Math Boot Camp

U.C. Berkeley, Department of Chemistry

Sonja Bumann, Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme, Aditya Singh, and Harrison Tuckman

Last Updated July 2022

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0.
This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

Contents

1 Multivariable calculus

Topics: partial derivatives, chain rule, total derivatives, change of coordinates

1.1 Motivation and approach

You are already familiar with the workhorses of calculus, derivation and integration. Skill at multivariable calculus is less about learning new mathematics and more about learning how to apply the tools of single variable calculus to a multivariable setting. Calculus breaks down difficult problems into infinitely many tiny pieces, solves the pieces, and reassembles them into a solution. Multivariable calculus, in analogy with calculus itself, breaks down problems into single-variable components, solves them, and then reassembles them into an answer. The main challenge is not mechanical, as the mechanics are familiar, but conceptual, because multivariable functions are much trickier to visualize and interpret than single variable functions. Understanding multivariable functions begins with visualizing them.

1.2 Visualizing multivariable functions

Conceptual article: Multidimensional Graphs by Khan Academy

Conceptual article: Transformations by Khan Academy

Graphs are a common way to visualize functions. To create them, we associate each point in the domain (the input space) with a height (the corresponding value in the output space).

Transformations are another way of visualizing the operation of functions. Instead of spatially encoding the relationship between the input and output spaces, like in a graph, we visualize both spaces separately and imagine the relation between them. Transformations can be a particularly useful way of thinking about linear algebra.

These visualization techniques are fuel for intuition. At more than three dimensions, full visualization becomes impossible. However, we can still appeal to our low-dimensional intuition to gain insight. We will regularly need this instinct when working in higher-dimensional spaces.

A note on rigor: we will presume that all functions are continuous and differentiable. Some statements require other unlisted assumptions to be mathematically rigorous. This is a crash course, not a math course.

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

1.3 Partial derivatives

The principles of multivariable and single variable derivatives are the same. Both cases are conceptually centered on finding a **tangent** for our non-linear function. In general, multivariable functions have tangent surfaces with more than one dimension, i.e. there are several different directions in which to calculate the derivative. Fortunately, we can simplify the problem and continue using the tools of single variable calculus.

1.3.1 Mechanics

Technical video: Introduction to Partial Derivatives by Khan Academy (10:55)

Technical video: (optional) Formal Definition of Partial Derivatives by Khan Academy (7:58)

Let's break down the multivariable derivative into a series of one dimensional derivatives. Rather than dealing with a multivariable function $f(x, y, z)$ in its gory detail, we will instead hold all but one variable constant and take the derivative with respect to that variable. For example, say

$$f(x, y, z) = x \tan\big(yz^2\big)e^{(y+z)} \tag{1.1}$$

to take the derivative with respect to x, we fix y and z as constants $y = y_0$ and $z = z_0$, and take the normal single variable derivative. So

$$f(x, y, z) \to f(x, y_0, z_0) \to f(x) = x \tan\big(y_0 z_0^2\big)e^{(y_0+z_0)}. \tag{1.2}$$

Which is a single variable function that is essentially x multiplied by a constant, $f(x) = xc$, where $c = \tan(y_0 z_0^2)e^{(y_0+z_0)}$. Taking the derivative is now much easier,

$$f'(x, y_0, z_0) = \tan\big(y_0 z_0^2\big)e^{(y_0+z_0)}. \tag{1.3}$$

1.3.2 Geometry

Conceptual video: Graphical View of Partial Derivatives by Khan Academy (6:55)

Geometrically, this can be interpreted as taking the derivative in the planar slice of the function where $y = y_0$ and $z = z_0$. In essence, it is a single variable derivative in disguise, referred to as a **partial derivatives**. They are distinguished from true single variable derivatives by the ∂ (del) notation. So, the partial derivative of the function $f(x, y, x)$ with respect to x is $\left(\frac{\partial f}{\partial x}\right)_{y,z}$, where the y and z subscripts indicate which variables are held constant. Indicating the constant variables may seem trivial in the current context, but it becomes important when more variables are present (such as in thermodynamics).

For the function

$$f(x, y) = y^2 e^{xy} \tag{1.4}$$

the partial derivatives would be

$$\left(\frac{\partial f}{\partial x}\right)_y = y^3 e^{xy}, \quad \left(\frac{\partial f}{\partial y}\right)_x = 2ye^{xy} + xy^2 e^{xy}. \tag{1.5}$$

Example 1.1: The volume of a rectangular box of width x, height y, and depth z is given by $V = xyz$.

1. Calculate $\frac{\partial V}{\partial x}$, $\frac{\partial V}{\partial y}$, and $\frac{\partial V}{\partial z}$
2. Note that each partial derivative gives the area of a side of the box. Draw a picture illustrating why the rate of change of volume with respect to x is equal to the area of one side of the box.

1.3.3 Symmetry

Technical video: Symmetry of Second Partial Derivatives by Khan Academy (7:02)

Multiple partial derivatives are a natural extension of higher degree single variables derivatives. A second degree partial derivative with respect to x encodes the curvature of the function in the x-direction. Mixed partials derivatives, where multiple partial derivatives are taken with respect to different variables, encode the change in the directional slope with respect to changes in the other direction.

$$\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)_y\right)_y = \left(\frac{\partial^2 f}{\partial x^2}\right)_y = \text{ curvature in x-direction.} \tag{1.6}$$

$$\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)_x\right)_y = \frac{\partial^2 f}{\partial x \partial y} = \text{ change in y-direction slope w.r.t. change in } x\text{-direction.} \tag{1.7}$$

The value of a mixed partial derivative is not dependent on the order of differentiation. This is referred to as the equality or symmetry of crossed partial derivatives, as shown below

$$\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)_y\right)_x = \left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)_x\right)_y = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}. \tag{1.8}$$

Notation:

Another common notation for partials differentials uses subscripts to indicate which partial derivatives have been taken, so if we have $f(x, y)$:

$$\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)_y\right)_x = f_{xy}, \left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)_y\right)_y = f_{xx}, \left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)_x\right)_x\right)_x = f_{yyy}. \tag{1.9}$$

1.3.4 A challenge

Now consider a more difficult situation. Say that still $f(x, y) = y^2 e^{xy}$, but now x and y themselves are functions, $x = x(u) = \cos u$ and $y = y(u) = \sin u$. We want to find $\frac{df}{du}$ (note

that this is not a partial derivative, why?). We can solve this in an ugly way by rewriting the whole equation in terms of u and taking the derivative

$$f(x(u), y(u)) = f(u) = \left(\sin^2 u \, e^{\sin u \cos u}\right). \tag{1.10}$$

After liberal use of the chain and product rules, an answer appears

$$\frac{\partial f}{\partial u} = 2\cos u \sin u \, e^{\sin u \cos u} + \sin^2 u \left(\cos^2 u - \sin^2 u\right) e^{\sin u \cos u}. \tag{1.11}$$

Fortunately, there is a better way to find $\frac{df}{du}$, the multivariable chain rule, but it will require a deeper dive into multivariable calculus.

1.4 Total differentials

The total differential is the full generalization of the derivative to a multivariable function.

1.4.1 Linear approximation

Conceptual video: What is a Tangent Plane by Khan Academy (3:19)
Conceptual video: Local Linearization by Khan Academy (9:12)

The total derivative of a multivariable function is the best linear approximation. For a multivariable function with two independent variables, this is a tangent plane. With more variables, it is a higher dimensional tangent surface. The general equation of a plane through (x_0, y_0, f_0) is:

$$0 = A(x - x_0) + B(y - y_0) + C(f - f_0) \tag{1.12}$$

dividing across by C, this yields

$$f - f_0 = \frac{A}{C}(x - x_0) + \frac{B}{C}(y - y_0). \tag{1.13}$$

where A/C and B/C are the changes in z with respect to changes in x and y. In other words, A/C and B/C are the slopes in the x and y directions (partial derivatives!). So to find a tangent plane through a point, we replace A/C and B/C with the partial derivatives with respect to x and y. Replacing $\Delta f = f - f_0$, $\Delta x = x - x_0$, $\Delta y = y - y_0$, and setting $A/C = \frac{\partial f}{\partial x}$, $B/C = \frac{\partial f}{\partial y}$ the tangent plane through a point becomes

$$f - f_0 = \left(\frac{\partial f}{\partial x}\right)_y \Delta x + \left(\frac{\partial f}{\partial y}\right)_x \Delta y. \tag{1.14}$$

1.4.2 The tangent plane

Technical video: (optional) Total differentials by MIT OpenCourseWare (21:00)

The **tangent plane** is a local approximation of the function, specifically, it is the best **linear** approximation. To approximate value of a non-linear function $f(x, y)$ about some

point (x_0, y_0), we evaluate the partial derivatives at (x_0, y_0) to find the slopes and use the equation above

$$\Delta f \approx \left(\frac{\partial f}{\partial x}\right)_y \Delta x + \left(\frac{\partial f}{\partial y}\right)_x \Delta y \tag{1.15}$$

$$\text{height of function w.r.t. } (x_0, y_0) \approx \text{height of plane w.r.t } (x_0, y_0). \tag{1.16}$$

Here, height references to the graph of the function, where the value of the function is represented by the height in the z axis.

As $\lim_{\Delta x, \Delta y \to 0}$, our approximation becomes exact. At this limit, we use the notation of differentials, dx, dy, and dz, to yield the equation for the **total differential** of a multivariable function

$$df(x, y) = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy. \tag{1.17}$$

Differentials are weird! They are not numbers and they cannot usually be manipulated algebraically (sometimes they can, Khan Academy on Differentials). You can think of them as infinitely small changes in the value of a variable. While infinitely small values are not useful on their own, we can get useful values by dividing one differential by another, leading us to the chain rule.

1.5 The multivariable chain rule

Conceptual video: The Chain Rule by 3Blue1Brown (8:45 - 14:35)

In practice, taking derivatives of multivariable functions is messy. Often one function is a function of another function and trying to naively use partial derivatives is infeasible. The chain rule provides a clear strategy for dealing with these messy problems.

1.5.1 Mechanics

Technical video: Multivariable Chain Rule by Khan Academy (9:32)

Conceptual video: Multivariable Chain Rule Intuition by Khan Academy (7:46)

The **multivariable chain rule**, or **total differential** has close ties to the tangent plane approximation (as discussed above). It encodes how changes in the independent variables $x_1, x_2, \ldots, x_n$ affect the value of a function $f(x_1, x_2, \ldots, x_n)$. We can use the chain rule to solve more complex derivation problems. Formally, the chain rule for $f(x_1, x_2, \ldots, x_n)$ is

$$df(x_1, x_2, \ldots, x_n) = df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \ldots + \frac{\partial f}{\partial x_n} dx_n. \tag{1.18}$$

As an example, we can use the chain rule to solve the partial derivative problem from earlier: $f(x, y) = y^2 e^{xy}$, where $x = x(u) = \cos u$ and $y = y(u) = \sin u$. Again, we want to find $\frac{df}{du}$. We do this by starting with the chain rule expansion of the function

$$df(x, y) = df = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy \tag{1.19}$$

and dividing across by du to yield an expression for $\frac{df}{du}$:

$$\frac{df}{du} = \left(\frac{\partial f}{\partial x}\right)_y \frac{dx}{du} + \left(\frac{\partial f}{\partial y}\right)_x \frac{dy}{du} \tag{1.20}$$

This may seem mathematically dubious (it is), but it is technically valid. If you yearn for a better justification, watch Total differentials by MIT OpenCourseWare.

1.5.2 Example

Earlier we defined $x(u) = \cos(u)$ and $y(u) = \sin(u)$ (such that $x = sin$ and $y = cos$), to yield $\frac{dx}{du} = -\sin u$ and $\frac{dy}{du} = \cos u$. Further, we used partial differentiation to find expressions for $\left(\frac{\partial f}{\partial x}\right)_y$ and $\left(\frac{\partial f}{\partial y}\right)_x$. Let's use those same x and y and plug our solutions for the partial differentials into the equation above. This puts x and y in terms of u, we find

$$\frac{\partial f}{\partial u} = \left(2\sin u\, e^{\sin u \cos u}\right)(-\sin u) + \left(2\sin u\ e^{\sin u \cos u} + \sin^2 u\ \cos u\ e^{\sin u \cos u}\right)(\cos u) \tag{1.21}$$

and after a little rearranging, we get

$$\frac{\partial f}{\partial u} = 2\cos u \sin u\, e^{\sin u \cos u} + \sin^2 u\,(\cos^2 u - \sin^2 u)\, e^{\sin u \cos u} \tag{1.22}$$

which is the same result we got earlier! The chain rule provides a cleaner path to the same answer, and critically, it does not require us to know the full functional form of our quantities. If we can learn $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{dx}{du}$, and $\frac{dy}{du}$ without deriving them analytically (perhaps using experiment or simulation), then we can determine $\frac{\partial f}{\partial u}$.

Example 1.2: Given $z = z(x, y)$ and $y = y(x)$ use the chain rule to show

$$\frac{dz}{dx} = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \frac{dy}{dx}\,. \tag{1.23}$$

1.6 Integration techniques

Conceptual article: Double Integrals by Khan Academy

Single variable integrals sum the value of a function over a one dimensional slice of the number line. Multivariable integrals sum over a higher dimensional domain. The following double integral, for example, finds the value of a function, $f(x, y)$, at every point in an area A and sums them together. This is equivalent to finding the volume under the graph of the function $f(x, y)$. We write

$$\iint_A f(x, y)\, dA. \tag{1.24}$$

Notation

The differential dA is commonly placed either right after the integral, or at the end of the expression. Both notations are identical in meaning and both will be used in this text. So keep in mind that

$$\iint_A f(x,y)\,dA = \iint_A dA\,f(x,y). \tag{1.25}$$

1.6.1 Integrals over non-rectangular regions

Conceptual article: Double Integrals over non-rectangular Regions by Khan Academy

Conceptual article: Double Integrals Beyond Volume

Fubini's Theorem is a central tool of multivariable integration. It states that an integral over higher dimensional space can be dissolved into a series of one dimensional integrals, as long as the one-dimensional domains of integration reflect the original domain. Here, X, and Y represent the domain of the function in the x and y dimensions. Fubini's Theorem also states that we can rearrange the order of integration, so

$$\int_A f(x,y)\,dA = \int_X \int_Y f(x,y)\,dy\,dx = \int_Y \int_X f(x,y)\,dx\,dy. \tag{1.26}$$

This means that solving a multivariable integral is only as difficult as identifying the domains of integration in each dimension and evaluating the integral! (though this is often still really hard).

As an example, let's set up an integral of the function $f(x,y)$ over the region of space A such that $0 < x < 1$ and $0 < y < x^2$. (can you figure out what this region of the plane look like?) Filling in the bounds, we get

$$\int_A f(x,y)\,dA = \int_A f(x,y)\;dA = \int_X \int_Y f(x,y)\;dydx = \int_0^1 \int_0^{x^2} f(x,y)\;dydx. \tag{1.27}$$

Because we are not integrating over a rectangular region, the bounds of the inner integral over y must adapt to the position of the outer integral over x. Once the integral is properly set up, it can be integrated as normal with sequential single variable integrations.

1.6.2 Coordinate transformations and spherical coordinates

Conceptual article: Triple Integrals by Khan Academy

Conceptual article: Spherical Coordinates by Khan Academy

Sometimes, Cartesian coordinates are not suited to the problem at hand. For example, say we want to integrate a function $f(x,y,z)$ over the a sphere of radius R. The integral would be set up as

$$\int_{-R}^{R} \int_{-\sqrt{R^2-z^2}}^{\sqrt{R^2-z^2}} \int_{-\sqrt{R^2-y^2-z^2}}^{\sqrt{R^2-y^2-z^2}} f(x,y,z)\;dx\,dy\,dz. \tag{1.28}$$

Terrible! To reflect an inherently non-rectangular domain, the bounds of integration must become incredibly convoluted. Instead, spherical coordinates represent the domain much more naturally. The same integral set up in spherical coordinates is much simpler!

$$\int_0^R \int_0^\pi \int_0^{2\pi} f(r,\phi,\theta)\, r^2 \sin\phi \; d\theta \, d\phi \, dr \tag{1.29}$$

To transform to spherical coordinates, we must change the variables in the function by substituting $x = r\sin\phi\cos\theta$, $y = r\sin\phi\cos\theta$, $z = r\cos\phi$ and replacing the integrand $dx\,dy\,dz \to r^2\sin\phi\; dr\, d\phi\, d\theta$. The first page of this MIT OPC handout has an explanation of where those substitutions come from Limits in Spherical Coordinates. The logic behind the factor of $r^2\sin\theta$ is more complex.

The objective of a change of coordinates is to choose a transformation so that the domain of integration becomes rectangular in the transformed space. Then the integral can be completed without convoluted boundaries.

1.6.3 The Jacobian

Conceptual video: Local Linearity for a Multivariable Function by Khan Academy (5:23)

Technical video: The Jacobian Matrix by Khan Academy (6:21)

Technical video: The Jacobian Determinant by Khan Academy (8:52)

> *Note:*
>
> This section is extremely terse, it is included for completion but is not necessary for CHEM 220A/221A.

The Jacobian exists for multivariable functions that map from n-dimensions to n-dimensions, known as transformations. For example, the spherical coordinate transformation is a transformation because it rescales 3-dimensional space and does not alter the dimensionality. The Jacobian of a transformation is its best linear approximation, it encodes how the transformation distorts space about one point. It is constructed from the partial derivatives of the transformation and its determinant encodes how the volume of space changes when the transformation is applied.

The Jacobian is used to scale the differential when changing variables. In direct substitution, which is a single variable change of variables, we apply a transformation $u = f(x)$ and use the derivative to substitute in a new differential $du = f'(x)dx$ to scale the new domain of integration. The determinant of the Jacobian is the multivariable equivalent, it scales the domain of integration so that the magnitude of the integral is correct. If T is the change of variables transformation and $\mathbf{J}T(\mathbf{x})$ is the Jacobian of our function at some point $\mathbf{x}$, then our differential is scaled by $dx_1...dx_n \to \det\left[\mathbf{J}T(\mathbf{x})\right] dy_1...dy_n$ where $x_1...x_n$ are the original coordinates and $y_1...y_n$ are the new coordinates. This may look familiar, recall $dx\,dy\,dz \to r^2\sin\phi\, dr\, d\phi\, d\theta$. Indeed, $r^2\sin\phi$ is just the determinant of the Jacobian of the spherical coordinate transformation!

1.6.4 Integration tricks

Solving integrals is a matter of knowing which trick to apply and when. Four particularly common tricks are briefly summarized: substitution, integration by parts, completing the square, and integration of hyperbolic functions. An integral lookup table is also included at the end, for your convenience.

- **Substitution** involves transforming the variable of integration into a more conveinient form. Substitution can be broadly used to simplify integrals. See: u-Substitution by Khan Academy. There are four steps:

  1. Identifying a useful substitute
  2. Adjusting the domain of integration
  3. Adjusting the differential
  4. Evaluating the integral

- **Integration by Parts** clevery rearranges the integral to make it more soluble. Most useful if the integrand is the product of two easily differentiable functions. The mechanics are described in Integration by Parts by Khan Academy. The integration by parts formula is shown below, (for a proof, see Proof of Integration by Parts)

$$\int udv = uv - \int vdu. \tag{1.30}$$

  Integration by parts has three steps, sometimes you will need to repeat the process multiple times.

  1. Identifying a useful u and v
  2. Rewriting the integral
  3. Evaluating the integral

- **Completing the Square** massages a polynomial into a more convenient form to ease integration. The term "completing the square" draws from a neat geometric analogy (see "Geometric Perspective" on the Wikipedia entry). This technique is useful when there is an "incomplete square" of the form $ax^2 + bx + c$. See Completing the Square by Khan Academy.

- **Integrands that are Derivatives of Hyperbolic Functions**. Noticing an integrand that is the derivative of a hyperbolic function is more art than science. Nonetheless, it is sometimes the only way forward. This can be particularly useful in conjunction with completing the square, see the completing the square link above for an example. There are many websites that list the derivatives of hyperbolic functions (or you can use Wolfram Alpha).

Example 1.3: Solve $\int e^x \cos x \, dx$.

1.7 Connections to physical chemistry

Calculus is ubiquitous in the physical sciences, physical chemistry being no exception. It's not exaggeration to say that you can't do physical chemistry without calculus. What can be difficult, however, is remembering that our mathematical intuition can drive our understanding of chemistry. Take the free energy equation from thermodynamics

$$dE = TdS - pdV + \mu dN. \tag{1.31}$$

Temperature, pressure, and chemical potential have physical meaning, we have intuition about their behavior and we can physically understand their role in free energy. Still, understanding their mathematical structure can yield deeper insight. By rewriting the free energy equation we can see it is really a tangent plane approximation!

$$dE = \left(\frac{\partial E}{\partial S}\right)_{V,N} dS + \left(\frac{\partial E}{\partial V}\right)_{S,N} dV + \left(\frac{\partial E}{\partial N}\right)_{S,N} dN \tag{1.32}$$

Temperature, pressure, and chemical potential are the partial derivatives of free energy. Physically interpretable constants transform into mathematical and geometric structures. By thinking about these familiar physical constants mathematically, we can intuit new facets of their behavior. Mathematics can add structure to the physical world.

1.8 Additional resources

Mary Boas, Mathematical Methods in the Physical Science has a excellent discussion of partial derivatives, the chain rule, and total differentials on pages 188-202.

1.9 Example problem solutions

Example 1.1: *The volume of a rectangular box of width x, height y, and depth z is given by $V = xyz$.*

1. *Calculate $\frac{\partial V}{\partial x}$, $\frac{\partial V}{\partial y}$, and $\frac{V}{\partial z}$*
2. *Note that each partial derivative gives the area of a side of the box. Draw a picture illustrating why the rate of change of volume with respect to x is equal to the area of one side of the box.*

1. $\frac{\partial V}{\partial x} = yz$, $\frac{\partial V}{\partial y} = xz$, $\frac{\partial V}{\partial z} = xy$
2. The partial derivatives are the areas of the opposing sides of the box.

Example 1.2: *Given $z = z(x, y)$ and $y = y(x)$ use the chain rule to show*

$$\frac{dz}{dx} = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \frac{dy}{dx}\,.$$

First, write down the total derivative for z.

$$dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy\,. \tag{1.33}$$

Then, divide across by the differential dz

$$\frac{dz}{dx} = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x dy\,, \text{ because } \frac{dx}{dx} = 1\,. \tag{1.34}$$

Example 1.3: *Solve $\int e^x \cos x\, dx$.*

Use integration by parts with $u = e^x$, $dv = \cos x$ to yield

$$\int e^x \cos x\, dx = e^x \sin x - \int e^x \sin x\, dx = e^x \sin x + e^x \cos x - \int e^x \cos x\, dx. \tag{1.35}$$

Then add $\int e^x \cos x\, dx$ to both sides

$$2\int e^x \cos x\, dx = e^x \sin x + e^x \cos x + c. \tag{1.36}$$

2 Probability and statistics

Topics: Random variables, probability distributions, mean, variance, counting, Gaussian distribution, delta function

2.1 Motivation and approach

The field of statistical mechanics, which connects the microscopic description of a system to macroscopic observables, is intricately tied with probability and statistics. When describing a large number (think Avogadro's number) of particles, it's not important (and too difficult) to know the details of each particle, so we often think of them as a statistical ensemble, or a probability distribution, over all possible states of the system (*i.e.* "microstates"). The mean and variance of this distribution correspond to observables like total energy, volume, temperature, or pressure.

Quantum mechanics also takes on a probabilistic view – predictions made by quantum mechanics are statistical by nature. Without getting too philosophical, observed quantities in quantum mechanics can be interpreted as the result of measuring an ensemble of infinite, identically-prepared systems.

Finally, statistics is important in analyzing and understanding any experimental data. The world around us is made up of random variables, and any time we measure these random variables or make predictions about how they will behave in the future, we are making use of probability and statistics.

This section describes random variables and probability distributions, providing formulae for computing means and variances, which are measures of probability distributions. It also introduces two distributions, the Gaussian distribution and delta function, which are commonly encountered in physics, and basic concepts of counting and combinatorics.

2.2 Random variables and probability distributions

Conceptual video: Random variables by Khan Academy (4:31)

Conceptual video: Density curves by Khan Academy (9:33)

A **random variable** is a variable whose values depend on the outcome of random events. The **probability distribution** of a random variable is the function that gives the probabilities that possible outcomes actually occur. Random variables can either be discrete (*i.e.* the result of a coin toss) or continuous (*i.e.* the amount of time you end up waiting for the bus).

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

If $p(x)$ is the **probability density function** of a continuous random variable x, then $p(x) \geq 0$ at all points x, as we know there cannot be negative probabilities! The probability that x is between the values a and b is given by

$$p(a < x < b) = \int_a^b p(x)dx\,, \tag{2.1}$$

and this probability should always be between 0 and 1.

Finally, all probability distributions should be **normalized**,

$$\int_{-\infty}^{\infty} p(x)dx = 1\,, \tag{2.2}$$

since we know that the probability of x being any value between negative and positive infinity should be 100%.

For a discrete random variable with **probability mass function** P, the probability that x is equal to a is given by $P(x = a)$ and the normalization condition is given by $\sum_i P(x_i) = 1$.

2.2.1 Mean and variance

Conceptual and technical video: Deriving the mean and variance of a continuous probability distribution by JB Statistics (7:21)

The **mean** or **expected value** of a random variable X is a value that captures the center of the random variable's distribution. If you perform many random events that come from the same probability distribution, the average value of all the outcomes is represented by the mean.

The expectation value $E[X]$ is often denoted as μ and can be computed as

$$\mu = E[X] = \int xp(x)dx \tag{2.3}$$

$$\mu = E[X] = \sum_i x_i P(x_i) \tag{2.4}$$

for continuous and discrete random variables, respectively.

The expectation is a linear operator, so

$$E[aX + b] = aE[X] + b\,. \tag{2.5}$$

The expected value of a constant is itself (*i.e.* $E[b] = b$) since the constant is not random and its value does not vary.

The **variance** measures the spread of a distribution, or the magnitude of fluctuations that occur around the mean. The variance is often denoted as σ^2 and can be computed as

$$\sigma^2 = \mathrm{Var}[X] = E\big[(X - \mu)^2\big] = E[X^2] - \mu^2\,. \tag{2.6}$$

For continuous and discrete random variables, this becomes

$$\sigma^2 = \int x^2 p(x) dx - \left(\int x p(x) dx \right)^2 \tag{2.7}$$

$$\sigma^2 = \sum_i x_i^2 P(x_i) - \left(\sum_i x_i P(x_i) \right)^2 . \tag{2.8}$$

Note that, unlike the expectation, the variance is not a linear operator. It does, however, follow the general rule:

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X] . \tag{2.9}$$

The square root of the variance is called the **standard deviation** and is often denoted as σ.

Example 2.1: Compute the mean and variance of a random variable X that is distributed according to the Bernoulli distribution, which is a discrete probability distribution with

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} , \tag{2.10}$$

where p is a parameter.

2.2.2 Multivariate distributions

A **multivariate distribution** or **joint distribution** is a distribution of multiple random variables. The measure of **correlation** between two random variables X and Y is given by

$$\sigma_{XY} = \langle XY \rangle - \langle X \rangle \langle Y \rangle . \tag{2.11}$$

Two random variables are said to be uncorrelated if $\langle XY \rangle = \langle X \rangle \langle Y \rangle$ so that $\sigma_{XY} = 0$. In this case, the joint probability distribution is simply a product of each probability distribution:

$$p(x, y) = p(x) p(y) . \tag{2.12}$$

In physics, we see that two things separated by shorter distances or timescales tend to be more correlated than things that are far away from each other in space or time.

To obtain the **marginal distribution** from the joint distribution, simply integrate over one of the random variables:

$$p(x) = \int p(x, y) dy . \tag{2.13}$$

2.3 Special distributions

2.3.1 Gaussian distribution

Conceptual video: Sampling distribution of the sample mean by Khan Academy (10:51)

Technical video: Standard error of the mean by Khan Academy (15:14)

Conceptual and technical video: Central limit theorem by Khan Academy (9:48)

The **Gaussian distribution** (*i.e.* **normal distribution** or **bell curve**) is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{2.14}$$

Why does the Gaussian distribution come up literally everywhere? The answer is the **central limit theorem**.

Often, we have a random variable for which we don't know the probability distribution, but we want to use data to estimate its mean and variance.

For example, we can conduct an experiment n times, the results of which are independent and identically distributed random variables: $X_1, X_2, \ldots, X_n$. From there, we can compute the **sample mean** as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{2.15}$$

The sample mean $\bar{X}_n$ is itself a random variable that has $E(\bar{X}_n) = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, where μ and σ^2 are the mean and variance of the underlying distribution from which the data X_i were drawn.

When plotting data corresponding to experiments we've repeated multiple times under the same conditions, we tend to use a point (the sample mean) and errorbars (the sample variance), which highlight both the result of our experiments and the uncertainty in our measurements.

According to the **law of large numbers**, the sample mean $\bar{X}_n$ converges to the true mean μ as $n \to \infty$. What's more, the central limit theorem says that for large n, the distribution of $\bar{X}_n$ approaches a Gaussian distribution with μ and variance σ^2/n, regardless of the underlying distribution for each of the X_i.

Conceptual and technical video: Why does pi show up here? The Gaussian integral, explained by vcubingx (5:45)

The Gaussian distribution can also be very helpful for integration. If you have a function of the form e^{x^2}, you know that it can be manipulated to match the exponential in Eq. (2.14) and then use the property of normalization to solve the integral.

The definite integral of any arbitrary Gaussian function is

$$\int_{-}^{\infty\,\infty} e^{-a(x+b)^2} dx = \sqrt{\frac{\pi}{a}}\,. \tag{2.16}$$

Example 2.2: Evaluate the following integral:

$$I = \int_0^\infty e^{-x^2/\sigma^2} dx\,. \tag{2.17}$$

Wikipedia has a very useful list of Gaussian integrals that is good to keep handy.

2.3.2 Delta function

Conceptual video: Dirac delta function by Khan Academy (17:47)

The delta function is a weird "function." Consider first the **Kronecker delta**:

$$\delta_{nm} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases} . \tag{2.18}$$

This is normalized (*i.e.* $\sum_m \delta_{nm} = 1$), so it can be thought of as a discrete probability distribution that picks out a particular value.

The continuous version of this distribution is the **Dirac delta function**:

$$\delta(x - x') = \begin{cases} \infty & \text{if } x = x' \\ 0 & \text{otherwise} \end{cases} . \tag{2.19}$$

The Dirac delta function is also normalized ($\int_{-\infty}^{\infty} \delta(x - x')dx = 1$), and it can also be used to pick out particular values:

$$\int f(x)\delta(x - x')dx = f(x') . \tag{2.20}$$

For this reason, the Dirac delta function is commonly used in physics to impose certain constraints, like the position of a particle or the total energy of a system.

The Dirac delta function comes up in other contexts, such as a Gaussian function with the variance tending to 0, in the Fourier transform of constants, and as the derivative of the Heaviside function.

2.4 Counting

Conceptual video: Introduction to combinations by Khan Academy (6:17)

For an experiment in which all outcomes are equally likely, the probability of an event occurring is given by

$$p = \frac{\text{number of outcomes satisfying event}}{\text{number of all possible outcomes}} . \tag{2.21}$$

Example 2.3: When considering a fair six-sided die, what is the probability of rolling an even number?

Sometimes, determining the numerator and denominator in the above probability expression can be challenging.

Often, we will need to figure out how many ways there are to order n objects. There are n ways to choose the first object, $(n - 1)$ ways to choose the second object, ..., and 1 way

to choose the last object. We can multiply all of these ways to obtain $n \cdot (n-1) \cdot \cdots \cdot 1 = n!$ ("n factorial") ways.

Other times, we will need to figure out how many ways there are to pick k objects from n total objects. If order matters, there are n ways to choose the first object, $(n-1)$ ways to choose the second object, ..., and $n-(k-1)$ ways to choose the k^{th} and final object. This means there are $n \cdot (n-1) \cdot \cdots \cdot n-(k-1) = n!/(n-k)!$ ways.

If order doesn't matter, then we can divide the above expression by the number of ways to choose those k objects ($k!$ ways) so we have $\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{(n-k)!k!}$ ("n choose k") ways.

Example 2.4: How many ways can n distinguishable particles be placed in k distinguishable boxes? Assume there are no restrictions on the number of particles that can be in each box.

2.5 Connections to physical chemistry

As noted in the Introduction and Motivation, both the fields of statistical mechanics and quantum mechanics are probabilistic in nature.

Counting and combinatorics can be helpful in determining the number of microstates of a system, *i.e.* the number of configurations of an ensemble of particles.

The probability that a macroscopic system will be in a certain microstate is given by a probability distribution. The specific probability distribution depends on the physical circumstances of the problem, but one of the most common is the Boltzmann probability distribution, for which the probability that the system is in microstate i with energy ε_i is given by $p_i \propto e^{-\beta\varepsilon_i}$, where β is related to the temperature of the system. This probability distribution can be used to calculate expectation values (the mean) and fluctuations (the variance) of observables like total energy, volume, temperature, or pressure.

In quantum mechanics, probability distributions appear frequently. One common example is that the square modulus of a particle's normalized wavefunction $|\psi(x)|^2$ gives the probability distribution of position measurements of that particle.

2.6 Additional text resources

Seeing Theory: A visual introduction to probability and statistics by Kunin, Guo, Devlin, and Xiang: Chapters 1 and 3

2.7 Example problem solutions

Example 2.1: *Compute the mean and variance of a random variable X that is distributed according to the Bernoulli distribution, which is a discrete probability distribution with*

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}, \tag{2.22}$$

where p is a parameter.

We can compute the mean directly via Eq. (2.4):

$$\begin{aligned} \mu &= \sum_i x_i P(x_i) && (2.23)\\ &= 0 \cdot P(0) + 1 \cdot P(1) && (2.24)\\ &= p\,. && (2.25) \end{aligned}$$

Similarly, we can use this result along with Eq. (2.8) to compute the variance:

$$\begin{aligned} \sigma^2 &= E[X^2] - \mu^2 && (2.26)\\ &= \sum_i x_i^2 P(x_i) - \mu^2 && (2.27)\\ &= 0^2 \cdot P(0) + 1^2 \cdot P(1) - \mu^2 && (2.28)\\ &= p - p^2\,. && (2.29) \end{aligned}$$

Example 2.2: *Evaluate the following integral:*

$$I = \int_0^\infty e^{-x^2/\sigma^2} dx\,. \tag{2.30}$$

First, it's important to notice that the integrand has the form of a Gaussian function centered at 0. Next, we notice that the limits of integration are from 0 to ∞. Since this Gaussian function is symmetric about its center at $x = 0$, this integral is half of that from $-\infty$ to ∞. Finally, we can use the result from Eq. (2.16) and divide by 2 to obtain

$$I = \frac{1}{2}\sqrt{\sigma^2 \pi}\,. \tag{2.31}$$

Example 2.3: *When considering a fair six-sided die, what is the probability of rolling an even number?*

Since all outcomes are equally likely, the probability of rolling an even number is the ratio of the number of outcomes resulting in an even roll to the number of all possible outcomes: $p = 3/6 = 1/2$.

Example 2.4: *How many ways can n distinguishable particles be placed in k distinguishable boxes? Assume there are no restrictions on the number of particles that can be in each box.*

First, we must ask ourselves if order matters. Since the particles are distinguishable, order does matter – we can tell which particles are in which boxes! The first particle can be arranged in k ways. Since there are no restrictions on the number of particles in each box, the second particle can also be arranged in k ways, and so on, for each of the n particles. Since each particle's box assignment is independent of all the others', we can multiply these n values to obtain k^n ways.

To consider further... Think about how this solution would be different if the particles were indistinguishable. In classical mechanics, identical particles are distinguishable, while they are indistinguishable in quantum mechanics. This difference has an implication in probabilities of events in classical versus quantum mechanics. Also, consider how this solution would be different if we put restrictions on the numbers of particles allowed in each box.

3 Linear Algebra 1: Vectors

Topics: Vector (sub)spaces, linear independence, projections, (orthonormal) basis decomposition, functions as basis vectors

3.1 Motivation and approach

Linear algebra can be thought of as the language of quantum mechanics, in some sense. While other formalisms can be used to express the definitions, axioms, and equations of quantum theory, the fundamental objects of linear algebra—vectors, matrices, etc.—are essential in understanding quantum mechanics for graduate coursework and research. This section reviews some of these basic ideas of linear algebra while also introducing notational and conceptual abstractions that are necessary to apply linear algebra to the study of quantum mechanics.

Linear algebra is often first introduced with a discussion of vectors, and we are taught to think of these objects as arrows that possess a magnitude and a direction. With this, we conceptually link vectors to objects we could encounter in our daily lives: a train moving a certain speed in a certain direction, the force acting on a ball that is falling to the ground, etc. While this approach is helpful in that it grounds the concept of a vector in physical reality, it doesn't provide as much clarity for other objects and operations that are necessary in more advanced applications of linear algebra.

Here, we're going to think about vectors in a slightly more abstract way that will hopefully make the subsequent topics of linear algebra more transparent and, later on, allow us to apply these ideas to think about more complicated systems, like those encountered in quantum mechanics. If you don't find this to be intuitive at first, that's okay! You can always return to an understanding or application of vectors that does make sense to you to work through it. The hope is that, with enough time, you'll begin to think about vectors in a more abstract way that will ultimately make some of the basic ideas of quantum mechanics make more sense and maybe even become intuitive. We'll first start with vectors and the build our way up.

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

3.2 Vectors and vector spaces

A note on notation:

Most undergraduate curricula use boldface ($\mathbf{v}$), arrows ($\vec{v}$), or underbars ($\underline{\mathrm{v}}$) to denote the vector v. We'll introduce Dirac notation and use what is called a "ket" to denote this vector, writing it as $|v\rangle$. At the beginning we'll use the notations $|v\rangle$ and $\mathbf{v}$ interchangeably. Gradually we will begin to use the $|\cdots\rangle$ notation exclusively.

3.2.1 Dirac Notation

Dirac notation (alternatively known as bra-ket notation) is utilized extensively in quantum mechanics due to its succinct yet informative form. If we have a vector v, we denote this vector in Dirac notation with the ket, $|v\rangle$, for example,

$$|v\rangle = \begin{bmatrix} v_x \\ v_y \end{bmatrix}, \qquad |v\rangle = f(x).$$

The bra of v, denoted by $\langle v|$, represents the complex conjugate of our vector v.

$$\langle v| = \begin{bmatrix} v_x^*, v_y^* \end{bmatrix}, \qquad \langle v| = f^*(x).$$

We'll talk about inner products in depth later on, but putting the bra and ket together, like $\langle v|v\rangle$, represents the inner product between the bra vector and ket vector, resulting in a scalar.

$$\langle v|v\rangle = \begin{bmatrix} v_x^* v_y^* \end{bmatrix} \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix}, \qquad \langle v|v\rangle = \int f^*(x) f(x) dx$$

To drive the point home, let's consider another example with vectors a and b from the Cartesian vector space. We denote our vectors as kets like

$$|a\rangle = \begin{bmatrix} a_x \\ a_y \end{bmatrix} \qquad |b\rangle = \begin{bmatrix} b_x \\ b_y \end{bmatrix}. \tag{3.1}$$

The complex conjugate of these vectors is represented with the bras as

$$\langle a| = \begin{bmatrix} a_x^* a_y^* \end{bmatrix} \qquad \langle b| = \begin{bmatrix} b_x^* b_y^* \end{bmatrix}. \tag{3.2}$$

Finally, the inner product between vectors a and b is represented with a bra-ket pair, and results in a scalar like so:

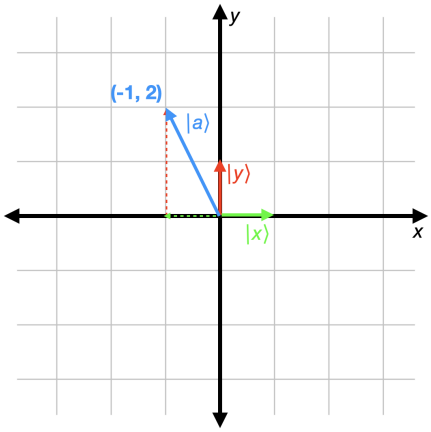
$$\begin{aligned} \langle a|b\rangle &= \begin{bmatrix} a_x^* a_y^* \end{bmatrix} \cdot \begin{bmatrix} b_x \\ b_y \end{bmatrix} \\ &= a_x^* b_x + a_y^* b_y \end{aligned} \tag{3.3}$$

3.2.2 Vector basics

Conceptual Video: Vectors, what even are they? by 3Blue1Brown (9:51)
Technical video: Vector intro for linear algebra by Khan Academy (5:49)

When we first learn to think about vectors, we are often taught to think of them as objects that have magnitude and direction, and we represent them as arrows in a Cartesian plane. We'll refer to these as *Cartesian vectors.* For sufficiently general understandings of what we mean by "magnitude" and "direction," this way of thinking will actually get us surprisingly far. We'll start by thinking about vectors in a two-dimensional, xy-plane, and then generalize our discussion to higher-dimensions and a broader understanding of what a vector is.

We'll start with some vector $\mathbf{a}$ (or $|a\rangle$) and superimpose some xy-plane on this vector:



There are lots of ways that we could describe this vector, but whatever we choose we will ultimately be indicating its direction and its magnitude. One straightforward way to describe its direction is with reference to the x- and y-axes. We can specify the direction of the vector $|a\rangle$ above as -1 unit on the x-axis and +2 units on the y-axis. We write this as

$$|a\rangle = -1\,|x\rangle + 2\,|y\rangle \qquad (\text{or} \quad \mathbf{a} = -1\,\hat{\mathbf{x}} + 2\,\hat{\mathbf{y}})\,, \tag{3.4}$$

where $|x\rangle$ and $|y\rangle$ (or $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$) are **unit vectors** (vectors of length 1) parallel to the x- and y-axes, respectively, and are indicated in the graph above. We can determine the length of the vector $\mathbf{a}$, which we write $\|\mathbf{a}\|$, using the Pythagorean theorem:

$$\|\mathbf{a}\| = \sqrt{(-1)^2 + (2)^2} = \sqrt{5}. \tag{3.5}$$

It is often convenient to work with normalized vectors, which have a length of 1. We can always normalize any vector my multiplying each of the coefficients by the reciprocal of the vector's length. For $\mathbf{a}$ this gives

$$|\tilde{a}\rangle = -\frac{1}{\sqrt{5}}\,|x\rangle + \frac{2}{\sqrt{5}}\,|y\rangle \qquad \left(\text{or} \quad \tilde{\mathbf{a}} = -\frac{1}{\sqrt{5}}\,\hat{\mathbf{x}} + \frac{2}{\sqrt{5}}\,\hat{\mathbf{y}}\right). \tag{3.6}$$

Before moving on, we need to define two operations for vectors spaces: scalar multiplication and vector addition. Multiplication of a general, two-dimensional Cartesian vector

$$|v\rangle = v_x\,|x\rangle + v_y\,|y\rangle\,, \tag{3.7}$$

by a (generally complex) scalar c is obtained my multiplying each of the coefficients in the vector by c, *viz.*

$$c\,|v\rangle = cv_x\,|x\rangle + cv_y\,|y\rangle \qquad (\text{or} \quad c\mathbf{v} = cv_x\,\hat{\mathbf{x}} + cv_y\,\hat{\mathbf{y}}) \tag{3.8}$$

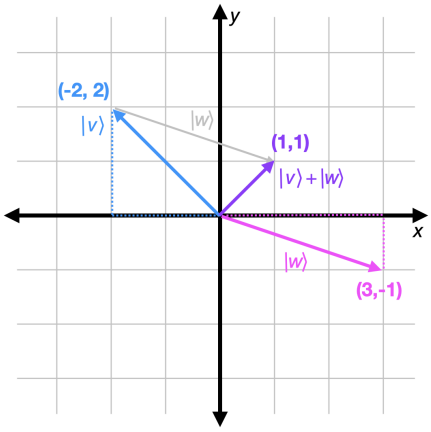
Conceptually, this amounts to rescaling the vector $|v\rangle$. Likewise, we can add $|v\rangle$ to some other general vector

$$|w\rangle = w_x\,|x\rangle + w_y\,|y\rangle \tag{3.9}$$

by adding together the coefficients along $|x\rangle$ and $|y\rangle$, yielding

$$|v\rangle + |w\rangle = (v_x + w_x)\,|x\rangle + (v_y + w_y)\,|y\rangle \qquad (\text{or} \quad \mathbf{v} + \mathbf{w} = (v_x + w_x)\,\hat{\mathbf{x}} + (v_y + w_y)\,\hat{\mathbf{y}})\,. \tag{3.10}$$

Graphically, the resulting vector $|\psi\rangle = |x\rangle + |y\rangle$ is obtained by first moving along the vector $|v\rangle$ and then from the end of $|v\rangle$ moving along $|w\rangle$. This is often referred to as placing the vectors "head-to-tail."



Closely related to the concept of vector addition is the concept of the **zero vector**, $|0\rangle = 0\,|x\rangle + 0\,|y\rangle$. From the definition of vector addition, it follows that

$$|v\rangle + |0\rangle = |0\rangle + |v\rangle = |v\rangle \tag{3.11}$$

for any vector $|v\rangle$.

Regardless of whether we use the $\mathbf{v}$ or the $|v\rangle$ notation, we can also write vectors as coefficients in columns, *viz.*

$$|v\rangle = \begin{bmatrix} v_x \\ v_y \end{bmatrix} \qquad |w\rangle = \begin{bmatrix} w_x \\ w_y \end{bmatrix}. \tag{3.12}$$

Here, the basis vectors are implicit in the notation, and we have to mentally keep track of the fact that the coefficients in the first and second rows correspond to $|x\rangle$ and $|y\rangle$, respectively. In this notation, the rules of scalar multiplication and vector addition apply as follows:

$$c\,|v\rangle = c\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} cv_x \\ cv_y \end{bmatrix} \tag{3.13}$$

$$|v\rangle + |w\rangle = \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \begin{bmatrix} v_x + w_x \\ v_y + w_y \end{bmatrix}. \tag{3.14}$$

3.2.3 Inner products

Technical video: Vector dot product and vector length by Khan Academy (9:09)

In quantum mechanics, the vectors of interest will also have a suitably defined inner product, which is an operation that takes to vectors and returns a (finite) scalar. Using the notation we have been developing, we write the inner product of two vectors $|v\rangle$ and $|w\rangle$ as $\langle v|w\rangle$. We won't worry about these too much now, but for reference, the inner product needs to satisfy three properties:

1. $\langle v|w\rangle = \langle w|v\rangle^*$
2. $\langle v|v\rangle \geq 0$
3. $\langle v| \left(|w_1\rangle + |w_2\rangle\right) = \langle v|w_1\rangle + \langle v|w_2\rangle$.

For Cartesian vectors, the inner product is often called the dot product (written $\mathbf{v} \cdot \mathbf{w}$), and it is computed as the sum of the product of the coefficients along each of the coordinates:

$$\mathbf{v} \cdot \mathbf{w} = v_x w_x + v_y w_y. \tag{3.15}$$

In the general case of a Cartesian space with N dimensions, we write the dot product as

$$\mathbf{v} \cdot \mathbf{w} = \sum_i^N v_i w_i \tag{3.16}$$

We commonly interpret the inner product of a vector with itself as the square of the length of the vector. This interpretation makes sense for Cartesian vectors.

Example 3.1: Compute the inner product of the vector $|a\rangle$ above with itself and verify the length you obtain matches that we determined.

The inner product of two different vectors gives a measure of the degree of similarity or "overlap" of these two vectors. If the inner product between $|v\rangle$ and $|w\rangle$ is zero, we say that $|v\rangle$ and $|w\rangle$ are *orthogonal*; for Cartesian vectors, this is the same thing as being perpendicular.

3.2.4 Vector spaces

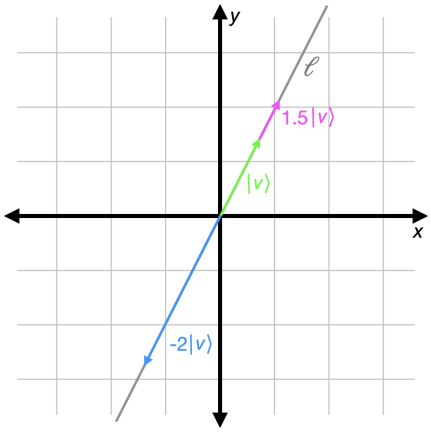
A **vector space** is a collection of vectors that satisfies a formal set of properties. The most conceptually important ones are as follows:

1. Closure under addition: or all vectors $|v\rangle$ and $|w\rangle$ in the vector space, the vector $|a\rangle = |v\rangle + |w\rangle$ must also be in the vector space
2. Closure under scalar multiplication: for all vectors $|v\rangle$ in the vector space and all coefficients c, the vector $c\,|v\rangle$ must also be in the vector space
3. It contains the zero vector, $|0\rangle$

An example is helpful. We mentioned above that scalar multiplication rescales the vector $|v\rangle$. Because of this property, $|v\rangle$ defines a line in the Cartesian plane. By this we mean that we can reach any point along the line ℓ that contains $|v\rangle$ by multiplying $|v\rangle$ by some scalar c. For any two vectors $|v_1\rangle = c_1 |v\rangle$ and $|v_2\rangle = c_2 |v\rangle$, the sum of these vectors is

$$|v_1\rangle + |v_2\rangle = (c_1 + c_2) |v\rangle \,, \tag{3.17}$$

which is also on the line ℓ. Since we have chosen the vector space to be the set of all vectors $c|v\rangle$, the set is closed under scalar multiplication. Finally, multiplying by $c = 0$ gives the zero vector $|0\rangle$, so this is contained in our space. Each of these things can be confirmed graphically as well:



The graph of the vector $|v\rangle$ and the line ℓ helps to illustrate what we mean by a vector space, which we can loosely take to be a self-contained set of vectors. Formally, we say that the vector $|v\rangle$ defines a **subspace** of the Cartesian plane. This subspace, which we'll call V, separates the Cartesian plane into points that we can reach by scalar multiplication of the vector $|v\rangle$ and points that we cannot reach by scalar multiplication of $|v\rangle$. Because it only takes a single vector ($|v\rangle$) to specify all of the points in the subspace, we say that the subspace is *one-dimensional*. The vector $|v\rangle$ is said to **span** the one-dimensional vector space V.

The Cartesian plane itself can be thought of as a vector space. To see this, we can define it the vectors of form

$$|\psi\rangle = c_x |x\rangle + c_y |y\rangle \tag{3.18}$$

for all (real) coefficients c_x and c_y. We can show that this vector space satisfies the axioms above, and we commonly refer to this vector space as $\mathbb{R}^2$. The expression for $|\psi\rangle$ above takes the form of multiplying two vectors ($|x\rangle$ and $|y\rangle$) by scalars and adding them together. This procedure is called **linear combination**. By forming these linear combinations of $|x\rangle$ and $|y\rangle$, we can reach any point in the Cartesian plane, and we say that $|x\rangle$ and $|y\rangle$ **span** $\mathbb{R}^2$, as above. Because we need two vectors to describe the points in $\mathbb{R}^2$, we say that this space is two-dimensional.

3.2.5 Linear independence

It is important to note that we cannot choose *any* two vectors from $\mathbb{R}^2$ at random and be guaranteed to have two vectors that span $\mathbb{R}^2$. If we choose two vectors that are both in the

subspace V, for instance, we will not be able to reach any point in the Cartesian plane, but only those along the line ℓ. This brings up the concept of **linear independence**. We say that two vectors are **linearly independent** if we cannot form one by rescaling the other. If we can form one vector by rescaling the other, we say that the two vectors are **linearly dependent**. The vectors $|x\rangle$ and $|y\rangle$ above are an example of linearly independent vectors. Importantly, any two orthogonal vectors must be linearly independent (although two vectors can be linearly independent and not be orthogonal).

We would say that a third vector is linearly independent from two other vectors if we could not form this third vector by forming linear combinations of the other two. The unit vector parallel to the z-axis in a three dimensional Cartesian plane ($|z\rangle$) is linearly independent from $|x\rangle$ and $|y\rangle$. Conversely, a third vector would be said to be linearly dependent on the first two vectors if you could write it as a linear combination of these two vectors. Any vector $|v\rangle$ of the form indicated above is therefore linearly dependent on $|x\rangle$ and $|y\rangle$. We would also say the set $\{|x\rangle, |y\rangle, |v\rangle\}$ is linearly dependent.

In general, the dimension of a vector space is given by the size of the largest set of linearly independent vectors we can form by taking vectors from the space.

3.3 Basis vectors

Conceptual video: Linear combinations, span, and basis vectors by 3 Blue 1 Brown (9:58)

We will often want a way to conveniently represent all of the vectors in a vector space that allows us to relate different vectors to each other and perform various operations on them. Above, we wrote the vectors in $\mathbb{R}^2$ as linear combinations of $|x\rangle$ and $|y\rangle$, *viz.*

$$|v\rangle = c_x |x\rangle + c_y |y\rangle . \tag{3.19}$$

Because $|x\rangle$ and $|y\rangle$ are two linearly independent vectors that allows us write any vector in $\mathbb{R}^2$ ("they span $\mathbb{R}^2$"), we say that $|x\rangle$ and $|y\rangle$ form a basis for $\mathbb{R}^2$. This choice of basis vectors is not unique, and there is almost always an infinite number of choices for our basis. For instance, we could replace the vector $|y\rangle$ with $|a\rangle = -1|x\rangle + 2|y\rangle$ from above, using $\{|x\rangle, |a\rangle\}$ as a basis for $\mathbb{R}^2$. In general, we could choose any two linearly independent vectors from $\mathbb{R}^2$ and use these as the basis for the space.

A basis for an N-dimensional vector space will have N-vectors, and we can write *any* vector $|v\rangle$ in this space as a linear combination of these basis vectors according to

$$|v\rangle = \sum_i^N c_i |b_i\rangle , \tag{3.20}$$

where $\{|b_i\rangle\}$ is the set of basis vectors and c_i are the coefficients along each of the basis vectors.

3.3.1 Orthonormal basis vectors

One class of basis vectors are especially easy to work with and make a lot of problems easier to work through: **orthonormal** basis vectors. A set of basis vectors must satisfy two criteria

to form an orthonormal basis: all vectors in the basis must be normalized, and all vectors in the basis must be mutually orthogonal.

The orthogonality property is particularly useful when we want to express a general vector in terms of our basis. For an orthogonal set of basis vectors, the expansion coefficients in Eq. (3.20) above can be efficiently determined using the dot product. It must be the case that

$$c_i = \langle b_i | v \rangle \tag{3.21}$$

for all of the coefficients and basis vectors. Because of this definition, you will often see equations like Eq. (3.20) written as

$$|v\rangle = \sum_i^N c_i \, |b_i\rangle = \sum_i^N |b_i\rangle \, \langle b_i | v\rangle \, . \tag{3.22}$$

Example 3.2: (a) Verify that the following three vectors form an orthogonal basis for $\mathbb{R}^3$

$$|w_1\rangle = \begin{bmatrix} -1 \\ 3 \\ 7 \end{bmatrix}, \qquad |w_2\rangle = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}, \qquad |w_3\rangle = \begin{bmatrix} 1 \\ 3 \\ -\frac{8}{7} \end{bmatrix}.$$

(b) Normalize these three vectors to obtain an orthonormal basis.
(c) Expand the vector $|a\rangle = (4, 5, 7)$ in terms of this basis.

3.3.2 Gram–Schmidt orthogonalization

Technical video: The Gram–Schmidt Process by Khan Academy (19:24)

Because of the usefulness of orthogonal basis vectors, it is helpful to have a procedure to orthogonalize two linearly independent vectors. In general there are multiple ways to do this, and the Gram–Schmidt orthogonalization procedure is a fairly common example. We will illustrate it using the vectors $|x\rangle$ and $|a\rangle = -1\,|x\rangle + 2\,|y\rangle$ from above in order to obtain a new vector $|b\rangle$ that is orthogonal to $|a\rangle$.

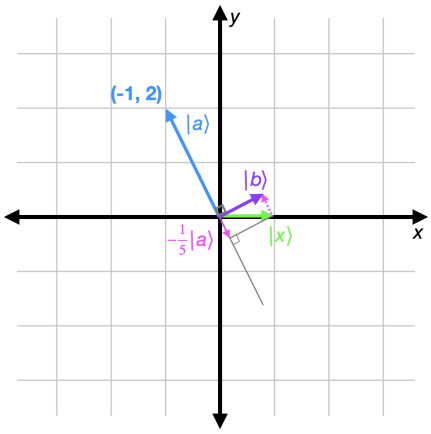
First we compute the **projection** of $|x\rangle$ along the vector $|a\rangle$. This is shown graphically in the figure below, and we accomplish it mathematically using the following formula

$$\text{proj}_{|a\rangle} \, |x\rangle = \left(\frac{|a\rangle \, \langle a|}{\langle a|a\rangle} \right) |x\rangle = \frac{\langle a|x\rangle}{\langle a|a\rangle} \, |a\rangle \, . \tag{3.23}$$

For the specific example we are using, this gives

$$\text{proj}_{|a\rangle} \, |x\rangle = -\frac{1}{5} \, |a\rangle \, . \tag{3.24}$$

As we can see, this gives the extent of $|x\rangle$ that lies along $|a\rangle$. Now, because we want a vector that is orthogonal to $|a\rangle$, we subtract off the amount of $|x\rangle$ that lies along $|a\rangle$, to



obtain a new vector

$$\begin{aligned} |b\rangle &= |x\rangle - \text{proj}_{|a\rangle} |x\rangle \\ &= |x\rangle + \frac{1}{5} |a\rangle \\ &= |x\rangle + \frac{1}{5} (-1 |x\rangle + 2 |y\rangle) \\ &= \frac{4}{5} |x\rangle + \frac{2}{5} |y\rangle . \end{aligned} \tag{3.25}$$

We can confirm that this vector is orthogonal to $|a\rangle$ by computing the inner product of $|a\rangle$ with $|b\rangle$:

$$\langle a|b\rangle = (-1 |x\rangle + 2 |y\rangle) \cdot \left(\frac{4}{5} |x\rangle + \frac{2}{5} |y\rangle \right) = -\frac{4}{5} + \frac{4}{5} = 0. \tag{3.26}$$

The vector $|b\rangle$ is also included in the figure. We can rescale this vector to normalize it, if we want to.

In the general case, we can orthogonalize a set of N vectors by iterating this process. With each successive vector, we must subtract off the projection along *all* proceeding vectors.

Example 3.3: Suppose we wanted to add a third dimension to our vector space and for some reason we began with the vector $|c\rangle = |x\rangle + |y\rangle + |z\rangle$. Use the Gram–Schmidt process to construct a third vector $|d\rangle$ from $|c\rangle$ that is orthogonal to both $|a\rangle$ and $|b\rangle$ in the preceding discussion. Normalize all three of these vectors to obtain an orthonormal basis for $\mathbb{R}^3$.

Note: This example is a little contrived, and it may or may not be obvious to you exactly what this vector will be immediately, and that's fine either way. You should still use the Gram–Schmidt process here to (re)familiarize yourself with it, as there are often abstract cases in quantum mechanics generally (and CHEM 221A specifically) that are not so obvious.

3.4 Generalizing to function spaces

Conceptual video: Abstract vector spaces by 3 Blue 1 Brown (16:45)

While it is common in introductory coursework to thing of basis vectors as Cartesian vectors, it is possible and useful to generalize the concept of vectors to include functions. In order to have an inner product space of functions, we need to define an inner product that has the required properties. For one-dimensional functions of x, a common definition of the inner product is

$$\langle f|g\rangle = \int_{-\infty}^{\infty} dx\, f^*(x)g(x). \tag{3.27}$$

Using this definition, a normalized vector satisfies

$$\int_{-\infty}^{\infty} dx\, f^*(x)f(x) = 1, \tag{3.28}$$

and we can normalize our vectors as long as the integral in the preceding equation is finite. We often require our functions to satisfy this condition, and a function that does so is called a **square integrable** function.

We can decompose some arbitrary function $\psi(x)$ into a linear combination of a complete set of basis functions $\{\phi_i(x)\}$ as

$$\psi(x) = \sum_i c_i\phi_i(x), \tag{3.29}$$

where the coefficients are determined through the inner product, as above, *viz.*

$$c_i = \langle\phi_i|\psi\rangle = \int_{-\infty}^{\infty} dx\, \phi_i^*(x)\psi(x). \tag{3.30}$$

Even when we are using functions (rather than arrows in a Cartesian space) as our basis vectors, we can write vectors in the standard notation we have worked with above, *i.e.* we make the association

$$\psi(x) = \sum_i^N c_i\phi_i(x) \qquad \rightarrow \qquad \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix}. \tag{3.31}$$

Here, just as above, the basis vectors are implicit in our construction of the column vector. We can choose them and define them at will, and we just need to keep track of how we have defined them. This generalization will prove incredibly important because it allows us to leverage the tools of linear algebra to solve a vast range of problems in quantum mechanics. We will use this framework of functions-as-vectors when we build further on our linear algebra in a later session.

Example 3.4: The functions $f_n(x) = x^n e^{-x^2}$ for $n = 0, 1, 2, \ldots$ form a basis for square integrable functions, meaning we can write any square integrable function $\phi(x)$ as an exact linear combination of the form

$$\phi(x) = \sum_{i=0}^{\infty} c_i f_i(x).$$

(a) Normalize the first two functions in the series above ($n = 0, 1$)
(b) Show that these two functions are orthogonal
(c) Find the expansion coefficient c_i for these two basis functions for

$$\phi(x) = xe^{-x^2/2}.$$

Note: You should feel free to consult the integral table in Appendix B for this problem.

3.5 Connections to physical chemistry

It would be difficult to overstate the relevance of vectors and vector spaces to quantum mechanics. As you will soon find out if you do not know already, we represent physical states with vectors in quantum mechanics. Hence, any type of mathematical analysis that we wish to do on a quantum mechanical system will require a deep knowledge of what vectors are and how we describe them. In particular, the concept of expanding a vector in terms of an arbitrary basis will be used repeatedly throughout your introductory coursework in quantum mechancis and throughout your research career.

3.6 Example problem solutions

Example 3.1: *Compute the inner product of the vector* $|a\rangle = -1\,|x\rangle + 2\,|y\rangle$ *with itself and verify the length you obtain matches that we determined.*

We compute the inner product of the vector $|a\rangle$ by computing the sum of the product of the coefficients for each basis vector in $|a\rangle$, *viz.*

$$\langle a|a\rangle = (-1) * (-1) + (2) * (2) = 5. \tag{3.32}$$

As discussed in Section 3.2.3, we interpret the inner product as the square of the length of the vector, and therefore find that the length of $|a\rangle$ is $\sqrt{5}$, which is what we had determined it to be using the Pythagorean theorem.

Example 3.2: *(a) Verify that the following three vectors form an orthogonal basis for* $\mathbb{R}^3$

$$|w_1\rangle = \begin{bmatrix} -1 \\ 3 \\ 7 \end{bmatrix}, \qquad |w_2\rangle = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}, \qquad |w_3\rangle = \begin{bmatrix} 1 \\ 3 \\ -\frac{8}{7} \end{bmatrix}.$$

(b) Normalize these three vectors to obtain an orthonormal basis.
(c) Expand the vector $|a\rangle = (4, 5, 7)$ *in terms of this basis.*

(a) We confirm that each of the three vectors are mutually orthogonal by performing the following dot products:

$$\langle w_1|w_2\rangle = (-1)(3) + (3)(-1) + (7)(0) = -3 + 3 + 0 = 0 \tag{3.33}$$

$$\langle w_1|w_3\rangle = (-1)(1) + (3)(3) + (7)\left(-\frac{8}{7}\right) = -1 + 9 - 8 = 0 \tag{3.34}$$

$$\langle w_2|w_3\rangle = (3)(1) + (-1)(3) + (0)\left(-\frac{8}{7}\right) = 3 - 3 + 0 \tag{3.35}$$

(b) We normalize each vector by dividing each of its coefficient by the square root of its inner product:

$$\langle w_1|w_1\rangle = (-1)^2 + (3)^2 + (7)^2 = 59 \quad \rightarrow \quad |\tilde{w}_1\rangle = \frac{1}{\sqrt{59}} \begin{bmatrix} -1 \\ 3 \\ 7 \end{bmatrix} \tag{3.36}$$

$$\langle w_2|w_2\rangle = (3)^2 + (-1)^2 + (0)^2 = 10 \quad \rightarrow \quad |\tilde{w}_2\rangle = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \tag{3.37}$$

$$\langle w_3|w_3\rangle = (1)^2 + (3)^2 + \left(-\frac{8}{7}\right)^2 = \frac{554}{49} \quad \rightarrow \quad |\tilde{w}_3\rangle = \frac{7}{\sqrt{554}} \begin{bmatrix} 1 \\ 3 \\ -\frac{8}{7} \end{bmatrix} \tag{3.38}$$

(c) The coefficient for the vector $|a\rangle$ along each of the basis vectors $\{|w_i\rangle\}$ can be determined by computing the inner product $\langle w_i|a\rangle$ for $i = 1, 2, 3$. Following this procedure, we have

$$c_1 = \langle w_1|a\rangle = \frac{1}{\sqrt{59}}\left(-1 \cdot 4 + 3 \cdot 5 + 7 \cdot 7\right) = \frac{54}{\sqrt{49}} \tag{3.39}$$

$$c_2 = \langle w_2|a\rangle = \frac{1}{\sqrt{10}}\left(3 \cdot 4 - 1 \cdot 5 + 0 \cdot 7\right) = \frac{7}{\sqrt{10}} \tag{3.40}$$

$$c_3 = \langle w_3|a\rangle = \frac{7}{\sqrt{554}}\left(1 \cdot 4 + 3 \cdot 5 - \frac{8}{7} \cdot 7\right) = \frac{77}{\sqrt{554}} \tag{3.41}$$

Thus, we have

$$|a\rangle = \frac{54}{\sqrt{49}}|w_1\rangle + \frac{7}{\sqrt{10}}|w_2\rangle + \frac{77}{\sqrt{554}}|w_3\rangle. \tag{3.42}$$

Example 3.3: *Suppose we wanted to add a third dimension to our vector space and for some reason we began with the vector* $|c\rangle = |x\rangle + |y\rangle + |z\rangle$. *Use the Gram–Schmidt process to construct a third vector* $|d\rangle$ *from* $|c\rangle$ *that is orthogonal to both* $|a\rangle = (-1, 2)$ *and* $|b\rangle = (4/5, 2/5)$ *in the preceding discussion. Normalize all three of these vectors to obtain an orthonormal basis for* $\mathbb{R}^3$

The Gram–Schmidt process orthogonalizes a set of vectors by starting with one vector and subtracting off its projections along other vectors. Thus, in long form, we can write the vector we're looking for as

$$|d\rangle = |c\rangle - \text{proj}_{|a\rangle}\,|c\rangle - \text{proj}_{|b\rangle}\,|c\rangle\,. \tag{3.43}$$

Using the definition of the projection operator given in Eq. (3.23), we have

$$\text{proj}_{|a\rangle}\,|c\rangle = \frac{\langle a|c\rangle}{\langle a|a\rangle}\,|a\rangle = \frac{1}{5}\,|a\rangle \tag{3.44}$$

$$\text{proj}_{|b\rangle}\,|c\rangle = \frac{\langle b|c\rangle}{\langle b|b\rangle}\,|b\rangle = \frac{6/5}{4/5}\,|b\rangle = \frac{3}{2}\,|b\rangle\,. \tag{3.45}$$

Hence, we have

$$|d\rangle = \begin{bmatrix}1\\1\\1\end{bmatrix} - \frac{1}{5}\begin{bmatrix}-1\\2\\0\end{bmatrix} - \frac{3}{2}\begin{bmatrix}4/5\\2/5\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}. \tag{3.46}$$

We can see by inspection that the vector $|d\rangle$ is already normalized. We normalize $|a\rangle$ and $|b\rangle$ in the usual way: dividing each vector by the square root of its inner product with itself:

$$|\tilde{a}\rangle = \frac{1}{\sqrt{\langle a|a\rangle}}\,|a\rangle = \frac{1}{\sqrt{5}}\,|a\rangle = \begin{bmatrix}-1/\sqrt{5}\\2/\sqrt{5}\\0\end{bmatrix} \tag{3.47}$$

$$\left|\tilde{b}\right\rangle = \frac{1}{\sqrt{\langle b|b\rangle}}\,|b\rangle = \frac{1}{\sqrt{4/5}}\,|b\rangle = \begin{bmatrix}2/\sqrt{5}\\1/\sqrt{5}\\0\end{bmatrix}. \tag{3.48}$$

Example 3.4: *The functions $f_n(x) = x^n e^{-x^2}$ for $n = 0, 1, 2, \ldots$ form a basis for square integrable functions, meaning we can write any square integrable function $\phi(x)$ as an exact linear combination of the form*

$$\phi(x) = \sum_{i=0}^{\infty} c_i f_i(x).$$

(a) Normalize the first two functions in the series above ($n = 0, 1$)
(b) Show that these two functions are orthogonal
(c) Find the expansion coefficient c_i for these two basis functions for

$$\phi(x) = xe^{-x^2/2}.$$

Note: *You should feel free to consult the integral table in Appendix B for this problem.*

(a) The metric (definition of the inner product) for this function space is given by

$$\langle n|n\rangle = \int_{-\infty}^{\infty} dx\; n^*(x)n(x), \tag{3.49}$$

and we can use this definition to normalize any of the functions $\{f_n(x)\}$ in the same way that we would normalize vectors. Starting with $n = 0$, we first obtain

$$\begin{aligned}\langle f_0|f_0\rangle &= \int_{-\infty}^{\infty} dx \left(e^{-x^2}\right)\left(e^{-x^2}\right) \\ &= \int_{-\infty}^{\infty} dx\; e^{-2x^2} \\ &= \sqrt{\frac{\pi}{2}},\end{aligned} \tag{3.50}$$

where we have used the general result for the integral of a Gaussian function over all space (this can be referenced in an integral table at any point). We now obtain the normalized basis function in the usual way:

$$\left|\tilde{f}_0\right\rangle = \frac{1}{\sqrt{\langle f_0|f_0\rangle}}\,|f_0\rangle = \left(\frac{2}{\pi}\right)^{1/4} e^{-x^2}. \tag{3.51}$$

Likewise, for the basis function with $n = 1$, we have

$$\begin{aligned}\langle f_1|f_1\rangle &= \int_{-\infty}^{\infty} dx \left(xe^{-x^2}\right)\left(xe^{-x^2}\right) \\ &= \int_{-\infty}^{\infty} dx\; x^2 e^{-2x^2} \\ &= \frac{1}{2}\sqrt{\frac{\pi}{8}} = \sqrt{\frac{\pi}{32}}\end{aligned} \tag{3.52}$$

and therefore

$$\left|\tilde{f}_1\right\rangle = \frac{1}{\sqrt{\langle f_1|f_1\rangle}}\,|f_1\rangle = \left(\frac{32}{\pi}\right)^{1/4} xe^{-x^2}. \tag{3.53}$$

(b) Two vectors are orthogonal if their inner product is equal to zero. We compute the inner product of $|f_0\rangle$ and $|f_1\rangle$ as

$$\begin{aligned}\langle f_0|f_1\rangle &= \int_{-\infty}^{\infty} dx \left[\left(\frac{2}{\pi}\right)^{1/4} e^{-x^2}\right]\left[\left(\frac{32}{\pi}\right)^{1/4} xe^{-x^2}\right] \\ &= \left(\frac{8}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} dx\; xe^{-x^2} = 0.\end{aligned} \tag{3.54}$$

As indicated, this integral is equal to zero. This must be the case because it is the product of an odd function (x) and an even function (e^{-x^2}), so it is an odd function, which will have an integral over all space that vanishes. Hence the two basis functions are orthogonal.

(c) The expansion coefficient for some target function along a given basis function is computed as the projection of the the target function onto the basis function, *viz.*

$$\langle f_n|\phi\rangle \tag{3.55}$$

Continuing with the inner product for this space, we have

$$\begin{aligned}c_0 &= \langle f_0|\phi\rangle \\ &= \int_{-\infty}^{\infty} dx \left[\left(\frac{2}{\pi}\right)^{1/4} e^{-x^2}\right]\left(xe^{-x^2/2}\right) \\ &= \left(\frac{2}{\pi}\right)^{1/4} \int_{-\infty}^{\infty} dx\; xe^{-3x^2/2} = 0\end{aligned} \tag{3.56}$$

because the integrand is an odd function. For the basis function with $n = 1$, we have

$$\begin{aligned}c_1 &= \langle f_1|\phi\rangle \\ &= \int_{-\infty}^{\infty} dx \left[\left(\frac{32}{\pi}\right)^{1/4} xe^{-x^2}\right]\left(xe^{-x^2/2}\right) \\ &= \left(\frac{32}{\pi}\right)^{1/4} xe^{-x^2} \int_{-\infty}^{\infty} dx\; x^2 e^{-3x^2/2} \\ &= \left(\frac{32}{\pi}\right)^{1/4} \left(\frac{1}{2}\sqrt{\frac{\pi}{(3/2)^3}}\right) \\ &= \left(\frac{128\pi}{27}\right)^{1/4}.\end{aligned} \tag{3.57}$$

4 Linear Algebra 2: Transformations

Topics: Transformations/operators, linear transformations, unitary transformations, projection operators, Hermitian operators

4.1 Motivation and approach

We referenced matrices in Sec. 3 above when we were thinking about how to write matrix/vector manipulations as summations in order to deal with arbitrarily sized vector spaces. Otherwise, matrices have been noticeably absent from our discussion of linear algebra. This may come as a surprise to you, especially if your previous exposure to linear algebra relied heavily on matrices. Our initial discussion in Sec. 3 focused exclusively on vectors because vectors are the fundamental object in linear and abstract algebras generally. This is also the case in quantum mechanics, where we associate the state of a system with a vector. Since the state of our quantum mechanical system can change for various reasons, we want to be able to mathematically describe what it means to go from one state to another; that is, we need to be able to relate one vector to another.

So how, in the context of linear algebra, do we get from one vector to another? We do this through a **transformation**, and we represent transformations with matrices. These matrices "transform" a vector through matrix–vector multiplication, as defined in Sec. 3, mapping this vector onto a new vector, which may or may not be in the same vector space as the initial vector. Both the conceptualization and the formalism of transformations carries over to systems where we use functions to represent our vectors, although here it is common to refer to transformations as **operators**.

We discuss both transformations and operators (which you should begin to think of as the same thing) in a general way below, and we also introduce specific types of transformations/operators that are of particular importance: linear transformations, unitary transformations, Hermitian operators, and projections. We introduce some of these as "transformations" and others as "operators," and the reasons for doing this are pedagogical and arbitrary. Hopefully using both frameworks will help you to practice thinking about how the ideas we're developing apply to different contexts. In any case, the concepts we develop below apply equally to Cartesian vectors, functions, and other types of vectors. That's part of what makes linear algebra so neat!

4.2 Transformations and operators

Conceptual and technical video: Linear transformations and matrices by 3Blue1Brown (10:58)
Technical video: Linear transformations as matrix vector products by Khan Academy (17:31)

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

The **transformation** F takes a vector $\mathbf{v}$ and moves it to another vector $\mathbf{w} = F(\mathbf{v})$. This is similar to how a function f takes one scalar a and returns another scalar $b = f(a)$. When we're speaking about **operators**, it's common to use a slightly different notation and phrasing, saying that the operator $\hat{O}$ acts on the vector $|v\rangle$ to return the vector $|w\rangle$, which we write altogether as $|w\rangle = \hat{O}\,|v\rangle$. We'll use both notations below.

One particular class of transformations/operators is of such importance mathematically and physically that it will be the exclusive focus of our discussion here, not to mention almost all of quantum mechanics: linear transformations.

4.2.1 Linear transformations

At its core, a **linear transformation** is a transformation that preserves the effects of both addition and scalar multiplication. Mathematically, this means that a linear transformation L satisfies

$$L(\mathbf{v} + \mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w}) \tag{4.1}$$
$$L(c\mathbf{v}) = cL(\mathbf{v}) \tag{4.2}$$

for any vectors $\mathbf{v}, \mathbf{w}$ and scalars c in the domain upon which the transformation is defined. We could also write this

$$\hat{L}\left(|v\rangle + |w\rangle\right) = \hat{L}\,|v\rangle + \hat{L}\,|w\rangle \tag{4.3}$$
$$\hat{L}(c\,|v\rangle) = c\hat{L}\,|v\rangle\,. \tag{4.4}$$

The importance of linear transformations is related to the idea of a linear combination, which allows us to write any vector in a vector space in terms of a set of basis vectors, as we saw in Sec. 3. Say we have some vector $|\psi\rangle$ that we write as a linear combination of basis vectors $\{|\phi_i\rangle\}$ for $i = 1, 2, \ldots, N$ according to

$$|\psi\rangle = \sum_{i=1}^{N} c_i\,|\phi_i\rangle \qquad \left(\text{or} \quad \mathbf{v} = \sum_{i=1}^{N} c_i \mathbf{b}_i\right). \tag{4.5}$$

The two properties of a linear operator indicated above then allow us to evaluate the action of the operator $\hat{L}$ on the arbitrary state $|\psi\rangle$ according to

$$\hat{L}\,|\psi\rangle = \hat{L}\left(\sum_{i=1}^{N} c_i\,|\phi_i\rangle\right) \tag{4.6}$$
$$= \sum_{i=1}^{N} c_i\left(\hat{L}\,|\phi_i\rangle\right) \qquad \left(\text{or} \quad L(\mathbf{v}) = \sum_{i=1}^{N} c_i L(\mathbf{b}_i)\right). \tag{4.7}$$

In words, this means that we can specify how the operator $\hat{L}$ acts on *any* vector $|\psi\rangle$ in the N-dimensional vector space by specifying its action on some set of N vectors that form a basis for the space.

This is incredibly important and bears repeating: we can determine the action of a linear operator $\hat{L}$ on an almost always infinite number of vectors by determining its action on a sometimes finite and always smaller number of basis vectors. For a concrete example, we can think of some transformation $\hat{F}$ that acts on the infinite set of vectors in $\mathbb{R}^3$. A basis of this space contains three vectors, and we'll take the standard Cartesian basis of $|x\rangle, |y\rangle$, and $|z\rangle$. If we are able to determine $\hat{F}|x\rangle, \hat{F}|y\rangle$, and $\hat{F}|z\rangle$, then what we have shown above means we know how to find $\hat{F}|v\rangle$ for any vector $|v\rangle$ in $\mathbb{R}^3$.

4.2.2 Representation as matrices

Linear transformations and operators can be effectively represented using matrices. Indeed, one helpful way to think about matrices is to think of them as the way to write a given transformation for some vector space of interest.

For a three-dimensional vector space, matrix–vector multiplication is defined such that placing the vectors $\hat{A}|x\rangle$, $\hat{A}|y\rangle$, and $\hat{A}|z\rangle$ in order in the columns of a matrix affords the transformation matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} | & | & | \\ \hat{A}|x\rangle & \hat{A}|y\rangle & \hat{A}|z\rangle \\ | & | & | \end{bmatrix}. \tag{4.8}$$

With this matrix in hand, we can determine the effect of this transformation on any vector $\mathbf{v} \in \mathbb{R}^3$ through matrix vector multiplication according to

$$\mathbf{v}' = \mathbf{A}\mathbf{v}, \tag{4.9}$$

where $\mathbf{v}'$ is the transformed vector.

Example 4.1: Suppose we are interested in rotating the vectors in $\mathbb{R}^3$ by some angle θ about the y-axis in a clockwise fashion. Determine the action of this transformation on each of the standard Cartesian basis vectors, and use this to construct the matrix that corresponds to this rotation.

Example 4.2: Consider representing all rank-3 polynomials in terms of a basis $|x^n\rangle$ for $n = 0, 1, 2, 3$, such that we can write any rank-3 polynomial as

$$f(x) = c_0 \left|x^0\right\rangle + c_1 \left|x^1\right\rangle + c_2 \left|x^2\right\rangle + c_3 \left|x^3\right\rangle . \tag{4.10}$$

Within this representation, we would write the function $g(x) = 3x^2 - 1$ as

$$|g\rangle = \begin{bmatrix} -1 \\ 0 \\ 3 \\ 0 \end{bmatrix} . \tag{4.11}$$

(a) Write the matrix representation of the operator $\hat{A} = \frac{d}{dx}$ in this basis.
(b) How would you write the operator $\hat{B} = \int dx$ in this basis? What does this tell you about the action of the operator $\hat{B}$ on vectors in this space?
(c) Write the matrix representation of the operator $\hat{B}$ that transforms a vector in the original basis into an expanded space that contains these four vectors and an additional vector $|x^4\rangle$. Is there some ambiguity in this transformation?

4.3 Special types of operators

Within the class of linear operators, there are a number of other types of operators that are worth mentioning because of their relevance to quantum mechanics. We will briefly discuss each of these operators and their matrix representations below.

4.3.1 The identity operator

The **identity operator**, $\hat{I}$, is trivial in the mathematical sense. It takes a vector $|v\rangle$ and returns the same vector $|v\rangle$, which we write

$$\hat{I} |v\rangle = |v\rangle . \tag{4.12}$$

In particular, this relationship holds for each of the basis vectors in any N-dimensional vector space, such that the N-dimensional identity matrix is given by

$$\mathbf{I}_N = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{4.13}$$

The main reason the identity operator/matrix is important here is that it allows us to define the concept of an **inverse** of a matrix. The inverse $\mathbf{A}^{-1}$ of a matrix $\mathbf{A}$ satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}. \tag{4.14}$$

4.3.2 Unitary operators

Unitary operators are norm-conserving operators, *i.e.* they leave the length of a vector unchanged when they operate on it. Rotating a Cartesian vector is a classic example of a unitary matrix, and this example is so intuitive that people will sometimes refer to *all* unitary transformations as rotations, regardless of whether they seem like "rotations" in the sense we are used to thinking about them.

Mathematically, a unitary matrix is defined by the property that its conjugate transpose (or *adjoint*) is equal to its inverse. We represent the conjugate transpose of a matrix $\mathbf{A}$ as $\mathbf{A}^\dagger$, and in the three-dimensional case we write

$$\mathbf{A}^\dagger = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}^\dagger = \begin{bmatrix} A_{11}^* & A_{21}^* & A_{31}^* \\ A_{12}^* & A_{22}^* & A_{32}^* \\ A_{13}^* & A_{23}^* & A_{33}^* \end{bmatrix}. \tag{4.15}$$

This means a unitary matrix satisfies

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}. \tag{4.16}$$

The identity matrix is the simplest example of a unitary matrix. Based on what we said above, the matrix representation of the rotation of a Cartesian vector must be a unitary matrix.

Example 4.3: Show that the rotation matrix you found in the example problem above is a unitary matrix.

4.3.3 Hermitian operators

Hermitian operators, represented as Hermitian matrices, are among the most important objects in quantum mechanics. They have lots of important properties that we won't discuss or prove here, but that you will consider in your coursework. In a sentence, a Hermitian matrix is equal to its adjoint (or conjugate transpose). We write this as

$$\mathbf{H} = \mathbf{H}^\dagger. \tag{4.17}$$

In the case of a vector space with strictly real coefficients, this definition means that a Hermitian matrix will be equal to its transpose, *i.e.* it is symmetric.

Example 4.4: This problem will use matrix algebra to help us think more about complex numbers. We can, in general, represent the complex number $z = a + ib$ (for $a, b \in \mathbb{R}$) as a matrix:

$$z = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \tag{4.18}$$

(a) Write the number i in this form, and show that $i^2 = -1$
(b) Show that the adjoint of this matrix is equal to the complex conjugate of z, which we write z^*
(c) Use matrix multiplication to find the square modulus ($|z|^2 = z^* z$) of z. What does the result indicate about the square modulus of a complex number?
(d) Under what condition will the matrix representation of z be Hermitian?
(e) Under what condition will the matrix representation of z be unitary?
(f) Euler's formula allows us to write the complex exponential $e^{ix} = \cos(x) + i\sin(x)$. Determine the square modulus of e^{ix} through matrix multiplication.
(g) It is also common to visualize complex numbers as vectors in the Cartesian plane, with the real component of the number, $\mathrm{Re}(z)$, along the x-axis, and the imaginary component, $\mathrm{Im}(z)$, along the y-axis. We'll call this vector $|z\rangle$. What is the effect of the operator $\hat{A} = e^{ix}$ on on the vector $|z\rangle$?

4.3.4 Projection operators

A **projection operator** acts on a vector and "flattens" it onto some vector space. Stated differently, a projection operator determines the portion of a vector that can be represented in a particular vector space. We already considered the projection of one vector onto another in Sec. 3, but here we can generalize this definition to include projections onto higher dimensional spaces. If the vectors $\{|\phi_i\rangle\}$ for $i = 1, 2, \ldots, N$ form a basis for some vector space V, then the projection onto this vector space, $\hat{P}_V$ is given by

$$\hat{P}_V = \sum_{i=1}^{N} \frac{|\phi_i\rangle \langle\phi_i|}{\langle\phi_i|\phi_i\rangle}. \tag{4.19}$$

It is common that each of the vectors $|\phi_i\rangle$ in our basis will be normalized, and in this case the definition of the projection operator simplifies to

$$\hat{P}_V = \sum_{i=1}^{N} |\phi_i\rangle \langle\phi_i|. \tag{4.20}$$

This is the form of the projection operator that we will use, and this operator acts on some vector $|\psi\rangle$ to give the projection of $|\psi\rangle$ onto V as

$$\hat{P}_V |\psi\rangle = \sum_{i=1}^{N} |\phi_i\rangle \langle\phi_i|\psi\rangle = \sum_{i=1}^{N} c_i |\phi_i\rangle, \tag{4.21}$$

where we have used the results from Sec. 3.3.1 to identify $\langle\phi_i|\psi\rangle$ as the coefficient of $|\psi\rangle$ along $|\phi_i\rangle$ in a linear combination that uses the vectors $\{\phi_i\}$ as (at least part of) its basis.

That was a lot of words, so we'll also approach this result from another angle to hopefully clear up any confusion. Instead of starting with the projection operator, suppose we started with some vector $|v\rangle$ that we wrote as a linear combination of the basis vectors $\{|\phi_i\rangle\}$ for $i = 1, 2, \ldots, N$ as

$$|\psi\rangle = \sum_{i=1}^{N} c_i \, |\phi_i\rangle \, . \tag{4.22}$$

Now we might be interested in what part of that vector $|\psi\rangle$ lies along, say, the vectors $\{|\phi_1\rangle \, , |\phi_3\rangle\}$, the span of which we'll call the vector space W. Knowing we have the linear combination above, we could conceptually write this projection of $|\psi\rangle$ as

$$\hat{P}_W \, |\psi\rangle = c_1 \, |\phi_1\rangle + c_3 \, |\phi_3\rangle \, . \tag{4.23}$$

Given this result, we want our projection operator to pluck out the coefficients along each of the vectors in the projection space and return a new vector with these coefficients. The projection operator as defined above does exactly this.

4.4 Connections to physical chemistry

As has been noted previously, quantum mechanical states are represented by vectors in a particular vector space. As a quantum system evolves in time, we can imagine the vector in our vector space moving. Transformations and operators that map an initial vector to some final vector correspond to various physical interactions between the quantum system and its environment.

Unitary operators take an important role in quantum mechanics because they are by definition norm-conserving (they do not change the length of the vector that they operate on). Briefly, the norm of a vector—or the particular components of a vector—correspond to the probability distribution of observing a particular state on measurement. The conservation of the overall norm of the state is critical within this interpretation, as whether or not there is, say, a particle in the system of interest should not change with time, even if where in the system we might find the particle does in fact change. Hence operators that correspond to the evolution of a quantum mechanical system in time will, in general, be unitary.

Hermitian operators play one of the most important roles in quantum mechanics, where we postulate that any observable property of a quantal system corresponds to some Hermitian operator.

Finally, projection operators are useful because they help us to determine the extent to which a given vector can be described by a certain subspace of the total space. This will prove to be a useful technique throughout various aspects of quantum mechanics, but it is particularly useful in perturbation theory and other approaches to improving the description of our quantum mechanical state, which will be the focus of the end of CHEM 221A.

4.5 Example problem solutions

Example 4.1: *Suppose we are interested in rotating the vectors in $\mathbb{R}^3$ by some angle θ about the y-axis in a clockwise fashion. Determine the action of this transformation on each of the standard Cartesian basis vectors, and use this to construct the matrix that corresponds to this rotation.*

The effect of this rotation, $\mathcal{R}(\theta)$, is the easiest to determine for the basis vector along the y-axis, because the rotation leaves it unchanged. Mathematically, we write this as

$$\mathcal{R}(\theta)\,|y\rangle = |y\rangle\,. \tag{4.24}$$

This rotation will mix the other basis vectors ($|x\rangle$ and $|z\rangle$) with each other to an extent that is determined by the magnitude of the angle θ. Specifically, we have

$$\mathcal{R}(\theta)\,|x\rangle = \cos(\theta)\,|x\rangle + \sin(\theta)\,|z\rangle \tag{4.25}$$
$$\mathcal{R}(\theta)\,|z\rangle = -\sin(\theta)\,|x\rangle + \cos(\theta)\,|z\rangle\,. \tag{4.26}$$

Altogether, these three expressions allow us to write the matrix representation of the rotation operator as the following matrix:

$$\mathcal{R}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}. \tag{4.27}$$

Example 4.2: *Consider representing all rank-3 polynomials in terms of a basis $|x^n\rangle$ for $n = 0, 1, 2, 3$, such that we can write any rank-3 polynomial as*

$$f(x) = c_0\left|x^0\right\rangle + c_1\left|x^1\right\rangle + c_2\left|x^2\right\rangle + c_3\left|x^3\right\rangle. \tag{4.28}$$

Within this representation, we would write the function $g(x) = 3x^2 - 1$ as

$$|g\rangle = \begin{bmatrix} -1 \\ 0 \\ 3 \\ 0 \end{bmatrix}. \tag{4.29}$$

(a) Write the matrix representation of the operator $\hat{A} = \frac{d}{dx}$ in this basis.
(b) How would you write the operator $\hat{B} = \int dx$ in this basis? What does this tell you about the action of the operator $\hat{B}$ on vectors in this space?
(c) Write the matrix representation of the operator $\hat{B}$ that transforms a vector in the original basis into an expanded space that contains these four vectors and an additional vector $|x^4\rangle$. Is there some ambiguity in this transformation?

(a) We determine the matrix representation of this operator by determining its action on each of the basis vectors. We have

$$\frac{d}{dx}\left|x^0\right\rangle = |0\rangle = 0 \tag{4.30}$$

$$\frac{d}{dx}\left|x^1\right\rangle = \left|x^0\right\rangle \tag{4.31}$$

$$\frac{d}{dx}\left|x^2\right\rangle = 2\left|x^1\right\rangle \tag{4.32}$$

$$\frac{d}{dx}\left|x^3\right\rangle = 3\left|x^2\right\rangle. \tag{4.33}$$

Hence, the matrix representation of $\hat{A} = \frac{d}{dx}$ in this basis is given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.34}$$

(b) This is a trick question: you *cannot* write the operator $\hat{B}$ in this basis. We can see this by determining the action of $\hat{B}$ on each of the basis vectors:

$$\int dx\left|x^0\right\rangle = \left|x^1\right\rangle + k_0\left|x^0\right\rangle \tag{4.35}$$

$$\int dx\left|x^1\right\rangle = \frac{1}{2}\left|x^2\right\rangle + k_1\left|x^0\right\rangle \tag{4.36}$$

$$\int dx\left|x^2\right\rangle = \frac{1}{3}\left|x^3\right\rangle + k_2\left|x^0\right\rangle \tag{4.37}$$

$$\int dx\left|x^3\right\rangle = \frac{1}{4}\left|x^4\right\rangle + k_3\left|x^0\right\rangle. \tag{4.38}$$

The problem is that $\hat{B}\left|x^3\right\rangle$ contains a contribution along $\left|x^4\right\rangle$, which is a vector that does not belong in the space as we have defined it. In other words, the operator $\hat{B}$ is not closed over the basis that we have chosen: it results in vectors that exist outside of the space.

(c) Using the equations from (b) above, we can write the operator $\hat{B}$ in the expanded basis as

$$\begin{bmatrix} k_0 & k_1 & k_2 & k_3 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}. \tag{4.39}$$

The transformation is ambiguous inasmuch as the integration constants (k_n, for $n = 0, 1, 2, 3$) are not determined. *Show that the rotation matrix you found in the example problem above is a unitary matrix.*

A matrix is unitary if the product of the matrix and its Hermitian conjugate yields the identity matrix. In the previous example problem, we found

$$\mathcal{R}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{4.40}$$

To determine if this matrix is unitary, we compute

$$\begin{aligned} \mathcal{R}^\dagger(\theta)\mathcal{R}(\theta) &= \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sin^2(\theta) + \cos^2(\theta) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{4.41}$$

Hence, we see that this rotation matrix is unitary.

Example 4.3: *This problem will use matrix algebra to help us think more about complex numbers. We can, in general, represent the complex number $z = a + ib$ (for $a, b \in \mathbb{R}$) as a matrix:*

$$z = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

(a) Write the number i in this form, and show that $i^2 = -1$
(b) Show that the adjoint of this matrix is equal to the complex conjugate of z, which we write z^*
(c) Use matrix multiplication to find the square modulus ($|z|^2 = z^* z$) of z. What does the result indicate about the square modulus of a complex number?
(d) Under what condition will the matrix representation of z be Hermitian?
(e) Under what condition will the matrix representation of z be unitary?
(f) Euler's formula allows us to write the complex exponential $e^{ix} = \cos(x) + i\sin(x)$. Determine the square modulus of e^{ix} through matrix multiplication.
(g) It is also common to visualize complex numbers as vectors in the Cartesian plane, with the real component of the number, $\mathrm{Re}(z)$, along the x-axis, and the imaginary component, $\mathrm{Im}(z)$, along the y-axis. We'll call this vector $|z\rangle$. What is the effect of the operator $\hat{A} = e^{ix}$ on on the vector $|z\rangle$?

(a) According to the definition given in the problem, i corresponds to $a = 0$ and $b = 1$, which we represent in a matrix as

$$i = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{4.42}$$

From this, we can evaluate

$$i^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.43}$$

This final matrix corresponds to $a = -1$ and $b = 0$, or just the number -1.

(b) The adjoint (Hermitian conjugate) of the general matrix z is given by

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}. \tag{4.44}$$

Meanwhile, the complex conjugate of z is given by

$$z^* = a - ib. \tag{4.45}$$

We can see that z^* is represented by the matrix written above, *i.e.* that taking the adjoint of our matrix representation corresponds to complex conjugation in the standard algebraic representation.

(c) For a general complex number z, we obtain the square modulus through matrix multiplication as

$$|z|^2 = z^* z = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}^\dagger \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{bmatrix}. \tag{4.46}$$

We note in particular that the off-diagonal elements of this matrix, which represent the imaginary portion of the number that corresponds to the matrix, are always zero. This is equivalent to saying that the square modulus of a complex number is always real.

(d) The matrix representation of z will be Hermitian when $b = -b$, *i.e.* whenever $b = 0$, or z is exclusively real.

(e) From the result in (c) above, we can see that the matrix representation of z will be unitary whenever $a^2 + b^2 = 1$.

(f) In our matrix formalism, we write $e^{ix} = \cos(x) + i\sin(x)$ as

$$e^{ix} = \begin{bmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{bmatrix}. \tag{4.47}$$

The square modulus of e^{ix} is then obtained as

$$\begin{aligned} \left|e^{ix}\right|^2 &= \begin{bmatrix} \cos(x) & \sin(x) \\ -\sin(x) & \cos(x) \end{bmatrix} \begin{bmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{bmatrix} \\ &= \begin{bmatrix} \cos^2(x) + \sin^2(x) & 0 \\ 0 & \cos^2(x) + \sin^2(x) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned} \tag{4.48}$$

(g) We have already determined the form of this matrix in part (f), and it is given by Eq. (4.47). From the form of this matrix (and by analogy to the roration matrix we determined in a previous example problem) we can see that the operator $\hat{A} = e^{ix}$ has the effect of rotating the vector $|z\rangle$ through the complex plane.

5 Linear Algebra 3: Matrix Algebra

Topics: Eigenvalues, eigenvectors, functions of matrices, summation notation

5.1 Motivation and approach

One of the major advantages of representing operators and transformations as matrices is that this allows us to take advantage of matrix algebra in order to perform computations. Many of the most important processes in matrix algebra have been streamlined for computational implementation; indeed, in some sense computers have literally been designed to do matrix algebra efficiently.

The fundamentals of this matrix algebra are the focus of this third and final module on linear algebra for the bootcamp. We'll discuss some of the theoretical underpinnings, briefly, and also illustrate how to work important types of problems.

5.2 Eigensystems

Conceptual and technical Video: Eigenvectors and eigenvalues by 3Blue1Brown (17:15)

5.2.1 Conceptual framework of an eigensystem

In this section will only discuss operators that act on a given N-dimensional vector space, or operators that can be represented by square N-by-N matrices. In general, the action of such an operator $\hat{A}$ on a vector $|v\rangle$ will yield a new vector $|w\rangle$ that is rotated relative to $|v\rangle$. However, operators and their representative matrices often have a set of vectors called **eigenvectors** for which the operator only acts to rescale the vector. If k is an eigenvector of $\hat{A}$, we have

$$\hat{A}\,|k\rangle = \lambda_k\,|k\rangle\,, \tag{5.1}$$

for some scalar λ_k, which we call the **eigenvalue** of $\hat{A}$ corresponding to the vector $|k\rangle$. Together, we refer to the eigenvalue-eigenvector pairs as the **eigensystem** of the matrix. Finding the eigensystem of an operator is an incredibly important operation in both linear algebra and quantum mechanics.

It is a fact, that we will not prove, that an Hermitian N-by-N matrix will have N eigenvector/eigenvalue pairs. These eigenvalues will be exclusively real, and the eigenvectors for different eigenvalues must be orthogonal. (You will prove these latter two statements in your quantum coursework, and you will employ all three of them repeatedly.)

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

5.2.2 Solving for an eigensystem

Technical video: Example solving for eigenvalues of a 2x2 matrix by Khan Academy (5:38)
Technical video: Finding eigenvectors and eigenspaces example by Khan Academy (14:33)
Technical video: 3 x 3 Determinant by Khan Academy (10:01)

We can follow a general procedure to solve for the eigensystem of a matrix according to Eq. (5.1). We can represent our operator and vector as a matrix and column vectors following the procedures above, such that we're ultimately interested in solving the matrix eigenvalue problem

$$\mathbf{A}\mathbf{k} = \lambda_k \mathbf{k} = \lambda_k \mathbb{1}_3 \mathbf{k}, \tag{5.2}$$

where $\mathbb{1}_3$ is the 3-by-3 identity matrix. This is equivalent to finding the eigenvalue λ_k and corresponding eigenvector $\mathbf{k}$ that satisfies

$$(\mathbf{A} - \lambda_k \mathbb{1}_N)\,\mathbf{k} = 0. \tag{5.3}$$

This equation is satisfied by values of λ_k that satisfy the **characteristic equation** of the matrix $\mathbf{A}$, which is given by

$$\det(\mathbf{A} - \lambda_k \mathbb{1}_3) = 0. \tag{5.4}$$

Once we solve for the values of λ_k that satisfy the characteristic equation, we can use them to find the eigenvectors they correspond to.

We won't belabor determinant theory here. We write the determinant of a matrix using vertical lines, and 2-by-2 and 3-by-3 matrices are obtained according to the following equations:

$$\det(\mathbf{M}_2) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{5.5}$$

$$\det(\mathbf{M}_3) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh \tag{5.6}$$

There is a systematic procedure for obtaining higher-order determinants, but we do not include it here.

Let's illustrate this by solving for the eigensystem of the following 3-by-3 matrix:

$$\mathbf{A} = \begin{bmatrix} 5 & 2i & 0 \\ -2i & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}. \tag{5.7}$$

We'll break the process of solving the eigensystem into a few steps:

1. Determine the characteristic equation of the matrix $\mathbf{A}$. We do this using Eq. (5.4),

which gives

$$\begin{aligned}\det(\mathbf{A}-\lambda_k\mathbb{1}_3) &= \det\left(\begin{bmatrix}5 & 2i & 0\\ -2i & 1 & 0\\ 0 & 0 & -3\end{bmatrix} - \lambda\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\right)\\ &= \begin{vmatrix}5-\lambda_k & 2i & 0\\ -2i & 1-\lambda_k & 0\\ 0 & 0 & -3-\lambda_k\end{vmatrix}\\ &= (5-\lambda_k)(1-\lambda_k)(-3-\lambda_k) - (2i)(-2i)(-3-\lambda_k)\\ &= (-3-\lambda_k)\left(\lambda_k^2-6\lambda_k+1\right) \end{aligned} \tag{5.8}$$

2. Find the roots of the characteristic equation. The roots of the equation are the values of λ_k that make the equation equal to zero. We can further factorize the equation determined above to write

$$(-3-\lambda_k)(\lambda_k-2)(\lambda_k+3)=0. \tag{5.9}$$

In this form, it is clear that the roots of the equation are

$$\lambda_k = 2,-3,-3. \tag{5.10}$$

Make note of the fact that the value of -3 appears twice. We call the redundancy of this eigenvalue **degeneracy**, saying things like "the eigenvalue -3 is two-fold degenerate."

3. Solve for the eigenvectors for each eigenvalue. We do this by plugging each of the eigenvalues into the characteristic equation in turn, labeling the basis vectors $|e_1\rangle$, $|e_2\rangle$, and $|e_3\rangle$ and labeling the eigenvectors $|v_1\rangle$, $|v_2\rangle$, and $|v_3\rangle$.

$\lambda_k = 2$

The eigenvalue equation for this eigenvalue is

$$\mathbf{A}|e_1\rangle = 2|e_1\rangle \rightarrow (\mathbf{A}-2)|e_1\rangle = 0. \tag{5.11}$$

In expanded form, this is written

$$\begin{aligned}\begin{bmatrix}5-2 & 2i & 0\\ -2i & 1-2 & 0\\ 0 & 0 & -3-2\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ c_3\end{bmatrix} &= \begin{bmatrix}0\\ 0\\ 0\end{bmatrix}\\ \begin{bmatrix}3 & 2i & 0\\ -2i & -1 & 0\\ 0 & 0 & -5\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ c_3\end{bmatrix} &= \begin{bmatrix}0\\ 0\\ 0\end{bmatrix}. \end{aligned} \tag{5.12}$$

If we perform the matrix multiplication defined from this equation, this results in the following system of equations

$$\begin{cases}3c_1+2ic_2 &= 0\\ -2ic_1-c_2 &= 0\\ -5c_3 &= 0.\end{cases} \tag{5.13}$$

From these equations, we must have $c_3 = 0$. We can obtain the other two coefficients by solving

$$\begin{aligned} 3c_1 + 2ic_2 &= -2ic_1 - c_2 \\ (3+2i)\,c_1 &= (-1-2i)\,c_2 \\ c_1 &= -\frac{1+2i}{3+2i}c_2 \\ &= -\frac{7+4i}{13}c_2. \end{aligned} \tag{5.14}$$

This equation defines the proportionality of c_1 and c_2. Their exact values are arbitrary, and we will choose them according to the requirement that the resulting eigenvector is normalized. Assuming our initial basis is orthonormal, the normalization condition is

$$\begin{aligned} |c_1|^2 + |c_2|^2 &= 1 \\ \left|-\frac{7+4i}{13}c_2\right|^2 + |c_2|^2 &= 1 \\ \left(\frac{65}{169} + 1\right) c_2^2 &= 1 \\ c_2 &= \pm\frac{13}{3\sqrt{26}} \end{aligned} \tag{5.15}$$

This means that we have

$$c_1 = \mp\frac{7+4i}{3\sqrt{26}}. \tag{5.16}$$

After all of that (I know it was a lot), we have determined that the normalized eigenvector that has eigenvalue 2 is

$$|v_1\rangle = \mp\frac{7+4i}{3\sqrt{26}}\,|e_1\rangle \pm \frac{13}{3\sqrt{26}}\,|e_2\rangle + 0\,|e_3\rangle\,. \tag{5.17}$$

The absolute sign of the coefficients is undefined by the matrix algebra. In physics, this absolute sign is referred to as a "phase," and it is normally the case that phase is arbitrary, physically.

$\lambda_k = -3$

Repeating this procedure for the two-fold degenerate eigenvalue $\lambda_k = 3$, we obtain the following matrix equation:

$$\begin{bmatrix} 8 & 2i & 0 \\ -2i & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.18}$$

The matrix multiplication for this system gives

$$\begin{cases} 8c_1 + 2ic_2 & = 0 \\ -2ic_1 + 4c_2 & = 0 \\ 0 & = 0 \end{cases} \tag{5.19}$$

It may seem odd that we wrote the final equation this way, but this is what the matrix algebra gives us, and it indicates something important: the vector $(c_1, c_2, c_3) = (0, 0, c_3)$ will satisfy this equation for *any* finite value of c_3. If we want the vector to be normalized, then we'll choose $c_3 = 1$. Hence, one of the eigenvectors with eigenvalue -3 is simply

$$|v_3\rangle = 0\,|e_1\rangle + 0\,|e_2\rangle + 1\,|e_3\rangle\,. \tag{5.20}$$

We can then solve for the other eigenvector, where we have

$$\begin{aligned} 8c_1 + 2ic_2 &= -2ic_1 + 4c_2 \\ (8+2i)\,c_1 &= (4-2i)\,c_2 \\ c_1 &= \frac{4-2i}{8+2i}c_2 \\ &= \frac{7-6i}{17}c_2 \end{aligned} \tag{5.21}$$

Then the normalization condition gives

$$\begin{aligned} \left|\frac{7-6i}{17}\right|^2 c_2^2 + c_2^2 &= 1 \\ \left(\frac{85}{289}+1\right)c_2^2 &= 1 \\ c_2 &= \pm\frac{289}{\sqrt{374}}. \end{aligned} \tag{5.22}$$

Hence, the other eigenvector with eigenvalue -3 is

$$|v_2\rangle = \pm\frac{119-102i}{\sqrt{374}}\,|e_1\rangle \pm \frac{289}{\sqrt{374}}\,|e_2\rangle + 0\,|e_3\rangle \tag{5.23}$$

The Final Results:

The eigensystem of **A** is as follows:

Eigenvalue	Eigenvector(s)
2	$\|v_1\rangle = \mp\frac{7+4i}{3\sqrt{26}}\,\|e_1\rangle \pm \frac{13}{3\sqrt{26}}\,\|e_2\rangle$
-3	$\|v_2\rangle = \pm\frac{119-102i}{\sqrt{374}}\,\|e_1\rangle \pm \frac{289}{\sqrt{374}}\,\|e_2\rangle$ $\|v_3\rangle = \|e_3\rangle$

Because the matrix **A** that we started with was Hermitian, these eigenvectors must be mutually orthogonal. The final orthogonal must be orthogonal to the first two because the basis vectors form an orthogonal set. It can be shown that the first two are orthogonal. (If they aren't, then I made a mistake in the diagonalization process.)

5.2.3 Expansion in a basis of eigenvectors

Solving for the eigensystem of an operator or matrix is sometimes referred to as **diagonalizing** the operator or matrix because an operator is diagonal in the basis of its eigenvectors. Suppose we have a 3-by-3 matrix $\mathbf{A}$ that has 3 eigenvectors with eigenvalues λ_1, λ_2, and λ_3 with corresponding eigenvectors $|\lambda_1\rangle$, $|\lambda_2\rangle$, and $|\lambda_3\rangle$, respectively. In the basis of its eigenvectors, we know we can write the matrix representation of $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} | & | & | \\ \hat{A}\,|\lambda_1\rangle & \hat{A}\,|\lambda_2\rangle & \hat{A}\,|\lambda_3\rangle \\ | & | & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{5.24}$$

where the second equality follows from the definition of an eigenvector, given in Eq. (5.1). Hence, we see the sense in which solving for the eigensystem diagonalizes the operator $\hat{A}$.

The logic of this goes both ways. If you are given (or determine) the matrix representation of some operator, and you see that it is a diagonal matrix, then the eigenvalues are the diagonal elements of the matrix and the eigenvectors are the basis vectors that define the matrix representation.

We can also make powerful statements about matrices that are **block diagonal**. Specifically, the eigenvalues (eigenvectors) of a block diagonal matrix can be obtained as the eigenvalues (eigenvectors) of each of the blocks individually.

In the case worked above, notice that the matrix takes the form of a 2-by-2 block followed by a 1-by-1 block. This form means that we can diagonalize each block individually. In particular, the diagonalizing the 1-by-1 block is trivial, and it has the eigenvalue -3 and the eigenvector $|v_3\rangle$, which, again, is what we saw above.

5.3 Functions of matrices

The function of an operator (or matrix) is defined in terms of the Taylor series expansion of the function:

$$f(\mathbf{A}) = \sum_{n=0}^{\infty} f^{(n)}(0)\frac{\mathbf{A}^n}{n!}. \tag{5.25}$$

The power of a matrix $\mathbf{A}$ is not a trivial thing to compute by hand unless the matrix $\mathbf{A}$ is a diagonal matrix. Powers of a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \tag{5.26}$$

are given by the power of the elements of the matrix, *viz.*

$$\mathbf{D}^n = \begin{bmatrix} (D_1)^n & 0 & 0 \\ 0 & (D_2)^n & 0 \\ 0 & 0 & (D_3)^n \end{bmatrix}. \tag{5.27}$$

You can convince yourself of this by evaluating the case of a smaller matrix, like $\mathbf{D}^2$ for a 3-by-3 matrix, or by rigorously proving it with methods we discuss below.

In any case, the importance of diagonal representations of matrices highlights the utility of the eigenbasis of an operator, because we have shown above that an operator is diagonal in its eigenbasis.

We can obtain the same result a different way by expanding a vector of interest $|\psi\rangle$ in the eigenbasis of the operator $\mathbf{A}$ as we have done for other bases previously. In the general case of an N-dimensional vector space, we can write

$$|\psi\rangle = \sum_k \lambda_k |k\rangle , \tag{5.28}$$

where the kets $\{|k\rangle\}$ are eigenvectors of the operator $\hat{A}$ with eigenvalues $\{\lambda_k\}$. Then the action of $\hat{A}$ on $|\psi\rangle$ is given by

$$\hat{A}^n |\psi\rangle = \sum_k c_k \hat{A}^n |k\rangle . \tag{5.29}$$

Because $|k\rangle$ is an eigenvector of $\hat{A}$,

$$\hat{A}^n |k\rangle = \lambda_k \hat{A}^{n-1} |k\rangle = \lambda_k^2 \hat{A}^{n-2} |k\rangle = \cdots = \lambda_k^n |k\rangle . \tag{5.30}$$

Hence, we have

$$\hat{A}^n |\psi\rangle = \sum_k \lambda_k^n |k\rangle . \tag{5.31}$$

In terms of evaluating a function of an operator, then, we have

$$\begin{aligned} f(\mathbf{A}) |\psi\rangle &= \sum_k^N c_k \sum_{n=0}^{\infty} f^{(n)}(0) \frac{1}{n!} \mathbf{A}^n |k\rangle \\ &= \sum_k^N c_k \left(\sum_{n=0}^{\infty} f^{(n)}(0) \frac{1}{n!} \lambda_k^n \right) |k\rangle \\ &= \sum_k^N c_k f(\lambda_k) |k\rangle . \end{aligned} \tag{5.32}$$

This is a compact expression that we can evaluate for all sorts of functions of operators, like $\exp(\mathbf{A})$, $\mathbf{A}^{-1}$, and others. This expression *only* holds for the eigenbasis of $\hat{A}$. Because of this, **it is incredibly common that you will need to change to the eigenbasis of the operator of interest in order to make any progress.** This is particularly true for problems involving time evolution (how a system changes with time), where the operator of interest is the Hamiltonian.

5.4 Summation notation

Our final topic in linear algebra is a little bit more technical, and it develops the tools necessary to do general proofs in linear algebra. We have, by and large, steered clear of proofs in the bootcamp curriculum, choosing to focus our time on underlying concepts and the mechanics of problem solving. Still, you will need to use linear algebra prove things in

your quantum mechanics coursework, and summation notation is the way to prove things in the general case in linear algebra.

This section takes the algorithms we commonly use to compute matrix/vector operations and formalizes them into **summation notation.** It can be tricky to get used to at first, and I recommend working out each of the operations below for small (2-dimensions is fine!) matrices and vectors to get the hang of what the expressions are saying.

The Kronecker delta. The Kronecker delta (δ_{ij}) is defined according to

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \tag{5.33}$$

This term comes up a lot when we are working with diagonal matrices, where we might write something like

$$D_{ij} = d_i \delta_{ij}, \tag{5.34}$$

to indicate that the element at the ith row and the jth column of the diagonal matrix $\mathbf{D}$ is zero unless the row number and the column number are the same (hence we only need one index, i in this case, to specify the value of the elements $\{d_i\}$.

It also comes up when we work with in an orthonormal basis, which we do often. In fact, it is actually common to write the condition of orthonormality with a Kronecker delta. If a set of basis vectors $\{|\phi_i\rangle\}$ is orthonormal, then the inner product

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \tag{5.35}$$

must hold for all possible values of i and j in the basis. In essence, this equation says that the value of the inner product is 1 if the two basis vectors in the inner product are the same (the basis is normalized) and that the value of the inner product is 0 if the two basis vectors are different (all pairs of vectors in the basis are orthogonal).

Mechanically, we can use the Kronecker delta to simplify and eliminate various sums. As a contrived example, suppose we have the sum over two indices i and j given below:

$$\sum_{ij} a_i b_j \delta_{ij}. \tag{5.36}$$

Because of the definition of the Kronecker delta, we will multiply the expression being summed by zero for all of the terms where $i \neq j$. The only terms that "survive" the Kronecker delta are those where $i = j$, and we can simplify the sum as

$$\sum_{ij} a_i b_j \delta_{ij} = \sum_{i} a_i b_i. \tag{5.37}$$

The upshot is this: when we sum over both indices of a Kronecker delta, we can replace the sum over two indices with a sum over one index, replacing the label of one of the indices with the other.

Inner product. The inner product of two N-dimensional vectors $\mathbf{v}$ and $\mathbf{w}$ is given by

$$\langle v|w\rangle = \sum_{i}^{N} v_i^* w_i. \tag{5.38}$$

Matrix–vector multiplication. For the matrix equation

$$\mathbf{A}\mathbf{v} = \mathbf{w},$$

where $\mathbf{A}$ is a matrix with elements A_{ij} at the ith row and the jth column and $\mathbf{v}$ and $\mathbf{w}$ are vectors with elements v_i and w_i, the ith element of $\mathbf{w}$ can be obtained as

$$w_i = \sum_{j} A_{ij} v_j. \tag{5.39}$$

Matrix–matrix multiplication. Likewise, if we want to multiply two matrices $\mathbf{A}$ and $\mathbf{B}$ together to obtain a third matrix $\mathbf{C}$ according to

$$\mathbf{A}\mathbf{B} = \mathbf{C}, \tag{5.40}$$

we can obtain each of the elements C_{ij} as

$$C_{ij} = \sum_{k} A_{ik} B_{kj}. \tag{5.41}$$

It can be difficult to remember which index should be summed over while you are getting used to this notation. It can be helpful to remember that indices that are adjacent to each other in the expression are usually summed over, while those on the outside of expressions are not.

Matrix trace. The trace of an N-by-N square matrix $\mathbf{A}$ is the sum of its diagonal elements, which we write as

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i} A_{ii}. \tag{5.42}$$

Example 5.1: Prove that the kth power of an N-by-N diagonal matrix $\mathbf{D}$ can be obtained by raising each of the elements of $\mathbf{D}$ to the kth power.

5.5 Connections to physical chemistry

Eigensystems are relevant to physical chemistry and quantum mechanics because in the latter we postulate that the measurement of some observable will always take on the value of one of the eigenvalues of operator that corresponds to this observable. Furthermore, we postulate that the state of the system after measurement corresponds to the eigenvector/eigenspace

associated with the eigenvalue that was measured. In this way, eigensystems show up any time we are interested in the properties of a quantum mechanical system—which is all the time! We will also commonly transform a vector into the eigenbasis of a particular operator in order to make our mathematical manipulations more straightforward.

Functions of matrices become an important topic when we start to use the Schrödinger equation to consider the time evolution of a quantum mechanical system. We will need to exponentiate the Hamiltonian matrix in order to consider this, and this is done using the Taylor series expansion that was described in Section. 5.3 above.

As indicated in the notes, summation notation is used to prove things in linear algebra, and by extension in quantum mechanics as well as statistical mechanics, although to a more limited extent. Facility with these concepts will enable you to better understand and work through important derivations in your coursework and in research.

5.6 Example problem solutions

Example 5.1: *Prove that the kth power of an N-by-N diagonal matrix $\mathbf{D}$ can be obtained by raising each of the elements of $\mathbf{D}$ to the kth power.*

We start with the example where $k = 2$ and then generalize from there. The ij'th element of the square of the matrix $\mathbf{D}$ is given by

$$(\mathbf{D}^2)_{ij} = \sum_k D_{ik} D_{kj}. \tag{5.43}$$

Because $\mathbf{D}$ is a diagonal matrix, we must have

$$D_{ij} = \delta_{ij} D_{ii}. \tag{5.44}$$

Hence, we can simplify the preceding expression to

$$(\mathbf{D}^2)_{ij} = \sum_k \delta_{ij}\delta_{kj} D_{ii} D_{kk} = D_{ii} D_{ii} = D_{ii}^2, \tag{5.45}$$

because the two Kroenecker delta's make every time in the sum where $k \neq i$ vanish. Based on induction, we can see that this result generalizes to any integral value of k. To see this, we can write

$$\begin{aligned}
(\mathbf{D}^3)_{ij} &= (\mathbf{D}^2 \cdot \mathbf{D})_{ij} \\
&= \sum_k (\mathbf{D}^2)_{ik} D_{kj} \\
&= \sum_k D_{ii}^2 D_{kj} \\
&= \sum_k D_{ii}^2 \delta_{kj} D_{kk} \\
&= D_{ii}^3
\end{aligned} \tag{5.46}$$

In words, we have shown that the kth power of an N-by-N diagonal matrix can be obtained by raising each of the elements of the matrix to the kth power for $k = 2$, and that the fact that this is true for k implies that it is true for $k + 1$. Hence, it must be true for all $k > 2$.

6 Fourier analysis

Topics: Function spaces, complex-valued functions, Fourier Series, Fourier transform

6.1 Motivation and approach

A fundamental step towards using math for describing physical and chemical systems is the clever use of **representations**. How do we easily describe the distribution of interstitial defects in a perovskite lattice? How do we represent the pulse of electromagnetic energy that is a laser hitting a sample? These are complicated functional fields, namely, a spatially dependent interstitial defect concentration $c(\mathbf{r})$ or a space and time dependent electric field $\mathbf{E}(\mathbf{r}, t)$. But leveraging periodic properties of a crystal lattice or of oscillatory electromagnetic waves, these complicated functions can be *represented* in a more simplified form, sometimes with only a few numbers (as in, constants)! The study of Fourier analysis gives us the tools to perform exactly this representation.

Fourier series is a way to represent periodic functions by using a countably infinite basis of sines and cosines. Fourier transform is a way to represent arbitrary functions using an uncountably infinite basis of free waves. Both are direct applications of linear algebra towards representing functions in a special way so that the information is densely contained in a few numbers or in simpler functions.

6.2 Recap about function spaces

- A function can be expanded in terms of basis vectors in the function space, for example, in terms of polynomials. A primary example is a power series expansion,

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots \tag{6.1}$$

- A Taylor expansion is a power series expansion around a given point x_0,

$$f(x) = f(x_0) + f^{'}(x_0)(x - x_0) + f^{''}(x_0)\frac{(x - x_0)^2}{2!} + \dots \tag{6.2}$$

  Putting $x_0 = 0$ gives back the original power series and lets us identify $a_0 = f(0), a_1 = f^{'}(0)$, *etc.* A Taylor expansion will terminate after a finite number of terms if $f(x)$ is a polynomial.

- Uniqueness: If two power series are equal over a range, they are equal at every term.

$$\sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} b_i x^i \text{ for some } u \leq x < v \text{ implies } a_i = b_i, \text{ for all } i. \tag{6.3}$$

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

6.3 Complex-valued functions

The Taylor expansions of the exponential, the sine and the cosine around $x = 0$ (*i.e.*, the power series expansions) are

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \tag{6.4}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \tag{6.5}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots \tag{6.6}$$

This helps us arrive at

$$e^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - ...\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - ...\right) = \cos x + i \sin x. \tag{6.7}$$

We define the **hyperbolic functions** $\sinh x$ and $\cosh x$ as

$$\cosh x = \frac{e^x + e^{-x}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots \tag{6.8}$$

$$\sinh x = \frac{e^x - e^{-x}}{2} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \tag{6.9}$$

A convenient way to recall the definition of the hyperbolic functions is the fact that a sine is an odd function and so is the hyperbolic sine, and a cosine is an even function and so the hyperbolic cosine.

6.4 Fourier series

Conceptual video: But what is a Fourier series? From heat flow to circle drawings by 3Blue1Brown (24:46)

A Fourier series is a **representation** of a **periodic** function using a periodic sinusoidal basis set. A function $g(x)$ periodic over $x \in [-\pi, \pi)$ can be expressed in terms of sines and cosines—this is called a Fourier series:

$$g(x) = \frac{a_0}{2} + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \dots \tag{6.10}$$

The expansion coefficients are obtained exactly similar to a basis set expansion of a vector in linear algebra. The definition of an inner product for periodic functions over $[-\pi, \pi)$ is an integral over that range. The sines and cosines are pairwise orthogonal for this definition of the inner product. Hence the coefficients can be solved by multiplying the Fourier series expansion on both sides by one of the basis functions ($\sin(mx)$ or $\cos(mx)$ for some integer m) and integrating over $[-\pi, \pi)$ such that all but one of the terms becomes zero. The

orthogonality relations for the Fourier basis are as follows for $m, n \neq 0$,

$$\int_{-\pi}^{\pi} dx \sin(mx)\sin(nx) = \begin{cases} \pi & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} \tag{6.11}$$

$$\int_{-\pi}^{\pi} dx \cos(mx)\cos(nx) = \begin{cases} \pi & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} \tag{6.12}$$

$$\int_{-\pi}^{\pi} dx \sin(mx)\cos(nx) = 0 \tag{6.13}$$

$$\int_{-\pi}^{\pi} dx \sin(mx) = \int_{-\pi}^{\pi} dx \cos(mx) = 0. \tag{6.14}$$

Because sines are odd functions and cosines are even functions, any periodic odd function has only sines in its Fourier series, and similarly any periodic even function has only cosines and the constant term.

Often we define an aperiodic function over a finite domain and define a periodic version of that function such that the domain is repeated many times. For example, if a function is defined as

$$f(x) = \begin{cases} -1 & \text{if } -\pi \leq x < 0 \\ 1 & \text{if } 0 \leq x < \pi. \end{cases} \tag{6.15}$$

and then $f(x)$ is defined to be periodic over $[-\pi, \pi)$, the new function has alternating values of -1 and 1 for every interval of length π. These functions can then be expanded in a Fourier Series.

Example 6.1: Find the Fourier series expansion of the function $f(x)$ periodic over $[-\pi, \pi)$ defined as

$$f(x) = \begin{cases} -1 & \text{if } -\pi \leq x < 0 \\ 1 & \text{if } 0 \leq x < \pi. \end{cases} \tag{6.16}$$

Using this expansion, find the sum of the infinite series

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots \tag{6.17}$$

Note that any finite domain $[a, b)$ over the real line can be mapped to $[-\pi, \pi)$ by scaling and shifting, so in effect any function defined over a finite domain can have a valid Fourier series.

An alternate definition for a Fourier series is

$$g(x) = c_o + (c_1 e^{ix} + c_{-1} e^{-ix}) + (c_2 e^{2ix} + c_{-2} e^{-2ix}). \tag{6.18}$$

This is slightly more general because it can handle a complex $g(x)$.

6.5 Representing operators in a Fourier series

Conceptual video: Abstract vector spaces by 3Blue1Brown (16:45)

An **operator** transforms a function into another function. Examples are the derivative operator $\frac{d}{dx}*$ that returns the derivative, and the multiplicative operator $x.*$ that multiplies a function by x and returns the product function.

Some operators like multiplication, taking derivatives or integrals, are **linear**, hence they can be expressed as matrices in function spaces. And hence in a Fourier Series, just like functions can be *represented* as vectors, operators can be *represented* as matrices.

For example, consider the Fourier space built with the basis vectors comprising of the complex exponentials in order $\{1, e^{ix}, e^{-ix}, e^{2ix}, e^{-2ix}, \ldots\}$. Define the multiplicative operator $\hat{O} = \cos x.* = \frac{e^{ix}+e^{-ix}}{2}.*$ where the asterisk $*$ denotes the position where the input function is multiplied. Examining the action on the basis gives

$$\cos x.1 = \frac{e^{ix} + e^{-ix}}{2} \tag{6.19}$$

$$\cos x.e^{ix} = \frac{e^{2ix} + 1}{2} \tag{6.20}$$

$$\cos x.e^{-ix} = \frac{1 + e^{-2ix}}{2} \tag{6.21}$$

$$\cos x.e^{2ix} = \frac{e^{3ix} + e^{ix}}{2} \tag{6.22}$$

$$\cos x.e^{-2ix} = \frac{e^{-ix} + e^{-3ix}}{2} \tag{6.23}$$

If the matrix representation is denoted by the boldface $\mathbf{O}$, the action of the matrix multiplication must conform to the above coefficients, such that

$$\mathbf{O}\begin{pmatrix}1\\0\\0\\0\\0\\\vdots\end{pmatrix} = \begin{pmatrix}0\\1/2\\1/2\\0\\0\\\vdots\end{pmatrix}, \quad \mathbf{O}\begin{pmatrix}0\\1\\0\\0\\0\\\vdots\end{pmatrix} = \begin{pmatrix}1/2\\0\\0\\1/2\\0\\\vdots\end{pmatrix}, \quad \mathbf{O}\begin{pmatrix}0\\0\\1\\0\\0\\\vdots\end{pmatrix} = \begin{pmatrix}1/2\\0\\0\\0\\1/2\\\vdots\end{pmatrix}, \quad \textit{etc.} \tag{6.24}$$

For these relations to be valid, these transformed vectors would have to form the columns of the matrix representation. So the multiplication operator in this space is represented by

$$\mathbf{O} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & \cdots \\ 1/2 & 0 & 0 & 1/2 & 0 & \cdots \\ 1/2 & 0 & 0 & 0 & 1/2 & \cdots \\ 0 & 1/2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1/2 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{6.25}$$

Example 6.2: Write the matrix form of the derivative operator $\frac{d}{dx}*$ and the integral operator $\int_0^x dx.*$ in this complex Fourier space.

6.6 Fourier transform

Conceptual video: But what is the Fourier Transform? by 3Blue1Brown (19:42)

A Fourier transform is a **linear** operator that relates two arbitrary (periodic or aperiodic) functions $f(t)$ and $g(\omega)$ through the relation

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt. \tag{6.26}$$

This operator is invertible through the relation

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega) e^{-i\omega t} d\omega. \tag{6.27}$$

t and ω have the conventional interpretations of time and frequency.

When talking about the positional space function $f(\mathbf{r})$ and its Fourier transformed function $g(\mathbf{k})$, $\mathbf{r}$ is said to exist in the **real space** and $\mathbf{k}$ is said to exist in the **Fourier space** or **reciprocal space**.

Fourier transforms are thus a way to *represent* any arbitrary function $f(t)$ as an expansion over the uncountably infinite basis set of free waves $e^{i\omega t}$ for all real values of ω. Any function can deconstructed in terms of these waves, stored and manipulated as necessary at specific frequencies, and then the original information can be extracted back through the inverse Fourier transform.

Example 6.3: Obtain an FT of $f(t) = e^{-\alpha|t|}$ for $\alpha > 0$. How does the resultant function behave when α is changed?

Example 6.4: Obtain both the Fourier series (defined with complex exponentials) and the Fourier transform for the function $f(t) = \sin \omega_0 t + \sin 2\omega_0 t$. Notice a pattern?

6.7 Connections to physical chemistry

A Fourier series is often used to understand standing waves (because the wave has to be defined within a closed periodic interval), examples being vibrational modes of solids and Bloch waves of mobile electrons inside a crystal. In electrical engineering, Fourier series is the way to analyze periodic signals and waveforms.

Fourier transforms are used to describe free-form (unconfined) waves and wave packets. Any scattering, diffraction and absorption spectrum involves interaction of matter with unconfined photons, hence the language of the Fourier transform must be used there. Spectroscopic information is often analyzed with the machinery of Fourier transforms.

6.8 Additional resources

Mathematical Methods for Physicists by Arfken, Weber and Harris- Chapters 19 and 20 (e-copy accessible through campus vpn).

6.9 Example problem solutions

Example 6.1: *Find the Fourier series expansion of the function $f(x)$ periodic over $[-\pi, \pi)$ defined as*

$$f(x) = \begin{cases} -1 & \text{if } -\pi \leq x < 0 \\ 1 & \text{if } 0 \leq x < \pi. \end{cases} \tag{6.28}$$

Using this expansion, find the sum of the infinite series

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots \tag{6.29}$$

We have defined the function $f(x)$ as

$$f(x) = \begin{cases} -1 & \text{if } -\pi \leq x < 0 \\ 1 & \text{if } 0 \leq x < \pi. \end{cases} \tag{6.30}$$

We write it in a Fourier series expansion as

$$f(x) = \frac{a_0}{2} + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \dots \tag{6.31}$$

The goal is to find these expansion coefficients. Integrating both sides of Eq. (6.31) from $-\pi$ to π gives,

$$\int_{-\pi}^{0} (-1)dx + \int_{0}^{\pi} (1)dx = \frac{a_0}{2} \int_{-\pi}^{\pi} dx + a_1 \int_{-\pi}^{\pi} \cos x dx + b_1 \int_{-\pi}^{\pi} \sin x dx + \dots \tag{6.32}$$

$$\implies -\pi + \pi = \frac{a_0}{2}.2\pi + a_1.0 + a_2.0 + \dots \tag{6.33}$$

$$\implies 0 = a_0 \tag{6.34}$$

where the relations in Eq. (6.14) have been used. In fact, since the function $f(x)$ is an odd function, *i.e.*, $f(-x) = -f(x)$, all integrals involving even basis functions will be 0, such that

$$\int_{-\pi}^{\pi} f(x)dx = \int_{-\pi}^{\pi} f(x) \cos x dx = \int_{-\pi}^{\pi} f(x) \cos 2x dx = \dots = 0. \tag{6.35}$$

Hence $a_0 = a_1 = a_2 = \dots = 0$.
For finding the b_m coefficients, we have to multiply the expansion in Eq. (6.31) on both sides

with $\sin mx$ and integrate from $-\pi$ to π. For b_1,

$$\int_{-\pi}^{0}(-\sin x)dx + \int_{0}^{\pi}(\sin x)dx = b_1\int_{-\pi}^{\pi}\sin^2 x dx + b_2\int_{-\pi}^{\pi}\sin 2x\sin x dx + \ldots \tag{6.36}$$

$$\implies [\cos x]_{-\pi}^{0} + [-\cos x]_{0}^{\pi} = b_1.\pi + b_2.0 + b_3.0 + \ldots \tag{6.37}$$

$$\implies \frac{4}{\pi} = b_1 \tag{6.38}$$

where the relations in Eqs. (6.11) have been used.
For finding b_2,

$$\int_{-\pi}^{0}(-\sin 2x)dx + \int_{0}^{\pi}(\sin 2x)dx = b_1\int_{-\pi}^{\pi}\sin x\sin 2x dx + b_2\int_{-\pi}^{\pi}\sin^2 2x dx + \ldots \tag{6.39}$$

$$\implies \frac{1}{2}[\cos 2x]_{-\pi}^{0} + \frac{1}{2}[-\cos 2x]_{0}^{\pi} = b_1.0 + b_2.\pi + b_3.0 + \ldots \tag{6.40}$$

$$\implies 0 = b_2 \tag{6.41}$$

Similarly for finding b_3,

$$\int_{-\pi}^{0}(-\sin 3x)dx + \int_{0}^{\pi}(\sin 3x)dx = b_1\int_{-\pi}^{\pi}\sin x\sin 3x dx + b_2\int_{-\pi}^{\pi}\sin 2x\sin 3x dx + b_3\int_{-\pi}^{\pi}\sin^2 3x dx + \ldots \tag{6.42}$$

$$\implies \frac{1}{3}[\cos 3x]_{-\pi}^{0} + \frac{1}{3}[-\cos 3x]_{0}^{\pi} = b_1.0 + b_2.0 + b_3.\pi + \ldots \tag{6.43}$$

$$\implies \frac{4}{3\pi} = b_3 \tag{6.44}$$

We see a pattern that the only non-zero coefficients are

$$b_{2n+1} = \frac{4}{(2n+1)\pi} \qquad \text{for } n = 0, 1, 2, \ldots \tag{6.45}$$

Hence the Fourier series expansion for the total function is

$$f(x) = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{\sin((2n+1)x)}{(2n+1)}. \tag{6.46}$$

For finding the sum of the series

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots, \tag{6.47}$$

we need to replace x in the Fourier series in Eq. (6.46) with $x = \pi/2$. This gives $f(x) = 1$ on the LHS, and rearranging both sides gives

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - ... = \frac{\pi}{4}. \tag{6.48}$$

Example 6.2: *Write the matrix form of the derivative operator* $\frac{d}{dx}*$ *and the integral operator* $\int_0^x dx.*$ *in this complex Fourier space.*

The matrix forms of the operators will become clear if we compute its action on a few of the basis vectors. Examining the action of the derivative operator,

$$\frac{d}{dx}1 = 0 \tag{6.49}$$

$$\frac{d}{dx}e^{\pm ix} = \pm ie^{ix} \tag{6.50}$$

$$\frac{d}{dx}e^{\pm mix} = \pm mie^{mix} \tag{6.51}$$

So $\frac{d}{dx}*$ is a diagonal matrix in the complex Fourier space $\{1, e^{ix}, e^{-ix}, e^{2ix}, e^{-2ix}, \ldots\}$ with diagonal elements being $\{0, i, -i, 2i, -2i, \ldots\}$.

Examining the action of the integral operator,

$$\int_0^x 1dx = x \tag{6.52}$$

$$\int_0^x e^{\pm ix}dx = \frac{e^{\pm ix}-1}{\pm i} = \left(\mp\frac{1}{i}\right)1 + \left(\pm\frac{1}{i}\right)e^{\pm ix} \tag{6.53}$$

$$\int_0^x e^{\pm mix}dx = \frac{e^{\pm mix}-1}{\pm mi} = \left(\mp\frac{1}{mi}\right)1 + \left(\pm\frac{1}{mi}\right)e^{\pm mix}. \tag{6.54}$$

We see that even though the integration operator maps the basis functions $\{e^{\pm mix}\}$ into a linear combination of themselves and 1, the basis function 1 (*i.e.*, e^{0ix}) is not mapped into a periodic function anymore. So the integral operator as defined is **not well-represented** inside this Fourier space. In order for the integral operator to be well-represented, the functions would have to be always bound within a finite domain, say $[-\pi, \pi)$, such that aperiodic functions like x can also be treated periodically by repetitions of the finite domain.

Example 6.3: *Obtain an FT of* $f(t) = e^{-\alpha|t|}$ *for* $\alpha > 0$. *How does the resultant function behave when* α *is changed?*

Using the definition of the Fourier transform in Eq. (6.26) we get

$$g(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\alpha|t|}e^{i\omega t}dt = \frac{1}{\sqrt{2\pi}}\left\{\int_{-\infty}^{0} e^{\alpha t+i\omega t}dt + \int_0^{\infty} e^{-\alpha t+i\omega t}dt\right\} \tag{6.55}$$

$$= \frac{1}{\sqrt{2\pi}}\left\{\left[\frac{e^{(\alpha+i\omega)t}}{\alpha+i\omega}\right]_{-\infty}^{0} + \left[\frac{e^{(-\alpha+i\omega)t}}{-\alpha+i\omega}\right]_0^{\infty}\right\} \tag{6.56}$$

$$= \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{\alpha+i\omega} - \frac{1}{-\alpha+i\omega}\right\} \tag{6.57}$$

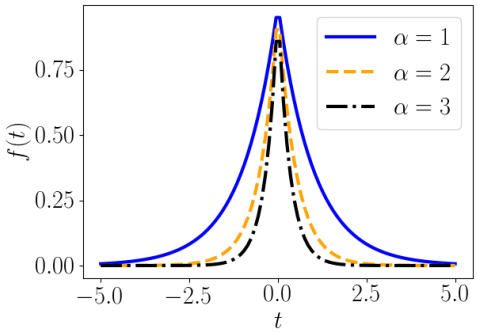


Figure 1: Exponential excitation in time

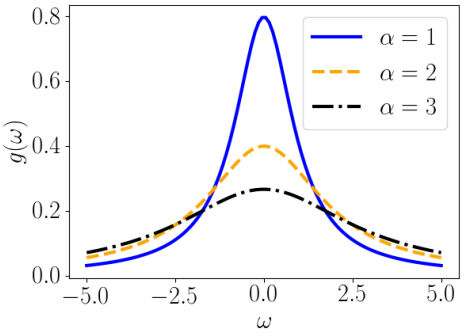


Figure 2: Lorentzian lineshape in frequency

where because $\alpha > 0$, $\exp(\alpha t + i\omega t) \to 0$ when $t \to -\infty$ and $\exp(-\alpha t + i\omega t) \to 0$ when $t \to \infty$, both becoming exponentials with large negative real parts in the exponent. Simplifying this gives the FT as

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \frac{2\alpha}{\alpha^2 + \omega^2} \tag{6.58}$$

This function has a **Lorentzian lineshape** in the frequency space. This lineshape often observed in spectroscopic signatures of real-time electronic or vibrational excitations. α is analogous to a **lifetime** of an excitation that decays exponentially with time. Plotting $f(t)$ and $g(\omega)$ for a few values of α gives us a better idea of the exact dependence. The **Full Width at Half Max (FWHM)** of a lineshape is defined as the width of the frequency-space peak at half its maximum value. From Eq. (6.58), the peak occurs at $\omega = 0$ and the peak value is $g(\omega) = \frac{1}{\sqrt{2\pi}} \frac{2}{\alpha}$. The function attains half that value at $\omega = \pm\alpha$, *i.e.*, $g(\pm\alpha) = \frac{1}{\sqrt{2\pi}} \frac{1}{\alpha}$. Hence the FWHM of the frequency space peak is 2α.
This means that with increasing α, the excitation decays faster in real-time (lifetime decreases), consequently, $g(\omega)$ becomes less sharply peaked. Notice that there is an inverse relation between the real space and Fourier space, such that when the function becomes sharply peaked in real-time, it becomes less sharply peaked in Fourier space.

Example 6.4: *Obtain both the Fourier series (defined with complex exponentials) and the Fourier transform for the function $f(t) = \sin \omega_0 t + \sin 2\omega_0 t$. Notice a pattern?*

The Fourier series can be written down by writing the sine functions using complex exponentials as

$$f(t) = \sin \omega_0 t + \sin 2\omega_0 t \tag{6.59}$$

$$= \frac{e^{i\omega_0 t} - e^{-i\omega_0 t}}{2i} + \frac{e^{2i\omega_0 t} - e^{-2i\omega_0 t}}{2i} \tag{6.60}$$

$$= \left(\frac{1}{2i}\right) e^{i\omega_0 t} + \left(-\frac{1}{2i}\right) e^{-i\omega_0 t} + \left(\frac{1}{2i}\right) e^{2i\omega_0 t} + \left(-\frac{1}{2i}\right) e^{-2i\omega_0 t}. \tag{6.61}$$

Hence the Fourier series has four non-zero terms when expressed with the sequence of complex exponentials $e^{im\omega_0 t}$ for integer m.

The Fourier transform is given by

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt. \tag{6.62}$$

Writing $f(t)$ as sum of four complex exponentials will simplify the problem as

$$\begin{aligned} g(\omega) = \frac{1}{\sqrt{2\pi}} \Bigg[& \frac{1}{2i} \int_{-\infty}^{\infty} e^{i(\omega_0+\omega)t} dt - \frac{1}{2i} \int_{-\infty}^{\infty} e^{i(-\omega_0+\omega)t} dt \\ & + \frac{1}{2i} \int_{-\infty}^{\infty} e^{i(2\omega_0+\omega)t} dt - \frac{1}{2i} \int_{-\infty}^{\infty} e^{i(-2\omega_0+\omega)t} dt \Bigg]. \end{aligned} \tag{6.63}$$

Each of the exponentials are rapidly oscillating with t and integrate to 0, unless the exponent is itself 0. This is a manifestation of the fact that complex exponentials integrate to Dirac delta functions that you will encounter in Section 10.3.2, and as explained below.

Dirac delta functions from complex exponentials:

The Dirac delta function comes out neatly from the general definition of the Fourier transform and its inverse, as follows:

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega') e^{-i\omega' t} d\omega' \right] e^{i\omega t} dt \tag{6.64}$$

$$= \int_{-\infty}^{\infty} g(\omega') \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega-\omega')t} dt \right] d\omega'. \tag{6.65}$$

The definition of the Dirac delta function $\delta(\omega - \omega')$ is through the integral identity that for any function $g(\omega)$ and any variables ω and ω',

$$g(\omega) = \int_{-\infty}^{\infty} g(\omega') \delta(\omega - \omega') d\omega'. \tag{6.66}$$

Comparing Eq. (6.66) with Eq. (6.65), we find

$$\int_{-\infty}^{\infty} e^{i(\omega-\omega')t} dt = 2\pi \delta(\omega - \omega'). \tag{6.67}$$

Applying this result to Eq. (6.63) gives us the Fourier transform of $\sin \omega_0 t + \sin 2\omega_0 t$ as

$$g(\omega) = \frac{\sqrt{2\pi}}{2i} [\delta(\omega + \omega_0) - \delta(\omega - \omega_0) + \delta(\omega + 2\omega_0) - \delta(\omega - 2\omega_0)]. \tag{6.68}$$

Hence the Fourier transform consists of four Dirac delta functions located at $\pm\omega_0$ and $\pm 2\omega_0$, just as the Fourier series comprised of four corresponding non-zero terms. This emphasizes that the Fourier series and the Fourier transform are inherently related. In fact the Fourier series expansion is a special case of the Fourier transform. The Fourier transform encodes

the amount of periodicity embedded in any arbitrary periodic or aperiodic function $f(t)$. In case the function does have a specific period (like $2\pi/\omega_0$ in our example), a Fourier series expansion is possible and it will involve only those terms where the Fourier transform is non-zero.

7 Functions and approximations

Topics: infinite series, Taylor expansions, finite difference, plotting

7.1 Motivation and approach

Chemistry and physics are messy. They are full of complicated problems and equations that are sometimes (literally) impossible to solve. In these situations, approximation methods can be incredibly useful. They can be used to obtain an estimate of your answer, which can often be quite accurate, or to build intuition for how a function might behave in certain limits, like at high temperatures or short times.

The following section introduces some of the most common methods, such as Taylor expansions and finite differences for approximating functions, and provides some tips and tricks for using graphing and plots to illustrate how these approximation methods work.

7.2 Infinite series

Conceptual video: Partial sums intro by Khan Academy (6:08)
Conceptual video: Infinite series as a limit of partial sums by Khan Academy (4:48)
Conceptual video: Convergent and divergent sequences by Khan Academy (4:59)

An **infinite series** is a sum formally consisting of an infinite number of terms. If there is a sequence of infinite terms $a_1, a_2, a_3, \ldots$, then the $\boldsymbol{n^{th}}$ **partial sum** is given by

$$s_n = \sum_{i=1}^{n} a_i \,. \tag{7.1}$$

If the partial sums converge to a finite value (*i.e.* $\lim_{n\to\infty} s_n = S$), then the infinite series $\sum_{i=1}^{\infty} a_i$ is said to be **convergent** and equal to S. A necessary (but not sufficient) criteria for convergence is that $\lim_{i\to\infty} a_i = 0$.

Other series are **divergent** if their sequences of partial sums approach $\pm\infty$ or are **oscillatory** if they oscillate between two values. In general, there are many different tests for convergence.

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

7.3 Taylor series and Taylor expansions

Conceptual and technical video: Taylor series by 3Blue1Brown (22:19)

A **Taylor expansion** is a power series representation of a function $f(x)$ that has a continuous n^{th} derivative on an interval $a \leq x \leq b$. The Taylor series of $f(x)$ around a point a is

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f'''(a) + \dots \tag{7.2}$$

$$= \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a)\,, \tag{7.3}$$

which gives the value of the function at x in terms of the value of the function and its derivatives at a reference point a.

In practice, Taylor expansions are often truncated at a finite number of terms, and the remainder after the n^{th} term is given by

$$R_n(x) = f(x) - \sum_{m=0}^{n} \frac{(x-a)^m}{m!} f^{(m)}(a)\,. \tag{7.4}$$

Truncated expansions are often identified by the order of the last term in the approximation. For example, an approximation of $e^x \approx 1 + x$ is often referred to as an "expansion to the first order" (since the last term is the one corresponding to $n = 1$) or a "linear approximation" (since the last term is linear in x).

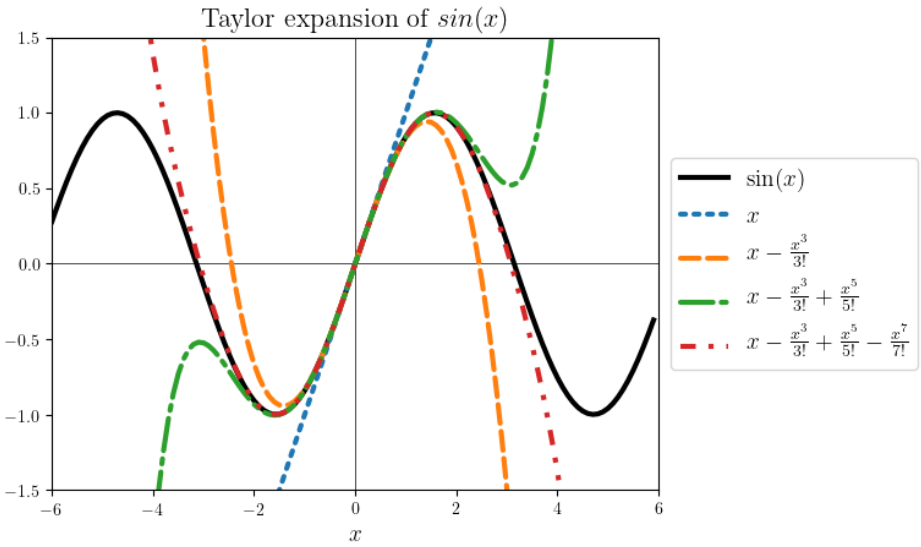
If we are approximating a function at a value x that is close to our reference point a, then $(x-a)$ is small, and powers of that difference (*i.e.* $(x-a)^n$) will shrink quickly as n increases. Therefore, we will likely be able to get a good approximation of $f(x)$ using a relatively small number of terms in our Taylor expansion. The further away from our reference point a we want to approximate $f(x)$, the more terms we will need in our expansion. This fact motivates how we might choose our reference point a.

The power series representation of a function is unique. Some examples of common Taylor series expansions are:

- $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$
- $\ln(1+x) = x - \frac{x^2}{2!} + \frac{x^3}{3!} - \dots = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n!}$
- $(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \dots = \sum_{n=0}^{\infty} \binom{m}{n} x^n$
  - This is known as the **binomial expansion**.
- $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$

Example 7.1: Determine the series expansion for $f(x) = \cos x$ by performing a Taylor expansion around the point $a = 0$.

Below is a graphical representation of the Taylor expansion of the sine function around $x = 0$. We see that, even the lowest order term, x, approximates the sine function quite well around $x = 0$. Higher order expansions provide more accurate estimates further away from our reference point.



Power series can be differentiated or integrated term by term. Also, two power series can be added, subtracted, multiplied, or divided (so long as the denominator is not zero at $x = 0$).

7.4 Finite difference methods

Recall that the **derivative** of a function is defined as

$$\frac{df(x)}{dx} = \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon} . \tag{7.5}$$

If ε is small, then we can approximate the derivative using the **forward finite difference**:

$$\frac{df(x)}{dx} \approx \frac{f(x+\varepsilon) - f(x)}{\varepsilon} . \tag{7.6}$$

We could also use the **backward finite difference**:

$$\frac{df(x)}{dx} \approx \frac{f(x) - f(x-\varepsilon)}{\varepsilon} \tag{7.7}$$

or the **central finite difference**:

$$\frac{df(x)}{dx} \approx \frac{f(x+\varepsilon) - f(x-\varepsilon)}{2\varepsilon} , \tag{7.8}$$

and we could use a similar approach to approximate the second derivative using, for example, the **second central difference**:

$$\frac{d^2 f(x)}{dx^2} \approx \frac{f(x+\varepsilon) - 2f(x) + f(x-\varepsilon)}{\varepsilon^2} . \tag{7.9}$$

7.4.1 Finite difference method for differential equations

Technical video: Finite differences introduction by Commutant (6:48)

If given a differential equation $f'(x)$ and an initial condition for the function $f(x_0)$, one can use the finite differences to approximate the function $f(x)$ by iteratively solving for points.

To do this, one can rearrange the forward finite difference equation to obtain

$$f'(x) = \frac{f(x+\varepsilon) - f(x)}{\varepsilon} \tag{7.10}$$

$$\varepsilon f'(x) = f(x+\varepsilon) - f(x) \tag{7.11}$$

$$f(x+\varepsilon) = f(x) + \varepsilon f'(x)\,. \tag{7.12}$$

This is called the **Euler formula**. Given a small enough step size ε and knowledge of the function and derivative evaluated a certain point, one can use this formula to approximate the function at a nearby point.

Other numerical schemes that use higher-order finite differences can be used for more accurate approximations .

Example 7.2: Given the differential equation $f'(x) = 2x$ and the initial condition $f(0) = 1$, use the Euler formula to approximate the function near $x = 0$. Use steps of $\varepsilon = 0.1$ and 0.5. Plot both of those approximations as well as the analytical solution on the same graph.

7.4.2 Stationary points

For a function $f = f(x_1, x_2, \ldots, x_n)$ defined in n dimensions, a **stationary point** $\boldsymbol{x}$ of f is one for which $\frac{\partial f}{\partial x_i}$ evaluated at $\boldsymbol{x}$ is equal to 0 for all $i = 1, \ldots, n$.

The stationary point is a **minimum** if $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2} > 0$ for all i, a **maximum** if $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i^2} < 0$ for all i, or otherwise a **saddle point** (*i.e.* it is a minimum in some directions and a maximum in others).

The **Hessian** is a square matrix of second-order partial derivatives of a function f, where the i, j matrix element is defined as

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}\,. \tag{7.13}$$

If the Hessian evaluated at a point $\boldsymbol{x}$ has all positive eigenvalues, then $\boldsymbol{x}$ is a minimum of f. If it has all negative eigenvalues, then $\boldsymbol{x}$ is a maximum, and if it has both positive and negative eigenvalues, then $\boldsymbol{x}$ is a saddle point.

Stationary points can be useful reference points when performing Taylor expansions.

7.5 Plotting functions

Conceptual and technical video: Solving linear systems by graphing by Khan Academy (8:30)

Sometimes, you are presented with a system of equations that are challenging to solve. Using a graphing tool, like the desmos graphing calculator, can allow you to quickly plot the equations and visually identify solutions as the intersection points.

7.5.1 Identifying functions by graphing

Often, you will encounter a function or data set to which you are trying to fit or assign a functional form. Plotting the function or data points is a good starting point, and changing the scale of the x and/or y axis to a log-scale (instead of a linear-scale) might be helpful.

First, let us review some useful properties of logarithms, regardless of the base, where a and b are constants:

- $\log(ab) = \log(a) + \log(b)$
- $\log(a/b) = \log(a) - \log(b)$
- $\log\left(a^b\right) = b\log(a)$
- $\log(1) = 0$

Additionally, for the natural log, $\ln(e) = 1$.

An **exponential function** has the general form:

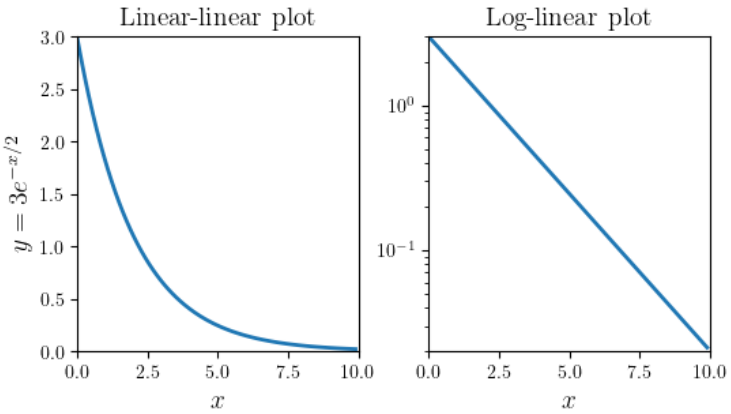
$$y = ae^{bx}\,. \tag{7.14}$$

Taking the natural log (ln) of both sides and simplifying gives:

$$\ln(y) = \ln\left(ae^{bx}\right) = \ln(a) + \ln\left(e^{bx}\right) = \ln(a) + bx\ln(e) \tag{7.15}$$

$$= \ln(a) + bx\,. \tag{7.16}$$

Therefore, if you plot an exponential function with the y-axis on a natural log scale, it will appear to have a linear relationship with the slope corresponding to the exponential factor b. Note that if you plot the function on a $\log_{10}$ scale, it will appear to have a linear relationship, but the slope will not correspond directly to b. Below is an illustration of the function $y = 3e^{-1/2}$ plotted on a linear-linear scale (left) and a log-linear scale (right).



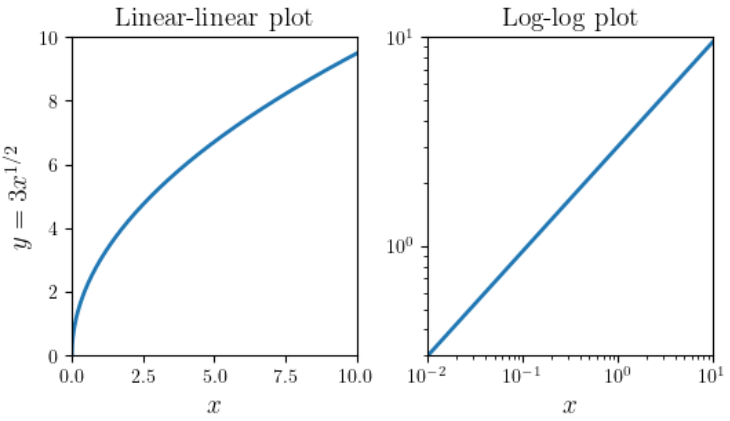
A **power law** has the general form:

$$y = ax^b \,. \tag{7.17}$$

Taking the log of both sides and simplifying gives:

$$\log(y) = \log\big(ax^b\big) = \log(a) + \log\big(x^b\big) \tag{7.18}$$
$$= \log(a) + b\log(x) \,. \tag{7.19}$$

In this case, plotting the power law with both the x and y axes on a log scale (of any base), it will appear to have a linear relationship with the slope corresponding to the power b. Below is an illustration of the function $y = 3x^{1/2}$ plotted on a linear-linear scale (left) and a log-log scale (right).



7.6 Connections to physical chemistry

The methods outlined here are common throughout physics and chemistry. For instance, functions of matrices or operators are defined in terms of their series expansions. One example is the quantum mechanical time-evolution operator, which describes how a quantum state evolves until time t. The time-evolution operator is given by:

$$e^{-iHt/\hbar} = 1 - \frac{i}{\hbar}Ht - \frac{1}{2\hbar^2}H^2t^2 - \dots \,, \tag{7.20}$$

where H is the Hamiltonian operator. If you want to evolve a state for a short time so that t is small, you may approximate the time-evolution by truncating the above expansion to first order, so that $e^{-iHt/\hbar} \approx 1 - iHt/\hbar$.

Additionally, Taylor expansions are a framework for thinking about **perturbation theory**, a method often used in the physical sciences to approximate the solution of a difficult problem through the use of a reference problem that is simple to solve. For example, consider the behavior of a particle under the effect of a weak force f. If you already know the equation for how the particle behaves without any force, you can add a term linear in the force (*i.e.* proportional to f) to approximate this more complex problem. Since the force is

weak, terms that are quadratic or higher in the force (*i.e.* proportional to $f^2, f^3, \dots$) can be ignored. This approximation to first order is called **linear response theory**.

The finite difference method for solving differential equations is used frequently, as many differential equations are too difficult or impossible to solve analytically. This method is also often used to numerically propagate an object in time, such as evolving a wavefunction in time according to the Schrödinger equation, its equation of motion.

Plotting can always be an extremely useful tool – when you are trying to understand how a function behaves with a changing parameter, when you are trying to make sense of your experimental data, and more.

7.7 Additional text resources

Mathematical Methods for Physicists (7th Edition) by Arkfen, Weber, and Harris: Sections 1.1, 1.2, 1.3, 1.5, 1.6, 1.8, and 1.9

7.8 Example problem solutions

Example 7.1: *Determine the series expansion for $f(x) = \cos x$ by performing a Taylor expansion around the point $a = 0$.*

To evaluate the Taylor expansion of $\cos(x)$ around the point $a = 0$, we will need to find the first few derivatives of $\cos(x)$ with respect to x and then evaluate them at $a = 0$:

$$\cos(x)\Big|_0 = 1 \tag{7.21}$$

$$\frac{d}{dx}\cos(x)\Big|_0 = -\sin(x)\Big|_0 = 0 \tag{7.22}$$

$$\frac{d^2}{dx^2}\cos(x)\Big|_0 = -\frac{d}{dx}\sin(x)\Big|_0 = -\cos(x)\Big|_0 = -1 \tag{7.23}$$

$$\frac{d^3}{dx^3}\cos(x)\Big|_0 = -\frac{d}{dx}\cos(x)\Big|_0 = \sin(x)\Big|_0 = 0 \tag{7.24}$$

$$\frac{d^4}{dx^4}\cos(x)\Big|_0 = \frac{d}{dx}\sin(x)\Big|_0 = \cos(x)\Big|_0 = 1\,. \tag{7.25}$$

Putting this all together gives

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots\,. \tag{7.26}$$

Examining these first few terms and recognizing a pattern, we can write this in summation notation as:

$$\cos(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}\,. \tag{7.27}$$

Example 7.2: *Given the differential equation $f'(x) = 2x$ and the initial condition $f(0) = 1$, use the Euler formula to approximate the function near $x = 0$. Use steps of $\varepsilon = 0.1$ and 0.5. Plot both of those approximations as well as the analytical solution on the same graph.*

We start with plugging in $f'(x) = 2x$ into Eq. (7.12), which gives:

$$f(x + \varepsilon) = f(x) + 2x\varepsilon \,. \tag{7.28}$$

Now, we can use our initial condition $f(x = 0) = 1$ to iteratively evaluate $f(x)$ away from $x = 0$ in steps of ε:

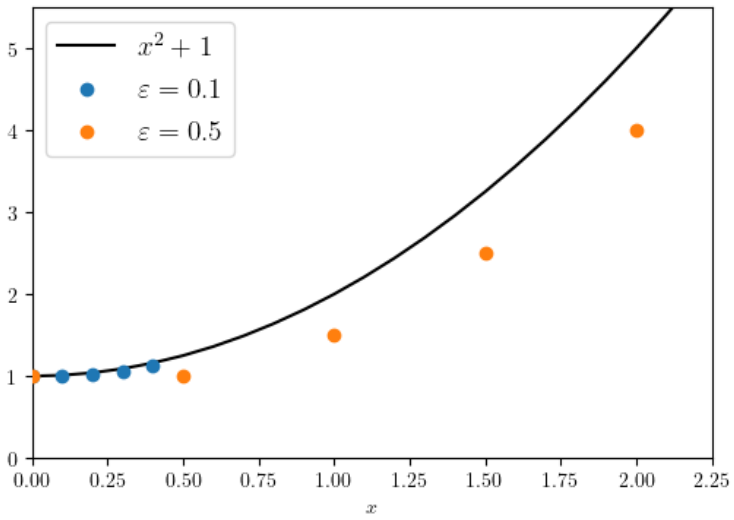
$$f(\varepsilon) = f(0) + 2 \cdot 0 \cdot \varepsilon = 1 \tag{7.29}$$
$$f(2\varepsilon) = f(\varepsilon) + 2 \cdot \varepsilon \cdot \varepsilon = 1 + 2\varepsilon^2 \tag{7.30}$$
$$f(3\varepsilon) = f(2\varepsilon) + 2 \cdot 2\varepsilon \cdot \varepsilon = 1 + 6\varepsilon^2 \tag{7.31}$$
$$f(4\varepsilon) = f(3\varepsilon) + 2 \cdot 3\varepsilon \cdot \varepsilon = 1 + 12\varepsilon^2 \,. \tag{7.32}$$

Analytically, we can directly integrate $f'(x)$ and then use in our initial condition to find $f(x)$:

$$f(x) = \int f'(x)dx = \int 2xdx = x^2 + C \tag{7.33}$$
$$f(0) = C = 1 \tag{7.34}$$
$$f(x) = x^2 + 1 \,. \tag{7.35}$$

We can plot the analytical solution along with our numerical estimates using step sizes of $\varepsilon = 0.1$ and 0.5. We see that, for both numerical estimates, the points closer to our initial condition are more accurate. Additionally, using a smaller step size provides a better estimate of the analytical solution.

If you want to learn how you can use Python as a tool to solve this problem and plot the solutions, see the next section!


$x^2 + 1$
$\varepsilon = 0.1$
$\varepsilon = 0.5$
0
1
2
3
4
5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
x

8 Differential equations: Numerical solutions

Topics: Power series and Fourier series solutions, mapping into linear algebra problems

8.1 Motivation and approach

In this section we will discuss ways to solve differential equations using two different types of series, Fourier and power. Aside from needing to know how to do this for some of your quantum mechanics homework, differential equations that are most easily solved by assuming a series solutions appear in several areas of science. For example, any time you have a differential equation whose solution you know is periodic (most types of spectroscopy), a good place to start solving for that function is by assuming it takes the form of a Fourier series.

8.2 Series solutions to ODEs

8.2.1 Power series

Here are a few resources with more complete derivations and extra practice problems:

- Paul's online notes for Power Series
- Paul's online notes for series solutions of ODE's
- Conceptual video: Video from the Organic Chemistry Tutor on representing functions as a power series (53:44)
- Technical video: solving a differential equation with a power series (18:28)

  This video is a bit higher level than what we will go over in the boot camp, but still a good video for the curious!

The power series problems in this section took inspiration from the above resources and this online text book.
Generally a **power series** is any series that can be expressed as

$$\sum_{n=0}^{\infty} c_n (x-a)^n. \tag{8.1}$$

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

The c_n's are coefficients for each term and a is the number you are expanding the series about. When we assume that the solution to our ODE takes this form we will know what we are expanding about (a) from the information given in the problem, but we will still need to solve for the coefficients. As you might imagine, solving for infinite coefficients would be tedious if it was even possible, so in most cases you will only need to fully solve for a few coefficients in the series. This can be for three different reasons:

1. After a certain point extra terms do not improve your result in any appreciable way
2. The coefficients after a certain point are provably zero
3. You discover a **recursion relationship** that relates the next coefficient you are trying to solve for one or several of the previous coefficients.

Like a Taylor series, the function is most accurate around the point you are expanding about (a), so if you remain near that region and don't need an amazingly accurate result, there's a good chance that even the eleventh term of the series will not change any of the significant figures you need, though this is dependent on the series. This would be a case where you just truncate the series after a certain point.

Power series have some interesting convergence properties. Power series have a **radius of convergence**, denoted R in which the series will converge. The series will converge if $|x-a| < R$ and will diverge if $|x-a| > R$. If $|x-a| = R$ nothing definitive can be said about whether the function will converge or diverge. If you would like to learn more about the radius of convergence or would like some practice determining the radius of convergence, Paul's online notes on power series has some great example problems.

For this material it's often easiest to learn by example, so let's start by using a power series to find the solution to $\frac{d^2y}{dx^2} - y = 0$.

First, we assume that the ODE can be solved with a power series. While it is annoying to start trying to solve for something by assuming your solution will work, power series often work well for ODE's so it's a good place to start. Differentiating a power series also just gives another power series (as we will see) so in cases like this where we have a second derivative of our function in our ODE, assuming this form can make it notationally simple to work through. Also, if you try it out and it doesn't work, you will have learned that your solution cannot be written as a power series, which will still be useful. So let's say that

$$y(x) = \sum_{n=0}^{\infty} c_n x^n. \tag{8.2}$$

Notice here that $a = 0$, this is out of convenience more than anything else. Math problems will usually tell you what to set a to, for problems you'll encounter in your physical chemistry courses the value of a will either be told to you or you will be able to determine a good value based on the surrounding information of the problem. For example, if you have an ODE describing electronic motion and you knew you wanted to know what the electron was doing in the region of 5 to 7, you could pick 6 as your value of a.

Next we will plug the power series into the ODE:

$$\frac{d^2}{dx^2}\left(\sum_{n=0}^{\infty} c_n x^n\right) - \sum_{n=0}^{\infty} c_n x^n = 0 \tag{8.3}$$

$$\frac{d}{dx}\left(\sum_{n=1}^{\infty} n c_n x^{(n-1)}\right) - \sum_{n=0}^{\infty} c_n x^n = 0 \tag{8.4}$$

$$\sum_{n=2}^{\infty} n(n-1) c_n x^{n-2} - \sum_{n=0}^{\infty} c_n x^n = 0 \tag{8.5}$$

To better see how taking a derivative leads to the new expression (note the change in the starting point for the sum!) we recommend writing out the first few terms of the series and taking the derivative and then putting it back into summation notation.

We can now match powers in the sums to determine what the coefficients are since we know that the whole equation must be zero. So we could say that $2c_2 + c_0 = 0$, $6c_3 = 0 + c_1 = 0$, *etc.* (I have just taken the coefficients from each sum that are attached to the same power of x and set them equal to zero.) Or I could change the index of one of the sums so that the sums range over the same values:

$$\sum_{n=0}^{\infty} n(n-1) c_n x^{n-2} - \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} (n+2)(n+1) c_{n+2} x^n - \sum_{n=0}^{\infty} c_n x^n \tag{8.6}$$

$$= \sum_{n=0}^{\infty} \left((n+2)(n+1) c_{n+2} - c_n\right) x^n \tag{8.7}$$

Once again, you should write out the first few of the terms of the sum to better see that the above is true. Now I can much more easily see that $(n+2)(n+1)c_{n+2} - c_n = 0$, or $c_{n+2} = \frac{c_n}{(n+2)(n+1)}$. Here we have found a **recursion relationship**. If I know c_0 and c_1 I know all of the coefficients in my series because of the above relation. c_0 and c_1 will need to be given to you to find numerical solutions to the coefficients. For example, if $c_0 = 2$ and $c_1 = 1$ I can find that $c_2 = 1$, $c_3 = \frac{1}{6}$, etc. using the recursion relation.

Example 8.1: Find a series solution around $x = 0$ for $y'' - xy = 0$. The initial conditions are that $y(0) = 1$ and $y'(0) = 2$.

8.2.2 Fourier Series

Here are a few resources with more explanations and practice problems:

- Wolfram MathWorld notes on Fourier Series
- Paul's online notes for Fourier Series
- Math24 extra practice for solving differential equations with Fourier series

Fourier series are amazing for representing **periodic functions** though they can be used to represent all types of continuous functions. The series is composed of sines and cosines to replicate the repeating behavior and the orthogonality relationships between trigonometric functions makes Fourier series particularly useful in certain types of problems. (See Wolfram notes linked above.)

A general Fourier series can be written as

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos\Big(\frac{n\pi x}{L}\Big) + \sum_{n=1}^{\infty} b_n \sin\Big(\frac{n\pi x}{L}\Big). \tag{8.8}$$

Where a_0, a_n, b_n are coefficients and L is related to the length of the region you are finding the series representation for. L will almost always be related to the periodicity of your function if your function is periodic. If your function is not periodic, it is just the region that your function is defined over. The reason why a_0 and a_n are both labeled with a's while the sine function coefficients are b's is because if the sums instead started from zero, the zeroth term would be zero for the sums over the sine functions, but the zeroth term of the cosine sum would jut be a constant, a_0. This is why you might see the general Fourier series written without a separate constant term.

To express any general function as a Fourier series all you need to do is find the coefficients which can be done using the following formulas:

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx \tag{8.9}$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\Big(\frac{n\pi x}{L}\Big)dx \tag{8.10}$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\Big(\frac{n\pi x}{L}\Big)dx \tag{8.11}$$

The derivation of these equations is outside of the scope of the boot camp, but the *Paul's online notes for Fourier series*, linked above, does go through this if you are curious. Also note that the Wolfram notes on Fourier series assume that L is equal to π, which is why their equations look a bit different, we are using the most general forms for the equations here. Note that these equations still assume that your integrating region is centered at zero. However, typically when we are solving a differential equation we are trying to solve for $f(x)$, we don't already know it, so then how do we get those coefficients?

Once again, to learn by example let's try to find the Fourier series solution to $y''+2y = 3x$ over the region $x = -1$ to $x = 1$ with the boundary conditions of $y(0) = y(1) = 0$. This problem was taken from this differential equations resource.

We will begin by assuming that our solution can be represented as a Fourier series and then substituting the general form into our differential equation. Before starting that substitution it will be useful to realize that:

$$\frac{d^2}{dx^2}\cos\Big(\frac{n\pi x}{L}\Big) = -\Big(\frac{n\pi}{L}\Big)^2\cos\Big(\frac{n\pi x}{L}\Big) \tag{8.12}$$

$$\frac{d^2}{dx^2}\sin\Big(\frac{n\pi x}{L}\Big) = -\Big(\frac{n\pi}{L}\Big)^2\sin\Big(\frac{n\pi x}{L}\Big) \tag{8.13}$$

Also, the right hand side of the equation, which I will denote as RHS from now on, is not written in terms of sines and cosines, which will make it difficult to match terms after doing the substitution in the ODE. So before worrying about the left hand side (LHS) of the ODE, let's first determine the Fourier series representation of $3x$:

$$a_0 = \frac{1}{2}\int_{-1}^{1} 3x dx = 0 \tag{8.14}$$

$$a_n = \int_{-1}^{1} 3x\cos(n\pi x)dx = 0 \tag{8.15}$$

$$b_n = \int_{-1}^{1} 3x\sin(n\pi x)dx = 2\int_{0}^{1} 3x\sin n\pi x dx = 6\frac{\sin(n\pi) - n\pi\cos(n\pi)}{n^2\pi^2} \tag{8.16}$$

Where we have used the above equations with $L = 1$, that integrating odd functions over a region symmetric about zero gives zero, and that integrating even functions over a region symmetric about zero is equal to twice the integral over half that region. The solution we found for b_n looks complicated, but it can be simplified quite a bit. Remembering that n is always a positive integer allows us to replace $\sin(n\pi)$ with 0 and $\cos(n\pi)$ with $(-1)^n$. So then

$$b_n = \frac{6(-1)^{n+1}}{n\pi} \tag{8.17}$$

$$3x = \sum_{n=1}^{\infty} \frac{6(-1)^{n+1}}{n\pi}\sin(n\pi x) \tag{8.18}$$

If you want to graphically see if this Fourier series really does represent $3x$ over the region of $-1 \leq x \leq 1$, see the first problem in the *Extra examples* section.

Back to the actual ODE we are trying to find a solution for, assume that our solution has the form:

$$y(x) = a_0 + \sum_{n=1}^{\infty} a_n\cos(n\pi x) + \sum_{n=1}^{\infty} b_n\sin(n\pi x) \tag{8.19}$$

$$\tag{8.20}$$

Then

$$y" + 2y = 3x \tag{8.21}$$

$$-\sum_{n=1}^{\infty} a_n(n\pi)^2\cos(n\pi x) - \sum_{n=1}^{\infty} b_n(n\pi)^2\sin(n\pi x) + 2\sum_{n=1}^{\infty} a_n\cos(n\pi x) \tag{8.22}$$

$$+ 2\sum_{n=1}^{\infty} b_n\sin(n\pi x) = \sum_{n=1}^{\infty} \frac{6(-1)^{n+1}}{n\pi}\sin(n\pi x) \tag{8.23}$$

To go from the first equation the assumed form of $y(x)$ was substituted into the ODE. Next we will start matching terms, we notice that the RHS of the equation has no constant terms, so we know that $a_0 = 0$. The RHS also has no terms that depend on $\cos(n\pi x)$, so we know

that $a_n = 0$ for all n. The coefficients for each sine mode must be equal to each other, so we can further match terms to end up with the relationship:

$$b_n \left[2 - (n\pi)^2\right]] = \frac{6(-1)^{n+1}}{n\pi} \tag{8.24}$$

$$b_n = \frac{6(-1)^{n+1}}{n\pi(2 - n^2\pi^2)} \tag{8.25}$$

So then

$$y(x) = \sum_{n=1}^{\infty} \frac{6(-1)^{n+1}}{n\pi(2 - n^2\pi^2)} \sin(n\pi x) \tag{8.26}$$

Notice that we didn't use the boundary conditions at all but still found a numerical solution for the coefficients? That will not always happen, we just got lucky here. It's still good to check that our boundary conditions are met though:

$$y(0) = \sum_{n=1}^{\infty} \frac{6(-1)^{n+1}}{n\pi(2 - n^2\pi^2)} \sin(n\pi * 0) = 0 \tag{8.27}$$

$$y(1) = \sum_{n=1}^{\infty} \frac{6(-1)^{n+1}}{n\pi(2 - n^2\pi^2)} \sin(n\pi) = 0 \tag{8.28}$$

Example 8.2: Find the solution to the differential equation $y' + 2y = \sin(x) + x^2 + 1$ over the range $-\pi \leq x \leq \pi$.

8.2.3 Summary

Here is the general procedure for using a series to solve a differential equation:

1. Assume your solution takes the form of a power/Fourier series.
2. Determine what the value you are expanding about (a) is if you are using a power series. Often this will be zero. If you are using a Fourier series, determine the region your function is defined over (L).
3. Plug your series into your ODE and take any derivatives/do any arithmetic you can.
4. Combine sums by changing the range of the sum, look for patterns, match terms, do any simplifying you can.
5. If boundary conditions or information about the first few coefficients in your power series is given, use those to solve for the coefficients completely. If those are not given, you can give your final answers in terms of c_n's

8.3 Connections to physical chemistry

The wave equation, heat equation, and Laplace's equation are all examples of differential equations that are most easily solved by assuming the solution takes the form of a Fourier series. The wave equation comes up regularly in quantum mechanics, the heat equation in statistical mechanics, and Laplace's equation is useful in studying fluid dynamics and heat conduction.

Bessel's equation is an example of when power series solutions become useful. Bessel's equation is routinely encountered when solving differential equations in spherical or cylindrical coordinate systems, you might encounter it when finding solutions to the radial Schrodinger equation for a free particle.

8.4 Extra examples

Example 8.3: Solve the ODE $xy' + x^2 \sin x = y$ by assuming the solution takes the form of a power series. Assume that you are interested in the region about $x = 0$. It might be useful to look at the Taylor expansion of sine about x=0 (the Maclaurin series of sine), is that a power series?

Example 8.4: Solve the ODE $y'' + y' + 4y = \sin(x) + \cos(2x)$ by assuming the solution takes the form of a Fourier series. The solution is periodic and repeats every 2π. This problem was taken from this differential equations resource.

Example 8.5: Connection to Python:
To see just how well Fourier series can represent functions and to get more practice plotting with Python, plot the Fourier series we found for $3x$ we found earlier in the notes along with $3x$ on the same plot. Make sure to give your plot a title, label your axes, and add a legend to your plot. Try using the following number of terms in the Fourier series: (a) 5 (b) 10 (c) 50. You should make use of functions in the creation of your Fourier series.
The solution can be found in `DiffEq2_9_5_solution.ipynb` on bCourses.

8.5 Example solutions

Example 8.1: *Find a series solution around $x = 0$ for $y'' - xy = 0$. The initial conditions are that $y(0) = 1$ and $y'(0) = 2$.*

Start by assuming your solution takes the form

$$y = \sum_{n=0}^{\infty} c_n x^n. \tag{8.29}$$

In order to plug this back into the ODE we need to know its second derivative,

$$y'' = \sum_{n=2}^{\infty} n(n-1) c_n x^{n-2}. \tag{8.30}$$

Using this in combination with the ODE gives us

$$\sum_{n=2}^{\infty} n(n-1) c_n x^{n-2} - \sum_{n=0}^{\infty} c_n x^{n+1} = 0. \tag{8.31}$$

From here all we have to do is change the ranges of the sums until we find a way to simplify this equation into a single summation:

$$\sum_{n=2}^{\infty} n(n-1) c_n x^{n-2} - \sum_{n=0}^{\infty} c_n x^{n+1} \tag{8.32}$$

$$= \sum_{n=0}^{\infty} (n+2)(n+1) c_{n+2} x^n - \sum_{n=0}^{\infty} c_n x^{n+1} \tag{8.33}$$

$$= \sum_{n=0}^{\infty} (n+2)(n+1) c_{n+2} x^n - \sum_{n=1}^{\infty} c_{n-1} x^n \tag{8.34}$$

$$= 2c_2 + \sum_{n=1}^{\infty} [(n+2)(n+1) c_{n+2} - c_{n-1}] x^n \tag{8.35}$$

In the last line we took out the first term of the first sum in order to have the sums start from the same place. Since the right hand side of the equation is just zero we know that

$$c_2 = 0 \tag{8.36}$$

$$c_{n+2} = \frac{c_{n-1}}{(n+2)(n+1)}. \tag{8.37}$$

Using our initial conditions of $y(0) = 1$ and $y'(0) = 2$, we can then find values for c_0 and c_1, which will then allow us to evaluate all the remaining terms via the recursion relationship above:

$$y(0) = c_0 = 1 \tag{8.38}$$

$$y'(0) = c_1 = 2 \tag{8.39}$$

$$c_2 = 0 \tag{8.40}$$

$$c_3 = \frac{c_0}{(3)(2)} = \frac{1}{6} \tag{8.41}$$

$$c_4 = \frac{c_1}{(4)(3)} = \frac{2}{12} \tag{8.42}$$

$$c_5 = \frac{c_2}{(5)(4)} = 0 \tag{8.43}$$

$$\cdots \tag{8.44}$$

Example 8.2: *Find the solution to the differential equation* $y' + 2y = \sin(x) + x^2 + 1$ *over the range* $-\pi \leq x \leq \pi$. *Assume that your solution can be written as a Fourier series.*

Instead of starting with the form of $y(x)$, since we know that our solution will take the form of a Fourier series we know that we need to rewrite x^2 in terms of sums of sines and cosines. Let's assume that

$$x^2 = c_0 + \sum_{n=1}^{\infty} c_n \cos(nx) + \sum_{n=1}^{\infty} d_n \sin(nx). \tag{8.45}$$

We can use equations 9.9-9.11 to find the forms for the coefficients:

$$c_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2 dx = \frac{\pi^2}{3} \tag{8.46}$$

$$c_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \cos(nx) dx = \frac{2(\pi^2 n^2 - 2)\sin(n\pi) + 4\pi n \cos(n\pi)}{n^3 \pi} = \frac{4(-1)^n}{n^2} \tag{8.47}$$

$$d_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \sin(nx) dx = 0 \tag{8.48}$$

We have used the fact that odd functions integrate to zero if the region of integration is overa symmetric region about zero to solve for d_n. From this we know that we can write x^2 as

$$x^2 = \frac{\pi^2}{3} + \sum_{n=1}^{\infty} \frac{4(-1)^n}{n^2} \cos(nx). \tag{8.49}$$

Now that the right hand side of the ODE is written entirely as constants and trigonometric functions, we can return to determining what $y(x)$ is. Assume

$$y = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx). \tag{8.50}$$

In order to plug this back into the ODE we need to determine the first derivative,

$$y' = -\sum_{n=1}^{\infty} n a_n \sin(nx) + \sum_{n=1}^{\infty} n b_n \cos(nx). \tag{8.51}$$

Then

$$\begin{aligned} -\sum_{n=1}^{\infty} n a_n \sin(nx) + \sum_{n=1}^{\infty} n b_n \cos(nx) + 2a_0 + 2\sum_{n=1}^{\infty} a_n \cos(nx) + 2\sum_{n=1}^{\infty} b_n \sin(nx) \\ = \sin(x) + \frac{\pi^2}{3} + \sum_{n=1}^{\infty} \frac{4(-1)^n}{n^2} \cos(nx) + 1 \end{aligned} \tag{8.52}$$

$$\begin{aligned} \Rightarrow 2a_0 + \sum_{n=1}^{\infty} (2b_n - n a_n)\sin(nx) + \sum_{n=1}^{\infty} (n b_n + 2a_n)\cos(nx) \\ = \frac{\pi^2 + 3}{3} + sin(x) + \sum_{n=1}^{\infty} \frac{4(-1)^n}{n^2} \cos(nx). \end{aligned} \tag{8.53}$$

Matching terms we can then get the following equations to solve for a_0, a_n, and b_n:

$$2a_0 = \frac{\pi^2 + 3}{3} \tag{8.54}$$

$$2b_1 + a_1 = 1 \tag{8.55}$$

$$2b_n - na_n = 0 \text{ for } n \geq> \tag{8.56}$$

$$nb_n + 2a_n = \frac{4(-1)^n}{n^2} \text{ for } n \geq 0 \tag{8.57}$$

Doing some arithmetic we can solve for a_1, b_1, and find general equations for a_n and b_n giving us

$$a_0 = \frac{\pi^2 + 3}{6} \tag{8.58}$$

$$a_1 = -\frac{9}{5} \tag{8.59}$$

$$b_1 = -\frac{2}{5} \tag{8.60}$$

$$a_n = \frac{8(-1)^n}{n^2(n^2+4)} \text{ for } n \geq 2 \tag{8.61}$$

$$b_n = \frac{4(-1)^n}{n(n^2+4)} \text{ for } n \geq 2 \tag{8.62}$$

Using these equations we can write down a final form for $y(x)$:

$$\begin{aligned} y(x) = \frac{\pi^2 + 3}{6} - \frac{9}{5}\cos(x) - \frac{2}{5}\sin(x) + 8\sum_{n=2}^{\infty} \frac{(-1)^n}{n^2(n^2+4)}\cos(nx) \\ + 4\sum_{n=2}^{\infty} \frac{(-1)^n}{n(n^2+4)}\sin(nx) \end{aligned} \tag{8.63}$$

Example 8.3: *Solve the ODE $xy' + x^2 \sin x = y$ by assuming the solution takes the form of a power series. Assume that you are interested in the region about $x = 0$. It might be useful to look at the Taylor expansion of sine about x=0 (the Maclaurin series of sine), is that a power series?*

First, assume that your solution has the form

$$y = \sum_{n=0}^{\infty} c_n x^n. \tag{8.64}$$

To substitute this back into our ODE, we need to know what the first derivative of this equation is:

$$y' = \sum_{n=1}^{\infty} c_n n x^{n-1} \tag{8.65}$$

Plugging these back into the ODE gives us

$$x\sum_{n=1}^{\infty} c_n n x^{n-1} + x^2 \sin x = \sum_{n=0}^{\infty} c_n x^n \tag{8.66}$$

$$\sum_{n=0}^{\infty} c_n (n-1) x^n = -x^2 \sin x. \tag{8.67}$$

Between the previous two equations I used the fact that changing the sum starting at $n = 1$ to starting at $n = 0$ would not change anything, as the $n = 0$ term is 0. From here we cannot immediately start solving for the coefficients of our series because $\sin x$ is not a polynomial. However, looking at the Maclaurin series for sine,

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}, \tag{8.68}$$

we notice that we *can* write $\sin x$ as an infinite power series. Substituting this form into our ODE now gives us

$$\sum_{n=0}^{\infty} c_n (n-1) x^n = -\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+3} \tag{8.69}$$

$$\sum_{n=-2}^{\infty} c_{n+2} (n+1) x^{n+2} = -\sum_{n=1,3,5,\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n!} x^{n+2}. \tag{8.70}$$

Here I have rewritten both sums to be able to more easily match powers of x. From this form, we can see that for even n c_n will be zero since the sum on the right hand side of the equation does not contain any even powers of x. We also can determine that c_1 must be zero since this would correspond to a term involving x^1 on the left hand side of the equation which does not exist on the right hand side of the equation. (This reasoning can also be used to determine that c_0 and c_2 are zero, but those coefficients were already covered with our statements about even terms.) Looking at the first few nonzero terms,

$$n = 1 \Rightarrow 2c_3 = -1 \Rightarrow c_3 = -\frac{1}{2} = -\frac{1}{2!} \tag{8.71}$$

$$n = 3 \Rightarrow 4c_5 = \frac{1}{3!} \Rightarrow c_5 = \frac{1}{4(3!)} = \frac{1}{4!} \tag{8.72}$$

$$n = 5 \Rightarrow 6c_7 = -\frac{1}{5!} \Rightarrow c_7 = -\frac{1}{6(5!)} = \frac{1}{6!}, \tag{8.73}$$

we begin to see a pattern forming. This pattern can be summarized by

$$c_n = \begin{cases} 0 & \text{for even } n \\ \frac{(-1)^{(n-1)/2}}{(n-1)!} & \text{for odd } n \end{cases} \tag{8.74}$$

or more simply

$$c_{2n+1} = \frac{(-1)^n}{(2n)!}. \tag{8.75}$$

Putting all of this information back together we finally get our final solution,

$$y(x) = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n+1} \tag{8.76}$$

Example 8.4: *Solve the ODE $y'' + y' + 4y = \sin(x) + \cos(2x)$ by assuming the solution takes the form of a Fourier series. The solution is periodic and repeats every 2π.*

This problem was taken from this differential equations resource.

Let's assume that our solution takes the form of

$$y = a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx). \tag{8.77}$$

Here L is 2π here because that is the range our solution is uniquely defined over. Next, in order to plug this solution into the ODE, we need to know its first and second derivatives:

$$y' = -\sum_{n=1}^{\infty} n a_n \sin(nx) + \sum_{n=1}^{\infty} n b_n \cos(nx) \tag{8.78}$$

$$y'' = -\sum_{n=1}^{\infty} n^2 a_n \cos(nx) - \sum_{n=1}^{\infty} n^2 b_n \sin(nx) \tag{8.79}$$

Plugging this into the ODE we get:

$$\begin{aligned} &-\sum_{n=1}^{\infty} n^2 a_n \cos(nx) - \sum_{n=1}^{\infty} n^2 b_n \sin(nx) - \sum_{n=1}^{\infty} n a_n \sin(nx) + \sum_{n=1}^{\infty} n b_n \cos(nx) \\ &+ 4a_0 + 4\sum_{n=1}^{\infty} a_n \cos(nx) + 4\sum_{n=1}^{\infty} b_n \sin(nx) = \sin(x) + \cos(2x) \end{aligned} \tag{8.80}$$

What we can immediately see from this is that the left hand side of the equation has a constant term but the right hand side does not, thus we can conclude that $a_0 = 0$. Grouping the sine and cosine sums to simplify the equation then gives us

$$\begin{aligned} &\sum_{n=1}^{\infty} (-n^2 a_n + n b_n + 4a_n) \cos(nx) + \sum_{n=1}^{\infty} (-n^2 b_n - n a_n + 4b_n) \sin(nx) \\ &\qquad = \sin(x) + \cos(2x). \end{aligned} \tag{8.81}$$

Matching the modes of the sine and cosine functions on each side of the equation then leads us to realize that the only term in the cosine sum that survives is when $n = 2$ and the only term in the sine sum that survives is when $n = 1$. This gives us a set of equations,

$$-4a_2 + 2b_2 + 4a_2 = 1 \tag{8.82}$$

$$-4b_2 - 2a_2 + 4b_2 = 0 \tag{8.83}$$

$$-a_1 + b_1 + 4a_1 = 0 \tag{8.84}$$

$$-b_1 - a_1 + 4b_1 = 1, \tag{8.85}$$

to use to solve for our coefficients. I looked at the $n = 1$ case in the cosine sum and the $n = 2$ case in the sine sum because those terms also use the a_n and b_n coefficients that show up in the terms that should have nonzero coefficients and I cannot solve for these coefficients without that extra information. Solving this set of equations gives us

$$a_1 = -\frac{1}{10} \tag{8.86}$$

$$a_2 = 0 \tag{8.87}$$

$$b_1 = \frac{3}{10} \tag{8.88}$$

$$b_2 = \frac{1}{2}. \tag{8.89}$$

Then our final solution is

$$y(x) = \frac{3}{10}\sin(x) + \frac{1}{2}\sin(2x) - \frac{1}{10}\cos(x). \tag{8.90}$$

9 Coding in Python

Topics: Python data types, variables, functions, importing libraries, connecting code to math and physical science.

9.1 Motivation and approach

Translating mathematical concepts into code is an amazingly useful skill for anything from data analysis and making plots to figures to simulating the behavior of a protein interacting with an enzyme. In this section we will go over an introduction to programming in Python followed by several detailed examples of how you can use these coding concepts to solve interesting problems. As coding is best learned when you are able to try it yourself, the coding notes are provided in python notebooks linked below. Follow the instructions in Appendix A to download copies of these notebooks to your UC Berkeley-provided JupyterHub accounts.

Introduction to Python

Scientific Python

Math Bootcamp Notes: Preparation for Graduate Physical Chemistry Courses ©2021 by Rachel Clune, Orion Cohen, Avishek Das, Dipti Jasrasaria, Elliot Rossomme is licensed under CC BY-NC 4.0. This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only.

A Jupyter Notebooks

A.1 Introduction

Jupyter Notebook is a useful tool that allows you to create an online notebook which can contain writing, equations, pictures, and code all in one place. Jupyter notebooks even allow you to run python code without installing any programs on your personal computer. In this tutorial we'll go through setting up your first Jupyter notebook, as well as some useful tips and tricks to make the most out of this powerful tool.

A.2 Creating a Notebook

Jupyter notebooks are available for free online, but have limited capabilities without a subscription. Fortunately for us, Berkeley provides everyone with a subscription so long as you log in through their interface here.

Once you've logged in, you'll be presented with your home page. Here you can find your previous notebooks, make a new notebook, or even upload another notebook that is saved to your computer. Let's start by making our own python Jupyter notebook. Click "new" in the top right corner and select a Python 3 notebook.

A.3 Working in a Notebook

In your new notebook, you'll be presented with a blank cell. Cells are where your content goes, and come in a few varieties. Code cells contain code, in this particular case python code because this is a python notebook, and are also able to execute this code. Markdown cells contain text and equations, and are even able to compile typesetting from your favorite typesetting language, such as LaTex. Let's start by practicing with code cells.

As stated before, code cells are where you will write and execute code. Let's start by writing a very simple python code. Putting "print('Hello World!')" into the cell (without the double quotations) will write a simple program which will display "Hello World!" when run. To run the code, simply press the play button on the toolbar or press ctrl + enter. You should now see the output printed below the cell! You also may have noticed the small circle in the top right corner. This circle tells you about the kernel, which is what actually running the code. Hovering over the kernel will tell you what the kernel is doing. Sometimes you may run into issues where the kernel disconnects or where you want to delete all the memory. To fix this, it is often helpful to restart the kernel. To do this, simply click the Kernel dropdown menu and click restart.

Now that you've run your code, you may find that you want to delete the output so that you can run your code again. To clear output, click the cells dropdown menu and select either Current Output → clear to clear output in only the selected cell or All Output → clear to clear output from all cells. Now let's create a new markdown cell and write an equation. To create a new cell, click the insert dropdown and then cell below, or just type shift + enter (note that this will also run the current cell).

By default, our new cell is a code cell. To change the cell to a markdown cell, click the dropdown on the toolbar that currently says code and select markdown. Now that we're in a markdown cell, we can type equations or text like in a typesetting program. For instance, if we type "$y=\frac{1}{3}x+b$" into our cell and run it, an equation will appear!

Keyboard shortcuts can save you a lot of time in Jupyter Notebooks, we suggest you check out this list: https://www.earthdatascience.org/courses/intro-to-earth-data-science/open-reproducible-science/jupyter-python/jupyter-notebook-shortcuts/. You can access a list of all keyboard shortcuts by hitting esc + h.

A.4 Closing Remarks

Jupyter Notebooks is a powerful tool that allows you to keep all your code, writing, and equations in one convenient place. It makes sharing and running code more accessible, and saves the hassle of installing a programming language onto your own personal device. While this tutorial was merely a bare minimum to create your first notebook and familiarize yourself with some of the essential tools, there are many more features, shortcuts, and hotkeys to take advantage of, so I encourage you all to search on the internet for more tools if you are interested in making extensive use of Jupyter Notebooks!

B Integral table

Here we include some formulae and tips for evaluating definite integrals that you should feel free to use as a reference as needed. Perhaps unlike previous stages in your education, you are strongly encouraged to use reference materials for things like this in graduate school. It is in general important for you to know the various tools in the integration toolbox (integration by parts, u-substitution, etc.; see Sec. 1), but you also do not need to reinvent the wheel every time. Over the course of your career as a graduate student and beyond, you very well may need to evaluate an integral that you *cannot* look up in a table, and when that happens it's important to know how to do it yourself. But for the other times, you should feel the freedom to use the resources at your disposal, which include the internet.

B.1 Symmetric integrals of even and odd functions

Many of the functions we work with in quantum mechanics can be classified as either **even functions** or **odd functions**. These functions satisfy the following requirements, respectively:

$$f(-x) = f(x) \qquad \text{(even function)} \tag{B.1}$$
$$f(-x) = -f(x) \qquad \text{(odd function).} \tag{B.2}$$

In general, you can determine whether a function is even or odd by making the substitution $x \to -x$ everywhere in the function and seeing whether this simplifies either to either $f(x)$ or $-f(x)$.

Even and odd functions exhibit many of the same properties under addition and multiplication as integers do. Specifically, each of the following conditions obtain:

1. The sum of two odd functions is an odd function
2. The sum of two even functions is an even function
3. The product of two even functions is an even function
4. The product of two odd functions is an even function
5. The product of an even function and an odd function is an odd function

Identifying the integrand in a symmetric integral, *i.e.* over some bounds $[-a, a]$, as an even or odd function is helpful because the following identities obtain

$$\int_{-a}^{a} dx\ f(x) = 2\int_{0}^{a} dx\ f(x), \qquad \text{when } f(x) \text{ is even} \tag{B.3}$$
$$\int_{-a}^{a} dx\ f(x) = 0, \qquad \text{when } f(x) \text{ is odd.} \tag{B.4}$$

Of particular use in quantum mechanics, we note that these equations still hold as $a \to \infty$

B.2 Some important Gaussian integrals

The following two integrals are routinely encountered in probability and statistics, as well as in basis expansion contexts in quantum mechanics. Related integrals outside the scope of the present notes can be found here.

$$\int_{-\infty}^{\infty} dx\, e^{-ax^2} = \sqrt{\frac{\pi}{a}} \qquad (a > 0) \tag{B.5}$$

$$\int_{-\infty}^{\infty} dx\, x^2 e^{-ax^2} = \frac{1}{2}\sqrt{\frac{\pi}{a^3}} \qquad (a > 0). \tag{B.6}$$