ***fields***. These are also ***multivariate*** functions, whereas the first only has an input of a single variable.

(1) Functions of the form

$$\vec{\boldsymbol{\gamma}}\colon \mathbb{R} \to \mathbb{R}^3$$

are ***curves***. Often times, we will have

$$\vec{\boldsymbol{\gamma}}\colon [a,b] \to \mathbb{R}^3$$

when we want curves with specific endpoints. We will concentrate first on curves, so I'll save the extra detail for a bit later.

(2) Functions of the form

$$f\colon \mathbb{R}^3 \to \mathbb{R}$$

are ***scalar fields*** . These functions are useful in describing quantities like temperature in space. In this example, at each point $\vec{\boldsymbol{x}}$ in space $\mathbb{R}^3$, we can assign a single real number output $f(\vec{\boldsymbol{x}})$ that tells us this temperature.

(3) Functions of the form

$$\vec{\boldsymbol{F}}\colon \mathbb{R}^3 \to \mathbb{R}^3$$

are called ***vector fields***. Roughly speaking, at each point $\vec{\boldsymbol{x}}$ in space $\mathbb{R}^3$, we can place a vector $\vec{\boldsymbol{F}}(\vec{\boldsymbol{x}})$ that is also in $\mathbb{R}^3$. These fields are very important in describing systems that have flow. For example, fluid flow or electromagnetism are vector field theories.

2 Curves in space

We will begin with curves as they are very useful tools and this will help us visualize results for the other types of functions.

Let us consider a curve

$$\vec{\boldsymbol{\gamma}}\colon \mathbb{R} \to \mathbb{R}^3.$$

We will specify a specific curve by supplying three functions $f_1(t)$, $f_2(t)$, and $f_3(t)$. Specifically, each of these functions f_i is a function $f_i\colon \mathbb{R} \to \mathbb{R}$. Then, we can say that

$$\gamma(t) = (f_1(t), f_2(t), f_3(t)) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix}.$$

Note, I will likely use these above notations for vectors interchangeably.

Each $f_i(t)$ (for the values $i = 1, 2, 3$) is called a **coordinate function**. Intuitively, the coordinate function tells us where the curve is at a time t. That is

$$\begin{aligned} f_1(t) &\quad \text{the } x\text{-position of } \gamma \text{ at time } t, \\ f_2(t) &\quad \text{the } y\text{-position of } \gamma \text{ at time } t, \\ f_3(t) &\quad \text{the } z\text{-position of } \gamma \text{ at time } t. \end{aligned}$$

The nice thing about these coordinate functions is we entirely know how to deal with their calculus since each is a function $f_i\colon \mathbb{R} \to \mathbb{R}$.

Now, we can find the **tangent vector** or **velocity vector** to the curve γ at a time t. We covered this in the complex case, and the idea here is arguably more straightforward.

Imagine that $\gamma(t)$ is the position of a small particle at the time t. Then the velocity is the time rate of change of this position. Specifically, we see how each of the coordinate functions changes, and this will tell us how the position changes! So, we have the following.

Definition 2.1: Tangent Vector to a Curve

Given a curve γ, the **tangent vector to γ at the time t** is

$$\gamma'(t) = \lim_{\delta \to 0} \frac{\gamma(t+\delta) - \gamma(t)}{\delta}.$$

It turns out that we find γ' is

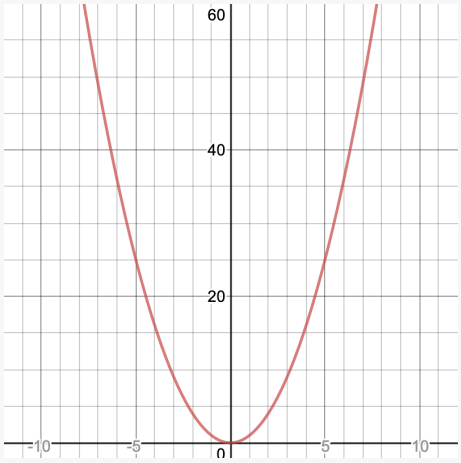
$$\gamma'(t) = (f_1'(t), f_2'(t), f_3'(t)) = \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{bmatrix}.$$

Example 2.1: Graph of a Quadratic Function

Consider the curve $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ given by

$$\gamma(t) = (t, t^2).$$

This curve looks exactly like the graph of the function $f(x) = x^2$ that we have drawn many times before.



What is the tangent vector at time t? We have

$$\gamma'(t) = (1, 2t).$$

If we take this y-value over the x-value we arrive at the same conclusion for the derivative to $f(x) = x^2$ (i.e., $f'(x) = 2x$).

Example 2.2: Circle Curve

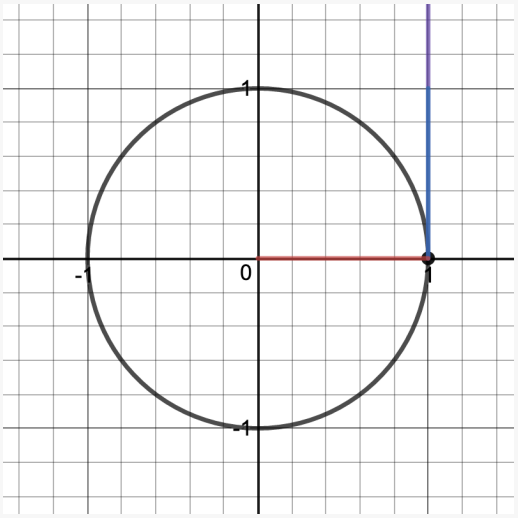
Consider the curve $\gamma\colon [0, 1] \to \mathbb{R}^2$ given by

$$\gamma(t) = (\cos(2\pi t), \sin(2\pi t)).$$

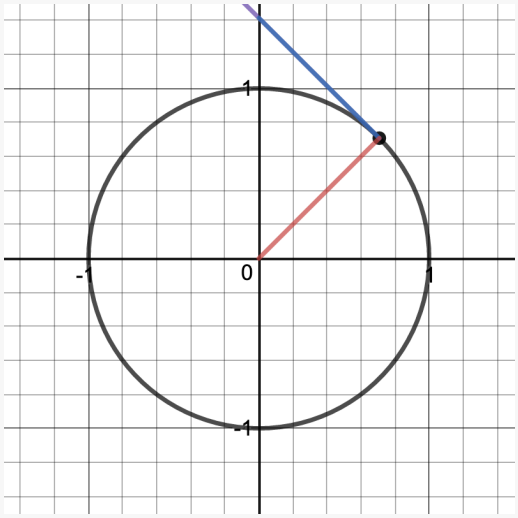
This curve is a circle of radius 1 centered at $(0, 0)$. We can find the tangent vector at a time t by

$$\gamma'(t) = (-2\pi \sin(2\pi t), 2\pi \cos(2\pi t)).$$

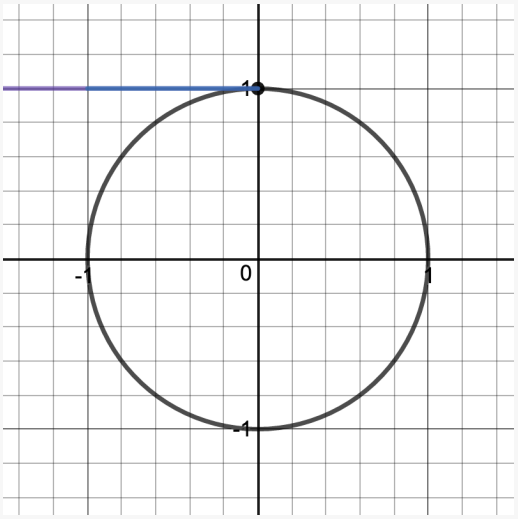
See the following graphs



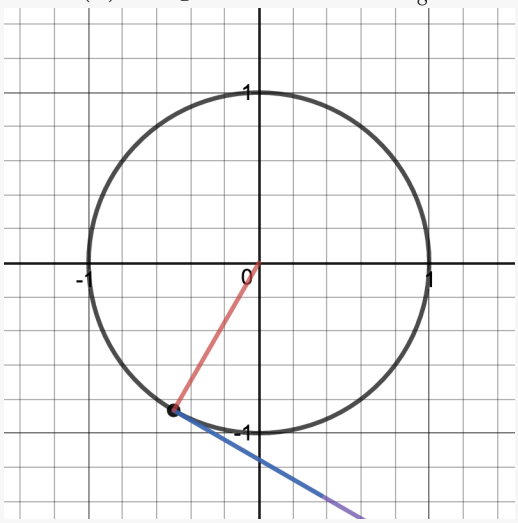
(a) Tangent vector at $t = 0$.



(b) Tangent vector at $t = \frac{1}{8}$.



(c) Tangent vector at $t = \frac{1}{4}$.



(d) Tangent vector at $t = \frac{2}{3}$.

We can also compute the **normal vector** or **acceleration vector** to a curve by taking another derivative. We write

$$\gamma''(t) = (-4\pi^2 \cos(2\pi t), -4\pi^2 \sin(2\pi t)).$$

Line integrals

In reality, we like to add up function values along curves. We have already done this before in the complex case with contours. You also did this in a calculus I course via the Riemann integral. We will get to this, but we need to know the correct functions to add up along a curve first. These will be the vector and scalar fields fields.

6 Vector Fields

1 Differentiation of fields

One may now wonder where the calculus comes in. We are almost there, but we should know what it means intuitively to use the calculus tools with these more general fields.

Curves will prove to be a great tool in the analysis of these scalar fields. We will also need to understand how vectors transform from point to point which will require us to recall the knowledge we gained on linear transformations.

Roughly speaking, we will have:

- **Curves:** Derivatives of curves $\gamma(t)$ are vectors. Specifically, we have already seen the tangent vector $\gamma'(t)$ and acceleration vector $\gamma''(t)$.
- **Scalar Fields:** Derivatives of scalar fields $f(x, y, z)$ will depend on which input of the function we change. We can collect the partial derivatives corresponding to finding derivatives in one input into a vector called the *gradient*.
- **Vector Fields:** Derivatives of vector fields $\mathbf{v}(x, y, z)$ will require us to see how each component of $\mathbf{v}$ changes based on how each individual input changes. In this case, this collection of derivatives will be put into a matrix known as the *total derivative*.

Remark 1.1. The important notion that we need to understand is the following.

The derivative of a function at a point is the best linear approximation to the function.

Based on this wording, it is intuitive to imagine derivatives as vectors or matrices.

2 Functions of Fields

In our world, we often care about combining different fields together. For example, we can take the electric field created by a single charged particle and add this field to another

field created by a different charged particle. What I mean, is we can write

$$\mathbf{v}(x,y,z) + \mathbf{u}(x,y,z)$$

and make sense of this. Just as we did with vectors, we add the components together! That is if we have

$$\begin{aligned}\mathbf{v}(x,y,z) &= (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z))\\ \mathbf{u}(x,y,z) &= (g_1(x,y,z), g_2(x,y,z), g_3(x,y,z)),\end{aligned}$$

then

$$\mathbf{v}(x,y,z) + \mathbf{u}(x,y,z) = (f_1+g_1, f_2+g_2, f_3+g_3).$$

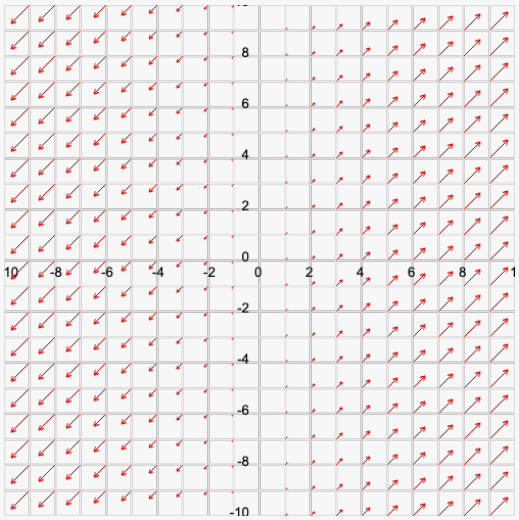
Intuitively, this just adds together the vectors at each point!

Example 2.1: Addition of vector fields

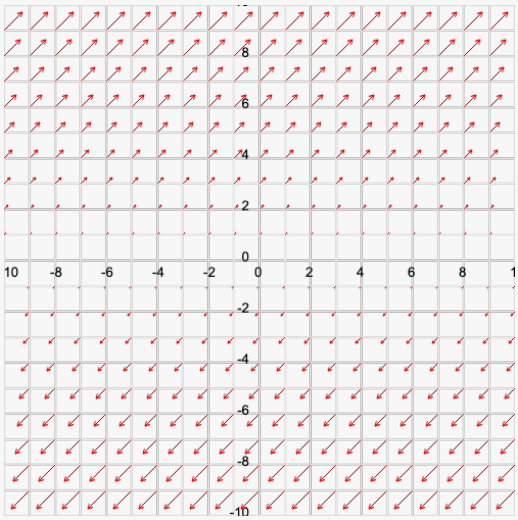
Consider the following vector fields

$$\begin{aligned}\mathbf{v}(x,y) &= (x,x)\\ \mathbf{u}(x,y) &= (y,y).\end{aligned}$$

These look like:



(a) Vector field $\mathbf{v}$.



(b) Vector field $\mathbf{u}$.

Adding these results in the field

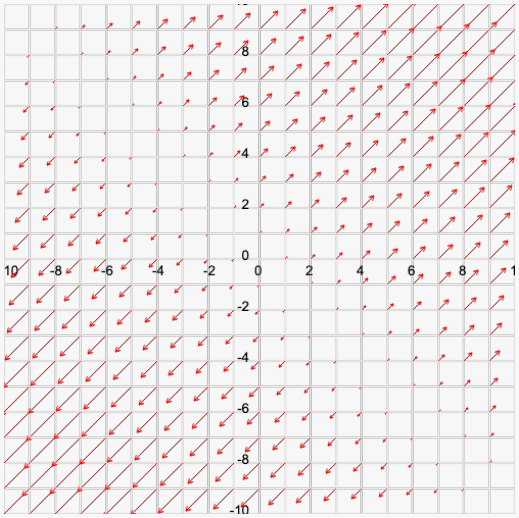


Figure 6.2: The vector field $\vec{\boldsymbol{U}} + \vec{\boldsymbol{V}}$.

Remark 2.1. If we do this all for vector fields, we can take curves and scalar fields as special cases.

- Adding together scalar fields is the same as adding functions. They just have more inputs to think about!
- Adding together curves is done in the same componentwise manner that we have shown here for vector fields.

We can also scale a field by a real number. This will stretch vectors at each point.

Example 2.2: Scaling of Vector Fields

Just take the vector field

$$\mathbf{v}(x, y, z) = (f_1(x, y, z), f_2(x, y, z), f_3(x, y, z))$$

then we can take

$$\lambda\mathbf{v} = (\lambda f_1, \lambda f_2, \lambda f_3).$$

Remark 2.2. There is many reasons why the above is important. It seems our physical world plays nicely with the above concept, for one.

One thing we can actually do, and will do a bit later, is find that there are two main types of vector fields in $\mathbb{R}^3$. These will be the curl fields and divergence fields! This is important in electromagnetism.

When we were working with complex numbers, we considered integrating a complex function $f(z)$ over a contour $\gamma(t)$ from $t = a$ to $t = b$. This involved computing

$$\int_\gamma f(\gamma)d\gamma = \int_a^b f(\gamma(t))\gamma'(t)dt.$$

I introduced this concept since it appears in the context of multivariate functions. However, in the multivariate case, we needed some more tools.

Now, the idea is as follows. We will be given a scalar field, $f(x, y, z)$ or a vector field $\mathbf{F}(x, y, z)$ and a curve $\gamma(t)$.

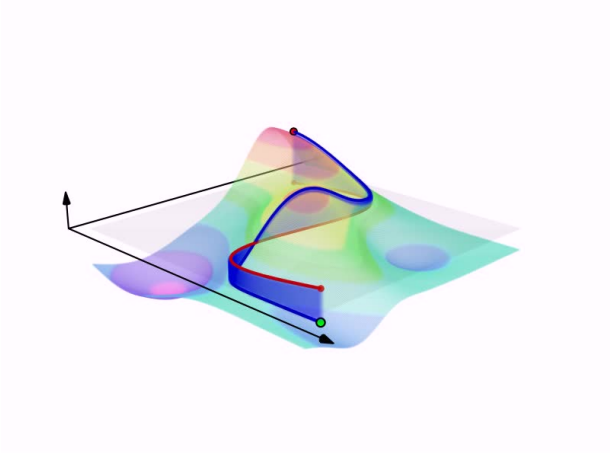
3 Line integrals of scalar fields

For the purpose of visualization, we will look at scalar fields of two variables and curves in the plane. Our set up will have $f(x, y)$ and $\gamma(t)$ over the time $t = a$ to $t = b$.

We want to understand the following:

$$\int_\gamma f(\gamma) d\gamma := \int_a^b f(\gamma(t)) \|\gamma'(t)\| dt.$$

Intuitively speaking, this integral finds the area under the curve γ along the graph of f. This is analogous to what we did in one dimension! See the following figure.



Example 3.1: Length of a curve

Let $f(x, y, z) = 1$ and $\gamma(t)$ be a curve over $t = a$ to $t = b$. Then the line integral

$$\int_\gamma f(\gamma) d\gamma = \int_a^b \|\gamma'(t)\| dt$$

is known as the *length* (sometimes *arclength* of the curve γ.

Example 3.2: Line Integral on a Paraboloid

Consider the function $f(x, y) = x^2 + y^2$ and $\gamma(t) = (t, t)$ over $t = 0$ to $t = 1$. Then the line integral

$$\int_\gamma f(\gamma) d\gamma = \int_0^1 f(\gamma(t)) \|\gamma'(t)\| dt.$$

We have

- $f(\gamma(t)) = t^2 + t^2 = 2t^2$.

- $\|\gamma'(t)\| = \|(1,1)\| = \sqrt{2}$.

So we have

$$\int_\gamma f(\gamma) d\gamma = \int_0^1 2\sqrt{2}t^2 dt.$$

This evaluates to $\frac{2\sqrt{2}}{3}$.

Exercise 3.1. Integrate $f(x,y) = x + y$ along the curve $\gamma(t) = (t, 0)$ from $t = 0$ to $t = 1$. What do you notice about this? Can we tie this to one-dimensional integration?

4 Line integrals of vector fields

There is also a type of line integral that works alongside vector fields. Roughly, the idea is to add up how much a vector field is pointing along the curve throughout the length of the curve.

Here we are given $\mathbf{F}(x,y,z)$ is a vector field, $\gamma(t)$ is a curve over the time $t = a$ to $t = b$. Then we can write

$$\int_\gamma \mathbf{F}(\gamma) \cdot d\gamma = \int_a^b \mathbf{F}(\gamma(t)) \cdot \gamma'(t) dt.$$

Example 4.1: Work Done on a Particle

The work done on a particle (or change in energy) is written as a line integral of this form.
Take for example, $\mathbf{F}(x,y) = (2x, 3y)$ and $\gamma(t) = (t, t^2)$ over the time $t = 0$ to $t = 1$. Then

$$\begin{aligned}\int_\gamma \mathbf{F}(\gamma) \cdot d\gamma &= \int_0^1 \mathbf{F}(t,t^2) \cdot (t,t^2) dt \\ &= \int_0^1 (2t, 2t^2) \cdot (t,t^2) dt \\ &= \int_0^1 2t^2 + 2t^4 dt.\end{aligned}$$

Which I'll leave to you to evaluate if you'd like.

Remark 4.1. This notion is extremely important in defining something called the *potential* of a vector field. This will show up in electrodynamics.

If the force in the example above is *conservative*, we will have a potential. This will correspond nicely to the vector field having no *curl*.

5 Derivatives

Having done some integral calculus, it's time to head back to differential calculus. We should say the following. This is far more abstract than we need, but it is an important

realization.

Definition 5.1: The Derivative

Given a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$, the **derivative of f at the point $\mathbf{x}_0$** is the best *linear approximation* to f at the point $\mathbf{x}_0$.
What this means is the following: If we zoom in to the point $\mathbf{x}_0$ and see what f does in this region, we'll notice that f is approximately a linear transformation.

Now, let us review the one-dimensional derivative that we are familiar with.

Definition 5.2: Derivative on $\mathbb{R}$

Given $f\colon \mathbb{R} \to \mathbb{R}$, we define the **derivative f' of f at the point** x_0 by

$$f'(x_0) \coloneqq \lim_{\delta \to 0} \frac{f(x_0+\delta) - f(x_0)}{\delta}.$$

This value $f'(x_0)$ is a number, and moreover is a 1×1-matrix! We often hide this idea at first.

It's a bit silly to say $f'(x_0)$ is a 1×1-matrix, but in the end it will help us to remember this.

Derivatives of scalar fields

We can investigate functions of multiple variables in more ways than the single variable case. Let us start with scalar fields.

Definition 5.3: Partial Derivatives

Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be a scalar field. Then we can consider derivatives with respect to changing each input. For example, we define the **partial derivative with respect to x at the point** (x_0, y_0, z_0), denoted $\frac{\partial f}{\partial x}$, and put

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0) \coloneqq \lim_{\delta \to 0} \frac{f(x_0+\delta, y_0, z_0) - f(x_0, y_0, z_0)}{\delta}$$

Remark 5.1. For partial derivatives, all but one variable are being held constant. So, when you are computing these, be sure to treat the proper variables as constant when necessary.

Exercise 5.1. Define $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ in a similar way to the above definition.

Exercise 5.2. Compute $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ for the function

$$f(x, y, z) = \sin(xyz) + x + 2y^2 + 3x^2 z.$$

It turns out that collecting the partial derivatives as a vector is the best linear approximation to a scalar function. We call this vector the gradient vector.

Definition 5.4: The Gradient

Given a scalar field $f(x, y, z)$, the **gradient of f at the point** (x_0, y_0, z_0), denoted $\nabla f(x_0, y_0, z_0)$ is given by

$$\nabla f(x_0, y_0, z_0) = \begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0, z_0) \\ \frac{\partial f}{\partial y}(x_0, y_0, z_0) \\ \frac{\partial f}{\partial z}(x_0, y_0, z_0) \end{bmatrix}.$$

Exercise 5.3. Compute the gradient for the function

$$f(x, y, z) = \sin(xyz) + x + 2y^2 + 3x^2 z.$$

Properties of partial derivatives and the gradient

As with the one-dimensional derivative, we have some properties that will be helpful.

Partial Derivatives:

(i) **Sum Rule:** Given $f(x, y, z)$ and $g(x, y, z)$, we have that

$$\frac{\partial}{\partial x}(f(x, y, z) + g(x, y, z)) = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}.$$

Of course, this holds for any partial derivative.

(ii) **Constant Multiple:** Given $\lambda \in \mathbb{R}$ and $f(x, y, z)$, we have that

$$\frac{\partial}{\partial x}(\lambda f(x, y, z)) = \lambda \frac{\partial f}{\partial x}.$$

Again, this holds for any partial derivative.

(iii) **Product Rule:** Given $f(x, y, z)$ and $g(x, y, z)$ we have that

$$\frac{\partial}{\partial x}(f(x, y, z)g(x, y, z)) = \frac{\partial f}{\partial x} g + f \frac{\partial g}{\partial x}.$$

Ths holds for all partial derivatives.

Remark 5.2. The chain rule will show up eventually, but not yet. As for the quotient rule, this also holds, but I don't show it here.

The Gradient:

(i) **Sum Rule:** Given $f(x, y, z)$ and $g(x, y, z)$, we have that

$$\nabla(f(x, y, z) + g(x, y, z)) = \nabla f(x, y, z) + \nabla g(x, y, z).$$

(ii) **Constant Multiple:** Given $\lambda \in \mathbb{R}$ and $f(x, y, z)$, we have that

$$\nabla(\lambda f(x, y, z)) = \lambda \nabla f(x, y, z).$$

(iii) **Product Rule:** Given $f(x, y, z)$ and $g(x, y, z)$ we have that

$$\nabla(f(x, y, z)g(x, y, z)) = (\nabla f(x, y, z))g(x, y, z) + f(x, y, z)(\nabla g(x, y, z))$$

We've learned how to compute partial derivatives and the gradient, but what are they really telling us? Remember that the derivative $\frac{d}{dx}$ of a function $f(x)$ tells us the rate of change of f as we move in the x-direction. This is very similar to what $\frac{\partial}{\partial x}$ tells us about a function $f(x, y, z)$. So we can say the following.

- $\frac{\partial f}{\partial x}$ tells us how f changes as we move in the x-direction.
- $\frac{\partial f}{\partial y}$ tells us how f changes as we move in the y-direction.
- $\frac{\partial f}{\partial z}$ tells us how f changes as we move in the z-direction.

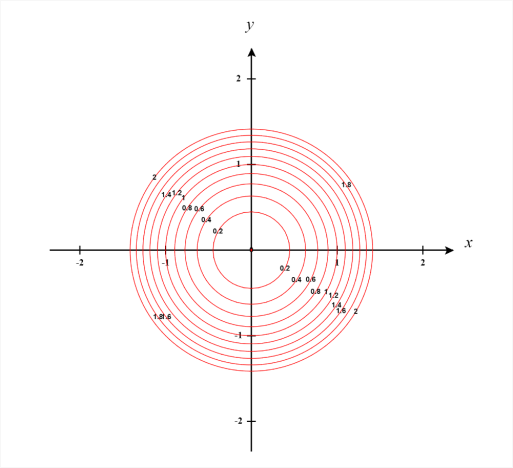
We can put these together into the gradient ∇f and know how f changes in each possible direction. Let's see how the gradient acts then.

Example 5.1: Gradients on the Paraboloid

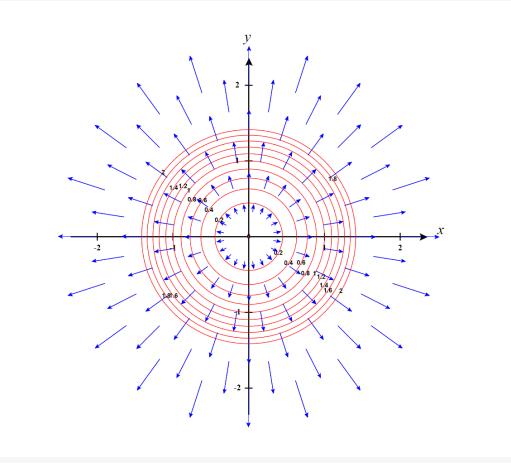
et us start with $f(x, y) = x^2 + y^2$. Then

$$\nabla f(x, y) = (2x, 2y).$$

Let us plot the level curves of this surface.



(a) Level curves of $f(x, y)$.



(b) Gradient vectors shown in blue.

- Notice that the gradient vectors point in a direction perpendicular to the level curves and the length corresponds to how close the nearest level curve is.
- The gradient is zero at the bottom of this surface.

Proposition 5.1: Gradient Points Uphill

The gradient $\nabla f(x, y, z)$ is the vector that points in the direction of greatest increase for a function $f(x, y, z)$.

How about second partial derivatives? What can we say here. We have each of the following for a function $f(x, y)$:

- $\frac{\partial^2 f}{\partial x^2}$
- $\frac{\partial^2 f}{\partial y^2}$
- $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}$
- $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}$

Recall what $\frac{d^2 f}{dx^2}$ meant for a function $f(x)$. This told us how f was curving (or what concavity f had). The story is similar for these partial derivatives.

- $\frac{\partial^2 f}{\partial x^2}$ tells us about the concavity (or curvature) of f as we move in the x direction.
- $\frac{\partial^2 f}{\partial y^2}$ tells us about the concavity (or curvature) of f as we move in the y direction.
- We actually have that $\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}$. This interpretation is a bit hard to deal with. Let's not worry too much about it.

Proposition 5.2: Partial Derivatives Commute

We have that

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}.$$

Moreover, for functions of more variables, we can say that the order we take derivatives does not matter.

Exercise 5.4. Given $f(x, y) = x^2 + y^2$, compute

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial y^2}, \quad \frac{\partial}{\partial y}\frac{\partial f}{\partial x}, \quad \frac{\partial}{\partial x}\frac{\partial f}{\partial y}.$$

What can we say about the curvature of f in the two directions? Does this make sense?

6 Optimization

In single variable calculus, we optimized functions $f(x)$ by finding the point x_0 where

$$f'(x_0) = 0.$$

We called this a *critical point*. We found if this optimizer x_0 was a maximizer or minimizer by checking the sign of second derivative $f''(x_0)$. We had

$$\text{Maximum: } f''(x_0) < 0$$
$$\text{Minimum: } f''(x_0) > 0.$$

In higher dimensions, this idea works similarly. We just have more to check.

Definition 6.1: Stationary Points

Given a function $f(x, y)$, we call a point (x_0, y_0) a **stationary point** if

$$\nabla f(x_0, y_0) = \mathbf{0}.$$

As before, we will use second derivatives to find out whether this is a maximum or a minimum.

Proposition 6.1: Maximizers and Minimizers

A stationary point (x_0, y_0) *is a*

$$\textit{Maximizer if } \frac{\partial^2 f}{\partial x^2} < 0 \quad \textit{and} \quad \frac{\partial^2 f}{\partial y^2} < 0,$$
$$\textit{Minimizer if } \frac{\partial^2 f}{\partial x^2} > 0 \quad \textit{and} \quad \frac{\partial^2 f}{\partial y^2} > 0,$$

Saddle if otherwise.

Exercise 6.1. Let

$$f(x, y) = \frac{xy}{e^{x^2+y^2}}.$$

(a) Find all stationary points for f.

(b) Determine whether these points are minimizers or maximizers.

Exercise 6.2. Show that $f(x, y) = xy$ has a saddle point at $(0, 0)$.

Lagrange Multipliers

Often times we are given a situation that we wish to find an optimum solution to, but we are somehow constrained. A biological example would be fixing a given volume for a red blood cell, and finding the optimum shape so that oxygen diffusion is maximized. A physical example would be finding the fastest path between two points when moving through a medium with varying viscosity.

The idea is relatively tame. We are given the function to optimize

$$f(x, y, z)$$

and the constraining function

$$g(x, y, z) = k.$$

Then we must solve the equation

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$$

where λ is called the **Lagrange multiplier**. Let's work through an example of solving this.

Example 6.1: Largest Box

Let's find the dimensions of the box with the largest volume if the total surface area is $64cm^2$. We must determine our function to optimize, which is the volume function

$$V(x, y, z) = xyz.$$

Our constraint is the surface area function $A(x, y, z)$ must be

$$g(x, y, z) = 2xy + 2xz + 2yz = 64,$$

which we will simplify as

$$xy + xz + yz = 32.$$

1. We first take

$$\nabla f(x, y, z) = \begin{bmatrix} yz \\ xz \\ xy \end{bmatrix}$$

and

$$\nabla g(x, y, z) = \begin{bmatrix} 2y + 2z \\ 2x + 2z \\ 2x + 2y \end{bmatrix}.$$

2. This gives us four equations to solve. Three are from the equation $\nabla f(x, y, z) = \nabla g(x, y, z)$

$$yz = \lambda(y + z) \quad \text{(Eq. 6.6.1)}$$
$$xz = \lambda(x + z) \quad \text{(Eq. 6.6.2)}$$
$$xy = \lambda(x + y), \quad \text{(Eq. 6.6.3)}$$

and one is from the constraint $g(x, y, z) = 64$

$$xy + xz + yz = 32.$$

3. Working through solving these can be nontrivial. In this case, we can do the following: Multiply (1) by x, (2) by y, and (3) by z. Which gives us

$$xyz = \lambda x(y + z)$$
$$xyz = \lambda y(x + z)$$
$$xyz = \lambda z(x + y)$$

Now we can set the first two equal to find

$$\lambda x(y + z) = \lambda y(x + z)$$

which simplifies to

$$\lambda(xz - yz) = 0$$

meaning that $\lambda = 0$ or $xz = yz$. Note that $\lambda = 0$ is not possible since that will end up giving us zero surface area and we won't satisfy the constraint.

Now $xz - yz = 0$ means that $x = y$, which we can substitute back into the equations later.

We can set the last two equal to find

$$\lambda y(x + z) = \lambda z(x + y)$$

which with similar work tells us $z = y$. So now we make note of $x = y$ and we have $x = y = z$.

So now we use the constraint equation with $x = y = z$ and find that we have

$$x^2 + y^2 + z^2 = 3x^2 = 32$$

which means that $x \approx 3.266$. Thus

$$V(3.266, 3.266, 3.266) \approx 34.8376$$

is the largest volume. *This means our ideal solution is a cube!*

7 Approximation and the Tangent Space

Sometimes it is helpful to know what a surface looks like up close. In this case, the surface is best approximated by a plane. This is analogous to how you can approximate functions of a single variable by a line.

Exercise 7.1. Compute the tangent line to $f(x) = 2x^2 + 5$ at the point $x_0 = 3$.

Equation for a Plane

We haven't worked much with planes in space yet, but we have seen surfaces. In some sense, planes are the easiest surfaces. They are, after all, linear objects.

Example 7.1: Plane and Normal

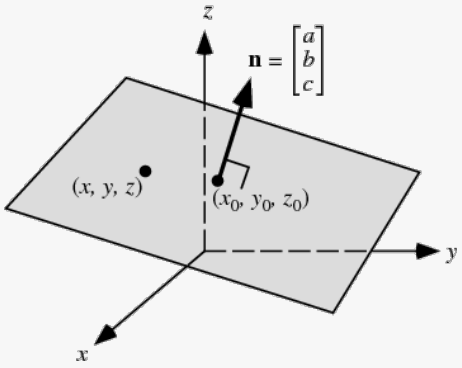
The equation for a plane is given by

$$ax + by + cz + d = 0.$$

Notice that this is a linear equation.
Then the *normal vector* to the plane is given by

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

We can see a diagram of this here.



Now, if we are given a surface (defined as a level surface or as the graph of a function), we can compute an approximation at a point called the *tangent plane*.

Example 7.2: Tangent Plane to Paraboloid

onsider the function

$$f(x, y) = -x^2 - y^2.$$

Then the graph of the function is given by plotting the points

$$(x, y, f(x, y)).$$

We compute the tangent plane by computing partial derivatives. We take

$$\frac{\partial f}{\partial x} = -2x$$
$$\frac{\partial f}{\partial y} = -2y.$$

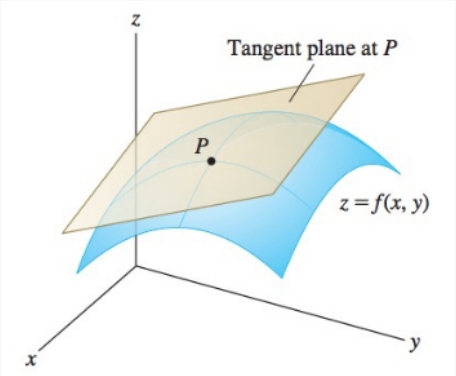
Then the equation for a tangent plane at the point $(x_0, y_0, f(x_0, y_0))$ is given by

$$z - f(x_0, y_0) = \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0).$$

So in our case, we have

$$z - (-x_0^2 - y_0^2) = -2x_0(x - x_0) - 2y_0(y - y_0).$$

Pictorially, it looks as follows (letting $p = (x_0, y_0, f(x_0, y_0))$):



Example 7.3: Tangent Vectors

Another way to understand this is to compute the *tangent vectors* at a point. Take the same function $f(x,y) = -x^2 - y^2$ and compute

$$\frac{\partial}{\partial x}\begin{bmatrix} x \\ y \\ f(x,y) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -2x \end{bmatrix}$$
$$\frac{\partial}{\partial y}\begin{bmatrix} x \\ y \\ f(x,y) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2y \end{bmatrix}.$$

Then take the cross product of these two vectors to get the normal vector to the tangent plane

$$\begin{bmatrix} 1 \\ 0 \\ -2x \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ -2y \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \\ 1 \end{bmatrix}.$$

Pick a point (x_0, y_0) and the normal to tangent plane is given by

$$\mathbf{n} = \begin{bmatrix} 2x_0 \\ 2y_0 \\ 1 \end{bmatrix}.$$

Which leads us to the following equation of a plane (but not exactly the tangent plane)

$$2x_0 x + 2y_0 y + z = 0.$$

This plane is parallel to the tangent plane and is often a nicer tool.

Exercise 7.2. Take the two equations for the planes above and simplify each to having a zero right hand side. Then subtract each of these equations from each other and see what the difference is.

Part VI

8 The Jacobian of a Vector Field

Recall we can define a three dimensional vector field $\mathbf{v}(x,y,z)$ by

$$\mathbf{v}(x,y,z) = (v_1(x,y,z), v_2(x,y,z), v_3(x,y,z)) = \begin{bmatrix} v_1(x,y,z) \\ v_2(x,y,z) \\ v_3(x,y,z) \end{bmatrix}$$

where we call each v_1, v_2, and v_3 the component functions. Notice that each component function is a scalar function!

Since each component function is a scalar function, we know how to compute the derivative of each by computing the gradient. This gives us a way to then talk about the derivative of the vector field as a whole

Definition 8.1: Jacobian

The **Jacobian** of a vector field $\mathbf{v}(x, y, z)$ is a matrix

$$J(x, y, z) \coloneqq \begin{bmatrix} \nabla v_1^T \\ \nabla v_2^T \\ \nabla v_3^T \end{bmatrix},$$

where the gradients are transposed (the superscript T) so they are are written as row vectors and placed in a matrix. More specifically, we can write this matrix as

$$J(x, y, z) = \begin{bmatrix} \frac{\partial v_1}{\partial x} & \frac{\partial v_1}{\partial y} & \frac{\partial v_1}{\partial z} \\ \frac{\partial v_2}{\partial x} & \frac{\partial v_2}{\partial y} & \frac{\partial v_2}{\partial z} \\ \frac{\partial v_3}{\partial x} & \frac{\partial v_3}{\partial y} & \frac{\partial v_3}{\partial z} \end{bmatrix}.$$

The Jacobian contains a lot of information. Intuitively, it tells us how each component of the vector field changes in each direction.

Example 8.1: Computing the Jacobian, 1

Let us consider the vector field

$$\mathbf{v}(x, y, z) = (x^2 + y^2, z, x + y + z).$$

Then we can write

$$\begin{aligned} v_1(x, y, z) &= x^2 + y^2 \\ v_2(x, y, z) &= z \\ v_3(x, y, z) &= x + y + z. \end{aligned}$$

So we compute the gradients of each

$$\nabla v_1 = \begin{bmatrix} 2x \\ 2y \\ 0 \end{bmatrix}$$

$$\nabla v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\nabla v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

So then the Jacobian is

$$asJ(x, y, z) = \begin{bmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

A related and important quantity comes in the form of the determinant of the Jacobian. We've previously talked about the determinant of a matrix as telling us the (signed) scaling of volume of a linear transformation. That is, for example, if we have

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

then $\det(A) = 8$ and we say that volumes of parallelopipeds are increased by a factor of 8 in this case.

When you are given a vector field, you can think of different regions of space being stretched differently. This is why we have that the Jacobian is a matrix that depends on the position (x, y, z). In this case, the volumes that are being stretched are very very tiny parallelopipeds. You can think of cubes with side lengths dx, dy, and dz.

Example 8.2: Computing the Jacobian, 2

e found that given

$$\mathbf{v}(x, y, z) = (x^2 + y^2, z, x + y + z)$$

that

$$J(x, y, z) = \begin{bmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

As we can see, this matrix depends on position. Let's compute the determinant

$$|J(x, y, z)| = \det(J(x, y, z)) = 2(y - x).$$

Something weird seems to happen with $y = x$ as the determinant $|J(x, y, z)|$ will be zero in this case. Otherwise, things are fine.

9 Divergence and Curl

Part VI

Often we do not need to use the whole Jacobian. We will find it to be necessary for integration, however.

For analysis of vector fields, we often wish to break them into their smaller pieces. Fundamentally, we can break vector fields into two parts:

- Sources and sinks,
- Rotations.

Sources, Sinks, and Divergence Fields

With some vector fields, we can make the analogy that some quantity (think air or water) is being added or removed from the system. We call these *sources* and *sinks* respectively. We want to quantify how much of some quantity is being added. This quantity is called the *divergence*.

Recall, we can write

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix}.$$

If we are also given a vector field

$$\mathbf{v}(x,y,z) = \begin{bmatrix} v_1(x,y,z) \\ v_2(x,y,z) \\ v_3(x,y,z) \end{bmatrix},$$

we can compute the **divergence** of $\mathbf{v}$ by

$$\nabla \cdot \mathbf{v}(x,y,z) = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}.$$

Notice that this quantity is a scalar! This scalar value, the divergence, tells us how much the vector field is diverging at a point (x,y,z). In other words, it tells us how much of a quanitity is being added or removed there.

Example 9.1: Source Field

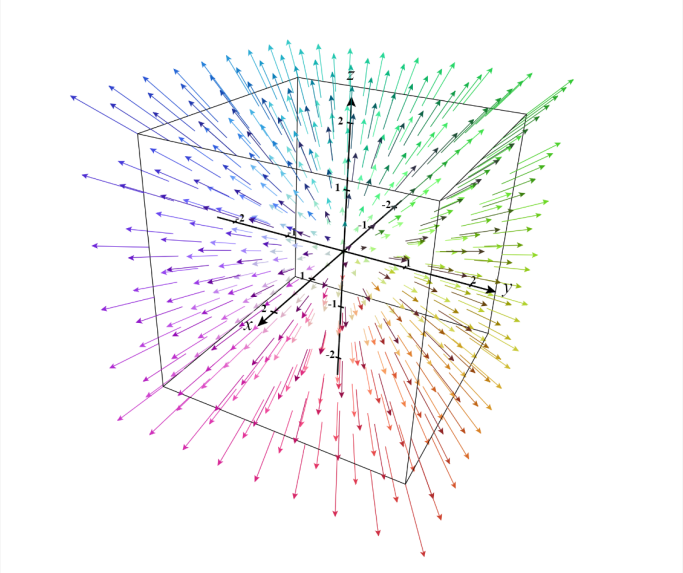
Consider

$$\mathbf{v}(x,y,z) = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Then the divergence is

$$\nabla \cdot \mathbf{v}(x,y,z) = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z} = 1 + 1 + 1 = 3.$$

We can think of a source of air being placed at each point that pumps in 3 units of air per second, or specifically being pumped in at the origin. The vector field looks like:



Rotation Fields

The divergence was the quantity that measured the outflow from a point for a vector field. The other quantity we can measure is the rotation of a vector field at a point.

We define the **curl** of a vector field $\mathbf{v}(x, y, z)$ to be

$$\nabla \times \mathbf{v}(x, y, z) = \begin{bmatrix} \frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z} \\ \frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x} \\ \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y} \end{bmatrix}.$$

Note that the curl is a vector! The curl is a vector that points orthogonally to the plane where rotation occurs and has magnitude relative to how quickly the field swirls.

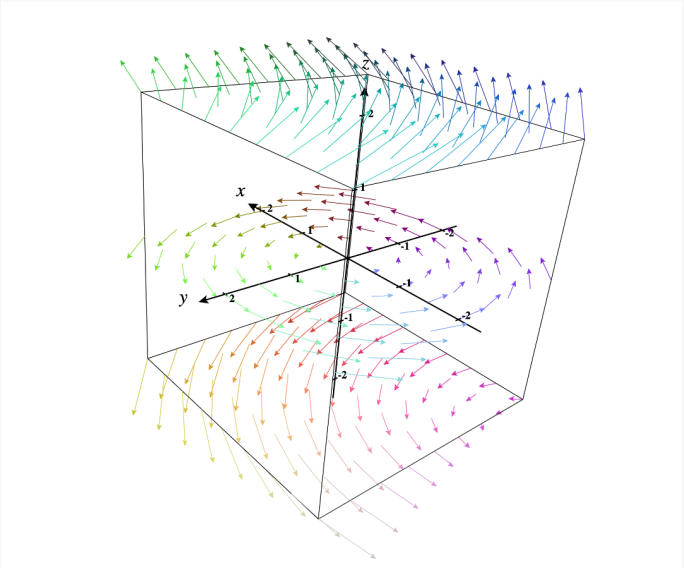
It's a bit involved to go through the work and see exactly why this is the correct quantity for seeing rotation of a vector field. However, you can recall that the cross product was useful in describing rotational motion of rigid bodies (that is, it showed up in angular velocity/momentum).

Example 9.2: A Rotation Field

Consider the vector field

$$\mathbf{v}(x, y, z) = (-y, x, z)$$

which looks like



We let

$$\begin{aligned} v_1(x, y, z) &= -y \\ v_2(x, y, z) &= x \\ v_3(x, y, z) &= z. \end{aligned}$$

If we look in this figure where $z = 0$, we can clearly see that this field swirls around the origin. If the curl is to measure rotation, we should see it nonzero here.

Let us compute the curl of this field. For this, we will all the other partial derivatives not contained in the divergence. That is, we need

$$\frac{\partial v_1}{\partial y} = -1 \qquad \frac{\partial v_1}{\partial z} = 0$$
$$\frac{\partial v_2}{\partial x} = 1 \qquad \frac{\partial v_2}{\partial z} = 0$$
$$\frac{\partial v_3}{\partial x} = 0 \qquad \frac{\partial v_3}{\partial y} = 0.$$

Then we have

$$\nabla \times \mathbf{v}(x,y,z) = \begin{bmatrix} 0-0 \\ 0-0 \\ 1-(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

We can decipher the meaning here by saying that the swirling occurs in planes parallel to the xy-plane since the direction of the curl is only in the z-direction. That is, curl is pointing perpendicularly to the plane of rotation. How quickly the field swirls is given by the magnitude of the curl which is 2 in this case. Using the right hand rule we discussed previously this tells us the direction of swirling as well. We have swirling counter-clockwise in the planes parallel to the xy-plane, and so we expect the curl to point in the positive z-direction.
Take a moment to analyze this using the figure provided or by plotting this yourself.

Example 9.3: Divergence of the Rotation Field

ne may also consider the divergent nature of the field

$$\mathbf{v}(x,y,z) = (-y,x,z)$$

from the previous example and find that

$$\nabla \cdot \mathbf{v}(x,y,z) = 1.$$

So, there is in some way divergence as well. This leads us to breaking the vector field into a part that swirls and a part that diverges as follows:

$$\mathbf{v}_{\text{swirl}}(x,y,z) = (-y,x,0)$$
$$\mathbf{v}_{\text{div}}(x,y,z) = (0,0,z).$$

This type of analysis can be very helpful when considering real world problems. It is especially important in electromagnetism.

Constant Vector Fields

Most of the understanding of vector fields was just covered by understanding the the part that diverges and the part that curls. However, you can always add constants to these vector fields and these constants will not change the divergence or curl. Why? Take the

following example.

Example 9.4: Constant Fields

Let

$$\mathbf{v}(x, y, z) = (c_1, c_2, c_3)$$

where c_1, c_2, and c_3 are constants. Then we can compute the Jacobian of $\mathbf{v}$

$$J(x, y, z) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the Jacobian holds all the partial derivative information, we can know from this that

$$\nabla \cdot \mathbf{v} = 0$$
$$\nabla \times \mathbf{v} = \mathbf{0}.$$

Exercise 9.1. Specifically show that the divergence and curl of a constant vector field (as in the previous example) are zero.

All of this is to say that aside from the addition of a constant vector field, we understand the behavior by looking at divergence and curl.

10 The Laplacian of a Scalar Field

In the study of partial differential equations (PDEs), we are often asked to find a function $u(x, y, z)$ that satisfies the following equation

$$\nabla \cdot \nabla u(x, y, z) = f(x, y, z)$$

for some given function $f(x, y, z)$. We will revisit this Part III, but for now we should see exactly what we mean by

$$\nabla \cdot \nabla u(x, y, z).$$

Exercise 10.1. Show that

$$\nabla \cdot \nabla u(x, y, z) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

Definition 10.1: Laplacian

We define the quantity

$$\nabla \cdot \nabla u(x, y, z)$$

to be the **Laplacian of** $u(x, y, z)$ and we often write

$$\Delta \coloneqq \nabla \cdot \nabla$$

and call this the **Laplacian operator**.

Intuitively, the Laplacian can be summed up in a few ways.

- The Laplacian is the *divergence* of the *gradient* of a scalar function.
- The Laplacian is the sum of "curvatures" in each direction.

Example 10.1: Computing the Laplacian

Let us consider the functions

$$f(x,y) = x^2 + y^2$$

and

$$g(x,y) = x^2 - y^2.$$

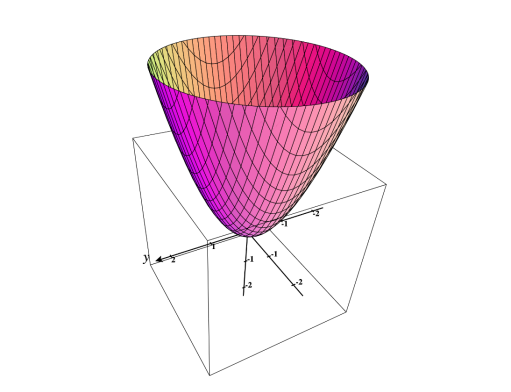
Then we can compute the gradients of each function to get

$$\nabla f(x,y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$
$$\nabla g(x,y) = \begin{bmatrix} 2x \\ -2y \end{bmatrix}.$$

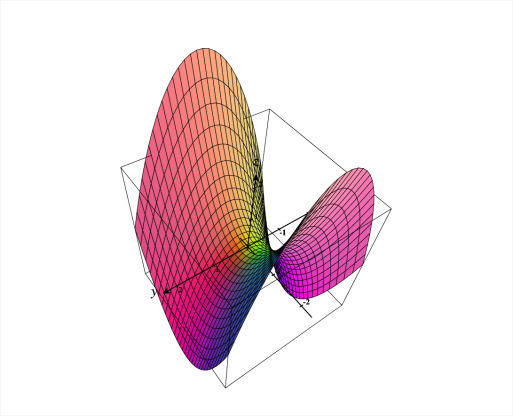
We can then compute the divergence of each of these and find

$$\nabla \cdot \nabla f(x,y) = 2 + 2 = 4$$
$$\nabla \cdot \nabla g(x,y) = 2 - 2 = 0.$$

Let us see what these two functions look like to get a bit of an intuitive feel.



(a) A plot of $f(x,y)$.



(b) A plot of $g(x,y)$.

Fundamentally we can see that these two functions are different. It seems that for $f(x,y)$ we are curving upward in both the x and y direction which is what allows the Laplacian to be positive. However, for $g(x,y)$ one direction curves the opposite direction as the other which cancels out and gives us that the Laplacian is zero.
It turns out that the Laplacian describes many phenomenon. Two examples would be soap films and temperature flow.

We have covered all of the differential calculus of multivariate functions that we need. That does not mean it won't show up again, but the material won't be new. We now move onto integration in multiple dimensions.

11 Integration of Scalar Fields

One Dimensional Case

In the case of one dimensional functions, we integrated to find the area under a curve. That is, we were given a function $f(x)$ and asked to find

$$\int_a^b f(x)dx$$

which gave us the *net* area under the curve. Let's briefly review this with an example.

Example 11.1: One-Dimensional Integral

et $f(x) = x^2 + 2$, $a = 1$, and $b = 2$. Then we want to find

$$\int_a^b f(x)dx = \int_1^2 x^2 + 2dx.$$

Then we use the *Fundamental Theorem of Calculus*. So, we find the antiderivative of the integrand and evaluate at the endpoints as follows

$$\begin{aligned}\int_1^2 x^2 + 2dx &= \left[\frac{x^3}{3} + 2x\right]_1^2 \\ &= \left(\frac{2^3}{3} + 2(2)\right) - \left(\frac{1^3}{3} + 2(1)\right) \\ &= \frac{13}{3}.\end{aligned}$$

Two Dimensional Case

Say we are now given a function $f(x, y)$ and bounds on both the x and y by $x_0 \leq x \leq x_1$ and $y_0 \leq y \leq y_1$. We then wish to evaluate

$$\int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y)dxdy.$$

You can think of this integral as being the *net* volume under the surface given by $f(x, y)$.

How do we compute such an integral? The answer is iteratively. Let's see how we do this with a concrete example.

Example 11.2: Two-Dimensional Integral

Let $f(x,y) = xy$, $x_0 = 1$, $x_1 = 2$, $y_0 = 3$ and $y_1 = 4$. So, we want to evaluate

$$\int_{y_0}^{y_1}\int_{x_0}^{x_1} f(x,y)dxdy = \int_3^4\int_1^2 xydxdy.$$

The way we do this is by first evaluating the integral with respect to x (holding y constant) and then integrate with respect to y (x will not appear here). So, we integrate from the inside out.
Let's start by integrating with respect to x. We take

$$\begin{aligned}\int_1^2 xydx &= \left[\frac{x^2y}{2}\right]_1^2 \\ &= \left(y\frac{2^2}{2}\right) - \left(y\frac{1^2}{2}\right) \\ &= \frac{3}{2}y.\end{aligned}$$

Now we take this function of y, and we integrate this with the bounds we are given.

$$\begin{aligned}\int_3^4 \frac{3}{2}ydy &= \left[\frac{3y^2}{4}\right]_3^4 \\ &= \frac{21}{4}.\end{aligned}$$

So we say that

$$\int_3^4\int_1^2 xydxdy = \frac{21}{4}.$$

Let's walk through the steps again. We did

$$\begin{aligned}\int_{y_0}^{y_1}\int_{x_0}^{x_1} f(x,y)dxdy &= \int_3^4\int_1^2 xydxdy \\ &= \int_3^4 \frac{3}{2}ydy \\ &= \frac{21}{4}.\end{aligned}$$

Three Dimensional Case

Integration here is performed in the same way. We are given a function $f(x,y,z)$ and bounds on x, y, and z such as $x_0 \le x \le x_1$, $y_0 \le y \le y_1$, and $z_0 \le z \le z_1$. Then we evaluate

$$\int_{z_0}^{z_1}\int_{y_0}^{y_1}\int_{x_0}^{x_1} f(x,y,z)dxdydz.$$

Let's work through an example.

Example 11.3: Three-Dimensional Integral

Let

$$f(x,y,z) = 2x + 8xyz + 3,$$

$x_0 = 0$, $x_1 = 1$, $y_0 = 2$, $y_1 = 3$, $z_0 = 4$, $z_1 = 5$. Then we want to find

$$\int_{z_0}^{z_1}\int_{y_0}^{y_1}\int_{x_0}^{x_1} f(x,y,z)dxdydz = \int_4^5\int_2^3\int_0^1 2x + 8xyz + 3dxdydz.$$

We do this iteratively. So we first evaluate the x integral holding the other variables constant for now.

$$\begin{aligned}\int_0^1 2x + 8xyz + 3dx &= \Big[x^2 + 4x^2yz + 3x\Big]_0^1 \\ &= \Big(1^2 + 4(1)^2yz + 3(1)\Big) - \Big(0^2 + 4(0)^2yz + 3(0)\Big) \\ &= 4yz + 4.\end{aligned}$$

We then take this, and integrate with respect to y.

$$\begin{aligned}\int_2^3 4yz + 4dy &= \Big[2y^2z + 4y\Big]_2^3 \\ &= (4(3)^2z + 4(3)) - (4(2)^2z + 4(2)) \\ &= 10z + 4.\end{aligned}$$

Lastly, we integrate with respect to z

$$\begin{aligned}\int_4^5 10z + 4dz &= \Big[5z^2 + 4z\Big]_4^5 \\ &= (5(5)^2 + 4(5)) - (5(4)^2 + 4(4)) \\ &= 49.\end{aligned}$$

So we say that

$$\int_4^5\int_2^3\int_0^1 2x + 8xyz + 3dxdydz = 49.$$

Again, let's walk through the steps a bit

$$\begin{aligned}\int_{z_0}^{z_1}\int_{y_0}^{y_1}\int_{x_0}^{x_1} f(x,y,z)dxdydz &= \int_4^5\int_2^3\int_0^1 2x + 8xyz + 3dxdydz \\ &= \int_4^5\int_3^4 4yz + 4dydz \\ &= \int_4^5 10z + 4dz \\ &= 49.\end{aligned}$$

12 Antiderivatives

In one variable calculus, we found that there is a relationship between the derivative and the indefinite integral. In fact, this led us to call the indefinite integral the antiderivative.

This relationship was that

$$\frac{d}{dx}\int f(x)dx = f(x)$$

and

$$\int \frac{df}{dx} = f(x) + C.$$

From this, we realized that the indefinite integral is almost an inverse operation of the derivative. It's just that in the case where we integrate a derivative, we only determine the function up to an additive constant.

Potential Functions

The higher dimensional analog happens to be a bit more nuanced but the idea remains the same. Let's say we are given a function $f(x, y, z)$ and we compute, for example,

$$\frac{\partial f}{\partial x}.$$

The issue now becomes this. Let's say that we let

$$f(x, y, z) = x + yz.$$

Then we have that

$$\frac{\partial f}{\partial x} = 1.$$

The terms with just a y, z dependence disappear. So if we were to try to undo this with an integral, we find that

$$\int \frac{\partial f}{\partial x}dx = \int 1dx = x + g(y, z).$$

That is to say, when we take an indefinite integral a multivariate function with respect to one variable, there could be a function of the residual variables that we cannot determine!

Integrating the Gradient

Let's say that we are given $\mathbf{v}(x, y, z) = \nabla f(x, y, z)$ and are asked to find the original function $f(x, y, z)$. This problem is called finding the **potential function** for $\mathbf{v}$. Remember that

$$\nabla f(x, y, z) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix}.$$

What we do is the following.

1. We integrate $\frac{\partial f}{\partial x}$ with respect to x and determine $f(x, y, z)$ up to adding a function of only y and z. That is we are able to recover what is essentially a third of the potential function $f(x, y, z)$.

2. We integrate $\frac{\partial f}{\partial y}$ with respect to y and determine $f(x, y, z)$ up to adding a function of only x and z.

3. We integrate $\frac{\partial f}{\partial z}$ with respect to z and determine $f(x, y, z)$ up to adding a function of only x and z.

4. Combine our knowledge from those three integrals we have determined $f(x, y, z)$ up to some additive constant!

Let's work through an example.

Example 12.1: Finding a Potential Function

Let's say that we are given the gradient of some function

$$\nabla f(x, y, z) = \begin{bmatrix} y + z \\ x + z \\ x + y \end{bmatrix}.$$

Then we follow the steps above

1. We integrate $\frac{\partial f}{\partial x}$ with respect to x. So we have

$$\int y + z dx = xy + xz + g(y, z).$$

   Here, $g(y, z)$ is a function of just y and z that we cannot determine yet.

2. We integrate $\frac{\partial f}{\partial y}$ with respect to y. So we have

$$\int x + z dy = xy + yz + h(x, z).$$

3. We integrate $\frac{\partial f}{\partial z}$ with respect to z. So we have

$$\int x + y dz = xz + yz + r(x, y).$$

4. Now we know that all of these functions should be equal (up to a constant). That is

$$xy + xz + g(y, z) = xy + yz + h(x, z) = xz + yz + r(x, y).$$

   Here, we can see that $g(y, z) = yz$, $h(x, z) = xz$, and $r(x, y) = xy$. So we have found that

$$f(x, y, z) = xy + xz + yz + C$$

   where the additive constant is there and is not something we can determine without a bit more information.

Requirements for Potentials

The main result here is the following.

Proposition 12.1: Curl of Gradient is Zero

We have that

$$\nabla \times \nabla f(x, y, z) = \mathbf{0}$$

for all $f(x, y, z)$.

Then with a bit more work, one can show this follows.

Theorem 12.1: Potential $\iff$ Curl Free

Let $\mathbf{v}(x, y, z)$ *be a vector field. Then if*

$$\nabla \times \mathbf{v} = \mathbf{0},$$

then $\mathbf{v}(x, y, z) = \nabla f(x, y, z)$. *That is, a curl-free vector field* $\mathbf{v}$ *is really the gradient of some scalar function* $f(x, y, z)$. *We call this scalar function* f *the* ***potential*** *for* $\mathbf{v}$.

This is what allows one to define the *voltage* in electrostatics. When charges are not moving, we have that the electric field $\mathbf{E}$ satisfies

$$\nabla \times \mathbf{E} = \mathbf{0}$$

and so it follows that

$$\nabla V(x, y, z) = \mathbf{E}$$

where V is the potential. We often call this electrostatic potential the voltage.