
MATHEMATICS FOR PHYSICISTS AND CHEMISTS

Course I

Part I: COMPLEX VARIABLES, FIRST AND SECOND ORDER ORDINARY DIFFERENTIAL EQUATIONS, QUALITATIVE ANALYSIS, BOUNDARY VALUE PROBLEMS, AND THE FREE QUANTUM PARTICLE IN A BOX.

Part II: SEQUENCES, RECURSION, INFINITE SERIES, POWER SERIES, TAYLOR SERIES, POWER SERIES SOLUTIONS TO DIFFERENTIAL EQUATIONS, SPECIAL POLYNOMIALS, AND THE QUANTUM HARMONIC OSCILLATOR.

Part III: VECTOR SPACES, SUBSPACES, BASES, LINEAR TRANSFORMATIONS, KERNELS AND IMAGES, GEOMETRY OF SPACE, LINEAR OPERATORS, INVERSES, SIMILARITY, DETERMINANTS, TRACES, SPECTRA, SELF ADJOINTNESS, DIAGONALIZATION, SYMMETRIES, AND GROUPS.

Course II

Part IV: CURVES, SCALAR FIELDS, VECTOR FIELDS, DIFFERENTIATION, GRADIENT, DIVERGENCE, CURL, LAPLACIAN, INTEGRATION, FLUX, POTENTIALS, STOKES' THEOREM, CURVED SURFACES, PARAMETERIZATION, AND COORDINATE SYSTEMS.

Part V: MODELING OF HIGH DIMENSIONAL ODES, FLOWS, CONTINUUM LIMIT OF DISCRETE SYSTEMS, PARTIAL DIFFERENTIAL EQUATIONS, HEAT EQUATION, LAPLACE EQUATION, WAVE EQUATION, D'ALEMBERT'S PRINCIPLE, AND SEPARATION OF VARIABLES.

Part VI: INNER PRODUCT SPACES, NORMS, BASES, LINEAR OPERATORS, ADJOINTS, DIFFERENTIAL OPERATORS, INTEGRAL OPERATORS, SYMMETRIES, FOURIER TRANSFORM, FUNDAMENTAL SOLUTIONS, CONVOLUTION, AND THE HYDROGEN ATOM.

By

COLIN ROBERTS

COLORADO STATE UNIVERSITY

Contents

1	Introduction & Review	2
1.1	Introduction	2
1.2	How to read this text	3
1.2.1	Mathematical language	4
1.3	Review	5
1.3.1	Computation	6
1.3.2	Algebra	9
1.3.3	Geometry	10
1.3.4	Trigonometry	10
1.3.5	Functions	11
1.3.6	Calculus	18
2	Tutorials	24
2.1	Wolfram Alpha	24
2.1.1	Solving Algebraic Expressions	24
2.1.2	Error Checking	25
I	Ordinary Differential Equations	26
3	The Complex Numbers	27
3.1	Introduction to Complex Numbers	27
3.2	Complex Number Algebra	29
3.3	Geometry of complex numbers	30
3.4	Polar coordinates	32
3.4.1	Euler's Formula	34
3.4.2	Polar Coordinates and Transformations	35

3.4.3	Multiplication of complex numbers in polar form	37
3.5	Periodic Motion	38
4	First and Second Order Dynamics	44
4.1	Ordinary Differential Equations	45
4.2	Solutions to an ODE	48
4.3	General and Particular Solutions	49
4.4	Separable Equations	52
4.5	Changing Variables and Symmetry	54
4.6	Qualitative analysis for first order equations	57
4.6.1	Autonomous equations	59
4.7	First Order Linear Differential Equations	62
4.8	Applications to Chemical Kinetics	65
4.9	Second Order Equations	71
4.9.1	Homogeneous and Constant Coefficients	71
4.9.2	Qualitative Analysis	75
4.9.3	Inhomogeneous Linear Equations	78
5	Boundary Value Problems	91
5.1	Laplace's and Poisson's Equations	92
5.2	A Particle in a 1-Dimensional Box	95
II	Elements of Analysis	104
6	Sequences and Series	105
6.1	Sequences	105
6.1.1	Convergence	106
6.1.2	Recursive Sequences	109
6.2	Series	110
7	Power Series	120
7.1	Power Series	120
7.2	Taylor Series	123
7.3	Operations with Power Series	126
7.4	Approximation with Power Series	128
7.5	Power Series Solutions to Differential Equations	132
7.6	Orthogonal Polynomials from Power Series Solutions	136
7.6.1	Legendre's Equation	136
7.7	Quantum Harmonic Oscillator	140
7.7.1	Quantum tunneling	142
III	Finite Dimensional Linear Algebra	151
8	Vectors and Vector Spaces	152
8.1	Vector Spaces	153
8.2	Vector Algebra	156
8.2.1	Linear Combinations	158

8.3	Vector Components	158
8.4	Vector Algebra with Components	159
8.5	Products of Vectors	164
8.5.1	The Dot Product	164
8.5.2	The Cross Product	168
9	Linear Transformations and Matrices	171
9.1	Linear Transformations	171
9.2	Matrix Representation of Linear Transformations	174
9.2.1	Matrix Algebra	175
9.2.2	Properties of Matrix Multiplication	177
9.3	Systems of Linear Equations	178
9.4	Solving Linear Systems of Equations	179
9.5	Linear Independence, Span, and Bases	184
9.6	The Determinant and Trace	185
9.6.1	Determinants and Systems of Equations	191
9.6.2	Properties of determinants	194
9.6.3	The Trace	196
9.7	Inverse Matrices and Similarity	197
9.7.1	Properties Related to Inverse Matrices	200
9.8	The Eigen-Problem	201
9.9	Diagonalization and Special Matrices	208
9.9.1	Hermitian Matrices	209
9.10	Groups and Symmetry	213
IV	Calculus in Higher Dimensions	220
10	Curves, scalar fields, and vector fields	221
10.1	Overview of higher dimensional functions	221
10.2	Curves in space	224
10.2.1	Derivatives of Curves	226
10.2.2	Lengths of Curves	228
10.3	Scalar fields	229
10.3.1	Partial derivatives of scalar fields	232
10.3.2	Directional Derivatives	236
10.4	Vector Fields	236
10.4.1	Vector field algebra	239
11	Vector calculus	242
11.1	The gradient of a scalar field	242
11.1.1	Optimization	244
11.2	Differentiation of Vector fields	245
11.2.1	The Jacobian of a vector field	246
11.2.2	Divergence and Curl	248
11.3	Laplace Operator	252
11.3.1	The Laplacian of a Scalar Field	253
11.3.2	Vector Laplacian	254

11.4 Composite Functions and Chain Rule	255
11.4.1 Composite Functions	256
11.5 Integration of Vector Fields	257
11.5.1 Line integrals of vector fields	258
11.6 Potential Functions	259
11.6.1 Integrating the Gradient	259
11.6.2 Requirements for Potentials	260
11.6.3 Properties of Fields with Potentials	261
11.7 Integrals of Scalar Fields	262
11.7.1 Line Integrals	267
11.7.2 Volume Integrals	268
12 Surfaces and Flux	269
12.1 Implicit surfaces	270
12.1.1 Normal Vectors to Implicit Surfaces	275
12.1.2 Implicit Function Theorem	276
12.2 Surfaces as Graphs	277
12.2.1 Equation for a Plane	278
12.2.2 Tangent Planes and Surface Normals	278
12.3 Flux and Surface Integrals	281
12.4 Area Form and Integration	283
13 Coordinate systems	288
13.1 Cylindrical Coordinates	288
13.1.1 Functions and Calculus in Cylindrical Coordinates	291
13.1.2 Cylindrical Unit Vector Fields	293
13.1.3 Integration in Cylindrical Coordinates	295
13.1.4 Derivatives in Cylindrical Coordinates	297
13.2 Spherical Coordinates	298
13.2.1 Integration in Spherical Coordinates	299
V Partial Differential Equations	300
14 Modeling and the continuum limit	301
14.1 Higher dimensional ODEs	301
14.2 Linear systems and flows	303
14.3 Partial Differential Equations	306
14.4 The heat equation	308
14.5 The Laplace Equation	313
14.6 The Wave Equation	314
15 Separation of variables	317
VI Further Topics in Linear Algebra	318
16 Complex functions and transformations	319
16.1 Introduction	319

16.2 Complex Functions	319
16.2.1 Cartesian and Polar Representations	320
16.2.2 Complex Valued Functions	323
17 Inner product spaces	331
17.1 Introduction	331
17.1.1 Infinite Dimensions	332
17.2 Function Spaces	332
17.3 Inner Products and Norms	333
17.3.1 Norms	335
17.4 Inner Product Spaces	336
17.5 Symmetries	337
17.6 Bases	338
17.6.1 Orthonormal Bases	339
17.6.2 Series and Integrals as Linear Combinations	341
17.6.3 Projection	342
18 Linear Operators	347
18.1 Matrices	347
18.2 Linear Operators	349
18.2.1 Adjoint	350
18.2.2 Hermitian Operators	352
18.3 Differential and Integral Operators	355
18.3.1 Integral Operators	356
18.4 Spectra	357
19 Fourier Theory	359
19.1 Fourier Series	359
19.1.1 Solving Equations	361
19.2 Fourier Transform	363
19.2.1 Fourier Transform Operator	365
19.3 The Dirac Delta Function	366
19.4 Computing Fourier Transforms	367
19.5 The Inverse Fourier Transform	369
19.6 Solving Differential Equations	370

Prerequisites

1

Introduction & Review

1.1 Introduction

This class and the sequel are concerned with covering the mathematics of thermodynamics, electromagnetism, and quantum mechanics. Of course, the mathematics supplied here is not limited to these fields, but it is hopefully presented in a way to align with those subjects. At the heart of chemistry are the interactions in and between atoms and molecules. On the small scale, their interactions and structures are well understood via experiment and quantum theory. Zoom out a bit and electromagnetic effects take over. Zoom out further and we find that ensembles of large amounts of chemicals behave due to thermodynamic laws.

All of scientific theory is verified by experiment, but the theory itself is based in mathematics. A working scientist tends to lean towards either theorist or experimentalist, but one should be knowledgeable in both respects. Think of mathematical theory as providing you with another form of analysis. Mathematical knowledge will undoubtedly help you in your future scientific endeavors. At the very least, it is an exercise in how to approach problems in a logical way.

Take this course as a first dive into many new areas of math. There will *immediately* be a wealth of new terms and methods presented to you. Learning what these are takes effort. Even keeping terminology correct or what type of object (scalar, vector, matrix, function, etc.) something is can be confusing! Keep your mind open to questions. Try not to fear math and it will work with you. Keep your head up and do not feel discouraged when the material feels tough. It is tough for everyone. Eventually you will improve your understanding. After this course and the sequel, you can look back at everything you have learned and smile!

1.2 How to read this text

What good is a textbook if you don't enjoy reading it? Or worse, what if you can't follow the material? Reading textbooks or scientific texts can be challenging, but you should think of it as another step in your career of learning!

To get the most out of this textbook I suggest the following:

1. *Pre-read.* Read through the relevant section once and try to see what the main point is. Don't fuss over details! Go for big picture here. Maybe write down questions that come to you.
2. *Re-read.* Read the relevant section again. Yes, I am asking for you to do this twice! This time, go slower. Go through with a fine tooth comb and make sure you can make as much sense of the material as you can. See if all the logic is true. If there are exercises, attempt them. Write down questions and bring them to class, office hours, or to a tutor.
3. *Review.* Come to class, office hours, or meet with other students/colleagues/tutors to go over the questions you have. As this says here, class should be more of a review than a first pass through. Having questions is great! It means learning is soon to follow.
4. *Practice.* Do the homework and work through other exercises present in the text. This is the start to mastering a subject. You may feel that you have gotten the material down but you can't *know* that until you try. Give the homeworks a try without notes or the text until you feel you need it! This aids in ability to recall.
5. *Correct.* None of us are perfect. Chances are that you may still miss something on an exercise or problem that you were sure that you had right. Realize the mistake and integrate the new knowledge. Sometimes this takes a bit of un-learning. That can be tough, so be patient and forgiving with yourself.
6. *Reference.* There will be times where you will need to revisit some previous material. Use this text as a reference for our class. The more times you see something, the more it will stick.

Sometimes there will be information stored away in problems, exercises, questions, answers, or remarks that is not otherwise explained in the text. This is done when a bit of information is better discovered by the reader as opposed to being given away. These problems will typically be labeled with an asterisk to denote their importance. I choose to use the following conventions.

- *Problems.* Not assigned to be turned in, but should be done to check your knowledge once you have spent time with a portion of the text. Problems will come after sections or subsections and important problems are marked with an asterisk.
- *Exercises.* Exercises come within the body of a section or subsection. These should be, at the very least, read along with the other material to see what can be asked. In general, these are not meant to take too long. They should mostly check your current understanding. Think of them as passage to the next material.

- *Questions.* Questions are put there for you to think about. Almost always the questions will come along with an answer. Though you have an answer, you should work to understand *why* it is the answer and if there is any more to be said.
- *Answers.* Given to most questions. Think about these and make sure they make sense to you.
- *Remarks.* These have nuggets of information that serve as a reminder or notify the reader about some subtleties. Note what these remarks mention and be sure to keep them in mind.

1.2.1 Mathematical language

Mathematical symbols

Here is a list of symbols you will find in our class (if more come, I will update this list):

- \mathbb{R} , The set of *real numbers*.
- \in , “Is a member of.” (i.e. $\sqrt{2} \in \mathbb{R}$ can be read as, “The square root of 2 is a member of the real numbers.”)
- \neq , Not equal to.
- \approx , “Approximately equal to.”
- \coloneqq , We are defining something, i.e., $\frac{df}{dx} := \lim_{h \rightarrow 0} \frac{f(x+h)-f(x)}{h}$.
- \implies , ”Implies.”
- \iff , ”If and only if.”
- \therefore , ”Therefore.”
- ∞ , Infinity. *Note: Infinity is not a number!*
- δ , A change. (i.e. δx represents a change in the variable x).
- \vec{v} , a boldfaced character with an arrow will represent a vector.
- (x_1, x_2, \dots, x_n) , represent a (row) vector with n components.
- $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, a vector with n components.

Mathematical presentation

I have provided you with some ideas on how to read this text which tend to be good practice in general. But, this is a mathematics text, how is it different? Well, mathematics tends to be presented in a specific way and it can come across as being utterly devoid of inspiration. In truth, mathematics is discovered almost entirely backwards from how a text is written. It's just the truth.

Like other mathematical texts, we will often motivate a topic and define some terms that seem to be helpful along the way. It's sometimes easiest to add new terms to a dictionary when they are often used. We just call this a *definition*. See for example, Definition 1.3.1. Definitions will be placed in CSU-themed green boxes.

Definitions are tautologically true since we have defined the very thing we would like. However, we sometimes wish to show that what we have defined plays nicely together and gives us some new useful information. These can come in many forms! We have *theorems*, *propositions*, *lemmas*, and *corollaries*; Each with its own specific meaning. All of these results will be placed in CSU-themed gold boxes as they are representing new information. Let's briefly say what each is.

- *Theorem*. An important logical result that needs proof to be accepted. Proof comes from logic that we previously develop. A very useful example is the Pythagorean theorem.
- *Proposition*. A smaller independent result, but important nonetheless.
- *Lemma*. A smaller result that tends to be working towards a larger one. Many times, a theorem is proved by using multiple lemmas beforehand.
- *Corollary*. A result that follows fairly easily from a lemma or theorem. Sometimes it's extra emphasis on using a very general theorem on something more specific, so it is literally not saying anything new but rather pointing the reader to this important consequence.

In this journey of mathematics we will come across each one. The important thing for the reader is to understand the logic along the way. Do you need to prove theorems? No, I will do that for you. More often than not, I will even omit proofs unless they are strategically enlightening. However, one should always be suspicious if some logic seems incorrect. I promise you, there are no lies in this text, but it doesn't mean things are obvious!

1.3 Review

The prerequisite classes for this course are Math 155, Math 159, or Math 160. Moreover, there is a requirement that the following topics are understood. To check your understanding, work through all the listed problems in this review section. I repeat! Work through each problem in the review! If you can work through each problem and, better yet, understand each step you take through a solution, then you are well on your way to learning new material in this course.

This course will require the following:

1. Computation.

2. Algebra.
3. Geometry.
4. Trigonometry.
5. Functions.
6. Calculus.

Below in each respective subsection are notes and example problems.

1.3.1 Computation

What is computation? Computation is what we have drilled into our heads at a young age. It's how many of us view mathematics up until we explore areas such as algebra. Given an expression, how can we interpret it? With this, we also need an agreed upon language.

Typically we mean to output a real (or possibly complex) number from an expression.¹ This number tells us what an expression means. For example, the expression is about making some type of measurement. In scientific fields, that is common.

PEMDAS

PEMDAS, or

Parentheses → Exponents → Multiplication → Division → Division → Addition → Subtraction,

gives the standard order of operations. This is simply a choice us humans have agreed on and made calculators or computer programming languages follow. Good for avoiding arguments, probably.

The reason I note this here is so that we can pay attention to properly computing expressions. Often times students will find expressions such as

$$-2^2 \quad \text{and} \quad (-2)^2$$

to be equal. If you found them equal, can you see how they differ?

Be cognisant of what the expression is *really* telling you to do. Use calculators or online tools like Wolfram Alpha as tools to check your answers. Often you will not have have a calculator available, so practicing without one is ideal.

Logarithms and exponents

Logarithms and exponents are extremely important classes of functions as well as tools. Logarithms originally assisted in helping mathematicians do large computations as they can turn multiplication into addition which is arithmetically less challenging. Both show up in nature quite often. Let us revisit the necessary rules we must know about these functions.

¹Really we just need outputs we agree upon. The numbers I mentioned are common but there's no inherent reason that those two number systems are what your expression calls for!

For the following table, log represents a logarithm of any base. A base is only specified when needed. Think about which rules are redundant, if any!

Rule	Logarithms	Exponents
Identity	$\log(1) = 0$	$a^0 = 1$
Inverse	$\log_a(a^x) = x$	$a^{\log_a(x)} = x$
Multiplication	$\log(ab) = \log(a) + \log(b)$	$a^x a^y = a^{x+y}$
Division	$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$	$\frac{a^x}{a^y} = a^{x-y}$
Power	$\log(a^p) = p \log(a)$	$(a^x)^p = a^{px}$

Challenge: Can you derive these rules based on what each function's definition is? That is, remember that

$$a^p = \underbrace{aa \cdots a}_{p \text{ times}}$$

²and

$$b^x = a \quad \text{means that} \quad \log_b(a) = x.$$

If you don't try this, that's okay. One mathematical exercise is to re-derive knowledge for yourself. Many people find it helps the learning process immensely!

Lastly, remember what the magnitude of different powers tell us. That is, I want you to be comfortable making the following identifications.

$$\begin{aligned} a \cdot a &= a^2, \\ \sqrt{a} &= a^{1/2}, \\ \sqrt[p]{a} &= a^{1/p}, \\ a^{-1} &= \frac{1}{a}, \\ a^{-p} &= \frac{1}{a^p}. \end{aligned}$$

Armed with your previous knowledge and the above information, work on the problems below.

Problems

Problem 1.1. Evaluate the following.

(a) $-3^2 - 2(3 + 2^2);$

(b) $(-3)^2 - 2(3 + 2)^2;$

(c) $\frac{2-(-3)^2}{2(3-2)^2}.$

²If p is not a positive integer, this is a bit harder to understand. Pretend that it is for this case.

Problem 1.2. Write the following as a simplified fraction:

$$\frac{1}{2} - \frac{3}{5}.$$

Problem 1.3. Rewrite the following in an equivalent way using log and exponent rules. If no rules apply, say so.

- (a) $\ln(2a)$;
- (b) $\ln(2 + a)$;
- (c) $\ln(2^a)$;
- (d) $\log_{10}(a^2)$;
- (e) $\log_2(2^a)$.

***Problem 1.4.** Recall that we can convert between different bases of logarithms. For example,

$$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}.$$

Convert the following logarithms to the *natural logarithm* (logarithm with base e) and evaluate the expression when possible.

- (a) $\log_2(2)$;
- (b) $\log_2(8)$;
- (c) $\log_8(2)$;
- (d) $\log_{10}(e)$;
- (e) $\log_3(1)$;
- (f) $\log_5(0)$.

If an expression can't be evaluated, can you explain why?

Remark 1.3.1. The name “natural logarithm” will make more and more sense as we see how natural Euler’s Number e is.

***Problem 1.5.** We can also convert between different bases for exponentials. Really, we just like to do this to convert to the exponential with Euler’s number e . We have that

$$b^x = e^{\ln(b)x}.$$

Can you show this using the rules above? Convert the following to their natural exponential ($e^{\text{something}}$) counterpart or convert to the form b^x .

- (a) 3^x ;
- (b) $e^{\ln(2)x}$.

1.3.2 Algebra

What is algebra? Algebra is Arabic for “reunion of broken parts.” It is a large and pervasive field of mathematics concerned with structure. There are many different flavors as well. Abstract algebra, linear algebra, and operator algebra are just a few examples. We will learn some of each of the named fields as we progress through the Math 271 and 272 courses.

At this point, we treat algebra as manipulation of equations. What one does to one side of an equation must do to the other in order to keep from changing the result or truth. We are all familiar with this concept. Clearly if

$$x = 5$$

and I only add 2 to the right side,

$$x = 5 + 2 = 7,$$

then our knowledge about x has changed. Albeit a simple example, this is *exactly* what happens when we make algebra mistakes in our work. Believe me, *all* of us make algebra mistakes.

Practice with the problems below to review your understanding of algebra. They are given in no particular order. Algebra is unbelievably fundamental in our pursuit of mathematical knowledge. We must make sure to have a strong foundation!

Problems

Problem 1.6. Solve the equation for p :

$$3p - (5 - 2p) = 2(p - 3).$$

Problem 1.7. Solve the equation for x :

$$12 < \frac{9 - x}{3}.$$

Problem 1.8. Which number(s) x are in the interval $[2, 6)$ and satisfy $x < 4$?

Problem 1.9. Simplify the following.

$$\frac{a^3 b^4}{(ab)^2} = ?$$

What can we say if a or b is equal to zero?

Problem 1.10. Show that the following equalities are true using the rules given on exponents, fractional, and negative powers.

$$(a) \sqrt{x^2 + x^2 y^2} = x\sqrt{1 + y^2}$$

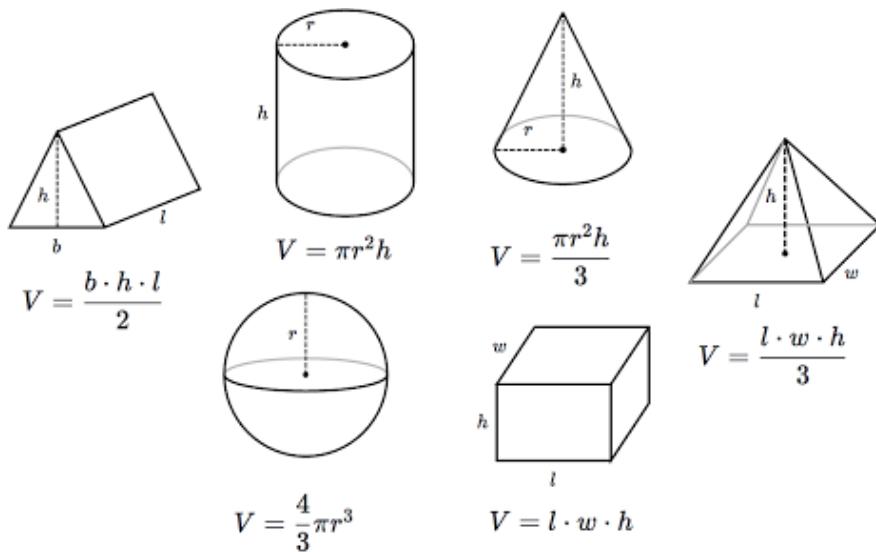
$$(b) \frac{x^2 + y^2}{x} = x + \frac{y^2}{x}$$

1.3.3 Geometry

What is geometry? Loosely speaking, geometry is the study of shape of space. Geometry is concerned with distances, angles, size, and shape. Many tools from more advanced geometry are distilled into a form that we will use throughout the time in Math 271 and Math 272. At times, we will explore some of these concepts in a more advanced way (see linear algebra and multivariate calculus).

For an introduction, knowing planar and spatial geometry is helpful. One should be familiar with calculating areas and volumes as well as lengths and angles. The most important shapes to recognize will be circles and triangles as they satisfy some nice symmetries or properties. We can use these two shapes along with some trigonometry to make geometrical inference.

Here are some common volumes that may show up in Math 271 or the sequel, Math 272.



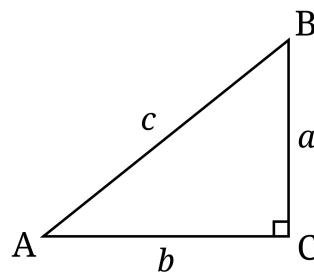
1.3.4 Trigonometry

What is trigonometry? Think of trigonometry as the mathematics behind analyzing planar triangles and circles. This comes down to using the well known trigonometric functions \cos , \sin , \tan , \sec , \csc , \cot . Of course, we often times care about the (pseudo) inverse functions of these as well.

One of the most important results in trigonometry is the Pythagorean Theorem. We will use it all throughout this course in one way or another. Let us state it below.

Theorem 1.3.1: Pythagorean Theorem

Let \triangle be a right triangle with side lengths a , b , and c shown below.



Then we have that $c^2 = a^2 + b^2$.

1.3.5 Functions

What are functions? Functions are machines that give a single output for a given input. This is something you may have heard called the “vertical line test” when you were looking at graphs of real values functions. But, Beware! That doesn’t mean that each single input corresponds to a single output. Many inputs can have the same output. More on this later.

We often use functions to represent relationships or measurements. Analyzing functions is of utmost importance when we wish to gain insight from them. For us, functions will typically be thought of as some physical variable (think temperature, velocity, or probability). We’ll also learn that functions can be solutions to certain kinds of expressions called *differential equations*.

There is a lot of terminology and notation surrounding functions. Let us cover the necessary details.

Notation and terminology

We typically represent a function (of a single variable) by $f(x)$ and say “ f of x .” Of course, the function f and input x can be represented by any symbols you’d like. Maybe you want to have a function $\odot(\uparrow)$ which we would say “smiley of up-arrow.” We tend to stick with notation others use to stay a bit more sane.

Mathematicians really like to specify a function by writing

$$f: D \rightarrow C$$

which can be read as “ f inputs variables from the set D and outputs an element from the set C .” Here, D is the *domain* and C is the *codomain*. This notation is handy when we deal with functions that have different inputs and outputs than we have handled before. So, mathematicians shouldn’t be the only ones to like this notation!

So far, what I have provided above does not specify what the function actually does. For this, we usually provide an expression like

$$f(x) = x^2$$

which tells us to square each input to get the output or we can specify the function to take certain values when given a certain input, i.e.,

$$f(2) = 3.$$

Domain, codomain, and range

The domain and range of a function are always specified. However, sometimes the specification is understood and we don't list it. The **domain** D of a function is the set of all input values. The **codomain** C of a function is the set that output values can live in. *Note, not all elements in the codomain are necessarily output by the function!* The **range** R of a function is the set of all output values.

Question 1.3.1. Is the codomain C a larger, smaller, or same size set as the range R ?

Answer 1.3.1. This was a bit of a trick question. The codomain is *at least as big* as the range as it contains all the output values the function *could* hit. The range may have the same exact elements of the codomain and the range could be smaller than the codomain.

Example 1.3.1: Codomain and Range for Finite Sets

- (i) Let f be a function with domain $D = \{1, 2, 3\}$ (i.e., D is the set of numbers one, two, and three) and codomain $C = \{1, 2, 3, 4\}$. We can say that

$$f: D \rightarrow C$$

and we completely define f by choosing an output for each input. We'll take

$$\begin{aligned} f(1) &= 1; \\ f(2) &= 1; \\ f(3) &= 4. \end{aligned}$$

As stated, the codomain is the set $\{1, 2, 3, 4\}$, but what is the range R ? Since we have specified an output for every input, we can just look at the list of outputs which is

$$f(1) = 1, \quad f(2) = 1, \quad f(3) = 4,$$

and so the range $R = \{1, 4\}$. Notice that the range is strictly smaller than the codomain. Also, this function does not output a unique value for each given input.

- (ii) Let g be a function on the same domain and codomain as in (i). That is,

$$g: D \rightarrow C.$$

Let us define g as we did with f by specifying the output for each input.

$$\begin{aligned} g(1) &= 2; \\ g(2) &= 3; \\ g(3) &= 4. \end{aligned}$$

Again, one should check that the range is smaller than the codomain. Also, check whether this function has a unique output for each input.

Question 1.3.2.

- Given an output value for the function g above, can one determine what the input value had to be?
- How about for f ?
- What is the difference between these two?

Answer 1.3.2.

- Yes we can. We call this association the *inverse function*.
- No, not quite. For example, what input gives us the output value of 1?
- g is a *one-to-one* function. It has a unique input value for each output value. f is not one-to-one. Inverting f is not really possible.

Example 1.3.2: Codomain and Range for \mathbb{R} -Functions

Let us work with functions we are likely more familiar with.

- (i) Let f be a real valued function of one real variable. That is,

$$f: \mathbb{R} \rightarrow \mathbb{R}.$$

We understand this as f takes in a real value and outputs a real value as well. This is just some extra notation and verbiage for something we already understand!

Since \mathbb{R} is infinite, us humans can't really specify an output value for every input. Instead, we provide an association. Let's define f by saying what f *does* to some input value. Let

$$f(x) = x^2.$$

Now, the codomain for f is all the real numbers \mathbb{R} but the range R for f is smaller. Since we are squaring the input to get the output, the range of f is $R = [0, \infty)$. We can't achieve negative values! Notice also that f is not one-to-one.

- (ii) Let's define a new function $g: \mathbb{R} \rightarrow \mathbb{R}$ by letting

$$g(x) = x^3.$$

The range in this case is \mathbb{R} and g is one-to-one. You should verify these facts.

Composition

If we have functions f and g , we may want to look at how we can string the two together. First, we should say what we mean by this and determine when it is even possible to do so. Let's say that we have

$$\begin{aligned} f: D_1 &\rightarrow C_1, \\ g: D_2 &\rightarrow C_2, \end{aligned}$$

and the range for f and g is R_1 and R_2 respectively. We want to consider the composition of functions

$$(f \circ g)(x) = f(g(x)), \\ (g \circ f)(x) = g(f(x)).$$

Question 1.3.3. When are we allowed to *compose* the function f with g as $f \circ g$? How about as $g \circ f$?

Answer 1.3.3. Let's think of this in an analogy. If f converts from Spanish to English and g converts from German to Spanish, how can we string the two translators together to convert from German to English?

Think of x as a word in German. If we then look at g translating x ,

$$g(x) = y,$$

then y is now some word in Spanish. Now, f is happy to translate this word right on over to English by

$$f(y) = z,$$

where z is our desired English word. So, we have really done the following:

$$(f \circ g)(x) = f(g(x)) = f(y) = z.$$

Now, in terms of functions on sets of numbers, we want to think like in this above analogy. The output of one function g better agree with the input f knows how to handle. More below.

If we have that the range R_1 is in the domain D_2 , then we can have the composite function

$$g \circ f,$$

and

$$g \circ f: D_1 \rightarrow C_2.$$

This may be fairly abstract and confusing. Don't worry too much! We'll keep seeing the point of this way of thinking over time. Let this knowledge grow slowly over time.

Exercise 1.3.1. Repeat the above if we have that the range R_2 is in the domain D_1 .

Inverse functions

We have all seen *inverse functions*, but let us re-familiarize ourselves with them just a bit while also giving a preview of the style of text to come.

Definition 1.3.1: Invertible Functions

Let $f: D \rightarrow C$ be a function. We say that f is **invertible** if for any value $y \in C$ there is a corresponding unique input $x \in D$. That is to say, if we have $f(x) = y$ then the inverse function, f^{-1} , satisfies

$$f^{-1}(y) = x.$$

The inverse function *undoes* the original function. So we have that

$$(f^{-1} \circ f)(x) = x$$

and

$$(f \circ f^{-1})(y) = y.$$

for every x and every y .

Remark 1.3.2. There's more subtleties here than meets the eye. I'm going to avoid them for now. But if you'd like more to think about, consider the composition function

$$f \circ f^{-1}.$$

What will we have to know about domains, ranges, or codomains? Careful.

Alluded to earlier was the idea of a function that is one-to-one. Let's define that here and now.

Definition 1.3.2: One-to-one Function

Let $f: D \rightarrow C$ be a function. We say that f is **one-to-one** if for each output value y there is a unique input value x .

Take a moment to soak this one in. Revisit Example 1.3.1 and Example 1.3.2. Think of a few examples of functions that are and are not one-to-one.

Another important notion is that of a function whose range is equal to the codomain. We made sure to see that there could be differences before, so let's define this notion.

Definition 1.3.3: Onto Function

Let $f: D \rightarrow C$ be a function with range R . If we have $R = C$, then f is **onto**.

While we're at it, we may as well introduce a proposition. This is not a groundbreaking result as the definitions we have really play nicely together.

Proposition 1.3.1: One-to-one and Onto Functions are Invertible

If $f: D \rightarrow C$ is a one-to-one and onto function if and only if it is also invertible.

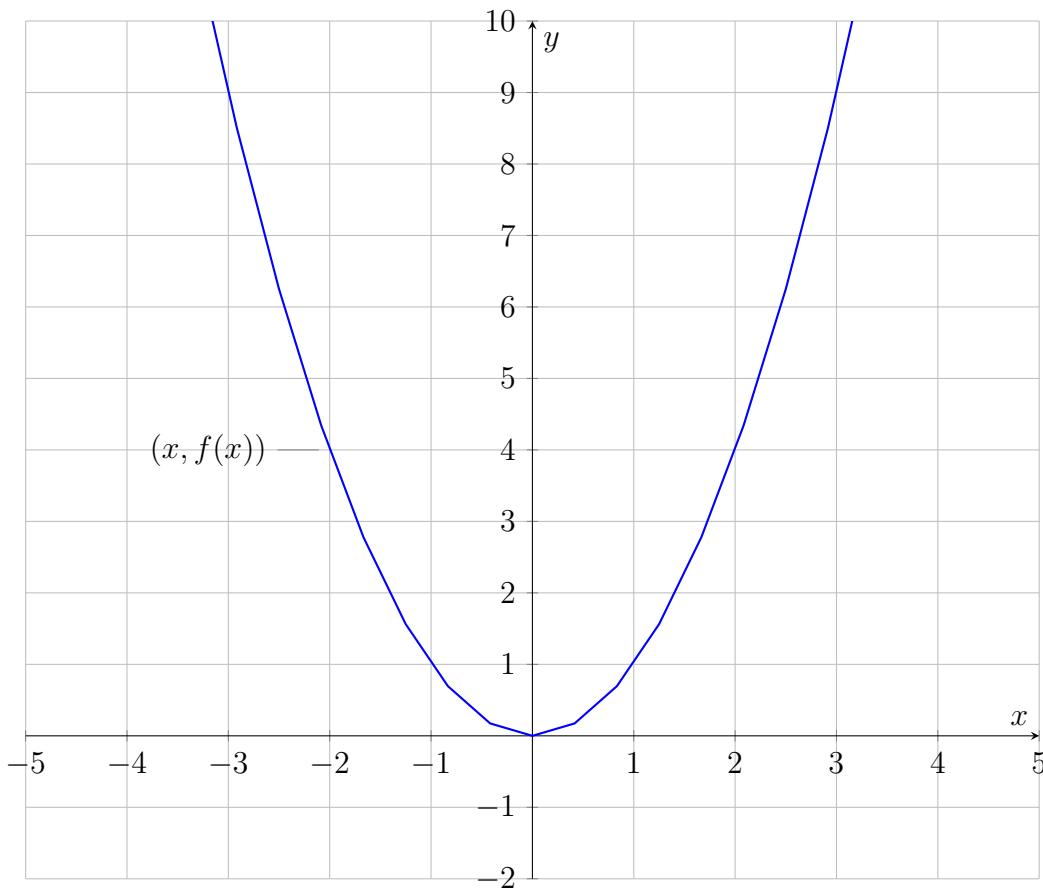
Right now, we're avoiding any proof. This one isn't too bad, but I'll restate that proofs are not part of this course. Proofs take time to understand and it's challenging to develop the skills to write them properly! That's why the mathematics majors at CSU take Math 235 as an introduction to mathematical reasoning. It truly is a challenge in its own right.

Visualizing functions

When working with real valued functions of one real variable (again, $f: \mathbb{R} \rightarrow \mathbb{R}$), we often draw the *graph* of the function in the plane. The function knows nothing about how we like to look at it, and it's important to realize that this is merely a human adaptation. We'll eventually learn *exactly* what it is that we do when we graph a function, but give it time. For the mean time, just revisit the idea yourself.

Example 1.3.3: Graphing a Function

Let's consider the function $f: \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x^2$. The the **graph** of f is the set of points $(x, f(x))$ that we highlight in the xy -plane. So, the graph for this f follows.



Similar techniques for visualizing functions in higher dimensions will come. For now, just practice with functions of one real variable.

Problems

Problem 1.11. Let $f(x) = x^2$.

- (a) Show that

$$f(x+y) = x^2 + 2xy + y^2.$$

Warning! There is a common error made here. Do not fall victim to believing that

$$f(x+y) = f(x) + f(y)$$

for all functions. Functions that have this property are *special*!

- (b) Show that

$$f(\lambda x) = \lambda^2 x^2.$$

Problem 1.12. Let $g(r) = \frac{t}{2}(r + r^2)$. Evaluate and simplify the following.

- (a) $g(a+b)$;
- (b) $g(\lambda a+x)$;
- (c) $g(r^2)$.

Problem 1.13. Let $r(\odot) = \frac{1}{\odot^2}$. Evaluate and simplify the following.

- (a) $[r(\odot)]^2$;
- (b) $r(-\odot)$;
- (c) $-r(\odot)$.
- (d) $r(\odot) + r(2\odot)$.
- (e) $r(\odot) + 2r(\odot)$.
- (f) $r(\odot^2) + r(2\odot) + 2r(\odot)$.

Remark 1.3.3. Just remember, the function and variable letters are arbitrary. Pay attention on how to properly evaluate a function with a given input despite aesthetic differences.

Problem 1.14. Find the domain and range of $f(x) = \sqrt{9-x}$.

Problem 1.15. Find the domain and range of $h(t) = \sqrt[3]{2-x^2}$.

Problem 1.16. Let $f(x) = 5x+1$ and $g(x) = x^2$.

- (a) Write the composite function $(g \circ f)(x)$.
- (b) Write the composite function $(f \circ g)(x)$.

Problem 1.17. Use the table to find $(f \circ g)(3)$

x	-2	-1	0	1	2	3
$f(x)$	0	6	4	-1	3	-2
$g(x)$	5	3	2	1	-1	0

Problem 1.18. State if the given functions are inverses of each other. Be sure to check both $(f \circ g)$ and $(g \circ f)$ to see that this is true.

- (a) $f(x) = 4 - \frac{3}{2}x$ and $g(x) = \frac{1}{2}x + \frac{3}{2}$;

$$(b) \ f(x) = 11x - 2 \text{ and } g(x) = \frac{x+2}{11}.$$

Problem 1.19. Find the inverse for the following functions. If one does not exist, explain why. If a domain isn't specified, assume that it is \mathbb{R} .

$$(a) \ f(x) = x^3.$$

$$(b) \ g(x) = x^2.$$

$$(c) \ h(x) = \sqrt{x}.$$

$$(d) \ l(x) = 8x + 3.$$

$$(e) \ p(x) = \sqrt[3]{x} - 3.$$

1.3.6 Calculus

So far, all of us have studied the calculus of functions of one real variable. It just so happens that the techniques developed there will generalize nicely to higher dimensional cases. However, it will take some careful thought along the way. What's the point anyway?

Calculus is the study of rates of change of functions. I like to think about it as the study of functions on the very small scale. On the small scale, you found that functions look almost linear. It's not hard to see this! Take any function of your liking and zoom in on a specific region of this function. Keep zooming in and you'll notice that at some point the function looks like a line. This idea allowed us to look at the *derivatives* of functions as well as help us compute *integrals* of functions! It's extremely important. It will also relate to a beautiful field of mathematics known as linear algebra.

Studying change is about all we ever do as scientists. Even systems that *don't* change over time can be thought of as just having no rate of change! As we turn knobs or mix chemicals we watch for change. We record this change and study the properties both before and after. It is undoubtedly important for all of us to know a nice way to analyze changes of systems. This is captured elegantly by calculus.

Derivatives

Derivatives were introduced to study change of functions. First, we looked at average rates of change of functions over an interval. Say we have a function $f: \mathbb{R} \rightarrow \mathbb{R}$ and we want to find its average rate of change from x_0 to x_1 , then we would compute

$$\frac{\delta f}{\delta x} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

But, what if we make this interval very *very* small? We would then be looking at functions on this very small scale where, graphically, they look like lines. As the interval shrank to zero, or we took

$$\lim_{x_1 \rightarrow x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

we arrive at an instantaneous rate of change.

Recall that a ***derivative*** is the instantaneous rate of change of a (single variable) function. We defined the derivative f' of f at the point x to be

$$f'(x) := \lim_{\delta x \rightarrow 0} \frac{f(x + \delta x) - f(x)}{\delta x}.$$

Can you see how the two limits given are the same? We also had the notation

$$\frac{d}{dx} f(x) = f'(x),$$

which is nice to use at times. It will be especially helpful in the future to use the latter (Leibniz) notation. Recall also that we had a handful of derivative rules. Keep these near and dear to your heart.

Rule	Derivative
Sum	$\frac{d}{dx}(f(x) + g(x)) = f'(x) + g'(x)$
Constant Multiple	$\frac{d}{dx}(cf(x)) = cf'(x)$
Product	$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$
Quotient	$\frac{d}{dx}(f(x)/g(x)) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$
Chain	$\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$
Power	$\frac{d}{dx}x^p = px^{p-1}$
Log	$\frac{d}{dx}\ln(x) = \frac{1}{x}$
Exponential	$\frac{d}{dx}e^x = e^x$
Sine	$\frac{d}{dx}\sin(x) = \cos(x)$
Cosine	$\frac{d}{dx}\cos(x) = -\sin(x)$

These will be all the necessary rules to know moving forward. This list is actually quite redundant. Can you see which you can find from others?

Integrals

Integration is the other fundamental technique learned in an introductory calculus course. The idea is to be able to add up function values over an interval. This helped to also develop the ability to find area under a curve. There are a few things to note here.

Remember that an **integral** can be in the form of **definite integrals** or as **indefinite integrals**. The former returns a number that tells one the *net* area under the graph of function. Indefinite integrals, on the other hand, return a function that we so joyfully refer to as the **anti-derivative**.

We have some integration rules as well, but it is a shorter list. For now.

- Sum rule: $\int f + g dx = \int f dx + \int g dx.$
- Constant multiple rule: $\int c f dx = c \int f dx.$

There is one important theorem that we all need to remember here and before we begin a new voyage. That is the following.

Theorem 1.3.2: Fundamental Theorem(s) of Calculus

Let $f: [a, b] \rightarrow \mathbb{R}$ be a differentiable function with derivative f' . Then we have the following:

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

We can also say that

$$\int f'(x) dx = f(x) + c$$

for some constant $c \in \mathbb{R}$ as well as

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

This theorem is very fundamental to the study of calculus, hence the name!

One should be familiar with how to perform **integration by substitution** or a *u*-substitution. That is, given an integral of the form

$$\int_a^b f(g(x))g'(x) dx$$

one can make a substitution letting $u = g(x)$ and noting that $du = g'(x)dx$. Of course, many integrals are not of this form, but surprisingly many are. Anyways, we find that with this substitution we have

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(u) du.$$

Example 1.3.4: Integration by Substitution

Consider the definite integral

$$\int_1^2 2x \sin(x^2) dx.$$

Notice that the $2x$ looks like the derivative of the inside of $\sin(x^2)$. So we let

$u = x^2 = g(x)$ and we have $du = 2x dx$. Thus we have

$$\begin{aligned}\int_1^2 2x \sin(x^2) dx &= \int_1^4 \sin(u) du \\ &= -\cos(4) + \cos(1).\end{aligned}$$

After this technique comes one that may not be seen in any prerequisite course for this class. This new technique is known as **integration by parts**. Integration by parts is a combination of the derivative product rule and the fundamental theorem of calculus. Given functions $u(x)$ and $v(x)$ we can write

$$(uv)' = u'v + uv'.$$

Then if we integrate both sides, we have

$$\int_a^b (uv)' dx = \int_a^b u'v dx + \int_a^b uv' dx.$$

Fundamental theorem of calculus gives us that

$$\int_a^b (uv)' dx = u(b)v(b) - u(a)v(a)$$

and thus

$$\int_a^b u'v dx = u(b)v(b) - u(a)v(a) - \int_a^b uv' dx.$$

Or, without bounds on the integral, we can write

$$\int u'v dx = uv - \int uv' dx.$$

I like to think of integration by parts as shifting the derivative from one function to another with a penalty term. Above, we swap a derivative on u to a derivative on v but have to correct with the function uv . This can all be derived in higher dimensions using Stokes' theorem. It's an excellent tool in the study of differential equations.

Example 1.3.5: Integration by Parts

Consider the indefinite integral

$$\int xe^x dx.$$

Now x is a function that gets “simpler” when we take a derivative since $\frac{d}{dt}x = 1$. So we wish to pick $v = x$ so $v' = 1$ and hence $u' = e^x$ and we then have $u = e^x$ as well. Plugging this in, we find

$$\begin{aligned}\int xe^x dx &= \int u'v dx = uv - \int uv' dx \\ &= xe^x - \int_e^x dx \\ &= xe^x - e^x + c.\end{aligned}$$

Problems

Problem 1.20. Compute derivatives for the following functions.

- (a) $f(x) = 18x^3 + 2x^2;$
- (b) $g(x) = \sqrt{x+5};$
- (c) $h(x) = \sqrt{\frac{1}{x}};$
- (d) $p(x) = \sqrt[5]{x^3 + 15x};$
- (e) $q(x) = \frac{1}{\sqrt[4]{2x}};$
- (f) $r(x) = \sqrt{x^2}.$ Hint: You can't cancel the square root and the square as they are not proper inverses! Is this function differentiable everywhere? Plot it if need be.

Problem 1.21. Compute derivatives for the following functions.

- (a) $f(x) = xe^x;$
- (b) $g(x) = e^{2x};$
- (c) $h(x) = b^x;$
- (d) $p(x) = (x^2 + 2x) \ln(x);$
- (e) $q(x) = \ln\left(\frac{1}{x}\right);$
- (f) $r(x) = x^x.$ Hint: This is tough. Try taking a natural log of both sides before differentiating.

Problem 1.22. Compute derivatives for the following functions.

- (a) $f(x) = \tan(x);$
- (b) $g(x) = \sec(x);$
- (c) $h(x) = \csc(x);$
- (d) $p(x) = \cot(x);$
- (e) $q(x) = \tan(\sin(\cos(x))).$

Problem 1.23. Compute the following.

- (a) $\frac{d}{dx} (15y + 15x);$
- (b) $\frac{d}{dy} (15y + 15x);$
- (c) $\frac{d}{dr} (15y + 15x);$

- (d) $\frac{d}{dx} (5x^2y + 32y^3 + 5xy^2);$
 (e) $\frac{d}{dy} (5x^2y + 32y^3 + 5xy^2).$

Problem 1.24. Compute the following antiderivatives. *Do not forget the +c!*

- (a) $\int 2xdx;$
 (b) $\int \cos(x)dx;$
 (c) $\int \sec^2(x)dx.$

Problem 1.25. Compute the following integrals.

- (a) $\int_0^2 x^3 dx;$
 (b) $\int_{-1}^1 \sin(x)dx;$
 (c) $\int_{-1}^1 |x|dx.$

***Problem 1.26.** The following is the definition of the natural logarithm:

$$\ln(x) := \int_1^x \frac{1}{x} dx.$$

- (a) Using the Fundamental Theorem of Calculus, show that

$$\frac{d}{dx} \ln(x) = \frac{1}{x}.$$

- (b) Show and explain why

$$\ln(1) = 0.$$

- (c) Show and explain why $\ln(0)$ is undefined.



Tutorials

2.1 Wolfram Alpha

Wolfram Alpha (WA) is an online tool that can perform many useful tasks. Essentially every problem posed in this text could be done using WA. It is therefore worthwhile to learn some basic input. Of course, much more can be done with WA than is presented here.

2.1.1 Solving Algebraic Expressions

An algebraic expression is an equation where one wishes to find a value for a variable (resp. variables) which make the equality in the equation (resp. equations) hold.

Example 2.1.1: One Variable Algebraic Expression

Consider the equation

$$3x^4 + 2x^3 + x - 5 = 0.$$

The goal is to find the value of x so that the left hand side is equal to zero. There is a formula like the quadratic formula one can use to solve this, but it is rather ugly (see: quartic formula). So, it is rather appealing to use technology to quickly solve this for us.

In the WA entry box, type

$$3x^4+2x^3+x-5=0$$

and press enter. Give the computation some time, and verify that you find the

following solutions:

$$\begin{aligned}x &\approx -1.4176 \\x &\approx 0.94346 \\x &\approx 0.09627 - 1.11216i \\x &\approx 0.09627 + 1.11216i.\end{aligned}$$

If we have multiple variables, we can perform a similar computation.

Example 2.1.2: Multivariable Algebraic Expressions

Consider the equations

$$\begin{aligned}xyz &= 1; \\x + y &= 2; \\y + z &= 1/2.\end{aligned}$$

The goal is to find values for x , y , and z so that each equation is satisfied simultaneously. These equations can take time to solve and can be prone to errors. In the WA entry box, type

`xyz=1; x+y=2; y+z=1/2`

and press enter. Once the computation has completed, verify that you find the following solutions

$$x \approx -0.253156, \quad y \approx 2.25316, \quad z \approx -1.75316$$

or

$$x \approx 1.87658 - 0.654667i, \quad y \approx 0.123422 + 0.654667i, \quad z \approx 0.376578 - 0.654667i$$

or

$$x \approx 1.87658 + 0.654667i, \quad y \approx 0.123422 - 0.654667i, \quad z \approx 0.376578 + 0.654667i.$$

2.1.2 Error Checking

There are some issues with the interpreting of input with WA. For example, WA often thinks certain symbols have implicit relationships. One case is that it tends to take y and x to be related in that $y(x)$ is a function of x .

Part I

Ordinary Differential Equations

3

The Complex Numbers

3.1 Introduction to Complex Numbers

Complex numbers are a wonderful part of mathematics. You, as the reader, have seen them before. However, there is a lot of power hiding within this number system. We will start our venture into this course with this topic as it is a set up for a lot to come. Quantum theory utilizes complex numbers right out of the gate, and we will as well.

There is a whole branch of study devoted to the calculus of complex numbers called *complex analysis*. Maybe you'll find this interesting and wish to take Math 419! We will not actually dive into the (slightly different) calculus for these numbers, but rather need them as a tool.

You see, we will always need to be able to factor polynomial equations. This is the biggest benefit to using complex numbers as we will see with the *Fundamental Theorem of Algebra*. We do not prove this theorem, but merely state it. One should always remember fundamental theorems.

Imaginary numbers are those we find by taking the square root of a negative number. By appending imaginary numbers to the set of real numbers \mathbb{R} , we create the complex numbers \mathbb{C} . For us, the ability to find a square root of -1 will be powerful. It will help us solve differential equations and do linear algebra.

Question 3.1.1. What is a complex number? What was the motivation for developing them?

Answer 3.1.1. Let me begin by answering the second part of the question.

Take a polynomial,

$$a_0 + a_1z + a_2z^2 + \cdots + a_nz^n.$$

When does this polynomial equal zero? In other words, what are the *roots* of this polynomial?

The fact of the matter is that the real numbers \mathbb{R} are just not large enough to guarantee that we have all possible roots. Take, for example, the polynomial

$$x^2 + 1.$$

Set this equal to zero and we have

$$\begin{aligned} x^2 + 1 &= 0 \\ \implies x^2 &= -1 \\ \implies x &= \pm\sqrt{-1}. \end{aligned}$$

It seems that we are missing a number in this case. Specifically,

$$\sqrt{-1} \notin \mathbb{R}.$$

Let us take this new number and denote

$$i := \sqrt{-1}.$$

Of course this also means that

$$i^2 = -1.$$

A complex number will be made with real numbers and this new additional member. We call i the **imaginary number**. It's too bad this number has been named imaginary as it is certainly very physical.

Definition 3.1.1: Complex Number; Real and Imaginary Part

We can generate any complex number by combining a real number with an imaginary number. Specifically, we can write a **complex number** z as

$$z = a + bi$$

where $a, b \in \mathbb{R}$. We call this form $z = a + bi$ the **cartesian** or **rectangular form** of a complex number. We also call a the **real part** of z and b the **imaginary part** of z . We will write

$$\operatorname{Re}(z) = a \quad \operatorname{Im}(z) = b,$$

and typically let z denote a complex number. Notice that we do not include i with the imaginary part!

Let us also denote the set of all complex numbers by \mathbb{C} (just as we denoted the real numbers by \mathbb{R}). It's important to remember that there is really no way to simplify this number further. There are other ways to specify a complex numbers but they will have real and imaginary parts like this.

The amazing fact about complex numbers is that they allow us to factor (or find the zeros of) *any* polynomial

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{n-1} z^{n-1} + a_n z^n$$

even when the coefficients a_i are complex themselves. This fact is known as the *fundamental theorem of algebra*.

Theorem 3.1.1: Fundamental Theorem of Algebra

Let $f(z)$ be an a polynomial of degree n . That is,

$$f(z) = a_0 + a_1z + a_2z^2 + \cdots + a_{n-1}z^{n-1} + a_nz^n.$$

Then $f(z)$ always has n complex roots.

3.2 Complex Number Algebra

With a new number system, we must ask how we treat it algebraically. Given two complex numbers $z_1 = a_1 + b_1i$ and $z_2 = a_2 + b_2i$ we can write the following:

- **Addition:** We add complex numbers by

$$z_1 + z_2 = (a_1 + b_1i) + (a_2 + b_2i) = (a_1 + a_2) + (b_1 + b_2)i.$$

That is to say that we add by adding the real parts together and the imaginary parts together.

- **Multiplication:** We can multiply complex numbers in the same way we multiply polynomials. In other words, we distribute. So we have

$$\begin{aligned} z_1 \cdot z_2 &= (a_1 + b_1i) \cdot (a_2 + b_2i) \\ &= a_1a_2 + a_1b_2i + a_2b_1i + b_1b_2(i)^2 \\ &= (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)i. \end{aligned}$$

Notice that we used the fact that $i^2 = -1$.

Exercise 3.2.1. What are the real and imaginary parts of the above two examples?

For real numbers x we had that

$$x + (-x) = 0.$$

We also could find a number called the inverse of x and denoted by x^{-1} so that

$$x \cdot x^{-1} = 1,$$

unless $x = 0$. We can mimic these properties with complex numbers as well. Before we do this, let me introduce one concept first that will prove to be very useful.

Definition 3.2.1: Complex Conjugate

Given a complex number $z = a+bi$ we define the *complex conjugate of z* (denoted z^*) by

$$z^* = a - bi.$$

- The negative of a complex number is just the negative of both the real and imaginary part. So if we are given $z = a + bi$ then the negative is $-z = -a - bi$. We can check that this gives us the desired property by

$$z + (-z) = (a + bi) + (-a - bi) = 0.$$

- To find the inverse of $z = a + bi$ which we call z^{-1} , we write $z^{-1} = \frac{1}{a+bi}$. This way we definitely have

$$z \cdot z^{-1} = \frac{a + bi}{a + bi} = 1.$$

However, we haven't written z^{-1} in our standard form! Let us fix this. Take $z = a + bi$ so that $z^{-1} = \frac{1}{a+bi}$. We can multiply the numerator and denominator by z^* and we find

$$\frac{z^*}{z^* \cdot (a + bi)} = \frac{a - bi}{(a - bi)(a + bi)} = \frac{a - bi}{a^2 + b^2}$$

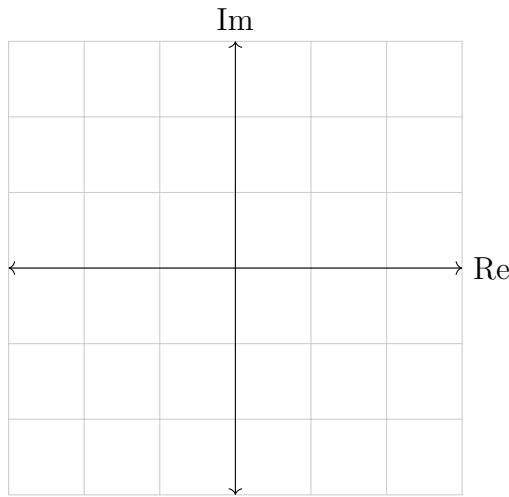
which means that we can write

$$z^{-1} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i.$$

This actually serves to motivate the geometry of complex numbers which we will visit next. As always, having more than one way of understanding a concept helps. One may find that they understand the algebraic or geometric point of view better. We're unique individuals after all!

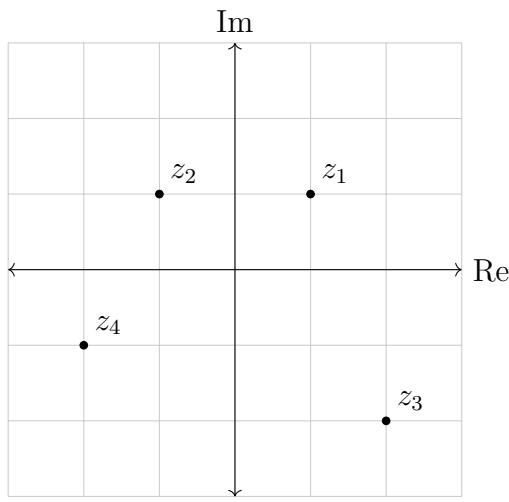
3.3 Geometry of complex numbers

With the algebra out of the way, we can concentrate on the geometry of \mathbb{C} for a bit. The complex numbers turn out to be wonderfully geometrical. We begin with the complex plane \mathbb{C} . The way we usually plot points in \mathbb{C} is by looking at the real and imaginary parts of a complex number z .



So, for example, let us plot a few points in the complex plane. We'll take

$$z_1 = 1 + i \quad z_2 = -1 + i \quad z_3 = 2 - 2i \quad z_4 = -2 - i$$



Let us investigate what our algebra was doing geometrically!

Addition of complex numbers: Recall that if we have $z_1 = a_1 + b_1i$ and $z_2 = a_2 + b_2i$ then

$$z_1 + z_2 = a_1 + a_2 + (b_1 + b_2)i.$$

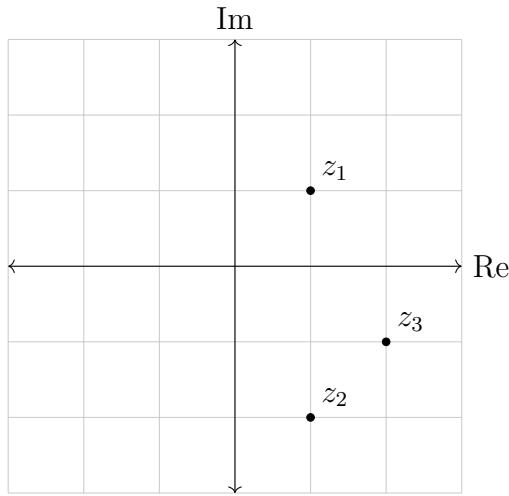
With an explicit example, let us take

$$z_1 = 1 + i \quad z_2 = 1 - 2i$$

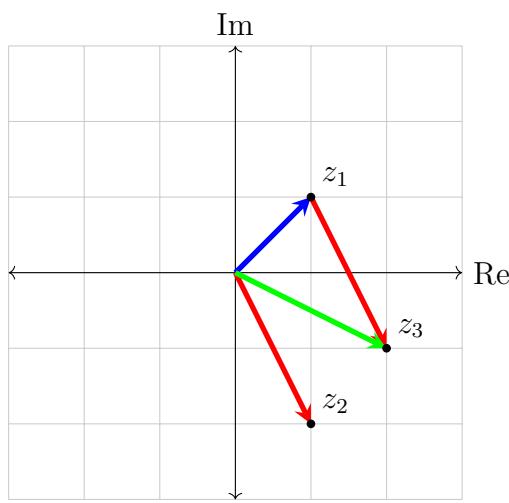
Then

$$z_3 = z_1 + z_2 = 2 - i.$$

We can plot these in the complex plane like so.

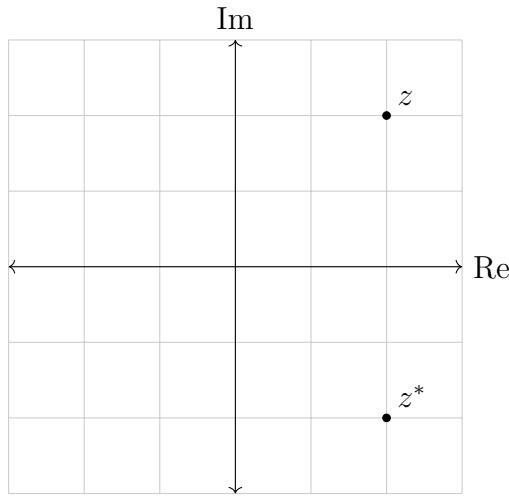


The other way of thinking about this addition is to think of attaching arrows to each other. If we draw an arrow from the origin $z = 0$ to our desired point z_i , we can then use these arrows to perform addition. If we take a copy of the arrow from the origin to z_2 and glue it to the head of the arrow from the origin to z_1 we arrive at our point z_3 .



How about the complex conjugate? What was this doing? Given a complex number $z = a + bi$, we know that $z^* = a - bi$. So, take the example

$$z = 2 + 2i \quad z^* = 2 - 2i.$$



Notice that complex conjugation is just reflection about the real axis!

3.4 Polar coordinates

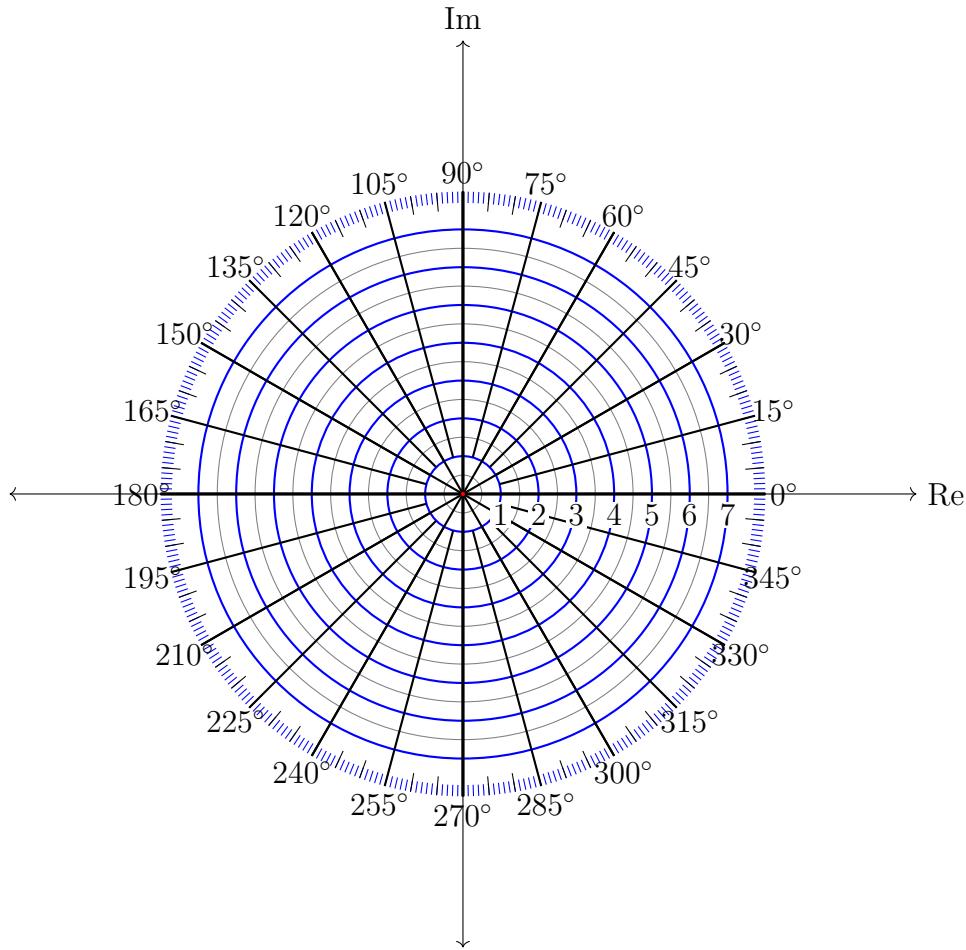
An extremely important notion in planar geometry is that of *polar coordinates*. Usually, we are happy to reference a point in the plane by giving the x -coordinates and y -coordinates. In the complex case, we can give the Re and Im parts of the complex number to specify its location in the complex plane \mathbb{C} .

There are many different ways you could choose to represent points in the plane, but almost all of them are rather silly to a human. However, one that provides a great bit of intuition is polar coordinates. The idea is as follows:

Definition 3.4.1: Polar Coordinates

Polar coordinates are the a coordinate system that uses r , the distance from the origin, and θ , the counter-clockwise angle measured from the Re axis in the complex plane \mathbb{C} (or x -axis in \mathbb{R}^2).

The following image may prove to be helpful.



In this picture you can see that coordinates are given by a distance from the origin and an angle. Of course, we remember that we tend to prefer radians, so in this case we would have

$$0^\circ = 0 \quad 45^\circ = \frac{\pi}{4} \quad 90^\circ = \frac{\pi}{2} \quad 180^\circ = \pi.$$

Question 3.4.1. Do you know where radians come from? Why are they a natural choice for angle?

Answer 3.4.1. If we consider the unit circle (a circle with radius one), then its circumference is 2π . Thus, the “angle” around the unit circle you travel can just be given by the amount of the circumference you have traveled around.

Exercise 3.4.1. Identify all the degrees on the above figure with their values in radians.

Question 3.4.2. If we are given a point in cartesian coordinates (specifying a Re and Im part in \mathbb{C} (or (x, y) in \mathbb{R}^2), how can we find the polar coordinates (r, θ) ? How about vice-versa?

Answer 3.4.2. Coming to you now!

3.4.1 Euler's Formula

Taylor Series*

This is a brief interlude on something we will get to later in this class. *Taylor series* are necessary for the *Euler's formula*. Right now I'm just writing it so that I avoid lying to you! Fundamentally, functions we care about can be approximated by their derivatives at a single point. For example, take the function e^x . It turns out, we can write this function in a new way. That is in the form of a *power series*. In fact, we have

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

More generally, for other functions (with some sufficient conditions I don't care to get into) we can write:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x - c)^n.$$

The latter is known as the *Taylor series for f centered at $x = c$* . You really may wonder why in the world I've just introduced this rather abstract concept, but it plays a fundamental role in the complex world.

Using the above Taylor series for e^x , we can plug in the following:

$$e^{i\theta} = \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!}.$$

It turns out that we find

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

This result is known as **Euler's formula**. We will do the rest of the work later! Just believe me for now.

It turns out that Euler's formula gives us all the necessary tools to represent numbers in the complex plane in a cartesian or polar way. So let's do a bit of work to reach this conclusion.

Definition 3.4.2: Modulus

Given a complex number z , the **modulus** $|z|$ (think *length*) of a complex number is given by

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

Letting $z = a + bi$ we find

$$|z| = \sqrt{a^2 + b^2},$$

which is not entirely shocking. Just like the lengths of vectors in the plane, the "length" of a complex number follows suit.

Question 3.4.3. Given any θ , what is the modulus of $z = e^{i\theta}$?

Answer 3.4.3. We take

$$z = e^{i\theta} = \cos \theta + i \sin \theta.$$

Then

$$zz^* = \cos^2 \theta + \sin^2 \theta = 1.$$

So

$$|z| = \sqrt{1} = 1.$$

It turns out that there is a fundamental relationship underlying the complex numbers. The moral of the story is that imaginary parts of complex numbers cause rotations while the real parts cause scaling. This isn't entirely the case, but it's an alright analogy. We'll be able to uncover this as we explore polar coordinates in more depth.

Remark 3.4.1. Something should be said here about the idea of coordinates in general. Points in a plane are just that – points. Where they are located relative to each other is the fundamental structure whereas how we choose to measure this distance is not.

Polar coordinates are just another way of specifying where a point is located. This choice of coordinates should not change the behavior of the algebra we have developed or any of the geometry. Does choosing to measure an object in inches versus centimeters change the length of the object? Of course not.

3.4.2 Polar Coordinates and Transformations

The fact that we can write

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

means that we can control the angle of a line that the complex number lives on. Since the above function lies on the unit circle in \mathbb{C} , we can scale this with a real number to reach any point in \mathbb{C} . This motivates the following.

Definition 3.4.3: Polar Coordinates

An equivalent way to represent a point in \mathbb{C} is to specify its **argument** (angle, denoted $\arg(z)$) θ and its **modulus** (distance from the origin, denoted r or $|z|$) r and write

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

We call this coordinate representation **polar coordinates**.

Before, we had always written this in a cartesian way, like

$$z = a + bi.$$

But we also have Euler's formula, which means that

$$a + bi = re^{i\theta} = r \cos \theta + ir \sin \theta.$$

This means that

$$a = r \cos \theta$$

and

$$b = r \sin \theta.$$

You can also see that this is really nothing more than the Pythagorean theorem.

Exercise 3.4.2. Pick a point z in the complex plane. Construct a right triangle where the hypotenuse goes from 0 to z and the other two sides run parallel to the Re and Im axes. Can you find r and θ from this drawing? (*Hint: you will need to use facts from the trigonometry section in the introduction.*)

Conversion from polar to cartesian: If you're given a point in \mathbb{C} in polar coordinates,

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

then in cartesian coordinates it will read

$$z = r \cos \theta + ir \sin \theta.$$

This is

$$a = r \cos \theta \quad b = r \sin \theta.$$

Conversion from cartesian to polar: If you're given a point in \mathbb{C} in cartesian coordinates,

$$z = a + bi,$$

then in polar coordinates it will read

$$z = |z|e^{i \arctan(b/a)},$$

when we have $a > 0$. When we have $a < 0$ then we have

$$z = -|z|e^{i \arctan(b/a)} = \|z\|e^{i(\arctan(b/a) + \pi)}.$$

That is that the modulus and argument follow

$$r = |z| = \sqrt{zz^*} \quad \theta = \arctan\left(\frac{b}{a}\right)$$

when $a > 0$ and again we have

$$\theta = \arctan\left(\frac{b}{a}\right) + \pi$$

when $a < 0$. We defined $|z|$ to be the distance z is from the origin, so this is clear. The $\arctan(b/a)$ can be realized from the trigonometric fact that

$$\tan \theta = \frac{b}{a}$$

and inverting. If we have that $a = 0$, then $\theta = \frac{\pi}{2}$ or $\theta = \frac{3\pi}{2}$. Since we have

$$e^{i\frac{\pi}{2}} = i \quad \text{and} \quad e^{i\frac{3\pi}{2}} = -i,$$

we need only look at the sign of b . If $b > 0$ then $\theta = \frac{\pi}{2}$ and if $b < 0$ then $\theta = \frac{3\pi}{2}$.

3.4.3 Multiplication of complex numbers in polar form

While we were perfectly happy to multiply complex numbers in Cartesian form, it is more illuminating to multiply them in polar form. Let us take two complex numbers

$$z_1 = r_1 e^{i\theta_1} \quad z_2 = r_2 e^{i\theta_2}.$$

Then we can write the product:

$$z = z_1 z_2 = (r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

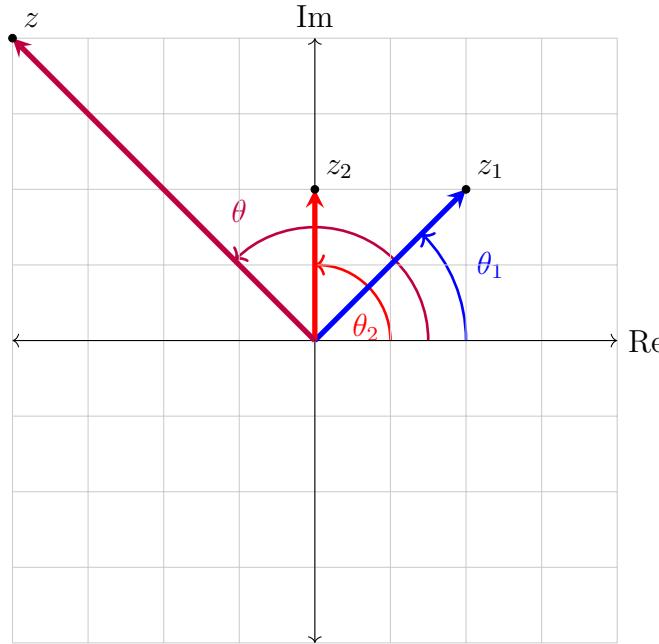
The thing to notice here is that the modulus of a product follows

$$|z| = |z_1||z_2| = r_1 r_2$$

and the argument of a product follows

$$\arg(z) = \arg(z_1) + \arg(z_2).$$

All of this is to say to the following: Multiplication of complex numbers is decomposed into rotation and scaling.



Inverse in polar form

We found before that given a complex number z , we can find the inverse $z^{-1} = 1/z$ in cartesian form. But this was a bit of a headache. In polar form, however, it is much easier. Recall that if we have

$$z_1 z_2 = (r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

Now, given $z = re^{i\theta}$ we can find fairly quickly that z^{-1} must be

$$z^{-1} = \frac{1}{r} e^{-i\theta}.$$

Exercise 3.4.3. Show that this is indeed the inverse. (*Hint: take our example z^{-1} and multiply it by $z = re^{i\theta}$ and see that you get one.*)

3.5 Periodic Motion

One of the greatest benefits to complex numbers is the ability to naturally capture rotations. When we wrote complex numbers in polar form, the rotation action became more clear. This ability makes complex numbers very useful in systems that rotate, oscillate, or have some kind of *periodic* motion.

Question 3.5.1. What is periodic motion?

Answer 3.5.1. Periodic motion is when a system undergoes a change that repeats itself. For example, a mass attached to a spring will vibrate back and forth. As the mass moves from one side to the other, it eventually gets back to where it started and repeats the process again (with no friction, of course).

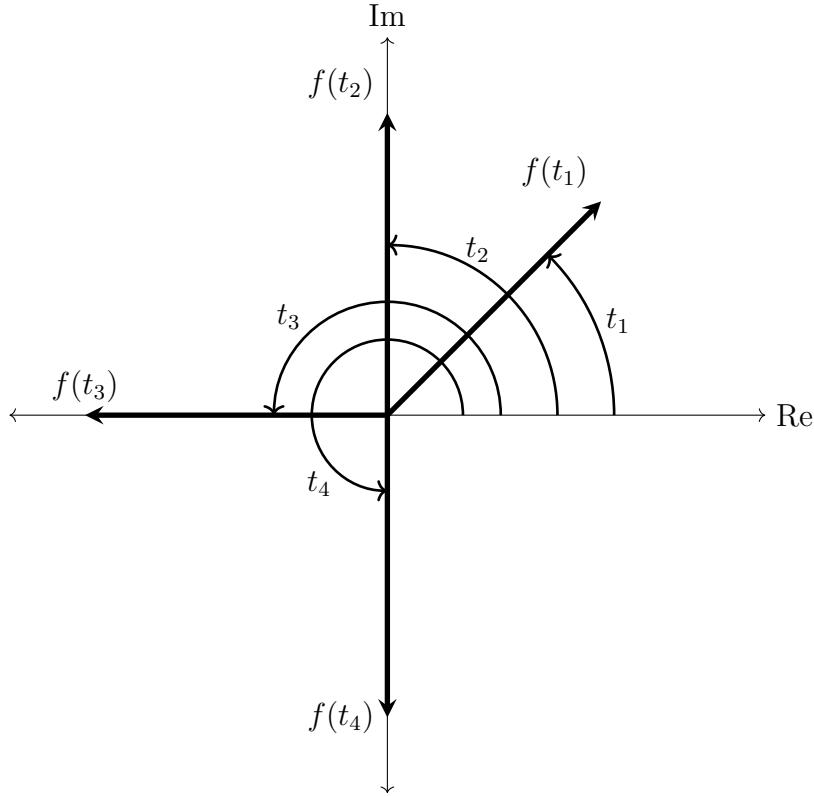
There is a great example of periodic motion in the complex numbers that we should take a look at.

Example 3.5.1: Circular Motion

Let us consider the function $f: \mathbb{R} \rightarrow \mathbb{C}$ given by

$$f(t) = e^{it}.$$

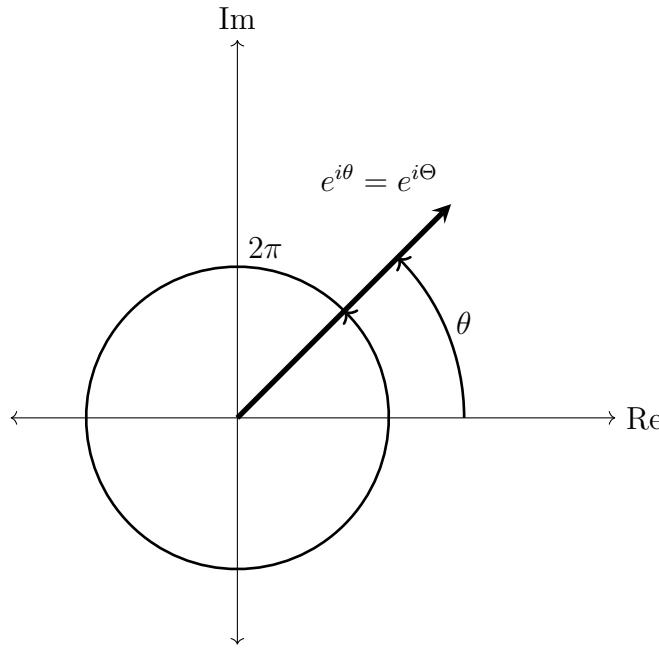
What happens as t increases? Let's let $0 < t_1 < t_2 < t_3 < t_4 < 2\pi$ for now and we can plot this



Notice that by choosing t in this way we have that t corresponds to the angle of rotation θ .

Question 3.5.2. What happens when we let $t > 2\pi$?

Answer 3.5.2. We come back around the circle. For example, if we take $\theta = \frac{\pi}{4}$ and $\Theta = 2\pi + \frac{\pi}{4}$ and plug these values in to e^{it} then we find:



We can see that if we add 2π to any angle θ we get back to the original angle.

Definition 3.5.1: Periodic Function

A function f with real inputs is **periodic** with period T if

$$f(x) = f(x + T)$$

for all x .

Exercise 3.5.1. Realize the function $f: \mathbb{R} \rightarrow \mathbb{C}$ given by

$$f(t) = e^{it}$$

as periodic with period 2π .

Exercise 3.5.2. Can you name two more periodic functions with period 2π ?

Exercise 3.5.3. Show that the function

$$g(t) = e^{i\frac{2\pi t}{T}}$$

is periodic with period T .

Definition 3.5.2: Frequency

Let f be a periodic function with period T . Then its *frequency* is

$$\nu = \frac{1}{T}.$$

Problems

Problem 3.1. Complete the exercises throughout the chapter.

Problem 3.2. Find the point(s) of intersection of the parabola $f(x) = 2x^2 + 2x + 2$ and the line $g(x) = 4x + 4$ using algebra. Plot this on Desmos to confirm your answer and include a picture of the two graphs.

Problem 3.3. Now take the same parabola $f(x) = 2x^2 + 2x + 2$ and the line $h(x) = 4x - 4$. Plot this on Desmos and include a picture to show that the graphs of these two functions do not intersect in the plane. Explain why this must be true.

We can find “complex intersections” by doing the same algebra as the previous problem. Find these complex intersections. (*Hint: set up an equation whose roots would give you the intersections of these two curves.*)

Problem 3.4. Convert the following numbers in Cartesian coordinates to polar coordinates and compute all pairwise products after converting.

(a) $z_1 = \frac{1}{2} - \frac{1}{2}i$;

(b) $z_2 = -1 + 3i$;

(c) $z_3 = -2 - 3i$.

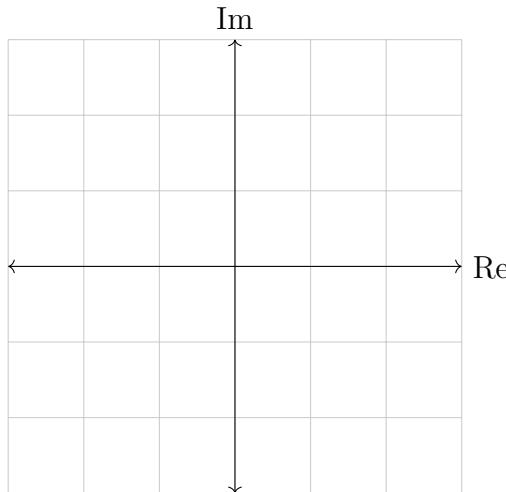
Problem 3.5. Find the square roots of $-i$ using a geometrical argument. *Hint: think of polar coordinates and rotation!*

Problem 3.6. Draw the unit circle in the complex plane. Plot the complex numbers z_1 and z_2 given above and find their inverses and conjugates. Explain what taking the inverse and conjugate does geometrically.

Problem 3.7. Add and multiply all possible pairs of the complex numbers

$$z_1 = 3 - 2i \quad z_2 = -1 - i \quad z_3 = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}} \quad z_4 = -\pi + \pi i.$$

Problem 3.8. Plot and label the above points on the graph below. Pick two points and draw their sum geometrically.

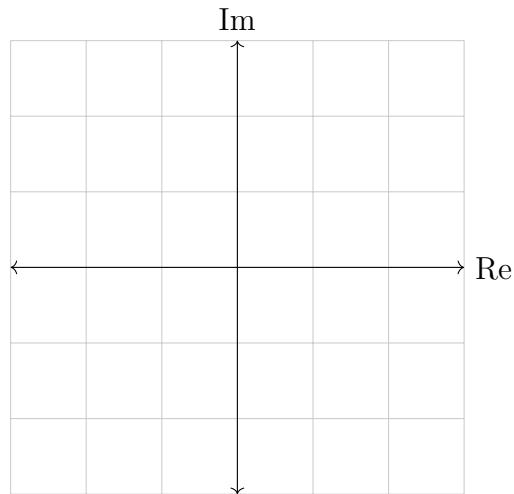


Problem 3.9. Convert all the above complex numbers in Cartesian form to polar form.

Problem 3.10. Multiply all possible pairs of complex numbers

$$w_1 = e^{i\pi} \quad w_2 = -\sqrt{2}e^{i\frac{\pi}{4}} \quad w_3 = 2e^{-i\frac{\pi}{3}} \quad w_4 = 3e^{i\frac{\pi}{2}}.$$

Problem 3.11. Plot and label the above points in the graph below. Draw the product w_1w_2 geometrically.



Problem 3.12. Show by a geometrical argument that $re^{i\theta} = re^{i(\theta+2n\pi)}$ for any integer value of n . Can you also show this by the typical conversion from polar to Cartesian?

Problem 3.13. Let $z = a + bi$. What is z^* ? What is z in polar coordinates? How about z^* ? Can you explain why zz^* will always be real using a geometrical (polar coordinate) argument?

Problem 3.14. Show that the functions $\sin(x)$ and $\cos(x)$ are periodic and determine the period. Is the function $\tan(x)$ periodic?

Problem 3.15. (Roots of unity) Given that $i^2 = -1$ we can factor equations in ways that are totally new to us. Solve the following. Geometrical reasoning may really help you here!

- (a) Find all solutions to $z^2 = 1$.
- (b) Find all solutions to $z^3 = 1$.
- (c) Find all solutions to $z^4 = 1$.
- (d) (Challenge) Find all solutions to $z^n = 1$.

Hint: each of the above has n solutions for z^n .

Problem 3.16. (Rotational symmetries) Consider a complex number $z = a + bi = re^{i\theta}$. What happens to this point as we repeatedly multiply by i ? Pick a specific z (i.e., fix a and b or r and θ) and multiply by i until you see a pattern. What is happening? *Note: this is an example of a symmetry group or a discrete dynamical system!*

Problem 3.17. (Bonus) The complex numbers can be thought of as a 2-dimensional number system. The **quaternions** are a 4-dimensional number system that are of the form

$$q = a + bi + cj + dk.$$

We define $i^2 = j^2 = k^2 = ijk = -1$. Pick another quaternion $\tilde{q} = \tilde{a} + \tilde{b}i + \tilde{c}j + \tilde{d}k$ and multiply each together. Is multiplication commutative? *Note: quaternions are extremely useful in 3-dimensional systems. They are used for describing computer graphics, robotics, relativity, spinning objects, or they also provide an algebraic way for formalizing the vector calculus in \mathbb{R}^3 which we will see in Math 272.*



First and Second Order Dynamics

Roughly stated, differential equations is the study of a system that undergoes change. This change can depend on time, space, or both. The history of differential equations begins with Newton's study of classical mechanics. Newton was investigating the motion of objects in space and derived the first examples of differential equations. The field quickly grew with the once other scientists noticed the wide scope of applicability of differential equations.

In the time of Newton's classical mechanics, we saw the advent of celestial, fluid, and continuum mechanics. Examples include the diffusion, advection, and wave equations. Other fields of science began to use these ideas from physics to model population dynamics and chemical reaction. As time moved forward, more complicated physical interactions brought even more uses of differential equations to the forefront. These were the dynamical theories of electromagnetism and thermodynamics. All of this happened before the turn of the 20th century.

As we moved into the 1900s, there was the boom of modern physics with work from Einstein and many other scientists. Einstein described the motion of atoms in a probabilistic manner that was also deeply related to differential equations of classical mechanics. Soon after, he then considered how spacetime itself behaves as a coupled dynamical continuum, much like the head of a drum that vibrates. It was shortly after this discovery that Schrödinger and Heisenberg independently developed the theories of quantum mechanics. Both stated the problem in different (but equivalent) ways. This study brought together the notion of motion of particles with that of waves.

Time passed, and in the mid 1900s computers were developed. This forever changed the study of differential equations. The problem was, we were finding that (other than very specific nice examples) most differential equations were extremely hard to solve. In fact, many are so wild that the dynamics we see is volatile to the point of the so called "butterfly effect." These volatile systems are known as chaotic and are abundant in nature. Weather, population, and chemical reaction are all areas where chaos can show up. However, the

ability to approximate solutions with computers allows us to make reasonable predictions and essentially solve problems that were previously deemed impossible.

The goal for us is not to learn to solve many differential equations with a handbook of techniques. These techniques can be readily found online and are very formulaic. If you do them once, you can do them again. Instead, our goal is to understand what differential equations model and what they say about systems. Of course, we will explicitly solve some and see a few techniques, but that is not the emphasis. If one pursues mathematical modelling, one will almost surely be working with computers to solve problems rather than by hand. It is with this mentality that we carry on to uncovering this structure of differential equations.

4.1 Ordinary Differential Equations

The first stop on the study of differential equations are the Ordinary Differential Equations (ODEs). ODEs are equations that involve a single independent variable t that we usually think of as time, a function (or dependent variable) $x(t)$, and derivatives of the function $x'(t)$, $x''(t)$, $x'''(t)$, up to n derivatives $x^{(n)}(t)$. We will not worry ourselves with the higher order equations yet.

Unlike previous problems where we solve for a variable, or compute derivatives, we wish to find a function that satisfies the differential equation. So, our aim is to find $x(t)$ given our understanding of how x changes over time.

Example 4.1.1: First Order ODE

Suppose we are asked to find a function $x(t)$ that satisfies the following Ordinary Differential Equation (ODE):

$$x'(t) = f(t, x(t)),$$

or with reduced notation

$$x' = f(t, x).$$

This is an example of the most general *first order* ODE.

- Written in English, this equation says, “what function $x(t)$ has a derivative that is equal to a function of t and $x(t)$? ”
- The function $f(t, x)$ is a function of two variables.

Often we will suppress some notation and let $x(t)$ be denoted by x . We cannot forget that x is a function of t , but it will be notationally more convenient to make this substitution. We should also say what a function of two variables is briefly.

For example, we can take a function $f(t, x)$ and specify what it does with input values t and x simultaneously. A specific example could be

$$f(t, x) = tx,$$

or

$$g(t, x) = t \sin(x^2),$$

or

$$h(t, x) = e^t \sqrt{x^2 + tx + t^2}.$$

Exercise 4.1.1. Write your own function $p(t, x)$.

If we let the right hand side be given by $f(t)$, we arrive at the most fundamental example of a differential equation. In fact, this is an equation that you have not only already seen, but it is one that you know how to solve. Indeed, consider the expression

$$x'(t) = f(t)$$

Here we see that we know the derivative of some function $x(t)$ is given by another function of t . If the desire is to determine an expression for $x(t)$, then we can solve these equations simply through integration. That is,

$$x(t) = \int f(t) dt.$$

Thus, $x(t)$ is simply the antiderivative of $f(t)$ in this case. Notice, that we will have a constant of integration in this expression.

However, not all differential equations come out this nicely. The techniques needed to solve more complicated expressions are going to be developed in the succeeding sections. Our attempts will be to classify the types of equations and find methods to find solutions for those types of equations. In general, we will stick to the most nicely behaved equations that appear throughout the natural world.

Definition 4.1.1: Order of an ODE

The *order* of an ODE is the highest derivative that appears in the ODE.

Though right now we will not investigate higher order ODEs, we should at least know what order means. Keep this in mind as we progress. It turns out (though we don't necessarily get to it in this course) that higher order ODEs will be equivalent to many first order ODEs.

Example 4.1.2: Exponential Growth and Decay

As opposed to a very general set up, let us consider the following problem statement.

The concentration of Plutonium in a vessel is measured over time. It's found that the rate of change of this concentration is proportional to the current concentration. What ODE models this situation?

The answer to the above question is

$$x' = kx.$$

- We let $x(t)$ represent the concentration of Plutonium at time t .
- The rate of change of x , $x'(t)$, is related to the current concentration x by a proportion k .

We like to choose our variable names in order to best communicate information. In some cases, x is not the best function name and t is not the best variable name. Try to understand the role of notation along the way. One may see equations written differently in specific contexts. Take a look at the next example.

Example 4.1.3: Mechanical Law

Newton's study of the motion of bodies brought him to say the following.

The change in velocity of a body is proportional to the force applied divided by the inertial mass of the body.

The equation that models this is

$$v'(t) = \frac{1}{m}F(t).$$

- The v represents the body's velocity at the time t and thus v' is the change in velocity.
- The change in velocity should be equal to the applied force, $F(t)$ but also dependent on the objects mass m .
- We could also describe this equation by noting the fact that v is the derivative of the position x . This gives

$$x''(t) = \frac{1}{m}F(t).$$

In other words, this is

$$ma = F.$$

Example 4.1.4: Harmonic Motion

There are many systems that are not first order. For example, we might have the following.

A spring has a rest length L . The force on a mass on a spring is proportional to the displacement from this rest length L in the direction opposite the displacement. The force causes an acceleration proportional to the force applied dived by the inertial mass of the body.

The governing ODE is

$$y''(t) = -\frac{k}{m}(y(t) - L).$$

The variable y was chosen here as this equation can be written in a more standard form with a substitution.

4.2 Solutions to an ODE

What do we mean when we say that we want to “solve an ODE?” This means we want to find a function whose derivatives satisfy our ODE. Let us see a few examples.

Example 4.2.1: Exponential Growth and Decay Solution

Previously, we were given the equation

$$x'(t) = kx(t).$$

I claim that

$$x(t) = Ae^{kt}$$

is a *general solution* to this ODE for any choice of A .

To verify this, we have to take the derivative of our claimed answer $x(t)$. We have,

$$x'(t) = \frac{d}{dt} (Ae^{kt}) = Ake^{kt} = kx(t),$$

so indeed $x(t)$ is a solution.

Example 4.2.2: Mechanical Law Solution

Let us suppose that there is no force acting on the object. We should all believe the object should move in a straight line. With the condition of no force, the equation reads

$$v'(t) = 0.$$

I claim that

$$v(t) = c,$$

with c a constant, is a solution to this equation.

To verify this, take

$$v'(t) = \frac{d}{dt}c = 0.$$

Indeed, this is a solution. The solution is that of a straight line in space. We can see this more easily by noting we also have the ODE

$$x'(t) = v(t),$$

since the rate of change of position is velocity. Again, I claim that

$$x(t) = ct$$

is a solution that is a straight line in space.

Exercise 4.2.1. Verify that $x(t)$ above is indeed a solution to

$$x'(t) = x(t).$$

Can you see how this means that a particle undergoing no force travels in a straight line?

Though we aren't solving equations yet, it is good practice to understand how to see when we have a solution. Follow my methods in previous examples and do the following exercise.

Exercise 4.2.2. Earlier, we were given the following

$$y''(t) = \frac{-k}{m}(y(t) - L).$$

Show that

$$y(t) = Ae^{i\sqrt{\frac{k}{m}}t}$$

solves the above differential equation.

4.3 General and Particular Solutions

In all previous examples, there were undetermined constants. These constants appear since, fundamentally, an antiderivative is determined up to a constant. Though not all ODE are solvable by direct integration, the constants are there due to this reason.

How do we determine the constants? It turns out we need a bit more information. The extra information is also very intuitive and physical. Take for example, a simplified version of the harmonic oscillator (spring-mass) system

$$u''(t) = -u(t).$$

Notice, we have just removed the constants. This is always possible to do by picking the right way to measure the problem! Then I stated that the *general solution* to this ODE is

$$u(t) = Ae^{it}.$$

But, we don't know A , which is a complex number. Here, having written the solution this way is possibly confusing. How is something that's oscillating really being described with complex numbers? It should be said that there are no "complex measuring sticks" in the real world (that we know of). So, our answer should end up being real valued in the end! It turns out, this solution is too general for what we think of as being physically real.

Definition 4.3.1: General Solution

A **general solution** to an n th order ODE is a solution with n (call these values c_1, c_2, \dots, c_n) undetermined constants. A general solution is in fact a whole family of solutions. That is, there is a solution for each different value of the constants c_1, c_2, \dots, c_n .

If we think about the situation, we have just determined the general oscillatory behavior of the system, but not the any particular *trajectory* of the system. What we need to know is where we pulled the mass to at the initial time $t = 0$. That is, we need

$$x(0)$$

but we also need to know how fast it was moving at that point

$$x'(0).$$

The analogy is as follows: If one is throwing a ball, one needs to know where it is released $\mathbf{x}(0)$, and the velocity at which it is released at $\mathbf{x}(0)$ in order to know where the ball will land.

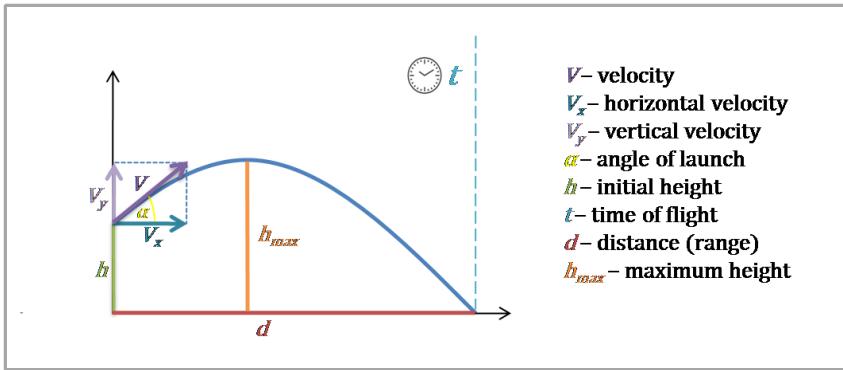


Figure 4.1: Throwing a ball as an example of needing initial data.

We call these values $x(0)$ and $x'(0)$ the *initial data*.

Definition 4.3.2: Initial Data

The *initial data* to an n th order ODE are the specific values for

$$x(0), x'(0), \dots, x^{(n-1)}(0),$$

where $x^{(k)}$ represents the k th derivative of $x(t)$.

It is from this initial data that we can figure out what a *particular solution* to a problem should be. Up to this point, we have found a general solution which is really a family of solutions. For real problems, we need one answer and not infinitely many.

Definition 4.3.3: Particular Solution

A *particular solution* is one member of the family of general solutions. That is, a solution where the constants c_1, c_2, \dots, c_n are all uniquely determined.

In other words, the particular solution is a single solution to a problem. Specifically, we call this type of problem where we have initial data present a *initial value problem*. We will want to make note of this name as we will see *boundary value problems* later on.

Proposition 4.3.1: Initial Data and Particular Solutions

In order to find a particular solution to an n th order ODE, one needs to know the initial values of the function, its derivative, its second derivative, and all derivatives up to the $(n - 1)$ th derivative

$$x(0), x'(0), \dots, x^{(n-1)}(0).$$

Now, let us work through this a bit. We can set up an initial value problem and from a general solution we can narrow in on a particular solution. Try to keep in mind what the initial data is really meaning based on the analogy given before.

Example 4.3.1: Particular Solution to Harmonic Oscillator

Consider

$$u''(t) = -u(t)$$

with initial data

$$u(0) = 1, u'(0) = 0.$$

This has the general solution

$$u(t) = Ae^{it}.$$

We can find the particular solution by rewriting x slightly using Euler's formula. In particular, we find

$$u(t) = a \cos(t) + ib \sin(t)$$

since A could be any complex number. You could also not make this substitution and find what complex number A has to be.

$$u(0) = a \cos(0) + ib \sin(0) = 1$$

from our initial data. Specifically, this gives us that

$$a =$$

Then note we also have

$$u'(0) = -a \sin(0) + ib \cos(0) = 0$$

which gives us that

$$ib = 0 \implies b = 0.$$

So our particular solution is

$$u(t) = \cos(t).$$

Exercise 4.3.1. Plot the solution. Can you determine what is happening if we think of the oscillator as a spring mass system? What does the initial data tell us about the spring and mass at the start?

Exercise 4.3.2. Instead of making the substitution using Euler's formula, solve for the

complex number A that shows up in $u(t) = Ae^{it}$.

The reason why we study these differential systems is to make predictions and models. Given that, our predictions must be sensible. This means if we are given a differential equation and initial data, there should only be one particular solution. This is known as **determinism**. Not all systems are deterministic. But it turns out the non-deterministic systems are either problematic as models or just very hard to deal with. When solving an ODE, we often call a particular solution a **trajectory**.

4.4 Separable Equations

The nicest possible ODEs come from equations that are *separable*. What this means is that, for example, we have a first order equation like

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

Why is this nice? Well, in particular, it means that we can simply integrate this equation to solve it. Many systems exhibit symmetry that allows for this type of separation, so this technique is crucial.

In general, we have

$$x' = \frac{dx}{dt}$$

and we put

$$\begin{aligned} \frac{dx}{dt} &= f(t)g(x) \\ \iff \frac{dx}{g(x)} &= f(t)dt. \end{aligned}$$

With this, we can integrate both sides, then solve for x . That is, we compute

$$\int \frac{dx}{g(x)} = \int f(t)dt,$$

and we will have an equation where we can isolate x

The nicest separable equations are those which we can integrate directly. Recall that the ODE

$$x'(t) = f(t)$$

can be solved directly via integration. That is,

$$x'(t) = \int f(t)dt.$$

Thus, the problem of finding an antiderivative in a first semester calculus course is just one example of a differential equation. Let's see an example of this.

Example 4.4.1: Antiderivative

Consider the equation

$$x'(t) = \sin(t).$$

This equation is separable since the right hand side is only a function of t . That is, $g(x) = 1$ and $f(t) = \sin(t)$. Thus, if we follow the separation method, we have

$$\begin{aligned} \frac{dx}{dt} &= \sin(t) \\ dx &= \sin(t)dt \\ \int dx &= \int \sin(t)dt \\ x &= -\cos(t) + c. \end{aligned}$$

This solution we have found is indeed just the antiderivative of $\sin(t)$. Notice that we also pick up an unknown constant of integration. This constant could be determined if we were given an initial condition.

Antiderivatives are then the easiest possible differential equations since finding their solution amounts to just taking an indefinite integral. But, separable equations in general will have the same basic technique but with an added $g(x)$ appearing. The next example will show how you can handle the additional inclusion of the dependent variable x on the right hand side.

Example 4.4.2: Separable ODE

Consider the following ODE where we are trying to find $x(t)$ that solves

$$x' = \frac{t}{x}.$$

Then we can put

$$\begin{aligned} \frac{dx}{dt} &= \frac{t}{x} \\ \iff xdx &= tdt. \end{aligned}$$

Then we can take the antiderivative of both sides and find

$$\begin{aligned} \int xdx &= \int tdt \\ \iff \frac{1}{2}x^2 &= \frac{1}{2}t^2 + c \\ \iff x &= \sqrt{t^2 + 2c}. \end{aligned}$$

So now we can verify that this is in fact a solution to our ODE. So we take

$$x'(t) = \frac{d}{dt}\sqrt{t^2 + 2c} = \frac{t}{\sqrt{t^2 + 2c}} = \frac{t}{x}.$$

Exercise 4.4.1. Solve the following ODE using separation

$$x'(t) = t.$$

Note that there will be an undetermined constant that we will learn how to handle next.

4.5 Changing Variables and Symmetry

When presented with a differential equation, it can often be in a form that is not the easiest to work with. Of course, if you do your modelling step correctly you will end up with a meaningful equation. We want to be able to translate the correct model equation into one that is more workable.

We found that separable equations were not too bad to solve as we really just need to integrate. Now, is it possible to turn an equation into a one that is separable? It turns out, we can when the equation satisfies a certain symmetry condition.

Proposition 4.5.1: Reduction to Separable Equations

Consider the first order equation

$$x' = f(x, t).$$

If we have that

$$f(x, t) = f(\lambda x, \lambda t)$$

for any number (or function) λ , then we can reduce the equation to a separable one by defining a new variable $u = \frac{x}{t}$.

The idea of changing variables is immensely important in solving real world problems. Often times, one can gain insight on the question at hand by searching for these “better” variables to work in.

Example 4.5.1: Reduction to Separable

Consider the differential equation

$$x' = \frac{x^2 + t^2}{xt}.$$

We can then define

$$f(x, t) = \frac{x^2 + t^2}{xt}.$$

Now, we can check that $f(x, t) = f(\lambda x, \lambda t)$ for any λ by plugging in. We have

$$\begin{aligned} f(\lambda x, \lambda t) &= \frac{(\lambda x)^2 + (\lambda t)^2}{(\lambda x)(\lambda t)} \\ &= \frac{\lambda^2(x^2 + t^2)}{\lambda^2 xt} \\ &= \frac{x^2 + t^2}{xt} \\ &= f(x, t). \end{aligned}$$

So our differential equation satisfies the necessary symmetry condition. So, we can let $u = \frac{x}{t}$ which also means that $x = tu$ and hence

$$f(x, t) = f(tu, t) = \frac{t^2 u^2 + t^2}{ut^2} = \frac{u^2 + 1}{u}.$$

Note that given $x = tu$ we have $x' = u + tu'$ which we can now plug into our original expression

$$x' = f(x, t)$$

to get

$$u + tu' = \frac{u^2 + 1}{u}.$$

Then we can rearrange to isolate u' on the right hand side

$$\begin{aligned} u + tu' &= \frac{u^2 + 1}{u} \\ tu' &= \frac{u^2 + 1}{u} - u \\ u' &= \frac{1}{tu}. \end{aligned}$$

This is now a separable equation! So, we can solve this in the typical way by noting we have $u' = \frac{du}{dt}$ and putting

$$udu = \frac{dt}{t}.$$

Then we can integrate and solve for u

$$\begin{aligned} \int u du &= \int \frac{dt}{t} \\ \frac{1}{2}u^2 &= \ln(t) + c \\ u^2 &= 2 \ln(t) + 2c \\ u &= \pm \sqrt{2 \ln(t) + 2c}. \end{aligned}$$

Recall that $u = \frac{x}{t}$ and so we have

$$\begin{aligned} \frac{x}{t} &= \pm \sqrt{2 \ln(t) + 2c} \\ x &= \pm t \sqrt{2 \ln(t) + 2c}. \end{aligned}$$

This x is our general solution to the original problem.

Another way that symmetry can help us is more from a modelling perspective. For example, it is possible to make more clever measurements of your system! This is a much needed tool for someone keen on doing experiments. For now, do your best to follow the logic of the following substitution. This is an example of how one can view a problem in a new way and make it easier. We should all love symmetry like this!

Example 4.5.2: Changing Variables in Harmonic Oscillator

Consider the harmonic oscillator equation given by

$$y'' = -\frac{k}{m}(y(t) - L).$$

This equation, being second order, is immediately more difficult to solve. What we can do, however, is make a change of variables to

$$x(t) = y(t) - L$$

and note that

$$x''(t) = y''(t)$$

but the ODE changes to

$$x''(t) = \frac{k}{m}x(t).$$

This is much easier to solve. The idea of changing variables is extremely helpful. In these new variables, we can see the following figure.

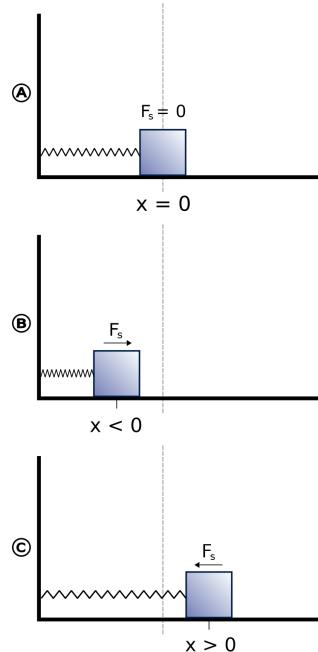


Figure 4.2: Spring-Mass system in the new variables. The dashed line represents the rest length L .

Interestingly enough, the rest length L does not enter the new equation now. As it turns out, it is essentially a useless parameter for understanding the problem.

However, an engineer would care about changing back to our original variable y as the length of a spring factors into design!

How could you know to make this change? Two ways. First, one can observe how a spring mass system oscillates. Set up the experiment and watch for yourself, if you'd like. What you'll see is that the mass oscillates in a symmetric way about the rest length of the spring. The experimentalist mindset will then suggest that you make the measurement about this position! The other way to observe this is as a mathematician would. In our original expression you can notice that we have $y(t) - L$ as a quantity and that

$$\frac{d}{dt}(y(t) - L) = y'(t).$$

This may be less intuitive to see, but this tells us that we can safely exchange the quantity $y(t) - L$ for a new quantity $x(t)$.

Earlier I claimed that this simplified equation

$$x'' = -\frac{k}{m}x$$

is equivalent to

$$u'' = -u$$

by changing the units in which we measure the problem. Indeed, consider the change of variables

$$u(t) = x(\sqrt{km}).$$

Then

$$u'(t) = \sqrt{km}x(\sqrt{km}) \quad \text{and} \quad u''(t) = \frac{k}{m}x(\sqrt{km}).$$

Exercise 4.5.1. Using the substitution shown above, show that the equation

$$x'' = -\frac{k}{m}x$$

is equivalent to

$$u'' = -u.$$

4.6 Qualitative analysis for first order equations

We have found that we are able to solve a specific type of ordinary differential equation. Of course, this is a great option when we are able to do so. But, not every equation will be nice or even possible to solve by hand. Take for example, the equation

$$x' = \sin(tx).$$

This equation is not separable and we will not develop a method for solving this type of equation in the future. However, we can attempt to understand what a solution to this equation may look like even if we do not arrive at a specific function as a solution. This

idea falls into the category of ***qualitative analysis*** since we are not getting an exact numerical answer but rather a reasonable guess to a solution that seeks to be qualitatively accurate.

The first method we will develop here is to plot the ***slope field*** of the system. Specifically, at each point in the tx -plane we plot a small segment of a line with a slope given by the value of x' at that point (t, x) . This is perhaps most easily seen with an example with a problem we've already solved.

Example 4.6.1: Slope field for exponential growth

Let us consider the exponential growth equation given by

$$x' = x.$$

We know that the general solution to this equation is

$$x = Ae^x,$$

and the particular solution is found when we specify an initial condition. For sake of example, let $x(0) = 1$ so that $A = 1$ and hence our particular solution is

$$x(t) = e^t.$$

Now, we can plot the slope field by plotting small line segments with slopes equal to the value of x' at each point (t, x) in the tx -plane. This gives us

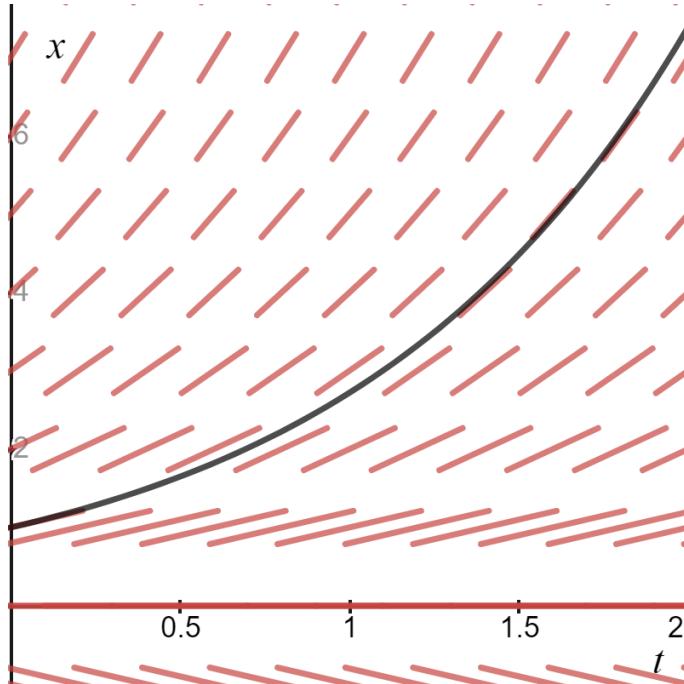


Figure 4.3: Slope field is plotted in red with the particular solution $x(t)$ plotted as the black curve.

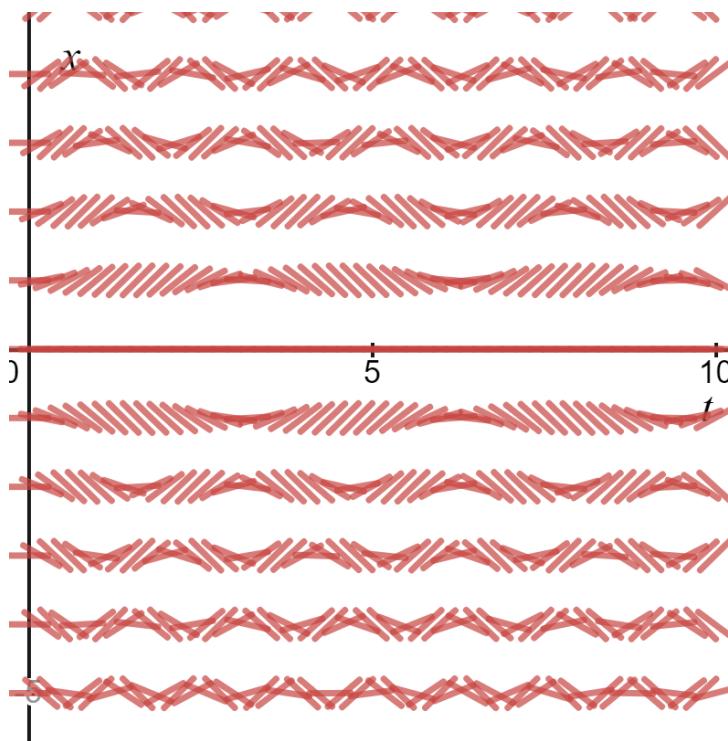
The take home idea is that the solution curve will have a slope that is equal to the slope field at each point (t, x) . In the above picture, you can see that the curve then

follows the slope field that we plotted. To do this yourself, you would pick an initial value which would specify where you start along the vertical axis, and you would draw a curve that follows the slope field.

To practice this yourself, let us return to the earlier example with the equation

$$x' = \sin(tx).$$

If we plot the slope field, we get the following figure.



Begin with the initial condition that $x(0) = 1$, and attempt to draw your best guess to the solution from this slope field.

We will revisit this idea later and in more detail when we talk about curves in higher dimensions. In a more general sense, what we are doing here is allowing our curve to follow a *vector field*. In other words, these solutions are truly the flow of a vector field. This is analogous to how a light particle is blown around in a wind – it flows along the direction of the wind current.

4.6.1 Autonomous equations

In physical systems that conserve some quantity in time (typically energy), we will see that the differential equations assume a certain form. This form is of an *autonomous* differential equation. In fact, this is another type of symmetry that can arise in a differential equation.

Definition 4.6.1: Autonomous ODE

A first order differential equation is ***autonomous*** if it can assume the form

$$x' = f(x).$$

A first order autonomous equation is also separable. So there is not much more here to say as far as finding solutions. However, we can also analyze the systems in a qualitative way. This is beneficial when the right hand side function $f(x)$ is something that we aren't able to integrate analytically.

When we are given a first order autonomous equation, we can make a plot of the system in the ***phase line***. That is, we can make a plot of $f(x)$ in the $x'x$ -plane. The reason why we do this is to seek out the following.

Definition 4.6.2: Equilibria

The ***equilibria*** of an autonomous differential equation is the set of points x where the derivative $x' = 0$. That is, these are the points where the system no longer changes over time.

Surprisingly, using the phase line will allow us to analyze the behavior of a function without having to arrive at a solution. Again, this is a type of qualitative analysis. We will seek to determine the equilibria of a system since these are points of interest. For systems that are evolving over time, knowing about the equilibria can tell us about the long time behavior of a system.

Take for example the exponential growth equation:

$$x' = x,$$

and note that if we have the initial condition $x(0) = 0$, then the particular solution to this initial value problem is $x(t) = 0$. So, the system stays at the same point for all possible times. There is no evolution happening!

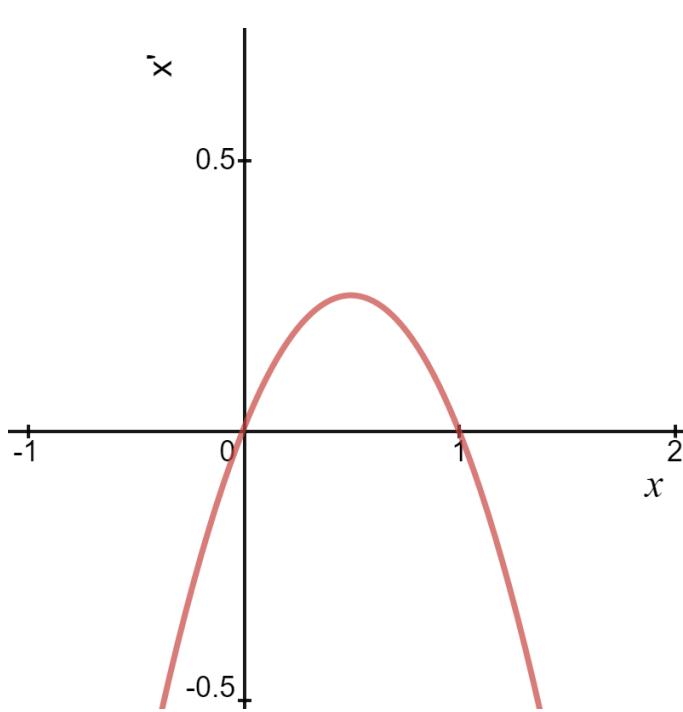
Now, how could we have deduced this qualitatively? Well, if we plot the graph of the function $f(x) = x$ in the $x'x$ -plane, we notice that $f(x) = 0$ when $x = 0$. It is exactly when $f(x) = 0$ that we have $x' = 0$ and thus there can no longer be motion for the system! Let's take a look at a more complicated example.

Example 4.6.2: Phase line for an autonomous equation

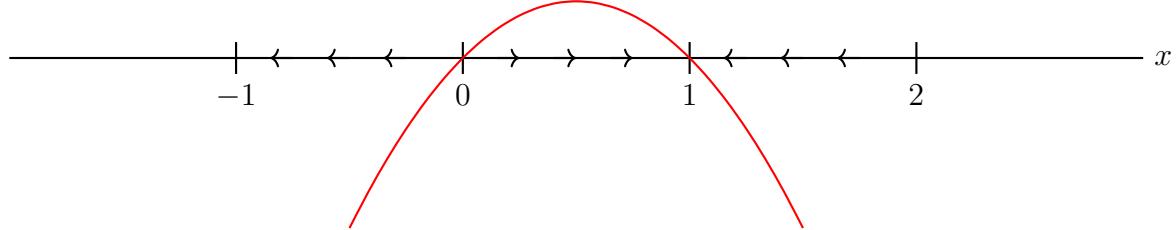
Let's consider the equation

$$x' = -\left(x - \frac{1}{2}\right)^2 + \frac{1}{4}.$$

This equation is indeed autonomous. However, finding a general solution analytically is rather difficult (but it is possible). Rather, what we can do is plot the graph of $f(x) = -\left(x - \frac{1}{2}\right)^2 + \frac{1}{4}$ in the $x'x$ -plane as follows.



What we can do next is mark along the x axis where the function is zero as these are the equilibria of the system. For this system, the set of equilibria is $x = 0$ and $x = -1$. Now, how do we know what happens over time? If we pick any value of x , we can know the value for x' since we can plug in the value of x into the autonomous equation. This leads us to the phase line below.



Here, I plot $f(x)$ again and we take note of the equilibria at $x = 0$ and $x = 1$. But, we can also see where $f(x) > 0$. For this f , the x values in the interval $(0, 1)$ lead to $f(x) > 0$. So, we draw arrows pointing in the positive x direction. Similarly, $f(x) < 0$ when $x < 0$ and $x > 1$, so we draw arrows pointing to the negative x direction in that case.

These arrows tell us which direction our solution will move over time. For example, if I take $x(0) = \frac{1}{2}$, then as I let $t \rightarrow \infty$, $x(t) \rightarrow 1$ since the arrows point towards the equilibrium at $x = 1$. If I instead took $x(0) = \frac{3}{2}$ then the arrows will dictate that our solution $x(t)$ approaches the equilibrium $x = 1$ as $t \rightarrow \infty$ as well. In this case, we say that the equilibrium $x = 1$ is **stable**. The last region of interest could be studied by choosing the initial condition $x(0) = \frac{-1}{2}$. Notice here that as $t \rightarrow \infty$ we have $x(t) \rightarrow -\infty$. This leads us to conclude that the equilibrium $x = 0$ is **unstable** since any initial condition around this equilibrium moves *away* from this equilibrium.

The prototypical example to keep in mind is a rigid swinging pendulum. There, we can

note that a pendulum hanging straight down is a stable equilibrium state. Over time, if we start swinging a pendulum, the pendulum will ultimately come to rest at the bottom. We can remark that this is much like the $x = 1$ case in the example system. Likewise, the pendulum could also rest perfectly in an upright position, but any small movement would cause the pendulum to swing and move away from that equilibrium position. This means that the top resting position is unstable. See the following figure.

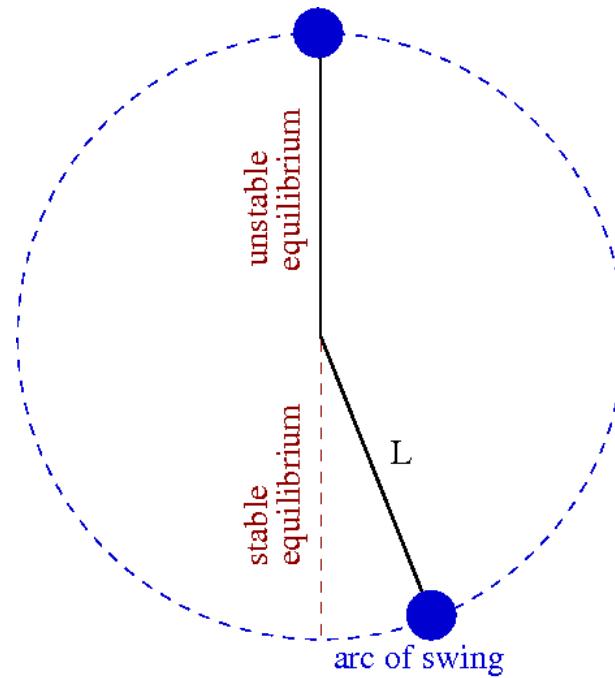


Figure 4.4: The two equilibria for a rigid pendulum.

This type of qualitative analysis is important. To get a feel for it yourself, you may want to consider analyzing the system

$$x' = \sin(x).$$

Exercise 4.6.1. Do the following exercises for the above system $x' = \sin(x)$.

- (a) Draw the phase line for this system.
- (b) Find all (infinitely many) equilibria.
- (c) Classify which equilibria are stable and which are unstable.

4.7 First Order Linear Differential Equations

Another nice set of equations that we can solve come in the form of linear equations. In particular, we first consider *first order linear ODEs* but will later consider the second order version as well. Generally, linear problems tend to be solvable whereas nonlinear problems are very hard. One can also learn how to approximate nonlinear problems via linearization, but we do not do that here.

Definition 4.7.1: First Order Linear ODE

A *first order linear ODE* is an equation that can assume the form

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

There is a general technique for solving this type of equation and, in fact, a general technique for solving higher order linear equations. More on that later. For now, let us work to solve this problem. The method we will use utilizes a function called an *integrating factor*. The idea is that we can use the format of the equation to our advantage.

Integrating Factor

Consider a first order linear ODE given by

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

Then, multiply the whole expression by a yet undetermined function $\mu(t)$ to get

$$\mu(t)x' + \mu(t)f(t)x = \mu(t)g(t). \quad (4.7.1)$$

The reason we have done this is that we can now take a look at the derivative of the product

$$(\mu(t)x)' = \mu'(t)x + \mu x'. \quad (4.7.2)$$

From here, we can set this product derivative (right hand side of 4.7.2) equals the left hand side of expression 4.7.1. This gives us

$$\mu'(t)x + \mu x' = \mu(t)x' + \mu(t)f(t)x$$

which means that

$$\mu'(t)x = \mu(t)f(t)x.$$

This is a separable ODE! So we can solve this for μ using the separation technique. This μ is the integrating factor.

Exercise 4.7.1. Solve the separable ODE

$$\mu'(t)x = \mu(t)f(t)x$$

for μ and show that you find

$$\mu(t) = e^{\int f(t)dt}.$$

From the exercise, we have that

$$\mu(t) = e^{\int f(t)dt}$$

and we can use this to complete the problem. Specifically, we have that the left hand side of 4.7.2 is equal to the right hand side of 4.7.1 to get

$$(\mu(t)x)' = \mu(t)g(t).$$

We can integrate both sides and solve for x to find

$$x = \frac{1}{\mu(t)} \int \mu(t)g(t)dt.$$

The above expression for x tells us the general solution to any first order linear ODE.

Example 4.7.1: Solving an ODE with Integrating Factor

Consider the first order equation

$$x' + \frac{2x}{t} = 2\cos(t).$$

Note that we can say $f(t) = \frac{2}{t}$ and then

$$\begin{aligned}\mu(t) &= e^{\int \frac{2}{t} dt} \\ &= e^{2\ln(t)} \\ &= t^2.\end{aligned}$$

Note, when computing the integrating factor μ we do not need to have a $+c$ after integrating. It will cancel later on if you do include it. Next, we note that

$$x = \frac{1}{\mu(t)} \int \mu(t)g(t)dt$$

where $g(t) = 2\cos(t)$ in this case. Thus

$$\begin{aligned}x &= \frac{1}{t^2} \int t^2 \cdot 2\cos(t)dt \\ &= \frac{2}{t^2} \left(t^2 \sin(t) + 2t \cos(t) - 2 \sin(t) + c \right),\end{aligned}$$

where the last equality involved using integration by parts twice. So we have found a general solution to our original ODE.

Exercise 4.7.2. Complete the integration by parts in the above example.

Just to introduce some terminology, we can note that if we have a first order linear ODE of the form

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

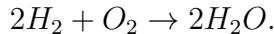
we call this equation **homogeneous**. Otherwise, we have that the expression with a nonzero right hand side

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

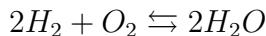
is called **inhomogeneous**. The homogeneous case for first order equations are simply separable equations and the distinction here is not really necessary.

4.8 Applications to Chemical Kinetics

As a chemist, one would probably like to understand how we can model elementary chemical reactions with mathematics. For example, maybe we would like to understand the rate of reaction of hydrogen H_2 and oxygen O_2 to create water H_2O . We typically write



Of course, we should actually be writing

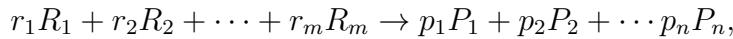


since real reactions have an equilibrium. The amount of back and forth in a reaction also depends on parameters like temperature, pressure, or concentration.

In thinking about these equations, one should consider how the rate of change of the reactants relates to the rate of change of the products. Likewise, the coefficients before a molecule describe how number of molecules that may be required to create a reaction or are produced from that reaction. For example, the creation of water requires two hydrogen molecules ($2H_2$) and just one oxygen molecule ($1O_2$). This means that we should expect to be using up twice as many molecules of hydrogen per second when we compare it to the rate of change of oxygen. Likewise, for every reaction that takes place, we produce two water molecules ($2H_2O$), and as such, we expect that the number of water molecules will be created at the same rate that we are using up hydrogen.

Question 4.8.1. Write down another chemical reaction of your choice. Before continuing on with this section, think about how the rates of the products and reactants will be related to each other. Can you see why we are studying this with differential equations?

For now, let us assume we are looking at the original model which can be written as a sum of m reactants R_i each with an amount r_i that give us an amount p_i of n different products P_j . That is, we have a reaction



that gives us an equation

$$\sum_{i=1}^m r_i R_i = \sum_{j=1}^n p_j P_j.$$

The second is just a mathematically succinct way of representing the quantities in the reaction. We will call the r_i and p_j the **stoichiometric variables**. These amounts r_i represent how many of the i^{th} reactants we need to create a single reaction and the p_j represent how many of the j^{th} products are produced given a single reaction has occurred.

We often write the equation above in the following form:

$$0 = \sum_{j=1}^n p_j P_j - \sum_{i=1}^m r_i R_i. \quad (4.8.1)$$

Later, we refer to this equation and it has been numbered so that we keep track of it as we derive an expression for the rate of change for each of the reactants and products.

Suppose we start with an initial amount of substance A , $N_A(0)$. Then we say that the number of molecules of species A at time t is $N_A(t)$. Note that N_A may represent either a

reactant or a product and we let a be the amount of species A necessary for the reaction to take place. That is, if A is a reactant, a is the number of molecules of species A required to form a reaction and if A is a product then a is the number of A produced by the reaction. In the case that A is a reactant, we will have $a < 0$ and if A is a product then $a > 0$. The amount of reaction observed is given by the *extent of reaction* ξ defined by

$$N_A(t) = N_A(0) + a\xi.$$

Here, ξ is some positive number that represents how much reaction has taken place. We will be using this variable later to relate different species involved in the reaction.

In 4.8.1, we can see that the total number of species has to be conserved. That is, our reaction is never losing any of the individual atoms in any way. Using this assumption, we have that the *rate of conversion* of reactants to products is

$$\begin{aligned}\rho &= \frac{d\xi}{dt} \\ &= \frac{1}{a} \frac{dN_a}{dt}.\end{aligned}$$

In other words, if we know how fast one species is changing, then we know how fast the others must be changing as well. Remember, the reaction is very strict. We always require the correct number of reactants to produce the given set of products. These rates must be related to one another.

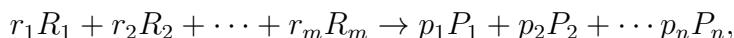
In the case that we care about concentrations of chemicals as opposed to the amount of molecules we let

$$[A] = \frac{N_A}{V}$$

where V is the volume the substance A is contained in. Then we have

$$v = \frac{\rho}{V} = \frac{1}{a} \frac{d[A]}{dt}.$$

Now, for a general reaction given by



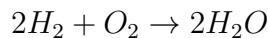
we must have that the concentrations of each species must change with matching rates. We can write this mathematically by

$$v = -\frac{1}{r_i} \frac{d[R_i]}{dt} = \frac{1}{p_j} \frac{d[P_j]}{dt},$$

for every product P_j and reactant R_i . Note that since reactants are utilized to create products, their rate of conversion is negative since their concentration must decrease. Similarly, products have a positive rate of conversion since they are being created by the reaction process.

Example 4.8.1: H_2 and O_2 to H_2O

If we are wishing to model



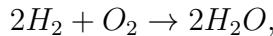
we will have the equations

$$v = -\frac{1}{2} \frac{d[H_2]}{dt} = -\frac{d[O_2]}{dt} = \frac{1}{2} \frac{d[H_2O]}{dt}.$$

At this point, we do not have enough data to create an equation that makes predictions for us. Indeed, we have the related rate expressions for each species, but we do not know what the rate of conversion v itself is. However, we can determine v from experimentation in a lab. There, we find that

$$v = k[R_1]^{r_1}[R_2]^{r_2} \dots$$

which allows us to complete our model. This expression here should make some sense. If we have the reaction



then we would have $v = k[H_2]^2[O_2]$ which is describing the likelihood of two hydrogen molecules coming into close contact with a single oxygen molecule. In this expression, we call k the **rate constant** of the reaction and note that we have $k > 0$. The numbers r_i determine that **order of the reaction**. We say that r_i is the order with respect to the reactant R_i . For example, we may have

- First order: $R \rightarrow$ Products;
- Second order: $2R \rightarrow$ Products;
- Second order: $R_1 + R_2 \rightarrow$ Products.

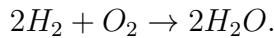
Thus, one can realize the order of the reaction by the expression

$$\text{Order} = \sum_{i=1}^m r_i.$$

That is, the order of the reaction is the total number of molecules of reactants required to create a single reaction.

Example 4.8.2: Creation of water

Consider the chemical reaction



This is a third order reaction since we require three reactant molecules to create products. The reaction rate is then

$$v = k[H_2]^2[O_2].$$

To find equations for rate of change the concentrations, we use the other expression

for v found in Example 4.8.1. Thus we have the equations

$$\begin{aligned}\frac{d[H_2]}{dt} &= -2k[H_2]^2[O_2] \\ \frac{d[O_2]}{dt} &= -k[H_2]^2[O_2] \\ \frac{d[H_2O]}{dt} &= +2k[H_2]^2[O_2].\end{aligned}$$

In the expression above, we would need to find an numerical value for the rate constant k . Once we have that, we then need to handle the fact that the concentrations are *coupled*. That is, for example, the rate of change of $[H_2]$ depends on the current amount of $[O_2]$. We will revisit this concept later. Lastly, we would need to know the initial concentrations of each species and we could determine a particular solution from there!

This formulation of chemical kinetics is powerful. It will give us a method to write down the rate of reaction for even the most complicated of chemical reactions. Whether we can solve these by hand is a different question, but nonetheless we can create a differential equation just from the structure of the reaction! The important constant in the expression is the rate constant k . This is a quantity that a chemist would measure in a lab by seeing how long a reaction takes to complete. Once this is determined for multiple reactions, we can then use this information to investigate more complicated kinetics.

The rest of this section is devoted to working through some example problems. Along the way, you may want to find reactions that fit the description provided. Look up the rate constant for those reactions and see if you can then take the general problems and apply them to your specific reaction. If done properly, this would be able to tell you how quickly you would expect the reaction to happen.

Example 4.8.3: First Order Reaction

Consider a reaction



with an initial concentration of R given by $[R]_0$. We then have

$$v = \frac{-d[R]}{dt} = k[R].$$

For ease of notation, let $x = [R]$ and we have

$$x' = -kx,$$

which we have solved before.

Exercise 4.8.1. Either solve the above differential equation again, or find the solution to the initial value problem somewhere in this text.

Example 4.8.4: Second Order Reaction

Consider a reaction



with an initial concentration $[R]_0$. Then we have

$$v = -\frac{1}{2} \frac{d[R]}{dt} = k[R]^2.$$

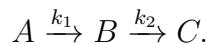
Again, let $x = [R]$ and we have the ODE

$$x' = -2kx^2.$$

Exercise 4.8.2. Find the particular solution to the initial value problem posed in the previous example.

Example 4.8.5: Several Reactions

Consider a chain of reactions as follows



Then we wish to describe the concentrations of each species A , B , and C over time. We know that we have

$$\frac{d[A]}{dt} = -k_1[A]$$

and it follows that

$$\frac{d[B]}{dt} = k_1[A] - k_2[B].$$

Then, at time $t = 0$ we have initial concentrations $[A]_0$, $[B]_0 = 0$, and $[C]_0$ so that we are only starting with species A . Later on, at time t , we have $[A] = [A]_0 - x$ and $[B] = y$ which means that we have $[C] = x - y$. Note that $[C]$ is created by $A \rightarrow B$ and from $B \rightarrow C$ so we must subtract off the concentration of B that has not converted to species C yet. This gives us the equations

$$\frac{d([A]_0 - x)}{dt} = -k_1([A]_0 - x) \quad (4.8.2)$$

$$\frac{dy}{dt} = k_1([A]_0 - x) - k_2y. \quad (4.8.3)$$

Note that 4.8.2 is a separable equation with solution

$$[A]_0 - x = [A]_0 e^{-k_1 t}.$$

We can then substitute in this solution to 4.8.3 to get

$$\frac{dy}{dt} = [A]_0 k_1 e^{-k_1 t} - k_2 y$$

which gives us the first order linear equation

$$y' + k_2 y = [A]_0 k_1 e^{-k_1 t}.$$

To solve this, we find the integrating factor

$$\mu(t) = e^{\int k_2 dt} = e^{k_2 t}.$$

Then we have that

$$y = \frac{1}{e^{k_2 t}} \int e^{k_2 t} [A]_0 k_1 e^{-k_1 t} dt = [A]_0 k_1 e^{-k_2 t} \int e^{(k_2 - k_1)t} dt.$$

This integral comes down to two cases; first when $k_1 = k_2$ and when $k_1 \neq k_2$. Thus we have

$$y = \begin{cases} \frac{[A]_0 k_1}{k_2 - k_1} e^{-k_1 t} + ce^{-k_2 t} & k_1 \neq k_2 \\ [A]_0 k_1 t e^{-k_1 t} + ce^{-k_2 t} & k_1 = k_2. \end{cases}$$

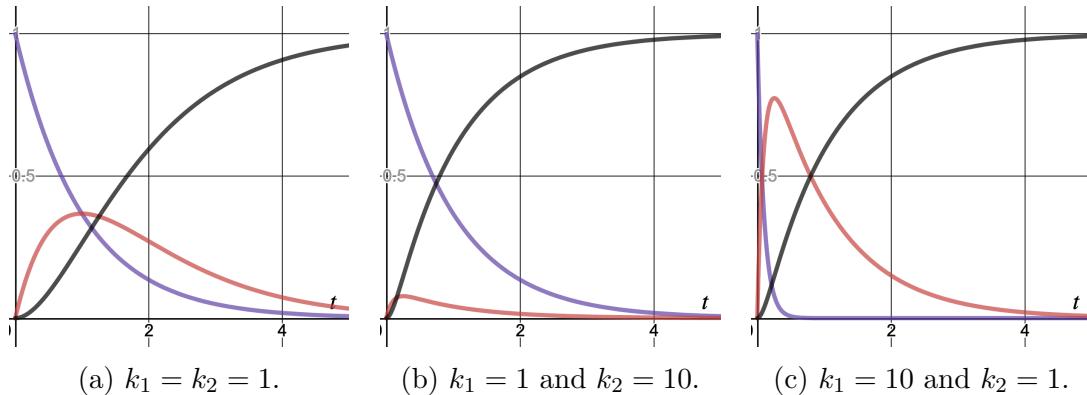
Using our initial conditions, we have $y = 0$ at time $t = 0$ which gives us the particular solution

$$y = \begin{cases} \frac{[A]_0 k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) & k_1 \neq k_2 \\ [A]_0 k_1 t e^{-k_2 t} & k_1 = k_2. \end{cases}$$

Now, we will have that each concentration can be written as

$$\begin{aligned} [A] &= [A]_0 e^{-k_1 t} \\ [B] &= \begin{cases} \frac{[A]_0 k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) & k_1 \neq k_2 \\ [A]_0 k_1 t e^{-k_2 t} & k_1 = k_2. \end{cases} \\ [C] &= [A]_0 - [A] - [B]. \end{aligned}$$

Now, we can plot the concentrations below letting $[A]$ be in purple, $[B]$ in red, and $[C]$ in black.



If instead, we wished to have arbitrary initial concentrations, we would let $[B]_0$ and $[C]_0$ be the initial concentrations for $[B]$ and $[C]$ respectively. This would give us

the equations

$$\begin{aligned}[A] &= [A]_0 e^{-k_1 t} \\ [B] &= \begin{cases} \frac{[A]_0 k_1}{k_2 - k_1} e^{-k_1 t} + \left(-\frac{[A]_0 k_1}{k_2 - k_1} + [B]_0 \right) e^{-k_2 t} & k_1 \neq k_2 \\ [A]_0 k_1 t e^{-k_1 t} + [B]_0 e^{-k_2 t} & k_1 = k_2. \end{cases} \\ [C] &= ([A]_0 - [A]) + ([B]_0 - [B]) + [C]_0.\end{aligned}$$

You should verify these that these new particular solutions are correct.

Exercise 4.8.3. Go through the previous example and fill in the missing steps.

4.9 Second Order Equations

Since Newtonian physics is governed by the equation

$$F = ma = mx''$$

we often see differential equations that are second order. The question is then, how can we solve these? Also, recall that in general, a second order equation can be written of the form

$$x'' = f(t, x, x').$$

We will immediately seek out special types of second order equations for which we can solve as we noticed we had to do this for the first order equations.

There are a two main forms of second order equations we will consider, but there are more out there. Let's define the class of equations we will stick with.

Definition 4.9.1: Second Order Linear ODEs

A second order ODE is *linear* if it can assume the form

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

We say that the equation is *homogeneous* if $h(t) = 0$ and otherwise it is *inhomogeneous*.

Remark 4.9.1. An inhomogenous second order linear ODE can be always be solved with the functions f , g , and h are smooth enough. We will not cover solving this general of a problem here.

4.9.1 Homogeneous and Constant Coefficients

Specifically, we care about homogeneous second order linear ODEs where the functions $f(t) = b$ and $g(t) = c$ are constant. These will be the most simple to solve yet fairly applicable. An equation like this can be rearranged to take the form

$$x'' + bx' + cx = 0.$$

A reasonable guess (or *ansatz*) for a solution is an exponential function of the form

$$x(t) = Ae^{\lambda t}$$

where A is an arbitrary constant. Since we can see that each successive derivative seems to only multiply our function by a constant, this is a good first guess. If we try to make this function x work to solve our ODE, we plug it in and see that we get

$$\begin{aligned} (Ae^{\lambda t})'' + b(Ae^{\lambda t})' + cAe^{\lambda t} &= 0 \\ \iff \lambda^2 Ae^{\lambda t} + \lambda bAe^{\lambda t} + cAe^{\lambda t} &= 0 \\ \iff Ae^{\lambda t} (\lambda^2 + b\lambda + c) &= 0 \\ \iff \lambda^2 + b\lambda + c &= 0. \end{aligned}$$

The last equality must be true since $Ae^{\lambda t}$ is never equal to zero. Thus, it seems we found a solution to the equation via the roots of this polynomial

$$\lambda^2 + b\lambda + c,$$

which we will refer to as the *characteristic polynomial*. The roots of the characteristic polynomial are then able to be found with the quadratic formula to yield

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

This leads us to the following.

Proposition 4.9.1: Solutions to Linear Constant Coefficient Second Order ODE

Consider the differential equation

$$x'' + bx' + cx = 0.$$

Then the corresponding characteristic polynomial is

$$\lambda^2 + b\lambda + c$$

and we let the solutions of this equation be equal to λ_1 and λ_2 . Then,

- *If $\lambda_1 \neq \lambda_2$, we have the general solution*

$$x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t},$$

where $C_1, C_2, \lambda_1, \lambda_2 \in \mathbb{C}$.

- *If $\lambda_1 = \lambda_2$, we have the general solution*

$$x(t) = C_1 e^{\lambda_1 t} + C_2 t e^{\lambda_1 t},$$

where $C_1, C_2, \lambda_1 \in \mathbb{C}$.

Question 4.9.1. Above we see that we have a sum of solutions and our ansatz only chose one. Is it always possible to have a sum of different solutions be a solution?

Answer 4.9.1. Yes. We will see by the following theorem that this is the case! This is an important result for studying quantum systems.

Theorem 4.9.1: Superposition of Solutions

Let x_1 and x_2 be general solutions to the equation

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

Then any **linear combination** (or **superposition**) of solutions

$$\alpha_1 x_1 + \alpha_2 x_2$$

with $\alpha_1, \alpha_2 \in \mathbb{C}$, is also a solution.

Proof. Let x_1 and x_2 be general solutions to the above ODE. Then consider the function

$$x = \alpha_1 x_1 + \alpha_2 x_2.$$

Then we have

$$\begin{aligned} x'' + f(t)x' + g(t)x &= (\alpha_1 x_1 + \alpha_2 x_2)'' + f(t)(\alpha_1 x_1 + \alpha_2 x_2)' + g(t)(\alpha_1 x_1 + \alpha_2 x_2) \\ &= \alpha_1 x_1'' + \alpha_2 x_2'' + \alpha_1 f(t)x_1' + \alpha_2 f(t)x_2' + \alpha_1 g(t)x_1 + \alpha_2 g(t)x_2 \\ &= \alpha_1(x_1'' + f(t)x_1' + g(t)x_1) + \alpha_2(x_2'' + f(t)x_2' + g(t)x_2) \\ &= 0, \quad \text{since } x_1 \text{ and } x_2 \text{ are solutions.} \end{aligned}$$

Hence $x = \alpha_1 x_1 + \alpha_2 x_2$ is also a solution. □

Remark 4.9.2. You may have heard of superposition in quantum mechanics. It turns out that this is exactly what is meant in the mathematical theory. Eventually, we'll see an example of what this physically means in a quantum system.

Example 4.9.1: Solving the Harmonic Oscillator

Consider the equation

$$mx'' = -kx$$

with the initial data $x(0) = 1$ and $x'(0) = 0$. We have shown that we have a solution to this equation before, but now we can explicitly solve it. We can rewrite the ODE as a second order linear homogeneous equation with constant coefficients by putting

$$x'' + \frac{k}{m}x = 0.$$

Then, for sake of notation, let $\omega = \sqrt{\frac{k}{m}}$ so that

$$x'' + \omega^2 x = 0.$$

Then the characteristic polynomial is

$$\lambda^2 + \omega^2$$

and the roots are

$$\begin{aligned}\lambda^2 + \omega^2 &= 0 \\ \lambda^2 &= -\omega^2 \\ \iff \lambda &= \pm\sqrt{-\omega^2} = \pm i\omega.\end{aligned}$$

Thus, the general solution is then

$$x = C_1 e^{i\omega t} + C_2 e^{-i\omega t}$$

where $C_1, C_2 \in \mathbb{C}$.

Now, we can find the particular solution from the initial data $x(0) = 1$ and $x'(0) = 0$ and letting

$$C_1 = a_1 + b_1 i \quad \text{and} \quad C_2 = a_2 + b_2 i.$$

We then have

$$\begin{aligned}1 = x(0) &= (a_1 + b_1 i)e^{i\omega \cdot 0} + (a_2 + b_2 i)e^{-i\omega \cdot 0} \\ &= (a_1 + b_1 i) + (a_2 + b_2 i) \\ &= (a_1 + a_2) + i(b_1 + b_2).\end{aligned}$$

Thus we must have

$$\begin{aligned}a_1 + a_2 &= 1 \\ b_1 + b_2 &= 0 \implies b_1 = -b_2.\end{aligned}$$

From the other initial condition, we have

$$\begin{aligned}0 = x'(0) &= \omega(a_1 + b_1 i)e^{i\omega \cdot 0} - \omega(a_2 + b_2 i)e^{-i\omega \cdot 0} \\ &= (a_1 - a_2) + i(b_1 - b_2).\end{aligned}$$

Thus we have

$$\begin{aligned}a_1 - a_2 &= 0 \implies a_1 = a_2 \\ b_1 - b_2 &= 0 \implies b_1 = b_2.\end{aligned}$$

Now we have $b_1 = -b_2$ and $b_1 = b_2$ which means $b_1 = b_2 = 0$. Then, we also have $a_1 = a_2$ which we can substitute into $a_1 + a_2 = 1$ to find that $a_1 = a_2 = 1/2$. Hence we have our particular solution

$$x(t) = \frac{1}{2}e^{i\omega t} + \frac{1}{2}e^{-i\omega t}.$$

This, we can rewrite to find a more familiar form of the solution by

$$\begin{aligned}x(t) &= \frac{1}{2}(\cos(\omega t) + i \sin(\omega t)) + \frac{1}{2}(\cos(-\omega t) + i \sin(-\omega t)) \\ &= \frac{1}{2}(\cos(\omega t) + i \sin(\omega t)) + \frac{1}{2}(\cos(\omega t) - i \sin(\omega t)) \\ &\qquad\qquad\qquad = \cos(\omega t).\end{aligned}$$

Note that I used the fact that sine is an odd function meaning that $\sin(-x) = -\sin(x)$ and that cosine is an even function meaning $\cos(-x) = \cos(x)$.

Exercise 4.9.1. Find where we claimed a solution to the harmonic oscillator earlier in this text and check that this solution matches that.

4.9.2 Qualitative Analysis

Solutions to homogeneous second order linear ODEs with constant coefficients only come in a few different classes of solutions. Essentially, they oscillate, exponentially grow or decay, or a combination of the two. There are some edge cases to be careful of, but they are not typical. Let's see why this is the case.

Recall that we have, in general, an equation

$$x'' + bx' + cx = 0$$

which gives rise to the characteristic polynomial

$$\lambda^2 + b\lambda + c.$$

The roots are then

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4c}}{2} = \frac{-b}{2} \pm \frac{\sqrt{b^2 - 4c}}{2}.$$

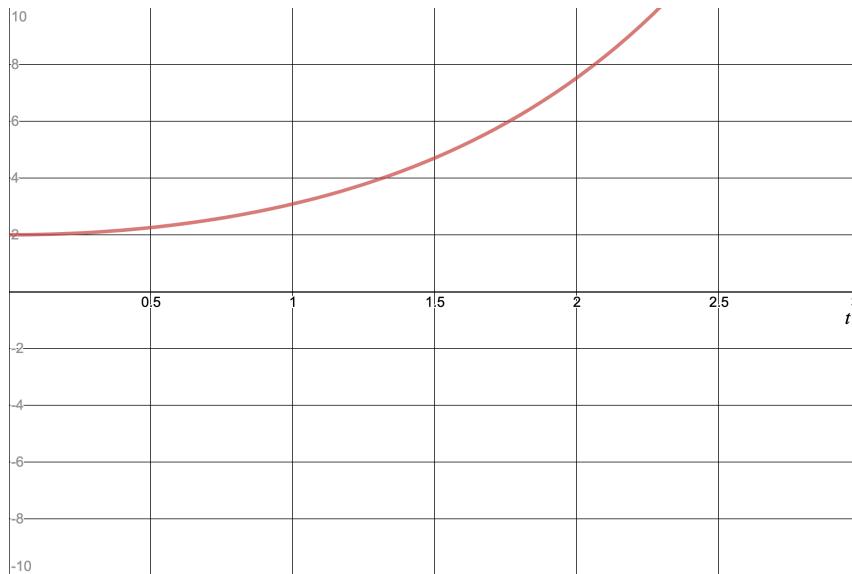
These roots can be real, complex, or purely imaginary. The real part can be greater than zero, or less. These facts encompass all the solutions we care about.

- **Case 1:** Consider the case where the roots of the characteristic polynomial are both real and denote the roots by λ_1 and λ_2 . If this is the case, then we must have that $b^2 > 4c$ so that the square root in the quadratic formula is not of a negative number. For this, we have two subcases.
 - **Subcase 1:** If we have two distinct real roots, λ_1 and λ_2 where $\lambda_1 \neq \lambda_2$, then our general solution to the equation is

$$x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}.$$

Note that λ_1 and λ_2 could be both positive, both negative, one negative one positive, or either could be zero as well.

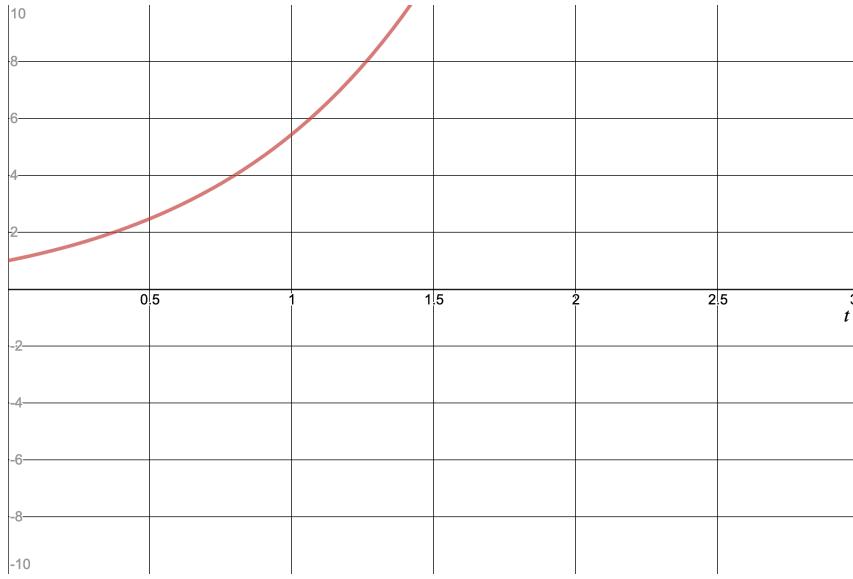
Let's say we have $\lambda_1 = 1$, $\lambda_2 = -1$ and let $C_1 = C_2 = 1$. Then the solution plotted in the xt -plane looks like



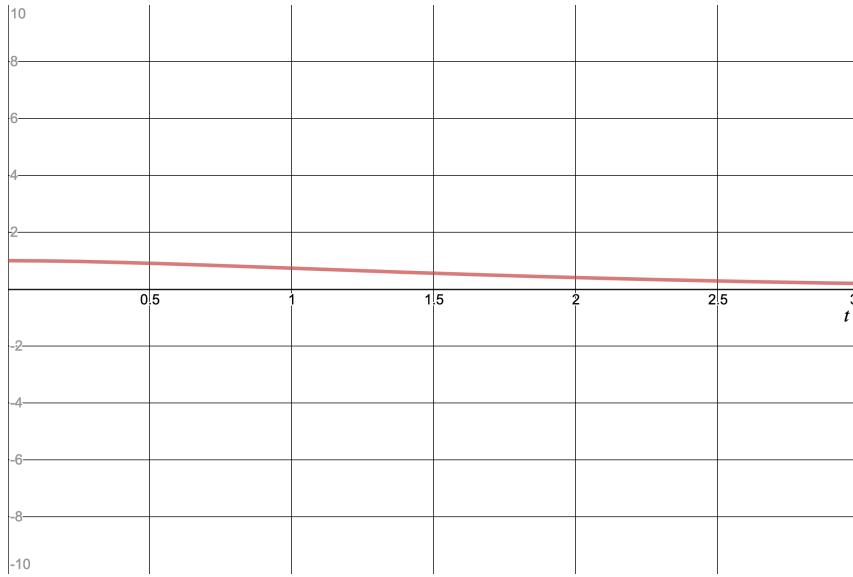
- **Subcase 2:** If we have two identical real roots, $\lambda = \lambda_1 = \lambda_2$ then the general solution is

$$x(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t}.$$

Here, we can take $\lambda = 1$ and let $C_1 = C_2 = 1$ and plot the solution:



We could also take $\lambda = -1$ with $C_1 = C_2 = 1$ to see:



- **Case 2:** Consider the case where the roots to the characteristic polynomial are complex valued. That happens when we have $b^2 - 4c < 0$ since then we will have a square root of a negative number appear with the quadratic formula. In this case, if we have a root λ , then λ^* is always the other root (take a look at the quadratic formula, and see if you can see why this is the case). Thus, we can put $\lambda = \alpha + \beta i$

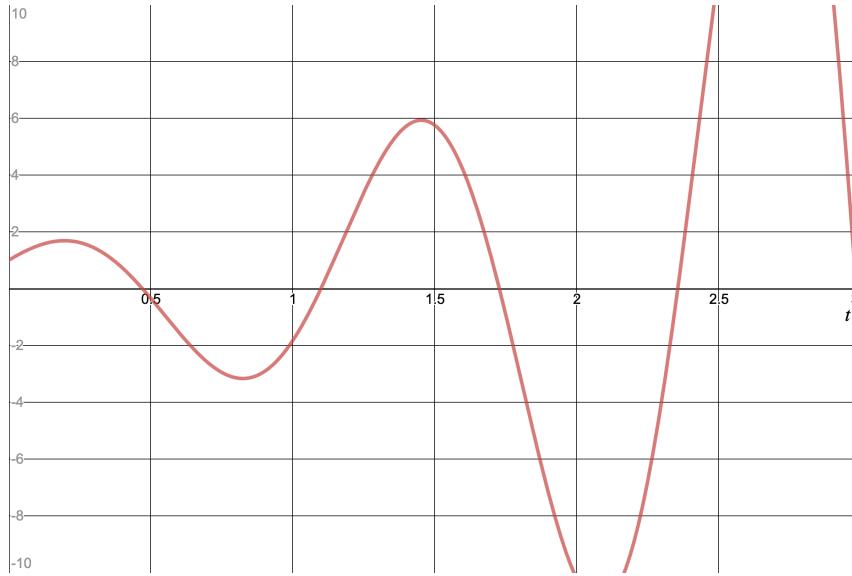
and then have another root $\lambda^* = \alpha - \beta i$. This gives us the general solution

$$\begin{aligned}x(t) &= C_1 e^{\lambda t} + C_2 e^{\lambda^* t} \\&= C_1 e^{\alpha t} e^{i\beta t} + C_2 e^{\alpha t} e^{-i\beta t} \\&= e^{\alpha t} (C_1 e^{i\beta t} + C_2 e^{-i\beta t}).\end{aligned}$$

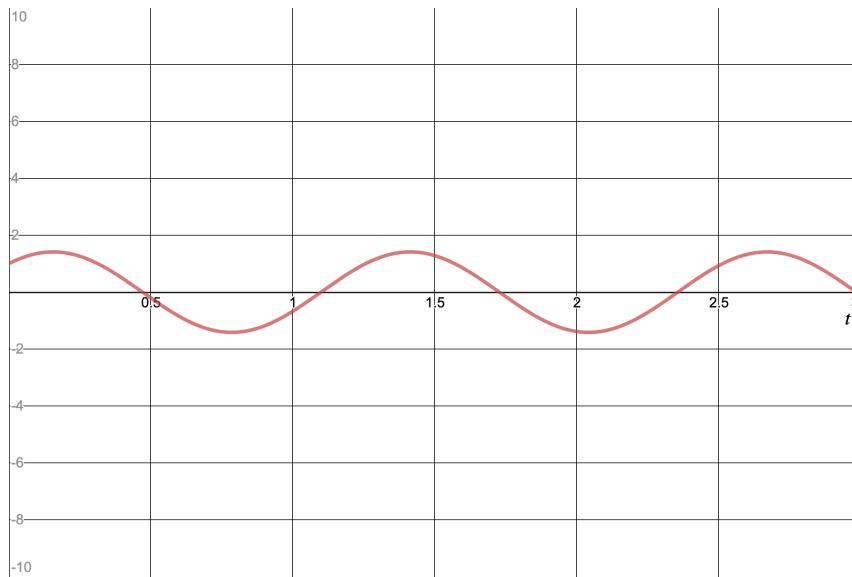
Here, we must have that C_1 and C_2 are complex numbers. We can also rewrite this general solution as

$$x(t) = e^{\alpha t} (C_1 \cos(\beta t) + C_2 \sin(\beta t)).$$

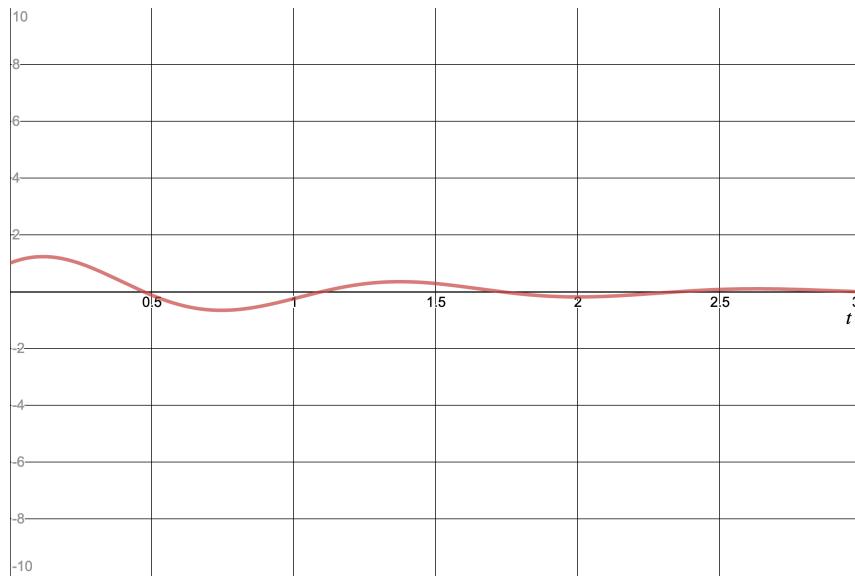
Here we can take $\alpha = 1$ and $\beta = 5$ with $C_1 = C_2 = 1$ and plot the solution in the xt -plane to see:



We could also take the case with $\alpha = 0$, $\beta = 5$, and $C_1 = C_2 = 1$ and plot:



Lastly, we could plot with $\alpha = -1$, $\beta = 5$ and $C_1 = C_2 = 1$ to see:



Exercise 4.9.2. Work out why if the roots to the characteristic polynomial are complex, that we always have the roots λ and λ^* .

Exercise 4.9.3. Show that the two general solutions

$$x(t) = e^{\alpha t}(C_1 e^{i\beta t} + C_2 e^{-i\beta t})$$

and

$$x(t) = e^{\alpha t}(C_1 \cos(\beta t) + C_2 \sin(\beta t))$$

are equivalent using Euler's formula.

4.9.3 Inhomogeneous Linear Equations

More often than not, we look at physical systems where there is some external force or potential that causes the system to change over time. In the case of second order linear equations, this corresponds to the inhomogeneous equation

$$x'' + bx' + cx = F(t)$$

where we think of $F(t)$ as an external force. For specific forcing terms $F(t)$, we can solve this equation exactly using the **method of undetermined coefficients**. The idea is that we can solve the homogeneous equation

$$x_h'' + bx_h' + cx_h = 0$$

to find a solution $x_h(t)$. This homogeneous solution can be understood as describing the dynamics of the system when undergoing no external forces. For example, if I have a mass on a spring, the system will oscillate even if I supply no external forces to it. We refer to this as the **homogeneous solution**. It does not, however, solve the equation on its own. We then need another function which we call the **particular integral** and denote by $x_p(t)$. This function is given by an ansatz based on what the forcing term is. This particular integral is essentially modeling the response that a system will have when it interacts with an external force. In this case, $x_p(t)$ solves the inhomogeneous equation

$$x_p'' + bx_p' + cx_p = F(t)$$

but is needs to be accompanied by a homogeneous solution. The solution to an inhomogeneous equation as above will then be

$$x = x_h + x_p.$$

This means that the dynamics for a system come down to a combination (or superposition) two different parts. First, the system itself has its own behavior whether acted on by a force or not. This is the x_h term. Then, if I act on the system with an external force, the system will have a reaction to that force and we see this reaction as the solution x_p . In total, the system will then behave in an additive way. The dynamics we see simply come from adding together the system behavior along with the system response.

How do we know that this $x = x_h + x_p$ is a solution? We can check by taking

$$\begin{aligned} x'' + bx' + cx &= (x_h + x_p)'' + b(x_h + x_p)' + c(x_h + x_p) \\ &= \underbrace{(x_h'' + bx_h' + cx_h)}_{=0} + \underbrace{(x_p'' + bx_p' + cx_p)}_{=F(t)} \\ &= F(t). \end{aligned}$$

So the sum of a particular integral and a homogeneous solution is also a solution to the inhomogeneous equation! Now, let's see an example.

Example 4.9.2: Inhomogeneous Linear Equation with Constant Force

If we assume our equation has a quadratic forcing term $F(t) = 2t^2$, then we have the equation

$$x'' + 3x' + 2x = 2t^2.$$

First, we find the solution to the homogeneous equation

$$x_h'' + 3x_h' + 2x_h = 0$$

which has the characteristic polynomial

$$\lambda^2 + 3\lambda + 2 = 0.$$

The roots are then $\lambda_1 = -1$ and $\lambda_2 = -2$. So we have that

$$x_h(t) = C_1 e^{-t} + C_2 e^{-2t}.$$

Next, we take an ansatz of $x_p(t) = a_0 + a_1 t + a_2 t^2$ and call the a_i coefficients the ***undetermined coefficients***. Then we have

$$x'_p = a_1 + 2a_2 t \quad \text{and} \quad x''_p = 2a_2.$$

We can plug these into our inhomogeneous equation

$$x'' + 3x' + 2x = 2a_2 + 3(a_1 + 2a_2 t) + 2(a_0 + a_1 t + a_2 t^2)$$

which we know should also be equal to the forcing term $F(t)$. This gives us the equation

$$2a_2 + 3(a_1 + 2a_2 t) + 2(a_0 + a_1 t + a_2 t^2) = 2t^2.$$

We rearrange the terms as follows

$$(2a_0 + 3a_1 + 2a_2) + (2a_1 + 6a_2)t + (2a_2)t^2 = 0 + 0t + 2t^2$$

which gives us the system of equations

$$\begin{aligned} 2a_0 + 3a_1 + 2a_2 &= 0 \\ 2a_1 + 6a_2 &= 0 \\ 2a_2 &= 2. \end{aligned}$$

We can solve these to find $a_2 = 1$, $a_1 = -3$ and $a_0 = \frac{7}{2}$. So the particular integral is $x_p(t) = t^2 - 3t + \frac{7}{2}$. Thus we have a solution

$$x = x_h + x_p = C_1 e^{-t} + C_2 e^{-2t} + t^2 - 3t + \frac{7}{2}.$$

If we were given initial data, we could solve for C_1 and C_2 .

Here's a table of the forms of $F(t)$ which will be solvable. You can use this to find an ansatz for the particular integral. Once you know the ansatz it comes down to solving for the coefficients as we did in the previous example.

Forcing Term $F(t)$	Particular Integral x_p
ke^{at}	Ce^{at}
$kt^n \quad (n = 0, 1, 2, \dots)$	$\sum_{j=0}^n a_j t^j$
$k \cos(at) \text{ or } k \sin(at)$	$K \cos(at) + M \sin(at)$
$ke^{at} \cos(bt) \text{ or } ke^{at} \sin(bt)$	$e^{at}(K \cos(bt) + M \sin(bt))$
$\left(\sum_{j=0}^n k_j t^j \right) \cos(bt) \text{ or } \left(\sum_{j=0}^n k_j t^j \right) \sin(bt)$	$\sum_{j=0}^n (Q_j t^j \cos(bt) + R_j t^j \sin(bt))$
$\left(\sum_{j=0}^n k_j t^j \right) e^{at} \cos(bt) \text{ or } \left(\sum_{j=0}^n k_j t^j \right) e^{at} \sin(bt)$	$e^{at} \left(\sum_{j=0}^n (Q_j t^j \cos(bt) + R_j t^j \sin(bt)) \right)$

Problems

Problem 4.1. Complete the exercises throughout the chapter.

Problem 4.2. What is an (ordinary) differential equation? Explain what it means to be a general and particular solution to a differential equation.

Problem 4.3. Look up an ordinary differential equation in chemistry that interests you. Write it down, and explain what it attempts to model. Why does it interest you?

Problems 4.4-4.9 are related.

Problem 4.4. (Newton's law of cooling) Write down a differential equation that models the following scenario:

The temperature of a substance in an ambient environment changes temperature over time proportionally to the difference of the temperature of the substance from the temperature of the ambient environment. Assume that the ambient environment is large enough to maintain a constant temperature.

Let $T(t)$ be the temperature of the substance, T_a be the ambient temperature, and k be the constant of proportionality.

Problem 4.5. With the equation found above, find a general solution.

Problem 4.6. With the parameter values $T_a = 100$, $k = 1$, and initial data $T(0) = 50$, find the particular solution. Find as well the particular solution when $T(0) = 55$ and when $T(0) = 150$. Plot each of the particular solutions and explain the results.

Problem 4.7. What happens instead if the initial temperature is equal to the ambient temperature? That is, when $T(0) = T_a = 100$. Does your solution reflect this? Does this make physical sense? Explain.

Problem 4.8. Let $\delta = (T_a - T)$. Show that the equation you found in Problem 1 reduces to

$$\delta' = -k\delta.$$

What is this equation describing physically? Explain.

Problem 4.9. The equation you arrived at earlier, $\delta'(t) = -k\delta(t)$, is *autonomous*. In this particular instance, it means that this problem has a derivative that is independent of time t . In fact, for this system, this essentially means that total energy is conserved!

- Using the slope field generator found at: <https://www.desmos.com/calculator/p7vd3cdmei>, plot the slope field in the $t\delta$ -plane and explore what happens as you vary k . What happens when $k = 0$? How about $k > 0$? How about $k < 0$?
- In this slope field plot, explain the symmetry. Can you see why this shows that we can always choose the initial time to be $t_0 = 0$ regardless of the value of the initial condition?
- k represents the conductivity of the object. Explain what the the solutions for $k < 0$, $k = 0$, and $k > 0$ mean physically. Should we think of objects with $k \leq 0$?

- (d) If instead we had an equation $y' = -kty$, can you see why we can no longer simply choose $t_0 = 0$ as the initial time?

Problems 4.10-4.12 are related.

Problem 4.10. (3 pts.) (Hooke's law) Write down an initial value problem based on the following statement.

"The rate of change of the rate of change of position of a mass is proportional to the position but in the opposite direction."

Problem 4.11. Show that $x = c_1 \sin(t) + c_2 \cos(t)$ is a general solution to the equation

$$x'' + x = 0.$$

Problem 4.12. Find the particular solution if $x(0) = 1$ and $x'(0) = 0$. Plot your solution in the t, x -plane and in the x, x' -plane.

Problem 4.13. Consider the differential equation

$$x' = \frac{x^2 + tx + t^2}{tx}.$$

- (a) Let $f(x, t) = \frac{x^2 + tx + t^2}{tx}$. Show that $f(\lambda x, \lambda t) = f(x, t)$.
- (b) Use the substitution $u = \frac{x}{t}$ in order to make the original equation separable.
- (c) Find the general solution to this separable equation in terms of u and t . You may use Wolfram Alpha to compute the necessary integral.
- (d) Find the solution to the original equation using the substitution $u = \frac{x}{t}$ and your solution from (c).

Problem 4.14. (Characteristic polynomial) Consider the following (differential) equation

$$ax''(t) + bx'(t) + cx(t) = 0.$$

This equation can be converted to a quadratic equation

$$a\lambda^2 + b\lambda + c = 0.$$

What are the roots to this equation?

Problem 4.15. (Why i ?) Consider the following (differential) equation

$$x''(t) = -x(t).$$

Now, let $x(t) = e^{it}$. Show that the above expression is true. (*Note: this is the same equation as in Problem 4.10, can you see why this must be true using Euler's formula?*)

Problem 4.16. Objects near Earth fall due to the force of gravity. The acceleration of an object due to gravity (regardless of mass) is then

$$y'' = g,$$

where $y(t)$ represents the height above the ground at time t and $g \approx -9.8 \frac{m}{s^2}$ is constant.

- (a) Find the general solution to the equation.
- (b) Given the initial data $y(0) = 0$ and $y'(0) = 1$, find the particular solution.
- (c) At what time $t > 0$ does the object first contact the ground?
- (d) Plot your solution only over the range of time that makes physical sense.

Problem 4.17. Consider the autonomous (Hamilton) equation

$$x' = ix$$

with initial condition $x(0) = 1$. *Hint: Think about what this is saying – the velocity of the particle x' is a rotation by $\pi/2$ of the position x .*

- (a) Find the particular solution to this equation.
- (b) Plot the function $x(t)$ in the complex plane.
- (c) Show that your solution $x(t)$ solves the ODE

$$x'' = -x$$

which is seen in Problems 4.10 and 4.14. *It is worth thinking about this too – the acceleration of a particle x'' is in the opposite direction of the position x , i.e., centripetal motion.*

Problem 4.18. Consider the following autonomous equation.

$$x' = -x^2.$$

- (a) Draw the phase line for this system. What are the equilibrium point(s)? Which equilibria are stable? Which are unstable? Explain.
- (b) Find a general solution to the ODE.
- (c) Explain how your general solution fits the qualitative behavior expected from the phase line. That is, can you show that limits of your general solution match your qualitative analysis?
- (d) Can $x(0) = 0$ be an initial condition? Explain. *Hint: your analysis from the phase line may prove to be more useful than the general solution you found.*

Problem 4.19. Let y be a function of x and consider the following differential equation.

$$y' = y \cos(x).$$

- (a) What is the order of this equation? Is the equation separable? Explain.
- (b) Plot an approximation of the slope field for this equation using this Desmos link: <https://www.desmos.com/calculator/e93gktwtfo>. Note that you will have to modify the $g(x, y)$ equation in that page.
- (c) Find the general solution to this equation.

- (d) Given the initial data $y(0) = 1$, find the particular solution.
- (e) Plot this function over your slope field. Explain how you could have approximated this solution using just the slope field.
- (f) Explain in words what the solution describes if we let $y(x)$ be the position of some object and x represents time.

Problem 4.20. Consider the differential equation

$$x' = \frac{x+t}{t}.$$

- (a) Let $f(x, t) = \frac{x+t}{t}$. Show that $f(x, t) = f(\lambda x, \lambda t)$.
- (b) Given (a) holds, use the change of variables $u = \frac{x}{t}$ to rewrite the differential equation as a separable equation in terms of u .
- (c) Find the general solution to the equation and write your solution in terms of the original variables t and x .

Problem 4.21. Find the general solution to the following equation.

$$tx' + 2x = \frac{\sin(t)}{t}.$$

Show that your solution is correct. (*Hint: can you use an integrating factor?*)

Problem 4.22. Write down the equations for each of the reactants and products for the following reactions.

- (a) $A + 3B + C \xrightarrow{k} 2D + 2E$.
- (b) $A \xrightarrow{k_1} B + C \xrightarrow{k_2} D$.

Problem 4.23. Consider the following reaction



For the following parts, use the link: <https://www.desmos.com/calculator/srrpeadlou>.

- (a) Compare and contrast the reactions that take place given the three different scenarios for initial conditions. Explain why what the graph displays makes sense and include your graphs.
- $[A]_0 = 1$, $[B]_0 = 0$, and $[C]_0 = 0$.
 - $[A]_0 = 0$, $[B]_0 = 1$, and $[C]_0 = 0$.
 - $[A]_0 = 0$, $[B]_0 = 0$, and $[C]_0 = 1$.
- (b) For the initial conditions $[A]_0 = 1$, $[B]_0 = 0$, and $[C]_0 = 0$, explain what happens when you let
- $k_1 = 0$ and $k_2 = 1$,

- $k_1 = 1$ and $k_2 = 0$.

Include plots for these cases as well.

- (c) Consider the initial conditions $[A]_0 = 1$, $[B]_0 = 0$, and $[C]_0 = 0$ and rate constants $k_1 = 1$ and $k_2 = 2$. Then, choose initial conditions of your own and compare your plots with the other initial conditions. Why do yours behave the way they do? Include your plots.

Problem 4.24. Consider the second order chemical reaction given by



- Write a *system* of differential equations to describe the concentration of the reactants A and B (this means write one for each).
- The concentrations of A and B can be related to each other in the following way: Let $A = A_0 - x$ and $B = B_0 - x$. Here, we think of x as the amount of each chemical that has reacted, and note that it depends on time t . Use this change of variables to rewrite the differential equation for chemical A in terms of x and t .
- Solve the differential equation in (b) with the initial condition $x(0) = 0$. You will need to use *partial fraction decomposition* to evaluate the integral.

Problem 4.25. If $x_1(t)$ and $x_2(t)$ are solutions to the differential equation

$$x'' + bx' + cx = 0$$

is $x = x_1 + x_2 + k$ for a constant k always a solution? Is the function $y = tx_1$ a solution? Explain.

Problem 4.26. Consider the following initial value problem:

$$x'' + 4x' + 3x = 0$$

with initial data $x(0) = 1$, $x'(0) = 0$.

- Find the solution.
- Sketch a plot of the solution.
- Explain in words what is happening to the solution as time goes on. What happens as $t \rightarrow \infty$?

Problem 4.27. Write down a homogeneous second-order linear differential equation where the system displays a decaying oscillation.

Problem 4.28. Consider the following differential equation:

$$x'' + 2x' + x = \sin(t)$$

- Find the homogeneous solution $x_H(t)$.
- Find the particular integral $x_P(t)$.

- (c) Find the specific solution corresponding to the initial data $x(0) = 0, x'(0) = 0$.
- (d) Plot the curve $(x(t), x'(t))$ in the plane (we often call this *phase space*). Use this link here: <https://www.desmos.com/calculator/ouqwcxj2xz>. Use the time range $t \in [0, 10\pi]$.
- (e) Describe what happens with this system over time. Does it seem to approach some kind of stable solution? Note that this stable solution could be periodic.

Problem 4.29. Let's revisit Newton's law of cooling and describe the equilibria of the system.

- (a) Is the system autonomous? Explain.
- (b) Draw a phase line for the system.
- (c) Can you explain why the value you find for the equilibrium makes sense? Is this equilibrium be stable? Does this make sense? Explain.

Problem 4.30. Are the following ODEs separable, autonomous, linear, nonlinear, or none of the above? Keep in mind that some may satisfy more than one property!

- (a) $x' = \sin(tx)$.
- (b) $x' = \sin(x)$.
- (c) $x' = t^2x$.
- (d) $e^t x' + tx = e^{-kt} \cos(\omega t)$.

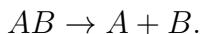
Problem 4.31. For the following problems, show that the equation is linear by writing the equation in a recognizable form.

- (a) $\frac{x'}{x} = t$.
- (b) $2xx' + x^2 = xt$.
- (c) $\tan(t)x' + \sin(t)x = \ln(t)$.

Problem 4.32. For the above linear equations, determine the integrating factor (even if you cannot compute the integral) and determine the solution $x(t)$ (again, even if you cannot compute the integral).

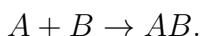
Problem 4.33. Find the solution to the equation $x' + x = t^2$.

Problem 4.34. Consider the dissociation chemical reaction



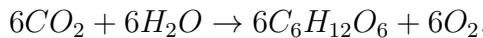
Write down an equation for each species A , B , and AB . Find a solution.

Problem 4.35. Compare and contrast the equations for the above reaction and the synthesis reaction



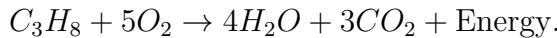
Hint: maybe this is a bit of an odd way to think, but is one just the reversal in time of the other?

Problem 4.36. Photosynthesis is an extremely important chemical reaction where plants convert carbon dioxide and water into glucose and oxygen. That is,



- (a) Write down the equations that describe the above equation.
- (b) Do we have any techniques to solve this type of equation (yet)?
- (c) The reaction also depends on the intensity of light. How can we implement this into the equation?

Problem 4.37. Everybody loves combustion. It keeps us warm and it looks cool on the 4th of July! Anyhow, combustion of propane in your grill at home is given by the equation



- (a) Write down the equations describing the above reaction (not including the energy term).
- (b) If a specific amount of energy is given off by each reaction (in this case, 2043kJ of energy per mole), include an equation for the total energy created if we begin with 1kg of propane in an abundance of oxygen.
- (c) Can you give some analogy to how much energy this is so that we may understand the usefulness a bit more? E.g., how many hot dogs could I cook on the 4th of July?

Problem 4.38. Write down the characteristic polynomial for the following equations. Then, find the roots to the characteristic polynomial and write down the general solution.

- (a) $x'' + x' + x = 0$.
- (b) $x'' - x' - x = 0$.
- (c) $x'' - x' + x = 0$.
- (d) $x'' + x' - x = 0$.

Problem 4.39. For the above solutions, analyze their behavior qualitatively. That is, do the solutions oscillate, grow, decay, or some combination of these, or something else entirely?

Problem 4.40. Consider the equation

$$x'' + bx' + cx = 0.$$

The roots to the characteristic polynomial are then

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

- (a) Explain why if $c > 0$ and $b = 0$ the solution $x(t)$ will be purely oscillatory.
- (b) Explain why if $b > 0$ and $b^2 < 4c$, the solution will oscillate and decay.

(c) Explain why if $b < 0$ and $b^2 < 4c$, the solution will oscillate and grow.

Problem 4.41. Write down a second order linear differential equation that oscillates and also decays over time.

Problem 4.42. Consider the following differential equation

$$x'' + x = 0.$$

- (a) Find the general solution to this equation.
- (b) Given the initial conditions $x(0) = 1$ and $x'(0) = 1$, find the particular solution.
- (c) Plot your particular solution.
- (d) Does the solution grow or decay over time?
- (e) What is $\lim_{t \rightarrow \infty} x(t)$?

Problem 4.43. Next, consider a related equation

$$x'' + x = t.$$

that has an additional linear external force.

- (a) What is the solution to the homogenous equation?
- (b) Find the particular integral with the given forcing term.
- (c) What is the specific solution to this equation?
- (d) Does the solution grow or decay over time?
- (e) What is $\lim_{t \rightarrow \infty} x(t)$?

Problem 4.44. Consider now the equation

$$x'' + x = F(t)$$

where the external force is $F(t) = \cos(t)$.

- (a) Find the particular integral with the given forcing term.
- (b) What is the specific solution to this equation?
- (c) What is $\lim_{t \rightarrow \infty}$? What does this mean about the growth or decay of the solution over time?

Problem 4.45. Consider the endothermic breakdown of a molecule x given by



where we let $x(t)$ denote the concentration of reactants. Since the reaction is endothermic, if we also heat up the solution over time, we get a factor of t^2 as well since the reaction occurs more readily in higher temperatures. The concentration decreases over time based on differential equation

$$x' = -kt^2(x - x_e).$$

where x_e is a constant that denotes the equilibrium concentration.

- Write an equivalent equation with the change of variables $\delta = x - x_e$.
- Find the general solution to this new equation.
- What is the general solution in terms of the original variables x ?
- Given the initial amount of x is $x(0) = 1$, the equilibrium concentration is $x_e = 1/2$, and $k = 1$, find the particular solution for $x(t)$.
- Does this reaction ever reach the equilibrium state?

Problem 4.46. Let the height above ground at time t be given by the function $y(t)$. A ball falling through air experiences gravitational acceleration and damping due to air friction. It follows the differential equation

$$y'' = -ky' - g.$$

- Describe the type of this equation (e.g., separable, autonomous, linear, etc.). What is the order?
- Find the solution to the homogeneous equation. Call this solution x_H .
- Find the particular integral to the inhomogeneous equation. Call this x_P .
- What is the general solution to the ODE given in this problem?
- Suppose that $k = 1$ and $g = 10$. Let $y(0) = 1,000$ and $y'(0) = 0$. Find the particular solution to this problem.
- Plot your particular solution. Plot the derivative of your solution $y'(t)$ as well. If your solution is correct, the derivative (velocity) y' should approach a constant value called the *terminal velocity*.

Problem 4.47. The spring/mass harmonic oscillator is given by the equation

$$x'' + \frac{k}{m}x = 0.$$

where m is the mass of the oscillating object and k is the spring constant.

- Show that

$$x(t) = A \cos \left(\sqrt{\frac{k}{m}} t \right).$$

solves the initial problem with initial data $x(0) = A$ and $x'(0) = 0$.

- (b) Since this system has no damping, the total energy is conserved at all times. In particular, the total energy is given by

$$E = \frac{1}{2}mx'(t)^2 + \frac{1}{2}kx(t)^2 = \text{constant.}$$

and is equal for all times $t \geq 0$. What is the total energy of the given particular solution $x(t) = A \cos(\sqrt{\frac{k}{m}}t)$?

Problem 4.48. Consider the inhomogeneous linear differential equation

$$x'' + \omega^2 x = \cos(\omega t),$$

where $x(t)$ is a function of t . This is an example of resonance.

- (a) Find the general homogeneous solution $x_h(t)$.
- (b) Show that your solution solves the homogeneous equation

$$x_h'' + \omega^2 x_h = 0.$$

- (c) For the particular integral, $x_p(t)$, take the ansatz

$$x_p(t) = C_1 t \cos(\omega t) + C_2 t \sin(\omega t)$$

and find the undetermined coefficients C_1 and C_2 .

- (d) Show that $x = x_h + x_p$ solves the inhomogeneous equation

$$x'' + \omega^2 x = \cos(\omega t).$$



Boundary Value Problems

There is another kind of problem we care to work with. All the problems we have considered have been initial value problems. However, there is a completely different type of problem we can solve. This will be called a ***boundary value problem***. Let us make a comparison.

Second Order Initial Value Problem

- Given a very general second order equation $x'' = f(t, x, x')$ and initial data $x(0) = x_0$ and $x'(0) = v_0$.
- Both conditions are given at the initial time $t = 0$. We have one condition $x(0)$ for position and the other $x'(0)$ for velocity
- Think of the solution $x(t)$ of tracking a particle over time.

Second Order Boundary Value Problem

- Given a very general second order equation $u'' = f(x, u, u')$ and a *region* $\Omega = [a, b]$ with *boundary values* $u(a) = u_a$ and $u(b) = u_b$.
- Conditions are given at the endpoints of a region $\Omega = [a, b]$. There will be one condition for every endpoint a , and b . The values will be given as $u(a) = u_a$ and $u(b) = u_b$. (*Note: we may also prescribe slightly different conditions later on. These are referred to as Dirichlet boundary conditions.*)
- Think of solution $u(x)$ as modeling a measurable quantity of an object (i.e., temperature, height, stress, or strain).

Boundary value problems are very important in the physical world. For example, an engineer may wish to understand how a rod is deformed under the force of gravity as well as whatever machinery is pulling and pushing on it. The engineer would hope to understand this deformation so that they can build sturdy structures. In the field of chemistry, we see boundary value problems arise in thermal or quantum transport, in chemical reactions, in waves or lattice vibrations, and more.

Not all boundary value problems are second order. However, we'll concentrate first on problems that are second order as they are more pertinent for us. The equations we'll start with will be Laplace's (or Poisson's) equation and then we'll move onto Schrödinger's equation.

5.1 Laplace's and Poisson's Equations

Let us set up the problem at hand. We'll define our first boundary value problem as such.

Definition 5.1.1: Boundary Value Problem

A **boundary value problem** is a differential equation (of possibly multiple variables) defined on a region in space Ω with prescribed **boundary data** (i.e., functions defined on the boundary of the region Ω).

For now, all regions will be $\Omega = [a, b]$, the **boundary** is the set $\{a, b\}$, and the differential equations will be second order.

Remark 5.1.1. Here, I will use the notation

$$\frac{d}{dx}u = u'$$

since we often like to think of $\frac{d}{dx}$ as an *operator*. We will get more into this later. Right now, it's just equivalent notation.

Example 5.1.1: Laplace's Equation: Dirichlet Boundary Data

Imagine a rod of length L attached at endpoints $x = 0$ and $x = L$. The height of the endpoints of the rod are $u(0) = 1$ and $u(L) = 0$ respectively. The rod is experiencing no external forces. What is the shape of the rod?

This problem statement leads us to consider the following boundary value problem

$$\frac{d^2}{dx^2}u(x) = 0$$

on the region $\Omega = [0, L]$ with boundary data $u(0) = 1$ and $u(L) = 0$. Here, we are providing the function value that $u(x)$ must satisfy on the boundary. Everywhere inside Ω (i.e., $(0, L)$), $u(x)$ must satisfy the differential equation as well.

We know how to solve this equation as we can just integrate twice. We find that the general solution is

$$u(x) = C_1x + C_2.$$

We can solve for the constants C_1 and C_2 by using our boundary data. We require

$$\begin{aligned} 1 &= u(0) = C_1(0) + C_2 = C_2 \\ 0 &= u(L) = C_1(L) + C_2. \end{aligned}$$

The first equation gives us $C_2 = 1$ which we can plug into the second to get

$$0 = C_1(L) + C_2$$

which means that $C_1 = -\frac{1}{L}$. Thus our particular solution to this boundary value problem is

$$u(x) = -\frac{1}{L}x + 1.$$

Exercise 5.1.1. Plot the above solution and verify it satisfies the ODE and the boundary values. What can we say about the shape of the rod? Is it what we would expect?

So, the set up to these problems is a bit different, but solving them is fundamentally the same idea. Let's see some other examples of different equations.

Definition 5.1.2: Laplace and Poisson Equations

On a region $\Omega = [a, b]$ we prescribe boundary data $u(a) = u_a$ and $u(b) = u_b$. On this region, **Laplace's equation** is

$$\frac{d^2}{dx^2}u(x) = 0$$

and **Poisson's equation** is

$$-\frac{d^2}{dx^2}(x) = F(x)$$

for some given function $F(x)$. If this $F(x)$ is of the form $ku(x)$ for some constant k , then we have the equation for **eigenfunctions of the Laplacian**

$$\frac{d^2}{dx^2}(x) = ku(x).$$

We'll see that this equation shows up in quantum mechanics.

If we read the two examples and the definition of the Laplace and Poisson equations, we can get an intuition on what they're describing. In a sense, both equations can describe the curvature of a 1-dimensional object. Why is that? The second derivative gives us information about the curvature of a function at each point! For the Laplace equation, the curvature is said to be zero everywhere, hence when we integrated and found a solution, we found that we got a linear equation with some unknown coefficients that we could find with boundary data. For the Poisson equation, however, there is some external force

function $F(x)$ that deforms the rod. So, the curvature should not be zero everywhere, which gives us a different set of solutions. One may also notice that the solution to the Laplace equation appeared within the solution to the Poisson equation (go back and reread both to see this). This is very much like solving homogeneous versus inhomogeneous second order linear ODEs with constant coefficients.

Example 5.1.2: Poisson's Equation: Dirichlet Boundary Data

Consider a cable of length 1 attached at endpoints $x = 0$ and $x = 1$. Now, the attachment points of the cable is prescribed by specifying $u(0) = 0$ and $u(1) = 0$. The cable is also being pushed down by a constant force (like gravity). That is, $F(x) = -1$ constantly pulls the cable down. We expect this cable to sag under this force much like power lines due under the force of gravity.

This gives us the boundary value problem

$$-\frac{d^2}{dx^2}(x) = -1$$

on the region $\Omega = [0, 1]$ with boundary data $u(0) = 0$ and $u(1) = 0$. We can integrate this equation twice to get

$$u(x) = \frac{1}{2}x^2 + bx + c.$$

Now, we can use our boundary data to try and solve this problem. We require

$$0 = u(0) = c,$$

and hence $c = 0$. Thus, at this point we have $u(x) = \frac{1}{2}x^2 + bx$. The second equation from the second boundary value is then

$$u(1) = \frac{1}{2} + b,$$

and so $b = -\frac{1}{2}$. Hence our particular solution to the boundary value problem is

$$u(x) = \frac{1}{2}x^2 - \frac{1}{2}x.$$

We can plot this solution and see if this makes physical sense.

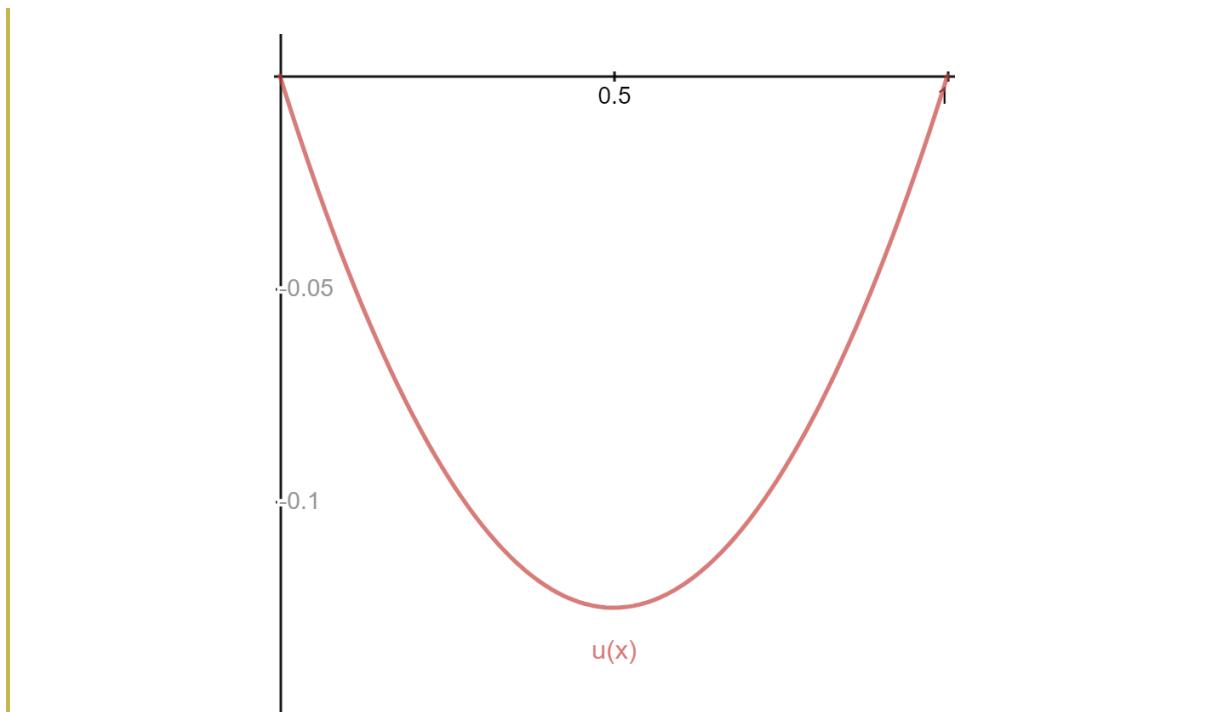


Figure 5.1: The profile of a cable under a constant force directed downward.

It seems like the curvature of this cable is close to what we would expect. How much deformation there is would depend on the strength of the force $F(x)$ as well as the elastic properties of the cable. These would all factor in to the equation if we considered it in greater detail.

5.2 A Particle in a 1-Dimensional Box

Now, let's enter into some quantum mechanics. In essence, quantum mechanics is governed by the *Schrödinger equation*. Think of this equation as a quantum version of Newton's laws. Right now, we have only considered ODEs, and as such we will consider a particular case of Schrödinger's equation which is independent of time. The ***time independent Schrödinger equation*** describes a ***stationary state*** of a particle which we denote by $\psi(x)$. If we put $\psi(x)$, we are thinking of this particle as not evolving in time, but having this function describing its position nonetheless. In general, we call a function $\Psi(x, t)$ that solves time dependent version of Schrödinger's equation a ***wavefunction***. Again, we will not approach the time dependent portion until we reach the sequel of this course. A wavefunction is not quite physically meaningful in its own right. However, it gives rise to a probability interpretation of the particle. We'll see that shortly.

Definition 5.2.1: Time Independent Schrödinger Equation

The *one dimensional time independent Schrödinger equation* with potential $V(x)$ is given by the differential equation

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

Here, Ψ is the wavefunction and in this case we will only find that the allowed wavefunctions can only be stationary states. \hbar is a physical constant called **Planck's reduced constant**, E is the energy, and m is the mass of whatever particle we are considering at the moment. Note that E is not fixed in this equation and is taken to be an unknown parameter (until we solve the equation). The quantity

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

is called the **Hamiltonian operator**.

If we think of a particle in a 1-dimensional box of length L , then the region the particle can exist is $[0, L]$. We're making it impossible for the particle to leave this box and we can do that using the potential $V(x)$. If we assign

$$V(x) = \begin{cases} 0 & \text{if } 0 < x < L \\ \infty & \text{if } x \leq 0 \text{ or } x \geq L. \end{cases}$$

This gives rise to the boundary value problem

$$-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} = E\Psi$$

with boundary conditions $\Psi(0) = 0$ and $\Psi(L) = 0$. This should look much like the eigenfunction equation for the Laplacian we saw previously. These boundary conditions will seem more sensible when we get a better interpretation of the wavefunction Ψ . If you'd like, you can essentially plug in $V(x) = \infty$ and see that the only possibility for a solution is that $\Psi(x) = 0$ (but this is a bit handwavy).

Now, let us solve this boundary value problem. We have a second order linear differential equation with constant coefficients. In fact, it is also homogeneous as we can write

$$\frac{d^2\Psi}{dx^2} + \frac{2mE}{\hbar^2}\Psi = 0.$$

Now, to solve this, find roots λ_1 and λ_2 to the characteristic polynomial

$$\lambda^2 + \frac{2mE}{\hbar^2} = 0.$$

We solve this by letting $\omega^2 = \frac{2mE}{\hbar^2}$ and putting

$$\begin{aligned} \lambda^2 &= -\omega^2 \\ \lambda &= \pm i\omega, \end{aligned}$$

so $\lambda_1 = i\omega$ and $\lambda_2 = \lambda_1^*$. This then gives us the general solution

$$\Psi(x) = C_1 e^{i\omega x} + C_2 e^{-i\omega x}.$$

Of course, it is also possible to write

$$\Psi(x) = C_1 \cos(\omega x) + C_2 \sin(\omega x),$$

as this is just an equivalent way to write out the general solution. Now, we have our boundary conditions $\Psi(0) = 0$ and $\Psi(L) = 0$ as well. Plugging these into our general solution gives us

$$\begin{aligned} 0 &= \Psi(0) = C_1 \cos(\omega \cdot 0) + C_2 \sin(\omega \cdot 0) \\ &= C_1, \end{aligned}$$

so $C_1 = 0$. Next, we have

$$0 = \Psi(L) = C_2 \sin(\omega \cdot L).$$

Now, how are we to solve this equation? We must have that input to the sin function must be an integer $n = \dots, -2, -1, 0, 1, 2, \dots$ copy of π as $\sin(n\pi) = 0$. Else, we force $C_2 = 0$ which gives us nothing! So, we require

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

Recall that $\omega = \frac{2mE}{\hbar^2}$ and that E is not determined (yet)! So now we have that $\omega = \frac{n\pi}{L}$ which gives us a general solution we will denote with a subscript n

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

Note that I have changed the constant just to C on ψ_n , and we'll solve for this constant soon enough. For each solution ψ_n there corresponds an energy E_n that we can solve for by

$$\begin{aligned} \omega_n^2 &= \left(\frac{n\pi}{L}\right)^2 \\ \frac{2mE_n}{\hbar^2} &= \left(\frac{n\pi}{L}\right)^2, \end{aligned}$$

which when we solve for E_n gives us

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

What have we found? We have found that to each solution ψ_n there corresponds an energy value E_n . We've also seen that there are infinitely many solutions but they correspond to **quantized** energy levels! We can disregard some of our solutions as well. In fact, we keep only $n = 1, 2, 3, \dots$ as $n = 0$ gives us $\psi_0(x) = 0$ which is not a physical solution (we'll see more on this in just a bit). Also, we have that $\psi_n(x) = -\psi_{-n}(x)$, so the negative values for n are simply redundant.

The interpretation of the ψ_n is that they are the allowable **stationary states** of the system. In this case, the system is the particle in a 1-dimensional box. In general, a **wavefunction** $\Psi(x, t)$ will be a sum of these states, but with an added time dependence that we did not talk about here. The wavefunction describes a **probability** of a particle being at a position. How so? Well, the following integral gives us the probability (a number $P \in [0, 1]$) of a particle with a wavefunction $\Psi(x, t)$ being the region $[a, b]$.

$$P([a, b] = \int_a^b |\Psi(x, t)|^2 dx = \int_a^b \Psi^* \Psi dx,$$

where $|\Psi(x, t)|$ represents the modulus of the complex wavefunction $\Psi(x, t)$. This interpretation of quantum mechanics is just another part of what makes the theory so different from classical theory! We can only make predictions in a probabilistic way! Truly, it's fascinating. Keep in mind, we really can only do this for the initial profile $\Psi(x, 0)$ since we have not introduced time dependence just yet.

In order to have this integral give us a valid probability at time $t = 0$, we must **normalize** the stationary states $\psi_n(x)$ by requiring that

$$\int_{-\infty}^{\infty} |C\psi_n(x)|^2 dx = 1.$$

We can compute this explicitly by finding C . So we have for the particle in a 1-dimensional box,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} |C\psi_n(x)|^2 dx \\ &= C^2 \int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx \quad \text{since } \psi_n = 0 \text{ outside of } [0, L] \\ &= C^2 \frac{L}{2} \end{aligned}$$

which gives us that the **normalization constant** $C = \sqrt{\frac{2}{L}}$ for any state ψ_n . So the normalized stationary states of the system are then

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

In general, one would hope to combine our states to create a more general wavefunction as

$$\Psi(x) = \sum_{j=1}^{\infty} a_j \psi_j.$$

We refer to this $\Psi(x)$ as a superposition of states. Without adding in time dependence, it is **not** true that a superposition of states solves the time independent Schrödinger equation!

Next, we can consider the **inner product** between states ψ_n and ψ_m which tells us how much of each state "points in the same direction"

$$\langle \psi_n, \psi_m \rangle := \int_{-\infty}^{\infty} \psi_n^*(x) \psi_m(x) dx.$$

If we evaluate this integral, we find that

$$\begin{aligned} \int_{-\infty}^{\infty} \psi_n^*(x) \psi_m(x) dx &= \frac{2}{L} \int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx \\ &= \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases}. \end{aligned}$$

Sometimes we denote this by putting

$$\langle \psi_n, \psi_m \rangle = \delta_{nm}$$

and saying that the ψ_n are *orthogonal*. In fact they are even *orthonormal* since

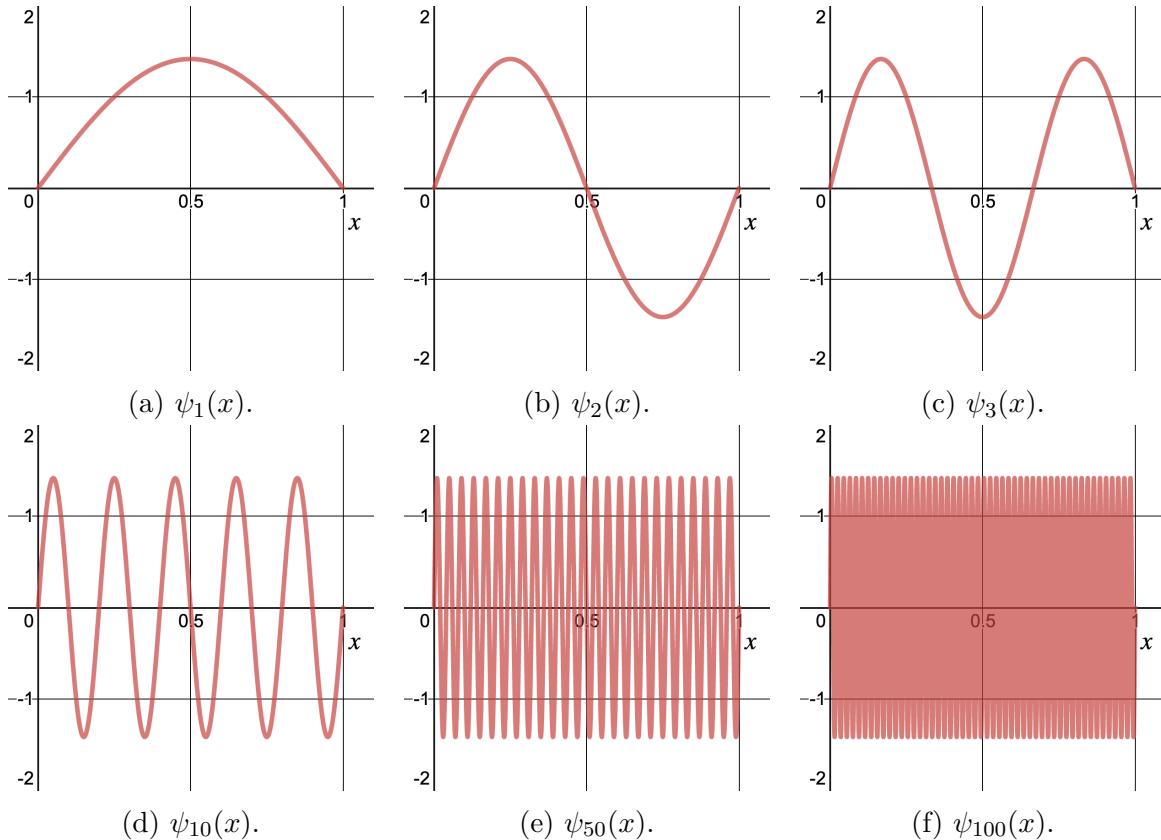
$$\langle \psi_n, \psi_n \rangle = 1,$$

since we have normalized each state.

Exercise 5.2.1. Evaluate the above integral for $\langle \psi_n, \psi_m \rangle$ to show that you get $\langle \psi_n, \psi_m \rangle = \delta_{nm}$. Hint: use the following substitution

$$\sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) = \frac{1}{2} \left[\cos\left(\frac{(n-m)\pi x}{L}\right) - \cos\left(\frac{(m+n)\pi x}{L}\right) \right].$$

What do these states of the system look like? We can plot them for different values of n . Let's take a look. Here we will let $L = 1$ for the plots.



One may notice that as we allow for larger and larger n values, the states seem to start to approach a uniform value on the region $[0, L]$. In fact, if we were to take the limit as $n \rightarrow \infty$, we would essentially have that $\psi_\infty(x) = \frac{1}{L}$ which would be equivalent to the case of a classic particle randomly placed in the box.

When we add in time dependence, it is possible to prepare this system in a *superposition* by, for example, considering the wavefunction

$$\Psi(x, 0) = \frac{1}{\sqrt{2}} (\psi_1(x) + \psi_2(x)).$$

More generally, due to the orthonormality of the stationary states, we can write a wavefunction $\Psi(x, 0)$ as

$$\boxed{\Psi(x, 0) = \sum_{j=1}^{\infty} a_j \psi_j(x) \quad \text{where} \quad \sum_{j=1}^{\infty} a_j^2 = 1.}$$

Summary of the 1-Dimensional Box

- The equation for the 1-Dimensional box is the same as the eigenfunctions of the Laplacian equation shown before. That is, we have $V(x) = 0$ and we allow for x to be in the range $[0, L]$ and get the equation

$$\begin{aligned} H\Psi &= E\Psi \\ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Psi &= E\Psi. \end{aligned}$$

- We solved this equation as a second order linear equation and used the boundary values to determine possible solutions. It turns out that we get a quantized set of solutions ψ_n each corresponding to an energy level E_n . These energy levels were given by $E_n = \frac{n^2 \hbar^2 \pi^2}{2mL}$ and so the energies increase proportionally to n^2 . We can see a plot of energies here.

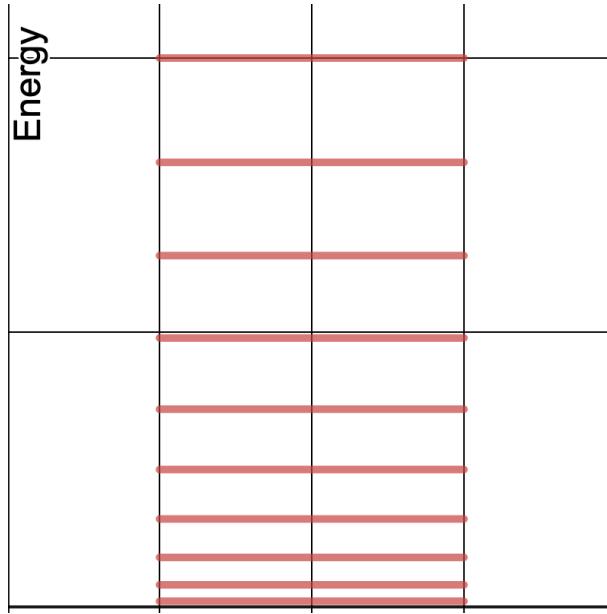


Figure 5.3: Energy levels of the 1-dimensional box.

- We normalized the solutions ψ_n and found that they give rise to a probability function that describe the probability of a particle being in a region $[a, b]$.
- We can take an initial profile for a wavefunction as a superposition

$$\Psi(x, 0) = \sum_{j=1}^{\infty} a_j \psi_j(x)$$

where $\sum_{j=1}^{\infty} a_j^2 = 1$. If more than one $a_j \neq 0$, then this wavefunction describes a superposition state.

- We can then find the probability of a particle with wavefunction $\Psi(x, t)$ being in the region $[a, b]$ at time t by

$$P([a, b]) = \int_a^b |\Psi(x, t)|^2 dx.$$

To do this to the full extent, we do need to add in time dependence which we do not consider at this moment! However, alluding to what we wrote for $\Psi(x, 0)$, we can compute this probability for a superposition of states ψ_n at time $t = 0$.

- As we let $n \rightarrow \infty$, we recover the case of the classical particle. We should expect this since we do not observe this odd quantum phenomenon on large scale!

Exercise 5.2.2. Look up common energies for something like throwing a baseball. Compare this to E_n for $m = 1[g]$ and $L = 1[m]$. You will have to use the measured value for Planck's reduced constant

$$\hbar \approx 1.0545718 \cdot 10^{-34} [kg \cdot m^2/s].$$

Remark 5.2.1. The time dependent aspect will be covered after we learn multivariate calculus in Math 272. There we will find that superposition of time dependent states does indeed give us a solution to the time dependent version of the equation.

Problems

Problem 5.1. Finish all the exercises in this chapter.

Problem 5.2. Consider the boundary value problem

$$x'' = g$$

with boundary values $x(0) = 0$, $x\left(-\frac{2}{g}\right) = 0$ and $g = -9.8[m/s^2]$. We can think of this as solving the *inverse problem* of one that we have seen in a homework. Specifically, think of this as knowing where a ball is launched and knowing where it lands and trying to find the speed it must have been thrown at.

Another interpretation is the shape of a rod bending due to gravity. x'' would measure the curvature of this rod, and this equation would say that the rod under the force of gravity would have a constant curvature. In this case, the dependent variable t should be thought of as spatial rather than temporal.

Finally, this equation above is referred to as *Poisson's equation*.

- (a) Find the general solution. If you already know it from the homework, just write it down.
- (b) Use the boundary values above to find the particular solution.
- (c) Is the solution unique?

Problem 5.3. Consider the *time independent Schrödinger equation* for a *free particle* constrained inside of a 1-dimensional box of length L . That is, we have the equation

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

on the unit interval $[0, L]$.

- (a) Find the general solution to this equation with no constraint.
- (b) Given the constraint, we have the boundary values $\psi(0) = \psi(L) = 0$. What are the general solutions given this constraint?
- (c) Show that the sum of two solution $\psi_1(x)$ and $\psi_2(x)$ is also a solution. When we have a particle whose state (or *wavefunction*) ψ is a sum of general solutions, we say that ψ is in a *superposition state*.
- (d) The wavefunction is not really a physically meaningful quantity. However, if we consider a region $[a, b]$ in the box $[0, L]$ the quantity

$$P([a, b]) = \int_a^b |\psi(x)|^2 dx$$

is meaningful. This expression tells us the *probability* that a particle will be observed in the region $[a, b]$. Take your general solutions you found in (b) (with the constraint) and solve for the constants that give you

$$\int_0^L |\psi(x)|^2 dx = 1.$$

We call this *normalization* and we must do so for each state so that we can interpret the integral $P([a, b])$ as a probability.

Problem 5.4. Consider the Schrödinger equation for a particle in a box of length 1,

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

with boundary conditions $\Psi(0) = 0$ and $\Psi(1) = 0$.

- (a) Find the general solution to the differential equation.
- (b) Apply the boundary conditions and write down a solution for each positive integer n . Recall that we call these solutions *states* and denote the states by ψ_n .
- (c) Determine the normalization constant for each ψ_n . Does this constant depend on n ?
- (d) Explain why a superposition of states is also a solution to the Schrödinger equation.

Problem 5.5. Schrödinger's equation is given by

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

where $V(x)$ is the potential at a point x , E is the total energy, and Ψ is the wave function.

- (a) For the free particle in a 1-dimensional box $[0, L]$, what are the boundary conditions for the wavefunction Ψ ?

- (b) Again, for the free particle in a 1-dimensional box, what is the potential inside of the box $(0, L)$?

- (c) If a wavefunction Ψ is normalized, then

$$\int_0^L \|\Psi(x)\|^2 dx = 1.$$

True or false?

- (d) If our wavefunction Ψ is normalized, how do we interpret the quantity

$$P([a, b]) = \int_a^b \|\Psi(x)\|^2 dx?$$

Part II

Elements of Analysis

6

Sequences and Series

6.1 Sequences

The next stop for us is sequences and series. In the realm of analysis, these are very important objects. It turns out that they are rather useful in applied sciences as well. Specifically, power series are undeniably useful. However, to gain understanding for those series, we need to start at the beginning with sequences. From there, we can realize a series as a sequence. Finally, we can turn functions into series, which, believe it or not, gives us the ability to solve problems like differential equations or to approximate harder problems in a nicer way.

Definition 6.1.1: Sequences

An (infinite) *sequence* $\{a_n\}_{n=1}^{\infty}$ is an (infinite) list of numbers typically written as follows:

$$\{a_n\}_{n=1}^{\infty} = a_1, a_2, a_3, \dots$$

While some speak of finite sequences, we will not. All sequences we consider are infinite, and so we drop the extra adjective.

Sequences can come up in all types of ways. For example, consider these two sequences related to the number π . We can take

$$a_n = \text{the } n^{\text{th}} \text{ digit of the number } \pi$$

and so

$$\{a_n\}_{n=1}^{\infty} = 3, 1, 4, 1, 5, \dots$$

We could also consider a sequence that successively approximates π as

$$b_n = \text{the } n^{\text{th}} \text{ decimal approximation of } \pi$$

which is

$$\{b_n\}_{n=1}^{\infty} = 3, 3.1, 3.14, 3.141, 3.1415, \dots$$

6.1.1 Convergence

When investigating series, we often care about whether they seem to trend towards a common value or not. This can be rather hard to deal with in some cases. But, the intuition on when a series converges is very important and shows up in various ways one should care about. For example, we often use computer approximations for functions (since most functions cannot be exactly calculated with a computer) and we should know when these approximations are accurate and tend towards the exact function we desire. This motivates the following definition.

Definition 6.1.2: Convergence of a Sequence

Consider a sequence $\{a_n\}_{n=1}^{\infty}$. We say that the sequence *converges* to its *limit* L if for any number $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for any $K \geq N$ we have

$$|a_K - L| < \epsilon.$$

We then write

$$\lim_{n \rightarrow \infty} a_n = L$$

or

$$a_n \rightarrow L.$$

If a sequence does not converge to any limit, then we say the sequence *diverges*.

Admittedly, this definition is a bit obtrusive. What is it really saying? Let's decode this a bit. Think of ϵ as a tolerance parameter (i.e., how accurate you would like a measurement to be) telling us how far from the ideal measurement L we are. Then, what this definition is saying is that if we look far enough along in our sequence we can be as close to the ideal L as we wish.

Example 6.1.1: Approximations with π

Consider the two sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ we generated for π . Which ones converge? Let us take a look first at

$$\{a_n\}_{n=1}^{\infty} = 3, 1, 4, 1, 5, \dots$$

Note that this sequence, if we keep looking at successive digits, does not approach any single value. In fact, if it did approach a single value it would have to approach a decimal $0, 1, \dots, 9$. Let's say that our sequence did converge to the decimal 0. In which case our definition of convergence would imply that if we chose a tolerance of

0.5, then at some point in our sequence we would have

$$|a_K - 0| = |a_n| = a_n < \epsilon = 0.5,$$

and so all a_n after some point must be equal to 0. But, if that was the case, then π would not have an infinite decimal expansion! And if we had chosen another decimal, this would still mean that π is a rational number (which we can prove it is not). Hence, a_n does not converge.

Considering $\{b_n\}_{n=1}^{\infty}$ may be a bit easier. We designed this sequence to continually approximate π as we put

$$\{b_n\}_{n=1}^{\infty} = 3, 3.1, 3.14, 3.141, 3.1415, \dots$$

One should believe then that $b_n \rightarrow \pi$. We can test this with our definition. Say we let $\epsilon = 0.1$, then we want

$$|b_K - \pi| < \epsilon = 0.1.$$

Now, b_n is defined to be the n^{th} approximation of π , so if we choose $N = 2$, then for $K \geq N = 2$, b_K has at least the first two digits of π correctly approximated. Then

$$|b_K - \pi| \leq |0.0415\dots| < 0.1 = \epsilon.$$

Similarly, we could choose $\epsilon = 0.01$ and let $N = 3$.

Often we want to define sequences in a functional way. That is, we like to specify what the n^{th} term of the sequence is by a function of n , $f(n)$. This is typically how sequences appear and it makes working with the definition of convergence a bit more simple as we can use rules from limits that we already know.

Example 6.1.2: A Sequence Given by a Function

Consider the sequence $\{a_n\}_{n=1}^{\infty}$ where

$$a_n = f(n) = \frac{1}{n}.$$

We can then write out the sequence

$$\{a_n\}_{n=1}^{\infty} = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$$

It seems like numbers in this sequence get smaller and smaller in magnitude, but are never negative. So, one may guess that $\lim_{n \rightarrow \infty} a_n = 0$. Let us prove that this is true. Fix some $\epsilon > 0$, then we want to find an N so that for all $K \geq N$ we have

$$|a_K - 0| = |a_K| = a_K < \epsilon.$$

Now, we have that

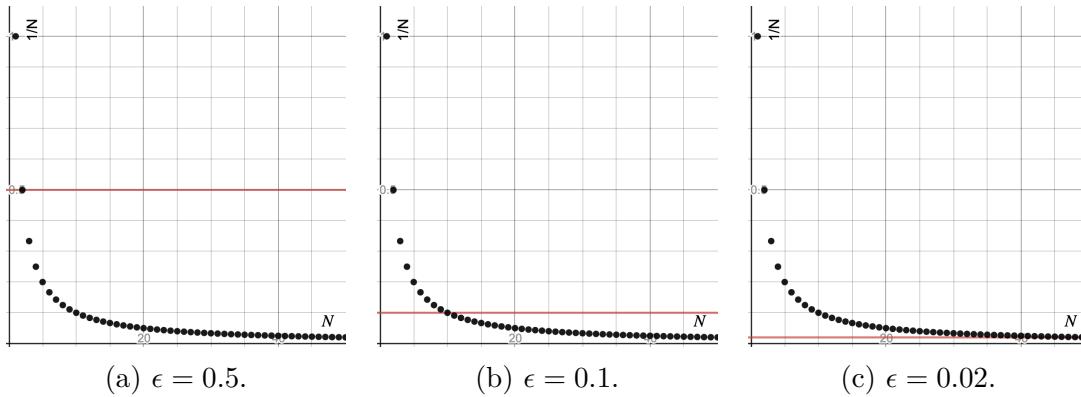
$$a_K = \frac{1}{K} < \epsilon$$

which we can rearrange

$$\begin{aligned} \frac{1}{K} &< \epsilon \\ \iff \frac{1}{\epsilon} &< K. \end{aligned}$$

Hence, if we pick a $N \geq K > \frac{1}{\epsilon}$ we guarantee that we satisfy our tolerance condition. Thus, since we found values of N that satisfy a generic tolerance condition, we know that $a_n \rightarrow 0$.

We can see a picture of what is going on here. I'll put the ϵ as a solid red line, and graph the points a_N versus N as (N, a_N) .



This definition of convergence is essentially the exact same definition as what we have seen for functions converging to their limit. Hence, all the results we have on functions extends to this case. This is why we wish to work with sequences defined by functions $f(n)$.

Proposition 6.1.1: Limit of Functions Gives the Limit for a Sequence

Consider a sequence $\{a_n\}_{n=1}^{\infty}$ with $a_n = f(n)$. Then the sequence converges to L if and only if $\lim_{n \rightarrow \infty} f(n) = L$.

Here's a short list of some limits you should feel free to use.

Function for a_n	Limit $\lim_{n \rightarrow \infty} a_n$
a^n	$\begin{cases} 0 & \text{if } a < 1 \\ 1 & \text{if } a = 1 \\ \text{Diverges} & \text{otherwise} \end{cases}$
n^p	$\begin{cases} 0 & \text{if } p \leq 0 \\ \text{Diverges} & \text{if } p > 0 \end{cases}$

It should be noted that not all sequences diverge in the same way. Let us see this with

an example of two different cases.

Example 6.1.3: Different Ways of Diverging

Consider the sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ given by

$$a_n = (-1)^n \quad \text{and} \quad b_n = n.$$

We can write out the first few terms of each series

$$\begin{aligned}\{a_n\}_{n=1}^{\infty} &= -1, 1, -1, 1, -1, 1, \dots \\ \{b_n\}_{n=1}^{\infty} &= 1, 2, 3, 4, 5, 6, \dots.\end{aligned}$$

Notice that $\{a_n\}_{n=1}^{\infty}$ never settles to a specific value as it just oscillates between -1 and 1 indefinitely. For $\{b_n\}_{n=1}^{\infty}$ we see that this sequence grows as n grows. We sometimes add that $\{b_n\}_{n=1}^{\infty}$ **diverges to infinity**. If we took $c_n = -n$, we would have that $\{c_n\}_{n=1}^{\infty}$ **diverges to minus infinity**.

Another way to describe a sequence converging is to show that the difference between successive values (i.e., differences between a_N and a_{N+1}) get arbitrarily small. Let us write this as a theorem. We'll define one term in this theorem as well.

Theorem 6.1.1: Cauchy Sequences Converge

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence satisfying that for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $K \geq N$ we have

$$|a_K - a_{K+1}| < \epsilon.$$

We call such a sequence $\{a_n\}_{n=1}^{\infty}$ a **Cauchy sequence**. Then if $\{a_n\}_{n=1}^{\infty}$ is a real (or complex) valued Cauchy sequence, then it is also convergent.

This fact will become useful later on when we study series as it leads us to the comparison test.

6.1.2 Recursive Sequences

Some sequences we will see are defined in a *recursive* way. Meaning that instead of supplying a function for the terms (i.e., $a_n = f(n)$), we tend to start with an initial term a_1 and define future terms based on that. For example, we could define a sequence $\{a_n\}_{n=1}^{\infty}$ by

$$a_1 = 1 \quad \text{and} \quad a_n = \frac{1}{2}a_{n-1}.$$

From this definition, we can find that

$$a_2 = \frac{1}{2}a_1 = \frac{1}{2} \cdot 1 = \frac{1}{2},$$

and

$$a_3 = \frac{1}{2}a_2 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

If we continued on, we could write out the sequence as

$$\{a_n\}_{n=1}^{\infty} = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$$

When we define a sequence in this way, we call it *recursive*.

Example 6.1.4: Fibonacci Sequence

One very important recursive sequence is the *Fibonacci sequence* $\{F_n\}_{n=1}^{\infty}$ defined by summing up the previous two terms in the sequence to get the new one. That is, we can define

$$F_1 = 0, \quad F_2 = 1, \quad \text{and} \quad F_n = F_{n-1} + F_{n-2}.$$

We can write the first few terms in this sequence by following the above rule with the initial values for F_1 and F_2 to get

$$\{F_n\}_{n=1}^{\infty} = 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \dots$$

This sequence shows up in many natural systems!

Exercise 6.1.1. Look up where this Fibonacci sequence shows up in nature.

6.2 Series

What else can we do with sequences? Well, instead of just keeping a sequence $\{a_n\}_{n=1}^{\infty}$ as a list of numbers, we could consider operations with these infinitely many numbers. For example, we could add the numbers as

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \dots$$

or even multiply them

$$\prod_{n=1}^{\infty} a_n = a_1 \cdot a_2 \cdot a_3 \cdots.$$

In our case, we will care about adding up terms in a sequence.

Definition 6.2.1: Series

Given a sequence $\{a_n\}_{n=1}^{\infty}$ we can create an (infinite) *series*

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

What does it really mean to add up infinitely many numbers? Can this even be a finite result? It turns out that both of these questions can be answered at the same time.

If we rethink what we are doing a bit, it turns out that we need no other definition than convergence of a sequence. Our goal now is to make a series into a sequence, and we will see how we can make sense of this infinite sum.

Consider a series $A = \sum_{n=1}^{\infty} a_n$. We can then consider ***partial sums*** of the series by looking at finite approximations to the infinite sum. That is, the N^{th} partial sum is given by

$$A_N = \sum_{n=1}^N a_n = a_1 + a_2 + a_3 + \cdots + a_N.$$

Now, we can collect all of the partial sums A_1, A_2, A_3, \dots and notice that this creates a sequence $\{A_n\}_{n=1}^{\infty}$.

Definition 6.2.2: Convergence of a Series

Given a series $\sum_{n=1}^{\infty} a_n$ and the sequence of partial sums $\{A_n\}_{n=1}^{\infty}$. We say that the series $\sum_{n=1}^{\infty} a_n$ ***converges*** to A if the sequence of partial sums converges to A and put

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

If the series does not converge, we say that it ***diverges***.

Remark 6.2.1. We should note that a series $\sum_{n=1}^{\infty} a_n$ cannot converge if the sequence $\{a_n\}_{n=1}^{\infty}$ does not converge to zero!

Example 6.2.1: Series Approximation of π

Previously, we had the sequences

$$\{a_n\}_{n=1}^{\infty} = 3, 1, 4, 1, 5, \dots$$

and

$$\{b_n\}_{n=1}^{\infty} = 3, 3.1, 3.14, 3.141, 3.1415, \dots$$

that we used to build up the number π . It turns out we can relate these two sequences in the following way. Consider the series

$$\sum_{n=1}^{\infty} 10^{-n+1} a_n = 3 + 0.1 + 0.04 + 0.001 + 0.0005 + \cdots.$$

Note, we are adding numbers of smaller and smaller magnitude. Now, notice if we take the sequence of partial sums, we find that we get back the sequence $\{b_n\}_{n=1}^{\infty}$

$$b_N = \sum_{n=1}^N 10^{-n+1} a_n$$

Since we worked with the sequence $\{b_n\}_{n=1}^{\infty}$ previously, we know that we have $b_n \rightarrow \pi$ and thus the series converges to π as well

$$\pi = \sum_{n=1}^{\infty} 10^{-n+1} a_n.$$

We are often provided each element of a sequence by a function. That is, $a_n = f(n)$. If this is the case for terms in a series as well, it turns out we can use other means to show a series converges. Let us consider the following series.

Example 6.2.2: Geometric Series

Consider a sequence defined by $a_n = ar^n$. Then we call the following a *geometric series*

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} ar^n$$

In order for the series to converge, we must have that the sequence $a_n \rightarrow 0$. This means that we must have $|r| < 1$ as if $|r| < 1$ then $\lim_{n \rightarrow \infty} a_n = 0$. If this is the case, then we can take our series

$$\sum_{n=1}^{\infty} ar^n = a \sum_{n=1}^{\infty} r^n = a(r + r^2 + r^3 + \dots).$$

So the N^{th} partial sum is

$$A_N = a(r + r^2 + r^3 + \dots + r^N).$$

Then we can subtract rA_N from both sides

$$\begin{aligned} A_N - rA_N &= a(r + r^2 + r^3 + \dots + r^N) - a(r^2 + r^3 + r^4 + \dots + r^{N+1}) \\ (1-r)A_N &= a(r - r^{N+1}) \\ A_N &= a \left(\frac{r - r^{N+1}}{1-r} \right). \end{aligned}$$

Now consider the limit for the sequence of partial sums

$$\begin{aligned} \lim_{N \rightarrow \infty} A_N &= \lim_{N \rightarrow \infty} a \left(\frac{r - r^{N+1}}{1-r} \right) \\ &= \frac{ar}{1-r}. \end{aligned}$$

Thus, we have that the geometric series converges

$$\boxed{\sum_{n=1}^{\infty} ar^n = \frac{ar}{1-r}.}$$

Now, not only have we shown that a whole set of series converges, but we have found what they converge to! And, as it turns out, geometric series show up quite often. We can use our formula to find the sum of following series.

Example 6.2.3: A Specific Geometric Series

Consider the geometric series

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

Note that this is a geometric series with $a = 1$ and $r = \frac{1}{2}$ as we can write the series as

$$\sum_{n=1}^{\infty} ar^n = \sum_{n=1}^{\infty} 1 \cdot \left(\frac{1}{2}\right)^n.$$

Our formula says that

$$\sum_{n=1}^{\infty} ar^n = \frac{ar}{1-r}$$

and hence we can plug in for both a and r to see

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1/2}{1 - 1/2} = 1.$$

How about convergence for other types of series? There is a test that we can perform that will tell us more information. We call this test the **ratio test**. However, the ratio test won't be a test that tells us if *any* series converges or diverges.

Proposition 6.2.1: Ratio Test

Consider the series $\sum_{n=1}^{\infty} a_n$. If we have that

$$L = \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1,$$

then the series converges (absolutely). If $L > 1$, then the series diverges. If $L = 1$, then we gain no information.

Proof. First, if $L > 1$, then each term is growing in magnitude for this series, and thus $a_n \not\rightarrow 0$ and the series cannot converge.

If $L < 1$, then at some point along our series, $N \in \mathbb{N}$, we have that $1 > r > \left| \frac{a_{n+1}}{a_n} \right|$

$$\begin{aligned} \sum_{n=1}^{\infty} |a_n| &= \sum_{n=1}^N |a_n| + \sum_{j=N}^{\infty} |a_j| \\ &= A_N + \sum_{j=1}^{\infty} |a_{N+j}| \\ &< A_N + \sum_{j=1}^{\infty} r^j |a_{N+1}| \\ &= A_N + |a_{N+1}| \sum_{j=1}^{\infty} r^j \\ &= A_N + \frac{|a_{N+1}|r}{1-r} < \infty. \end{aligned}$$

Hence the series converges (absolutely). □

Remark 6.2.2. A series $\sum_{n=1}^{\infty} a_n$ **converges absolutely** if

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

converges.

The ratio test is a helpful tool, but it does have its shortcomings. We can take a look at an example where the test tells us pertinent information and one where it does not.

Example 6.2.4: Series for e

One can actually define the number e by the series

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

where $n!$ is the factorial function

$$n! = n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1$$

and we define $0! = 1$. Now, by saying the series above equals e , we should think that the series converges. But we can prove this is true using the ratio test. Consider the terms

$$a_n = \frac{1}{n!} \quad \text{and} \quad a_{n+1} = \frac{1}{(n+1)!} = \frac{1}{(n+1)n!}.$$

Then if we consider

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| &= \lim_{n \rightarrow \infty} \left| \frac{\frac{1}{(n+1)n!}}{\frac{1}{n!}} \right| \\ &= \lim_{n \rightarrow \infty} \left| \frac{n!}{(n+1)n!} \right| \\ &= \lim_{n \rightarrow \infty} \left| \frac{1}{n+1} \right| \\ &= 0. \end{aligned}$$

Hence, by the ratio test with $L < 1$, we know the series converges. And as stated above, it converges to Euler's number e .

Below is an important class of series called *p-series*. These and geometric series are some of the most common that we will see. It turns out that some *p*-series converge and others diverge depending on the value for p . However, the ratio test gives us no information on any of them.

Example 6.2.5: p -Series

Consider a general p -series which has the form

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

where p can be any real number. Now, if $p \leq 0$, then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^p} > 0$$

and the series cannot possibly converge. However, with the ratio test for any p , we see that

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n \rightarrow \infty} \left| \frac{\frac{1}{(n+1)^p}}{\frac{1}{n^p}} \right| = \lim_{n \rightarrow \infty} \left| \frac{n^p}{(n+1)^p} \right| = 1.$$

So, for any value of p , the ratio test tells us nothing. However, we can show that the p -series does converge for certain values of p .

Exercise 6.2.1. Investigate the limit in the ratio test

$$\lim_{n \rightarrow \infty} \left| \frac{\frac{1}{(n+1)^p}}{\frac{1}{n^p}} \right|$$

in more detail to show that it indeed is equal to one.

There is another test we can use to determine convergence of series which will work in some cases when the ratio test does not. We call this the **integral test**.

Proposition 6.2.2: Integral Test

Consider the series $\sum_{n=N}^{\infty} a_n$ with $a_n = f(n)$. Then the series converges if and only if

$$\int_N^{\infty} f(x) dx$$

is finite. If the integral diverges, the series does as well. Note, the integral is likely not equal to the series. It only tells us whether the series converges or not!

Example 6.2.6: p -Series Integral Test

Going back to the p -series, we need to check whether for $p > 0$ the series converges as we already ruled out $p \leq 0$ cannot. So our series in question is

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

so our $f(n) = \frac{1}{n^p}$ and our $N = 1$. Thus we investigate the integral

$$\int_1^\infty \frac{1}{x^p} dx.$$

Now, if $0 < p < 1$, then we have

$$\begin{aligned} \int_1^\infty \frac{1}{x^p} dx &= \left[\frac{1}{-p+1} x^{-p+1} \right]_1^\infty \\ &= \infty, \end{aligned}$$

since $-p + 1 > 0$ and so x^{-p+1} approaches infinity as $x \rightarrow \infty$ and this integral diverges. Hence, the p -series also diverges for $0 < p < 1$.

For $p = 1$, we find that the integral is

$$\int_1^\infty \frac{1}{x} dx = [\ln(x)]_1^\infty$$

and $\ln(x) \rightarrow \infty$ as $x \rightarrow \infty$ so this integral diverges as well.

For $p > 1$, we have

$$\int_1^\infty \frac{1}{x^p} dx = \left[\frac{1}{-p+1} x^{-p+1} \right]_1^\infty$$

where $-p + 1 < 0$. Thus, in this case, $x^{-p+1} \rightarrow 0$ as $x \rightarrow \infty$ and this integral is finite! So, we have the following result.

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \quad \begin{cases} \text{Diverges} & \text{if } p \leq 1 \\ \text{Converges} & \text{if } p > 1 \end{cases}.$$

One last useful test is to compare to series we know more about. We call this the **comparison test**

Proposition 6.2.3: Comparison Test

Suppose that we have two series $\sum_{n=N_1}^{\infty} a_n$ and $\sum_{n=N_2}^{\infty} b_n$ with $a_n, b_n \geq 0$ and for some $N \in \mathbb{N}$ we have that for all $K \geq N$ that $a_n \leq b_n$. Then

- If $\sum_{n=N_2}^{\infty} b_n$ is convergent, then so is $\sum_{n=N_1}^{\infty} a_n$.
- If $\sum_{n=N_1}^{\infty} a_n$ is divergent, then so is $\sum_{n=N_2}^{\infty} b_n$.

Problems

Problem 6.1. Complete all the exercises in this chapter.

Problem 6.2. Write down the first few terms in the sequence for the following:

- (a) $a_n = n$;

(b) $b_n = \frac{1}{n^2}$;

(c) $c_n = 2^{-n}$.

Problem 6.3. For the above sequences, state whether each converges or diverges. If they converge, state the limit.

Problem 6.4. Consider the recursive sequence

$$a_n = \frac{1}{2}a_{n-1} + 1$$

with $a_1 = 1$.

(a) Write the first few terms in the sequence.

(b) Can you write a_n as a function $f(n)$? If so, what is $f(n)$?

(c) Does this sequence converge or diverge? Can you show why with a limit $\lim_{n \rightarrow \infty} f(n)$?

(d) Can you show that this is a Cauchy sequence?

Problem 6.5. Consider the sequence

$$a_n = ar^n.$$

(a) If $|r| < 1$, show that this sequence $\{a_n\}_{n=0}^{\infty}$ converges to zero.

(b) Consider now the *geometric series*

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} ar^n.$$

Show that the N^{th} partial sum for this series satisfies

$$\sum_{n=0}^N ar^n = a \left(\frac{1 - r^{N+1}}{1 - r} \right).$$

(c) Does the geometric series converge for all r ? For $|r| < 1$? When it converges, what does it converge to?

Problem 6.6. Consider the following sequences,

$$a_n = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^n}, \dots,$$

and

$$b_n = 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \dots, \frac{1}{n!}, \dots$$

(a) For what values of N do we need for $a_N < 0.01$ and $b_N < 0.01$? Note, these will be different values for N .

(b) Compute $\lim_{n \rightarrow \infty} a_n$.

(c) Compute $\lim_{n \rightarrow \infty} b_n$.

- (d) Which sequence converges more quickly to its limit? (*Hint: consider a ratio of the terms of the sequences and take a limit. Part (a) should help you think about this.*)

Problem 6.7. With the same a_n from Problem 6.6, consider the series

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

- (a) Write down the N^{th} partial sum A_N for this series.
- (b) Does this sequence of partial sums converge? If so, to what?
- (c) Note that this is an *geometric series* with $a = 1$ and $r = \frac{1}{2}$. However, we start from $n = 1$ instead of $n = 0$. Show the value that this series converges to using the formula for a geometric series.

Problem 6.8. With the same b_n from Problem 6.6, consider the series

$$B = \sum_{n=0}^{\infty} b_n.$$

- (a) Use the ratio test to show that this series converges.
- (b) Approximate the value the series converges to by considering larger and larger partial sums.
- (c) What number does this series converge to?

Problem 6.9. Consider the *p-series*:

$$\sum_{n=1}^{\infty} \frac{1}{n^p}.$$

- (a) For $p = 1$, show that the ratio test is inconclusive.
- (b) For $p = 2$, show that the ratio test is again inconclusive.
- (c) Look up the sum of the series for $p = 1$ and $p = 2$. Notice how the ratio test is not perfect!

Problem 6.10 (Explicit Euler Method). Consider the differential equation $x' = kx$ where $k \in \mathbb{C}$ is a complex parameter for the system. Note that the solution to this equation given the initial condition $x(0) = 1$ is $x = e^{kt}$.

- (a) Suppose we want to find an approximation to a solution using a computer. Let t_0 be some arbitrary time, define δt to some fixed change in the input variable t and let $\delta x = x(t_0 + \delta t) - x(t_0)$ be the corresponding change in the output x . Compute the first order Taylor approximation of x at the point t_0 to see that

$$x(t_0 + \delta t) \approx x(t_0) + x'(t_0)(\delta t) \quad (6.2.1)$$

from which you can then note that

$$x'(t_0) \approx \frac{\delta x}{\delta t} \quad (6.2.2)$$

- (b) Define the *explicit Euler approximation sequence* $\{x_\tau\}_{\tau=0}^T$ so that $x(t_0) = x_0$ and at later times $x(t_0 + \tau\delta t) = x_\tau$. Show using the previous equations, the ODE itself, and the fact that $x_{\tau-1} + \delta x = x_\tau$ means can make a sequence

$$x_\tau = x_{\tau-1} + kx_{\tau-1}\delta t.$$

- (c) Let $k = 1$, $\delta t = 0.01$, $t_0 = 0$, and let $x(0) = 1$. Plot the explicit Euler approximation sequence using the following URL <http://www.calcul.com/show/calculator/recursive>. Compare this graph to the solution $x(t) = e^t$.
- (d) Let $k = -1$, $\delta t = 2$, $t_0 = 0$, and let $x(0) = 1$. Plot the approximation again. Is there something wrong? How does this compare to what the solution should be?

Problem 6.11. Sequences show up in the realm of approximation all the time. One of the first sequences one may encounter comes from *Newton's method* for calculating roots to a function (see this week's discussion for more). The idea is that we are handed a function $f(x)$ and pick an initial guess x_0 as a candidate for a root to $f(x)$. Then, we can compute a tangent line approximation to $f(x)$ at the point x_0 . That is,

$$f(x) \approx f'(x_0)(x - x_0) + f(x_0).$$

From there, we can see where this tangent line intersects the $y = 0$ axis which gives us a new x -value we call x_1 . If we solve for x_1 , we find that we get

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

From here, we can generate a recursive sequence given by

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}.$$

that *can possibly* give a better and better approximation to a root of $f(x)$, meaning that x_n will tend to be an better approximation to a root of $f(x)$ than x_{n-1} . This method is in fact exactly how your TI-XX calculator finds zeros to a function (notice that you give bounds and an initial guess when you do this)!

Feel free to use a calculator for this problem.

- (a) Consider the function $f(x) = x^2 - 1$ with the starting point of $x_0 = 2$, draw a picture that illustrates the first two iterations of Newton's method. (*Hint: This requires drawing a tangent line to $f(x)$ at two different points and telling me what the intersections of these lines with the x -axis represent.*)
- (b) Write the first four terms in the sequence for Newton's method. That is, find x_0, \dots, x_3 and write
- $$\{x_n\}_{n=0}^{\infty} = x_0, x_1, x_2, x_3, \dots$$
- (c) This sequence approaches $+\sqrt{1} = 1$ since this is the nearest root to $f(x)$ compared to our starting value. What is the first value of N where the error $|x_N - 1| < 10^{-6}$?
- (d) If we instead have $g(x) = x^3 - 2x + 2$ with an initial guess of $x_0 = 0$, write the first four terms of the sequence x_0, \dots, x_3 for Newton's method. Does this sequence seem to approach the real root $x \approx -1.7693$?



Power Series

Since we have constructed different sequences and series, we can move forward into making them more useful for us. The biggest use for series in our case will be to try and think of functions as a series. Specifically, we may be able to represent a function as a *power series*. This has many advantages! For one, we can differentiate and integrate a series term by term which simplifies calculations immensely. Another advantage is the ability to approximate complicated functions. This is especially useful if one ever plans to use a computer to perform computations.

7.1 Power Series

Formally, a *power series* is a series with a variable x (typically $x \in \mathbb{R}$ or $z \in \mathbb{C}$) written as

$$\sum_{n=0}^{\infty} a_n x^n \quad \text{or} \quad \sum_{n=0}^{\infty} a_n z^n.$$

An immediate question is whether or not this series converges and if that convergence depends on the choice of x we are inputting into the expression. The tools from the previous chapter allow us to determine this. But, before that, it is important to get a little intuition on what we're doing here.

A power series is much like a polynomial function

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N$$

except for that a power series continues on indefinitely. That is, there is no highest power of x for a power series! However, if we consider a partial sum of a power series we can realize that as a polynomial. For example, using the series above the N^{th} partial sum is

equivalent to the polynomial above

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

So, a finite part of a power series is just a polynomial. In all honesty, this is essentially the idea behind the construction to a power series. We tend to build these power series as polynomial approximations to functions we already know, or use power series to determine functions in a new way.

Example 7.1.1: Partial Sums for Power Series

Consider the power series with $a_n = \frac{1}{n!}$. We write this as

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This is an extremely important power series as we will repeatedly. We can also write this series as

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots.$$

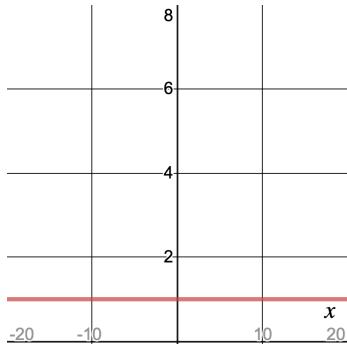
Now, let us take the following partial sums and see what the graphs look like.

$$\sum_{n=0}^0 \frac{x^n}{n!} = 1$$

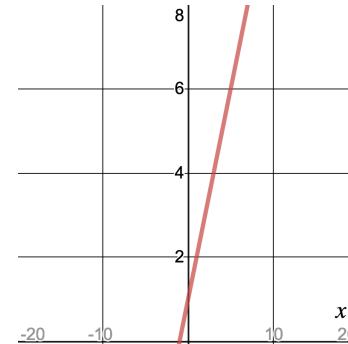
$$\sum_{n=0}^1 \frac{x^n}{n!} = 1 + x$$

$$\sum_{n=0}^6 \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720}.$$

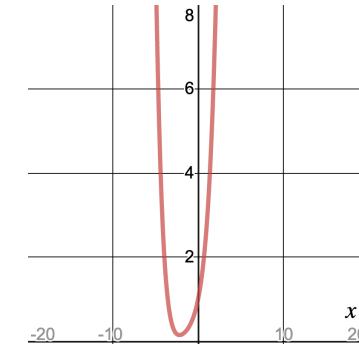
Notice how terms of higher powers of x have increasingly smaller contributions to the sum as the denominator gets larger and larger. Below are plots of these.



(a) Partial sum with $N = 0$.



(b) Partial sum with $N = 1$.



(c) Partial sum with $N = 6$.

Before we carry on and see this used in a more applicable way, we just need to know how to talk about convergence of a power series. A power series is no different than a

series, it's just that there is a variable input that may affect convergence. For example, for large n , if x is large, then x^n will be very large. Thus we must have that the denominator in each term of the power series also grows rapidly. We can test this with the ratio test and determine the values of x (or even $z \in \mathbb{C}$) in which the series converges.

Definition 7.1.1: Radius of Convergence

Consider a power series given by $\sum_{n=0}^{\infty} a_n x^n$. We define the *radius of convergence* to be the largest value for $|x|$ such that

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| < 1.$$

In the above definition we simply used the ratio test to define this. One should also note that this can be generalized slightly to allow for complex valued input $z \in \mathbb{C}$. All that has to be done is to exchange the absolute value for the modulus.

Example 7.1.2: Infinite Radius of Convergence

Consider again the power series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$. We wish to find the radius of convergence. So consider the limit

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}} \right| &= \lim_{n \rightarrow \infty} \left| \frac{x}{n+1} \right| \\ &= 0. \end{aligned}$$

Since x does not factor into this limit, it does not matter what values of x we plug in. That is, the series will converge no matter our choice for x . Fundamentally, this is because $n!$ grows rapidly even compared to x^n . Thus, the radius of convergence is infinite.

Example 7.1.3: Finite Radius of Convergence

Consider the power series

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots$$

This power series is similar to a p -series for $p = 1$. Hence, we know that if we take $x = 1$, the series does not converge. However, one should note that

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = \ln(2).$$

Given that, we should expect this series to converge for some values of x . So, let us

use our ratio test

$$\begin{aligned}\lim_{n \rightarrow \infty} \left| \frac{\frac{x^{n+1}}{n+1}}{\frac{x^n}{n}} \right| &= \lim_{n \rightarrow \infty} \left| \frac{nx}{n+1} \right| \\ &= |x| \cdot \lim_{n \rightarrow \infty} \frac{n}{n+1} \\ &= |x|.\end{aligned}$$

Hence, if we want the above limit to be less than one, then we must have $x < 1$. Thus the radius of convergence is one.

One often talks about functions being even or odd by using the following relationships:

- $f(x)$ is **even** if $f(-x) = f(x)$.
- $f(x)$ is **odd** if $f(-x) = -f(x)$.

This is captured very nicely by a power series. We often say that a power series defines a function and we write

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

and we consider the domain of $f(x)$ to be the x -values in which the series converges.

Proposition 7.1.1: Even and Odd Functions

Consider a function $f(x)$ given by a power series so that

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

where the series converges. We then say $f(x)$ is **even** if all odd coefficients $a_{2n+1} = 0$ and the function is **odd** if all even coefficients $a_{2n} = 0$.

In other words, if $f(x)$ is even, then

$$f(x) = \sum_{n=0}^{\infty} a_{2n} x^{2n}$$

and if $f(x)$ is odd, then

$$f(x) = \sum_{n=0}^{\infty} a_{2n+1} x^{2n+1}.$$

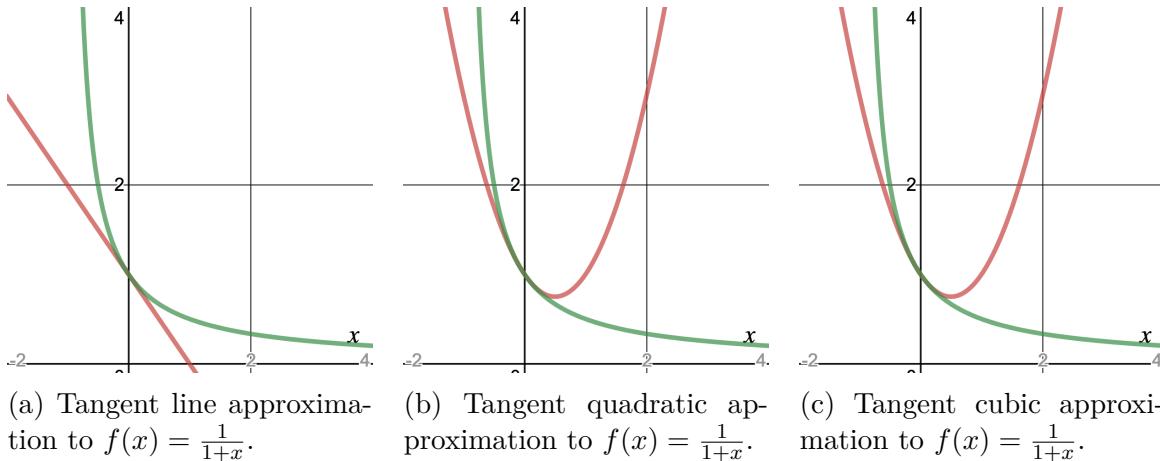
7.2 Taylor Series

Knowing when a power series converges is helpful, but we have yet to see a way to create relevant power series. The point of developing these power series is to give a different (and more useful) representation of functions. One way of doing this is called a *Taylor series*.

When first talking of derivatives, we think of tangent line approximations to functions. Specifically, the derivative is the slope of the tangent line. Recall, that if we are given a function $f(x)$, we can approximate $f(x)$ with a tangent line passing through the point a by

$$y = f'(a)(x - a) + f(a).$$

Now, we should think of the tangent line approximation as the beginning of a power series that approximates $f(x)$. The intuition to have now is that we could, in some sense, create a tangent quadratic approximation, and a tangent cubic approximation, and so on.



Let us think for a moment about how we may construct a series that gives us the above approximations and more. If we are to pick a point $x = a$ to build our approximation from, then the approximation should have the same value as the function at the point a . So, the zeroth order approximation (i.e., the zeroth term in the series we're building) should be $f(a)$. Notice that $f(a)$ occurs in the tangent line approximation. Next, if we take a first order approximation, this should give us an equation for a line. The best approximation around a to the function $f(x)$ with a line would be the tangent line approximation. The first derivative tells us the information about the slope. Now, notice the tangent line approximation above contains the zeroth order approximation as well. Functions also tend to have some curvature to them and this is captured nicely by the second derivative of the function at the point a . Now we can add in this second derivative information to get a tangent quadratic

$$f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2.$$

This would be our second order approximation. Similarly, we can build the tangent cubic (or third order approximation) by

$$f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3.$$

Taking this pattern in mind, we can create the N^{th} order approximation to be

$$f(a) + f'(a)(x - a) + \frac{f''(a)}{2!} + \cdots + \frac{f^{(N)}(a)}{N!}(x - a)^N.$$

As it turns out, for many functions we can continue this process infinitely and create a power series that exactly matches the function we started with.

Definition 7.2.1: Taylor Series

Given a function $f(x)$ that is *analytic* in a region around the point $x = a$, we can construct the **Taylor series centered at $x = a$**

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

which is equal to the function $f(x)$ (on a sufficiently small interval $(a - \epsilon, a + \epsilon)$). If we let $a = 0$, we call this the **Maclaurin series**.

Though the definition above only guarantees that the Taylor series is equal to the function on a small interval, many functions we care about will satisfy

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

for large intervals or even all real numbers.

Example 7.2.1: Maclaurin Series for e^x

Consider the function $f(x) = e^x$. We consider constructing the Maclaurin series (Taylor series centered at $x = 0$) for e^x by

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n.$$

Now, note that $f^{(n)}(x) = e^x$ and so $f^{(n)}(0) = 1$ for every n . Hence we have that the Maclaurin series for e^x is

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

It turns out that we have

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

for all real numbers.

It's nice to see a few examples of this construction, so let us consider another example.

Example 7.2.2: Maclaurin Series for $\ln(1 + x)$

Consider the function $f(x) = \ln(1 + x)$. We can build the Maclaurin series for $f(x)$ using

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n.$$

Now we should take a few derivatives of $f(x)$ to see the pattern we need.

$$\begin{aligned} f(x) &= \ln(1+x) \implies f(0) = 0 \\ f'(x) &= \frac{1}{1+x} \implies f'(0) = 1 = 0! \\ f''(x) &= -\frac{1}{(1+x)^2} \implies f''(0) = -1 = -1! \\ f'''(x) &= \frac{2}{(1+x)^3} \implies f'''(0) = 2 = 2! \\ f^{(4)}(x) &= -\frac{6}{(x+1)^4} \implies f^{(4)}(0) = -6 = -3!. \end{aligned}$$

So we have that $f^{(n)}(0) = (-1)^{n-1}(n-1)!$ for $n \geq 1$. So our series takes the form

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}(n-1)!}{n!} x^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n.$$

It turns out here that we have

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(n-1)!}{n!} x^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$$

for $x \in (-1, 1)$.

Exercise 7.2.1. Determine the radius of convergence of the power series for $\ln(1+x)$.

Taylor series granted us the ability to create power series representations of functions. However, not all functions have a useful Taylor series. The prototypical example is the *bump* function

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

If one computes $f^{(n)}(0)$ they find that it is zero for each n . So the Taylor series centered at 0 is the zero series! Even though the function is infinitely differentiable at $x = 0$, it isn't analytic there.

7.3 Operations with Power Series

Since we created power series to essentially be polynomials, we hope to gain some of the nice properties of polynomials as well. Even though the sums are infinite, it turns out that we still get integral and derivative properties like the sum and product rule. So, given a power series representation for a function (where the x are within the radius of convergence)

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

we want to consider

$$\frac{d}{dx} f(x) \quad \text{and} \quad \int f(x) dx.$$

Since the derivative and integral are *linear operators* (more on this later) we know they satisfy the sum and constant multiple rules which leads us to

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

and

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

Let us concentrate first on differentiating the power series. So, we have

$$\begin{aligned} \frac{d}{dx} \sum_{n=0}^{\infty} a_n x^n &= \frac{d}{dx} (a_0 + a_1 x + a_2 x^2 + a_3 x^3 \dots) \\ &= a_1 + 2a_2 x + 3a_3 x^2 + \dots . \end{aligned}$$

Notice now that the zeroth term in the series has been deleted, and what we have is then

$$\boxed{\frac{d}{dx} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=1}^{\infty} n a_n x^{n-1}.}$$

One could rearrange this series and continue starting it at zero. We simply have to make sure that the series remains the same after this process. So, equivalently, one could write

$$\frac{d}{dx} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n.$$

Exercise 7.3.1. Verify that the reindexing above makes sense.

Exercise 7.3.2. How can we compute second derivatives of a power series? What terms do we lose? What about n^{th} derivatives?

And now we turn to integration. Approaching this in the same way we have

$$\begin{aligned} \int \sum_{n=0}^{\infty} a_n x^n dx &= \int (a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots) dx \\ &= C + a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \frac{a_3}{4} x^4 + \dots . \end{aligned}$$

Thus we have that

$$\boxed{\int \sum_{n=0}^{\infty} a_n x^n dx = C + \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1},}$$

where C is the undetermined constant from integration.

Exercise 7.3.3. Try integrating a power series twice.

Example 7.3.1: Differentiating and Integrating e^x

We have the power series representation for e^x given by

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Now, if we consider

$$\begin{aligned} \frac{d}{dx} \sum_{n=0}^{\infty} \frac{x^n}{n!} &= \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n!} \\ &= \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} \\ &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ &= e^x, \end{aligned}$$

which is what we expect. Similarly,

$$\begin{aligned} \int \sum_{n=0}^{\infty} \frac{x^n}{n!} dx &= C + \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)n!} \\ &= C + \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)!}. \end{aligned}$$

Now, this should be equal to $e^x + C$ for some undetermined constant C . So notice that if we replace $C = D + 1$ (which is fine, since C is undetermined), we have

$$\begin{aligned} \int \sum_{n=0}^{\infty} \frac{x^n}{n!} dx &= D + 1 + \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)!} \\ &= D + 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots \\ &= D + \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ &= D + e^x, \end{aligned}$$

which does give us what we expect.

Remark 7.3.1. With integration and differentiation of series, one just has to be very careful. It often helps to write out part of the series (as shown above) to analyze the problem further. This is why we begin with examples of functions we already understand well!

7.4 Approximation with Power Series

A big application of power series is the ability to approximate functions with polynomials. We did indeed begin developing the Taylor series in order to approximate functions in

this way after all. If we have a function $f(x)$ that is N times differentiable, then we can approximate $f(x)$ around the value $x = a$ as

$$f(x) \approx \sum_{n=0}^N \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(N)}(a)}{N!}(x-a)^N$$

which is just a truncation of the Taylor series for f . Of course, we may not have that the power series is totally useful (see the end of Section 6.2), but we also have relaxed the assumption that we need to take infinitely many derivatives of $f(x)$. In fact, it turns out that having two derivatives tends to be plenty as a quadratic approximation works remarkably well.

Definition 7.4.1: Order of an Approximation

Given an approximation of $f(x)$ about the point $x = a$

$$f(x) \approx \sum_{n=0}^N \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

we refer to the ***order of the approximation*** as the highest power of x that appears. In the above case, we would say that this is an N^{th} order approximation of $f(x)$.

It's worth seeing some examples of different orders of approximations to see just how well they fare for different functions.

Example 7.4.1: Approximating e^x

Consider the function e^x which has the Taylor series centered at $a = 0$ given by

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Then we have the approximations

$$\begin{aligned} 0^{\text{th}} : e^x &\approx 1 \\ 1^{\text{st}} : e^x &\approx 1 + x \\ 2^{\text{nd}} : e^x &\approx 1 + x + \frac{x^2}{2}. \end{aligned}$$

Example 7.4.2: How Accurate are Approximations of $\sin(x)$

Consider the function $\sin(x)$ which has the Taylor series centered at $a = 0$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

Then we can consider the approximations

$$\begin{aligned} 0^{\text{th}} : \sin(x) &\approx 0 \\ 1^{\text{st}} : \sin(x) &\approx x \\ 2^{\text{nd}} : \sin(x) &\approx x \\ 3^{\text{rd}} : \sin(x) &\approx x - \frac{x^3}{3!}. \end{aligned}$$

Due to the fact that $\sin(x)$ is an odd function, we have that the even order approximations don't do any better than the previous odd order approximation. We can compare the accuracy of the first and third order approximations to the real function values.

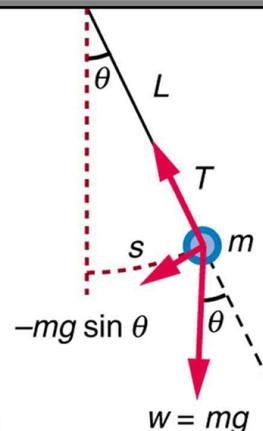
Input	True Value	3^{rd} Order Approx.
x	$\sin(x)$	$x - \frac{x^3}{3!}$
0.001	≈ 0.000999999	0.00099999983
0.01	≈ 0.00999983	0.00999983
0.1	≈ 0.0998334	0.09983
0.5	≈ 0.479426	0.47916
1.0	≈ 0.841471	0.83

Note that the leftmost column is also the first order approximation. One can see that to six significant digits, the third order approximation does a great job even up to $\sin(1)$. The upshot is that the third order approximation is far easier to work with than the function $\sin(x)$ itself (which we tend to use the whole series for).

Since these approximations can do such a great job, it's worthwhile to see how we can use these approximations in practice

Example 7.4.3: The Pendulum Problem

Consider a pendulum system mounted to a frictionless pivot with a bob of mass m a distance L from the pivot. If we pull the bob up by an angle θ , the force of gravity mg is directed downward and thus causes a force of $-mg \sin \theta$ on the bob. We can draw this system out as follows:



Using Newton's second law, we have that the force is equal to the mass times acceleration which leads us to the differential equation

$$mL\theta'' = -mg \sin \theta.$$

This can be well approximated by using the first order approximation for $\sin \theta$. That is, we let

$$\sin \theta \approx \theta$$

to arrive at the differential equation

$$\theta'' = -\frac{g}{L}\theta,$$

which is the harmonic oscillator equation with $\omega^2 = \frac{g}{L}$.

Exercise 7.4.1. What is the solution to the above approximation of the pendulum equation?

Example 7.4.4: Classical Diatomic Vibration

The Morse potential models the attraction between nuclei in a diatomic molecule. Specifically, the potential is given by

$$V(R) = D_e \left[1 - e^{-a(R-R_e)} \right]^2.$$

Here, R is the distance between nuclei, R_e is the equilibrium distance, D_e is the dissociation energy of the molecule, and a is a constant. Stable molecules in low energy states only make small displacements $R - R_e$ from equilibrium and thus we can expand $V(R)$ in a series by

$$\begin{aligned} V(R) &= D_e \left[a^2(R - R_e)^2 - a^3(R - R_e)^3 + \dots \right] \\ &\approx a^2 D_e (R - R_e)^2. \end{aligned}$$

Then, Newton's second law with a potential is given by

$$F = -\frac{dV}{dR}$$

and so

$$F \approx -2a^2 D_e (R - R_e).$$

Letting $x = (R - R_e)$ and $\omega^2 = \frac{2a^2 D_e}{m}$, we have the differential equation

$$x'' = -\omega^2 x$$

which is again the simple harmonic oscillator.

Exercise 7.4.2. Verify that the expansion of $V(R)$ about the point $R = R_e$ above is correct.

7.5 Power Series Solutions to Differential Equations

Power series provided a tool to approximate functions in order to simplify differential equations, but they also provide a way to solve differential equations as well. It is in fact very close to the method of undetermined coefficients. The major difference is we (essentially) have to solve for infinitely many coefficients. We briefly touched on recursive sequences here, and we'll find that these appear here as well. That is to say, if we consider a function given by a power series

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

then the a_n coefficients tend to depend on one another. There are a handful of important functions in physics derived from power series solutions to certain differential equations. For example, we have Bessel functions, Legendre polynomials, and Laguerre polynomials. Bessel functions are, for example, found in solving for the wave modes on a circular drum head. Legendre polynomials are found when solving Laplace's equation in spherical coordinates (we have seen Laplace's equation and will see spherical coordinates in the sequel). Specifically, one sees this when solving for states for the electron in a hydrogen atom. The Laguerre polynomials also arise in quantum mechanics for the radial states of the Hydrogen atom. Not to mention, if one uses the Morse potential in the Schrödinger equation, one will find solutions are forms of Laguerre polynomials.

To begin, it's easiest to consider some more simple examples of power series solutions. The two equations we will work with for now are the population growth or decay equation

$$f'(x) = kf(x)$$

and the harmonic oscillator equation

$$f''(x) = -\omega^2 f(x).$$

As we already know the solutions to this equation, we don't actually need to solve this equations this way. But, it does provide us a few working examples before we move onto

an equation like Legendre's equation. In general, the idea is to take the ansatz that

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

and to take the necessary derivatives of the series and plug it into the given differential equation. From there, one is able to determine the coefficients a_n which determines the function $f(x)$. If we are given an initial value (values if the differential equation is higher than first order), then we can explicitly determine every a_n exactly. Let us see this in action.

Example 7.5.1: Power Series Solution for the Population Model

Consider the initial value problem

$$f'(x) = kf(x)$$

with $f(0) = 1$. Then, we know the particular solution to this initial value problem is

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

which we can find by separation. So, if power series solutions are to work, we should achieve the same particular solution. We take the ansatz that

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

which we can take a derivative of

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

Now, we can plug both into our differential equation

$$\sum_{n=1}^{\infty} n a_n x^{n-1} = k \sum_{n=0}^{\infty} a_n x^n.$$

It's usually easier to write out the terms in the series a bit so that we can match them. So we have

$$a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots = k(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots).$$

From this, we can see that we have

$$a_1 = ka_0$$

$$a_2 = \frac{k}{2} a_1$$

$$a_3 = \frac{k}{3} a_2$$

$$a_4 = \frac{k}{4} a_3$$

\dots

$$a_n = \frac{k}{n} a_{n-1}.$$

Let us also use our initial condition that $f(0) = 1$, which means

$$\begin{aligned} 1 &= f(0) = \sum_{n=0}^{\infty} a_n 0^n \\ &\implies 1 = a_0. \end{aligned}$$

Now, we can plug this into our above expressions for the coefficients which gives us

$$\begin{aligned} a_1 &= k \\ a_2 &= \frac{k^2}{2!} \\ a_3 &= \frac{k^3}{3!} \\ a_4 &= \frac{k^4}{4!} \\ &\vdots \\ a_n &= \frac{k^n}{n!}. \end{aligned}$$

This gives us the particular solution

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{k^n}{n!} x^n \\ &= \sum_{n=0}^{\infty} \frac{(kx)^n}{n!} \\ &= e^{kx}. \end{aligned}$$

Thus we have found exactly what we expected!

Exercise 7.5.1. Carefully go through each step above and work out any details you feel are missing.

Example 7.5.2: Power Series Solution for the Harmonic Oscillator

Consider the differential equation

$$f''(x) = -\omega^2 f(x).$$

We wish to find the general solution via a power series solution. We already know the general solution for this equation is of the form

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

and so we should verify the power series allow us to find the same general solution. So, we have the power series ansatz

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

and thus

$$\begin{aligned} f'(x) &= \sum_{n=1}^{\infty} n a_n x^{n-1} \\ f''(x) &= \sum_{n=2}^{\infty} n(n-1) x^{n-2}. \end{aligned}$$

We then plug in $f(x)$ and $f''(x)$ to the differential equation to get

$$\sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = -\omega^2 \sum_{n=0}^{\infty} a_n x^n.$$

Once again, I believe it is helpful to write out some of the terms of the series

$$2a_2 + 6a_3x + 12a_4x^2 + 20a_5x^3 + \dots = \omega^2(a_0 + a_1x + a_2x^2 + a_3x^3 + \dots).$$

This gives us the equations

$$\begin{aligned} a_2 &= \frac{-\omega^2}{2} a_0 \\ a_3 &= \frac{-\omega^2}{6} a_1 \\ a_4 &= \frac{-\omega^2}{12} a_2 \\ a_5 &= \frac{-\omega^2}{20} a_3. \end{aligned}$$

Then we can determine the even coefficients a_{2n} in terms of a_0 and the odd coefficients a_{2n+1} in terms of a_1 . So we have

$$\begin{array}{ll} a_2 = \frac{-\omega^2}{2} a_0 & a_3 = \frac{-\omega^2}{6} a_1 \\ a_4 = \frac{\omega^4}{24} a_0 & a_5 = \frac{\omega^4}{120} a_1 \\ \vdots & \vdots \\ a_{2n} = \frac{(-1)^n \omega^{2n}}{(2n)!} a_0 & a_{2n+1} = \frac{(-1)^n \omega^{2n}}{(2n+1)!} a_1. \end{array}$$

Hence, the solution to the differential equation can be written as an even power series plus an odd power series as

$$f(x) = a_0 \sum_{n=0}^{\infty} \frac{\omega^{2n}}{(2n)!} x^{2n} + \frac{a_1}{\omega} \sum_{n=0}^{\infty} \frac{\omega^{2n+1}}{(2n+1)!} x^{2n+1}.$$

If we rename $C_1 = a_0$ and $C_2 = \frac{a_1}{\omega}$ then we have

$$\begin{aligned} f(x) &= C_1 \sum_{n=0}^{\infty} \frac{(\omega x)^{2n}}{(2n)!} + C_2 \sum_{n=0}^{\infty} \frac{(\omega x)^{2n+1}}{(2n+1)!} \\ &= C_1 \cos(\omega x) + C_2 \sin(\omega x). \end{aligned}$$

Again, we found exactly what we needed from the power series ansatz. Notice I made the substitution that $C_2 = \frac{a_1}{\omega}$. This made the functions come out to being exactly what we want. This is fine to do since a_1 is undetermined, so $\frac{a_1}{\omega}$ is undetermined as well.

Exercise 7.5.2. Again, verify each step in the above solution and fill in any work you need.

7.6 Orthogonal Polynomials from Power Series Solutions

Throughout physics and chemistry there are equations that arise again and again. A prime example would be the simple harmonic oscillator equation. It would be a reasonable thought to believe that any oscillating system could be approximated by a harmonic oscillator. There are of course other types of systems that will continue to pop their heads out in new ways.

We have solved some boundary value problems in 1-dimensional space, but when moving to higher dimensions, especially, when dealing with a specific geometry such as a rectangle, circle, disk, cylinder, or sphere for example, new equations begin to appear. It seems that the underlying geometry is very important for these boundary value problems. In fact, different geometrical objects may not even have a boundary! Take for example the unit circle which is the set of all points

$$\text{Unit Circle} = \left\{ e^{i\theta} \mid \theta \in [0, 2\pi] \right\}.$$

There is in fact no boundary for this shape. If you imagine walking along a circle, you'll never find the end of it.

7.6.1 Legendre's Equation

Eventually, we will try to understand the quantum mechanical solution to the hydrogen atom problem. In the sequel, we build up the ability to do calculus in multiple dimensions and will solve some partial differential equations in higher dimensions. Blackboxing some notation and terminology for now, the Schrödinger equation that describes the electron orbiting a proton (i.e., a hydrogen atom) has the Hamiltonian

$$H = -\frac{\hbar^2}{2\mu} \nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 r}$$

where r is the radial coordinate in the spherical coordinate system. Recall Schrödinger's equation is then

$$H\Psi(r, \theta, \phi) = E\Psi(r, \theta, \phi),$$

where r, θ, ϕ are the three spherical coordinates. The above equation turns out to be solvable using separation of variables, and the equation for the polar angle $\Theta(\theta)$ turns out to be

$$\frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + A \sin^2 \theta - B = 0.$$

If one takes the change of variables $P(\cos \theta)$ with $x = \cos \theta$ then the equation reads

$$(1 - x^2) \frac{d^2 P}{dx^2} - 2x \frac{dP}{dx} + \left(A - \frac{m^2}{1 - x^2} \right) P = 0.$$

This is the ***associated Legendre's equation***. This equation turns out to be a bit more involved than ***Legendre's equation*** which is given by

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

We will solve Legendre's equation instead.

Solving Legendre's Equation

Legendre's equation arises from boundary value problems. Specifically, we have

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

where we allow x to be in the region $\Omega = [-1, 1]$ and force m to be a non-negative integer $m = 0, 1, 2, 3, \dots$. The fact m takes on just those values is that, for example, other requirements in the other variables present in Schrödinger's equation place a restriction on m . The parameter m being restricted to non-negative integers is once again an example of quantization. If we relate this differential equation to the true physics, this is saying that we have a discrete amount allowed states in the system. We've seen this before with the particle in the 1-dimensional box. We will also require the solutions to this equation for any m to be *normalized*. This is defined below, but in essence it is the requirement that the integral of each of the solution functions (for each m) is finite (and in fact equal to one). This solves the problem uniquely for each m .

Much of the above is extra detail that one will eventually spend more time analyzing. For now, we want to solve this equation and we take the ansatz that $f(x)$ can be written as a power series

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

Then we can take the necessary derivatives

$$\begin{aligned} f'(x) &= \sum_{n=1}^{\infty} n a_n x^{n-1} \\ f''(x) &= \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2}. \end{aligned}$$

We can plug this into Legendre's equation to get

$$\begin{aligned} (1 - x^2)f''(x) - 2xf'(x) + m(m + 1)f(x) &= 0 \\ (1 - x^2) \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} - 2x \sum_{n=1}^{\infty} n a_n x^{n-1} + m(m + 1) \sum_{n=0}^{\infty} a_n x^n &= 0 \\ \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} - \sum_{n=2}^{\infty} n(n-1) a_n x^n - \sum_{n=1}^{\infty} 2na_n x^n + \sum_{n=0}^{\infty} m(m+1) a_n x^n &= 0. \end{aligned}$$

It is helpful to reindex the sums here to have all the powers of x be the same so that we can relate each term slightly easier. So we have

$$\sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n - \sum_{n=2}^{\infty} n(n-1) a_n x^n - \sum_{n=1}^{\infty} 2na_n x^n + \sum_{n=0}^{\infty} m(m+1) a_n x^n = 0.$$

Then if we make all the starting terms agree as well we have

$$\begin{aligned} 0 &= \sum_{n=0}^1 (n+2)(n+1)a_{n+2}x^n - \sum_{n=1}^1 2na_nx^n + \sum_{n=0}^1 m(m+1)a_nx^n \\ &\quad + \sum_{n=2}^{\infty} [(n+2)(n+1)a_{n+2} - n(n-1)a_n - 2na_n + m(m+1)a_n] x^n. \end{aligned}$$

By setting the infinite sum equal to zero, we find a_{n+2} in terms of a_n by

$$a_{n+2} = -\frac{(m-n)(m+n+1)}{(n+2)(n+1)}a_n.$$

The finite sums above must also be zero and so

$$\begin{aligned} 0 &= \sum_{n=0}^1 (n+2)(n+1)a_{n+2}x^n - \sum_{n=1}^1 2na_nx^n + \sum_{n=0}^{\infty} m(m+1)a_nx^n \\ &= 2a_2 + 6a_3x - 2a_1x + m(m+1)a_0 + m(m+1)a_1x \\ &= (2a_2 + m(m+1)a_0) + (6a_3 + (m(m+1)-2)a_1)x \end{aligned}$$

and thus we also have the equations

$$\begin{aligned} a_2 &= \frac{m(m+1)}{2}a_0 & a_3 &= \frac{m(m+1)-2}{6}a_1 \\ &= \frac{m(m+1)}{2!}a_0 & &= \frac{(m-1)(m+2)}{3!}a_1. \end{aligned}$$

We can thus solve for the even terms and odd terms from a_0 and a_1 respectively by

$$\begin{aligned} a_{2n} &= (-1)^n \frac{m(m-2)\cdots(m-2n+2)(m+1)(m+3)\cdots(m+2n-1)}{(2n)!} a_0 & n \geq 1 \\ a_{2n+1} &= (-1)^n \frac{(m-1)(m-3)\cdots(m-2n+1)(m+2)(m+4)\cdots(m+2n)}{(2n+1)!} a_1 & n \geq 1. \end{aligned}$$

So for example, the first few even and first few odd terms are

$$\begin{aligned} a_2 &= -\frac{m(m+1)}{2!}a_0 & a_3 &= -\frac{(m-1)(m+2)}{3!}a_1 \\ a_4 &= -\frac{m(m-2)(m+1)(m+3)}{4!}a_0 & a_5 &= \frac{(m-1)(m-3)(m+2)(m+4)}{5!}a_1. \end{aligned}$$

This splits up our solution into two linearly independent (even and odd) solutions by

$$y_1(x) = \sum_{n=0}^{\infty} a_{2n}x^{2n} \quad \text{and} \quad y_2(x) = \sum_{n=0}^{\infty} a_{2n+1}x^{2n+1}.$$

We then have a polynomial for every choice of m we make. We will denote these polynomials by $f_m(x)$. For example, notice if we take $m = 0$, then we have $a_{2k} = 0$ for $k \geq 1$ and so we have that $y_1(x) = a_0$. For this case $y_2(x)$ will not converge unless $a_1 = 0$. Thus, when $m = 0$ we have the polynomial

$$f_0(x) = a_0.$$

If we take $m = 1$, then we find that $a_{2k+1} = 0$ for $k \geq 1$ and $y_1(x)$ diverges unless $a_0 = 0$. Then we have

$$f_1(x) = a_1 x.$$

We can continue this process to find, for example,

$$f_2(x) = a_0(1 - 3x^2) \quad \text{and} \quad f_3(x) = a_1 \left(x - \frac{5x^3}{3} \right).$$

There is a polynomial for each m , but I have only listed a few here. In general, these polynomials are known as the **Legendre polynomials**

Orthogonality and Normalization

It turns out that the polynomials created from this process are **orthogonal** since

$$\int_{-1}^1 f_i(x) f_j(x) dx = 0$$

when $i \neq j$. Previously we saw a relationship like this when we studied a particle in the 1-dimensional box. The idea is the same. Since we have the undetermined constants a_0 and a_1 present for each polynomial (where each is also attached to a different value of m), we can **normalize** the polynomials above by requiring that

$$\int_{-1}^1 |f_i(x)|^2 dx = 1.$$

In doing this process, we find that the normalize polynomials are

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

One could continue generating normalization constants for each polynomial.

We can then say that this set of normalized polynomials is **orthonormal** which means that

$$\int_{-1}^1 f_i(x) f_j(x) dx = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

It turns out to be orthonormal polynomials that are most useful for us. These polynmomiials correspond to states for a quantum system, and thus they should accurately represent a probability. Hence, the normalization. The fact that the polynomials are also orthogonal is important in quantum mechanics as well, but it is a topic we will revisit in the sequel. In the next part of this text, we discuss linear algebra where we begin to explore the concepts of vectors in more generality. In that sense, these polynomials are vectors in a vector space. It's just that the vector space here is infinite dimensional, and is a bit harder to deal with than the finite dimensional case.

7.7 Quantum Harmonic Oscillator

The utility of power series for solving differential equations is becoming clear. We can now use this technique to study an extremely important example in quantum mechanics. Specifically, the ***Quantum Harmonic Oscillator*** (QHO). As we did previously with the free particle in a 1-dimensional box, we will set up a boundary value problem with Schrödinger's equation as the underlying differential equation. In this case, there will actually be no boundaries for our domain as we will imagine the particle is allowed to be anywhere along the real line \mathbb{R} . That is, the particle could take on any position in the domain $(-\infty, \infty)$. In order to make the system more interesting, we will now set the potential $V(x)$ to be nonzero. Thus, this particle is no longer free as it is being constrained by some type of external force (which is the derivative of the potential).

We take $V(x) = \frac{1}{2}kx^2$ which is the same potential we have seen for a spring on a mass and it is found via Hooke's law. Recall the Hamiltonian is then

$$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2}kx^2.$$

Recall as well that we write the independent Schrödinger equation as $H\Psi = E\Psi$ and for the QHO we have

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2}kx^2 \right) \Psi(x) = E\Psi(x), \quad (7.7.1)$$

where $\Psi(x)$ is the wavefunction. To simplify our notation, we will let $\omega = \sqrt{\frac{k}{m}}$ and the change of variables $y = \sqrt{\frac{m\omega}{\hbar}}x$. Our equation then reads

$$\left(-\frac{d^2}{dy^2} + y^2 \right) \Psi(y) = \frac{2E}{\hbar\omega} \Psi(y).$$

We can rearrange this equation as

$$\frac{d^2\Psi}{dy^2} + \left(\frac{2E}{\hbar\omega} - y^2 \right) \Psi(y) = 0.$$

To start understanding the form of our solution, we look at the behavior of the solution for large y values. As y gets large, the term $\frac{2E}{\hbar\omega}$ is negligible. What we can now do falls under the name of ***asymptotic analysis***. First, we can note that for larger y values, we have

$$\frac{2E}{\hbar\omega} - y^2 \approx -y^2.$$

Thus we have the approximate equation

$$\frac{d^2\phi}{dy^2} - y^2\phi = 0.$$

If we then take $\phi(y) = Ae^{-\frac{y^2}{2}}$ then we can note

$$\frac{d\phi}{dy} = y\phi.$$

Likewise,

$$\frac{d^2\phi}{dy^2} = (-1 + y^2)\phi \approx y^2\phi.$$

This analysis leads us to an ansatz $\Psi(y) = f(y)\phi(y)$ for the QHO equation. This technique is very similar to how we derived the integrating factor technique. Often times we seek to deconstruct solutions, and asymptotic analysis is just one way of doing that.

Taking this ansatz, we can plug this into the QHO equation to yield a new differential equation for the function $f(y)$

$$e^{-\frac{y^2}{2}} \left(\frac{d^2f}{dy^2} - 2y \frac{df}{dy} + \left(\frac{2E}{\hbar\omega} - 1 \right) f \right) = 0.$$

Exercise 7.7.1. Fill in the missing steps to arrive at the above equation. That is, take the ansatz given and plug it into the QHO differential equation (Equation 7.7.1) and simplify until you reach this result.

Notice now that we can divide by $e^{-\frac{y^2}{2}}$ to get the equation

$$\frac{d^2f}{dy^2} - 2y \frac{df}{dy} + \left(\frac{2E}{\hbar\omega} - 1 \right) f = 0. \quad (7.7.2)$$

With this equation, we can now assume a power series ansatz for $f(y)$ so that $f(y) = \sum_{n=0}^{\infty} a_n y^n$. Taking the necessary derivatives of this power series and plugging into Equation 7.7.2, we arrive at

$$\sum_{n=2}^{\infty} n(n-1)a_n y^{n-2} + \sum_{n=1}^{\infty} 2na_n y^n + \left(\frac{2E}{\hbar\omega} - 1 \right) \sum_{n=0}^{\infty} a_n y^n = 0.$$

Simplifying further, we get

$$\sum_{n=0}^{\infty} \left[(n+2)(n+1)a_{n+2} + \left(\frac{2E}{\hbar\omega} - 1 - 2n \right) a_n \right] y^n = 0.$$

Exercise 7.7.2. Work through the missing steps to achieve the simplified expression above.

Since all coefficients must be zero for the above equation to be true, we get the following relationship for the coefficients

$$a_{n+2} = \frac{2n+1 - \frac{2E}{\hbar\omega}}{(n+2)(n+1)} a_n.$$

Though I will not show this here, we require that only finitely many of these a_n terms be nonzero. The reason why has to do with requiring that $\Psi(x)$ represent a valid wave function. In particular, we must have

$$\int_{-\infty}^{\infty} \Psi^2(x) dx = 1.$$

In order for these a_n terms to eventually be zero, we need the numerator for the relationship for a_{n+2} in terms of a_n to be zero and so

$$2n+1 - \frac{2E}{\hbar\omega} = 0,$$

which leads to determining the energy eigenvalues

$$E = \left(n + \frac{1}{2} \right) \hbar\omega.$$

Amazingly, the energy eigenvalues grow linearly! So each state will have an energy level $\hbar\omega$ greater than the previous.

Going back to our original variable, this gives rise to the states

$$\psi_n(x) = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{2}} \frac{1}{\sqrt{2^n n!}} H_n \left(\left(\frac{m\omega}{\hbar} \right)^{\frac{1}{2}} x \right) e^{-\frac{m\omega x^2}{2\hbar}},$$

where $H_n(x)$ are the **Hermite polynomials**. Specifically, we can write the Hermite polynomials as

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} (e^{-x^2}).$$

So for every n , we have a state with a corresponding energy level. Using the Hermite polynomials, the first few are

$$\begin{array}{lll} n = 0 & E = \frac{\hbar\omega}{2} & \psi_0(x) = a_0 e^{-\frac{m\omega x^2}{2\hbar}} \\ n = 1 & E = \frac{3\hbar\omega}{2} & \psi_1(x) = a_0 \left(2\sqrt{\frac{m\omega}{\hbar}} x \right) e^{-\frac{m\omega x^2}{2\hbar}} \\ n = 2 & E = \frac{5\hbar\omega}{2} & \psi_2(x) = a_0 \left(4\frac{m\omega}{\hbar} x^2 - 2 \right) e^{-\frac{m\omega x^2}{2\hbar}} \end{array}$$

As with the particle in the box, we can determine a_0 via the normalization process by requiring

$$1 = \int_{-\infty}^{\infty} \psi_n^2(x) dx,$$

to get

$$a_0 = \left(\frac{m\omega}{\pi\hbar} \right)^{\frac{1}{4}}.$$

Finally, it should again be recognized that the states $\psi_n(x)$ are mutually orthogonal when integrated together over the whole domain. Specifically, we have

$$\int_{-\infty}^{\infty} \psi_m(x) \psi_n(x) dx = \delta_{mn}.$$

7.7.1 Quantum tunneling

One fundamental difference between the solutions to the QHO and the particle in a box is readily apparent. Let us examine the ground state $\psi_0(x)$ in relation to the potential $V(x)$. Briefly letting $m = k = \hbar = 1$, we can plot both to yield

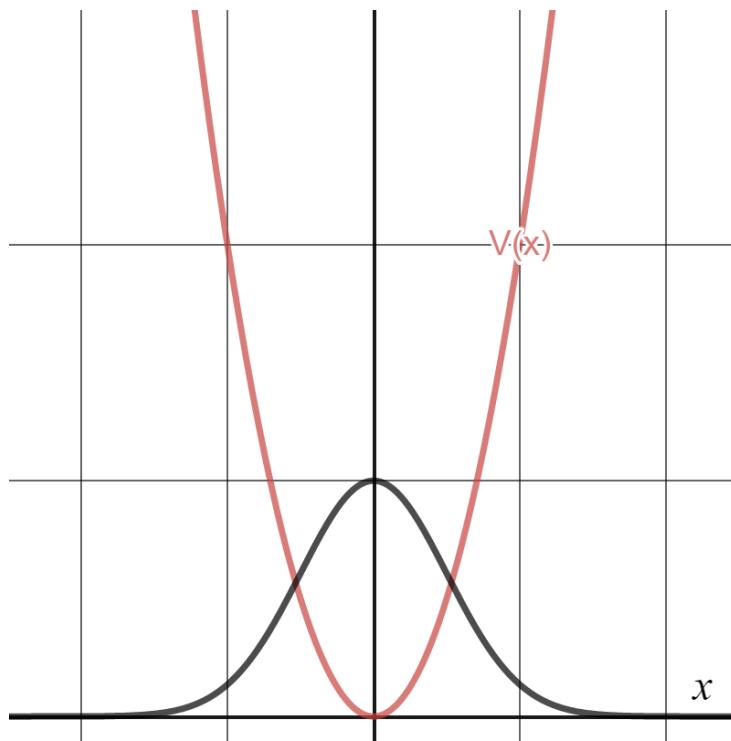


Figure 7.3: Ground state and potential for the QHO.

For a classical particle with energy E , we can never observe the particle outside the region of x where $\frac{1}{2}kx^2 = E$. That is, we must always have the particle between $[-\sqrt{\frac{2E}{k}}, \sqrt{\frac{2E}{k}}]$. This is due to the fact that the particle does not have enough energy to move to any further out x values. However, looking at the graph above, one sees that there is area under the curve outside the classically allowed region. Indeed, there is even a nonzero area under the curve no matter how far out in x values you look! What this means is that we could possibly observe that a particle passes through this potential energy barrier. This phenomenon is known as ***quantum tunneling***. This is yet another fascinatingly non-classical result we find when we study quantum mechanics. It is deeply important!

Problems

Problem 7.1. Complete all the exercises in this chapter.

Problem 7.2. Often we wish to think about functions being represented by series. For example, we can consider the function

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

where $n!$ is read as “ n -factorial” and

$$n! = n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1.$$

Then $1! = 1$ and we define $0! = 1$ as well.

- (a) Consider $f(1)$. Use a tool like WolframAlpha to compute the series

$$f(1) = \sum_{n=0}^{\infty} \frac{1}{n!}$$

- (b) For any value of x , this series converges. So this defines a function on all real numbers. In fact, the series converges even for complex numbers. Simplify the series into its real and imaginary parts. Note,

$$f(ix) = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!}.$$

- (c) We can take derivatives of the function $f(x)$ by differentiating the series *term by term*. That is,

$$\frac{d}{dx} f(x) = \frac{d}{dx} \sum_{n=0}^{\infty} \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{d}{dx} \left(\frac{x^n}{n!} \right).$$

- (d) Show that $\frac{d}{dx} f(x) = f'(x)$.

- (e) What is your guess for what function $f(x)$ is?

Problem 7.3. Find the radius of convergence for the following power series.

(a) $\sum_{n=1}^{\infty} \frac{x^n}{n};$

(b) $\sum_{n=1}^{\infty} \frac{x^n}{n^2};$

(c) $\sum_{n=0}^{\infty} (-1)^n x^n;$

- (d) Repeat (c) but for $z \in \mathbb{C}$ instead of $x \in \mathbb{R}$.

- (e) How does the convergence in (a) and (b) compare with the convergence of the typical p -series?

Problem 7.4. A *Taylor series centered at a* is a power series for a function f is computed by

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

When $a = 0$, we call this a *Maclaurin series*.

- (a) Compute the Maclaurin series for e^x .
- (b) Compute the Taylor series for e^x centered at $a = 1$.
- (c) Compute the Maclaurin series for $\cos(x)$.
- (d) Compute the Maclaurin series for $\sin(x)$.

Problem 7.5. Consider the recursive sequence generated by $a_n = \frac{a_{n-1}}{n}$ with $a_0 = 1$. We can generate a power series for e^x we have seen before from this data

$$e^x = \sum_{n=0}^{\infty} a_n x^n$$

from this data.

- (a) Write down the first five terms of the sequence $\{a_n\}_{n=0}^{\infty}$ to show that it is the same as the sequence $\left\{\frac{1}{n!}\right\}_{n=0}^{\infty}$.
- (b) Using some test for series convergence, show that the power series for e^x converges for any $x \in \mathbb{R}$. This defines the function e^x on all real numbers.

Problem 7.6. How might you arrive at the above recursive sequential definition for e^x ? Consider defining e^x to be the solution to the initial value problem

$$\frac{d}{dx} e^x = e^x \quad \text{and} \quad e^0 = 1.$$

We have seen that we can differentiate a power series term by term and this will help us here.

- (a) Let $e^x = \sum_{n=0}^{\infty} a_n x^n$ where we are pretending (for the moment) that we don't know what a_n is. Using the series definition of e^x , write out the differential equation above.
- (b) Now, by matching coefficients on the same powers of x , can you find the relationship on a_n posed in Problem 1?

Problem 7.7. Recall that we can also take e^z with complex numbers $z \in \mathbb{C}$. In the last worksheet, we found we can write

$$\begin{aligned} e^{ix} &= \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \\ &= \cos(x) + i \sin(x). \end{aligned}$$

Using the above fact and work from Problem 1, show that e^z converges for any $z \in \mathbb{C}$.

Problem 7.8. We have differentiated series term by term which will prove to be useful for solving ODEs. However, we can also integrate them term by term which proves useful for computing integrals of functions that have no “closed-form” anti-derivative. Take for example, the *Gaussian function* (or *normal distribution*)

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

This describes a probability distribution on the real line with mean μ and standard deviation σ . The semi-colon notation for $f(x; \mu, \sigma)$ just means μ, σ are parameters (as opposed to variables) for the function.

- (a) Using the power series for e^x , write the power series for the Gaussian $f(x; \mu, \sigma)$.

- (b) Now, suppose we want to find the probability of an event occurring between $\mu \pm \sigma$ (that is, within one standard deviation of the mean). For the sake of ease, let $\mu = 0$ and $\sigma = 1$ and integrate the series term by term.
- (c) Using the anti-derivative you made in (b), compute an approximation to the definite integral on the region $[-1, 1]$ which will tell us the probability of an event occurring within one standard deviation of the mean.
- (d) Look at a z -score table to see how close your approximation is.

Problem 7.9 (Euler's Formula). Given that

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} \quad \text{and} \quad \sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!},$$

- (a) Plot the approximations of both $\cos(x)$ and $\sin(x)$ versus the original function for order 1,5,20 over the domain $[-4\pi, 4\pi]$.

- (b) Show that

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

using the power series representation for the exponential function e^x .

- (c) Show that cosine is even, $\cos(-x) = \cos(x)$, and that sine is odd $\sin(-x) = -\sin(x)$.
- (d) Compute $\frac{d}{dx} e^{ix}$ using the series representation and show that $\frac{d}{dx} \cos(x) = -\sin(x)$ and $\frac{d}{dx} \sin(x) = \cos(x)$.

Problem 7.10. p -series are actually related to a very important function called the *Riemann zeta function*. This function is involved in a million dollar math problem! If you're interested in other million dollar problems, look up the Clay Institute Millennium Problems. The Riemann zeta function is given by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

- (a) Use the integral test to show that the p -series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

converges. Look up what this series converges to and write it down. This is $\zeta(2)$.

- (b) Use the comparison test to show that the p -series

$$\sum_{n=1}^{\infty} \frac{1}{n^3}$$

converges. This converges as well to $\zeta(3)$. Look up what this approximate value is.

Problem 7.11. Find the radius of convergence for the following power series

$$(a) \sum_{n=1}^{\infty} \frac{x^n}{n(n+1)};$$

$$(b) \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}.$$

Problem 7.12. Consider the function

$$f(x) = \frac{1}{1-x}.$$

- (a) Compute the Taylor series centered at $a = 0$ for the function.
- (b) Find the antiderivative $\int \frac{dx}{1-x}$ using the Taylor series for $f(x)$ found in (a).
- (c) Write down the Taylor series for $\ln(1-x)$ centered at $a = 0$ and compare to your answer in (b).

Problem 7.13.

- (a) Compute the Taylor series centered at $a = 0$ for $f(x) = e^{-\frac{x^2}{2}}$.
- (b) Use the Taylor series for e^x and modify it to find a power series for $f(x)$. Is this the same as the series in (a)?
- (c) Plot the original function $f(x)$ compared to the first, second, third, and fourth term approximation for the series on the same graph.

Problem 7.14. How can we approximate a (possibly complicated) function by using a power series? Why is this useful (specifically for computation on a computer)?

Problem 7.15. Consider the function

$$f(x) = \sqrt{x}.$$

- (a) **(2 pts.)** Find the second order Taylor approximation to $f(x)$ at the point $x = 1$.
- (b) **(2 pts.)** Consider the nonlinear 2nd order autonomous equation

$$x'' + x' + \sqrt{x} = 0.$$

How could we make an approximate ODE using Taylor approximations?

Problem 7.16. Consider the differential equation

$$f'(x) = \frac{1}{\sqrt{1-x^2}} f(x).$$

- (a) Write down the 2nd order Taylor approximation to $\frac{1}{\sqrt{1-x^2}}$ centered at zero.
- (b) Using this second order approximation, find the general solution to the differential equation using separation.
- (c) The solution you find using the approximation doesn't have an issue at $x = 1$, but I claim the original equation does. What is wrong at $x = 1$? Our approximation is then only reasonable in the window $[0, 1)$ (and really isn't that accurate near 1 either).

Problem 7.17. Consider the differential equation

$$f'(x) = xf(x)$$

with initial condition $f(0) = 1$.

- (a) Find the particular solution to this separable differential equation.
- (b) What is the Taylor series centered at zero for this solution?
- (c) Now, assume that the solution $f(x)$ can be written as a power series

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

Determine all of the coefficients a_n which will give us the power series representation for $f(x)$. *Hint: use your solution from (a) to help you.*

Problem 7.18. Consider the differential equation

$$(x - 1)f'(x) + f(x) = 0$$

with initial condition $f(0) = 1$.

- (a) Find the solution to this equation using separation.
- (b) Find the Taylor series centered at zero for your solution in (a).
- (c) Again, suppose that the solution can be written as a power series and determine all the coefficients a_n so that we find the power series representation for $f(x)$. *Hint: use your solution from (a) to help you.*

Problem 7.19. We derived two linearly independent (even and odd) solutions to *Legendre's equation*

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

which were

$$f(x) = \sum_{n=0}^{\infty} a_{2n} x^{2n} \quad \text{and} \quad f(x) = \sum_{n=0}^{\infty} a_{2n+1} x^{2n+1}.$$

- (a) Look up where this equation shows up in quantum mechanics and write it down.
- (b) If we add boundary conditions then we get a finite polynomial for each choice of $\alpha = 0, 1, 2, 3, \dots$. Using this, the first four polynomials are

$$\begin{array}{ll} f_0(x) = 1 & f_1(x) = x \\ f_2(x) = 1 - 3x^2 & f_3(x) = x - \frac{5x^3}{3}. \end{array}$$

Show that these above polynomials are *orthogonal* by showing

$$\int_{-1}^1 f_i(x) f_j(x) dx = 0$$

when $i \neq j$.

Problem 7.20. Bessel functions can arise as the vibrational states of disk shaped objects (e.g., ripples that form in drum heads and cymbals when struck), or as electromagnetic waves in a cylindrical waveguide, or in diffraction from DNA. The *Bessel equation of order zero* is given by

$$x^2 f'' + x f' + x^2 f = 0,$$

where f is a function of x .

These parts will walk you through finding a solution to the equation. I give you the answer in part (a) (which you will check is a solution), but you can (and should) use this to help guide you through the remaining parts. *The trick for this problem is to be careful and to write out the first few terms of series when needed. Wolfram Alpha can help you write partial sums.*

- (a) Show that the power series

$$f(x) = a_0 \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{2n}(n!)^2} x^{2n},$$

is a general solution to the Bessel equation of order zero.

- (b) Take a power series ansatz for $f(x)$. Then, using that ansatz in the ODE, show that you arrive at the condition

$$a_1 x + \sum_{n=2}^{\infty} [n^2 a_n + a_{n-2}] x^n = 0.$$

- (c) Argue now that all the odd terms of the series for your solution must be odd. Do not use the given answer in (a) as your reasoning!
- (d) Determine a general form for a_{2n} and show that you get the same solution as in part (a).

Problem 7.21. Consider the two example systems from quantum mechanics. First, for a particle in a box of length 1 we have the equation

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

with boundary conditions $\Psi(0) = 0$ and $\Psi(1) = 0$.

Second, the Quantum Harmonic Oscillator (QHO)

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2} kx^2 \right) \Psi = E\Psi$$

- (a) Write down the states for both systems. What are their similarities and differences?
- (b) Write down the energy eigenvalues for both systems. What are their similarities and differences?
- (c) Plot the first three states of the QHO along with the potential for the system.
- (d) Explain why you can observe a particle outside of the “classically allowed region”. *Hint: you can use any state and compute an integral to determine a probability of a particle being in a given region.*

Problem 7.22.

- (a) **(3 pts.)** Give a definition of a sequence and explain the difference between a sequence and series. (*This does not need to be overly precise!*)
- (b) **(3 pts.)** Give a definition for a sequence to converge to a limit L . How can you use this to give a definition for a series to converge? (*Again, this does not need to be overly precise!*)
- (c) **(2 pts.)** Explain why the following statement is true or false. To show the statement is true, provide a sound argument. To show a statement is false, you need only provide a counter example.

If the sequence $a_n \rightarrow 0$, then $\sum_{n=1}^{\infty} a_n$ converges.

Problem 7.23. Note that the power series for $\sin(x)$ is given by

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

- (a) **(2 pts.)** To third order, we have

$$\sin(x) \approx x - \frac{x^3}{6}.$$

Approximate $\sin(\pi)$ and compare to the true answer ($\sin(\pi) = 0$). *Hint: you can assume $\pi \approx 3$.*

- (b) **(2 pts.)** Write down the power series for $\sin(3x^3)$.

Problem 7.24.

- (a) **(2 pts.)** Consider the recursive sequence

$$a_n = \frac{1}{2}a_{n-1}$$

with the base case $a_0 = 1$. Show that the formula for a_n is

$$a_n = \frac{1}{2^n}.$$

- (b) **(2 pts.)** Consider the power series with these coefficients

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \frac{x^n}{2^n}.$$

Argue that

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

Hint: can you think of this as a geometric series?

- (c) **(2 pts.)** Compute the derivative of $f(x)$ using the power series representation.

Part III

Finite Dimensional Linear Algebra



Vectors and Vector Spaces

Introduction

Often times a single number is not adequate for describing a quantity. Take for example, position and velocity. In order to describe the position or velocity of a particle, we need to know where in space the particle lies as well as the direction in space the particle travels along with its speed. This takes three numbers to describe (since we move in a 3-dimensional space). Compare this to temperature. At any given point in space, we can describe the temperature with a single number.

The quantity being described above is called a vector. Formally, a vector is an object that can be scaled and added to other vectors of the same type. This allows one to think of more abstract objects as vectors. For example, we can think of solutions to Schrödinger's equation (which are functions) as being vectors in their own type of vector space.

Another way to think of a vector is as a list of numbers. Computer scientists tend to think this way, and we will adopt this point of view for this part of the class. The vector space for solutions to differential equations behaves slightly differently. Here, we cover only finite dimensional spaces which is the key difference between the point of view that a vector is a list of numbers versus the view that vectors satisfy some rules of addition and scaling.

The study of vectors, the spaces they reside in, and how they transform falls under the name *linear algebra*. Whether you have been told or not, linear algebra has dictated much of the mathematics you have learned. This is due to the fact that computations in linear algebra are generally very tractable. We saw examples of linear differential equations previously. One then saw approximation of functions via Taylor series. Fundamentally, linear algebra underlies both of these subjects. Together, one could turn any non-linear differential equation, into a linear one in which we could solve using a power series.

Linear algebra itself has a huge amount of application. It is widely used in computa-

tion and graphics. One also sees its use in optimization or machine learning. Quantum mechanics is deeply rooted in linear algebra as well. However, aside from spin systems, quantum systems have an underlying vector space which is infinite dimensional! For now, we will stick with the finite dimensional case.

8.1 Vector Spaces

Vector spaces are the worlds in which vectors reside. Of course, before we talk about the spaces these objects live in, we should describe the objects as well. The two important quantities in vectors are the vectors themselves and the numbers that scale them. We'll define these immediately.

Definition 8.1.1: Scalar

A **scalar** is an element of a *field* \mathbb{F} . Typically, we take scalars to be numbers $x \in \mathbb{R}$ or $z \in \mathbb{C}$. In other words, we choose our field \mathbb{F} to be equal to \mathbb{R} or \mathbb{C} .

I won't define what a field is here, since we don't need that definition at all. We will always be working with scalars that are real or complex numbers and we will avoid other scalars. That isn't to say they aren't important, we just don't need them for our applications.

Exercise 8.1.1. If you are interested, look up what a field is and show that \mathbb{R} and \mathbb{C} are fields.

With that in place, we can describe vectors and their associated spaces in a single definition.

Definition 8.1.2: Vectors and Vector Spaces

A **vector space** V over a field \mathbb{F} is a set of elements we call **vectors** that satisfy the following properties. Take vectors $\vec{u}, \vec{v}, \vec{w} \in V$ and scalars $\alpha, \beta \in \mathbb{F}$. Then, we require that $\alpha\vec{v} + \beta\vec{u} \in V$ and

- (i) (Commutativity of vector addition)

$$\vec{u} + \vec{v} = \vec{v} + \vec{u}$$

- (ii) (Associativity of vector addition)

$$\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w};$$

- (iii) (Neutral element) There exists $\vec{0}$ such that

$$\vec{0} + \vec{v} = \vec{v};$$

- (iv) (Inverse element) For each \vec{v} there exists $-\vec{v}$ such that

$$\vec{v} + (-\vec{v}) = \vec{0};$$

(v) (Assosciativity of scalar multiplication) We have

$$\alpha(\beta\vec{v}) = (\alpha\beta)\vec{v};$$

(vi) (Distribution over vector addition) we have

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

(vii) (Distribution over field addition) we have

$$(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v};$$

(viii) (Unit element) There exists $1 \in \mathbb{F}$ such that

$$1\vec{v} = \vec{v}.$$

One can try to remember the above rules by the acronym *CANIADDU*. The first four have to do with the addition operation and the last four have to do with the scalar multiplication.

Remark 8.1.1. It is a common practice to denote vectors as bold symbols or with arrows over the top. Here we use both to make them more distinguished. For example, above we used \vec{u} and \vec{v} to denote a vector. We also tend to use Greek letters to denote scalars. Above we used α and β for scalar elements.

Now, the above definition is very abstract and we will continually revisit it. But, for now, we should think of vectors as being real geometrical and physical quantities. In short, a vector space allows for addition of elements we call vectors and scaling of vectors by numbers we calls scalars.

The geometrical picture of a vector is often most helpful. Typically, we represent a vector \vec{v} as an arrow starting with a tail at the origin 0, and head at the desired point. We often do not distinguish a vector \vec{v} from the point at which the tip lies. Let us consider the following.

Example 8.1.1: Vectors in the Plane \mathbb{R}^2

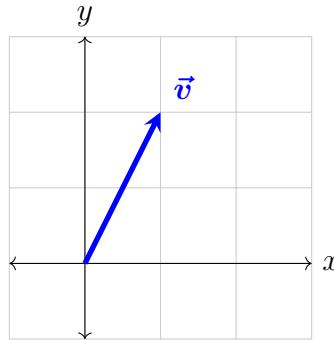
We denote the 2-dimensional real plane by the symbol \mathbb{R}^2 . This is a vector space of 2-dimensions with underlying field the real numbers \mathbb{R} . The naming convention follows that of the 1-dimensional real line \mathbb{R} . The plane simply has two copies of the real line that provide coordinates for points in the plane.

When specifying a point in the plane, we must provide two coordinates. We saw this previously with complex numbers when we had to provide both a real and imaginary part. Here in the plane, we tend to think of providing an x -value and a y -value in a pair (x, y) .

For example, consider the point $(1, 2) \in \mathbb{R}^2$. Then, we can imagine a vector \vec{v} with

tip at $(1, 2)$ and tail at the origin $(0, 0)$ being given by

$$\vec{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$



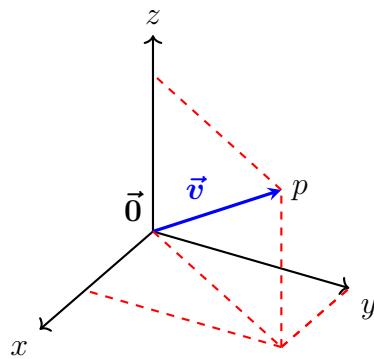
Notice that we are barely distinguishing a vector \vec{v} from the point at which it lies at. We simply are providing a different notation for it, in some sense.

Example 8.1.2: Positions in 3-Dimensional Space

The next step up from planar vectors would be vectors in space. Specifically, when we say space we mean a 3-dimensional space \mathbb{R}^3 . This is the space we live in in our day to day lives. If you are standing in place, you can measure the position of an object and note that it lies at the point $p = (3, 5, 4)$ relative to yourself $\vec{0}$. This position can be described with a vector and the vector has a bit more structure than just points in space. This vector begins at your body, and ends at the object. Moreover, it provides you the oriented line segment exiting your eyes and ending at the object. Similarly to the previous example, we would write

$$\vec{v} = \begin{pmatrix} 3 \\ 5 \\ 4 \end{pmatrix}.$$

We refer to this notation for a vector as a **column vector**. Here, the first entry provides the x -coordinate, the second provides the y -coordinate, and the third provides the z -coordinate. We can picture this vector as follows.



If p moves over time, then our vector \vec{v} changes over time as well. We'll often denote this $\vec{v}(t)$. It's possible the position of an object changes due to interactions with the environment or other objects. All to come later.

Great, we have some objects, but what can we do with them? As it turns out, we can do quite a bit. For the most part, anything we could do with numbers, we can do with vectors. However, we have to be comfortable looking at things in new ways.

8.2 Vector Algebra

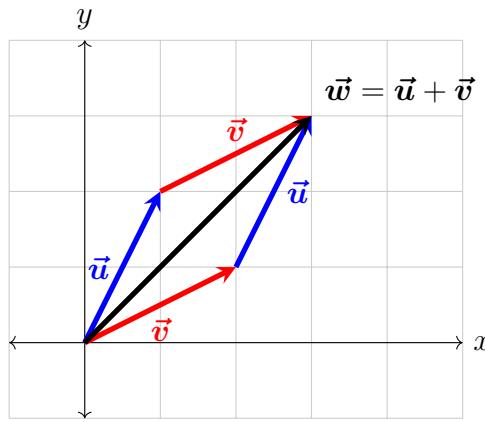
Described in the vector space definition is an operation called ***vector addition***. Given two vectors \vec{u} and \vec{v} , we can create a new vector

$$\vec{w} = \vec{u} + \vec{v}$$

We then required that addition is a commutative operation and so we also have that

$$\vec{w} = \vec{v} + \vec{u}$$

Pictorially, what do we do when we add vectors? We take \vec{u} and attach the tail of \vec{v} to the head of \vec{u} .



From this diagram, you can see why the operation is commutative. Both paths, $\vec{u} + \vec{v}$ and $\vec{v} + \vec{u}$, lead to the same \vec{w} .

As always, repeated addition gives us a form of multiplication. What will this mean here?

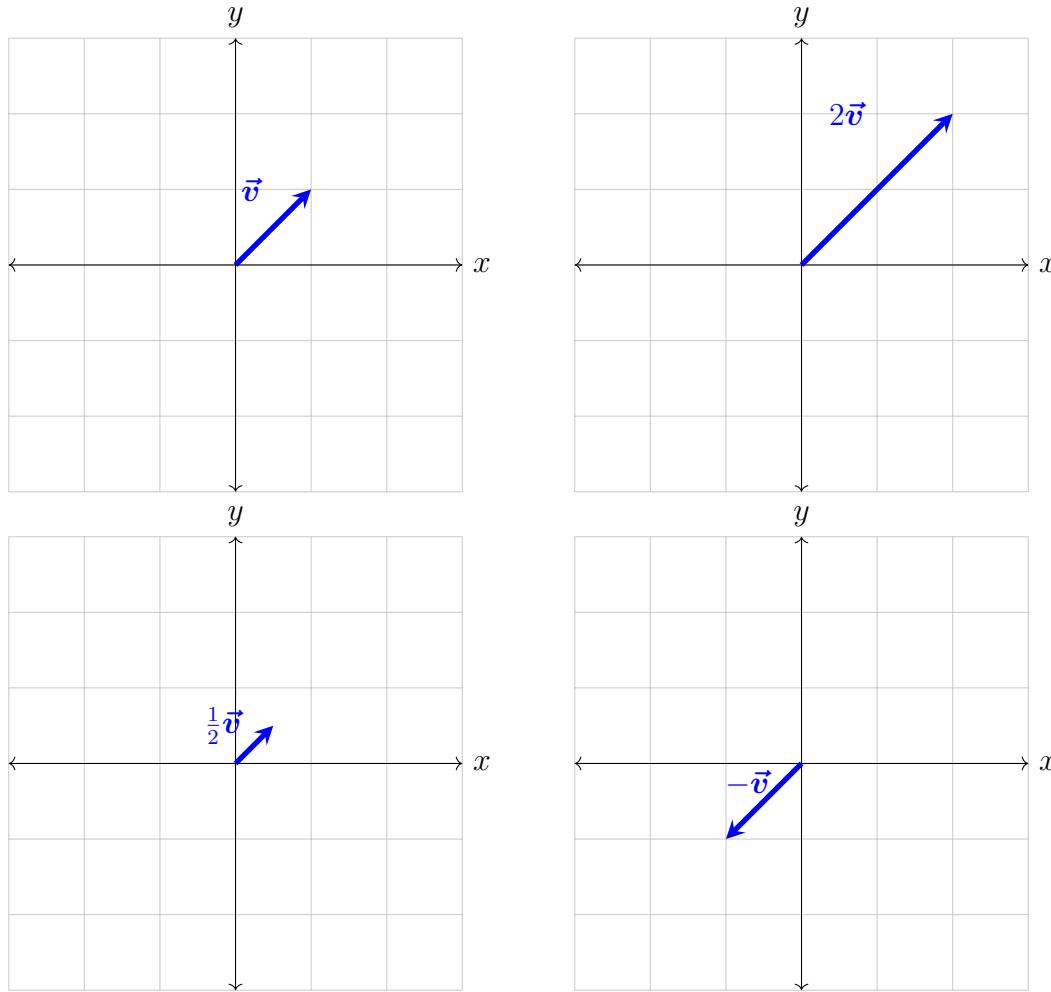
Exercise 8.2.1. Draw a 2-dimensional coordinate system (x and y axes), and draw some vector \vec{v} . Using vector addition, what does $\vec{v} + \vec{v} = 2\vec{v}$ look like? Given this, what do you think $\frac{1}{2}\vec{v}$ will look like?

When dealing with a vector, we are allowed to scale the length of the arrow. As in the definition for a vector space, we call this ***scalar multiplication***. Since the vector $2\vec{v}$ has twice the length of \vec{v} , we would expect $\frac{1}{2}\vec{v}$ to have half the length of \vec{v} . All of these vectors point in the same direction though. We have merely scaled their lengths.

Question 8.2.1. What happens if we take $-\vec{v}$ (i.e., $-1 \cdot \vec{v}$)? Hint: Consider what happens for numbers on a number line when multiplied by -1 .

Answer 8.2.1. It flips the direction of the vector.

We can take a look at all of this.



In the definition for a vector space, we also required that there exists an inverse element for the vector \vec{v} . That is, we wanted to find a vector \vec{u} so that $\vec{v} + \vec{u} = \vec{0}$. Notice that if we take $\vec{u} = -\vec{v}$, then we have $\vec{v} + (-\vec{v}) = \vec{0}$ using the geometrical notion of vector addition shown above. Thus, additive inverses of vectors are just vectors of the same length pointing in the opposite direction. This also gives us a notion of subtracting vectors.

Question 8.2.2. Given two arbitrary vectors \vec{u} and \vec{v} , how can we define $\vec{u} - \vec{v}$?

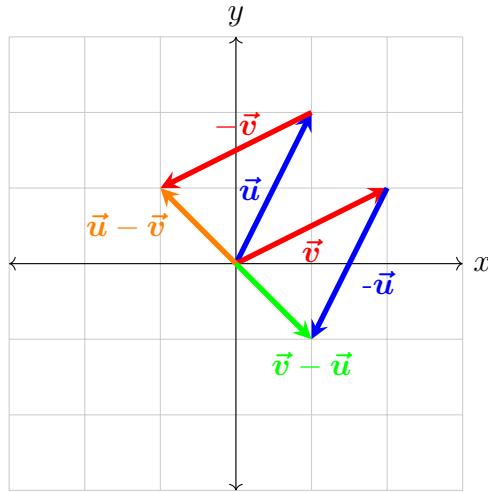
Answer 8.2.2. As we did above for the inverse case. We write $\vec{u} + (-\vec{v})$ and use the rules for vector addition and scalar multiplication together.

Exercise 8.2.2. Draw two different vectors \vec{u} and \vec{v} and draw the following:

- $\vec{u} + \vec{v}$,
- $\vec{u} - \vec{v}$,

- $\vec{v} - \vec{u}$.

Now see the picture below.



8.2.1 Linear Combinations

The most important notion in linear algebra is that of a *linear combination* or *superposition*. Each and every axiom we required in the definition is put there in order to make sense of linear combinations. Given two vectors $\vec{u}, \vec{v} \in V$ and two scalars $\alpha, \beta \in \mathbb{F}$ from the underlying field \mathbb{F} , we call

$$\alpha\vec{u} + \beta\vec{v}$$

a linear combination.

The requirements in our definition allow us to geometrically interpret scalar multiplication and vector addition as we do, and we can string both of these concepts together in one linear combination. Vector addition itself is a linear combination where the scalars are chosen to both be one, and scalar multiplication could just be a linear combination with one of the vectors the zero vector.

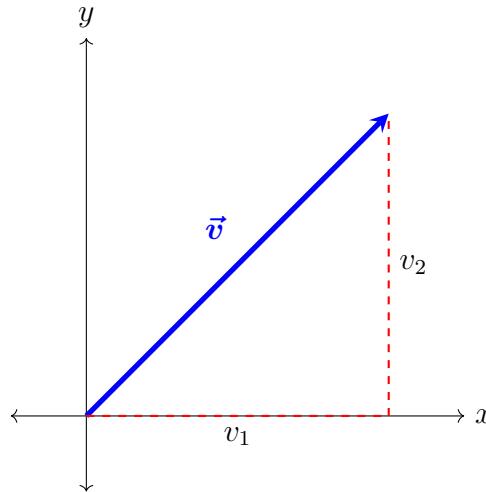
Linear combinations then allow us to understand how to break up vectors and better use coordinates to describe them. Similar to coordinates on Earth (i.e., latitude and longitude) we need numbers to describe a vector. But, we also need to know what these numbers are specifying. We know what latitude and longitude describe on Earth, but these are not the only choice in coordinates we could use!

8.3 Vector Components

The previous work has still been fairly abstract and has left us without a way to explicitly compute quantities! To work with vectors more effectively, it's necessary to break them down into *components*. We can think of components as coming from coordinates on the vector space. Let us consider the following.

Example 8.3.1: Components of a Vector

Let us fix an arbitrary vector \vec{v} in \mathbb{R}^2 (i.e., the xy -plane).



We can break down this vector into the portions that are in the x -direction and y -direction. We say that the component of the vector in the x -direction is v_1 and the component of the vector in the y -direction is v_2 . Using the Pythagorean theorem, it follows that the length of the vector is $\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}$. This is exactly how we defined the modulus of a complex number $\|z\|$.

Hence, we would write

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

In the previous two examples we represented vectors by ordered lists of numbers. This notation is more than adequate, but we do often see vectors presented a different way. When we think of vectors, we tend to think of coordinates presented in the ordered list. However, we will sometimes wish to change coordinates and so we must consider the other form of notation. These coordinates would then give us different components for a vector. We will eventually see this.

8.4 Vector Algebra with Components

Let us work with vectors in space \mathbb{R}^3 . These vectors each have three components, so it suffices to specify \vec{v} as

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

One can then also specify the length $\|\vec{v}\|$ of \vec{v} by writing

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2},$$

which is much like the Pythagorean theorem. If a vector \hat{n} has length one, that is, if $\|\hat{n}\| = 1$, then we say that \hat{n} is a **unit vector**. One may also say that \hat{n} is a **normalized**

vector. We also tend to place hats on unit vectors as opposed to arrows to denote the unit length.

When we write these coordinates for a vector, we are really saying that \vec{v} points an amount v_1 in the x -direction, v_2 in the y -direction, and v_3 in the z -direction. So, if we consider unit vectors that point in these different directions, then we can write \vec{v} as a linear combination of these vectors. So, we set

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \hat{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We call the above vectors the ***unit basis vectors*** since they are of length one and we can write any vector in \mathbb{R}^3 as a linear combination of the three of them. That is, we can write

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

Now, we want to be able to understand vector algebra in these components. Let

$$\vec{u} = u_1 \hat{x} + u_2 \hat{y} + u_3 \hat{z} \quad \text{and} \quad \vec{v} = v_1 \hat{x} + v_2 \hat{y} + v_3 \hat{z},$$

then we can note the following:

- **Equality:** Vectors are equal when their components are equal,

$$\vec{u} = \vec{v} \text{ if } u_1 = v_1, u_2 = v_2 \text{ and } u_3 = v_3.$$

- **Addition:** The sum $\vec{u} + \vec{v}$ is done by adding components together,

$$\vec{u} + \vec{v} = (u_1 + v_1) \hat{x} + (u_2 + v_2) \hat{y} + (u_3 + v_3) \hat{z}.$$

- **Scalar Multiplication:** The product $\alpha \vec{v}$ is obtained by multiplying each component of \vec{v} by α ,

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

Example 8.4.1: Vector Algebra in Space

Consider the vectors

$$\vec{u} = \hat{x} + \hat{y} + \hat{z} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \vec{v} = 2\hat{x} + \hat{y} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \vec{w} = \hat{x} + \hat{z} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Then we can compute

$$\vec{u} + \vec{v} = (\hat{x} + \hat{y} + \hat{z}) + (2\hat{x} + \hat{y}) = 3\hat{x} + 2\hat{y} + \hat{z}.$$

Or, we can add the column vectors

$$\vec{u} + \vec{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Of course, we could also sum all three vectors and find

$$\vec{u} + \vec{v} + \vec{w} = 4\hat{x} + 2\hat{y} + 2\hat{z} = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix}.$$

We can also scale the vectors. Take for example

$$5\vec{u} = 5\hat{x} + 5\hat{y} + 5\hat{z} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}$$

or

$$-\vec{v} = -2\hat{x} - \hat{y} = \begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix}.$$

Then, we could consider a linear combination of the vectors like

$$\begin{aligned} 5\vec{u} - \vec{v} + 2\vec{w} &= (5\hat{x} + 5\hat{y} + 5\hat{z}) + (-2\hat{x} - \hat{y}) + (2\hat{x} + 2\hat{z}) \\ &= 5\hat{x} + 4\hat{y} + 7\hat{z} \\ &= \begin{pmatrix} 5 \\ 4 \\ 7 \end{pmatrix}. \end{aligned}$$

Exercise 8.4.1. Draw the picture for the subtraction $\vec{u} - \vec{v}$.

Exercise 8.4.2. Given $\vec{u} = (2, 3, 1)$, $\vec{v} = (1, -2, 0)$, and $\vec{w} = (5, 2, -1)$, find

- (a) $\vec{d} = 2\vec{u} + 3\vec{v} - \vec{w}$,
- (b) $\|\vec{d}\|$ (the length of \vec{d}),
- (c) Let α be a scalar. Does it make sense to consider $\vec{r} = \vec{u} + \alpha$? Why or why not?

Given a vector space V , we want to find a list of vectors that, by taking a linear combination of said vectors, can form any vector in the vector space V . We call this set of vectors a **basis**. Given a set of vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, we say that the **span** of the vectors is the set of all vectors that can be written as linear combinations of the vectors in the set. Specifically, we are looking for all vectors \vec{u} so that

$$\vec{u} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n.$$

Then we can define the span (which is a set) to be

$$\text{Span}(\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}) = \{\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n \mid \alpha_i \in \mathbb{F}\}.$$

Take for example vectors in the real plane \mathbb{R}^2 . We have that the unit basis vectors \hat{x} and \hat{y} form a basis for \mathbb{R}^2 since any vector \vec{v} can be written as a linear combination as

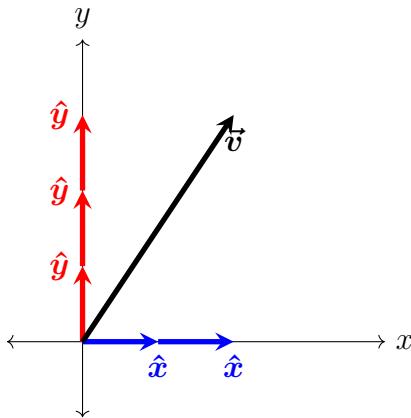
$$\vec{v} = v_1 \hat{x} + v_2 \hat{y}.$$

Similarly, a vector \vec{u} in space \mathbb{R}^3 can be written as a linear combination of the unit basis vectors \hat{x} , \hat{y} , and \hat{z} as

$$\vec{u} = u_1 \hat{x} + u_2 \hat{y} + u_3 \hat{z}.$$

Hence the unit basis vectors for \mathbb{R}^3 are also a basis. One shouldn't have been surprised based on the name we chose!

We can visualize these basis vectors in the plane and make sense of the combinations above. Let us take the vector $\vec{v} = 2\hat{x} + 3\hat{y}$, then we have:



Example 8.4.2: Another Basis for \mathbb{R}^3

Consider again the vectors

$$\vec{u} = \hat{x} + \hat{y} + \hat{z} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \vec{v} = 2\hat{x} + \hat{y} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \vec{w} = \hat{x} + \hat{z} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

then these vectors form a basis for \mathbb{R}^3 .

Question 8.4.1. How do we show that this is a basis?

Answer 8.4.1. We must show that any vector $\vec{r} \in \mathbb{R}^3$ can be written as a linear combination of the three given vectors.

So, let $\vec{r} = r_x \hat{x} + r_y \hat{y} + r_z \hat{z}$, and we have

$$\begin{aligned} \vec{r} &= \alpha_1 \vec{u} + \alpha_2 \vec{v} + \alpha_3 \vec{w} \\ u_1 \hat{x} + u_2 \hat{y} + u_3 \hat{z} &= \alpha_1(\hat{x} + \hat{y} + \hat{z}) + \alpha_2(2\hat{x} + \hat{y}) + \alpha_3(\hat{x} + \hat{z}) \\ u_1 \hat{x} + u_2 \hat{y} + u_3 \hat{z} &= (\alpha_1 + 2\alpha_2 + \alpha_3)\hat{x} + (\alpha_1 + \alpha_2)\hat{y} + (\alpha_1 + \alpha_3)\hat{z}. \end{aligned}$$

This gives us three equations. One for each component of the vector \vec{u} . Specifically, we have

$$\begin{aligned} u_1 &= \alpha_1 + 2\alpha_2 + \alpha_3 \\ u_2 &= \alpha_1 + \alpha_2 \\ u_3 &= \alpha_1 + \alpha_3. \end{aligned}$$

The goal is now to determine whether or not this system of equations can be solved in general!

Now that we have some structure, it's worth seeing one example of an application for vectors. There are plenty more reasons, but with the tools we already have, this example is not too hard to work with. Plus, it is rather fundamental.

Example 8.4.3: Center of Mass

If we have n point masses each with a mass m_i and position \vec{r}_i , then the center of mass can be written

$$\vec{R}_{cm} = \frac{1}{M}(m_1\vec{r}_1 + m_2\vec{r}_2 + \cdots + m_n\vec{r}_n) = \frac{1}{M} \sum_{i=1}^n m_i \vec{r}_i$$

where

$$M = \sum_{i=1}^n m_i$$

is the total mass.

The way to think about this is as averaging the position of these particles but keeping track of how much each weighs. For example, imagine two masses m_1 and m_2 in one dimension. If $\vec{r}_1 = 1$ and $\vec{r}_2 = -1$, then if $m_2 > m_1$, the center of mass should be closer to m_2 and thus negative.

The most important vector space for us will be \mathbb{R}^n where n is a positive integer. \mathbb{R}^2 is familiar to us, as it is usually called the xy -plane. \mathbb{R}^3 is as well, as we usually think of the space surrounding us as being 3-dimensional.

Remark 8.4.1. For us, we will always assume that the tail of vectors starts at $\vec{0}$ (the origin) unless otherwise stated. When we get to vector fields, this will be a bit different.

Example 8.4.4: Examples of Vectors

When the field of scalars are real numbers, we have some example vector spaces. Note, for these, all the scalars v_x, v_y, v_z , or v_i are all real numbers.

- A vector $\vec{v} \in \mathbb{R}^1 = \mathbb{R}$ is just a real number.
- A vector $\vec{v} \in \mathbb{R}^2$ is written as

$$\vec{v} = v_1 \hat{x} + v_2 \hat{y} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

- A vector $\vec{v} \in \mathbb{R}^3$ is written as

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

- In general, a vector $\vec{v} \in \mathbb{R}^n$ is written as

$$\vec{v} = v_1 \hat{e}_1 + v_2 \hat{e}_2 + \cdots + v_n \hat{e}_n = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

where \hat{e}_i is the i th standard unit basis vector for \mathbb{R}^n .

It is possible that the field of scalars is also the complex numbers, in which case we have analogous vector spaces. Here, we tend to not have special notation for two or three dimensional space and we just put

- A vector $\vec{u} \in \mathbb{C}^n$ is written as

$$\vec{u} = u_1 \hat{e}_1 + u_2 \hat{e}_2 + \cdots + u_n \hat{e}_n,$$

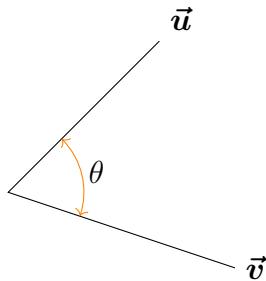
where all the u_i are complex numbers.

8.5 Products of Vectors

Given two vectors, one would like to make products between them in order to find out how related the two vectors are. Also, we should be able to decompose any vectors in a vector space into more fundamental components we understand. We've done this above, but now we would like a way to do this in more generality.

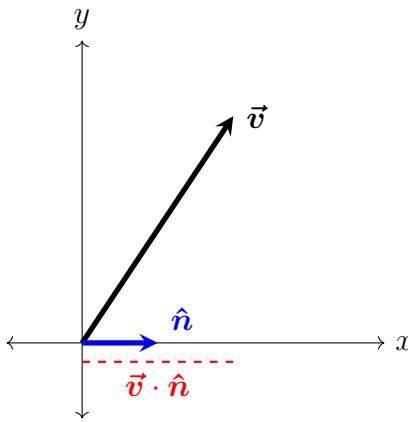
8.5.1 The Dot Product

Question 8.5.1. Given two vectors \vec{u} and \vec{v} based at the same point, can we determine the angle θ between them?



Answer 8.5.1. Yes. In fact, the way we find this gives us a lot more than just an angle. It will be a fundamental concept that we will use often.

Let's consider the following related question first. If we have a vector \vec{v} and a unit vector \hat{n} , how much of the vector \vec{v} lies across the vector \hat{n} ? We will denote this quantity by $\vec{v} \cdot \hat{n}$.) The physical way to think about this is to imagine that we are shining a light perpendicularly to \hat{n} and we measure the length of the shadow that \vec{v} would leave on \hat{n} ? Take for example the following illustration.



Above, one should note that this operation of $\vec{v} \cdot \hat{n}$ is not giving us a new vector, rather it is just providing the length of this shadow being cast.

When the angle is $\theta = 0$, then we expect this quantity $\vec{v} \cdot \hat{n}$ to be maximized as the shadow cast by projecting \vec{v} onto \hat{n} will be as long as \vec{v} is itself. When $\theta = \frac{\pi}{2}$, the quantity will be zero as there will be no shadow cast at all. When $\theta = \pi$, then our quantity would be reversed as to tell us that we would be measuring a shadow pointing the opposite direction of \hat{n} . This is then the negative of the result that we arrived at when $\theta = 0$. Continuing, if $\theta = \frac{3\pi}{2}$, we would get 0 and lastly if $\theta = 2\pi$ this is no different than $\theta = 0$.

What we arrive at with a bit more work is that

$$\vec{v} \cdot \hat{n} = \|\vec{v}\| \cos \theta.$$

We call this the **dot product** or **inner product** of the vectors \vec{v} and \hat{n} . However, we are not required to form this product between a vector and a unit vector. The only difference we see is that the length of the vector we are casting a shadow onto also plays a role in the output of the dot product. Specifically, we have for vectors \vec{v} and \vec{u} that the inner product is given by

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \theta,$$

where θ is again the angle between the two vectors

It turns out that the inner product can be computed in another way. This fact makes the inner product an indispensable tool. For two vectors given by coordinates $\vec{u} = (u_x, u_y, u_z)$ and $\vec{v} = (v_x, v_y, v_z)$ that

$$\boxed{\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z}.$$

Both definitions imply that we have commutivity for this product of vectors. That is,

$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}.$$

They also imply that if we take three vectors, we have

$$(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}.$$

Remark 8.5.1. If we had vectors \vec{u} and \vec{v} that were both elements of \mathbb{R}^n , we would compute the dot product just as above. That is, if

$$\vec{u} = u_1 \hat{e}_1 + u_2 \hat{e}_2 + \cdots + u_n \hat{e}_n \quad \vec{v} = v_1 \hat{e}_1 + v_2 \hat{e}_2 + \cdots + v_n \hat{e}_n,$$

then we have

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^n u_i v_i.$$

The inner product gives us special relationships to vectors. It also gives us a natural notion of direction and lets us compare these directions to one another. Specifically, the inner product lets us see when two vectors point in two perpendicular directions.

Definition 8.5.1: Orthogonal

Two vectors \vec{u} and \vec{v} are *orthogonal* if

$$\vec{u} \cdot \vec{v} = 0.$$

Orthogonal and perpendicular are synonymous.

Example 8.5.1: Unit Basis Vectors

Consider the unit basis vectors in \mathbb{R}^3 ,

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then each of these unit basis vectors are mutually orthogonal to one another. Specifically, if we take

$$\begin{aligned}\hat{x} \cdot \hat{y} &= 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0 \\ \hat{x} \cdot \hat{z} &= 1 \cdot 0 + 0 \cdot 0 + 0 \cdot 1 = 0 \\ \hat{y} \cdot \hat{z} &= 0 \cdot 0 + 1 \cdot 0 + 0 \cdot 1 = 0.\end{aligned}$$

Exercise 8.5.1. Take two vectors $\vec{u} = 2\hat{x} + \hat{y}$ and $\vec{v} = \hat{x} + \hat{y}$. Compute the dot product in both ways. That is

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 = \|\vec{u}\| \|\vec{v}\| \cos \theta.$$

Verify that they give the same answer.

Example 8.5.2: Work by Constant Force

The *work* W done by a constant force \vec{F} on a mass m displaced from position $\vec{u} = u_1\hat{x} + u_2\hat{y} + u_3\hat{z}$ to $\vec{v} = v_1\hat{x} + v_2\hat{y} + v_3\hat{z}$ is given by

$$W = \vec{F} \cdot \vec{d}$$

where

$$\vec{d} = \vec{v} - \vec{u}.$$

Length of a Vector from the Inner Product

The inner product of vectors gives us a natural way to compute the length of a vector. By the way we defined it, we have that the inner product describes the length of a vector

lying along another vector. So, if we allow both vectors input into the inner product to the same vector, we will find this relates to the length. Take a vector $\vec{v} \in \mathbb{R}^2$, then we can compute

$$\vec{v} \cdot \vec{v} = \|\vec{v}\| \|\vec{v}\| \cos \theta,$$

but the angle between the vectors is zero and hence

$$\boxed{\vec{v} \cdot \vec{v} = \|\vec{v}\|^2}.$$

Using the other means of computing the dot product, we have

$$\boxed{\vec{v} \cdot \vec{v} = \|\vec{v}\|^2 = v_1^2 + v_2^2}$$

which is just the Pythagorean theorem! This of course generalizes to 3-dimensions and higher.

The dot product has lead us to one last important definition that we should carry with us throughout the rest of this text and the sequel.

Definition 8.5.2: Orthonormal

Given a set of vectors $\{\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n\}$ in \mathbb{R}^n we say that this set of vectors is **orthonormal** if each vector \hat{e}_i is a unit vector and if each pair of vectors \hat{e}_i and \hat{e}_j are mutually orthogonal. That is, we have

$$\|\hat{e}_i\| = 1 \quad \text{for every } i$$

and

$$\hat{e}_i \cdot \hat{e}_j = 0 \quad \text{if } i \neq j.$$

This can all be succinctly written as

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

We call δ_{ij} the **Kronecker delta symbol**.

Components via the Inner Product

One may also wish to recover components of a vector via the dot product. For example, if one wants to recover the x -component of a vector $\vec{v} \in \mathbb{R}^3$, then one can compute

$$\vec{v} \cdot \hat{x} = v_1 \cdot 1 + v_2 \cdot 0 + v_3 \cdot 0 = v_1.$$

Then we can do this for each component as well by

$$\begin{aligned} \vec{v} \cdot \hat{y} &= v_1 \cdot 0 + v_2 \cdot 1 + v_3 \cdot 0 = v_2 \\ \vec{v} \cdot \hat{z} &= v_1 \cdot 0 + v_2 \cdot 0 + v_3 \cdot 1 = v_3. \end{aligned}$$

8.5.2 The Cross Product

In 3-dimensional space it's possible to define another very special product of vectors. As to why this only works in 3-dimensions, we will see a bit in the next section. Anyways, let's take a look.

We can define a product between vectors as follows. We take two vectors, $\vec{u}, \vec{v} \in \mathbb{R}^3$ and we put $\vec{u} \times \vec{v}$ which outputs a new vectors orthogonal to both \vec{u} and \vec{v} with length

$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\| \|\vec{v}\| \sin \theta,$$

where θ is the angle between the two vectors \vec{u} and \vec{v} . We will also require that

$$\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$$

in order to make this product well defined. All this means is that with this rule, this all work out properly! We will call this vector product the *cross product*. Remember, the cross product only works in 3-dimensions.

Question 8.5.2. With all we've stated above, can you then find what $\vec{u} \times \vec{u}$ is equal to?

Question 8.5.3. Yes. Note that

$$\vec{u} \times \vec{u} = -\vec{u} \times \vec{u}$$

which implies that

$$\vec{u} \times \vec{u} = \vec{0}$$

as the only vector satisfying the relationship above is the $\vec{0}$. We can see that since $-\vec{0} = \vec{0}$.

Computing the Cross Product

If we define the cross product on the basis elements \hat{x} , \hat{y} , and \hat{z} , we will know how to do this with any vector. We define

$$\begin{array}{lll} \hat{x} \times \hat{y} = \hat{z} & \hat{y} \times \hat{z} = \hat{x} & \hat{z} \times \hat{x} = \hat{y} \\ \hat{x} \times \hat{x} = \vec{0} & \hat{y} \times \hat{y} = \vec{0} & \hat{z} \times \hat{z} = \vec{0}. \end{array}$$

The last item we need in order to have the cross product fully defined is to note that we can take the cross product of a linear combination of vectors as follows. We let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$ and $\alpha, \beta, \gamma \in \mathbb{R}$ and we have

$$(\alpha \vec{u} + \beta \vec{v}) \times (\gamma \vec{w}) = \alpha \gamma \vec{u} \times \vec{w} + \beta \gamma \vec{v} \times \vec{w}.$$

One can also note that with three vectors we have

$$(\vec{u} + \vec{v}) \times \vec{w} = \vec{u} \times \vec{w} + \vec{v} \times \vec{w}.$$

With these rules in place, we can now attempt the next exercise.

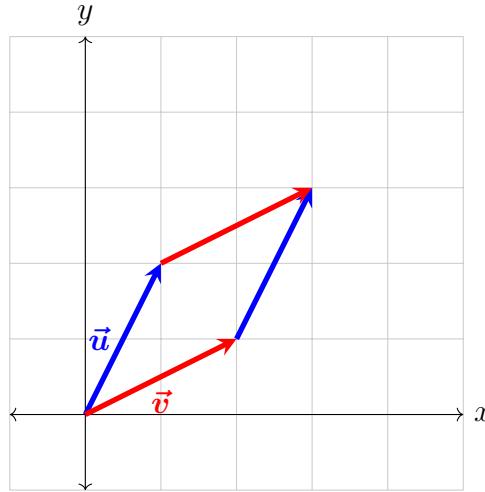
Exercise 8.5.2. Show that for $\vec{u} = u_1 \hat{x} + u_2 \hat{y} + u_3 \hat{z}$ and $\vec{v} = v_1 \hat{x} + v_2 \hat{y} + v_3 \hat{z}$ that we have

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2) \hat{x} + (u_3 v_1 - u_1 v_3) \hat{y} + (u_1 v_2 - u_2 v_1) \hat{z}.$$

The cross product has a nice geometrical interpretation and usefulness. The inner product measures how parallel two vectors are, the cross product measures how perpendicular two vectors are. The cross product also has the added advantage of being a vector quantity as opposed to the scalar quantity we receive as an output from the inner product.

Example 8.5.3: Parallelogram given by two vectors

Given two vectors \vec{u} and \vec{v} we can define a parallelogram. For example, we let $\vec{u} = \hat{x} + 2\hat{y}$ and $\vec{v} = 2\hat{x} + \hat{y}$. The parallelogram looks like:



We can then compute the area of this parallelogram by first computing the cross product $\vec{u} \times \vec{v}$. So we have

$$\begin{aligned}\vec{u} \times \vec{v} &= (\hat{x} + 2\hat{y}) \times (2\hat{x} + \hat{y}) \\ &= 2\hat{x} \times \hat{x} + \hat{x} \times \hat{y} + 4\hat{y} \times \hat{x} + 2\hat{y} \times \hat{y} \\ &= \vec{0} + \hat{z} - 4\hat{z} + \vec{0} \\ &= -3\hat{z}.\end{aligned}$$

Then we have that the area of the parallelogram is

$$A = \|\vec{u} \times \vec{v}\| = \| -3\hat{z} \| = 3.$$

Exercise 8.5.3. Compute the area of the parallelogram defined by $\vec{u} = 3\hat{x} + \hat{y} - \hat{z}$ and $\vec{v} = \hat{x} + 2\hat{y} - 3\hat{z}$.

Example 8.5.4: Angular Velocity and Right-Hand Rule

For a particle moving along a circle with radius r we can define a quantity called the *angular velocity* and denote it by $\vec{\omega}$. Then, for example, the time it takes for the particle to travel around the whole circle (the *period*) is

$$\tau = \frac{2\pi r}{\|\vec{\omega}\|}.$$

It turns out that the angular velocity of this particle at any point is given by

$$\vec{\omega} = \frac{\vec{r} \times \vec{v}}{\|\vec{r}\|^2}.$$

We can also find \vec{v} from $\vec{\omega}$ and \vec{r} . Take a look at the figure below and note the orientation of $\vec{\omega}$ relative to the direction the particle travels around the circle.

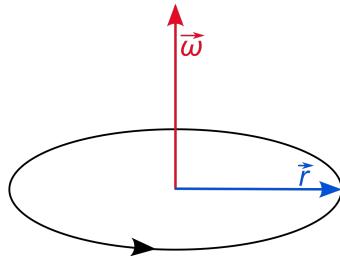


Figure 8.1: A particle moving around a circle of radius $\|\vec{r}\|$ in a counter-clockwise motion.

Note that \vec{v} would be tangent to this circle at the point \vec{r} and pointing along the direction of travel.

It turns out that $\vec{\omega}$ is a vector pointing perpendicularly to the plane that the circle the particle traverses is in. Which way does $\vec{\omega}$ point? We need the *right-hand rule*!

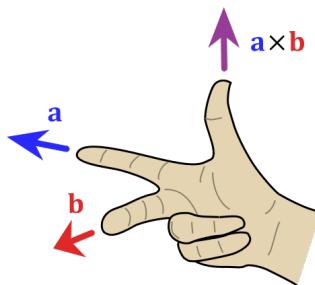


Figure 8.2: The right-hand rule.

To see how this works, we let $\vec{a} = \vec{r}$ be our index finger, then $\vec{b} = \vec{v}$ be our middle finger. The resulting direction of $\vec{r} \times \vec{v} = \vec{\omega}$ is then pointing in the direction we see from Figure 8.1.

9

Linear Transformations and Matrices

9.1 Linear Transformations

Now that we have set the stage for vectors and the products between them, we would like to investigate how we can transform these vectors. Specifically, we will first care about functions that are *linear*. These will be functions that stretch and rotate vectors and possibly change dimension all while leaving the origin alone.

Definition 9.1.1: Linear Transformation

A *linear transformation* is a function

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

that satisfies the following requirements:

- (i) $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$,
- (ii) $T(\alpha\vec{v}) = \alpha T(\vec{v})$,
- (iii) $T(\vec{0}) = \vec{0}$.

Note that requirement (iii) follows from (i) and (ii), but it is something that is easy to check, so we place it here as well.

Remark 9.1.1. When writing transformations of vectors, we often omit the extra paren-

theses. That is, instead of

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

we will just put

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Remark 9.1.2. These rules should seem similar to the properties of the derivative and integral. We'll find that what we're building here will let us properly talk about derivatives in multiple dimensions.

Example 9.1.1: Scaling is Linear

Consider $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha x \\ \beta y \end{pmatrix}.$$

This transformation scales the x -component of our vector by α and scales the y -component by β .

To see that this is linear, we just check that it satisfies the three necessary conditions. First we will show (i). So, if we take two vectors

$$\vec{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix},$$

then we have

$$\begin{aligned} T(\vec{v}_1 + \vec{v}_2) &= T \left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right) \\ &= T \left(\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \right) \\ &= \begin{pmatrix} \alpha(x_1 + x_2) \\ \beta(y_1 + y_2) \end{pmatrix} \\ &= \begin{pmatrix} \alpha x_1 + \alpha x_2 \\ \beta y_1 + \beta y_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha x_1 \\ \beta y_1 \end{pmatrix} + \begin{pmatrix} \alpha x_2 \\ \beta y_2 \end{pmatrix} \\ &= T(\vec{v}_1) + T(\vec{v}_2). \end{aligned}$$

Then for (ii), take $\vec{v} = x\hat{x} + y\hat{y}$ and we have

$$\begin{aligned} T(\mu\vec{v}) &= T\left(\mu \begin{pmatrix} x \\ y \end{pmatrix}\right) \\ &= T\left(\begin{pmatrix} \mu x \\ \mu y \end{pmatrix}\right) \\ &= \begin{pmatrix} \alpha\mu x \\ \beta\mu y \end{pmatrix} \\ &= \mu \begin{pmatrix} \alpha x \\ \beta y \end{pmatrix} \\ &= \mu T(\vec{v}). \end{aligned}$$

Since the qualities (i) and (ii) imply (iii), we don't necessarily need to check it. But, we can anyway. So we take

$$\begin{aligned} T(\vec{0}) &= T\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \\ &= \begin{pmatrix} \alpha \cdot 0 \\ \beta \cdot 0 \end{pmatrix} \\ &= \vec{0}. \end{aligned}$$

So this function T is indeed linear.

Exercise 9.1.1. Which of the following are linear transformations? Why or why not?

- (a) $f: \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = \lambda x$.
- (b) $g: \mathbb{R} \rightarrow \mathbb{R}$ given by $g(x) = 2x + 1$.
- (c) $h: \mathbb{R} \rightarrow \mathbb{R}$ given by $h(x) = x^2$.
- (d) $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ given by

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} y \\ x \\ z \end{pmatrix}.$$

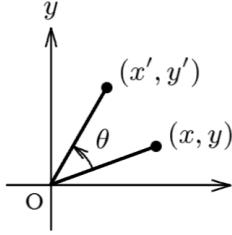
Exercise 9.1.2. Pick a vector in the plane and draw a picture of the scaling transformation.

Example 9.1.2: Rotation is Linear

Consider the following linear transformation $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

This transformation rotates a vector by θ in the counter-clockwise direction.



9.2 Matrix Representation of Linear Transformations

The salient fact of linear transformations is how we can represent them. As it turns out, any linear transformation $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be understood by seeing how the vectors

$$\hat{\mathbf{x}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \hat{\mathbf{x}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \hat{\mathbf{x}}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

This is due to the fact that A is linear and any vector $\vec{u} \in \mathbb{R}^n$ can be written as

$$\vec{u} = u_1 \hat{\mathbf{x}}_1 + u_2 \hat{\mathbf{x}}_2 + \dots + u_n \hat{\mathbf{x}}_n = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix}.$$

Thus, we would have

$$T(\vec{u}) = u_1 T(\hat{\mathbf{x}}_1) + u_2 T(\hat{\mathbf{x}}_2) + \dots + u_n T(\hat{\mathbf{x}}_n).$$

Let us say that

$$T(\hat{\mathbf{x}}_1) = \vec{A}_1, \quad T(\hat{\mathbf{x}}_2) = \vec{A}_2, \dots, T(\hat{\mathbf{x}}_n) = \vec{A}_n,$$

Then we can create a matrix representation for A , denoted by $[A]$, by taking the j th column of this matrix to be the output of $T(\hat{\mathbf{x}}_j)$. Specifically, this gives us

$$[A] = \begin{pmatrix} | & | & | \\ \vec{A}_1 & \vec{A}_2 & \vec{A}_3 \\ | & | & | \end{pmatrix}.$$

Hence, this arbitrary linear transformation is captured entirely by the ***matrix*** of numbers

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix},$$

where we can think of the columns as

$$\vec{A}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad \vec{A}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \quad \vec{A}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Now, our study moves to that of matrices since we have seen that matrices capture all that we need in order to describe linear transformations. They aren't necessary to use, but they make computation and understanding a bit easier. It turns out that we can also think of vectors as special cases of matrices, which makes the idea of studying matrices themselves all that much better. From here on out, we will restrict ourselves to matrix representations.

Often, one may be handed a linear transformation $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ as above. When one wants to represent A as a matrix, we may put

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nm} \end{pmatrix}.$$

However, we will often just assume that a linear transformation is given as a matrix a priori. As one continues to learn more linear algebra, it becomes clear why we wish to make this distinction, but for now this is alright. In the sequel, we will care about understanding *linear operators* which will not always have some matrix representation.

9.2.1 Matrix Algebra

Just as we did with vectors, we want to understand what we can do with these matrices algebraically. These matrices must allow us to perform linear transformations. Given that linear transformations are special types of functions, we also want to be able to compose linear functions as well. All that we need will be captured in an algebraic way through matrix multiplication.

We will call a matrix with n -rows and m -columns an $n \times m$ -matrix (read: n by m matrix). These take the form:

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nm} \end{pmatrix}.$$

We use a capital letter in brackets to denote a matrix. Each entry of the matrix will be given by a lowercase letter with subscripts a_{ij} . The subscripts will tell you which row and column the entry is located. For example, a_{25} would be the entry in the 2nd row and 5th column.

- **Equality:** Matrices are equal if each entry is equal. That is, given $[A]$ and $[B]$, we say $[A] = [B]$ if $a_{ij} = b_{ij}$ for each pair i, j . Clearly if these matrices are not the same “shape” (meaning, they have a different number of rows or columns), then they cannot be the same. Just as a 2-dimensional vector cannot be the same as a 3-dimensional one. They are distinctly different objects.
- **Addition:** We can add matrices of the same shape. We write $[A] + [B] = [C]$ and create the new matrix $[C]$ by adding the entries. That is, $c_{ij} = a_{ij} + b_{ij}$.
- **Scalar Multiplication:** We can also scale matrices. We do this by scaling the entries. So, if we have $\alpha[A] = [B]$, then we know the entries of $[B]$ are given by $b_{ij} = \alpha a_{ij}$.
- **Matrix Multiplication:** It is also possible to multiply two matrices together. Recall that matrices are how we capture the information of a linear transformation. Matrix multiplication will capture the idea of composing two linear transformations. Consider two linear transformations

$$A: \mathbb{R}^n \rightarrow \mathbb{R}^m \quad \text{and} \quad B: \mathbb{R}^m \rightarrow \mathbb{R}^p.$$

Then we have that $[A]$ is an $m \times n$ -matrix and $[B]$ is a $p \times m$ matrix. We can create the composite linear transformation

$$B \circ A: \mathbb{R}^n \rightarrow \mathbb{R}^p,$$

and hence we would like to understand how to represent this composite function with the two matrices $[A]$ and $[B]$. We define the matrix multiplication

$$[B][A] = [C]$$

to properly capture the composite linear transformation $A \circ B$. Hence, we can multiply matrices $[A]$ and $[B]$ if

the number of columns of $[A]$ = the number of rows of $[B]$.

If $[A]$ is an $m \times n$ -matrix and $[B]$ is an $p \times m$ -matrix, then $[C] = [B][A]$ is an $m \times p$ matrix. You can remember this helpful fact:

$$(p \times m) \cdot \underbrace{(m \times n)}_{\text{the same}}$$

and

$$\underbrace{(p \times m)}_{\text{the same}} \cdot (m \times n)$$

gives the dimensions of the resulting matrix.

How we perform this matrix multiplication looks a bit ugly at first, but it ends up being slightly easier after digesting this a bit. We have that the components of $[C] = [B][A]$ are

$$c_{ij} = \sum_{k=1}^m b_{ik} a_{kj}.$$

Let us take the example of letting $[B]$ be an $1 \times n$ -matrix and $[A]$ a $n \times 1$ -matrix. Then

$$[B][A] = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} = b_{11}a_{11} + b_{12}a_{21} + \cdots + b_{1n}a_{n1} = \sum_{k=1}^n b_{1k}a_{k1}.$$

This is the dot product of two vectors! As it turns out, we can decompose matrix multiplication into a bunch of dot products.

In general, if we look at the multiplication described above for the component c_{ij} , we can break this down as follows. Note that the i th row of a matrix $[B]$ is a vector with m elements, the j th column of a matrix $[A]$ is a vector with m elements, and the dot product of these two vectors gives the entry c_{ij} of the matrix $[B][A] = [C]$.

Remark 9.2.1. This all means it is possible to take linear combinations of matrices as well!

Proposition 9.2.1: $n \times m$ -Matrices Form a Vector Space

Let $\mathcal{M}_{n \times m}$ be the set of all $m \times n$ -matrices with real or complex entries. Then this set forms a vector space with the scalar multiplication and matrix addition described above.

Exercise 9.2.1. Show that the above proposition is true.

Example 9.2.1: Multiplying Matrices

Let us multiply the following matrices:

$$[A] = \begin{pmatrix} 2 & 0 & -3 \\ 1 & 1 & -2 \end{pmatrix} \quad [B] = \begin{pmatrix} 2 & 3 & 4 & 1 \\ 1 & 2 & 2 & 0 \\ 0 & -1 & 2 & 0 \end{pmatrix}.$$

Verify that you get

$$[A][B] = \begin{pmatrix} 4 & 9 & 2 & 2 \\ 3 & 7 & 2 & 1 \end{pmatrix}.$$

9.2.2 Properties of Matrix Multiplication

Matrices will behave in the following ways:

- **Associativity:** The order in which you choose to multiply matrices does not matter. That is

$$[A]([B][C]) = ([A][B])[C] = [A][B][C].$$

- **Distributivity:** We can multiply matrices over sums. That is

$$[A]([B] + [C]) = [A][B] + [A][C].$$

- **(non)-Commutativity:** In general, we have

$$[A][B] \neq [B][A].$$

However, there are certain types of matrices that do commute with each other.

Right now we have the ability to write linear transformations as matrices. We can also multiply these matrices. Keep in mind that this in some way mimics composing functions. In essence, we have entirely captured the functional behavior of linear spaces through matrices, since, we can realize vectors as special types of matrices as well.

9.3 Systems of Linear Equations

Often times we are handed a system of equations to solve. In this case, we have more than one variable, and the same number of equations is required in order to determine values for these variables uniquely. For example, one may have three equations for the variables x , y , and z where each equations simultaneously equal zero. That is,

$$\begin{aligned} f(x, y, z) &= 0 \\ g(x, y, z) &= 0 \\ h(x, y, z) &= 0. \end{aligned}$$

In the most general case where f , g , and h are potentially nonlinear, these equations may be very difficult to solve together. In the case that each of the above functions is linear, it is much easier to determine a solution (if one exists).

When the equations are linear, we can write them as a matrix times a vector

$$[A]\vec{x} = \vec{y}$$

where A is an $n \times m$ -matrix, \vec{x} is an m -dimensional vector, and \vec{y} is an n -dimensional vector. In this case, we know the vector \vec{y} , but we wish to determine the correct vector \vec{x} that satisfies the equation.

Let us write out what this looks like:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

gives us a set of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= y_2 \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= y_n. \end{aligned} \quad \vdots$$

We call these equations a ***system of linear equations***.

In the most general case for a system of linear equations, it may be that $n \neq m$, and finding solutions is far from guaranteed. In this case, a system may be ***overdetermined*** if $n > m$. This is the case where there are more equations than there are variables, hence the

name. When a system is overdetermined, there is likely no solution to the problem. The other case where $n < m$ has fewer equations than the amount of unknown variables in the vector \vec{x} , and we call this system ***underdetermined***. When a system is undetermined, we tend to expect a solution, but it won't be unique.

We will tend to concentrate on the case that is properly determined. That is, we will look at systems of equations that have n equations and n unknowns, which means that \vec{x} is a vector of length n , \vec{y} is a vector of length n , and $[A]$ is an $n \times n$ -matrix.

9.4 Solving Linear Systems of Equations

The problem at hand is that we are handed a matrix $[A]$ and an output vector \vec{y} and we are asked to find a vector \vec{x} such that the equation

$$[A]\vec{x} = \vec{y}$$

is satisfied. If \vec{y} is equal to the zero vector $\vec{0}$, then we call this a ***homogeneous*** linear system. Otherwise, we call the system ***inhomogeneous***.

To solve these equations requires an algorithmic approach. The algorithm we will use is known as ***row reduction***. Row reduction is a list of a few rules we are allowed to use in order to modify a matrix and determine a solution to either a homogeneous or inhomogeneous equation.

Definition 9.4.1: Row Operations

We call the following list of operations the ***elementary row operations***.

- **Row scaling:** We can scale the rows of a matrix by any scalar value α .
- **Row addition:** We can add scalar multiples of any rows to another row.
- **Row swapping:** We can swap any two rows.

These operations are elementary in the sense that they do not change the “character” of the matrix. This means that these will not affect the solution to our system. Also, the operations will make determining the solution doable as well.

When given a system of linear equations, it becomes handy to shorten what we have to write and place the whole system into a matrix. However, when we do this, we want to keep separate the matrix $[A]$ from the solution vector \vec{y} . In this matrix, the vector we are solving for, \vec{x} , will not appear. But, when we finish the algorithm, we will have determined \vec{x} .

Definition 9.4.2: Augmented Matrix

Given an equation $[A]\vec{x} = \vec{y}$, we have the ***augmented matrix*** $[M]$. $[M]$ is a matrix with one extra column than $[A]$, and it is created by letting $[A]$ fill the left most columns, and \vec{y} take the spot in the right most column. For example, for a

3×3 -matrix $[A]$, and 3-dimensional vector \vec{y} , we have:

$$[M] = \left(\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & y_1 \\ a_{21} & a_{22} & a_{23} & y_2 \\ a_{31} & a_{32} & a_{33} & y_3 \end{array} \right).$$

We often use the vertical line to distinguish the output vector \vec{y} column from the matrix $[A]$ inside of $[M]$.

Our goal here is to use elementary row operations to reduce our augmented matrix to the following *row reduced echelon form*. For the example shown in the definition above, this will look like:

$$[M] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & x_1 \\ 0 & 1 & 0 & x_2 \\ 0 & 0 & 1 & x_3 \end{array} \right]$$

When we have reduced the augmented matrix to this form, we are able to read off the answer for the vector \vec{x} as

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Thus, the work in solving the linear system just comes down to working in a clever manner with row operations.

Remark 9.4.1. Note that not every linear system will have ones along the diagonal portion as shown above. It is possible that some of the ones above could be zero, or if we started with a matrix that was not square (i.e., not $n \times n$), then there will be more to deal with. However, the goal should always be to try and reduce the matrix to this point.

Example 9.4.1: A System of Inhomogeneous Equations

Let us consider the following matrix equation:

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 2 & 3 \\ 4 & 4 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

If we multiply this out, we get the following system of linear equations:

$$\begin{aligned} 1x + 0y + 2z &= 1 \\ 2x + 2y + 3z &= 1 \\ 4x + 4y + 1z &= 1. \end{aligned}$$

You can solve this by hand, but it is tedious. In fact, each of the tricks you would use to solve this are captured by elementary row operations anyway. So, we will use the row reduction technique instead. Let us create the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 4 & 4 & 1 & 1 \end{array} \right).$$

We can then do row operations on the whole matrix (including the added vector column) to get our result

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 4 & 4 & 1 & 1 \end{array} \right) \xrightarrow{-2R_2 \text{ from } R_3} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 0 & 0 & -5 & -1 \end{array} \right)$$

then

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 0 & 0 & -5 & -1 \end{array} \right) \xrightarrow{-2R_1 \text{ from } R_2} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & -5 & -1 \end{array} \right).$$

Now, we can continue,

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & -5 & -1 \end{array} \right) \xrightarrow{\div R_3 \text{ by } -5} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & 1 & 1/5 \end{array} \right),$$

then

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & 1 & 1/5 \end{array} \right) \xrightarrow{+R_3 \text{ to } R_2} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & -4/5 \\ 0 & 0 & 1 & 1/5 \end{array} \right),$$

and next

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & -4/5 \\ 0 & 0 & 1 & 1/5 \end{array} \right) \xrightarrow{\div R_2 \text{ by } 2} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -2/5 \\ 0 & 0 & 1 & 1/5 \end{array} \right),$$

and lastly,

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -2/5 \\ 0 & 0 & 1 & 1/5 \end{array} \right) \xrightarrow{-2R_3 \text{ from } R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3/5 \\ 0 & 1 & 0 & -2/5 \\ 0 & 0 & 1 & 1/5 \end{array} \right),$$

Notice now that this corresponds to the equations

$$\begin{aligned} x + 0y + 2z &= \frac{3}{5} \\ 0x + y - 1z &= \frac{-2}{5} \\ 0x + 0y + z &= \frac{1}{5}. \end{aligned}$$

These provide us the answers

$$z = \frac{1}{5} \implies y = \frac{-2}{5} \implies x = \frac{3}{5}.$$

Double check my work above. But we can also plug in our vector now to verify our result. That is

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 2 & 3 \\ 4 & 4 & 1 \end{pmatrix} \begin{pmatrix} 3/5 \\ -2/5 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Remark 9.4.2. One should keep in mind that not every system of equations will have a solution. The one above does, but that does not mean that they all will!

The case for solving homogeneous systems is slightly different. It turns out that there is an extra *degree of freedom* in the homogeneous equations (when they have a solution). If one is able to find a solution to a homogeneous equation, then any scalar times that solution vector will also be a solution. It is also possible to find multiple vectors whose linear combinations are also a solution.

Example 9.4.2: System of Homogeneous Equations

Consider the matrix

$$[A] = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 4 & 3 \\ 1 & 2 & 1 \end{pmatrix}$$

and solve for the vector \vec{x} so that $[A]\vec{x} = \vec{0}$. Then the augmented matrix is

$$[M] = \left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 1 & 4 & 3 & 0 \\ 1 & 2 & 1 & 0 \end{array} \right).$$

Now we can perform row operations. It will be good to keep track of what you do as I do.

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 1 & 4 & 3 & 0 \\ 1 & 2 & 1 & 0 \end{array} \right) \xrightarrow{-1/2R_1 \text{ from } R_2 \text{ and } R_3} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & 5/2 & 5/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{array} \right)$$

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & 5/2 & 5/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{array} \right) \xrightarrow{-1/5R_2 \text{ from } R_3} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & 5/2 & 5/2 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Notice that the whole last row is all 0 now. This corresponds to the equation:

$$0x + 0y + 0z = 0.$$

In this case, we can plug in any value for x , y , or z and still have a solution. However, if we look at the other two equations

$$2x + 3y + 1z = 0$$

$$\frac{5}{2}y + \frac{5}{2}z = 0,$$

we can notice that we do have restrictions on the variables. If we reduce further, we will have

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

Thus we must have that

$$x = -y \quad \text{and} \quad y = -z.$$

Thus, we can see that choosing a value for z determines the values for the other variables. In this case, we call z a **free variable** since we could choose any value for it. To denote this arbitrary choice for z , I'll let $z = t$. Hence we have that $y = -t$ and $x = t$ which means the solution here is the vector

$$\vec{x} = \begin{pmatrix} t \\ -t \\ t \end{pmatrix},$$

where t is *any* real number. Let's check this:

$$\begin{pmatrix} 2 & 3 & 1 \\ 1 & 4 & 3 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} t \\ -t \\ t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

So we are happy. We found a solution! In fact, we found infinitely many solutions. It turns out that anything on this *line* is a solution.

Remark 9.4.3. Note that the zero vector $\vec{0}$ is always a solution to homogeneous equations! So we tend to look for *more* solutions than just the zero vector.

Finding the homogeneous solutions for a matrix $[A]$ is special enough to warrant a name of its own. We will also use this terminology later on.

Definition 9.4.3: Nullspace

Given a (possibly rectangular) matrix $[A]$, we call the set of all solutions to the homogeneous equation

$$[A]\vec{x} = \vec{0}$$

the **nullspace** of the matrix $[A]$ and denote this by $\text{Null}([A])$.

The nullspace of a matrix tells us a lot about the homogeneous solutions to equations with that matrix. Note that the zero vector $\vec{0}$ is always a member of the nullspace for any matrix $[A]$. If we have that $\text{Null}([A]) = \{\vec{0}\}$, then we say that the nullspace for $[A]$ is **trivial**. Also, if we find any set of vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ is in the nullspace of a matrix $[A]$, then any linear combination of those vectors is in the nullspace as well.

Systems of Equations as Linear Combinations of Columns

There is also another way of thinking about these matrix/vector equations. If we have, for example, a 3-dimensional system of equations given by

$$[A]\vec{x} = \vec{y}$$

then we can think of the columns of $[A]$ as being 3-dimensional vectors. Remember, this was how we first introduced the matrix representation for a linear transformation! That

is, we can put:

$$\begin{pmatrix} \vec{A}_1 & \vec{A}_2 & \vec{A}_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

where we have

$$\vec{A}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} \quad \vec{A}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} \quad \vec{A}_3 = \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}.$$

Note that if we perform this multiplication above we have

$$x_1 \vec{A}_1 + x_2 \vec{A}_2 + x_3 \vec{A}_3 = \vec{y}.$$

So indeed this matrix/vector multiplication is describing how to take a linear combination of the columns of the matrix $[A]$ to obtain the vector \vec{y} .

Question 9.4.1. Is it always possible to obtain the vector \vec{y} given any matrix $[A]$?

Answer 9.4.1. No, it is not. We will see a useful tool in the next section that will help us figure out whether this is possible or not.

9.5 Linear Independence, Span, and Bases

Say we are looking at the vectors space \mathbb{R}^3 . What are the smallest sets of vectors that we need in order to create any vector in that space? Clearly, we found that the unit basis vectors \hat{x} , \hat{y} , and \hat{z} allowed us to do this. When we read these as column vectors

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

this becomes more obvious. Since if we are given a vector

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix},$$

we can put

$$\vec{v} = v_x \hat{x} + v_y \hat{y} + v_z \hat{z}.$$

However, these unit basis vectors are not the only vectors that we can use to build the vector \vec{v} . This motivates the following two definitions.

Definition 9.5.1: Linearly Independent Vectors

Consider a list of vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ in the vector space \mathbb{R}^n . We say that the list of vectors is **linearly independent** if

$$\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_m \vec{v}_m = \vec{0}$$

if and only if each $\alpha_i = 0$. If the list of vectors is not linearly independent, we say that it is *linearly dependent*.

It quickly follows that if we are given too many vectors the set must be linearly dependent. For example, if handed a set of four vectors in the vector space \mathbb{R}^3 , it is guaranteed that they are linearly dependent.

Proposition 9.5.1: More Vectors than Dimension

Consider the list of vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ in the vector space \mathbb{R}^m . Then if $m > n$, the list of vectors is linearly dependent.

The above proposition is similar to why solving underdetermined systems does not have a unique solution. Similarly, if we do not have enough vectors in our list, it may be a linearly dependent list, but we may not be able to form any vector in the space from that list of vectors. This is analogously similar the case of an overdetermined system. So we would like to know if a list of vectors suffices to generate any vector in the vector space.

Definition 9.5.2: Span

Given a list of vectors $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ in the vector space \mathbb{R}^n we say that the *span* of the vectors in the set is all of the possible vectors that can be written as a linear combination of the elements in the set.

When a set of vectors spans the whole space we are interested in, then this suffices to give us a way to build any vector in that space from a linear combination. We often care to have the smallest set of vectors that does this as well.

Definition 9.5.3: Basis

A *basis* for the vector space \mathbb{R}^n is the smallest list of vectors such that the span of those vectors is \mathbb{R}^n .

It follows that for a space of n -dimensions that we must have a basis of n -elements. No more, or no less. That is exactly why we have three unit basis vectors \hat{x} , \hat{y} , and \hat{z} for the vector space \mathbb{R}^3 .

9.6 The Determinant and Trace

Previously we saw matrix vector equations and posed the question in whether the equations or solvable. We also so that it is an equivalent question to ask whether the set of vectors making up a matrix $[A]$ can be put in a linear combination to yield a desired output vector. In the previous chapter we also explored the geometry of vectors in space and we were

able to compute areas and distances. Now, with one extra tool we will be able to tie all of these concepts together and perform these calculations in arbitrary dimension.

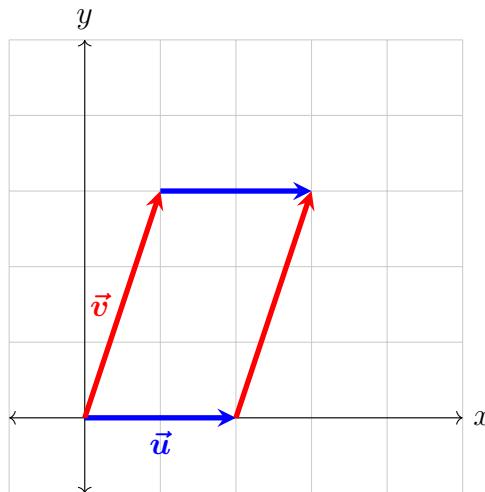
To start, say that we want to compute the area of a parallelogram. Before, we used the cross product to do this, but what will we do if we want to compute volume in higher dimension? Let us explore how we may think of doing this in general.

Example 9.6.1: Area of a parallelogram

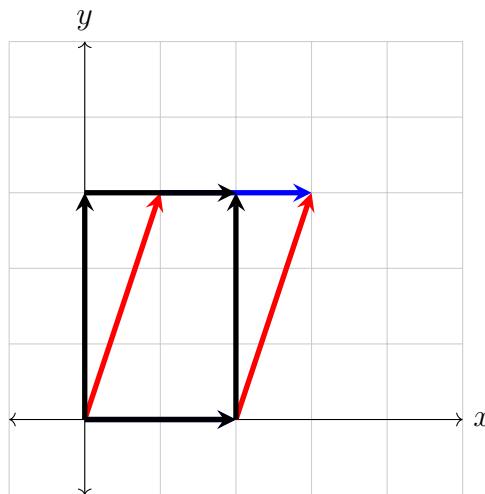
Say we have the vectors

$$\vec{u} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

What is the area of the parallelogram that these vectors create? (See previous notes for how the cross product can do this. We'll see that the cross product is highly related to determinants.)



The trick to finding the area here is to take move a triangle and fill in a gap. We do this like so:



In the above, we've moved a triangle from the right, to the left, to make the black rectangle. This rectangle then has an area of $2 \cdot 3 = 6$.

Let us see the way we can do this that does not require this extra work. Let us place these vectors into a matrix:

$$A = \begin{pmatrix} | & | \\ \vec{u} & \vec{v} \\ | & | \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}.$$

Then $\det(A)$ will give us the area of this parallelogram! We have

$$\det(A) = 2 \cdot 3 - 1 \cdot 0 = 6.$$

This begs the question, what is this formula in general? Note, that if you placed these vectors in \mathbb{R}^3 by adding a zero z -component, then

$$\|\vec{u} \times \vec{v}\| = 6,$$

as well.

In order to move to higher dimensions we must first settle the case in two dimensions. So, to compute the **determinant** of a 2×2 -matrix

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

we have

$$\det([A]) = ad - bc.$$

You can remember this by thinking that you multiply top left with bottom right, then subtract the product of top right with bottom left. Also, we will often use the notation of vertical bars around the matrix of numbers to denote a determinant. That is,

$$\det([A]) = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

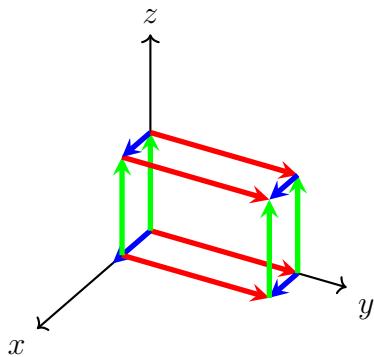
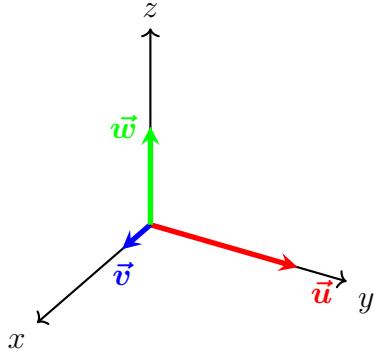
Now, the generalization of a parallelogram to 3-dimensional space is called a **parallelepiped**. Later on when we are doing calculus in 3-dimensional space it will become extremely important to be able to compute infinitesimal volumes (just like we have infinitesimal lengths dx) and they will be found via the volume of a parallelepiped.

Example 9.6.2: Volume of a Parallelepiped

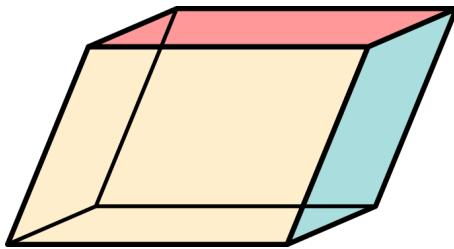
In 3-dimensional space, we can take three vectors

$$\vec{u} = \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{w} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

These, when combined properly, enclose a volume of a parallelepiped.



Of course, I picked an easy example for the illustration. But what is the volume? Here we again know we can compute this the usual way and we get that the volume is $1 \cdot 1 \cdot 2 = 2$. If we instead created a more complicated parallelepiped like:



we must perform the same type of process as we did in Example 9.6.1. The determinant is designed to do this process for us.

Now, let us consider computing the determinant for a 3×3 -matrix $[A]$. We should think of this matrix as being created from three vectors \vec{A}_1 , \vec{A}_2 , and \vec{A}_3 as columns. So we have

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} | & | & | \\ \vec{A}_1 & \vec{A}_2 & \vec{A}_3 \\ | & | & | \end{pmatrix},$$

we want $\det(A)$ to reflect the the volume of a parallelepiped generated by \vec{A}_1 , \vec{A}_2 , and \vec{A}_3 . It turns out the way we compute $\det(A)$ comes from using the determinant of 2×2 -matrices. Let me explain further.

Let me write this matrix as a tool:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}.$$

What we will use this for is a memory tool when we do something called the **cofactor expansion**. What I will do is choose a row or column to expand along. For this example, let's choose the top row to expand along. If we use this above matrix with pluses and minuses, we note that the top row will have signs $+ - +$. These signs show up in our computation. We will also have to take into account the element in the original matrix and multiply by that.

We start with the top left of $[A]$ and we will have a $+$ sign. The top left of $[A]$ is element a_{11} and so we remove the first row and first column from $[A]$ to give the **cofactor matrix**

$$\text{Cof}_{11}[A] = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}.$$

Since this is a 2×2 -matrix, we can compute the determinant by

$$\det(\text{Cof}_{11}[A]) = a_{22}a_{33} - a_{23}a_{32}.$$

Then we continue expanding along this row and move on to the top middle of $[A]$ and will have a $-$ sign. The top middle element is a_{12} and so we remove the first row and second column of $[A]$ to get

$$\text{Cof}_{12}[A] = \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix}.$$

Then we have that

$$\det(\text{Cof}_{12}[A]) = a_{21}a_{33} - a_{31}a_{13}.$$

Lastly, we move on to the top right of $[A]$ and will have a $+$ sign. The top right element is a_{13} and so we remove the first row and third column of $[A]$ to give

$$\text{Cof}_{13}[A] = \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

Then we have that

$$\det(\text{Cof}_{13}[A]) = a_{21}a_{32} - a_{22}a_{31}.$$

We then define the determinant of $[A]$ to be

$$\det([A]) = +a_{11}\det(\text{Cof}_{11}[A]) - a_{12}\det(\text{Cof}_{12}[A]) + a_{13}\det(\text{Cof}_{13}[A]).$$

Similarly, if we expanded along the second column, for example, we would have that

$$\det([A]) = -a_{12}\det(\text{Cof}_{12}[A]) + a_{22}\det(\text{Cof}_{22}[A]) - a_{32}\det(\text{Cof}_{32}[A]).$$

Example 9.6.3: Volume of a Parallelepiped from the Determinant

Consider the vectors from the previous example

$$\vec{u} = \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{w} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

Then we can place these three vectors into a matrix $[A]$ by

$$[A] = \begin{pmatrix} | & | & | \\ \vec{u} & \vec{v} & \vec{w} \\ | & | & | \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Then we can compute the determinant of $[A]$ by expanding along any row or column. Typically, one wants to expand along a row or column with the most zeros. This will reduce the amount of work. Here, it does not matter, so let us expand across the top row. We get

$$\begin{aligned} \det([A]) &= a_{11} \det(\text{Cof}_{11}[A]) - a_{12} \det(\text{Cof}_{12}[A]) + a_{13} \det(\text{Cof}_{13}[A]) \\ &= 0 \cdot \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix} - 1 \cdot \begin{vmatrix} 3 & 0 \\ 0 & 2 \end{vmatrix} + 0 \cdot \begin{vmatrix} 3 & 0 \\ 0 & 0 \end{vmatrix} \\ &= -1 \cdot (3 \cdot 2) \\ &= -6. \end{aligned}$$

This determinant gave us exactly the negative of the volume of the parallelepiped.

From the previous example we saw that the determinant need not always be positive. So we must say that the determinant of a matrix $[A]$ provides us the *signed volume* of the parallelepiped generated by its column vectors. The orientation is much like the orientation we saw for the cross product.

Exercise 9.6.1. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

- (a) Find $\text{Cof}_{11}(A)$.
- (b) Find $\text{Cof}_{22}(A)$.
- (c) Find $\text{Cof}_{23}(A)$.
- (d) Compute $\det(A)$.

Question 9.6.1. What does it mean if $\det(A) = 0$ for some matrix A ?

Answer 9.6.1. It means that we have a parallelopiped with zero volume. Which means our three vectors actually lie in a plane, or even on a single line. This is an important fact that will come up when we solve systems of equations!

Remark 9.6.1. Determinants can be computed in arbitrary dimension using the same process above. But, this will not be a worry for us at the moment.

9.6.1 Determinants and Systems of Equations

In defining the determinant, we found it useful for computing the volume of a parallelepiped. However, it does not stop there. If we adopt the convention that systems of linear equations such as

$$[A]\vec{x} = \vec{y}$$

where $[A]$ is a square matrix, then this is related to the question of whether the columns of $[A]$ form a spanning list of vectors for the vector space.

Since the determinant $\det([A])$ tells us the volume of the parallelepiped of the three column vectors creating $[A]$, if the volume is zero, then we know that the columns of $[A]$ must be linearly dependent. Since, if they were nonzero, the vectors would generate some volume. Thus we have the following two propositions.

Proposition 9.6.1: Solutions to Homogeneous Equations

Consider the homogeneous equation

$$[A]\vec{x} = \vec{0}.$$

Then we have the solutions:

- (*Trivial*) *The only solution for \vec{x} is $\vec{x} = \vec{0}$ if and only if $\det(A) \neq 0$;*
- *There are multiple nonzero solutions for \vec{x} if and only if $\det(A) = 0$.*

The other way of wording the above proposition is that a matrix $[A]$ with a nonzero determinant equal has a nullspace that is trivial. In other words, if

$$\det([A]) = 0$$

then $\text{Null}([A]) = \{\vec{0}\}$. Otherwise, if the determinant is zero, then $\text{Null}([A])$ contains more than just the zero vector.

There is a similar result for the inhomogeneous equations as well. However, be careful to note the differences here. Afterwards we can discuss geometrically what is happening.

Proposition 9.6.2: Solutions to Inhomogeneous Equations

The inhomogeneous equation

$$[A]\vec{x} = \vec{y}$$

(with $\vec{y} \neq \vec{0}$) have the following solutions:

- *A unique solution for \vec{x} if and only if $\det(A) \neq 0$;*
- *None or possibly infinitely many solutions for \vec{x} if $\det(A) = 0$.*

Exercise 9.6.2. Take the determinant and then attempt to solve the following homogeneous equations.

(a) Let

$$[A] = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 5 \\ 0 & 1 & 2 \end{pmatrix}$$

and solve $A\vec{x} = \vec{0}$.

(b) Let

$$[B] = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 4 & 3 \\ 1 & 2 & 1 \end{pmatrix}$$

and solve $[B]\vec{x} = \vec{y}$.

When thinking about these propositions, one should consider the case for functions $f: \mathbb{R} \rightarrow \mathbb{R}$. In this case, we may have been handed a function f , a y (output) value, and were asked to find what input x corresponds to that y value. That is, we want

$$f(x) = y.$$

Here we could sometimes find an inverse function f^{-1} that would satisfy

$$f^{-1}(y) = x.$$

However, for example, if we had the function $f(x) = x^2$, then there was no inverse function. Why? Well, say we let $y = 1$, then $x = \pm 1$ are both solutions. This is not a valid function (since it does not pass the vertical line test). Also, if we provide $y = -1$, then there is no input value to achieve that output.

We can also think of solving systems of linear equations in a geometrical way. Let us consider two examples. First, for the homogeneous case, we can take the homogeneous equation

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Here we can see that this gives the equations

$$x = 1, \quad y = 1, \quad z = t.$$

That is, z is a free variable. Hence here we find that the nullspace for the matrix

$$[A] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is given by

$$\text{Null}([A]) = \text{Span} \left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right),$$

which is the z -axis. If we think of columns of $[A]$ as vectors, then we find that the span of the columns is the xy -plane. Thus, any z -component of an input vector does not affect the output.

If we take this same matrix at consider an inhomogeneous problem, then the following equation has no solution

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

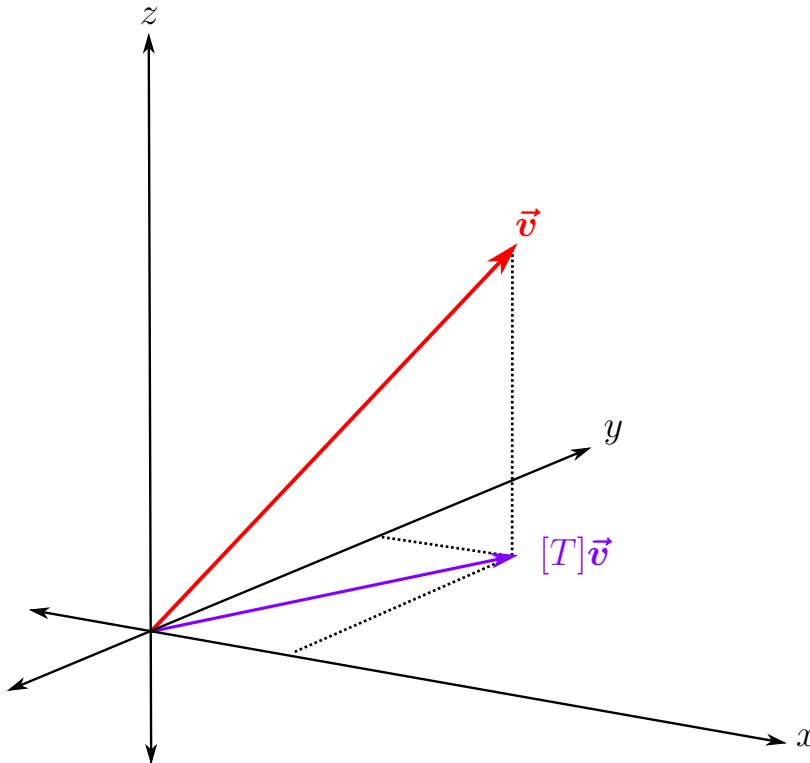
Why is that? Well, this gives rise to the equations

$$x = 1, \quad y = 1, \quad 0 \cdot z = 1,$$

which has no solution for z . This is exactly because of the fact that the columns of $[A]$ only span the xy -plane. Taking a linear combination of the columns cannot give a resulting vector with a z -component. We can visualize this as follows. Let

$$[T] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \vec{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

and we can see geometrically what this transformation does to a vector. In a sense, this transformation removes a whole dimension from \mathbb{R}^3 . It takes any z -component and removes it, hence we see that the transformation really takes the shadow of the input vector on the xy -plane.



That is, we visually see that

$$[T]\vec{v} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Though this was a special circumstance, this is the exact picture to have in mind when solving systems of linear equations. If we are given a point \vec{y} , a matrix $[A]$, and are asked to find the input vector \vec{x} that solves

$$[A]\vec{x} = \vec{y},$$

we are being asked if \vec{y} is in the span of the columns of $[A]$ and \vec{x} describes the coefficients for this linear combination. And as we have seen, the columns of $[A]$ may or may not span the necessary amount of space to be able to create the output vector \vec{y} .

The cross product and the determinant are related in some ways. Since the formula for a cross product is not easy to memorize, and applying the cross product to basis vectors is tedious, it suffices to remember how to compute determinants of 3×3 -matrices and stop there. It is possible to compute the cross product by using the proper determinant.

Example 9.6.4: Cross product from the Determinant

You can compute the cross product from the determinant. (You should be warned: this is an abuse of notation and a somewhat weird coincidence.) Let us choose two vectors \mathbf{v} and \mathbf{w} . We place them in a matrix $[A]$ as follows:

$$A = \begin{bmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix}.$$

Then $\det([A])$ will give us the cross product $\mathbf{v} \times \mathbf{w}$. Just know that you are briefly ignoring the fact that $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ are actually vectors when you compute this. Treat them as numbers until you get your final result.

9.6.2 Properties of determinants

There are a few nice properties of determinants to keep in mind.

- **Transposition:** If we exchange rows for columns in a matrix (that is, to take the *transpose matrix*, then the value of the determinant is the same. Given

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

we write the transpose matrix

$$[A]^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

So $a_{ij} \mapsto a_{ji}$. Then

$$\det([A]) = \det([A]^T).$$

- **Multiplication by constants:** If we multiply a row or column by a scalar, then the determinant is also multiplied by that scalar. So we have

$$[A]_\alpha = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Then

$$\det([A]_\alpha) = \alpha \det([A]).$$

And this holds no matter what row or column we scale.

- **Linear combinations of rows or columns:** If we add a linear combination of columns (or rows) to any column (or row), then the determinant does not change. That is, if we have

$$[A] = \begin{pmatrix} \vec{A}_1 & \vec{A}_2 & \vec{A}_3 \end{pmatrix},$$

then

$$\det([A]) = \begin{vmatrix} \vec{A}_1 + \alpha \vec{A}_2 + \beta \vec{A}_3 & \vec{A}_2 & \vec{A}_3 \end{vmatrix}.$$

Here we just showed adding a linear combination of two columns to the first column, but this holds in general for adding linear combinations of any columns to other columns or adding linear combinations of rows to any row.

- **Exchanging rows or columns:** If we swap rows or columns, the determinant is multiplied by -1 . So we have

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{vmatrix}$$

Again, this holds for swapping any two rows (or any two columns).

- **Linearly-dependent rows or columns:** If a column (resp. row) of a matrix can be written as a linear combination of the other two columns (resp. rows), then the determinant is zero. This is basically saying that all three vectors lie in a single plane and so the volume of the parallelepiped given by the three vectors (as columns in the matrix) create no volume.

Say we can write

$$\vec{A}_1 = \alpha \vec{A}_2 + \beta \vec{A}_3$$

Then $\det(A) = 0$. That is, if one column (or row) is in the span of the other columns (or rows), then the determinant of the matrix will be zero.

Another nice property of the determinant comes from the product of two matrices. If we consider two square $n \times n$ -matrices $[A]$ and $[B]$, then we can multiply both $[A]$ and $[B]$. Then we have that the determinant is **multiplicative** in that

$$\det([A][B]) = \det([A]) \det([B]).$$

This turns out to be a very useful property of the determinant. If one thinks of the determinant as describing how much the basis vectors \hat{x} , \hat{y} , and \hat{z} are stretched by the composition transformation $[A][B]$, then this is saying that the product transformation stretches the vectors by first stretching by $[A]$ then by $[B]$.

9.6.3 The Trace

The determinant was an example of an *invariant* of a matrix $[A]$. That is, it is a quantity that does not change even when certain operations are performed. These operations that leave the determinant invariant were the row operations.

Given a square matrix $[A]$, we can associate to it another invariant called the *trace* which is rather easy to compute. The trace of a matrix is simply the sum of the diagonal entries. That is, given

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

then the trace of $[A]$ is

$$\text{tr}([A]) = a_{11} + a_{22} + a_{33} = \sum_{i=1}^3 a_{ii}.$$

Example 9.6.5: Trace of a Matrix

Consider the matrix

$$[A] = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Then we have

$$\text{tr}([A]) = 1 + 5 + 9 = 15.$$

The trace has a few nice properties as well. First, the trace is *additive* so that if we have two $n \times n$ -matrices $[A]$ and $[B]$ then

$$\text{tr}([A] + [B]) = \text{tr}([A]) + \text{tr}([B]).$$

However, the trace is not multiplicative like the determinant!

But, if we have a product of matrices $[A]$, $[B]$, and $[C]$, then we have

$$\text{tr}([A][B][C]) = \text{tr}([B][C][A]) = \text{tr}([C][A][B]),$$

which is a way of saying that we can *cyclically permute* products of matrices with the trace. For a product of two matrices this means

$$\text{tr}([A][B]) = \text{tr}([B][A]).$$

Lastly, if we scale a matrix by a constant, then that scales the trace by a constant as well. That is,

$$\text{tr}(\alpha[A]) = \alpha \text{tr}([A]).$$

The invariant actions for the trace are a require a bit more knowledge that we will develop in the next section.

9.7 Inverse Matrices and Similarity

As previously mentioned, we can be handed an equation such as

$$f(x) = y,$$

and attempt to find an inverse function f^{-1} so that we have

$$f^{-1}(y) = x.$$

This is generally a very useful tactic to use. In this case, we can quickly find an input x for a given y without having to solve a new set of equations each time.

Since matrices are representations of linear functions, we can attempt to find an inverse for matrices as well. If we are given the equation

$$[A]\vec{x} = \vec{y},$$

then we can try to determine the matrix $[A]^{-1}$ so that

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

Hence, if we are given any output vector \vec{y} , we can quickly find the corresponding input vector \vec{x} through matrix multiplication. We call this special matrix $[A]^{-1}$ the **inverse matrix**. In this case $[A]$ must be square or else this is not at all possible! If a given matrix $[A]$ has an inverse matrix we say that $[A]$ is **invertible**.

Since $[A]$ is a function (specifically, a linear transformation), we do not always have an inverse. However, the determinant carries enough data to tell us when a matrix is invertible. Suppose that $[A]$ is an $n \times n$ -matrix, it turns out that $[A]^{-1}$ exists when the columns of $[A]$ span \mathbb{R}^n . Based on our knowledge of the determinant, we can phrase this another way.

Proposition 9.7.1: Existence of Inverse

If $\det([A]) \neq 0$, then $[A]^{-1}$ exists and is unique.

When speaking of inverses, we must also realize what function acts as the **identity**. For example, if we have a function $f(x) = y$ that is invertible. Then we know

$$f(f^{-1}(y)) = y \quad \text{and} \quad f^{-1}(f(x)) = x.$$

In other words, the composite functions

$$f \circ f^{-1} \quad \text{and} \quad f^{-1} \circ f$$

give the same output value as the given input value. So, we must ask what matrix (or linear transformation) gives the same output value for a given input value.

Definition 9.7.1: Identity Matrix

We call the matrix $[I]$ with entries

$$[I]_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

the ***identity matrix***. For example, the 3×3 identity matrix is

$$[I] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since we can multiply matrices and vectors, we should see how the identity matrix acts on each. Just as we had with functions, the identity should fix the input value and thus give the same value as the output. In other words, this is the matrix that “does nothing.”

Proposition 9.7.2: Identity Matrix Fixes Vectors

For any vector \vec{v} we have that

$$[I]\vec{v} = \vec{v}.$$

In fact, for any equal sized square matrix $[A]$, we have that

$$[I][A] = [A][I] = [A].$$

Exercise 9.7.1. Let

$$\vec{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Show that

$$[I]\vec{v} = \vec{v}.$$

Similarly, let

$$[A] = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Show that

$$[I][A] = [A].$$

The most important property of the inverse is that given an invertible matrix $[A]$ we have that

$$[A]^{-1}[A] = [A][A]^{-1} = I.$$

This is exactly like the inverse of a function (as it should be since $[A]$ is just a special type of function). To think of this geometrically, we also know that if $[A]$ is an $n \times n$ -matrix,

the columns of $[A]$ must span \mathbb{R}^n . If this were not the case, say with the matrix

$$[A] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

then we cannot invert the matrix. See for example, this matrix above is the function that forgets the z -component of an input vector \vec{x} . So if we take

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

we see that we cannot possibly recover the z -input. Why? We could have given any value for the z -input, but the output is always zero which gives us no means of recovering the true input value. By this argument, we can see that if a matrix $[A]$ has a nontrivial nullspace, we cannot invert it. This is equivalent to the statement that $\det([A])$ must be nonzero in order to construct $[A]^{-1}$.

Inverting a matrix

Since we know a matrix can be inverted, we should also work on constructing the inverse. So, for example, let $[A]$ be a 3×3 invertible matrix then we can create the augmented matrix

$$[M] = \left(\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array} \right),$$

where we have augmented the identity matrix $[I]$ along with the matrix $[A]$. From here, we can row reduce the left portion until we have

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \tilde{a}_{11} & \tilde{a}_{12} & \tilde{a}_{13} \\ 0 & 1 & 0 & \tilde{a}_{21} & \tilde{a}_{22} & \tilde{a}_{23} \\ 0 & 0 & 1 & \tilde{a}_{31} & \tilde{a}_{32} & \tilde{a}_{33} \end{array} \right),$$

where \tilde{a}_{ij}^{-1} is the entry of the matrix A^{-1} .

Note that in the 3×3 case that we have

$$[I] = \begin{pmatrix} | & | & | \\ \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ | & | & | \end{pmatrix}.$$

Hence, the idea of the augmented matrix above is that we are checking that if $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, or $\hat{\mathbf{z}}$ are given as output vectors, then we can construct an input vector corresponding to those output values.

Exercise 9.7.2. Let

$$[A] = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

(a) Find $[A]^{-1}$.

(b) Compute $\det([A])$ and $\det([A]^{-1})$. What do you notice about $\det([A]^{-1})$ compared to $\det([A])$?

9.7.1 Properties Related to Inverse Matrices

Matrices that are invertible are rather special and with that comes some related properties. For example, we have that

$$\det([I]) = 1.$$

If we take that along with the fact that

$$\det([A][B]) = \det([A]) \det([B]),$$

then we have the following proposition.

Proposition 9.7.3: Determinant of Inverse Matrix

We have that

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Also, the inverse of an inverse matrix is the original matrix. That is,

$$([A]^{-1})^{-1} = [A].$$

Lastly, one may consider the inverse of a product of matrices. If we have the product of two matrices

$$[A][B],$$

then to invert this product, we have to undo the transformations in the proper order. We refer to this as *socks and shoes* since this is analogous to how we put on socks then shoes and take them off in the reverse order! Specifically, we have that

$$([A][B])^{-1} = [B]^{-1}[A]^{-1},$$

which we can see is correct by

$$([A][B])^{-1}[A][B] = [B]^{-1}[A]^{-1}[A][B]^{-1} = [B]^{-1}[B] = [I].$$

Invertible matrices are also matrices that can be used to simplify other matrices. Typically, when we are handed a matrix $[A]$, we only care about the main invariant properties of that matrix. That is, properties which are captured through a certain type of operation.

Definition 9.7.2: Similar Matrices

Let $[A]$ be an $n \times n$ -square matrix, and $[P]$ an $n \times n$ invertible matrix. Then we say that the matrix $[Q]$ is **similar** to the matrix $[A]$ if we have

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

As it turns out that there is typically a “nicest looking” matrix that is similar to a given matrix $[A]$. This will be one of the main goals of the Eigenvalue problem which we will see in the next section. The two main invariant quantities we associate to a matrix are the trace and determinant and thus, we should hope they remain invariant for any similar matrix. This motivates the following proposition which we shall prove.

Proposition 9.7.4: Trace and Determinant are Invariant

Let $[A]$ and $[Q]$ be similar matrices. Then we have that

$$\det([A]) = \det([Q]) \quad \text{and} \quad \text{tr}([A]) = \text{tr}([Q]).$$

Proof. Since $[A]$ and $[Q]$ are similar, there exists an invertible matrix $[P]$ so that

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

Then we have that

$$\begin{aligned}\det([Q]) &= \det([P]^{-1}[A][P]) \\ &= \det([P]^{-1}) \det([A]) \det([P]) \\ &= \frac{1}{\det([P])} \det([A]) \det([P]) \\ &= \det([A]).\end{aligned}$$

Similarly for the trace we have,

$$\begin{aligned}\text{tr}([Q]) &= \text{tr}([P]^{-1}[A][P]) \\ &= \text{tr}([P][P]^{-1}[A]) \\ &= \text{tr}([I][A]) \\ &= \text{tr}([A]).\end{aligned}$$

□

9.8 The Eigen-Problem

Given a square matrix $[A]$, what is the “nicest looking” matrix $[\Lambda]$ that is similar to $[A]$? This is a question of fundamental importance in many ways. For one, it allows us to simplify down a linear transformation into the form that is the most easy to understand. Also, the problem is very physical as it turns out to be exactly the description of Schrödinger’s equation. All of this can be stated in the following way.

Question 9.8.1. Given a square matrix $[A]$, does there exist a scalar λ and a corresponding vector \vec{e} such that

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

This type of equation is called an **eigenvalue equation**. We can immediately compare it to Schrödinger’s equation which takes the form

$$H(x)\Psi(x) = E\Psi(x)$$

where $H(x)$ is the Hamiltonian, and E is the energy value for the wavefunction $\Psi(x)$. It turns out that this is an eigenvalue equation, but in this case $H(x)$ can be something other than a matrix of constant numbers (which makes it more complicated to solve).

Definition 9.8.1: Eigenvalues and Eigenvectors

Let $[A]$ be a square matrix. Then if we have a scalar λ and \vec{e} satisfying

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

we call λ an *eigenvalue* and \vec{e} the corresponding *eigenvector*.

Remark 9.8.1. It's important to note that an eigenvalue *corresponds* to an eigenvector. They are inseparable partners!

Question 9.8.2. Why in the world should we care about this? What are the applications?

Answer 9.8.1.

- First off, mathematically what this is saying is we can find vectors that are affected by a linear transformation (a matrix $[A]$) in the simplest possible way. That is, the matrix just scales the vector! If we are able to find a basis of eigenvectors, then we can understand how a matrix $[A]$ acts entirely by scaling certain directions. This last part is not always possible, but we are not concerned with this.
- The applications are extremely far ranging.
 - Data analysis: Understanding the structure of a data set.
 - Mechanics: Understanding rotational motion.
 - Quantum Mechanics: The whole theory of “first quantization” is about solving the eigen-equation called *Schödinger’s equation*. There, it turns out that eigenvectors are the observable states for a quantum system.
 - Molecular Orbitals: Molecular orbitals are eigenvectors of the Fock operator.
 - Differential Equations: We can often recast differential equations as eigen-problems which are, in general, fairly easy to solve in comparison. Here eigenvectors are functions.
 - Finance: Breaking down portfolios based on risk and returns.
 - Geology: The study of glacial till.
 - Singular Value Decomposition and Principal Component Analysis: A generalization of the eigen-problem.

Great, so now we are motivated. But, how can we hope to find these eigenvalues and eigenvectors?

We begin with the equation

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

and we subtract $\lambda\vec{e}$ from both sides to yield

$$[A]\vec{e} - \lambda\vec{e} = \vec{0}.$$

Remember that we can multiply by the identity matrix and not change a single thing. We do this like so.

$$\begin{aligned}[A]\vec{e} - I\lambda\vec{e} &= \vec{0} \\ ([A] - \lambda[I])\vec{e} &= \vec{0}.\end{aligned}$$

This means we have reduced the problem to solving the homogeneous equations for a slightly different matrix $[A] - \lambda[I]$.

Recall that we had a proposition that said these homogeneous equations have nontrivial solutions if

$$\det([A] - \lambda[I]) = 0.$$

This determinant condition gives us an equation that allows us to determine a value for λ . Specifically, this is a polynomial equation in the variable λ that we refer to as the ***characteristic polynomial***. Since this is a polynomial equation, we will always be able to find roots over the field of complex numbers.

When we then know λ , we can solve for \vec{e} by plugging in the known λ value and solving the homogeneous equations. Let me put this down in an algorithmic way as follows.

1. Let λ be a variable. Then we make the matrix $[A] - \lambda[I]$.
2. Take the determinant of $[A] - \lambda[I]$ and set it equal to zero. That is,

$$\det([A] - \lambda[I]) = 0.$$

This will give you a polynomial equation with the variable λ .

3. You will find (possibly) multiple values of λ from the determinant above.
4. For each λ_i , you will find an eigenvector \vec{e}_i by solving the homogeneous equation

$$([A] - \lambda_i[I])\vec{e}_i = \vec{0}.$$

In other words, \vec{e}_i is in the nullspace $\text{Null}([A] - \lambda_i[I])$.

Example 9.8.1: Eigenvalues of a Diagonal Matrix

The easiest example of the eigenvalue problem is finding the eigenvectors and eigenvalues of a *diagonal* matrix. In this case, let

$$[A] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

1. Take λ to be a variable and put

$$A - \lambda I = \begin{bmatrix} 1 - \lambda & 0 & 0 \\ 0 & 2 - \lambda & 0 \\ 0 & 0 & 3 - \lambda \end{bmatrix}.$$

2. Now we take the determinant of $[A] - \lambda[I]$

$$\det([A] - \lambda[I]) = (1 - \lambda)(2 - \lambda)(3 - \lambda).$$

We set this equal to zero

$$(1 - \lambda)(2 - \lambda)(3 - \lambda) = 0.$$

3. Note that this has solutions $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$.

4. • $\lambda_1 = 1$: We take

$$([A] - 1[I])\vec{e}_1 = \vec{0}.$$

We solve this by making the augmented matrix:

$$[M] = \left(\begin{array}{ccc|c} 1-1 & 0 & 0 & 0 \\ 0 & 2-1 & 0 & 0 \\ 0 & 0 & 3-1 & 0 \end{array} \right).$$

We can then perform the row reduction steps to get this augmented matrix to RREF. In our case, this is essentially done and the system of equations reads:

$$\begin{aligned} 0x + 0y + 0z = 0 &\implies x = t \\ 0x + 1y + 0z = 0 &\implies y = 0 \\ 0x + 0y + 2z = 0 &\implies z = 0. \end{aligned}$$

So we have a solution

$$\vec{e}_1 = \begin{bmatrix} t \\ 0 \\ 0 \end{bmatrix}.$$

We usually prefer our eigenvectors to be normalized, so in this case we take $t = 1$ and say that the eigenvector is

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

• $\lambda_2 = 2$: I'll suppress the work here but the eigenvector is

$$\vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

• $\lambda_3 = 3$: Again, the eigenvector is

$$\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The above case is the easiest case for solving the eigenvalue problem, and it is also one of the reasons for the eigenvalue problem which we will see in the next section. We can

consider another example.

Example 9.8.2: A 2×2 Eigen-problem

Consider the matrix

$$[A] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Now, we take

$$\begin{aligned} \det([A] - \lambda[I]) &= \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} \\ &= \lambda^2 - 1. \end{aligned}$$

Hence, setting the determinant equal to zero we find that

$$\lambda = \pm 1,$$

and we put $\lambda_1 = 1$ and $\lambda_2 = -1$.

- $\lambda_1 = 1$: We take

$$([A] - 1[I])\vec{e}_1 = \vec{0}$$

which gives us the augmented matrix

$$[M] = \left(\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array} \right),$$

which we can reduce to

$$\left(\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right),$$

which gives the equation

$$x = y,$$

hence we can take $x = 1$ which gives $y = 1$ and hence

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- $\lambda_2 = -1$: We take

$$([A] + 1[I])\vec{e}_1 = \vec{0}$$

which gives us the augmented matrix

$$[M] = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array} \right),$$

which we can reduce to

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right),$$

which gives the equation

$$x = -y,$$

hence we can take $x = 1$ which gives $y = -1$ and hence

$$\vec{e}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus we have our eigenvalue and eigenvector pairs

$$\lambda_1 = 1 \quad \vec{e}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and

$$\lambda_2 = -1 \quad \vec{e}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Every eigenvalue problem will behave this same way, however, since we need to factor polynomial equations it is best to now consider vectors that have complex entries in them. That is, we can more easily work over the vector space \mathbb{C}^n where vectors are given by

$$\vec{v} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \quad \text{with each } z_i \in \mathbb{C}.$$

Let us see why this is necessary with an example.

Example 9.8.3: A Complex 2×2 Eigen-problem

Consider the matrix

$$[A] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and take

$$\det([A] - \lambda[I]) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1.$$

Then after setting the determinant equal to zero we find that

$$\lambda = \pm i.$$

This tells us the eigenvalues for this matrix are complex! Hence the necessity for working with complex numbers as opposed to reals.

Now, we put $\lambda_1 = i$ and $\lambda_2 = -i$.

- $\lambda_1 = i$: We take

$$([A] - i[I])\vec{e}_1 = \vec{0}$$

which gives us the augmented matrix

$$[M] = \left(\begin{array}{cc|c} -i & -1 & 0 \\ 1 & -i & 0 \end{array} \right),$$

which we can reduce to

$$\left(\begin{array}{cc|c} i & 1 & 0 \\ 0 & 0 & 0 \end{array} \right),$$

which gives the equation

$$ix = -y,$$

hence we can take $x = 1$ which gives $y = -i$ and hence

$$\vec{e}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

- $\lambda_2 = -1$: We take

$$([A] + 1[I])\vec{e}_1 = \vec{0}$$

which gives us the augmented matrix

$$[M] = \left(\begin{array}{cc|c} i & -1 & 0 \\ 1 & i & 0 \end{array} \right),$$

which we can reduce to

$$\left(\begin{array}{cc|c} -i & 1 & 0 \\ 0 & 0 & 0 \end{array} \right),$$

which gives the equation

$$ix = y,$$

hence we can take $x = 1$ which gives $y = i$ and hence

$$\vec{e}_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Thus we have our eigenvalue and eigenvector pairs

$$\lambda_1 = i \quad \vec{e}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

and

$$\lambda_2 = -i \quad \vec{e}_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Now, let us revisit something we have seen before and relate this to eigenvalues. First, we must realize that not all eigenvalues for a matrix will be unique. Take for example,

$$[I] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which has an eigenvalue of 1 repeated three times. We call the number of times an eigenvalue is repeated the ***algebraic multiplicity***.

Exercise 9.8.1. Verify this above statement.

Theorem 9.8.1: Determinant, Trace, and Eigenvalues

Let $[A]$ be an $n \times n$ -matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ that occur m_1, m_2, \dots, m_k times respectively (i.e., the algebraic multiplicity), then we have

$$\det([A]) = \lambda_1^{m_1} \lambda_2^{m_2} \cdots \lambda_k^{m_k}$$

and

$$\text{tr}([A]) = m_1\lambda_1 + m_2\lambda_2 + \cdots + \lambda_k m_k.$$

The way we can think of this is that the eigenvalues tell us how different directions are stretched. The directions that are stretched are exactly the directions defined by the eigenvalues. Hence, it makes sense that the determinant is related to the eigenvalues.

[insert figure](#)

9.9 Diagonalization and Special Matrices

The eigenvalue problem allows us to do a bit more with matrices. We mentioned similar matrices previously, and one may wonder if given a matrix, if there is a similar matrix that is easier to understand. The easiest matrices to understand are the *diagonal matrices* which take the form

$$[\Lambda] = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

To see why these matrices are nice, we can take an example of a 3×3 -matrix times an arbitrary 3-dimensional vector. So we have

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \lambda_1 x \\ \lambda_2 y \\ \lambda_3 z \end{pmatrix}.$$

As we can see, this matrix simply scales each of the inputs separately. This gives us a transformation without rotation since it is only scaling. No components are mixed together either.

It was exactly this that the eigenvalue problem went to solve. Recall, we were given a matrix $[A]$ and were tasked with finding a eigenvector \vec{e} and corresponding eigenvalue λ so that

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

which says that the transformation $[A]$ simply scales the eigenvector \vec{e} by the eigenvalue λ . Hence, if we are given a matrix $[A]$, we can find the eigenvalues and eigenvectors and construct a diagonal matrix! Let us see how this shall work.

Consider for now the 3×3 -matrix case for $[A]$. Suppose that we find three eigenvalues λ_1, λ_2 , and λ_3 that correspond to the eigenvectors \vec{e}_1, \vec{e}_2 , and \vec{e}_3 respectively. Then we

can construct the matrix

$$[P] = \begin{pmatrix} | & | & | \\ \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ | & | & | \end{pmatrix}$$

Now, in order for this matrix to be invertible we must have that the eigenvectors are linearly independent which we can verify by checking that $\det([P]) \neq 0$. Then, when $[P]$ is invertible we can consider the following similar matrix

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

Let us see what happens when we consider multiplication by an arbitrary vector. We get

$$\begin{aligned} [P]^{-1}[A][P]\vec{v} &= [P]^{-1}[A] \begin{pmatrix} | & | & | \\ \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= [P]^{-1}[A](x\vec{e}_1 + y\vec{e}_2 + z\vec{e}_3) \\ &= [P]^{-1}(x[A]\vec{e}_1 + y[A]\vec{e}_2 + z[A]\vec{e}_3) \\ &= [P]^{-1}(x\lambda_1\vec{e}_1 + y\lambda_2\vec{e}_2 + z\lambda_3\vec{e}_3). \end{aligned}$$

Now, we just need to know what the matrix $[P]^{-1}$ does. To see what $[P]^{-1}$ does, we can just investigate $[P]$ acting on the unit basis vectors \hat{x} , \hat{y} , and \hat{z} . We have that

$$[P]\hat{x} = \vec{e}_1 \quad [P]\hat{y} = \vec{e}_2 \quad [P]\hat{z} = \vec{e}_3$$

and thus we have

$$\hat{x} = [P]^{-1}\vec{e}_1 \quad \hat{y} = [P]^{-1}\vec{e}_2 \quad \hat{z} = [P]^{-1}\vec{e}_3.$$

Thus we have

$$[P]^{-1}[A][P]\vec{v} = \lambda_1 x\hat{x} + \lambda_2 y\hat{y} + \lambda_3 z\hat{z} = \begin{pmatrix} \lambda_1 x \\ \lambda_2 y \\ \lambda_3 z \end{pmatrix}.$$

Hence it must be that $[P]^{-1}[A][P]$ is a diagonal matrix! Specifically we have

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

where

$$[\Lambda] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

When a matrix $[A]$ is similar to a diagonal matrix $[\Lambda]$ we say that $[A]$ is *diagonalizable*.

9.9.1 Hermitian Matrices

Some types matrices have extra properties that make them worth studying on their own. Specifically, when studying the physical world, we may want eigenvalues for a matrix to be real. This isn't always the case, but in quantum mechanics, for example, we will want energy eigenvalues to be real numbers. This property we care about arises in the symmetries of the inner product. We defined the dot product for the vector spaces \mathbb{R}^n , but we have a more general inner product for the vector spaces \mathbb{C}^n .

Definition 9.9.1: Inner Product for Complex Vectors

Let $\vec{u}, \vec{v} \in \mathbb{C}^n$ be n -dimensional vectors with complex entries. We define the **Hermitian inner product** of \vec{u} and \vec{v} to be

$$\langle \vec{u}, \vec{v} \rangle := \sum_{i=1}^n u_i v_i^*,$$

where v_i^* is the complex conjugate of the component v_i .

Take for example the vectors $\vec{u}, \vec{v} \in \mathbb{C}^2$ given by

$$\vec{u} = \begin{pmatrix} 2i \\ 1+i \end{pmatrix} \quad \text{and} \quad \vec{v} = \begin{pmatrix} 1-i \\ 3 \end{pmatrix}.$$

Then we can compute

$$\begin{aligned} \langle \vec{u}, \vec{v} \rangle &= \sum_{i=1}^2 u_i v_i^* \\ &= (2i)(1-i)^* + (1+i)(3)^* \\ &= (2i)(1+i) + (1+i)(3) \\ &= 2i - 2 + 3 + 3i \\ &= 5i + 1. \end{aligned}$$

One can note that this gives back the dot product when the components of the vectors are purely real. So, we are satisfied with taking this as our standard notion of an inner product. The symmetries of the inner product can be after defining the following.

Definition 9.9.2: Adjoint of a linear transformation

Given a linear transformation $A: \mathbb{C}^n \rightarrow \mathbb{C}^m$, the **adjoint** to A denoted by A^\dagger is given by the transformation satisfying

$$\langle A\vec{u}, \vec{v} \rangle = \langle \vec{u}, A\vec{v} \rangle.$$

This definition alludes intuition just slightly at the moment and begs the question as to what does the adjoint mean for a matrix. We let $[A]$ be the matrix representing the transformation A above and we have that each component is complex ($a_{ij} \in \mathbb{C}$). It's worth noting that since real numbers are merely a subset of the complex numbers, everything done here works for the real case as well. Given a matrix $[A]$, we define the of $[A]$ **transpose** to be the matrix $[A]^T$ whose entries are a_{ji} . That is, we swap rows with columns. For example, we have

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \Rightarrow \quad [A]^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

If we have that

$$[A] = [A]^T$$

we say that $[A]$ is ***symmetric***. Similarly, we have that the entries in the adjoint $[A]^\dagger$ are given by a_{ji}^* which means we take the transpose of $[A]$ and then complex conjugate the entries. That is, we have

$$[A] = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \implies [A]^\dagger = \begin{pmatrix} a_{11}^* & a_{21}^* & a_{31}^* \\ a_{12}^* & a_{22}^* & a_{32}^* \\ a_{13}^* & a_{23}^* & a_{33}^* \end{pmatrix}$$

When a matrix $[A]$ is equal to its adjoint $[A]^\dagger$ we say that $[A]$ is ***Hermitian*** or ***self adjoint***. More formally, a self adjoint linear transformation satisfies

$$\langle A\vec{u}, \vec{v} \rangle = \langle \vec{u}, A\vec{v} \rangle.$$

In the case that $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, this statement is true for symmetric matrices as real symmetric matrices are self adjoint!

Hermitian and symmetric matrices play an important role in physics and mathematics primarily due to their extremely nice spectral properties. In particular we have the following.

Theorem 9.9.1: Hermitian Matrices Have Real Eigenvalues

Let $[A]$ be an $n \times n$ -matrix with complex entries. We have that $[A]$ is Hermitian if and only if $[A]$ is diagonalizable and has only real eigenvalues.

Theorem 9.9.2: Orthogonal Eigenvectors for Different Eigenvalues

Let $[A]$ be Hermitian an $n \times n$ -matrix with complex entries. Then let \vec{e}_i and \vec{e}_j be eigenvectors corresponding to eigenvalues λ_i and λ_j such that $\lambda_i \neq \lambda_j$. Then we have that \vec{e}_i and \vec{e}_j are orthogonal (i.e., $\langle \vec{e}_i, \vec{e}_j \rangle = 0$).

Both of these theorems turn out to be exactly the conditions needed to properly build a quantum theory. Hence, studying Hermitian matrices is an important task all on its own.

Example 9.9.1: Hermitian Matrix

Consider the matrix

$$[A] = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Then we can show that $[A]$ is diagonalizable with real eigenvalues and orthogonal eigenvectors.

So we can find the characteristic polynomial by

$$\det([A] - \lambda[I]) = \lambda^2 - 1.$$

Hence we set this equal to zero and find that we get the two real eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$. Then we can find the eigenvector \vec{e}_1 by solving

$$([A] - \lambda_1[I])\vec{e}_1 = 0.$$

Specifically, we have the augmented matrix

$$[M] = \left(\begin{array}{cc|c} -1 & -i & 0 \\ i & -1 & 0 \end{array} \right)$$

Which we find gives a solution for the eigenvector

$$\vec{e}_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Similarly, we can get that the eigenvector corresponding to $\lambda_2 = -1$ is given by

$$\vec{e}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Then, we can see that these eigenvectors are orthogonal by taking

$$\langle \vec{e}_1, \vec{e}_2 \rangle = (-i)(i)^* + 1 = -1 + 1 = 0.$$

Now, letting

$$[P] = \begin{pmatrix} | & | \\ \vec{e}_1 & \vec{e}_2 \\ | & | \end{pmatrix} = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix},$$

we can compute and find

$$[P]^{-1} = \begin{pmatrix} \frac{i}{2} & \frac{1}{2} \\ \frac{-i}{2} & \frac{1}{2} \end{pmatrix}$$

Then we can compute

$$\begin{aligned} [P]^{-1}[A][P] &= \begin{pmatrix} \frac{i}{2} & \frac{1}{2} \\ \frac{-i}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ &= [\Lambda]. \end{aligned}$$

Let us briefly revisit the particle in the 1-dimensional box. There, we found states of the system were given by

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

We stated that the states were orthogonal when we considered that the integral

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

were $m \neq n$. This integral is exactly the same as the Hermitian inner product we defined above, but the underlying vector space is quite a bit different. This leads to an integral

versus a sum, and functions versus vectors. However, once one becomes more comfortable with the mathematics of vector spaces, the identification between these two seemingly different quantities becomes clear.

9.10 Groups and Symmetry

In the previous section we looked at a few special classes of matrices. These were the diagonalizable matrices, the symmetric matrices, and the Hermitian matrices. We found that they were all related in the following way.

- Real symmetric matrices are Hermitian matrices.
- All hermitian matrices are diagonalizable.

Thus, the largest class of matrices we discussed were the diagonalizable and the symmetric and Hermitian matrices are closely related. The main difference between the two is the entries we choose (i.e., real values versus complex values).

However, not all special matrices have to do with this type of structure. In fact, another large class of matrices has to do with the geometry of space. Recall that we built matrices in order to capture linear transformations. It turns out that linear transformations are operations that merely scale, reflect, or rotate vectors. The study of the eigenvalue problem seeks to breakdown this behavior some, but we can also take the approach to restrict our transformations from the beginning.

Take for example the unit circle in the real plane \mathbb{R}^2 . What symmetries does this circle have? In other words, what operations can we do to the whole plane that transform the circle back onto itself? For one, we can rotate the circle. We can also reflect the circle about any line through the origin. So, we may now ask, what matrices give us these symmetries of the circle

think about this with eigenvalues, eigenvectors, axes of an ellipse, and stuff

To visit this a bit, we can consider Example 9.1.2. With our tools now, we can say that the matrix for that linear transformation is given by

$$[\text{Rot}]_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

rotates the the unit circle by an angle θ in the counterclockwise direction. Similarly, we can define a reflection matrix

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

which reflects the circle about a line L passing through the origin that makes an angle θ with the x -axis.

insert visual for both of these two transformations.

One can think logically about the rotations and reflections of the plane \mathbb{R}^2 and notice that if we take two different angles to rotate by (θ and φ), then we can apply both the transformations to get

$$[\text{Rot}]_\theta [\text{Rot}]_\varphi = [\text{Rot}]_{\varphi+\theta}.$$

Exercise 9.10.1. Perform the matrix multiplication above to see that the statement is true.

In fact, we can actually show the following are also true.

$$\begin{aligned} [\text{Ref}]_\theta[\text{Ref}]_\varphi &= [\text{Rot}]_{2(\theta-\varphi)} \\ [\text{Rot}]_\theta[\text{Ref}]_\varphi &= [\text{Ref}]_{\varphi+\theta/2} \\ [\text{Ref}]_\varphi[\text{Rot}]_\theta &= [\text{Ref}]_{\varphi-\theta/2}. \end{aligned}$$

What we notice here is that a product of reflections is a rotation, a product of rotations is a rotation, and a product of a rotation with a reflection gives a reflection. We can also, given any rotation or reflection, undo the process to get back to where we started. That is also to say that the identity matrix $[I]$ is a trivial reflection or rotation.

This structure that is being described is that of a **group**. A group is a set G that arises as the symmetries of some object or set. Said in a more mathematical way, a group is:

- A set G of elements.
- There is a group operation $*$ between elements $g, h \in G$ that give a new element in the group. That is

$$g * h \in G.$$

- There is a unique identity element $e \in G$ such that $eg = ge$ for any $g \in G$.
- For each element $g \in G$ there exists g^{-1} such that

$$gg^{-1} = g^{-1}g = e.$$

Proposition 9.10.1: Rotations and Reflections Form a Group

Consider the set G of rotation and reflection matrices described above. Then this set G is a group with group operation given by matrix multiplication (hence we do not write the asterisk in this case). In particular, this group is the **orthogonal group** $O(2)$.

Proof. By the rules for multiplication described above, we have that if we take two matrices $[g], [h] \in O(2)$ then $[g][h] \in O(2)$ since it is also a reflection or rotation matrix. Then, we have that the identity element is given by the identity matrix $[I] \in O(2)$ and we have already shown that the identity matrix fixes matrices and vectors. Lastly, using the multiplication rules above, if given a $[g] \in O(2)$, we can construct a $[g]^{-1} \in O(2)$. Indeed, let $[g] = [\text{Rot}]_\theta$, then $[g]^{-1} = [\text{Rot}]_{-\theta}$. Similarly, if $[g] = [\text{Ref}]_\theta$, then $[g]^{-1} = [\text{Ref}]_{-\theta}$ as well. Thus, $O(2)$ is a group. \square

One may wonder why this group of matrices is referred to as orthogonal. The reason why this is the case is that the columns of a matrix in $O(2)$ are orthonormal. We can show this explicitly by first taking the columns of $[\text{Rot}]_\theta$. This gives us the two vectors

$$\vec{u} = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}.$$

Then note that we get

$$\vec{u} \cdot \vec{v} = -\cos \theta \sin \theta + \sin \theta \cos \theta = 0.$$

Hence the columns are orthogonal. But, moreover the columns are also normalized. We can take

$$\vec{u} \cdot \vec{u} = \cos^2 \theta + \sin^2 \theta = 1,$$

and we see this is true for the second column \vec{v} as well. This can all be summed up by the following rule.

- A matrix $[g]$ is in $O(2)$ if and only if

$$[g][g]^T = [I] = [g]^T[g].$$

This statement are analogous to having orthonormal columns. Why is that? Well, when we multiply a matrix $[g]$ by its transpose, the product matrix is exactly the dot products of all the columns. That is, if we have

$$[g] = \begin{pmatrix} | & | \\ \vec{u} & \vec{v} \\ | & | \end{pmatrix},$$

then we have

$$[g]^T[g] = \begin{pmatrix} \vec{u} \cdot \vec{u} & \vec{u} \cdot \vec{v} \\ \vec{v} \cdot \vec{u} & \vec{v} \cdot \vec{v} \end{pmatrix}.$$

Hence if this is the identity matrix, it must be that the columns are orthonormal! It also follows that the determinant of an $O(2)$ matrix is equal to ± 1 .

Living inside of $O(2)$ is the group of rotation matrices. This was due to the fact that a product of rotations generated another rotation. However, reflections do not form a group on their own right since a product of reflections can generate a rotation. We denote the group of rotations of the plane by $SO(2)$ which is the *special orthogonal group*. It turns out that all the elements of $SO(2)$ are the matrices in $O(2)$ that have a determinant of $+1$. We say that $SO(2)$ is a *subgroup* of $O(2)$.

While groups can be constructed in complete generality, they are more often viewed as matrices since the computation with matrices is straightforward. Not to mention, the tools from linear algebra assist us heavily with the theory of groups. Groups are part of a field of mathematics known as *abstract algebra* in which you will see structures such as rings, fields (these were briefly mentioned before), modules, and algebras (which square matrices are an example of). There are other structures that are also studied, but are not quite as common.

One may ask about what other meaningful groups there are for a chemist to study? The answer is yes. There is a lot to be gained by understanding the symmetries of systems, and when we're studying symmetries we are studying groups. In the field of organic chemistry, understanding molecular symmetries is important.

Take for example the symmetries of a square lying in the plane. Since the square is a subset of the plane, we can make the guess that the symmetries of the square would form as a subgroup to $O(2)$ which is exactly the case. Let us see what operations are symmetries of the square.

insert figure here

So we have that the rotational symmetries of square are the identity transformation $[I]$, a $\pi/2$ rotation, a π rotation, a $3\pi/2$ rotation. Rotating by any other angles would not leave the square invariant! Hence these are given by the matrices

$$[I], \quad [\text{Rot}]_{\pi/2}, \quad [\text{Rot}]_\pi, \quad [\text{Rot}]_{3\pi/2}.$$

We also have the reflections about the horizontal and vertical bisecting lines as well as the diagonal bisecting lines. These would be given by the reflection matrices

$$[\text{Ref}]_0, \quad [\text{Ref}]_{\pi/2}, \quad [\text{Ref}]_{\pi/4}, \quad [\text{Ref}]_{3\pi/4}.$$

Hence, there are eight elements in the symmetry group of the square! This group is known as the **dihedral group of order 8** D_8 . Note that the dihedral group of order $2n$ is the symmetry group of a regular n -gon in the plane (e.g., the dihedral group of order 6 is the symmetry group of an equilateral triangle).

Exercise 9.10.2. Using the multiplication rules for the $O(2)$ matrices, can you deduce that we can create any element in D_8 from multiplying two specific elements in different ways. For example, we can create the $[\text{Rot}]_\pi$ matrix by taking $[\text{Rot}]_{\pi/2}[\text{Rot}]_{\pi/2}$. We call these two elements you find here the **generators** of a group.

We are also not limited to two spatial dimensions. For example, we can create any rotation in 3-dimensional space by using the matrices

$$\begin{aligned} [\text{Rot}_x]_\theta &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \\ [\text{Rot}_y]_\theta &= \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \\ [\text{Rot}_z]_\theta &= \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \end{aligned}$$

which rotate vectors by an angle θ about the x -axis, y -axis, and z -axis respectively. Using these three matrices, one can create any element of the group $SO(3)$ which are the rotations of 3-dimensional space. In other words, this is the group of rotational symmetries of the two dimensional sphere. In general, we can think of the rotations and reflections of n -dimensional space as living in the group $O(n)$. The matrices in $O(n)$ with determinant +1 form the $SO(n)$ group, which are again just rotations.

The largest matrix group we can create would be the group of invertible $n \times n$ -matrices. We refer to this group as the **general linear group** and write $GL(n)$. One usually specifies the entries that go into the matrix (i.e., real or complex) in order to fully describe what matrices we are referring to. We can see why this is a group fairly easily since we know $[I]$ is an invertible matrix and we know that every matrix has an inverse. In fact, we have that any matrix group sits inside of $GL(n)$ as a subgroup!

The last groups that are worth mentioning are the complex versions of the groups $O(n)$ and $SO(n)$. These groups are the **unitary group** $U(n)$ and the **special unitary group** $SU(n)$. Instead of having real numbers as inputs, they have complex numbers. Roughly speaking, $U(n)$ essentially describe the rotations and reflections not of \mathbb{R}^n , but of \mathbb{C}^n and $SU(n)$ captures just the rotations.

We have that a matrix $[g] \in U(n)$ if

$$[g]^\dagger [g] = [I] = [g][g]^\dagger.$$

This is telling us that the columns of $[g]$ are orthonormal but with respect to the Hermitian inner product! This is the only difference between $\mathrm{U}(n)$ and $\mathcal{O}(n)$. It follows that for $[g] \in \mathrm{U}(n)$ that $\det([g])$ is a complex number with modulus one. That is,

$$\det([g]) = e^{i\theta},$$

for some θ . Then, we can define the $\mathrm{SU}(n)$ matrices to be the matrices in $\mathrm{U}(n)$ whose determinants are equal to one.

Groups are their own field of study and they are massively useful objects. We have only briefly scraped the surface, and the spectrum of application is broad. Though you may not explicitly do group theory, groups will still show up. Learning more about them and expanding your horizons will make the journey much more interesting and fulfilling!

Remarks for Math 271

So far, we have covered a broad spectrum of topics. Many of these topics are not typical for a calculus II student, while some were. Instead of taking the typical path through calculus II, then calculus III, we are venturing in a way that allows us to tell a story and build a firm foundation in mathematics specifically used throughout chemistry (and especially physical chemistry). Typically, students in the calculus sequence learn differentiation and integration in a first course. In a second course, students learn about sequences, series, power series, and Taylor series. The third course entails learning calculus in higher dimensions. Specifically, one concentrates on calculus in 3-dimensional space since this is (arguably) the most physically meaningful to us. Following the calculus sequence comes a course in differential equations where students are granted a handbook of techniques for solving different types of equations.

We approached this in a completely different way. We began by studying complex numbers as this field of numbers allows us to factor polynomials. The complex numbers also formed a vector space, and we briefly investigated this structure. However, the main goal was to use these complex numbers in order to be able to solve many different first and second order differential equations. These differential equations arose by studying systems that change over time. For example, we saw the harmonic oscillator equation arise from a spring/mass system. We also saw that chemical reactions were nicely described by differential equations. Those equations were describing systems that evolved over time and were given with initial function values at time zero. Another way differential equations arose was via boundary value problems. In particular, we studied the free particle in a one-dimensional box and came across many new concepts that resurface later.

Quickly, one sees that the techniques we used to solve differential equations were not all powerful. Thus, we sought out a new technique to solve more equations. Eventually, we arrived at power series which provided us with newfound abilities to compute. For example, one needs a tool such as a power series to compute values to the function $\sin(x)!$ Power series then proved as indispensable tools for solving more differential equations

than we were previously able to work with. When we still had trouble with a specific equation, we could then use a Taylor series to find polynomial approximations of terms in a differential equation and from there we could solve this equation using power series. This part of the class was closer to a typical calculus II course but came with the added bonus of solving differential equations as well. Even in a first course on ordinary differential equations, one may not solve equations using a power series!

Finally, the last portion of this class came as a preparation to deal with higher dimensional spaces. Understanding vector spaces and how they transform is a necessary building block to studying calculus in higher dimensions. However, linear algebra can be done in more generality than we have covered here. In this generality, we can see how topics we have previously worked with fit in as well. For example, we spoke of linear differential equations and one can show that the set of solutions to a linear differential equation forms a vector space! Either way, the lessons that linear algebra teaches us about geometry and using geometrical tools to solve problems is highly important. This can lead one to considering abstracting algebra further and seeing if that can help solve problems as well.

Next, we move onto the sequel of Math 271 which is Math 272 where we will spend time learning calculus in higher dimensions. In order to model the physical world, this is completely necessary. The topics covered in 271 will lead straight into the new mathematics awaiting us in 272. Keep these previous chapters as a reference as they are now prerequisite material!

Part IV

Calculus in Higher Dimensions

10

Curves, scalar fields, and vector fields

10.1 Overview of higher dimensional functions

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

$$\begin{aligned}\vec{\gamma}: \mathbb{R} &\rightarrow \mathbb{R}^3 \\ f: \mathbb{R}^3 &\rightarrow \mathbb{R} \\ \vec{V}: \mathbb{R}^3 &\rightarrow \mathbb{R}^3.\end{aligned}$$

Of these three functions, the first is called a *curve* the latter two are often referred to as *fields* due to their inherent attachment to geometric objects. These fields are also *multivariate* functions since their inputs depend on points in \mathbb{R}^3 , whereas the first only has an input of a single variable that we typically think of as time.

A point \vec{x} in space \mathbb{R}^3 , or more generally, a point in \mathbb{R}^n is given by specifying n coordinates. For example, in space we may take

$$\vec{x} = (x, y, z) = x\hat{x} + y\hat{y} + z\hat{z} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

All of these forms of notation are equivalent in our case. The latter two were what we typically used in the prequel, but we find utility in the slightly more compact notation (x, y, z) when specifying input to a function.

- (1) Once again, functions that assume the form

$$\vec{\gamma}: \mathbb{R} \rightarrow \mathbb{R}^3$$

are curves. More generally, we can have a curve take input values from a restricted domain and the output can lie in an arbitrary dimension. Hence, we may put

$$\vec{\gamma}: [a, b] \rightarrow \mathbb{R}^n.$$

Another specific example would be a *planar curve* for which we take $n = 2$. Note, we can always take $a = 0$ and $b = 1$ to yield a planar curve

$$\vec{\gamma}: [0, 1] \rightarrow \mathbb{R}^2.$$

Curves arise in many scenarios. We plot curves when we first learn about functions. We solve for curves when we find solutions to ODEs. Curves also provide a mathematician a way for exploring a space of interest. Often times we think of $\vec{\gamma}(t)$ as describing a particle or observer's position at the time t .

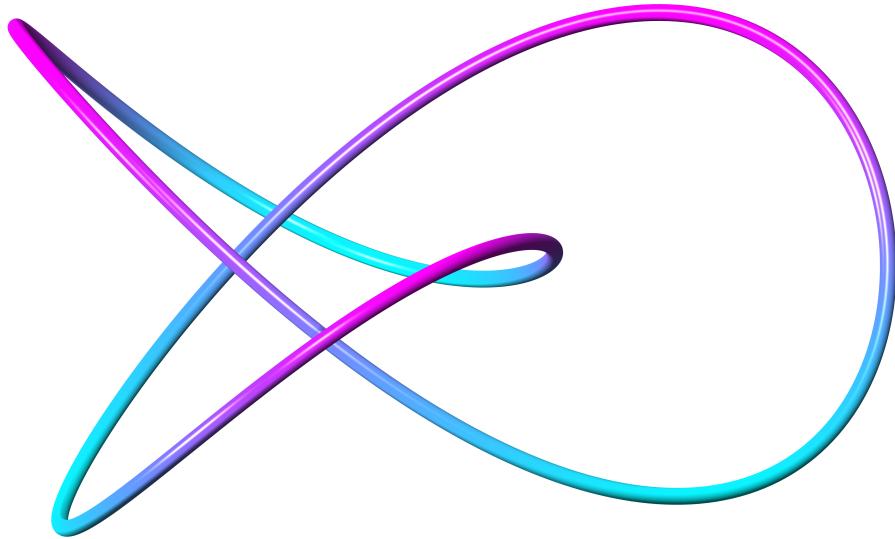


Figure 10.1: A curve called the trefoil knot. Notice how the curve wraps around itself in an interesting way (notice the lighting on each portion of the curve to see these overlaps). The reader can try to recreate this curve with a string!

- (2) Functions of the form

$$f: \mathbb{R}^3 \rightarrow \mathbb{R}$$

are *scalar fields*. These types of functions are useful in describing quantities like temperature in space. In this example, at each point $\vec{x} = (x, y, z)$ in 3-dimensional space \mathbb{R}^3 , we can assign a single real number output $f(\vec{x}) = f(x, y, z)$ that tells us this temperature. Or, we may find that we restrict ourselves to a specific domain Ω that lives in space. In that case, Ω may represent a physical object for which we want to describe temperature. Hence, we have

$$f: \Omega \rightarrow \mathbb{R}$$

is a scalar field defined on Ω . For a point $(x, y, z) \in \Omega$, the value of f at that point is $f(x, y, z)$. This is where we see a slight distinction between functions and fields. Fields arise from a geometrical object (e.g., space itself or an object embedded in space). Once again, if we are thinking of f as describing temperature, we may want to describe how this temperature field change over time. Abstractly, our function then assumes the form

$$f: \Omega \times [0, \infty) \rightarrow \mathbb{R},$$

and the value of f at $(x, y, z) \in \Omega$ and time t is $f(x, y, z, t)$. We will return to the time dependent fields later as we visit Partial Differential Equations (PDEs).

To advance our intuition of functions we have used techniques for visualization. For example, with a function $f: \mathbb{R} \rightarrow \mathbb{R}$ we drew the graph of the function by plotting the points $(x, f(x))$ for all values of x . Sadly, full visualization of 3-dimensional scalar fields would require us to visualize 4-dimensions. Many examples we will do will instead take 2-dimensional scalar functions in order to build intuition.

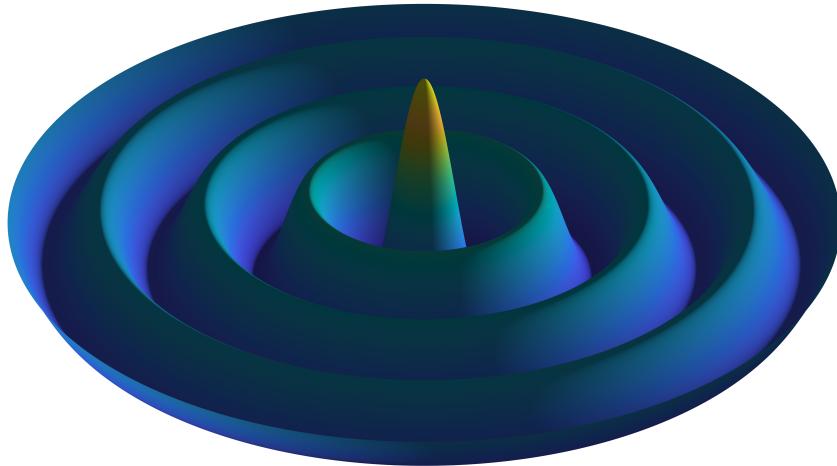


Figure 10.2: The graph of a radial Bessel function $J_0(\rho)$ where $\rho = \sqrt{x^2 + y^2}$. This is a surface whose height is the value of $J_0(\rho)$. This function can be used to describe ripples of water.

(3) Functions of the form

$$\vec{V}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

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

Vector fields can also be defined on regions Ω in space or in lower or dimensional settings. Visualization of 2-dimensional vector fields is easiest, but we lose out on lots of more interesting configurations and applications.

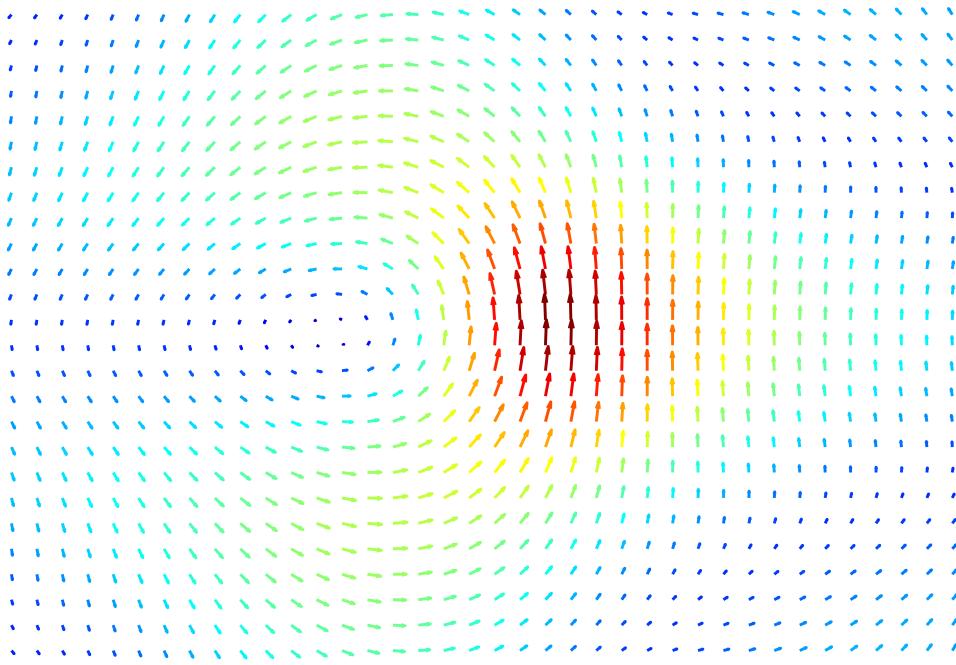


Figure 10.3: A plot of a vector field using color to help emphasize the magnitude of the vectors at each point. Note the swirling created by the changing direction of the vectors and the spots where the arrows are longer and where they are shorter.

The notation we use on the types of functions is telling. Curves $\vec{\gamma}$ are indeed vectors; at a time t they tell us the position of a point in space $\vec{\gamma}(t)$. Scalar fields are scalar valued though their input is a vector. So $f(x, y, z)$ is simply a number assigned to a point in space. Finally, a vector field inputs and outputs a vector. At a point (x, y, z) we will place a new vector $\vec{V}(x, y, z)$.

We can wrap all of these up in a fluid example. Curves are the trajectories of single particles, scalars describe quantities of fluids (pressure, temp, height), and vectors show the whole group flow

10.2 Curves in space

Recall that a curve in space is a function of the form

$$\vec{\gamma}: \mathbb{R} \rightarrow \mathbb{R}^3.$$

We will specify a specific curve by supplying three functions $\gamma_1(t)$, $\gamma_2(t)$, and $\gamma_3(t)$. Specifically, each of these functions γ_i is a function $\gamma_i: \mathbb{R} \rightarrow \mathbb{R}$. So, a curve in 3-dimensions is made up of three single variable functions. Then, we can say that

$$\vec{\gamma}(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \gamma_3(t) \end{pmatrix}.$$

Each $\gamma_i(t)$ (for the values $i = 1, 2, 3$) represents the components of a vector that changes with respect to the variable t . Intuitively, we like to think of the components as evolving

in time, which is why we make use of the variable t . To say this more explicitly, we can write

- $\gamma_1(t)$ the x -position of γ at time t ,
- $\gamma_2(t)$ the y -position of γ at time t ,
- $\gamma_3(t)$ the z -position of γ at time t .

Example 10.2.1: A Planar Curve

To see an example, let us take a curve $\vec{\gamma}: [0, 1] \rightarrow \mathbb{R}^2$ by defining

$$\vec{\gamma} = \begin{pmatrix} t \\ t \end{pmatrix}.$$

This gives us the components

$$\gamma_1(t) = t \quad \text{and} \quad \gamma_2(t) = t.$$

We can plot this curve in the plane by plotting the vector for each time t . This gives us

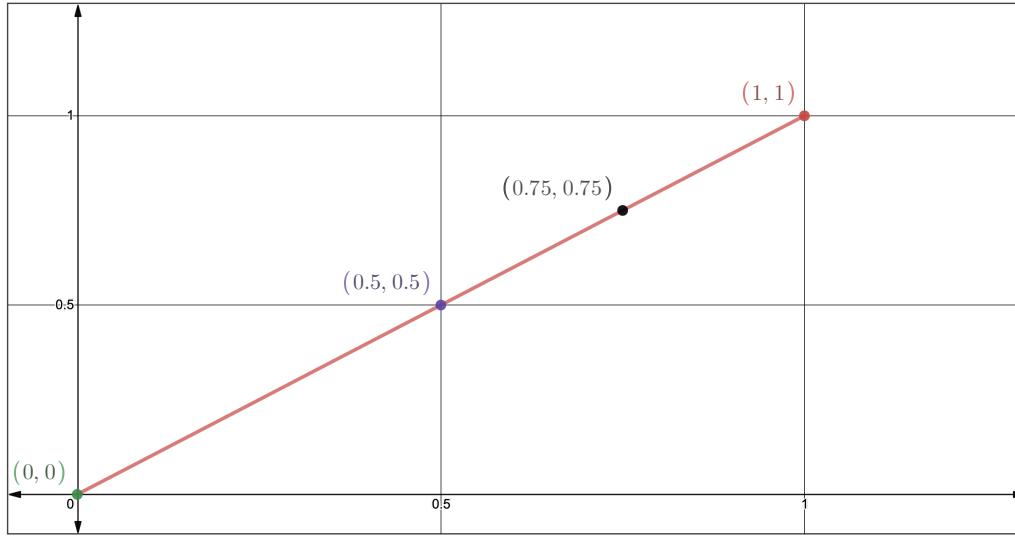


Figure 10.4: The curve $\vec{\gamma}(t)$ with labeled points at different times.

Replace figure with a color varying curve In this picture, we can see that the curve begins at the the origin and ends at the point $(1, 1)$. Specifically, we have

$$\vec{\gamma}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \vec{\gamma}(1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This curve does not extend indefinitely as our input values can only take on the values in $[0, 1]$.

From the example, we can see that curves also have a direction associated to them. As we increase the parameter t , we move along the curve in a certain way. This is an important concept! We can use curves to describe the motion of an object (or other quantities) over time. This also brings to mind the notion of velocity or speed for a curve.

10.2.1 Derivatives of Curves

The nice thing about curves is that each component function is a single variable function $\gamma_i: \mathbb{R} \rightarrow \mathbb{R}$. This means that we can compute the derivatives of each component with the knowledge we have from single variable calculus. In fact, the derivative of a curve can be computed using the typical difference quotient that we learn in a first semester calculus course.

Imagine that a curve $\vec{\gamma}(t)$ describes the position of a particle at the time t . Then, one could ask for the velocity of this curve which is the first time derivative of position. Here, we can compute this first derivative at the time t by

$$\dot{\vec{\gamma}}(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{\gamma}(t + \Delta t) - \vec{\gamma}(t)}{\Delta t}.$$

We use the overdot notation for $\dot{\vec{\gamma}}$ to represent a time derivative of the curve. We then refer to $\dot{\vec{\gamma}}(t)$ as the **tangent** or **velocity vector** to the curve $\vec{\gamma}$ at a time t . Now, let us compute the tangent vector explicitly. We take

$$\begin{aligned}\dot{\vec{\gamma}}(t) &= \lim_{\Delta t \rightarrow 0} \frac{\vec{\gamma}(t + \Delta t) - \vec{\gamma}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left(\begin{pmatrix} \gamma_1(t + \Delta t) \\ \gamma_2(t + \Delta t) \\ \gamma_3(t + \Delta t) \end{pmatrix} - \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \gamma_3(t) \end{pmatrix} \right) \\ &= \lim_{\Delta t \rightarrow 0} \begin{pmatrix} \frac{\gamma_1(t + \Delta t) - \gamma_1(t)}{\Delta t} \\ \frac{\gamma_2(t + \Delta t) - \gamma_2(t)}{\Delta t} \\ \frac{\gamma_3(t + \Delta t) - \gamma_3(t)}{\Delta t} \end{pmatrix} \\ &= \begin{pmatrix} \dot{\gamma}_1(t) \\ \dot{\gamma}_2(t) \\ \dot{\gamma}_3(t) \end{pmatrix}.\end{aligned}$$

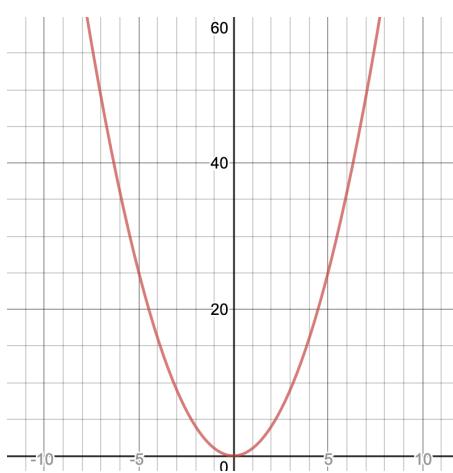
What this shows is that the tangent vector to a curve $\vec{\gamma}$ is the vector containing the derivative of the components of $\vec{\gamma}$. Intuitively, this says that the velocity of the curve is found by computing the rate of change of the x -component, y -component, and z -component of $\vec{\gamma}$.

Example 10.2.2: Graph of a Quadratic Function

Consider the curve $\gamma: \mathbb{R} \rightarrow \mathbb{R}^2$ given by

$$\vec{\gamma}(t) = \begin{pmatrix} t \\ t^2 \end{pmatrix}$$

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



What is the tangent vector at time t ? We have

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} 1 \\ 2t \end{pmatrix}.$$

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

Include better visualization with tangent vectors

The two examples of curves we have seen so far mimic graphs of real valued functions we have seen before. But, curves can be much more general. There is no need for a curve to pass the horizontal line test, and curves can even have self intersection! For example, a particle that orbits a large body such as the Earth orbits the Sun will move in an elliptical path.

Example 10.2.3: Circle Curve

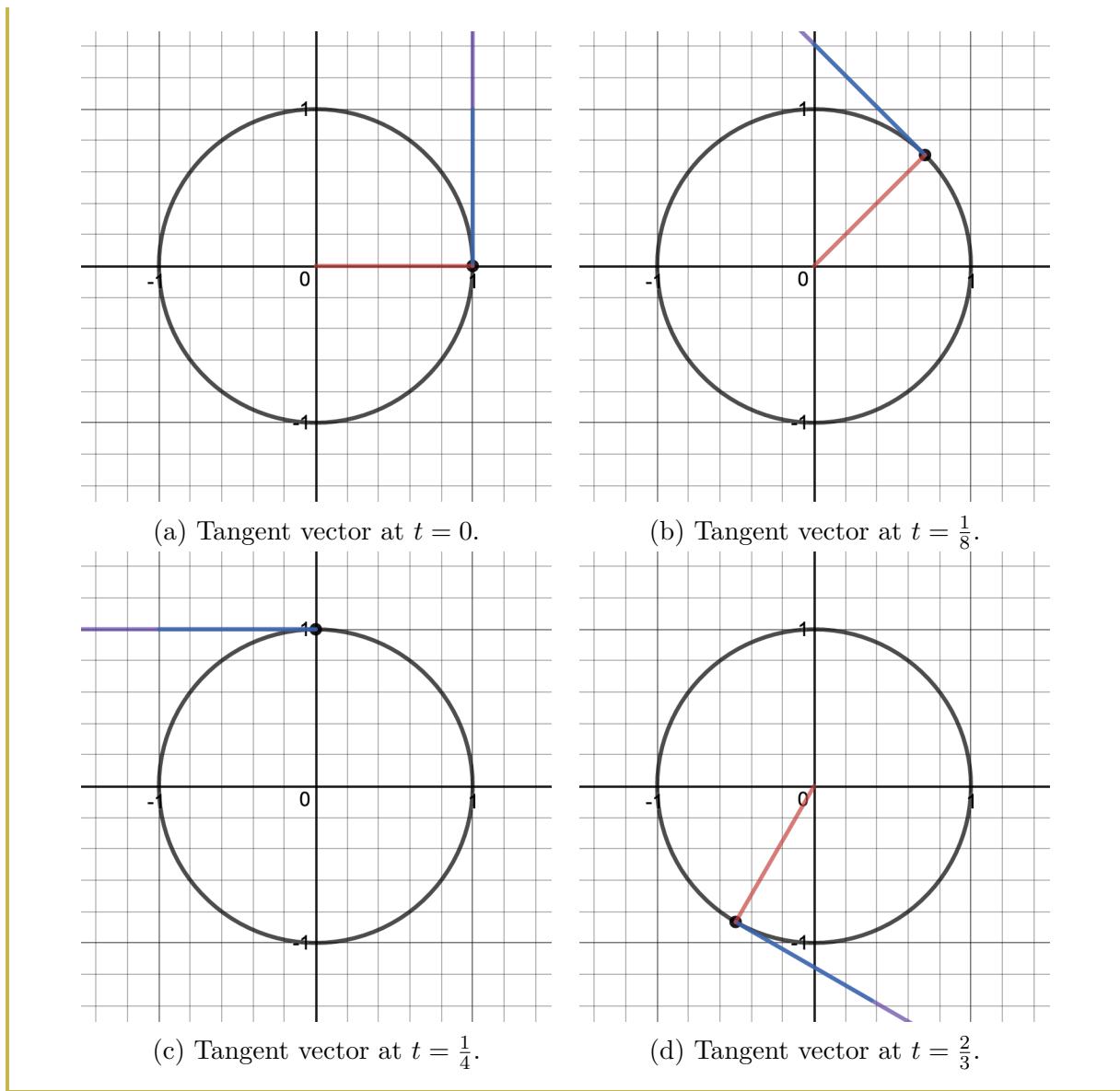
Consider the curve $\gamma: [0, 1] \rightarrow \mathbb{R}^2$ given by

$$\vec{\gamma}(t) = \begin{pmatrix} \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix}.$$

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

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} -2\pi \sin(2\pi t) \\ 2\pi \cos(2\pi t) \end{pmatrix}.$$

See the following graphs



There is nothing stopping us from computing the second time derivative of a curve as well. We repeat the process for computing the first derivative above, but instead we take the time derivative of the tangent vector $\vec{\gamma}$. This yields,

$$\ddot{\vec{\gamma}}(t) = \begin{pmatrix} \ddot{\gamma}_1(t) \\ \ddot{\gamma}_2(t) \\ \ddot{\gamma}_3(t) \end{pmatrix}.$$

We then refer to this quantity as the *normal* or *acceleration vector* to the curve $\vec{\gamma}$.

Exercise 10.2.1. Compute the normal vector to the circle curve in the previous exercise. If you know of the notion of centripetal force, how does this quantity relate?

10.2.2 Lengths of Curves

Integration is a tool we use to add up values of functions. In the 1-dimensional case, we found that the integral of a function $f(x)$ from $x = a$ to $x = b$ computed the net area

under the graph of the function. Similarly, we can use integration to compute the length of a curve.

If we take a curve $\vec{\gamma}: [a, b] \rightarrow \mathbb{R}^3$, then we can compute the tangent vector at time t , $\dot{\vec{\gamma}}(t)$. Again, the tangent vector is analogous to the velocity of a curve (at some point in time), and so the length of this tangent vector is the speed. So, we put that the speed of the curve at time t is $|\dot{\vec{\gamma}}(t)|$. Note that $|\dot{\vec{\gamma}}(t)|$ is a single variable function so we do not need any special tools to integrate this.

Computing the length of a curve amounts to adding up the speed of the curve over the total amount of time. In this perspective, this just takes into account the total distance a particle has moved, even if the particle back-tracks at some point. We can write this length as

$$\ell(\vec{\gamma}) = \int_a^b |\dot{\vec{\gamma}}(t)| dt.$$

Example 10.2.4: Circumference of a Circle

Take for example the circle curve $\vec{\gamma}: [0, 1] \rightarrow \mathbb{R}^2$ given by

$$\vec{\gamma}(t) = \begin{pmatrix} \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix}.$$

Then, we found

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} -2\pi \sin(2\pi t) \\ 2\pi \cos(2\pi t) \end{pmatrix}.$$

Taking the length of the tangent vector at time t yields

$$|\dot{\vec{\gamma}}(t)| = \sqrt{4\pi^2 \sin^2(2\pi t) + 4\pi^2 \cos^2(2\pi t)} = 2\pi.$$

The length is then

$$\ell(\gamma) = \int_0^1 2\pi t dt = 2\pi,$$

which is indeed the circumference of a circle of radius one.

10.3 Scalar fields

The next major class of functions we will consider are the scalar fields. That is, functions that take the form

$$f: \mathbb{R}^3 \rightarrow \mathbb{R}.$$

We may also find it helpful to visualize functions by considering instead

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}$$

and looking at the **graph** of f much like we consider the graph of functions $f: \mathbb{R} \rightarrow \mathbb{R}$. Let us break down this idea in the 1-dimensional case first.

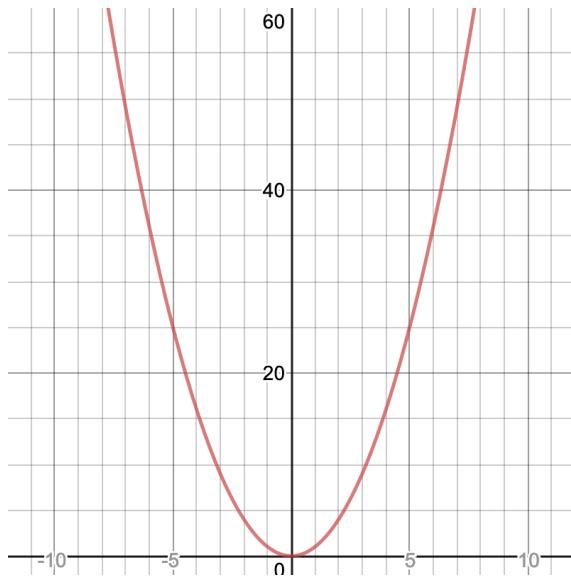
Example 10.3.1: Graph of a function

Whenever we talk of a function of the form

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

we inherently tend to draw the graph of the function f . By graph, I mean that given f , we usually just draw the curve $(x, f(x))$ in the plane. Note that this is indeed a curve by the definition we saw prior.

Take for example, $f: \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = x^2$. We usually draw the plane \mathbb{R}^2 and graph the curve $(x, f(x))$ which looks like



All of this is to say that graphs of these types of functions are special kinds of curves. In other words, they are curves that pass the vertical line test.

The 1-dimensional examples are rather redundant. In a vague sense, it can be hard to discern a difference between 1-dimensional curves, scalar fields, and vector fields. However, if we consider graphing higher dimensional scalar fields, we will immediately see a difference. Our new point of view will be to look at the graph of 2-dimensional scalar fields in order to gain intuition on the 3-dimensional scalar fields.

Graphing a 1-dimensional scalar field f involved plotting the graph

$$(x, f(x)).$$

Here, we put the y -component of the output as the function $f(x)$ which allows us to see the function values at a point x as the height above (or below) the x -axis. Similarly, if we have a scalar field $h(x, y)$, we can plot the set of points

$$(x, y, h(x, y)),$$

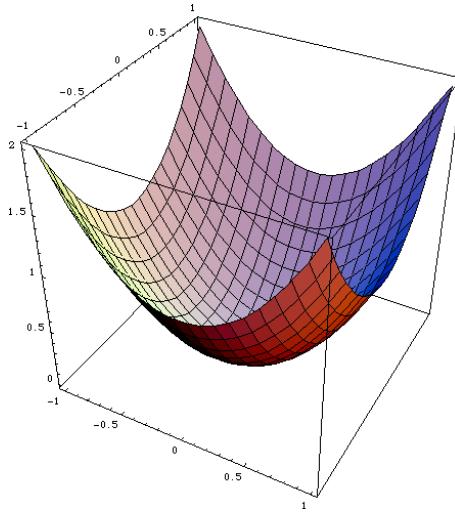
in 3-dimensional space. Then, what we receive is a picture that places $z = h(x, y)$ so that we see the function h describing the height of a rubber sheet above the xy -plane. One can imagine that the function h describes how this rubber sheet is deformed. This graph is an example of a *surface* which we will discuss later in this text.

Example 10.3.2: The Paraboloid

Let $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ be given by

$$h(x, y) = x^2 + y^2.$$

We can plot the graph of the function by plotting $(x, y, h(x, y))$ in \mathbb{R}^3 . This will look like:



Then we can analyze this function in a few nice ways. For example, if we fix a value for x or y , then we will be able to look at h as a function of just a single variable. This will also give us a cross section of the graph. We can imagine that, say setting $y = 0$ looks at the slice of the graph where $y = 0$ and x is allowed to vary.

- Let us take $y = 0$, then we have

$$h(x, 0) = x^2.$$

So along the $y = 0$ line, the function is just the parabola we are used to! Feel free to repeat this for other values of y .

- Similarly, we can force $x = 0$ and arrive at

$$h(0, y) = y^2$$

is also a parabola. Again, you should repeat this for various values of x .

- But we are not limited to these choices. We could have chosen $y = 5$ and we would have

$$h(x, 5) = x^2 + 25$$

which is a parabola shifted upwards by 25 units.

- Again, we could also choose yet another “slice” of this function and let $x = y$ which would give us

$$h(x, x) = x^2 + x^2 = 2x^2.$$

So, along the $x = y$ line, the parabola is scaled by 2. Any number of options are available to you here.

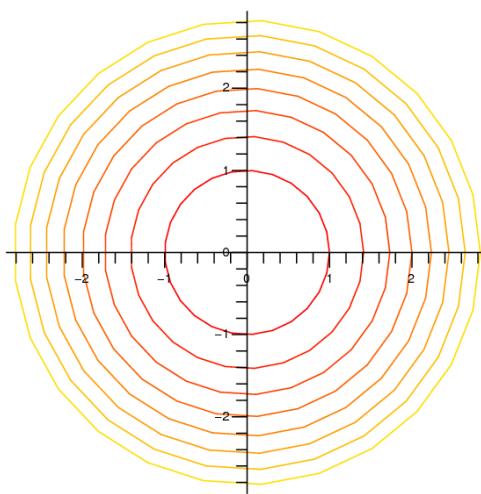
One other method of analyzing this function would be to find what the *level curves* of this function are. What is the set of points (x, y) that satisfy the equation $h(x, y) = c$? We call these level curves much in the way that a topographical map plots curves along the areas with equal height. For the previous example with $h(x, y) = x^2 + y^2$, consider the set of points (x, y) so that $h(x, y) = 1$. This means

$$f(x, y) = 1 = x^2 + y^2.$$

Since we have

$$x^2 + y^2 = 1$$

that means each level curve is a circle! Here is a “topographical map” for this function (i.e., a plot of the level curves for this paraboloid.)



Here, each circle represents $h(x, y) = c$ for different values of c . This really *is* a topographical map!

Exercise 10.3.1 (Plane). Repeat this analysis for yourself for the following function:

$$f(x, y) = x + y.$$

An important remark is that we should only use this idea to gain intuition. For one, it fails when our function is given by $f: \mathbb{R}^n \rightarrow \mathbb{R}$ where $n > 2$. The objects for calculus that we develop for scalar fields do not take into account the graph of the function, but just the function itself.

Replace previous figures and discuss how coloring can be used to build intuition (based on same idea as level curves)

10.3.1 Partial derivatives of scalar fields

Functions of multiple variables are quite a bit more interesting than their single variable counterparts. As the previous example shows, the visualization becomes more involved and we find that we can, for example, break down the scalar fields into “slices”. This means that we must develop more ways to analyze these higher dimensional functions.

If we take a scalar field $f(x, y, z)$, then we can investigate the rate of change that this function incurs in all possible directions. This of course brings us to the notion of taking

derivatives. In particular, we will take derivatives with respect to the individual input variables.

Definition 10.3.1: Partial Derivatives

Let $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ be a scalar field. We define the ***partial derivative with respect to x at the point (x_0, y_0, z_0)*** , denoted $\frac{\partial f}{\partial x}(x_0, y_0, z_0)$, and put

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0) := \lim_{\delta \rightarrow 0} \frac{f(x_0 + \delta, y_0, z_0) - f(x_0, y_0, z_0)}{\delta}$$

Remark 10.3.1. For partial derivatives, all but one variable are being held constant. So, when you are computing these, be sure to treat the proper variables as constant when necessary.

To make notation easier to deal with, we will often just write

$$\frac{\partial f}{\partial x}$$

in place of

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

When we specify a point at which this partial derivative should be computed, then we must use the notation

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0),$$

so that the meaning is unambiguous. This definition is readily extended to the other input variables.

Exercise 10.3.2. Define $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$ in a similar way to the above definition.

Example 10.3.3: Computing partial derivatives

Consider again the function $h(x, y) = x^2 + y^2$ which we analyzed in depth earlier. Then, we can compute the partial x and partial y derivatives.

To compute $\frac{\partial h}{\partial x}$, we will hold the variable y constant and take the derivative of h with respect to x . Thus, we have

$$\frac{\partial h}{\partial x} = 2x.$$

Since y was constant, the derivative must be zero. On top of this, we looked at the behavior of h when we held y constant and found that we get a parabolic function. Indeed, the function $f(x) = x^2$ has a derivative $2x$ just like we found $\frac{\partial h}{\partial x} = 2x$. The partial derivative is describing the profile of the graph of h when y is held constant! We repeat this but with respect to the input variable y now. Thus,

$$\frac{\partial h}{\partial y} = 2y$$

since x is treated as a constant. Once again, we see the function $h(x, y)$ looks like a parabola as we slice along constant values of x !

Exercise 10.3.3. Compute $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ for the function

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

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

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

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

- $\frac{\partial^2 f}{\partial x^2}$ tells us about the concavity (or curvature) of f as we move in the x direction.
- $\frac{\partial^2 f}{\partial y^2}$ tells us about the concavity (or curvature) of f as we move in the y direction.
- For nice functions, we actually have that $\frac{\partial}{\partial y} \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} \frac{\partial f}{\partial y}$. This interpretation is a bit harder to deal with. Let us not worry too much about it at the moment. We call a second partial derivative of this type a ***mixed partial derivative***.

Proposition 10.3.1: Partial Derivatives Commute

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a scalar field with input variables x_1, x_2, \dots, x_n . Then if all second partial derivatives of f are continuous, we have that

$$\frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i},$$

for any i and j .

The interpretation of the proposition is that for sufficiently nice functions (functions whose second partial derivatives are all continuous), the order in which we take the partial derivatives does not matter. So, for example, if we have a 2-dimensional scalar field $f(x, y)$ with continuous second partial derivatives, it must be that

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

Exercise 10.3.4. Given $f(x, y) = x^2 + y^2$, compute

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

Show through explicit computation that the mixed partial derivatives are equal. What can we say about the curvature of f in the two directions? Does this make sense?

Properties of partial derivatives

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

Partial Derivatives:

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

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

Of course, this holds for any partial derivative.

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

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

Again, this holds for any partial derivative.

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

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

This holds for all partial derivatives.

Remark 10.3.2. The chain rule will show up eventually, but not yet. As for the quotient rule, this also holds, but I don't show it here.

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

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

10.3.2 Directional Derivatives

We are not just limited to taking derivatives of scalar fields with respect to the chosen input variables. Much like we can slice up a scalar field in any direction we would like, we can take derivatives of a scalar field in any direction that we would like. Let us take a scalar field $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ with input $\vec{x} = (x, y, z)$. Recall that a unit vector \hat{n} has length $|\hat{n}| = 1$. Every unit vector corresponds to a direction in space, and as such allows us to define the *directional derivative*

$$\frac{\partial f}{\partial \hat{n}} := \lim_{\delta \rightarrow 0} \frac{f(\vec{x} + \delta \hat{n}) - f(\vec{x})}{\delta}.$$

This derivative then tells us the rate of change of the scalar field f in the direction \hat{n} . In particular, if we choose a specific point $\vec{x}_0 = (x_0, y_0, z_0)$, then $\frac{\partial f}{\partial \hat{n}}(\vec{x}_0)$ is the rate of change of f in the direction of \hat{n} at the point \vec{x}_0 . We will return to this notion later to determine a more simplistic manner in computing this derivative.

10.4 Vector Fields

The final object in our new set of functions and fields are the vector fields. Succinctly, a vector field is a function that inputs a vector and outputs a vector. For a vector field in 3-dimensional space $\vec{V}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ we can write

$$\vec{V}(x, y, z) = \begin{pmatrix} V_1(x, y, z) \\ V_2(x, y, z) \\ V_3(x, y, z) \end{pmatrix}.$$

Try to notice the differences and similarities between vector fields, scalar fields, and curves. For one, we can see that a vector field is made up of a multiple scalar fields and in this particular instance, we have that \vec{V} has three scalar field components. Just as before, it will be nice to visualize many vector fields in the xy -plane to avoid drawing in 3-dimensions. Of course, we can use technology to make nice plots in space. Intuitively, one should imagine that a vector field assigns an arrow at each point in space (or at each point on the domain). This gives us a simple (but tedious by hand) method of visualization.

Example 10.4.1: Constant eastward wind

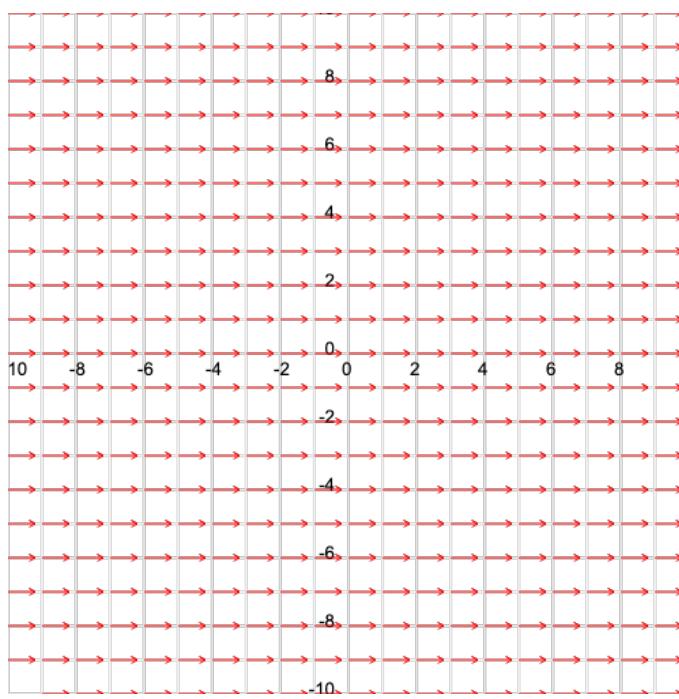
Let us define a planar vector field

$$\vec{V}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

by

$$\vec{V}(x, y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

In this example, the components are $V_1(x, y) = 1$ and $V_2(x, y) = 0$. At each point (x, y) , we are assigning a vector that points a distance 1 in the x -direction. This leads us to the following figure.



To start drawing these fields by hand, the task is as follows:

1. Pick a point (x_0, y_0) and evaluate $\vec{V}(x_0, y_0)$ to get a vector.
2. Take that output vector and attach the tail of that vector at the point (x_0, y_0) in space.
3. Repeat steps 1 and 2 until you are satisfied with the illustration you have.

The figure in this example took the set of points (x_0, y_0) where x_0 and y_0 were all integers from -10 to 10 . Then, the program computed the output vector at each of those points and glued the tails to those points. One could instead pick different points to draw these arrows at and one can also “scale” the lengths of the arrows to look nicer since the qualitative look of the vector field is often the most dominant bit of information.

The vector field in the previous example was constant. As the name suggests, one can imagine this as a constant eastward wind. This way of thinking can help one develop an intuition for why we study vector fields in the first place. Fluid motion (such as motion of the atmosphere) is just one prototypical example that we tend to identify with. Challenge yourself to think of other scenarios you have experienced.

Exercise 10.4.1. Think of the map of the world with a large category five hurricane present in the south Atlantic. Draw a vector field that describes what the wind motion would look like in that case.

We should not limit ourselves solely with macroscopic fields that we can feel blowing against us. The physical world is full of vector fields that create other dynamics that we see. Atoms and molecules predominantly interact due to their construction from charged

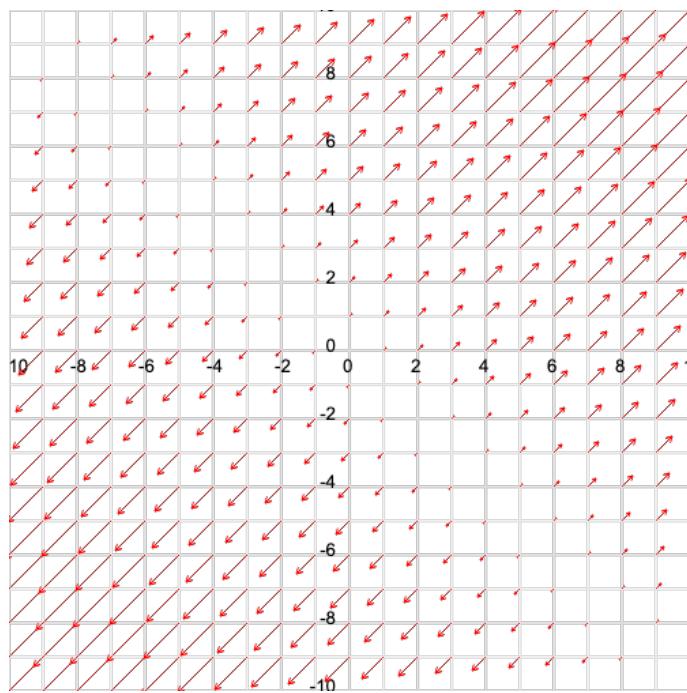
particles. These charged particles move due to the electromagnetic field. We will discuss this field later. For now, let us just see more examples.

Example 10.4.2: Line source

Consider the vector field in the plane given by

$$\vec{V}(x, y) = \begin{pmatrix} x + y \\ x + y \end{pmatrix}.$$

See the following figure.



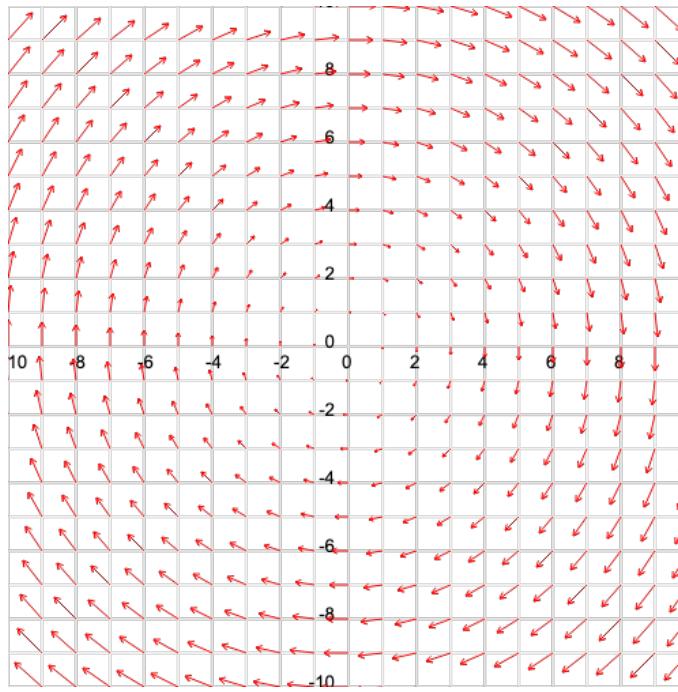
This vector field is zero when $y = -x$. So, along this line, we see no arrows. As we move away from this line, the arrows increase in length and each arrow is pointing in a direction perpendicular to the line $y = -x$.

Example 10.4.3: Vortex

Consider the vector field in the plane given by

$$\vec{V}(x, y) = \begin{pmatrix} y \\ -x \end{pmatrix}.$$

See the following figure.



Perhaps this field starts to make you think of a swirling hurricane. In this case, the field seems to swirl in a clockwise direction and as we move further away from the origin, the lengths of each vector increases.

10.4.1 Vector field algebra

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

$$\vec{V}(x, y, z) + \vec{U}(x, y, z)$$

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

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

then

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

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

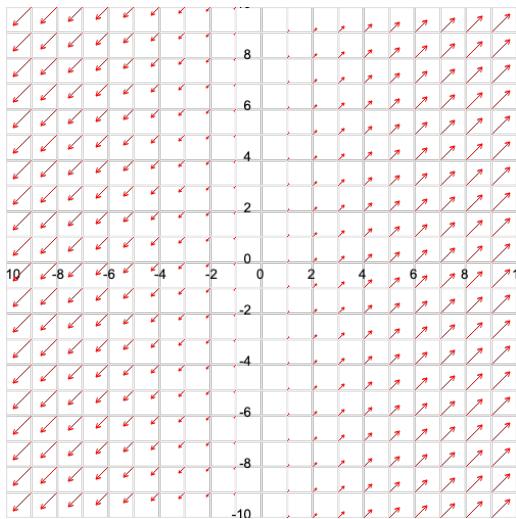
Example 10.4.4: Addition of vector fields

Consider the following vector fields

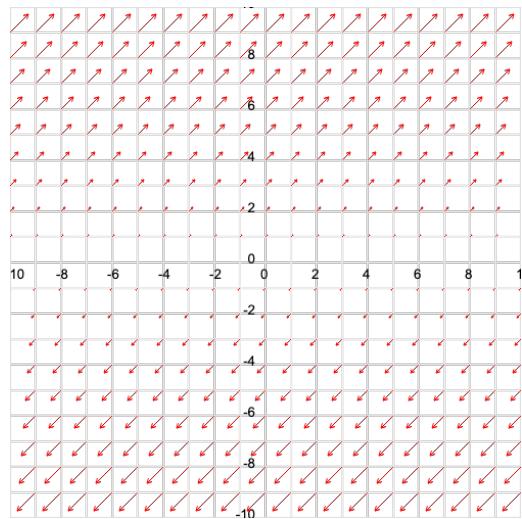
$$\mathbf{v}(x, y) = (x, x)$$

$$\mathbf{u}(x, y) = (y, y).$$

These look like:



(a) Vector field \mathbf{v} .



(b) Vector field \mathbf{u} .

Adding these results in the field

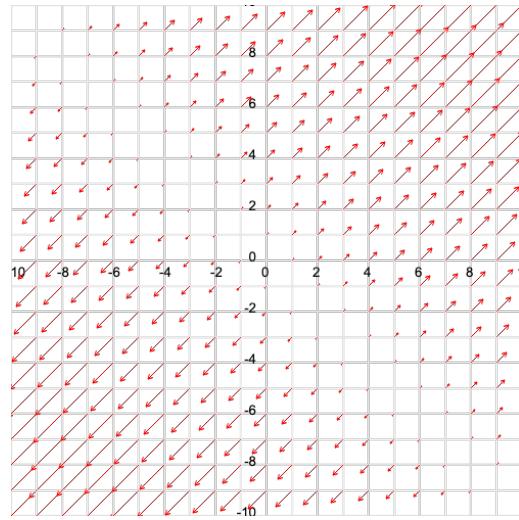


Figure 10.7: The vector field $\vec{U} + \vec{V}$.

Remark 10.4.1. If we do this all for vector fields, we can take curves and scalar fields as special cases.

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

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

Just take the vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} V_1(x, y, z) \\ V_2(x, y, z) \\ V_3(x, y, z) \end{pmatrix}$$

then we can take

$$\lambda \vec{V} = \begin{pmatrix} \lambda V_1 \\ \lambda V_2 \\ \lambda V_3 \end{pmatrix}.$$

In fact, we can scale vector fields by scalar fields! For instance, if we have the scalar field $f(x, y, z)$, then we can create the vector field

$$f(x, y, z) \vec{V}(x, y, z) = \begin{pmatrix} fV_1 \\ fV_2 \\ fV_3 \end{pmatrix}.$$

This is just computing a scalar multiple of a vector where the scalar is allowed to change from point to point. For instance, wind flow may also depend on temperature or pressure at that point.

Remark 10.4.2. There is many reasons why the above is important. It seems our physical world plays nicely with the above concept, for one.

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



Vector calculus

11.1 The gradient of a scalar field

Related to the notion of partial differentiation and the gradient is that of finding the derivative of a function in a given direction. Say that we are given a scalar field $f(x, y, z)$, and we are also given a unit vector $\hat{\mathbf{n}}$. Then, we can compute the *directional derivative of f in the direction $\hat{\mathbf{n}}$* by

$$\frac{\partial f}{\partial \hat{\mathbf{n}}} = \hat{\mathbf{n}} \cdot \vec{\nabla} f.$$

There is another way to define this derivative, but it requires a bit more work to derive.

One can see that we can recover partial derivatives via this notion as well. For example, if we let $\hat{\mathbf{n}} = \hat{\mathbf{x}}$, then we have

$$\frac{\partial f}{\partial \hat{\mathbf{n}}} = \hat{\mathbf{x}} \cdot \vec{\nabla} f = \frac{\partial f}{\partial x}.$$

We can think of directional derivatives as generalizations of partial derivatives where we allow for computing the derivative of our function f in a direction other than the ones given by the chosen basis.

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

Definition 11.1.1: The Gradient

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

$\vec{\nabla} f(x_0, y_0, z_0)$ is given by

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

In fact, we often will consider the gradient as a vector itself. That is, we will put

$$\vec{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}.$$

This will allow us to compute derivatives of vector fields by use of the cross and dot products.

Exercise 11.1.1. Compute the gradient for the function

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

The Gradient:

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

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

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

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

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

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

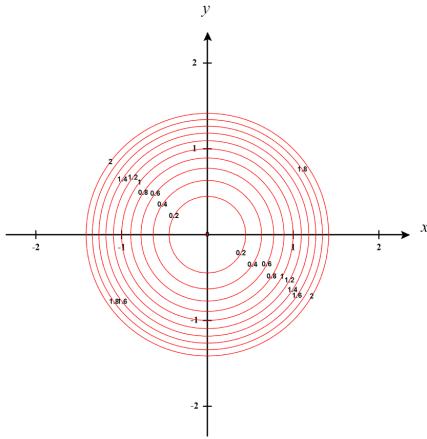
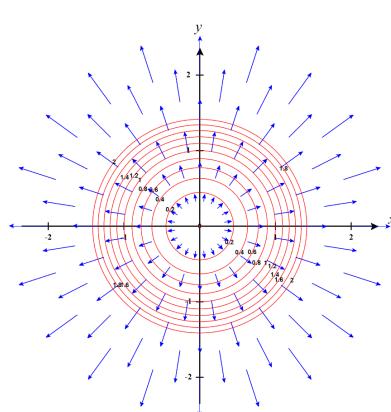
We can put these together into the gradient ∇f and know how f changes in each possible direction. Let's see how the gradient acts then.

Example 11.1.1: Gradients on the Paraboloid

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

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

Let us plot the level curves of this surface.

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

(b) Gradient vectors shown in blue.

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

Proposition 11.1.1: Gradient Points Uphill

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

11.1.1 Optimization

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

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

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

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

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

Definition 11.1.2: Stationary Points

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

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

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

Proposition 11.1.2: Maximizers and Minimizers

A stationary point (x_0, y_0) is a

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

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

Saddle if otherwise.

Exercise 11.1.2. Let

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

- (a) Find all stationary points for f .
- (b) Determine whether these points are minimizers or maximizers.

Exercise 11.1.3. Show that $f(x, y) = xy$ has a saddle point at $(0, 0)$.

Some optimization problems cannot be solved just using this technique. If you are interested, consider reading about *Lagrange multipliers*.

11.2 Differentiation of Vector fields

We have already seen how we can differentiate curves and scalar fields. Combining these two notions should give us what we need as far as building a derivative of a vector field goes. Recall, given a curve $\vec{\gamma}$, that the derivative is the tangent vector

$$\dot{\vec{\gamma}} = \begin{pmatrix} \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \dot{\gamma}_3 \end{pmatrix}.$$

Likewise, if we are given a scalar field f , then the derivative is the gradient vector

$$\vec{\nabla} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}.$$

A vector field is essentially a combination of a curve and scalar field in that it is as if we built a curve from scalar fields. To be precise, a vector field truly defines the tangent vector to many curves that depend on their spatial position. Imagine that a vector field describes the wind that can blow a particle from point to point.

Remark 11.2.1. Though we have not explicitly stated this before, this may be the most important take-away for calculus.

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

Based on this wording, it is intuitive to imagine derivatives as vectors or matrices. The type of matrix that we receive as a derivative depends on the type of function that we wish to differentiate.

11.2.1 The Jacobian of a vector field

Take a 3-dimensional vector field $\vec{V}(x, y, z)$ by

$$\vec{V}(x, y, z) = \begin{pmatrix} V_1(x, y, z) \\ V_2(x, y, z) \\ V_3(x, y, z) \end{pmatrix},$$

where we call each V_1 , V_2 , and V_3 the component functions. Notice that each component function is a scalar field!

Since each component function is a scalar field, we know how to compute the derivative of each by computing the gradient. This gives us a way to then talk about the derivative of the vector field as a whole

Definition 11.2.1: Jacobian

The **Jacobian** of a vector field $\mathbf{v}(x, y, z)$ is a matrix

$$[J](x, y, z) := \begin{pmatrix} \vec{\nabla} V_1^T \\ \vec{\nabla} V_2^T \\ \vec{\nabla} V_3^T \end{pmatrix},$$

where the gradients are transposed (the superscript T) so they are written as row vectors and placed in a matrix. More specifically, we can write this matrix as

$$[J](x, y, z) = \begin{pmatrix} \frac{\partial v_1}{\partial x} & \frac{\partial v_1}{\partial y} & \frac{\partial v_1}{\partial z} \\ \frac{\partial v_2}{\partial x} & \frac{\partial v_2}{\partial y} & \frac{\partial v_2}{\partial z} \\ \frac{\partial v_3}{\partial x} & \frac{\partial v_3}{\partial y} & \frac{\partial v_3}{\partial z} \end{pmatrix}.$$

The Jacobian contains a lot of information. Intuitively, it tells us how each component of the vector field changes in each direction. In fact, the Jacobian essentially contains all the possible 1st order differential information that we could ever use for a vector field. It will also allow us to compute coordinate transformations and volumes!

Example 11.2.1: Computing the Jacobian, 1

Let us consider the vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} x^2 + y^2 \\ z \\ x + y + z \end{pmatrix}.$$

Then we can write

$$V_1(x, y, z) = x^2 + y^2$$

$$V_2(x, y, z) = z$$

$$V_3(x, y, z) = x + y + z.$$

So we compute the gradients of each

$$\vec{\nabla}V_1 = \begin{pmatrix} 2x \\ 2y \\ 0 \end{pmatrix}$$

$$\vec{\nabla}V_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\vec{\nabla}V_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

So then the Jacobian is

$$[J](x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

This also tells us how a generalized product rule should work in this case. Recall that we can take a scalar field $f(x, y, z)$ and multiply it by a vector field $\vec{V}(x, y, z)$ to get $f\vec{V}$. Then, the Jacobian of this field is given by

$$[J](x, y, z) = \begin{pmatrix} \vec{\nabla}(fV_1)^T \\ \vec{\nabla}(fV_2)^T \\ \vec{\nabla}(fV_3)^T \end{pmatrix},$$

where we can then use the product rule for the gradient to compute the above statement.

A related and important quantity comes in the form of the determinant of the Jacobian. We've previously talked about the determinant of a matrix as telling us the (signed) scaling of volume of a linear transformation. That is, for example, if we have

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

then $\det(A) = 8$ and we say that volumes of parallelopipeds are increased by a factor of 8 in this case.

When you are given a vector field, you can think of different regions of space being stretched differently. This is why we have that the Jacobian is a matrix that depends on the position (x, y, z) . In this case, the volumes that are being stretched are very very tiny parallelopipeds. You can think of cubes with side lengths dx , dy , and dz .

Example 11.2.2: Computing the Jacobian, 2

We found that given

$$\vec{V}(x, y, z) = \begin{pmatrix} x^2 + y^2 \\ z \\ x + y + z \end{pmatrix}$$

that

$$[J](x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

As we can see, this matrix depends on position. Let's compute the determinant

$$\det([J](x, y, z)) = 2(y - x).$$

Something weird seems to happen with $y = x$ as the determinant $|[J](x, y, z)|$ will be zero in this case. In that case, it seems as if the volume along the plane given by $y = x$ is contracted to zero.

11.2.2 Divergence and Curl

Often we do not need to use the whole Jacobian. We will find it to be necessary for integration, however. For analysis of vector fields, we often wish to break them into their smaller pieces. Fundamentally, we can break vector fields into two parts:

- Sources and sinks,
- Rotations.

Sources, Sinks, and Divergence Fields

With some vector fields, we can make the analogy that some quantity (think air or water) is being added or removed from the system. We call these *sources* and *sinks* respectively. We want to quantify how much of some quantity is being added. This quantity is called the *divergence*.

Recall, we can write

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).$$

If we are also given a vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} V_1(x, y, z) \\ V_2(x, y, z) \\ V_3(x, y, z) \end{pmatrix},$$

we can compute the *divergence* of \vec{V} by

$$\vec{\nabla} \cdot \vec{V}(x, y, z) = \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial y} + \frac{\partial V_3}{\partial z}.$$

Notice that this quantity is a scalar! This scalar value, the divergence, tells us how much the vector field is diverging at a point (x, y, z) . In other words, it tells us how much of a quantity is being added or removed there. In short, the divergence is the dot product of the gradient operator $\vec{\nabla}$ with the vector field \vec{V} . We can also realize the divergence as the trace of the Jacobian $\text{tr}([J])$.

Example 11.2.3: Source Field

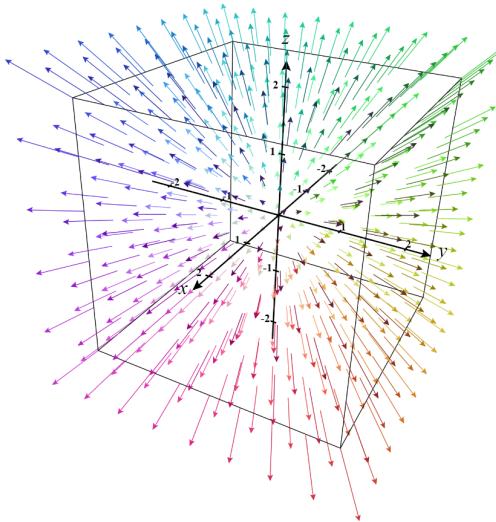
Consider

$$\vec{V}(x, y, z) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Then the divergence is

$$\vec{\nabla} \cdot \vec{V}(x, y, z) = \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial y} + \frac{\partial V_3}{\partial z} = 1 + 1 + 1 = 3.$$

We can think of a source of air being placed at each point that pumps in 3 units of air per second, or specifically being pumped in at the origin. The vector field looks like:



Try to imagine that air is being pumped in from the origin, and the vector field shows the flow of the air out from the origin.

Say that we again take a scalar field $f(x, y, z)$ and multiply it by $\vec{V}(x, y, z)$. How does this affect the divergence? Well, we can compute this using a more generalized notion of the product rule (or by realizing the divergence as the trace of the Jacobian via the earlier example). We will find

$$\vec{\nabla} \cdot (f\vec{V}) = (\vec{\nabla} f) \cdot \vec{V} + f(\vec{\nabla} \cdot \vec{V}).$$

Exercise 11.2.1. Take any vector field and any scalar field and compute the above product rule.

Rotation Fields

The divergence was the quantity that measured the outflow from a point for a vector field. The other quantity we can measure is the **vorticity** of a vector field at a point. That is, how much a vector field looks like a “vortex” around any given point.

We define the **curl** of a vector field $\vec{V}(x, y, z)$ to be

$$\vec{\nabla} \times \vec{V}(x, y, z) = \begin{pmatrix} \frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z} \\ \frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x} \\ \frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \end{pmatrix}.$$

Note that the curl is a vector! The curl is a vector that points in a direction orthogonal to the plane where rotation in a field occurs and has magnitude relative to how quickly the field swirls. The direction of the curl also tells us whether the field rotates clockwise or counterclockwise in the plane of rotation as well.

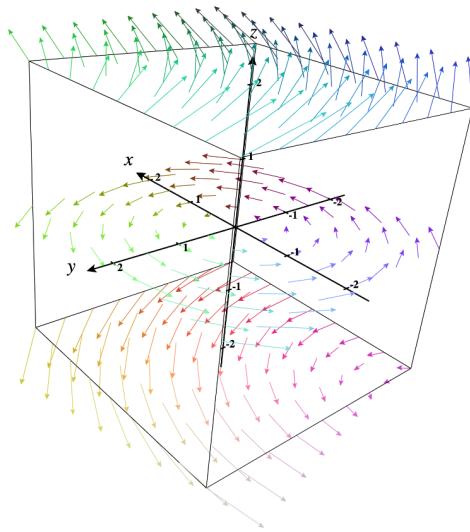
It's a bit involved to go through the work and see exactly why this is the correct quantity for seeing rotation of a vector field. However, you can recall that the cross product was useful in describing rotational motion of rigid bodies (that is, it showed up in angular velocity/momentum).

Example 11.2.4: A Rotation Field

Consider the vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} -y \\ x \\ z \end{pmatrix}$$

which looks like



We let

$$V_1(x, y, z) = -y$$

$$V_2(x, y, z) = x$$

$$V_3(x, y, z) = z.$$

If we look in this figure where $z = 0$, we can clearly see that this field swirls around the origin. If the curl is to measure rotation, we should see it nonzero here. Let us compute the curl of this field. For this, we will all the other partial derivatives not contained in the divergence. That is, we need

$$\begin{aligned}\frac{\partial V_1}{\partial y} &= -1 & \frac{\partial V_1}{\partial z} &= 0 \\ \frac{\partial V_2}{\partial x} &= 1 & \frac{\partial V_2}{\partial z} &= 0 \\ \frac{\partial V_3}{\partial x} &= 0 & \frac{\partial V_3}{\partial y} &= 0.\end{aligned}$$

Then we have

$$\vec{\nabla} \times \mathbf{v}(x, y, z) = \begin{pmatrix} 0 - 0 \\ 0 - 0 \\ 1 - (-1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

We can decipher the meaning here by saying that the swirling occurs in planes parallel to the xy -plane since the direction of the curl is only in the z -direction. That is, curl is pointing perpendicularly to the plane of rotation. How quickly the field swirls is given by the magnitude of the curl which is 2 in this case. Using the right hand rule we discussed previously this tells us the direction of swirling as well. We have swirling counter-clockwise in the planes parallel to the xy -plane, and so we expect the curl to point in the positive z -direction.

Exercise 11.2.2. Plot the above field yourself and do a visual analysis of the vector field.

Example 11.2.5: Divergence of the Rotation Field

One may also consider the divergent nature of the field

$$\vec{V}(x, y, z) = \begin{pmatrix} -y \\ x \\ z \end{pmatrix}$$

from the previous example and find that

$$\vec{\nabla} \cdot \vec{V}(x, y, z) = 1.$$

So, there is in some way divergence as well. This leads us to breaking the vector field into a part that swirls and a part that diverges as follows:

$$\begin{aligned}\vec{V}_{\text{swirl}}(x, y, z) &= \begin{pmatrix} -y \\ x \\ 0 \end{pmatrix} \\ \vec{V}_{\text{div}}(x, y, z) &= \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}.\end{aligned}$$

This type of analysis can be very helpful when considering real world problems. It is especially important in electromagnetism.

Remark 11.2.2. There is a general statement for the decomposition of vector fields given above. It is known as the *Helmholtz decomposition*. This decomposition is very important in the study of electromagnetism as it allows for one to fully decompose the symmetry group of the electromagnetic theory.

Once again, we may wonder how this can be affected if we multiply our vector field by a scalar field. That is, how do we evaluate $\vec{\nabla} \times (f\vec{V})$? Again, this comes from a more general product rule, and we have

$$\vec{\nabla} \times (f\vec{V}) = (\vec{\nabla} f) \times \vec{V} + f(\vec{\nabla} \times \vec{V}).$$

Constant Vector Fields

Most of the understanding of vector fields was just covered by understanding the part that diverges and the part that curls. However, you can always add constants to these vector fields and these constants will not change the divergence or curl. Why? Take the following example.

Example 11.2.6: Constant Fields

Let

$$\mathbf{v}(x, y, z) = (c_1, c_2, c_3)$$

where c_1, c_2 , and c_3 are constants. Then we can compute the Jacobian of \mathbf{v}

$$J(x, y, z) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the Jacobian holds all the partial derivative information, we can know from this that

$$\nabla \cdot \mathbf{v} = 0$$

$$\nabla \times \mathbf{v} = \mathbf{0}.$$

Exercise 11.2.3. Specifically show that the divergence and curl of a constant vector field (as in the previous example) are zero.

All of this is to say that aside from the addition of a constant vector field, we understand the behavior by looking at divergence and curl.

11.3 Laplace Operator

Underlying the dynamics of many differential equations is an operator known as the *Laplacian* or *Laplace operator*. There are many reasons why this operator is so special, but one fundamental reason is it describes the sums of curvatures at each point (at least with scalar functions). In studying second order ODEs we came across the operator

$$-\frac{d^2}{dx^2},$$

which is the Laplace operator in one dimension. This operator described the curvature of a 1-dimensional rod. We will now seek to generalize this operator so that we may describe the curvature of a higher dimensional membrane. We will also define an analogous operator for vector fields.

Fundamentally, in physics, we take the belief that fields tend to be *minimal* in some sense. When it comes to scalar fields, we can imagine their graph, and find scalar fields that have minimal deformation energy given that the function must have certain boundary conditions. This is known as the minimal surface problem. This is where the scalar Laplace operator enters. Likewise, to find minimal vector fields, we should have analogous notion, and we do. If we were to place a number of sinks and sources in space, we could find the vector field that minimizes the energy of this configuration. This amounts to finding the solution to an equation involving the vector Laplacian.

11.3.1 The Laplacian of a Scalar Field

In the study of partial differential equations (PDEs), we are often asked to find a function $u(x, y, z)$ that satisfies the following equation

$$\vec{\nabla} \cdot \vec{\nabla} u(x, y, z) = f(x, y, z)$$

for some given function $f(x, y, z)$. We will revisit this in the next part, but for now we should see exactly what we mean by

$$\vec{\nabla} \cdot \vec{\nabla} u(x, y, z).$$

Exercise 11.3.1. Show that

$$\vec{\nabla} \cdot \vec{\nabla} u(x, y, z) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

Definition 11.3.1: Laplacian

We define the quantity

$$\vec{\nabla} \cdot \vec{\nabla} u(x, y, z)$$

to be the *scalar Laplacian* of $u(x, y, z)$ and we often write

$$\Delta := \vec{\nabla} \cdot \vec{\nabla}.$$

Intuitively, the Laplacian can be summed up in a few ways.

- The Laplacian is the *divergence* of the *gradient* of a scalar function.
- The Laplacian is the sum of “curvatures” in each direction.

Example 11.3.1: Computing the Laplacian

Let us consider the functions

$$f(x, y) = x^2 + y^2$$

and

$$g(x, y) = x^2 - y^2.$$

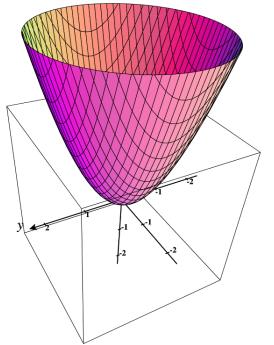
Then we can compute the gradients of each function to get

$$\begin{aligned}\vec{\nabla} f(x, y) &= \begin{bmatrix} 2x \\ 2y \end{bmatrix} \\ \vec{\nabla} g(x, y) &= \begin{bmatrix} 2x \\ -2y \end{bmatrix}.\end{aligned}$$

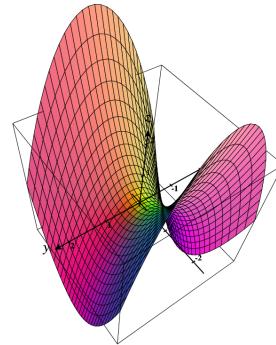
We can then compute the divergence of each of these and find

$$\begin{aligned}\vec{\nabla} \cdot \vec{\nabla} f(x, y) &= 2 + 2 = 4 \\ \vec{\nabla} \cdot \vec{\nabla} g(x, y) &= 2 - 2 = 0.\end{aligned}$$

Let us see what these two functions look like to get a bit of an intuitive feel.



(a) A plot of $f(x, y)$.



(b) A plot of $g(x, y)$.

Fundamentally we can see that these two functions are different. It seems that for $f(x, y)$ we are curving upward in both the x and y direction which is what allows the Laplacian to be positive. However, for $g(x, y)$ one direction curves the opposite direction as the other which cancels out and gives us that the Laplacian is zero. It turns out that the Laplacian describes many phenomenon. Two examples would be soap films and temperature flow.

11.3.2 Vector Laplacian

As stated in the introduction to this section, there exists an analogous operator known as the ***vector Laplacian***. Given a vector field $\vec{V}(x, y, z)$, we define the vector Laplacian by

$$\vec{\Delta} \vec{V} := \vec{\nabla}(\vec{\nabla} \cdot \vec{V}) - \vec{\nabla} \times (\vec{\nabla} \times \vec{V}).$$

In our cartesian coordinate system, we can write this as

$$\vec{\Delta} \vec{V} = \begin{pmatrix} \Delta V_1 \\ \Delta V_2 \\ \Delta V_3 \end{pmatrix}.$$

Let us compute this with an example.

Example 11.3.2: Electric Field

Say we place a point charge of 1 at the origin. The electric field given by this point charge is then

$$\vec{E}(x, y, z) = \begin{pmatrix} \frac{-x}{(x^2+y^2+z^2)^{3/2}} \\ \frac{-y}{(x^2+y^2+z^2)^{3/2}} \\ \frac{-z}{(x^2+y^2+z^2)^{3/2}} \end{pmatrix}.$$

Then, we can compute the vector Laplacian of this field

$$\vec{\Delta}\vec{E} = \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \vec{\nabla} \times (\vec{\nabla} \times \vec{E}).$$

Note that we have $\vec{\nabla} \times \vec{E} = 0$ (which is known as Faraday's law). Hence, we need only compute

$$\vec{\Delta}\vec{E} = \vec{\nabla}(\vec{\nabla} \cdot \vec{E}).$$

We have that

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial y} + \frac{\partial V_3}{\partial z} \\ &= \frac{-2x^2 + y^2 + z^2}{(x^2 + y^2 + z^2)^{5/2}} + \frac{x^2 - 2y^2 + z^2}{(x^2 + y^2 + z^2)^{5/2}} + \frac{x^2 + y^2 - 2z^2}{(x^2 + y^2 + z^2)^{5/2}} \\ &= 0, \text{ everywhere except at } (x, y, z) = 0. \end{aligned}$$

Thus, it must be that

$$\vec{\Delta}\vec{E} = 0, \text{ except for at the origin.}$$

This means that, in some sense, the electric field is the minimal field given this source configuration. We will ignore the origin in this computation, but the solution is related to the delta function. More on this later.

Exercise 11.3.2. Verify that the above computations are correct. Specifically, compute $\vec{\nabla} \times \vec{E}$ and compute the vector Laplacian by taking the scalar Laplacian of each component.

11.4 Composite Functions and Chain Rule

We have introduced curves $\vec{\gamma}$, scalar fields f , and vector fields \vec{V} as the fundamental types of functions in the space \mathbb{R}^3 . Given these types of functions, we can also consider composite functions such as

$$f \circ \vec{\gamma}, \quad \vec{V} \circ \vec{\gamma}, \quad f \circ \vec{V}.$$

We will work through each example, and realize the derivative of the composite functions as well. Hence, we are seeking to generalize the chain-rule from single variable calculus.

Earlier, we briefly remarked that *the derivative of a function is the best linear approximation to that function*. So, let us briefly revisit a few things.

First, if we have a curve $\vec{\gamma}: \mathbb{R} \rightarrow \mathbb{R}^3$, we computed the derivative of this curve as $\dot{\vec{\gamma}}$. Keep in mind, we may restrict the domain of a curve to a subset of \mathbb{R} such as $[a, b]$, but this does not change the argument we are about to make. If we have

$$\vec{\gamma}(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \gamma_3(t) \end{pmatrix},$$

then

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} \dot{\gamma}_1(t) \\ \dot{\gamma}_2(t) \\ \dot{\gamma}_3(t) \end{pmatrix}.$$

But, we can note that $\dot{\vec{\gamma}}: \mathbb{R} \rightarrow \mathbb{R}^3$ as well. At the time t_0 , the derivative $\dot{\vec{\gamma}}(t_0)$ is thus a 3×1 -matrix (a column vector).

Likewise, we can take a scalar field $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ and note that the derivative of f was the gradient $\vec{\nabla}: \mathbb{R}^3 \rightarrow \mathbb{R}$. In other words, we should have that $\vec{\nabla}f$ is a 1×3 -matrix given by

$$\vec{\nabla}f = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \quad \frac{\partial f}{\partial z} \right).$$

However, we wrote the gradient as a column vector. This distinction is not extremely important, but the important part is that at some point (x_0, y_0, z_0) , one can supply a column vector \vec{u} to $\vec{\nabla}f(x_0, y_0, z_0)$, and receive a number. How so? Well,

$$\begin{aligned} \vec{\nabla}f(x_0, y_0, z_0)\vec{u} &= \left(\frac{\partial f}{\partial x}(x_0, y_0, z_0) \quad \frac{\partial f}{\partial y}(x_0, y_0, z_0) \quad \frac{\partial f}{\partial z}(x_0, y_0, z_0) \right) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \\ &= u_1 \frac{\partial f}{\partial x}(x_0, y_0, z_0) + u_2 \frac{\partial f}{\partial y}(x_0, y_0, z_0) + u_3 \frac{\partial f}{\partial z}(x_0, y_0, z_0). \end{aligned}$$

Finally, we took a full derivative of a vector field $\vec{V}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, and noted that this gave us a matrix $[J]_{\vec{V}}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Then, at each point in space (x_0, y_0, z_0) , the matrix $[J]_{\vec{V}}(x_0, y_0, z_0)$ is merely a 3×3 -matrix of numbers. We could of course supply a 3-dimensional vector \vec{u} , and $[J]_{\vec{V}}(x_0, y_0, z_0)\vec{u}$ will be another 3-dimensional vector. We spent quite some time analyzing 3×3 -matrices in the prequel, and it is worth revisiting them again.

11.4.1 Composite Functions

First, let us consider the function

$$f \circ \vec{\gamma}.$$

Notice f will have three inputs, and $\vec{\gamma}$ has one. But, $\vec{\gamma}$ outputs three quantities, and thus this composite function makes sense.

For example, if we have

$$f(x, y, z) = 2x + 3y + 4z \quad \text{and} \quad \vec{\gamma}(t) = \begin{pmatrix} t \\ t^2 \\ t^3 \end{pmatrix},$$

then

$$f \circ \vec{\gamma} = f(\vec{\gamma}(t)) = 2t + 3t^2 + 4t^3.$$

But, how should we compute the derivative of this composite function?

Exercise 11.4.1. Let a vector field $\vec{V}(x, y, z) = \begin{pmatrix} x^2 \\ x - y - z \\ z^2 \end{pmatrix}$ and compute the composite functions

$$\vec{V} \circ \vec{\gamma} \quad \text{and} \quad f \circ \vec{V}.$$

In 1-dimension, if we had the functions $g, h: \mathbb{R} \rightarrow \mathbb{R}$, we had

$$(g \circ h)' = (g(h(x)))' = g'(h(x))h'(x).$$

Our result should mimic this, but our result should also be the best linear function of the correct dimensions. For example, the composite function $f \circ \vec{\gamma}: \mathbb{R} \rightarrow \mathbb{R}$ is a function from 1-dimension to 1-dimension, and so $(f \circ \vec{\gamma})'$ should be a linear function from \mathbb{R} to \mathbb{R} . So, following our nose, the chain rule yields

$$\begin{aligned} (f \circ \vec{\gamma})' &= f'(\vec{\gamma}(t))\dot{\vec{\gamma}}(t) \\ &= \vec{\nabla}f(\vec{\gamma}(t))\dot{\vec{\gamma}}(t) \\ &= \left(\frac{\partial f}{\partial x}(\vec{\gamma}(t)) \quad \frac{\partial f}{\partial y}(\vec{\gamma}(t)) \quad \frac{\partial f}{\partial z}(\vec{\gamma}(t)) \right) \begin{pmatrix} \dot{\gamma}_1(t) \\ \dot{\gamma}_2(t) \\ \dot{\gamma}_3(t) \end{pmatrix} \\ &= \frac{\partial f}{\partial x}(\vec{\gamma}(t))\dot{\gamma}_1(t) + \frac{\partial f}{\partial y}(\vec{\gamma}(t))\dot{\gamma}_2(t) + \frac{\partial f}{\partial z}(\vec{\gamma}(t))\dot{\gamma}_3(t). \end{aligned}$$

Repeating this, we can take

$$(\vec{V} \circ \vec{\gamma})' = [J]_{\vec{V}}(\vec{\gamma}(t))\dot{\vec{\gamma}}(t).$$

And lastly, we have

$$(f \circ \vec{V})' = \vec{\nabla}f(\vec{V}(x, y, z))[J]_{\vec{V}}(x, y, z).$$

For this, one should think of $\vec{\nabla}f$ as a 1×3 -matrix, and perform the matrix product above.

Exercise 11.4.2. Compute the derivatives of the composite functions using the f , $\vec{\gamma}$, and \vec{V} given in this section.

11.5 Integration of Vector Fields

Adding up fields over curves, surfaces, or volumes is commonplace. We have already done this for scalar fields, but we would also like to do this for vector fields. The problem is, we don't have the ability to integrate vectors. We only know how to integrate quantities such

$$f(x, y, z)dx \quad f(x, y, z)dxdy \quad f(x, y, z)dxdydz.$$

Of course, we could also allow for different permutations like $f(x, y, z)dydz$ or any addition of the above.

What we must do is get scalar quantities from vector fields. The primary way of doing this is to project the vector field using the dot product. This will allow us to build three notions of integrating a vector field. First, we can integrate the amount of a vector field that flows along a curve. Second, we can integrate the amount of a vector field flowing through a surface. Finally, we can integrate the divergence of a vector field over a volume.

11.5.1 Line integrals of vector fields

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

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

$$\int_{\vec{\gamma}} \vec{F} \cdot d\vec{\gamma} = \int_a^b \vec{F}(\vec{\gamma}(t)) \cdot \dot{\vec{\gamma}}(t) dt.$$

An intuitive notion of this integral is that we are computing how much our vector field is aligning with the tangent vector of the curve at each point. When the fields are in great alignment, the integral will be large. If the vector field is perpendicular to the tangent vector to the curve at each point, then the integral will be zero. If the fields are anti-aligned (e.g., $\vec{F} = -\dot{\vec{\gamma}}$), then the integral will be large in magnitude but negative.

Example 11.5.1: Work Done on a Particle

The work done on a particle (or change in energy) is written as a line integral of this form.

Take for example, $\vec{F}(x, y) = \begin{pmatrix} 2x \\ 3y \end{pmatrix}$ and $\vec{\gamma}(t) = \begin{pmatrix} t \\ t^2 \end{pmatrix}$ over the time $t = 0$ to $t = 1$.

Then we can note that

$$\dot{\vec{\gamma}}(t) = \begin{pmatrix} 1 \\ 2t \end{pmatrix}.$$

This yields,

$$\begin{aligned} \int_{\vec{\gamma}} \vec{F} \cdot d\vec{\gamma} &= \int_0^1 \vec{F}(t, t^2) \cdot \begin{pmatrix} 1 \\ 2t \end{pmatrix} dt \\ &= \int_0^1 \begin{pmatrix} 2t \\ 3t^2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2t \end{pmatrix} dt \\ &= \int_0^1 2t + 6t^3 dt. \end{aligned}$$

Exercise 11.5.1. Redo the above work and compute the integral.

Remark 11.5.1. This notion is extremely important in defining something called the *potential* of a vector field. This will show up in electrodynamics. If the force in the example above is *conservative*, we will have a potential. This will correspond nicely to the vector field having no *curl*.

11.6 Potential Functions

In one variable calculus, we found that there is a relationship between the derivative and the indefinite integral. In fact, this led us to call the indefinite integral the antiderivative. This relationship was that

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

and

$$\int \frac{df}{dx} = f(x) + C.$$

From this, we realized that the indefinite integral is almost an inverse operation of the derivative. It's just that in the case where we integrate a derivative, we only determine the function up to an additive constant.

The higher dimensional analog happens to be a bit more nuanced but the idea remains the same. Let's say we are given a function $f(x, y, z)$ and we compute, for example,

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

The issue now becomes this. Let's say that we let

$$f(x, y, z) = x + yz.$$

Then we have that

$$\frac{\partial f}{\partial x} = 1.$$

The terms with just a y, z dependence disappear. So if we were to try to undo this with an integral, we find that

$$\int \frac{\partial f}{\partial x} dx = \int 1 dx = x + g(y, z).$$

That is to say, when we take an indefinite integral a multivariate function with respect to one variable, there could be a function of the residual variables that we cannot determine!

11.6.1 Integrating the Gradient

Let's say that we are given $\vec{V}(x, y, z) = \vec{\nabla} f(x, y, z)$ and are asked to find the original function $f(x, y, z)$. This problem is called finding the ***potential function*** or just ***potential*** for \vec{V} . Remember that

$$\vec{\nabla} f(x, y, z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}.$$

What we do is the following.

1. We integrate $\frac{\partial f}{\partial x}$ with respect to x and determine $f(x, y, z)$ up to adding a function of only y and z . That is we are able to recover what is essentially a third of the potential function $f(x, y, z)$.
2. We integrate $\frac{\partial f}{\partial y}$ with respect to y and determine $f(x, y, z)$ up to adding a function of only x and z .

3. We integrate $\frac{\partial f}{\partial z}$ with respect to z and determine $f(x, y, z)$ up to adding a function of only x and z .
4. Combine our knowledge from those three integrals we have determined $f(x, y, z)$ up to some additive constant!

Finding a potential is very helpful as there are some nice properties of these types of functions.

Example 11.6.1: Finding a Potential Function

Let's say that we are given the gradient of some function

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

Then we follow the steps above

1. We integrate $\frac{\partial f}{\partial x}$ with respect to x . So we have

$$\int y + z dx = xy + xz + g(y, z).$$

Here, $g(y, z)$ is a function of just y and z that we cannot determine yet.

2. We integrate $\frac{\partial f}{\partial y}$ with respect to y . So we have

$$\int x + z dy = xy + yz + h(x, z).$$

3. We integrate $\frac{\partial f}{\partial z}$ with respect to z . So we have

$$\int x + y dz = xz + yz + r(x, y).$$

4. Now we know that all of these functions should be equal (up to a constant). That is

$$xy + xz + g(y, z) = xy + yz + h(x, z) = xz + yz + r(x, y).$$

Here, we can see that $g(y, z) = yz$, $h(x, z) = xz$, and $r(x, y) = xy$. So we have found that

$$f(x, y, z) = xy + xz + yz + C$$

where the additive constant is there and is not something we can determine without a bit more information.

11.6.2 Requirements for Potentials

The main result here is the following.

Proposition 11.6.1: Curl of Gradient is Zero

We have that

$$\vec{\nabla} \times \vec{\nabla} f(x, y, z) = \vec{0}$$

for all $f(x, y, z)$.

Exercise 11.6.1. Prove that the above statement is true.

Then with a bit more work, one can show this follows.

Theorem 11.6.1: Potential \iff Curl Free

Let $\vec{V}(x, y, z)$ be a vector field. Then if

$$\vec{\nabla} \times \vec{V} = \vec{0},$$

then $\vec{V}(x, y, z) = \vec{\nabla} f(x, y, z)$. That is, a curl-free vector field \vec{V} can be written as the gradient of some scalar function $f(x, y, z)$.

This is what allows one to define the voltage $\phi(x, y, z)$ in electrostatics. When charges are not moving, we have that the electric field \vec{E} satisfies

$$\vec{\nabla} \times \vec{E} = \vec{0}$$

and so it follows that

$$\vec{\nabla} \phi = \vec{E}.$$

Hence, the voltage ϕ is the potential function for \vec{E} . Thus the name of the electrostatic potential.

11.6.3 Properties of Fields with Potentials

Let \vec{V} be a vector field that has a potential ϕ . Then, we refer to the vector field \vec{V} as **conservative**. The reason why we give the vector field this name is due to the fact that in physics, these types of vector fields tend to (in some way) conserve energy. Mathematically, we can state this result as the following theorem.

Theorem 11.6.2: Curve Integration of Conservative Vector Fields

Let $\vec{\gamma}: [a, b] \rightarrow \mathbb{R}^3$ be a curve and let \vec{V} be a conservative vector field. Then, if $\tilde{\gamma}$ is another curve satisfying $\vec{\gamma}(a) = \tilde{\gamma}(a)$ and $\vec{\gamma}(b) = \tilde{\gamma}(b)$ we have that

$$\int_{\vec{\gamma}} \vec{V} \cdot d\vec{\gamma} = \int_{\tilde{\gamma}} \vec{V} \cdot d\tilde{\gamma}$$

In other words, the integral of a conservative vector field over a curve only depends on the endpoints of the curve.

We have seen the name potential arise for us before. Namely, in the Schrödinger equation, we saw that the potential was a term in the Hamiltonian. Namely, the Hamiltonian was

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

and $V(x)$ was the potential. In this case, $V(x)$ is exactly the potential for some vector field, and in the realm of quantum mechanics, this vector field is often coming from one caused by the electromagnetic field. We have only viewed quantum mechanics in one dimension, but later on we will revisit Schödinger's equation in higher dimensions.

11.7 Integrals of Scalar Fields

In one dimension, we integrated functions in order to sum up values over a given interval. Geometrically, this gave us the net area under a curve. In this case, we took a function $f(x)$ and an interval $[a, b]$ and we wrote

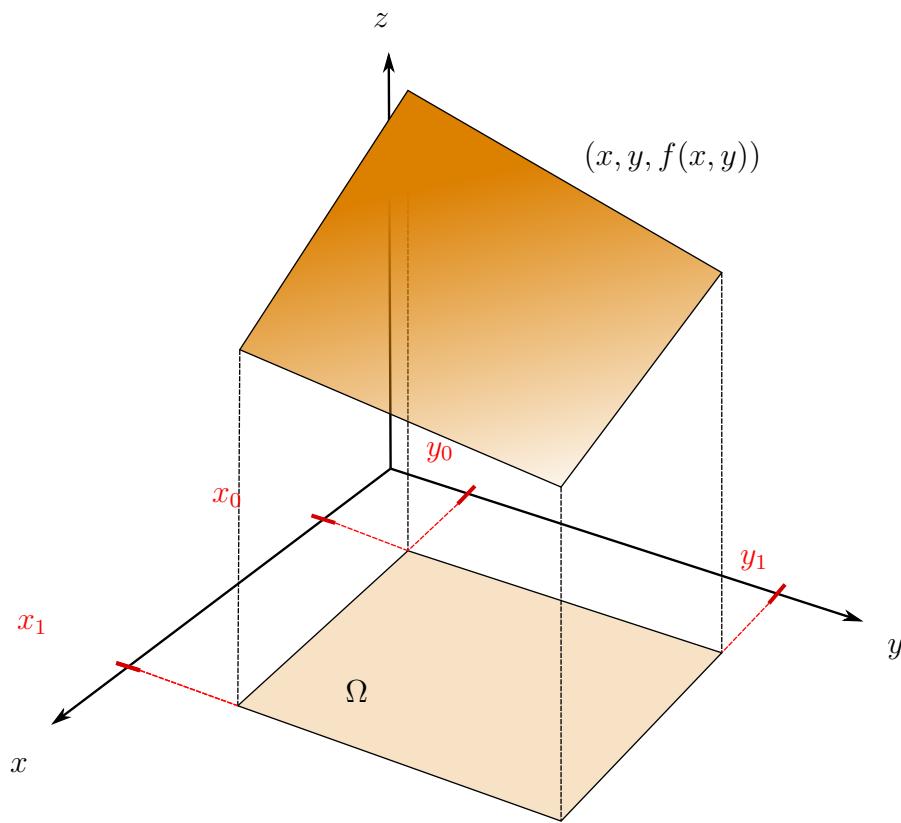
$$\int_a^b f(x) dx.$$

When we work in higher dimensions, we have to be a bit more careful. Let us instead rephrase this problem by instead specifying a region $\Omega = [a, b]$ and a function f , then we would put

$$\int_{\Omega} f d\Omega = \int_a^b f(x) dx.$$

The difference we wish to establish is that the integral we are performing depends on the coordinates in which we choose. Later on, this will become very apparent when we consider new forms of coordinate systems (e.g., polar coordinates).

What we put above is simply a change of notation that allows for greater versatility. Imagine we have a region in the plane Ω and a scalar field $f: \Omega \rightarrow \mathbb{R}$, then we can plot the graph of that function in Cartesian coordinates by thinking of the function f as taking in an (x, y) pair, and placing $f(x, y)$ as the height above the xy -plane.



A region Ω is given in the plane with respect to some coordinates. For example, we may specify the region in the plane by $x_0 \leq x \leq x_1$ and $y_0 \leq y \leq y_1$, and so we write our function with respect to these choices. Then, say we want to compute the volume under the surface given by the graph $(x, y, f(x, y))$, we would then compute

$$\int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y) dx dy.$$

This is known as a **double integral**. But, we can phrase this double integral without reference to any coordinate choice by taking

$$\iint_{\Omega} f d\Omega.$$

We will stick with this notation as we continue onward.

Nothing changes as we increase the dimension; just the number of integrals we must compute will increase. Below, we will compute examples in varying dimensions.

One Dimensional Case

Briefly, let us compute an example in one dimension so that we can see how this analogy generalizes.

Example 11.7.1: One-Dimensional Integral

Let $f(x) = x^2 + 2$, $a = 1$, and $b = 2$. Then we want to find

$$\int_a^b f(x)dx = \int_1^2 x^2 + 2dx.$$

Then we use the *Fundamental Theorem of Calculus*. So, we find the antiderivative of the integrand and evaluate at the endpoints as follows

$$\begin{aligned} \int_1^2 x^2 + 2dx &= \left[\frac{x^3}{3} + 2x \right]_1^2 \\ &= \left(\frac{2^3}{3} + 2(2) \right) - \left(\frac{1^3}{3} + 2(1) \right) \\ &= \frac{13}{3}. \end{aligned}$$

Two Dimensional Case

Say we are now given a function $f(x, y)$ and bounds on both the x and y by $x_0 \leq x \leq x_1$ and $y_0 \leq y \leq y_1$. We then wish to evaluate

$$\int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y)dxdy.$$

You can think of this integral as being the *net* volume under the surface given by $f(x, y)$.

How do we compute such an integral? The answer is iteratively. Let's see how we do this with a concrete example.

Example 11.7.2: Two-Dimensional Integral

Let $f(x, y) = xy$, $x_0 = 1$, $x_1 = 2$, $y_0 = 3$ and $y_1 = 4$. So, we want to evaluate

$$\int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y)dxdy = \int_3^4 \int_1^2 xydxdy.$$

The way we do this is by first evaluating the integral with respect to x (holding y constant) and then integrate with respect to y (x will not appear here). So, we integrate from the inside out.

Let's start by integrating with respect to x . We take

$$\begin{aligned} \int_1^2 xydx &= \left[\frac{x^2y}{2} \right]_1^2 \\ &= \left(y \frac{2^2}{2} \right) - \left(y \frac{1^2}{2} \right) \\ &= \frac{3}{2}y. \end{aligned}$$

Now we take this function of y , and we integrate this with the bounds we are given.

$$\begin{aligned}\int_3^4 \frac{3}{2}y dy &= \left[\frac{3y^2}{4} \right]_3^4 \\ &= \frac{21}{4}.\end{aligned}$$

So we say that

$$\int_3^4 \int_1^2 xy dx dy = \frac{21}{4}.$$

Let's walk through the steps again. We did

$$\begin{aligned}\int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y) dx dy &= \int_3^4 \int_1^2 xy dx dy \\ &= \int_3^4 \frac{3}{2}y dy \\ &= \frac{21}{4}.\end{aligned}$$

Three Dimensional Case

Integration here is performed in the same way. We are given a function $f(x, y, z)$ and bounds on x , y , and z such as $x_0 \leq x \leq x_1$, $y_0 \leq y \leq y_1$, and $z_0 \leq z \leq z_1$. Then we evaluate

$$\int_{z_0}^{z_1} \int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y, z) dx dy dz.$$

Let's work through an example.

Example 11.7.3: Three-Dimensional Integral

Let

$$f(x, y, z) = 2x + 8xyz + 3,$$

and say we want to integrate over the rectangular prism given by $x_0 = 0$, $x_1 = 1$, $y_0 = 2$, $y_1 = 3$, $z_0 = 4$, $z_1 = 5$. Then we want to find

$$\int_{z_0}^{z_1} \int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y, z) dx dy dz = \int_4^5 \int_2^3 \int_0^1 2x + 8xyz + 3 dx dy dz.$$

We do this iteratively. So we first evaluate the x integral holding the other variables constant for now.

$$\begin{aligned}\int_0^1 2x + 8xyz + 3 dx &= \left[x^2 + 4x^2yz + 3x \right]_0^1 \\ &= (1^2 + 4(1)^2yz + 3(1)) - (0^2 + 4(0)^2yz + 3(0)) \\ &= 4yz + 4.\end{aligned}$$

We then take this, and integrate with respect to y .

$$\begin{aligned}\int_2^3 4yz + 4dy &= \left[2y^2 z + 4y \right]_2^3 \\ &= (4(3)^2 z + 4(3)) - (4(2)^2 z + 4(2)) \\ &= 10z + 4.\end{aligned}$$

Lastly, we integrate with respect to z

$$\begin{aligned}\int_4^5 10z + 4dz &= \left[5z^2 + 4z \right]_4^5 \\ &= (5(5)^2 + 4(5)) - (5(4)^2 + 4(4)) \\ &= 49.\end{aligned}$$

So we say that

$$\int_4^5 \int_2^3 \int_0^1 2x + 8xyz + 3 dx dy dz = 49.$$

Again, let's walk through the steps a bit

$$\begin{aligned}\int_{z_0}^{z_1} \int_{y_0}^{y_1} \int_{x_0}^{x_1} f(x, y, z) dx dy dz &= \int_4^5 \int_2^3 \int_0^1 2x + 8xyz + 3 dx dy dz \\ &= \int_4^5 \int_3^4 4yz + 4dy dz \\ &= \int_4^5 10z + 4dz \\ &= 49.\end{aligned}$$

Remark 11.7.1. There is nothing stopping us from integrating, say, a 3-dimensional scalar field over a 2-dimensional domain Ω . The definitions above work in this situation.

Example 11.7.4: Integrating a Scalar Fields on Subsurfaces

Consider the same scalar field $f(x, y, z) = 2x + 8xyz + 3$, but instead integrated over the rectangle Ω defined by $0 \leq x, y \leq 1$ and $z = 1$. Then, we have

$$\begin{aligned}\iint_{\Omega} f(x, y, z) d\Omega &= \int_0^1 \int_0^1 f(x, y, 1) dx dy \\ &= \int_0^1 (x^2 + 4x^2 y + 3x) |_0^1 dy \\ &= \int_0^1 1 + 4y + 3dy \\ &= (2y^2 + 4y) |_0^1 \\ &= 6.\end{aligned}$$

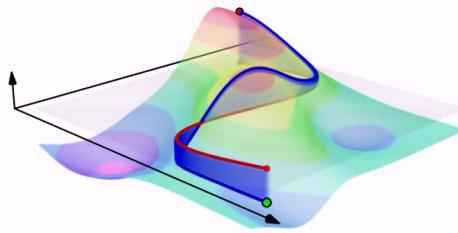
11.7.1 Line Integrals

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

We want to understand the following:

$$\int_{\vec{\gamma}} f d\vec{\gamma} := \int_a^b f(\vec{\gamma}(t)) |\dot{\vec{\gamma}}(t)| dt.$$

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



Example 11.7.5: Length of a curve

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

$$\int_{\vec{\gamma}} f d\vec{\gamma} = \int_a^b |\dot{\vec{\gamma}}(t)| dt$$

which is the length of the curve $\vec{\gamma}$.

Example 11.7.6: Line Integral on a Paraboloid

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

$$\int_{\vec{\gamma}} f d\vec{\gamma} = \int_0^1 f(\vec{\gamma}(t)) |\dot{\vec{\gamma}}(t)| dt.$$

We have

- $f(\vec{\gamma}(t)) = t^2 + t^2 = 2t^2$.
- $|\dot{\vec{\gamma}}(t)| = \left| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right| = \sqrt{2}$.

So we have

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

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

Exercise 11.7.1. Integrate $f(x, y) = x + y$ along the curve $\vec{\gamma}(t) = \begin{pmatrix} t \\ 0 \end{pmatrix}$ from $t = 0$ to $t = 1$. What do you notice about this? Can we tie this to one-dimensional integration?

11.7.2 Volume Integrals

Our final installment of vector integration is that over a volume in space. The geometric interpretation here is that we are integrating over the sources and sinks inside a volume and adding up their contributions. In other words, we can take a vector field $\vec{V}(x, y, z)$ and note that the divergence of this field $\vec{\nabla} \cdot \vec{V}$ describes the source/sink behavior of \vec{V} . Note as well that this is a scalar quantity we can integrate. So, given a volume Ω in space, we can take

$$\iiint_{\Omega} \vec{\nabla} \cdot \vec{V} d\Omega.$$

Once again, the way in which we integrate this depends on how we describe Ω .

Example 11.7.7: Source in a Volume

Take again the field $\vec{V}(x, y, z) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and let Ω be the region given by $0 \leq x \leq 1$, $2 \leq y \leq 3$, and $4 \leq z \leq 5$. Then we can compute

$$\begin{aligned} \iiint_{\Omega} \vec{\nabla} \cdot \vec{V} d\Omega &= \int_4^5 \int_2^3 \int_0^1 \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial y} + \frac{\partial V_3}{\partial z} dx dy dz \\ &= \int_4^5 \int_2^3 \int_0^1 3 dx dy dz. \end{aligned}$$

Exercise 11.7.2. Finish evaluating the above integral.

Remark 11.7.2. There is a result known as the *divergence theorem* which states that given a volume Ω with the boundary surface Σ that

$$\iiint_{\Omega} \vec{\nabla} \cdot \vec{V} d\Omega = \iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma.$$

Physically, this amounts to one being able to measure how much of a substance is being pumped in or out in a region by measuring the in/outflow along the boundary.

This is not a core result for this class, but it is a rather beautiful one.



Surfaces and Flux

The real world is composed of geometrical objects and our models for physics, chemistry, and biology rely heavily on utilizing them. But, how can one describe certain objects such as a sphere, a donut, or a helical curve? We have seen these types of surfaces arise in a few ways, but we need something that provides us a more pragmatic approach to constructing such objects. The two means we will look at in this chapter are that of *implicit* and *explicit parameterizations*. First and foremost, a *parameterization* is a means in which one can describe a geometrical object via variables. For example, one could describe the unit sphere as the level set to $x^2 + y^2 + z^2 = 1$ or one could describe a paraboloid via the graph of the function $z = x^2 + y^2$. The former is an example of an implicit parameterization while the latter is explicit. Let's briefly discuss the differences between these approaches.

Let us now consider surfaces in \mathbb{R}^3 as the prototypical example for a need for providing a parameterization. To be a bit more precise, a surface is a 2-dimensional membrane that lives inside of space. It is not a bad idea to picture a sphere, a donut, or a wiggly sheet of rubber. In general, an implicit parameterization of a surface is given by a list of equations of functions of more than two variables. For example, if we consider

$$f(x, y, z) = x^2 + y^2 + z^2,$$

then we take

$$f(x, y, z) = 1,$$

this gives us a spherical surface. Notice how $f(x, y, z)$ depends on three variables, and setting this equation equal to a constant gives us a surface. There is a far more general approach where we may consider a list of functions $f_1(x_1, x_2, \dots, x_n)$, $f_2(x_1, x_2, \dots, x_n)$, \dots , $f_m(x_1, x_2, \dots, x_n)$ where we set each of these functions equal to a constant. For

example,

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= c_1 \\ f_2(x_1, x_2, \dots, x_n) &= c_2 \\ &\vdots \\ f_m(x_1, x_2, \dots, x_n) &= c_m. \end{aligned}$$

However, we will not need such abstraction.

In the explicit case, we provide a different perspective. Namely, we seek to describe the two dimensional membrane using two variables u and v , and we take the functions $x(u, v)$, $y(u, v)$, and $z(u, v)$ to describe the points in space that live on this surface. That is, if $u \in [u_0, u_1]$ and $v \in [v_0, v_1]$, then a surface can be described by the set of points

$$(x(u, v), y(u, v), z(u, v)).$$

This description is a bit more abstract than we wish. For us, we will take surfaces to be broken up into graphs. For example, we could take the surface given by the points

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

where $f(x, y) = x^2 + y^2$ to receive the paraboloid surface. Notice that here we have swapped u and v for x and y since we are able to construct this as a graph. Graphs are simply the nicest examples of explicit surfaces.

12.1 Implicit surfaces

In the planar case, we would often have to specify a set of points by providing an equation

$$x^2 + y^2 = 1.$$

The tuples of points (x, y) which satisfy the above equations define a nice set of points in the plane that we know as the unit circle.

Exercise 12.1.1. Verify that the set of points (x, y) is indeed the unit circle. *Hint: you can take the points to be $(\cos(\theta), \sin(\theta))$ for $\theta \in [0, 2\pi)$ and notice that these points all live on the circle. Is it possible to find other points?*

This concept should not be totally foreign to us as we previously worked with level sets of scalar fields. The above example is nothing but the 1 level set to the function $f(x, y) = x^2 + y^2$. In higher dimensions, we can do the same in order to visualize functions of the form $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ written as $f(x, y, z)$. If we pick some constant c and write

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

then we will get level surfaces for the function $f(x, y, z)$. We should also remark that surfaces are common objects in space, and we would like a means of describing them anyways. We refer to this description of level surfaces as **implicit**. Some even refer to these surfaces as **implicit surfaces**. Later, we will also consider explicit parameterizations.

Example 12.1.1: A Sphere as a Level Surface

We can consider the following function

$$f(x, y, z) = x^2 + y^2 + z^2.$$

If we take the one level set, that is the points (x, y, z) that satisfy

$$x^2 + y^2 + z^2 = 1$$

then we get the unit *sphere*. These are the set of points that are all a distance one from the origin.

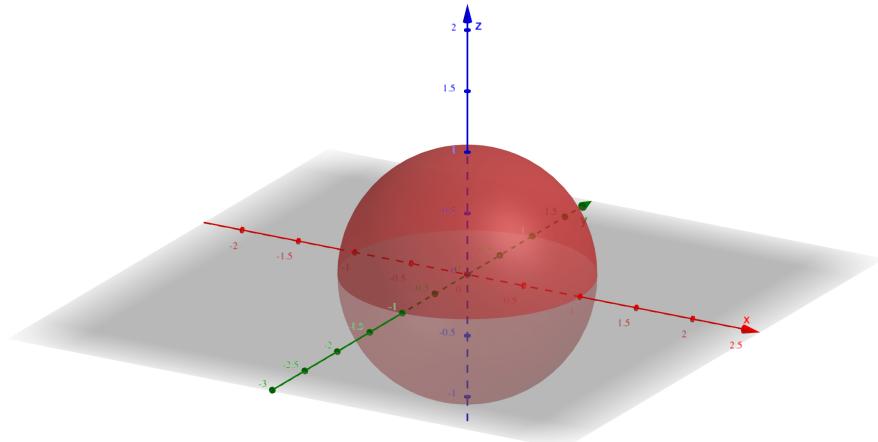


Figure 12.1: The unit sphere in \mathbb{R}^3 as the solution to $x^2 + y^2 + z^2 = 1$.

If we change the constant from $c = 1$, the surface will also change. That is, if we consider the equation

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

then each c describes a different set of points. Let's investigate and see what our options are.

It's quite clear that $c \geq 0$ since a c value less than zero will have no real solutions. One should make sure this is clear. For the value $c = 0$, we have

$$x^2 + y^2 + z^2 = 0,$$

which has a single solution of $(x, y, z) = (0, 0, 0)$. Thus, the level set for $c = 0$ is not a surface but a single point. For values $c > 0$, we will define a sphere with radius \sqrt{c} . Note that the distance from the origin to any point in the plane (x, y, z) is defined by

$$\sqrt{x^2 + y^2 + z^2},$$

and so this description of the sphere is capturing the distance from the origin to a point on the sphere. Calling this distance

$$r = \sqrt{x^2 + y^2 + z^2},$$

we have that

$$r^2 = c \implies r = \sqrt{c}.$$

The sphere will be a go-to example of an implicit surface. The function that describes the sphere behaves rather nicely and the sphere also provides a great visualization. Later on, we will cover a new system of coordinates in \mathbb{R}^3 that is based upon the sphere as well.

One may wonder what happens at the origin in the previous example. To see the issue, we can consider the gradient of the scalar field that defined the surface, $f(x, y, z) = x^2 + y^2 + z^2$. Taking the gradient yields,

$$\vec{\nabla} f = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}.$$

If we evaluate the gradient at the origin, we can note

$$\vec{\nabla} f(0, 0, 0) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is the zero vector! This leads us to the following.

Theorem 12.1.1: Implicit Surfaces iff $\vec{\nabla} f \neq 0$

Given a scalar field $f(x, y, z)$, the equation $f(x, y, z) = c$ defines an implicit surface if the gradient $\vec{\nabla} f$ on the level set of points ($\{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = c\}$) does not vanish. Thus, if we take a level set and consider the points where the gradient is nonzero, we have a level surface.

Example 12.1.2: The Hyperboloids

If we change the function above just slightly, we will receive a whole family of surfaces given by the level sets of

$$f(x, y, z) = x^2 + y^2 - z^2.$$

Again, we will have to consider what values for c define level surfaces.

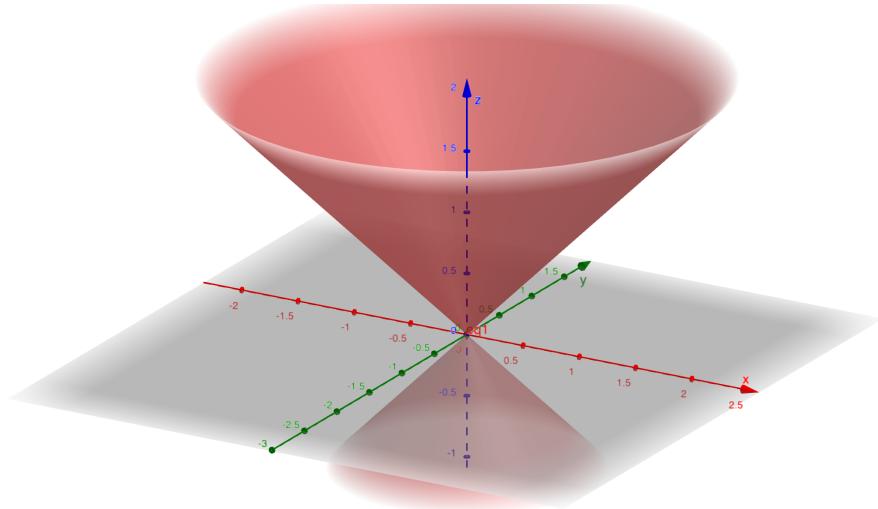
- If we take $c = 0$ and set

$$x^2 + y^2 - z^2 = 0.$$

then we find the 0 level surface. In this case, we can do a bit of work to find

$$z = \pm \sqrt{x^2 + y^2}.$$

Notice, if we pick any value for z , that we get a circle at that level! When $z = 0$, we get a single point. It turns out that we get the **(double) cone** surface which looks like



hy

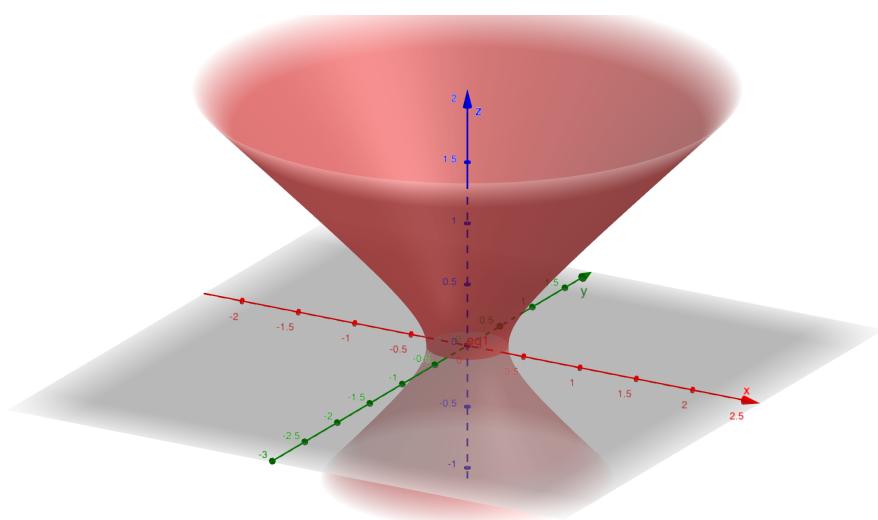
Figure 12.2: (Double) cone surface given by the level set $c = 0$.

At the origin, $(x, y, z) = (0, 0, 0)$, we can see that our surface degenerates to a single point. Again, one can see that the gradient at that point $\vec{\nabla}f(0, 0, 0) = \vec{0}$. One should take some time to work out visualizing this both with software like GeoGebra as well as by hand.

- If we take $c = 1$ and set

$$x^2 + y^2 - z^2 = 1$$

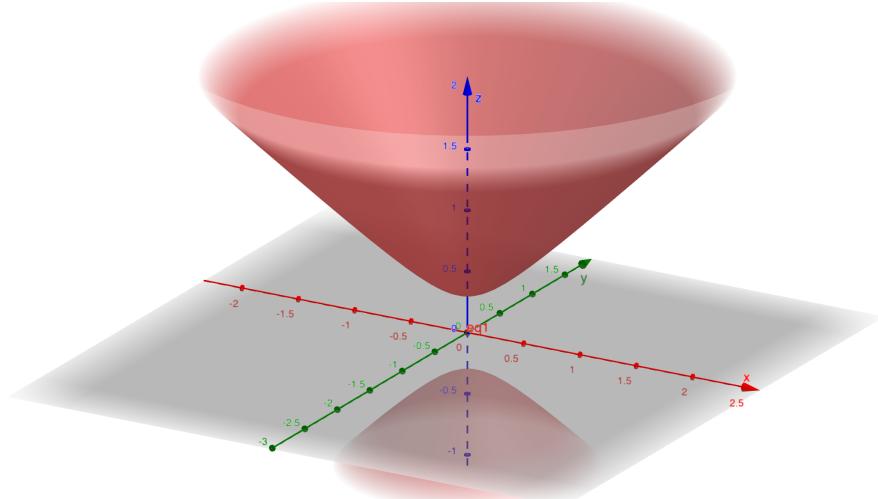
we get the *hyperboloid of one sheet*. This looks like

Figure 12.3: Hyperboloid of 1-sheet defined by $c = 0.1$.

- If we take $c < 0$ and set

$$x^2 + y^2 - z^2 = c$$

we get a *hyperboloid of two sheets*. This looks like

Figure 12.4: Hyperboloid of 2-sheets defined by $c = -0.1$.

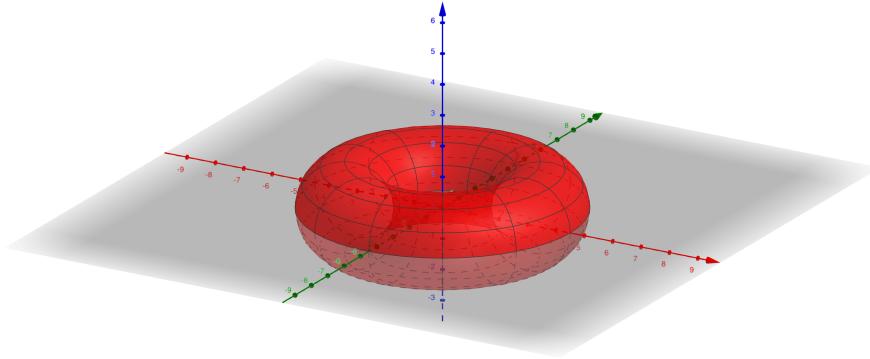
Example 12.1.3: The Torus

For this example, I will choose specific nice numbers, but this is a yet another case

of a level surface. Take

$$\left(3 - \sqrt{x^2 + y^2}\right)^2 + z^2 = \frac{3}{2}.$$

This gives us the *torus* with inner radius (the radius from the center of the donut hole to the center of the tube) $3 = R$ and tube radius $\frac{3}{2} = r$. This looks like



12.1.1 Normal Vectors to Implicit Surfaces

Let $f(x, y, z)$ be a scalar field. Then, recall that the gradient of this scalar function $\vec{\nabla} f$ points perpendicularly to the level sets of the scalar field. Thus, if we have a level set of points for $f(x, y, z) = c$ and $\vec{\nabla} f \neq \vec{0}$ at any of these points, then we have a means of describing the vector $\hat{\mathbf{n}}$ that is normal to the surface at each point. Specifically, if the point (x_0, y_0, z_0) are on the implicitly defined surface $f(x, y, z) = c$, then the normal vector at that point is given by

$$\hat{\mathbf{n}}(x_0, y_0, z_0) = \frac{\vec{\nabla} f(x_0, y_0, z_0)}{|\vec{\nabla} f(x_0, y_0, z_0)|}.$$

Previously, we wished to compute flux integrals of vector fields through surfaces. But, we took the simplest cases of surfaces (specifically where the surfaces were flat and had a constant normal vector). Now, we could conceivably compute flux integrals across implicitly defined surfaces!

One more step must be taken into account. We must know how the implicit function “stretches” out our Cartesian coordinates. Say, for example, we can solve for z in our equation for the level surface Σ . That is, if we have $f(x, y, z) = c$, then we can solve for z in terms of x and y to find $z = g(x, y)$. Then, the area is stretched by a factor

$$d\Sigma = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} dx dy.$$

We will cover this in more depth later when we consider explicit surfaces.

Example 12.1.4: Surface Normal to Unit Sphere

Consider the unit sphere defined implicitly by the level set to

$$f(x, y, z) = x^2 + y^2 + z^2 = 1.$$

Then, we can compute the surface normal by taking

$$\vec{\nabla} f = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}.$$

Then, we also have that

$$|\vec{\nabla} f| = \sqrt{4x^2 + 4y^2 + 4z^2} = 2\sqrt{x^2 + y^2 + z^2}.$$

Note that we are requiring that $x^2 + y^2 + z^2 = 1$ by definition, and thus

$$|\vec{\nabla} f| = 2,$$

along the unit sphere. Hence the surface normal to the unit sphere is given by

$$\hat{\mathbf{n}} = \frac{\vec{\nabla} f}{|\vec{\nabla} f|} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Thus, the surface normal to the sphere points in the same direction as the point on the sphere does!

Exercise 12.1.2. Plot the normal vectors to the unit sphere as well as the sphere.

12.1.2 Implicit Function Theorem

There is a great relationship between implicit surfaces and graphs of functions. In general, graphs of functions provide the nicest possible description of a surface. Thus, one would appreciate the ability to be able to work with graphs while still defining a surface implicitly.

Theorem 12.1.2: Implicit Function Theorem

Let $f(x, y, z) = c$ define an implicit surface. Then where $\frac{\partial f}{\partial z} \neq 0$, we can describe the surface as the (possibly many) graph of functions $z = g(x, y)$. That is, the set of points $(x, y, g(x, y))$ describes a portion of the implicit surface.

Remark 12.1.1. There is in fact nothing special about the choice of graphing a function over the xy -plane. One could instead have taken $\frac{\partial f}{\partial x} \neq 0$ and produced a graph $(g(y, z), y, z)$ instead! Given that, one could graph the surface over any plane in \mathbb{R}^3 .

This would not be a complete treatment if we did not provide an example for this. So let us work through this description for the unit sphere.

Example 12.1.5: The Sphere as Two Graphs

Consider the unit sphere defined by

$$f(x, y, z) = x^2 + y^2 + z^2 = 1.$$

Then we have

$$\frac{\partial f}{\partial z} = 2z,$$

and hence as long as $z \neq 0$, we can build the unit sphere via graphs. Specifically, we have

$$z = \pm \sqrt{1 - x^2 - y^2}.$$

Taking the graph $(x, y, \sqrt{1 - x^2 - y^2})$ yields the following.

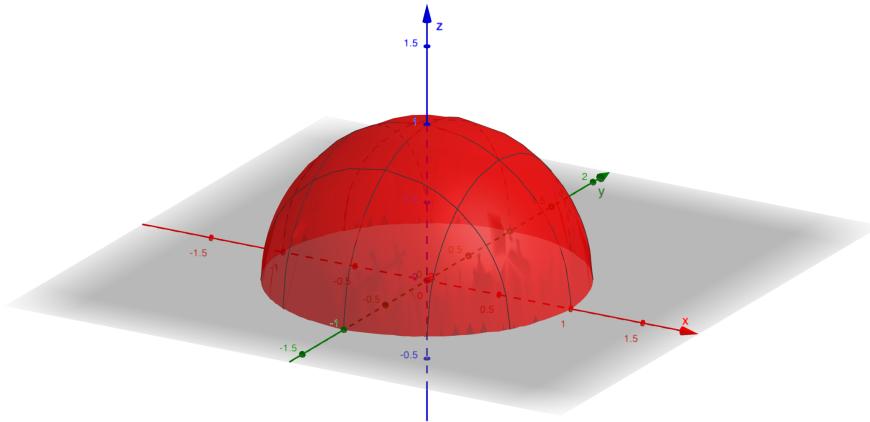


Figure 12.5: The northern hemisphere given by the graph $(x, y, \sqrt{1 - x^2 - y^2})$.

Exercise 12.1.3. Graph the other portion of the sphere by plotting $(x, y, -\sqrt{1 - x^2 - y^2})$ to see that the combination of these two graphs covers the whole sphere.

12.2 Surfaces as Graphs

Implicit surfaces provided a concise description for a surface as a level set of some scalar field. One could then compute the surface normal. However, to do further computations, it tends to be easiest to use an explicit description of a surface. Luckily, the implicit function theorem provides us a means of construction a surface as a graph over the xy -plane (or really any other plane in \mathbb{R}^3). So, it is worth continuing working with graphs of

functions.

Suppose that we have a function $g(x, y)$ and we construct the surface Σ given by the graph $(x, y, g(x, y))$. One would like to construct the surface normal $\hat{\mathbf{n}}$ and compute the **area form** $d\Sigma$. Up until now, we have had no trouble defining $d\Sigma$ as our surfaces have been rather nice.

Our methodology for constructing these quantities will be to work very locally on a surface. In other words, it is helpful to know what a surface looks like up very close. In this case, the surface is best approximated by a plane. This is analogous to how you can approximate functions of a single variable by a line.

Exercise 12.2.1. Compute the tangent line to $f(x) = 2x^2 + 5$ at the point $x_0 = 3$.

12.2.1 Equation for a Plane

We haven't worked much with planes in space yet, but we have seen surfaces. In some sense, planes are the easiest surfaces. They are, after all, linear objects.

Example 12.2.1: Plane and Normal

The implicit equation for a plane is given by

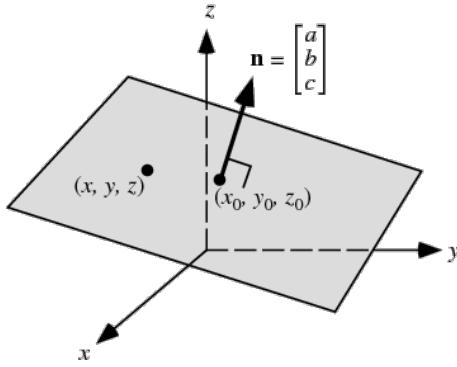
$$ax + by + cz + d = 0.$$

Notice that this is a linear equation.

Then the surface normal to the plane is given by

$$\hat{\mathbf{n}} = \frac{\vec{\nabla} f}{|\vec{\nabla} f|} = \frac{1}{\sqrt{a^2 + b^2 + c^2}} \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

We can see a diagram of this here.



12.2.2 Tangent Planes and Surface Normals

Now, if we are given a surface (defined as a level surface or as the graph of a function), we can compute an approximation at a point called the **tangent plane**. The means in which we compute this comes from computing the tangent vectors to curves along the surface. If we have Σ given by $(x, y, g(x, y))$, then we can create a curve along the surface aligned

in the x -direction by letting y be constant. So, choose $y = k$ and we then have a curve given by the points

$$\vec{\gamma}_1(x) = \begin{pmatrix} x \\ k \\ g(x, k) \end{pmatrix}.$$

Then, the tangent vector is given by

$$\frac{d}{dx} \vec{\gamma}_1(x) = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial g}{\partial x}(x, k) \end{pmatrix} = \dot{\vec{\gamma}}_1.$$

In other words, we can simply compute a tangent vector in the x -direction on the graph by taking

$$\frac{\partial}{\partial x} \begin{pmatrix} x \\ y \\ g(x, y) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial g}{\partial x} \end{pmatrix}.$$

Similarly, we can compute the tangent vector in the y -direction by

$$\frac{\partial}{\partial y} \begin{pmatrix} x \\ y \\ g(x, y) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial g}{\partial y} \end{pmatrix}.$$

This gives us a means of computing the tangent plane in two ways. First, we have that the equation for the tangent plane to the graph of a function $g(x, y)$ at the point $(x_0, y_0, g(x_0, y_0))$ is given by

$$z - g(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0)(y - y_0).$$

Notice how this equation mimics the equation for a tangent line. Similarly, we can compute the unit vector normal to the tangent plane by taking

$$\hat{n} = \frac{\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2}{|\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2|} = \frac{1}{\sqrt{1 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2}} \begin{pmatrix} -\frac{\partial g}{\partial x} \\ -\frac{\partial g}{\partial y} \\ 1 \end{pmatrix}.$$

It then follows that the vector normal to the tangent plane is indeed the surface normal at that point. Thus, we have a nearly full description of the surface. Also, notice the denominator is very much related to the area form $d\Sigma$. We'll work on computing this quantity next.

Example 12.2.2: Tangent Plane and Normal to a Paraboloid

onsider the function

$$f(x, y) = -x^2 - y^2.$$

Then the graph of the function is given by plotting the points

$$(x, y, f(x, y)).$$

We compute the tangent plane by computing partial derivatives. We take

$$\begin{aligned}\frac{\partial f}{\partial x} &= -2x \\ \frac{\partial f}{\partial y} &= -2y.\end{aligned}$$

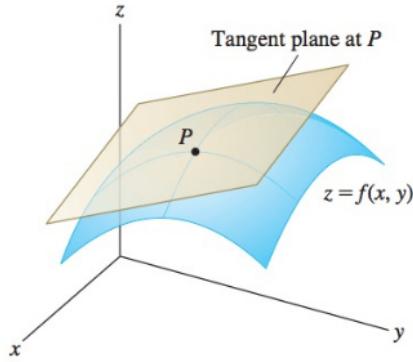
Then the equation for a tangent plane at the point $(x_0, y_0, f(x_0, y_0))$ is given by

$$z - f(x_0, y_0) = \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0).$$

So in our case, we have

$$z - (-x_0^2 - y_0^2) = -2x_0(x - x_0) - 2y_0(y - y_0).$$

Pictorially, it looks as follows (letting $p = (x_0, y_0, f(x_0, y_0))$):



To compute the surface normal, we compute

$$\dot{\gamma}_1 = \begin{pmatrix} 1 \\ 0 \\ -2x \end{pmatrix} \quad \text{and} \quad \dot{\gamma}_2 = \begin{pmatrix} 0 \\ 1 \\ -2y \end{pmatrix}.$$

Computing the cross product yields,

$$\begin{pmatrix} 2x \\ 2y \\ 1 \end{pmatrix},$$

and we can normalize this vector by first computing

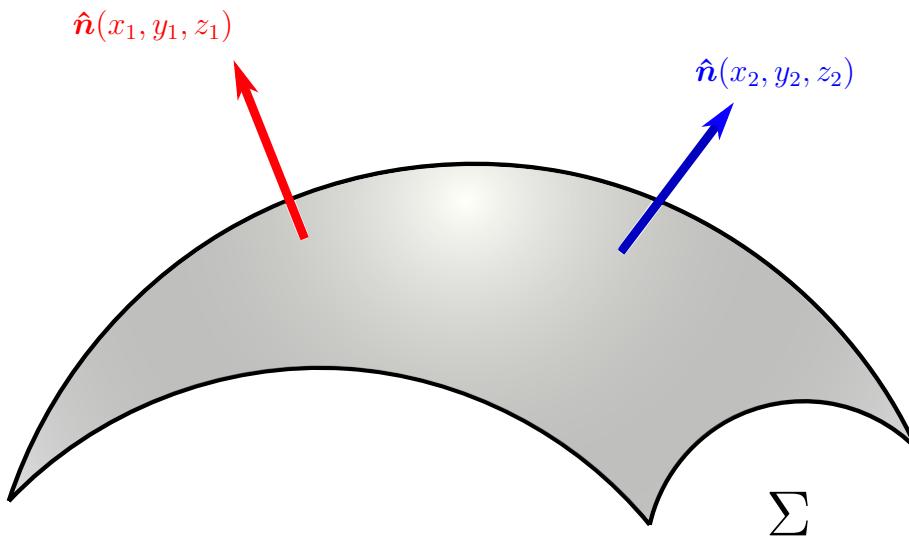
$$\sqrt{1 + (-2x)^2 + (-2y)^2} = \sqrt{1 - 4x^2 - 4y^2}.$$

Thus, our surface normal is given by

$$\hat{n} = \frac{1}{\sqrt{1 - 4x^2 - 4y^2}} \begin{pmatrix} 2x \\ 2y \\ 1 \end{pmatrix}.$$

12.3 Flux and Surface Integrals

There was nothing too special about integrating over a curve, just as we saw with integration of scalar fields. In that case, we could integrate over any type of region that we would like. Here, we can essentially do the same (one must be careful when the dimension not equal to three). Fundamentally, what one would like to measure in the case of integration of a vector field \vec{V} over a surface Σ is how much of the vector field passes through this surface. If we have a surface Σ , then there is a unique (outward) normal vector to this surface denoted by \hat{n} . That is, there is a unique vector that is perpendicular to the surface at that point. We can see this in the following figure.



Keep in mind that the normal vector to a surface may change over time! For example, the normal vector to a sphere will change, whereas the normal vector to a plane will not. This means that, in general, the normal vector is really a vector field $\hat{n}(x, y, z)$. We will concentrate more on surfaces later, so for now we will take our surfaces to be planar.

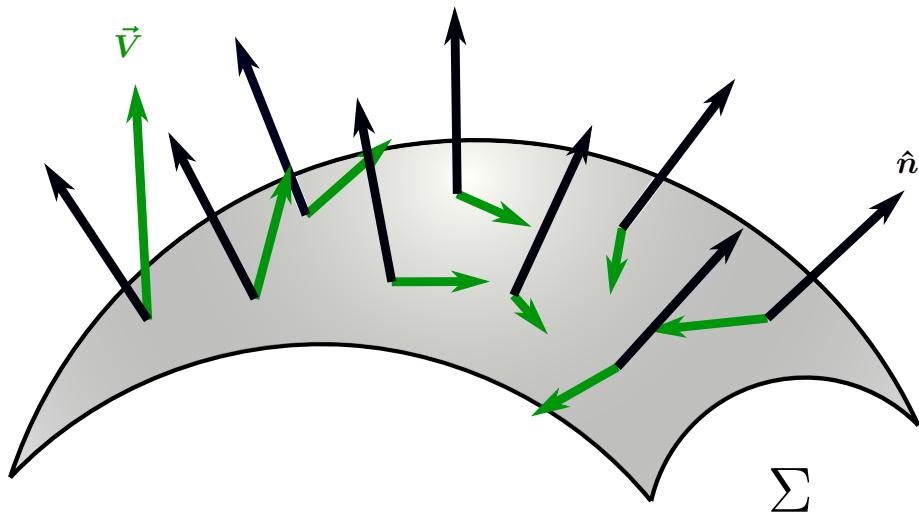
The amount the vector field passes through the surface at the point (x, y, z) is then given by the scalar quantity

$$\vec{V}(x, y, z) \cdot \hat{n}(x, y, z).$$

Then, if we integrate over the whole surface Σ we have

$$\iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma.$$

As we have seen, how we compute this will depend on the surface Σ . The picture for flux can be seen as follows.



Above, the green arrows represent vectors coming from the vector field \vec{V} and the black arrows represent the surface normal \hat{n} at the corresponding point. By taking the dot product of the vector field with the unit vector at each point ($\vec{V}(x, y, z) \cdot \hat{n}(x, y, z)$) we compute the amount of the field flowing through that point.

Example 12.3.1: Total Flux Through a Plane

Consider the vector field $\vec{V}(x, y, z) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and the planar surface Σ given by $0 \leq x \leq 1$, $2 \leq y \leq 3$ and $z = 5$. Then, the normal to this surface is given by $\hat{n}(x, y, z) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \hat{z}$. Note that we could have instead taken $-\hat{n}$, but we chose

the other orientation, and orientation is always a choice we can make. Here, the normal vector does not change as we change our position in space.

We now wish to compute

$$\iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma.$$

Now, given our surface defined above, we can put

$$\begin{aligned} \iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma &= \int_2^3 \int_0^1 \vec{V}(x, y, 5) \cdot \hat{z} dx dy \\ &= \int_2^3 \int_0^1 5 dx dy \\ &= \int_2^3 5 dy \\ &= 5. \end{aligned}$$

12.4 Area Form and Integration

Let us begin this section by computing the area of the unit disk in the plane. Specifically, this is the set of points inside the unit circle given by the equation $x^2 + y^2 \leq 1$. How should we compute this area?

Example 12.4.1: Area of Unit Disk

Let the surface Σ be the unit disk given by the set of points in the plane $x^2 + y^2 \leq 1$. We know that the area should be π , and so if we integrate

$$\iint_{\Sigma} d\Sigma,$$

we should find this to be equal to the area of the unit disk. In this case, since our surface is flat and in the xy -plane, we have that $d\Sigma = dx dy$, and so we just need to find our bounds of integration. If we solve for x in the given equation, we have

$$\sqrt{x^2} \leq \sqrt{1 - y^2},$$

and note that $\sqrt{x^2} = |x|$ and thus this amounts to

$$-\sqrt{1 - y^2} \leq x \leq \sqrt{1 - y^2}.$$

So, as y varies from -1 to 1 , this describes how x should vary. So, we can write our integral as

$$\iint_{\Sigma} d\Sigma = \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dx dy.$$

We can compute this integral in the standard way

$$\begin{aligned} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dx dy &= \int_{-1}^1 2\sqrt{1-y^2} dy \\ &= \pi. \end{aligned}$$

This is the correct area of the unit disk.

Exercise 12.4.1. Fill in the gaps in the integral for the previous example.

Lastly, we need to describe the area form $d\Sigma$. The idea of $d\Sigma$ is to capture the “stretching” that occurs when we graph a function. Geometrically, what we are doing when we take the graph of a function is we are taking a region in the plane, and applying a function to this region that lifts the area up and stretches it into the desired surface. For example, if we take the unit disk in the plane given by the set of points (x, y) satisfying $x^2 + y^2 \leq 1$. Then, we can create the northern hemisphere surface by taking the graph $(x, y, \sqrt{1 - x^2 - y^2})$. Through this process, the unit disk is transformed into the northern hemisphere. One can note that the area of the unit disk is π and the area of the northern

hemisphere of the unit sphere is 2π . So, it must be that this act of graphing deforms area!

We want to know how the area is deformed at every single point on our surface, and so we need a way to compute this. Recall that the cross product of two vectors in \mathbb{R}^3 gives us the area of the parallelogram defined by the two vectors. Specifically, if we have $\vec{u}, \vec{v} \in \mathbb{R}^3$, then the area of the parallelogram is given by $A = |\vec{u} \times \vec{v}|$. Thinking along these lines, if we can compute an infinitesimal parallelogram on our surface, we can compute what the area is. Specifically, what we will do is take an infinitesimal movement along our surface x -direction and y -direction, and see what the area given by these infinitesimal vectors will be. In other words, we want to compute an infinitesimal version of

$$|\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2|,$$

where $\dot{\vec{\gamma}}_1$ and $\dot{\vec{\gamma}}_2$ are tangent vectors at the point of interest. Recall that

$$\dot{\vec{\gamma}}_1 = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial g}{\partial x} \end{pmatrix} \quad \text{and} \quad \dot{\vec{\gamma}}_2 = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial g}{\partial y} \end{pmatrix}.$$

We can make these of infinitesimal lengths by taking the x -coordinate to be the infinitesimal dx , the y -coordinate to dy , and the z -coordinate to be dz . Hence, we want to compute

$$\left| \begin{pmatrix} dx \\ 0 \\ dz \end{pmatrix} \times \begin{pmatrix} 0 \\ dy \\ dz \end{pmatrix} \right|.$$

However, we can note that $z = g(x, y)$ is a function. Thus, we can write dz in terms of dx and dy . Specifically, we can compute the **differential**

$$dz = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy.$$

Let us pause briefly to define this.

Definition 12.4.1: Differential of a Scalar Field

Given a scalar field $f(x, y, z)$, the *differential* df is given by

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

Returning to our work before, we were wishing to compute

$$\left| \begin{pmatrix} dx \\ 0 \\ \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy \end{pmatrix} \times \begin{pmatrix} 0 \\ dy \\ \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy \end{pmatrix} \right|.$$

Computing this cross product yields

$$\left| \begin{pmatrix} -\frac{\partial g}{\partial x} dy dz \\ \frac{\partial g}{\partial y} dx dz \\ dxdy \end{pmatrix} \right|.$$

Let us compute further,

$$\begin{pmatrix} -\frac{\partial g}{\partial x} dy dz \\ \frac{\partial g}{\partial y} dx dz \\ dxdy \end{pmatrix} = \begin{pmatrix} -\frac{\partial g}{\partial x} dxdy - \frac{\partial g}{\partial y} dy^2 \\ \frac{\partial g}{\partial x} dx^2 + \frac{\partial g}{\partial y} dxdy \\ dxdy \end{pmatrix}.$$

Now, if we are trying to compute area, then the area generated by dx^2 is 0 since no area is contained in this infinitesimal parallelogram. Likewise $dy^2 = 0$ as well. Hence, our result is

$$\begin{pmatrix} -\frac{\partial g}{\partial x} dxdy \\ \frac{\partial g}{\partial y} dxdy \\ dxdy \end{pmatrix}.$$

Then, computing the length of this vector yields

$$\begin{aligned} \left| \begin{pmatrix} -\frac{\partial g}{\partial x} dxdy \\ \frac{\partial g}{\partial y} dxdy \\ dxdy \end{pmatrix} \right| &= \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 (dxdy)^2 + \left(\frac{\partial g}{\partial y}\right) (dxdy)^2 + (dxdy)^2} \\ &= \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} dxdy. \end{aligned}$$

Thus, after this tedious computation, we have that for a surface defined by a graph

$$d\Sigma = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} dxdy.$$

Example 12.4.2: Area form on the Northern Hemisphere

Consider the graph of the northern hemisphere given by

$$(x, y, \sqrt{1 - x^2 - y^2}),$$

for x and y values satisfying $x^2 + y^2 \leq 1$. Letting $g(x, y) = \sqrt{1 - x^2 - y^2}$, we have that

$$\frac{\partial g}{\partial x} = -\frac{x}{\sqrt{1 - x^2 - y^2}} \quad \text{and} \quad \frac{\partial g}{\partial y} = -\frac{y}{\sqrt{1 - x^2 - y^2}}.$$

Then we have

$$\begin{aligned} d\Sigma &= \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1} dxdy \\ &= \sqrt{\frac{1}{1 - x^2 - y^2}} dxdy. \end{aligned}$$

If we have computed this correctly, then we should have that the surface area of the northern hemisphere is given by

$$\iint_{\Sigma} d\Sigma.$$

Now, we require that $x^2 + y^2 \leq 1$, and we can find

$$\sqrt{x^2} \leq \sqrt{1 - y^2},$$

which means that

$$-\sqrt{1 - y^2} \leq x \leq \sqrt{1 - y^2},$$

as y ranges from -1 to 1 . Hence, we have that

$$\iint_{\Sigma} d\Sigma = \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sqrt{\frac{1}{1-x^2-y^2}} dx dy = 2\pi,$$

which is indeed the area of the northern hemisphere.

Remark 12.4.1. The above integral is rather hard to compute in Cartesian coordinates. Thus, we will be motivated to choose a different way of computing that integral to simplify this problem.

Now, if we are solely integrating along a surface defined by a graph, then notice that we have the following nice relationship. We have that

$$\begin{aligned}\hat{\mathbf{n}} d\Sigma &= \frac{\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2}{|\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2|} |\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2| dx dy \\ &= (\dot{\vec{\gamma}}_1 \times \dot{\vec{\gamma}}_2) dx dy \\ &= \begin{pmatrix} -\frac{\partial g}{\partial x} \\ -\frac{\partial g}{\partial y} \\ 1 \end{pmatrix} dx dy.\end{aligned}$$

Putting this together with the previous section, we can compute a nice example.

Example 12.4.3: Flux Through Northern Hemisphere

Consider the vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

and the graph of the northern hemisphere surface Σ given by $(x, y, g(x, y))$ where

$$g(x, y) = \sqrt{1 - x^2 - y^2}.$$

We want to integrate

$$\iint_{\Sigma} \vec{V} \cdot \hat{\mathbf{n}} d\Sigma.$$

Then, we have computed the normal vector to the whole unit sphere to be

$$\hat{\mathbf{n}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

as well as the area form

$$d\Sigma = \sqrt{\frac{1}{1-x^2-y^2}} dx dy.$$

Along this surface. Thus, we have

$$\begin{aligned} \iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma &= \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \sqrt{\frac{1}{1-x^2-y^2}} dx dy \\ &= \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} (x^2 + y^2 + z^2) \sqrt{\frac{1}{1-x^2-y^2}} dx dy. \end{aligned}$$

Note that along the unit sphere we have that $x^2 + y^2 + z^2 = 1$, and thus our integral reduces to

$$\iint_{\Sigma} \vec{V} \cdot \hat{n} d\Sigma = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sqrt{\frac{1}{1-x^2-y^2}} dx dy.$$

To evaluate this integral, we will learn to use different coordinates!

In summary, if a surface is given as a graph, we have all the tools to integrate. One should then revisit the idea of integrating over a implicit surface. Notice in that case, we can define the surface using multiple graphs, and so this reduces to integrating along each necessary graph.

13

Coordinate systems

In the prequel, we discussed the complex plane and its two natural coordinate representations. These were the Cartesian and polar coordinates. Recall that if we were given a complex number $z \in \mathbb{C}$, then we could write

$$z = x + iy,$$

which were the Cartesian coordinates for z . We could also represent z as

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

which was the polar coordinate representation. One found that, for example, addition was nice to compute in the Cartesian representation as

$$z_1 + z_2 = (x_1 + iy_1) + (x_2 + iy_2) = x_1 + x_2 + i(y_1 + y_2).$$

However, multiplication of complex numbers was much nicer to compute in polar coordinates as

$$z_1 z_2 = (r_1 e^{i\theta_1}) (r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

The take-home message is that different choices of coordinates provide a (potentially) nicer outlook for certain operations or problems. It is with this in mind that we extend our representation of \mathbb{R}^3 to new coordinate systems.

13.1 Cylindrical Coordinates

Our first glance at new coordinates comes in the form of *cylindrical coordinates*. Let us quickly define the coordinates as

$$\begin{aligned}x &= \rho \cos \theta \\y &= \rho \sin \theta \\z &= z.\end{aligned}$$

Here, the variables are defined over the ranges $\rho \geq 0$, $0 \leq \theta < 2\pi$, and $-\infty < z < \infty$. Ignoring z , this is analogous to the polar representation of the complex plane.

Exercise 13.1.1. Compare the real and imaginary Cartesian components of the polar representation in \mathbb{C} to the x and y components of the cylindrical coordinates in \mathbb{R}^3 .

Thus, we may now specify a vector in the plane by providing

$$\vec{v} = \begin{pmatrix} \rho \\ \theta \\ z \end{pmatrix}.$$

The type of coordinates we will be using to represent a vector will be clear from context. Most instructive is to provide an image of what the coordinates look like. If we indeed take the vector above, we can plot this in \mathbb{R}^3 as follows.

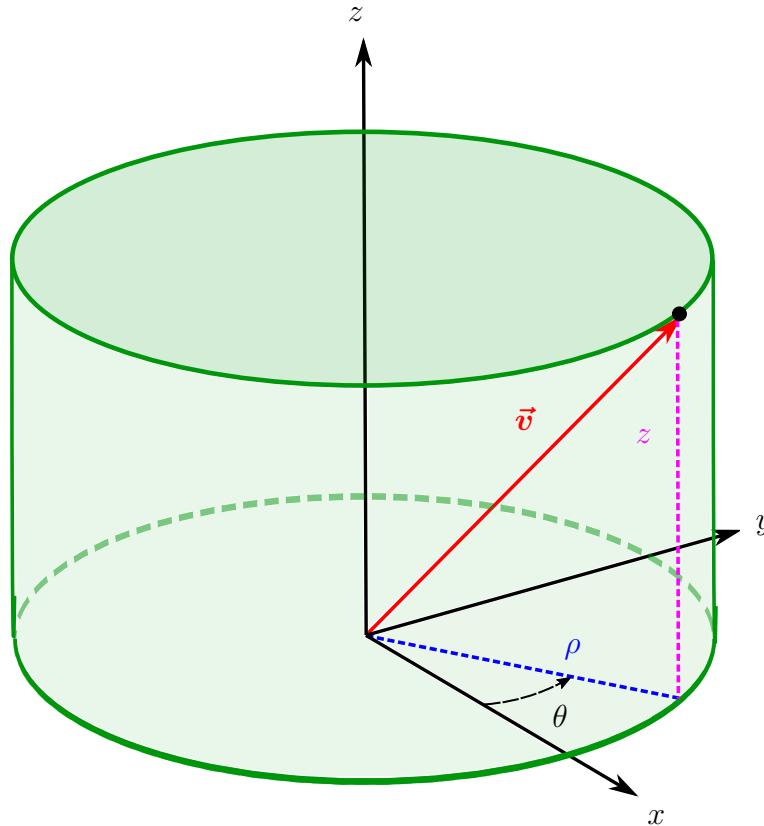


Figure 13.1: A diagram of cylindrical coordinates in \mathbb{R}^3 .

The idea is that at every height z above the xy -plane, we are (essentially) using polar coordinates. In the Figure ??, one can see how a vector \vec{v} can be represented in these coordinates geometrically. We need to supply the angle θ that increases from the x -axis in a counter-clockwise manner; we supply the distance from the z -axis as the variable ρ ; we supply the height above the xy -plane as the variable z (which is no different than the Cartesian coordinates).

We can also invert these coordinates so that we could construct a cylindrical represen-

tation of a vector given in Cartesian coordinates. That is, we have

$$\begin{aligned}\rho &= \sqrt{x^2 + y^2} \\ \theta &= \arctan\left(\frac{y}{x}\right) \text{ if } x > 0, \\ \theta &= \arctan\left(\frac{y}{x}\right) + \pi \text{ if } x < 0, \\ z &= z.\end{aligned}$$

Exercise 13.1.2. Either re-derive these facts for yourself, or revisit the section in the prequel on polar coordinates.

It's worth comparing the cylindrical coordinate system to the Cartesian coordinate system by first seeing what objects we can create in a natural way. If we take the variables x , y , and z to be over some range, say

$$x_0 \leq x \leq x_1, \quad y_0 \leq y \leq y_1, \quad z_0 \leq z \leq z_1,$$

then this will generate a rectangular prism.

[insert figure](#)

We could also see more natural objects by holding certain variables constant. In the case for cartesian coordinates, if we hold the variable x to be constant, and let the variables y and z be free to vary, then this will construct a plane parallel to the yz -plane.

[Insert figure](#)

Likewise, we can hold the variable y or z constant and receive something analogous.

Exercise 13.1.3. In Cartesian coordinates, first take the variable y to be constant and draw a picture of the object (plane) that you will get. Repeat this by then taking z to be constant.

We could also hold two variables constant at once and generate another natural object. Say, for example, we hold both x and y constant, then this will generate a line parallel to the z axis that passes through a point (x, y) in the xy -plane.

[insert figure](#)

This same analysis could be performed in cylindrical coordinates. Say that we let the variables range over

$$\rho_0 \leq \rho \leq \rho_1, \quad \theta_0 \leq \theta \leq \theta_1, \quad z_0 \leq z \leq z_1,$$

then what we receive is a cut-out of a solid cylinder.

[insert figure](#)

We could then hold ρ constant, and note that this will create a cut-out of a surface of a cylinder.

[insert figure](#)

If we hold θ constant, then this will cut-out a plane that is perpendicular to the xy -plane.

[insert figure](#)

Finally, if we hold z constant, then we will generate a disk at the given height z .

[insert figure](#)

Exercise 13.1.4. Try holding two variables constant and seeing what shapes will be created in those cases. Realize the following.

- If ρ and θ are constant, this will create a line parallel to the z axis.
- If ρ and z are constant, this will create a circle of radius ρ at height z .
- If θ and z are constant, this will create a ray.

13.1.1 Functions and Calculus in Cylindrical Coordinates

We defined three types of functions previously. These were curves $\vec{\gamma}$, scalar fields f , and vector fields \vec{V} . But, our description for these was always provided in Cartesian coordinates. Now, we could specify a curve by

$$\vec{\gamma}(t) = \begin{pmatrix} \rho(t) \\ \theta(t) \\ z(t) \end{pmatrix},$$

a scalar field by

$$f(\rho, \theta, z),$$

and a vector field by

$$\vec{V}(\rho, \theta, z) = V_1(\rho, \theta, z)\hat{x} + V_2(\rho, \theta, z)\hat{y} + V_3(\rho, \theta, z)\hat{z}.$$

Or, indeed,

$$\vec{V}(\rho, \theta, z) = V_1(\rho, \theta, z)\hat{\rho} + V_2(\rho, \theta, z)\hat{\theta} + V_3(\rho, \theta, z)\hat{z}.$$

Perhaps it is confusing to see what we are describing in the case of the curve and vector field.

For the curve $\vec{\gamma}(t)$ in cylindrical coordinates, we will describe how the coordinates ρ , θ , and z change with respect to the variable t . But, for a vector field $\vec{V}(\rho, \theta, z)$, we will see how the x , y , and z components of the vector field depend on ρ , θ , and z . One could generate new basis vectors $\hat{\rho}$, $\hat{\theta}$, and \hat{z} , but we will get to this in a bit. The key point of using new coordinates is to simplify expressions that are not well suited for the Cartesian coordinate system. Let's take an example.

Example 13.1.1: A Cylindrical Curve

Take for example the following curve in space

$$\vec{\gamma}(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \\ 0 \end{pmatrix} = \cos(t)\hat{x} + \sin(t)\hat{y}.$$

We have seen this curve before, and it parameterizes the unit circle in the plane. However, the description of this curve can be drastically simplified. Notice that the unit circle satisfies $\rho(t) = 1$ since the distance from the z -axis never changes over time. Also, one can see that $\theta(t) = t$, since θ changes constantly over time. Specifically, we notice that we complete a full revolution when $t = 2\pi$.

Recall that $x = \cos(\theta)$ and $y = \sin(\theta)$, and for a curve in Cartesian coordinates we

provide

$$x(t), \quad y(t), \quad z(t).$$

Since $x(t) = \rho(t) \cos(\theta(t))$ and $y(t) = \rho(t) \sin(\theta(t))$, we can see that we must have $\rho(t) = 1$ $\theta = t$. Lastly, $z(t) = 0$, which does not change under these new coordinates. So, one could write the curve in cylindrical coordinates as

$$\vec{\gamma}(t) = \begin{pmatrix} 1 \\ t \\ 0 \end{pmatrix},$$

which is considerably more simple.

In the previous example, we wrote

$$\vec{\gamma}(t) = \begin{pmatrix} 1 \\ t \\ 0 \end{pmatrix},$$

which through context specifies $\rho(t) = 1$, $\theta(t) = t$, and $z(t) = 0$. But, we have written this as a vector, and thus we should be considering some basis vectors in which this is inherently using. It turns out that we can put

$$\vec{\gamma}(t) = 1\hat{\rho} + t\hat{\theta} + 0\hat{z},$$

where $\hat{\rho}$ and $\hat{\theta}$ are the cylindrical unit vector(fields). Again, we will discuss these vectors in a bit.

In general, we can describe a curve in cylindrical coordinates by providing a $\rho(t)$, $\theta(t)$, and $z(t)$. Then, in order to compute the tangent vector, we do need to take into account how the time derivatives of the vector components $\hat{\rho}$, $\hat{\theta}$, and \hat{z} change in relation to one another. Without going into the derivation, it turns out that given a curve $\vec{\gamma}(t) = \rho(t)\hat{\rho} + \theta(t)\hat{\theta} + z(t)\hat{z}$ we have the tangent (velocity) vector

$$\dot{\vec{\gamma}}(t) = \dot{\rho}(t)\hat{\rho} + \rho(t)\dot{\theta}(t)\hat{\theta} + \dot{z}(t)\hat{z}. \quad (13.1.1)$$

It is important to see that the θ direction needs to be multiplied by the distance from the z -axis, ρ , in order to have the proper units! Then, normal (acceleration) vector is

$$\ddot{\vec{\gamma}}(t) = (\ddot{\rho}(t) + r\dot{\theta}^2)\hat{\rho} + (2\dot{\rho}(t)\dot{\theta} + r\ddot{\theta}^2)\hat{\theta} + \ddot{z}(t)\hat{z}. \quad (13.1.2)$$

The basis vector fields $\hat{\rho}$, $\hat{\theta}$, and \hat{z} are always orthonormal, so computing the speed yields

$$|\dot{\vec{\gamma}}| = \sqrt{\dot{\rho}^2 + \rho^2\dot{\theta}^2 + \dot{z}^2}. \quad (13.1.3)$$

Example 13.1.2: A Cylindrical Scalar Field

Consider the scalar field

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

One can claim that this scalar field is much more natural in cylindrical coordinates. Let's see why that is. Recall that $x = \rho \cos(\theta)$ and $y = \sin(\theta)$, and hence if we plug these substitutions into the field, we get

$$f(\rho, \theta, z) = \frac{z}{\sqrt{\rho^2 \cos^2(\theta) + \rho^2 \sin^2(\theta)}} = \frac{z}{\rho}.$$

Thus, this field is really just coming from a ratio of z to ρ . In Cartesian coordinates, this is not as clear.

Though the scalar field example may seem somewhat contrived, it is not so. Indeed, there are many descriptions of our world that are actually described most easily using this coordinate system from the beginning. For example, if one has a constantly charged wire aligned with the z -axis (which you could always choose the axis of alignment), then the strength of this field would fall off according to a function of just ρ . Hence, the cylindrical description is not only adequate, but it is natural.

13.1.2 Cylindrical Unit Vector Fields

In Cartesian coordinates, we took the basis vectors to be \hat{x} , \hat{y} , and \hat{z} . At any point in space, we chose to use this same basis in order to describe curves and vector fields. In principle, however, one does not have to choose the same basis vectors at each point in space! Thinking this way is new, but it is important.

Recall that if we are given a point in space in cylindrical coordinates, that we can convert this back to Cartesian coordinates via

$$\begin{aligned}\rho(x, y, z) &= \sqrt{x^2 + y^2} \\ \theta(x, y, z) &= \arctan\left(\frac{y}{x}\right) \text{ if } x > 0, \\ \theta(x, y, z) &= \arctan\left(\frac{y}{x}\right) + \pi \text{ if } x < 0, \\ z(x, y, z) &= z.\end{aligned}$$

As we've written above, we can think of each of these as scalar fields of x , y , and z . That is, for example, we took $\rho(x, y, z) = \sqrt{x^2 + y^2}$. From this perspective, we can generate new vector fields by taking the gradient of these coordinate functions. For example,

$$\vec{\nabla} \rho(x, y, z) = \frac{x}{\sqrt{x^2 + y^2}} \hat{x} + \frac{y}{\sqrt{x^2 + y^2}} \hat{y}.$$

Note that this vector field is normalized since

$$|\vec{\nabla} \rho(x, y, z)| = 1,$$

at every point in space. Thus, we will refer to this unit vector field as

$$\hat{\rho}(x, y, z) = \frac{x}{\sqrt{x^2 + y^2}} \hat{x} + \frac{y}{\sqrt{x^2 + y^2}} \hat{y}.$$

This vector field points away from the z -axis in the direction of increasing ρ at every point in space!

Likewise, we can compute the other unit vectors

$$\hat{\theta}(x, y, z) = \frac{\vec{\nabla}\theta}{|\vec{\nabla}\theta|} = \frac{-y}{\sqrt{x^2 + y^2}}\hat{x} + \frac{x}{\sqrt{x^2 + y^2}}\hat{y}.$$

and

$$\hat{z}(x, y, z) = \hat{z}.$$

Notice that \hat{z} is constant much like we found \hat{x} and \hat{y} to be constant in Cartesian coordinates.

Exercise 13.1.5. Compute the vector fields above. Then, plot those vector fields.

A primary concern for the basis vectors \hat{x} , \hat{y} , and \hat{z} was that these vector fields were orthonormal at every point. This meant that we could decompose vectors into this basis very nicely. It also meant that we could more easily compute a cross product. The same is true for the vector fields $\hat{\rho}$, $\hat{\theta}$, and \hat{z} .

Exercise 13.1.6. Show that the vector fields $\hat{\rho}$, $\hat{\theta}$, and \hat{z} are orthonormal.

insert figure

These vector fields come into play immediately. For example, we considered the curve given by $\rho(t) = 1$, $\theta(t) = t$ and $z(t) = 0$ which we wrote as

$$\vec{\gamma}(t) = \begin{pmatrix} 1 \\ t \\ 0 \end{pmatrix}.$$

But, this was really implying that we should write

$$\vec{\gamma}(t) = \rho(t)\hat{\rho}(x(t), y(t), z(t)) + \theta(t)\hat{\theta}(x(t), y(t), z(t)) + z(t)\hat{z} = \hat{\rho} + t\hat{\theta}.$$

Using the substitutions above, we have that

$$x(t) = \rho(t) \cos(\theta(t)) = \cos(t), \quad y(t) = \rho(t) \sin(\theta(t)) = \sin(t), \quad z(t) = 0,$$

and thus we arrive at

$$\vec{\gamma}(t) = \cos(t)\hat{x} + \sin(t)\hat{y},$$

which was the curve we started with.

Remark 13.1.1. It may seem a bit circular to describe quantities in this manner, but it turns out to simplify our expressions quite a bit.

Likely the most useful application is to use new orthonormal basis vectors to describe a vector field. Some vector fields in nature simply seem to align themselves along these cylindrical basis elements. For example, we saw that a charged wire generated a scalar field that could be described nicely in terms of these coordinates. One could also find that a charged wire generates a vector field in terms of these coordinates as well. In the next example, we can see a vector field that sits in this coordinate system nicely. One can interpret this vector field as the magnetic field as being related to one generated by a current carrying wire.

Example 13.1.3: A Cylindrical Vector Field

Consider the vector field

$$\vec{V}(x, y, z) = \begin{pmatrix} \frac{-y}{x^2+y^2} \\ \frac{x}{x^2+y^2} \\ 0 \end{pmatrix},$$

previously. If one plots this field, you can see that this vector field rotates in the θ -direction. Said another way, this vector field curls around the z -axis. Due to this nature, it may be apt to describe this field not in Cartesian coordinates, but in cylindrical coordinates. Let us write

$$\vec{V}(x, y, z) = \frac{-y}{x^2 + y^2} \hat{x} + \frac{x}{x^2 + y^2} \hat{y}.$$

Now, recall that

$$\hat{\theta} = \frac{-y}{\sqrt{x^2 + y^2}} \hat{x} + \frac{x}{\sqrt{x^2 + y^2}} \hat{y}, \quad \text{and} \quad \rho(x, y, z) = \sqrt{x^2 + y^2},$$

and we have

$$\vec{V}(\rho, \theta, z) = \frac{1}{\rho} \hat{\theta}.$$

Hence, we can see that this vector field rotates in the θ -direction with decreasing magnitude as we move further from the z -axis.

Exercise 13.1.7. Confirm that the substitution above is correct.

13.1.3 Integration in Cylindrical Coordinates

The next installment of our treatment of the cylindrical coordinate system is to determine how we can integrate functions. For the same reasoning as before, one may find that using a new coordinate system simplifies the problem at hand. But, we must be careful in how we set this up.

For a region in space Ω , we used a triple integral to integrate a function over this region. Specifically, in Cartesian coordinates we would put

$$\iiint_{\Omega} f d\Omega = \iiint_{\Omega} f(x, y, z) dx dy dz.$$

However, if our function f , for example, is a function of ρ , θ , and z , then we would need to integrate with respect to those variables. Likewise, we must then be able to describe the region Ω in these variables as well. We can make note of the fact that some regions (e.g., a solid cylinder) are more easily described in these coordinates anyway.

We must then determine the **volume form** $d\Omega$ in cylindrical coordinates. Let us think of the coordinate transformations as functions. That is, we have $x(\rho, \theta, z)$, $y(\rho, \theta, z)$, and $z(\rho, \theta, z)$. We can then collect these functions into a single transformation

$$\vec{T}(\rho, \theta, z) = \begin{pmatrix} x(\rho, \theta, z) \\ y(\rho, \theta, z) \\ z(\rho, \theta, z) \end{pmatrix} = \begin{pmatrix} \rho \cos(\theta) \\ \rho \sin(\theta) \\ z \end{pmatrix}.$$

Then, if we take the Jacobian of this transformation, we have

$$[J]_{\vec{T}} = \begin{pmatrix} \frac{\partial T_1}{\partial \rho} & \frac{\partial T_1}{\partial \theta} & \frac{\partial T_1}{\partial z} \\ \frac{\partial T_2}{\partial \rho} & \frac{\partial T_2}{\partial \theta} & \frac{\partial T_2}{\partial z} \\ \frac{\partial T_3}{\partial \rho} & \frac{\partial T_3}{\partial \theta} & \frac{\partial T_3}{\partial z} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\rho \sin(\theta) & 0 \\ \sin(\theta) & \rho \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The Jacobian is then the matrix (for a linear transformation) that tells us how our coordinates x , y , and z change as we change the variables ρ , θ , and z infinitesimally. Remember, the determinant of a matrix describes the stretching of space caused by the transformation, and here we have

$$\det([J]_{\vec{T}}) = \rho.$$

Intuitively, this tells us that our space is more stretched in these coordinates as we move further out in ρ .

Exercise 13.1.8. Confirm that the above determinant is correct.

It is a fact that this leads us to the volume form by

$$d\Omega = |\det([J]_{\vec{T}})| d\rho d\theta dz = \rho d\rho d\theta dz,$$

with the ordering $d\rho d\theta dz$ chosen due to the order of these coordinates that we started with. Thus, in cylindrical coordinates we would find

$$\iiint_{\Omega} f d\Omega = \iiint_{\Omega} f(\rho, \theta, z) \rho d\rho d\theta dz.$$

We can picture this as follows. [insert figure](#)

Example 13.1.4: Integral in Cylindrical Coordinates

Consider the scalar field $f(x, y, z) = \frac{z}{\sqrt{x^2+y^2}}$ over the region Ω which is the solid cylinder of radius $\rho = 1$ centered around the z -axis with $z \in [0, 5]$. Thus, we also have that $\theta \in [0, 2\pi)$. We want,

$$\iiint_{\Omega} f d\Omega.$$

We have converted $f(x, y, z)$ into cylindrical coordinates to get $f(x, y, z) = \frac{z}{\rho}$. This yields the integral

$$\int_0^5 \int_0^{2\pi} \int_0^1 \frac{z}{\rho} \rho d\rho d\theta dz.$$

We can then compute this integral

$$\begin{aligned} \int_0^5 \int_0^{2\pi} \int_0^1 \frac{z}{\rho} \rho d\rho d\theta dz &= \int_0^5 \int_0^{2\pi} \int_0^1 z d\rho d\theta dz \\ &= \int_0^5 \int_0^{2\pi} z d\theta dz \\ &= \int_0^5 2\pi z dz \\ &= \pi z^2 \Big|_0^5 \\ &= 25\pi. \end{aligned}$$

One can notice that we get rid of a pesky ∞ in this coordinate system as well. That is, when $x = y = 0$, we have $f(0, 0, z) = +\infty$. But, when it came to integration, this was not even noticed!

Exercise 13.1.9. One should challenge themselves to integrate the above function over the same region but in Cartesian coordinates.

13.1.4 Derivatives in Cylindrical Coordinates

In Cartesian coordinates we defined a few notions of derivatives depending on what types of functions (curves, scalar fields, or vector fields) we wanted to analyze. Not only that, but for vector fields, we saw a few different derivatives such as the Jacobian, divergence, and curl. Underlying these definitions was the Cartesian coordinate system. Fundamentally, each coordinate in the Cartesian system was given the same weighting. We can see this by looking at the volume form which was $dxdydz$. However, in cylindrical coordinates, the volume form has dependence on the current position in space. Specifically, we saw that the volume form was $\rho d\rho d\theta dz$.

We will choose now to concentrate on scalar fields and their derivatives. Recall that given a scalar field in Cartesian coordinates, that we have

$$\vec{\nabla} f(x, y, z) = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \quad \frac{\partial f}{\partial z} \right) = \frac{\partial f}{\partial x} \hat{x} + \frac{\partial f}{\partial y} \hat{y} + \frac{\partial f}{\partial z} \hat{z}.$$

But, in cylindrical coordinates, will it be as simple? The answer is no. Specifically, it's due to the same fact that the area swept out as we change in θ is given by $\rho d\theta$. So, it is not as simple as just taking partial derivatives. Also, we will have to write the gradient in terms of the cylindrical basis vectors $\hat{\rho}$, $\hat{\theta}$, and \hat{z} .

By doing some work, one can derive that the gradient in cylindrical coordinates is given by

$$\vec{\nabla} f(\rho, \theta, z) = \frac{\partial f}{\partial \rho} \hat{\rho} + \frac{1}{\rho} \frac{\partial f}{\partial \theta} \hat{\theta} + \frac{\partial f}{\partial z} \hat{z}.$$

The major difference is the inclusion of the $1/\rho$ factor in front of the partial derivative with respect to θ . It is important to remember that the gradient of a scalar field is a vector field!

We also discussed the Laplace operator Δ which was given by

$$\Delta f(x, y, z) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

Here, the Laplacian of a scalar field is once again a scalar field. But, seeing as we found the Laplacian as

$$\vec{\nabla} \cdot (\vec{\nabla} f(x, y, z)) = \Delta f(x, y, z),$$

we should expect that the Laplacian in cylindrical coordinates undergoes a change as well. Indeed, the cylindrical Laplacian is given by

$$\Delta f(\rho, \theta, z) = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial f}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}.$$

13.2 Spherical Coordinates

Perhaps one of the most useful coordinate systems is the *spherical coordinates* given by the variables r , θ , and ϕ . Let us first write the coordinate transformations by

$$x = r \cos \theta \sin \phi$$

$$y = r \sin \theta \sin \phi$$

$$z = r \cos \phi.$$

The picture to keep in mind with these coordinates is the following.

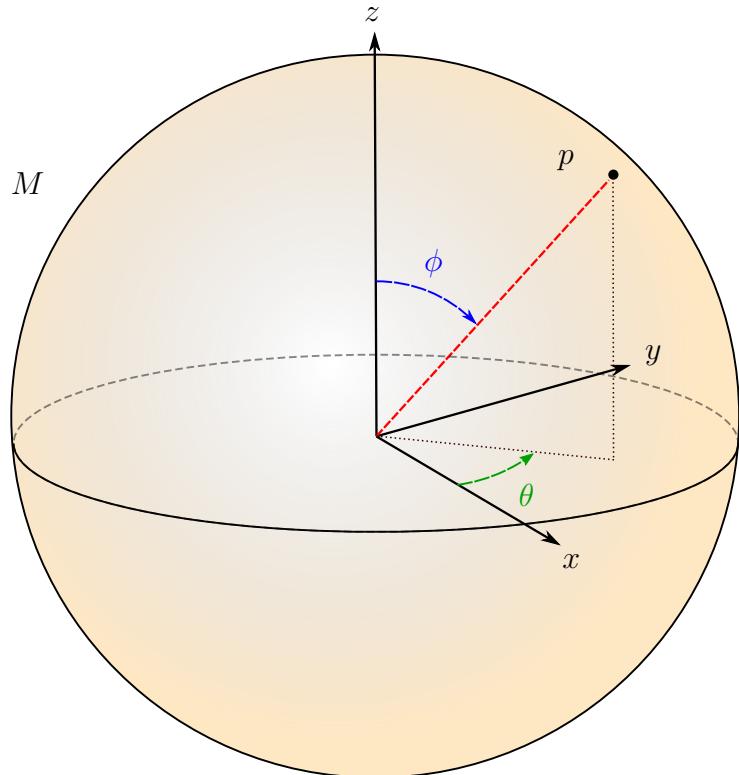


Figure 13.2: A diagram of spherical coordinates in \mathbb{R}^3 .

13.2.1 Integration in Spherical Coordinates

We begin by determining the volume form in spherical coordinates. Take the coordinate transformation

$$\vec{T}(r, \theta, \phi) = \begin{pmatrix} x(r, \theta, \phi) \\ y(r, \theta, \phi) \\ z(r, \theta, \phi) \end{pmatrix} = \begin{pmatrix} r \cos \theta \sin \phi \\ r \sin \theta \sin \phi \\ r \cos \phi \end{pmatrix}.$$

Then, if we take the Jacobian of this transformation, we have

$$[J]_{\vec{T}} = \begin{pmatrix} \frac{\partial T_1}{\partial r} & \frac{\partial T_1}{\partial \theta} & \frac{\partial T_1}{\partial \phi} \\ \frac{\partial T_2}{\partial r} & \frac{\partial T_2}{\partial \theta} & \frac{\partial T_2}{\partial \phi} \\ \frac{\partial T_3}{\partial r} & \frac{\partial T_3}{\partial \theta} & \frac{\partial T_3}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \phi & 0 & -r \sin \phi \end{pmatrix}$$

The determinant of a matrix describes the stretching of space as our coordinates change in r , θ , and ϕ to yield

$$\det([J]_{\vec{T}}) = r^2 \sin \phi$$

Thus we arrive at the volume form

$$d\Omega = r^2 \sin \phi dr d\theta d\phi.$$

Part V

Partial Differential Equations



Modeling and the continuum limit

In the prequel, we took an approach to modeling problems that evolved over time through Ordinary Differential Equations (ODEs). Though these equations were interesting in their own right, they make up but a tiny fraction of dynamical problems that scientists find interesting. Moreover, our equations typically consisted of a single dependent variable x which depended on the parameter t which we thought of as time. Given that we have sufficient understanding of higher dimensional spaces, it behooves us to revisit ODEs in a new light.

First, we take a cursory view of higher dimensional ODEs. The theory here will naturally lead us to the richer theory of Partial Differential Equations (PDEs). This path is actually rather direct as a handful of important PDEs can be explicitly derived from higher dimensional analogs of example ODEs we have already encountered. Once a satisfying bridge is built, we will forgo the need of explicit derivation for sake of understanding more elaborate equations that describe the fabric of nature.

14.1 Higher dimensional ODEs

Though ODE consist of a single variable, it's possible that many ODE interact with each other and form a system. Another case of interest would be finding trajectories as curves in higher dimensions. We may find ourselves thinking of tracking particles in space, but the components of a system need not be the same components of physical space. For example, the system could represent atmospheric dynamics along lines of latitude as in the Lorenz '96 model.

For now, let us take an example with a system of two first order ODEs. In general,

this assumes the form

$$\dot{x}_1(t) = f_1(x_1, x_2, t), \quad (14.1.1)$$

$$\dot{x}_2(t) = f_2(x_1, x_2, t). \quad (14.1.2)$$

This is an example of a first order equation with only two dependent variables, but in general we can allow for n dependent variables and there is no restriction on the order of equations.

Definition 14.1.1: System of ODEs

A ***system of ODEs*** is a collection of differential equations of various order. We call the system ***coupled*** if one of the differential equations is dependent on another.

In the above equations, the coupling can be seen in each equation quite explicitly. If we set $f_1(x_1, x_2, t) = x_1 + x_2$, then we can note that the t derivative of x_1 , \dot{x}_1 , explicitly has a dependence on the current value of both x_1 and x_2 . Hence, the values for x_1 are coupled to the other dependent variable of the system, x_2 .

Example 14.1.1: SIR Model in Ecology

A biological example of a system of first order equations comes from modeling the spread of disease in a colony. We let $S(t)$ denote the number of susceptible animals, $I(t)$ denote the number of infected animals, and $R(t)$ denote the animals resistant to the disease. The model is a system of ODE of the form

$$\begin{aligned}\dot{S} &= -\frac{\beta IS}{N}, \\ \dot{I} &= \frac{\beta IS}{N} - \gamma I, \\ \dot{R} &= \gamma I.\end{aligned}$$

We let $N = S(t) + I(t) + R(t)$ denote the constant population, and β and γ are other measured parameters.

This is a coupled system since, for example, the equation \dot{S} contains the function I . We also see that the equation for \dot{I} contains the function S . Lastly, the equation for \dot{R} contains the function I .

This notation can be condensed quite readily. We can simply condense the quantities of the system, x_1, x_2, \dots, x_n into a vector \vec{x} such that

$$\vec{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}. \quad (14.1.3)$$

We refer to this vector \vec{x} as the ***state*** of the system. Likewise, the t derivative of the state is the tangent vector $\dot{\vec{x}}$. The functions for the derivatives on the right hand side of

eqs. (14.1.1) and (14.1.2), f_1, f_2, \dots, f_n , can also be vectorized by

$$\vec{f}(\vec{x}, t) = \begin{pmatrix} f_1(\vec{x}, t) \\ f_2(\vec{x}, t) \\ \vdots \\ f_n(\vec{x}, t) \end{pmatrix}. \quad (14.1.4)$$

This then leads to a notationally compact version of a system of differential equations by

$$\dot{\vec{x}}(t) = \vec{f}(\vec{x}, t). \quad (14.1.5)$$

Finally, we should mention the caveat that these higher dimensional ODEs must be supplied with initial conditions in order to determine a unique curve that solves the problem at hand. We supply this as

$$\vec{x}(0) = \vec{x}_0, \quad (14.1.6)$$

where \vec{x}_0 is some chosen vector.

One may ask about higher order systems, but truthfully all systems of ordinary differential equations can be reduced to a system of this form. This shall suffice for now. It also allows us to supply an initial condition of the state with no reference to the t derivatives of the state. In higher order problems, we would also have to supply, for example, first order derivative information. Once again, we appeal to the fact that we can always describe any system as first order.

14.2 Linear systems and flows

One special case of higher dimensional ODEs are those that are linear. Fortunately, these systems are completely solvable and well understood. We will not discuss the explicit solutions (though the reader may find them interesting) and concentrate on our bridge to PDEs.

Definition 14.2.1: Linear System of ODE

A system of first order differential equations is *linear* if it can be expressed as a matrix equation

$$\dot{\vec{x}} = [A(t)]\vec{x},$$

otherwise the system is *nonlinear*.

Looking back at the system in example 14.1.1, we can note that it is nonlinear. The reason why is due to the infection terms where we see IS appear. This multiplication between the dependent variables I and S cannot be captured through matrix multiplication.

Example 14.2.1: Heat system

If we have three particles with temperatures u_1, u_2 , and u_3 that interact with one

another. We can model the system as

$$\dot{u}_1 = -k_1 u_1 + k_1 u_2 \quad (14.2.1)$$

$$\dot{u}_2 = -k_1 u_2 - k_2 u_2 + k_1 u_1 + k_2 u_3 \quad (14.2.2)$$

$$\dot{u}_3 = -k_2 u_3 + k_2 u_2. \quad (14.2.3)$$

Hence, we can put

$$\dot{\vec{u}} = [L]\vec{u}, \quad (14.2.4)$$

where we have the matrix $[L]$ given by

$$[L] = \begin{pmatrix} -k_1 & k_1 & 0 \\ k_1 & -(k_1 + k_2) & k_2 \\ 0 & k_2 & -k_2 \end{pmatrix}. \quad (14.2.5)$$

One can see that in this system, the temperature u_1 depends on the temperature u_2 , whereas the temperature u_2 depends on both u_1 and u_2 , and finally u_3 depends on u_2 . Pictorially, this would place the object 2 between the objects 1 and 3 and heat would flow between the objects in contact. The constants k_j describe the ability to transport heat between the different objects. In this case, k_1 is the transport medium between objects 1 and 2 and k_2 is the transport medium between objects 2 and 3. We revisit this in more detail later.

Vector fields play an intimate role in describing the evolutions of systems. Much of our interpretations of vector fields were given under the assumption that they describe a system in motion. For example, a vector field with divergence will cause straight line acceleration of a particle placed in the field. Curl, on the other hand, causes particles to rotate in a plane. Of course, vector fields can have both divergence and curl and the values for both can be spatially dependent. To complete this description, we provide the following definition.

Definition 14.2.2: Flow

Let \vec{x} be the state of a system in \mathbb{R}^n and let \vec{V} be an n -dimensional vector field. Then the **flow of the vector field** \vec{V} is

$$\dot{\vec{x}}(t) = \vec{V}(\vec{x}(t)). \quad (14.2.6)$$

The interpretation here is that the tangent vector of the curve is equal to the vector field at all points in time t so that the curve follows the vector field. In principle, the vector field \vec{V} could also depend on the variable t explicitly for which we would arrive at an equation of the form eq. (14.1.5). This would happen when, for example, if the wind in a region changes over time, we would have both spatial dependence from \vec{x} and time dependence from t so that we must put $\vec{f}(\vec{x}, t)$.

Example 14.2.2: Helical flow

Consider the vector field $\vec{V} = \begin{pmatrix} -y \\ x \\ \frac{1}{4}z \end{pmatrix}$. Note that

$$\vec{\nabla} \times \vec{V} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \quad (14.2.7)$$

$$\vec{\nabla} \cdot \vec{V} = \frac{1}{10}. \quad (14.2.8)$$

Hence, we should expect rotation in a plane except when $x = y = 0$ along with acceleration in a straight line at any point in space aside from where $z = 0$. To this end, we consider the initial value problem for the flow of \vec{V} by

$$\dot{\vec{x}} = \vec{V}(\vec{x}) \quad \vec{x}(0) = \begin{pmatrix} 1 \\ 0 \\ \frac{1}{2} \end{pmatrix}. \quad (14.2.9)$$

The above equation is also linear since it can be written in terms of a matrix. This makes it possible to find a close form analytical solution. We note the solution to the initial value problem is

$$\vec{x}(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \\ \frac{1}{2}e^{\frac{t}{4}} \end{pmatrix}. \quad (14.2.10)$$

In general, it may only be possible to approximate a solution computationally, but this is a rather quick task as well. Let us view this solution along with the vector field.

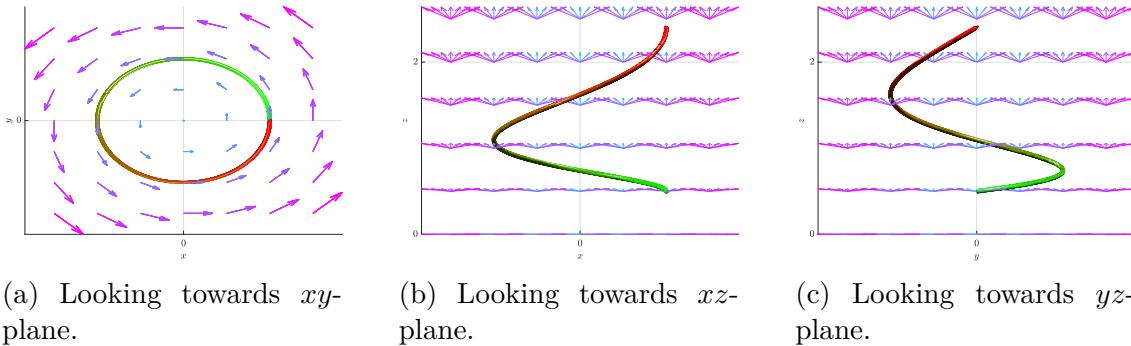


Figure 14.1: The curve \vec{x} and the vector field \vec{V} from different angles.

Exercise 14.2.1. Show that example 14.2.2 is linear by constructing a matrix $[A]$ such that $\dot{\vec{x}} = [A]\vec{x}$.

One other interesting and widely applicable type of systems are the gradient systems. These systems are used to describe physical processes such as fluid flow or heat flow. Not only that, but for problems where one is seeking to find a minimal or optimal solution, one can use a gradient system as a means of approaching the optimum. In particular, this

method is used in machine learning in the gradient descent algorithm.

Definition 14.2.3: Gradient flow

Let \vec{x} be the state of a system in \mathbb{R}^n and let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a scalar field. Then the flow

$$\dot{\vec{x}}(t) = \vec{\nabla} f(\vec{x}(t)) \quad (14.2.11)$$

is called *gradient flow*.

The interpretation of a gradient flow is that the state tends towards a local maximum of the function f . If we took the flow defined by the negative of the gradient

$$\dot{\vec{x}}(t) = -\vec{\nabla} f(\vec{x}(t)), \quad (14.2.12)$$

then the state would tend towards the local minimum value of f . This is sometimes referred to as *gradient descent*. Examples of physical scenarios involving gradient descent include fluid dynamics and heat transfer.

14.3 Partial Differential Equations

We now want to investigate a larger class of differential equations. These are the *partial differential equations* (PDEs). These equations become yet more complicated to solve, but are very prevalent in the study of the physical world. Fundamentally, these are time-varying differential equations of vector and scalar fields of many variables. The goal for us is to be able to recognize a few specific example equations and understand their behavior. We will also be able to solve a few equations with our tools from studying ODEs. However, it is easy to pose a PDE that is virtually impossible to solve.

Definition 14.3.1: Scalar Partial Differential Equation

A *partial differential equation of a scalar field* of three spatial variables x, y, z and a time variable t is an expression of a scalar function $u(x, y, z, t)$, the partial derivatives of $u(x, y, z, t)$, and other functions.

Definition 14.3.2: Vector Partial Differential Equation

A *partial differential equation of a vector field*

$$\vec{V}(x, y, z, t) = \begin{bmatrix} V_1(x, y, z, t) \\ V_2(x, y, z, t) \\ V_3(x, y, z, t) \end{bmatrix}$$

is an equation containing \vec{V} , the (component) derivatives of \vec{V} , and other vector fields.

It is worthwhile to take a look at a few important example equations before we move onto derivations or methods of solutions. The realms of PDEs are so widespread, it is rather difficult to give a comprehensive list. It is even difficult to list the applications of a single equation. To begin, we will consider three intimately related scalar equations. Each of which, we will derive in some way. Finally, we list the vector equations for the laws of electrodynamics and the time dependent Schrödinger equation.

Example 14.3.1: Heat equation

The scalar *heat equation* is given by

$$\frac{\partial u}{\partial t}(\vec{x}, t) - \vec{\nabla} \cdot (k(\vec{x}) \vec{\nabla} u(\vec{x}, t)) = f(\vec{x}, t), \quad (14.3.1)$$

where \vec{x} represents a position in space (typically \mathbb{R}^3). This equation models the diffusion of heat in a region of space, hence the name. We think of $u(\vec{x}, t)$ being the temperature at the point \vec{x} at the time t .

Example 14.3.2: Laplace (Poisson) equation

The scalar *Laplace* (sometimes *Poisson*) *equation* is given by

$$-\Delta u(\vec{x}) = f(\vec{x}). \quad (14.3.2)$$

Notice, there is no dependence on time! This equation can be realized as the long term behavior of the heat equation. If $u(\vec{x})$ describes temperature, then the solution to this equation tells you the equilibrium temperature. Since this is an equilibrium solution, the time component is gone.

Example 14.3.3: Wave equation

The scalar *wave equation* is given by

$$\frac{\partial^2 u}{\partial t^2}(\vec{x}, t) - c^2 \Delta u(\vec{x}, t) = f(\vec{x}, t). \quad (14.3.3)$$

The solutions here are wavelike. Think of plucking a guitar string, or the ripples on the surface of a lake after a rock has been tossed in, or the vibrating cymbal or drum head.

Example 14.3.4: Maxwell's equations

Maxwell's equations describe the electric \vec{E} and magnetic \vec{B} fields that permeate space due to charged particles. These equations turn out to be coupled PDE. They

read

$$\vec{\nabla} \cdot \vec{E}(\vec{x}, t) = \frac{\rho(\vec{x}, t)}{\epsilon}, \quad (14.3.4)$$

$$\vec{\nabla} \cdot \vec{B}(\vec{x}, t) = 0, \quad (14.3.5)$$

$$\vec{\nabla} \times \vec{E}(\vec{x}, t) = -\frac{\partial \vec{B}}{\partial t}(\vec{x}, t), \quad (14.3.6)$$

$$\vec{\nabla} \times \vec{B}(\vec{x}, t) = \mu \vec{J} + \mu \epsilon \frac{\partial \vec{E}}{\partial t}(\vec{x}, t). \quad (14.3.7)$$

The equations above are best realized when studying spacetime as a single entity. We will not cover this in this text, but one can find that electrodynamics is a rich subject deeply imbued into the fabric of the cosmos.

Example 14.3.5: Time dependent Schrödinger equation

The time dependent Schrödinger equation models the state of a quantum system and its evolution in time. Succinctly, one sees this equation written in terms of operators by

$$H\Psi(\vec{x}, t) = E\Psi(\vec{x}, t), \quad (14.3.8)$$

where H is the Hamiltonian operator given by

$$-\frac{\hbar^2}{2m}\Delta + V(\vec{x}) \quad (14.3.9)$$

and E is the energy operator given by

$$i\hbar \frac{\partial}{\partial t}. \quad (14.3.10)$$

Hence, we have

$$\left(-\frac{\hbar^2}{2m}\Delta + V(\vec{x})\right)\Psi(\vec{x}, t) = i\hbar \frac{\partial}{\partial t}\Psi(\vec{x}, t) \quad (14.3.11)$$

We will return to this equation in more depth later.

14.4 The heat equation

Imagine at this moment, that we are to generalize example 14.2.1 to a system with n particles. If we then take the limit as $n \rightarrow \infty$, we should expect nothing less than to arrive at the heat equation we see in example 14.3.1 with the caveat that we would be taking a 1-dimensional approach. Keep this in mind as we proceed – you know where we are starting and where we are heading. Let us now explicitly consider this scenario. Let us imagine a rod lined with n equally spaced particles with a heat conductive medium placed between the interior particles. This gives us the following picture.



We label the position of the j^{th} particle as x_j and the thermally conductive media between particle j and $j + 1$ will be given a value of k_j . The value of k_j describes how easily thermal energy is transported between particle j and $j + 1$. Larger k_j means more heat can be transferred at any given moment. k_j can be referred to as the **diffusivity** of the thermal transport medium. We then let $u(x, t)$ be the temperature at the point x and time t along the rod. For now, we will define u only on the particles and note that we will put $u(x_j, t) = u_j(t)$. We can place additional sources of heat on each particle by taking a function $f(x, t)$ and defining $f_j(t) := f(x_j, t)$. The function $f(x, t)$ describes the rate of thermal energy added at the point x at the time t .

Exercise 14.4.1. Run a thought experiment in your head. Pick a small value of n , choose some temperatures u_j for all of particles, and ignore the heat sources f_j for the time being. What do you think will happen as time moves forward?

In this rod, we can refer to the endpoints x_1 and x_n as the boundaries and the other $n - 2$ points as the interior. Let us deduce the equations for the boundary and interior. First, particle 1 radiates heat through the medium k_1 while particle 2 does the same. This process is random, but on average, we find that the rate of change of temperature of particle 1 depends on how much is lost and gained in the interaction with particle 2. The law of mass action then states

$$\dot{u}_1 = -k_1 u_1 + k_1 u_2 + f_1 = k_1(u_2 - u_1) + f_1. \quad (14.4.1)$$

The second equality shows that this is equivalent Newton's law of cooling with an additional source of heat from the rate due to f ! The equation for particle n follows a completely analogous argument and we have

$$\dot{u}_n = -k_{n-1} u_n + k_{n-1} u_{n-1} + f_n. \quad (14.4.2)$$

The interior points are only more complicated in the sense that each has two neighbors. The interior particle j will interact with both particles $j - 1$ and $j + 1$ by

$$\dot{u}_j = -k_{j-1} u_j - k_j u_j + k_{j-1} u_{j-1} + k_j u_{j+1} + f_j = k_j(u_{j+1} - u_j) - k_{j-1}(u_j - u_{j-1}) + f_j. \quad (14.4.3)$$

Since this process of heat is random, we can imagine that quanta of heat are passed between the particles and so, if the rod has a conductivity κ , then $k = \frac{\kappa}{\delta x^2}$ will describe how much heat flow between any two particles. Here, δx is the distance between any two particles (i.e., $\delta x = x_j - x_{j-1}$). This is due to the fact that, on average, a quanta of heat will travel a distance proportional to δx^2 . Hence, we shall replace the instances of k_j with $\frac{\kappa_j}{\delta x^2}$ where and we have

$$\dot{u}_j = \frac{\kappa_j(u_{j+1} - u_j) - \kappa_{j-1}(u_j - u_{j-1})}{\delta x^2} + f_j. \quad (14.4.4)$$

If we now take the limit that $n \rightarrow \infty$, then $\delta x \rightarrow 0$ and we have for interior points $x \in (x_1, x_n)$

$$\frac{\partial}{\partial t} u(x, t) = \frac{\partial}{\partial x} \left(\kappa(x) \frac{\partial}{\partial x} u(x, t) \right) + f(x, t). \quad (14.4.5)$$

One should now compare this equation to that in example 14.3.1 in the case where the spatial dimension is equal to one.

Stopped updating here. Mention gradient flows and linear systems. Write the above as a linear system?

One of the most illuminating examples of PDEs is the heat equation. Let us work through a specific example of the heat equation and keep in mind the physical intuition throughout.

In order to move forward, we need to also properly specify the problem we want to solve.

The one-dimensional source-free heat equation is a great starting point to begin our process. We are given the following data:

- A region Ω in space that we are concerned with. For example, in one dimension, we can consider the interval $\Omega = (0, 1)$.

- A PDE

$$\frac{\partial u}{\partial t}(x, t) - k \frac{\partial^2 u}{\partial x^2} = 0.$$

- Boundary conditions. These can come in a few forms, but we will concentrate on just one. We must specify $u(0, t) = a$ and $u(1, t) = b$. These boundary conditions correspond to fixing the temperature at the ends of a rod constant.

- Initial conditions. We specify the initial temperature distribution

$$u(x, 0) = u_0(x).$$

Example 14.4.1: Solving the Heat Equation

Let us consider the simplified one-dimensional source free (i.e., the right hand side is zero) heat equation given by the following:

$$\frac{\partial u}{\partial t}(x, t) - \frac{\partial^2 u}{\partial x^2} = 0.$$

We can require boundary conditions and initial conditions later on.

Let us assume that the solution function $u(x, t)$ can be written as

$$u(x, t) = f(x)g(t).$$

We call this approach the *separation of variables*. We then plug in this assumption to our PDE.

$$\begin{aligned} \frac{\partial}{\partial t}(f(x)g(t)) - \frac{\partial^2}{\partial x^2}(f(x)g(t)) &= 0 \\ f(x)\frac{\partial g}{\partial t} - g(t)\frac{\partial^2 f}{\partial x^2} &= 0 \\ fg' - f''g &= 0. \end{aligned}$$

We can then do a bit more algebra.

$$\begin{aligned} fg' - f''g &= 0 \\ fg' &= f''g \\ \frac{g'(t)}{g(t)} &= \frac{f''(x)}{f(x)}. \end{aligned}$$

Now, notice that both sides depend on different variables. We have successfully separated this equation into an equation for each variable. This is to say, since each side of the equation depends on a different variable, each side must be equal to a constant λ ! So we have two equations.

$$\begin{aligned} \frac{g'(t)}{g(t)} &= \lambda \\ \frac{f''(x)}{f(x)} &= \lambda. \end{aligned}$$

We can then solve both of these as ODE. Note, it will be helpful to instead choose $-\lambda$ as the constant.

Exercise 14.4.2. What are the general solutions to the above ODE?

Exercise 14.4.3. Given those general solutions, what is the general solution to the heat equation?

Answer 14.4.1. We get

$$u(x, t) = f(x)g(t) = Ae^{-\lambda y} \sin(\sqrt{\lambda}x) + Be^{-\lambda t} \cos(\sqrt{\lambda}x).$$

Previously we found the general solution to the heat equation

$$\frac{\partial u}{\partial t}(x, t) - \frac{\partial^2 u}{\partial x^2}(x, t) = 0$$

is

$$u(x, t) = Ae^{-\lambda t} \sin(\sqrt{\lambda}x) + Be^{-\lambda t} \cos(\sqrt{\lambda}x).$$

However, this solution is very general. We have the undetermined constants λ , A , and B . We need more information to get a particular solution.

Example 14.4.2: Particular Solution to the 1D Heat Equation

We will stick with the one-dimensional case but we must pick the following.

- Domain: Let $\Omega = (0, 1)$.
- Initial Conditions: $u(x, 0) = \sin(\pi x)$.
- Boundary Conditions: $u(1, t) = u(0, t) = 0$.

This list of requirements gives us enough information to solve the equation explicitly for a particular solution.

First, let us take the boundary conditions. We impose these on our general solution:

$$u(x, t) = Ae^{-\lambda t} \sin(\sqrt{\lambda}x) + Be^{-\lambda t} \cos(\sqrt{\lambda}x).$$

Thus we require

$$0 = u(0, t) = Ae^{-\lambda t} \sin(0) + Be^{-\lambda t} \cos(0)$$

which gives us that

$$B = 0.$$

The other boundary condition is

$$0 = u(1, t) = Ae^{-\lambda t} \sin(\sqrt{\lambda}).$$

Specifically, this means that $A = 0$, which gives us a trivial solution or that we have

$$\sqrt{\lambda} = n\pi$$

for any integer n . This is because $\sin(n\pi) = 0$ when n is an integer. Thus our solution now reads

$$u(x, t) = Ae^{-n^2\pi^2 t} \sin(n\pi x).$$

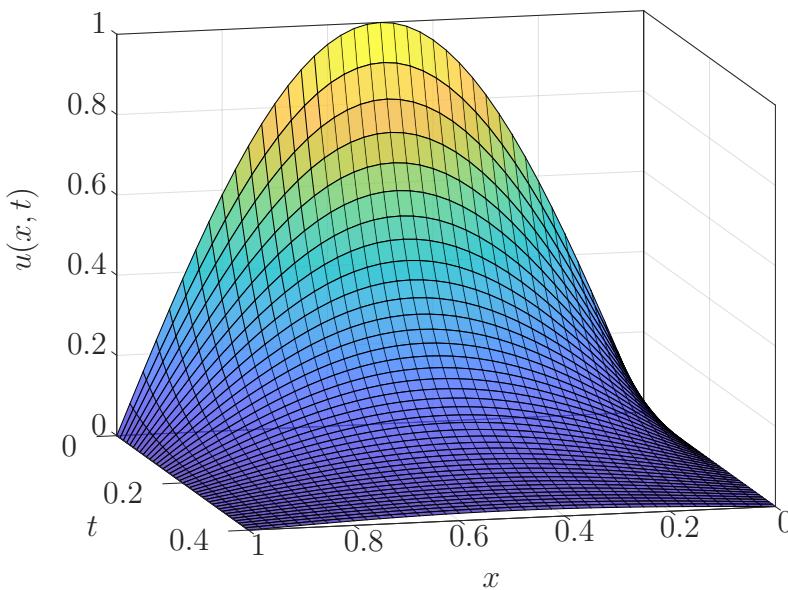
Lastly, we match our initial conditions. So we have

$$\sin(\pi x) = u(x, 0) = Ae^0 \sin(n\pi x)$$

and so we find that $n = 1$. Thus, our solution is

$$u(x, t) = e^{-\pi^2 t} \sin(\pi x).$$

We can plot this solution as follows.



Exercise 14.4.4. Can you interpret what this is physically describing as t gets larger?

14.5 The Laplace Equation

In the long time limit ($t \rightarrow \infty$) or steady-state of the (source free) heat equation, one arrives at the so called Laplace equation

$$-\Delta u = 0.$$

In the one-dimensional case, this equation reads

$$-\frac{d^2u}{dx^2}(x) = 0.$$

Note there is no dependence on time as this is the steady-state behavior for the heat equation.

Exercise 14.5.1. This is an ODE in the variable x . You can solve this and find a general solution by integration.

Answer 14.5.1. The general solution to the one-dimensional Laplace equation is the equation for a line

$$u(x) = Ax + B.$$

One can show that this is indeed a solution by taking two derivatives of $u(x)$ and finding that you get zero.

Example 14.5.1: Particular Solution to the 1D Laplace Equation

Through integration, we find

$$u(x) = Ax + B$$

where A and B are undetermined constants. In order to specify these constants, we must provide the following.

- Domain: Let $\Omega = (0, 1)$.
- Boundary Conditions: $u(0) = 0$ and $u(1) = 0$.

Note that we do not need initial conditions since there is no time dependence in this PDE.

Now, to find the particular solution, we apply the boundary conditions to our general solution. So we have

$$0 = u(0) = A(0) + B,$$

so $B = 0$. Then the other condition

$$0 = u(1) = A,$$

so $A = 0$. Thus, our solution is

$$\boxed{u(x) = 0}.$$

Now, compare this to the solution to the heat equation previously

$$u(x, t) = e^{-\pi^2 t} \sin(\pi x).$$

We claimed the Laplace equation is the long-time solution of the heat equation and indeed if we look at $t \rightarrow \infty$, we have

$$\lim_{t \rightarrow \infty} u(x, t) = 0.$$

14.6 The Wave Equation

The wave equation is studied when one wants to find the oscillatory behavior of some medium. For example, one can pluck a guitar string or hit a drum head. These actions induce vibrations in the medium (the string or head) and it is the vibrations that one hears. The equation that models these phenomenon is the wave equation

$$\frac{\partial^2 u}{\partial t^2}(x, y, z, t) - c^2 \nabla \cdot \nabla u(x, y, z, t) = f(x, y, z, t).$$

Example 14.6.1: Solving the 1D Wave Equation

In one-dimension, the simplified source free wave equation reads

$$\frac{\partial^2 u}{\partial t^2}(x, t) - \frac{\partial^2 u}{\partial x^2}(x, t) = 0.$$

It turns out we can solve the 1D wave equation in the same way we did the heat equation. So, we assume a separation of variables approach in that

$$u(x, t) = f(x)g(t).$$

We plug this into the PDE to find

$$\begin{aligned} \frac{\partial^2}{\partial t^2}(f(x)g(t)) - \frac{\partial^2}{\partial x^2}(f(x)g(t)) &= 0 \\ f(x)\frac{\partial^2 g}{\partial t^2} - g(t)\frac{\partial^2 f}{\partial x^2} &= 0 \\ f(x)g''(t) - f''(x)g(t) &= 0. \end{aligned}$$

We then wish to make the left hand side and right hand side functions of different input variables

$$\begin{aligned} f(x)g''(t) - f''(x)g(t) &= 0 \\ f(x)g''(t) &= f''(x)g(t) \\ \frac{g''(t)}{g(t)} &= \frac{f''(x)}{f(x)}. \end{aligned}$$

Since each side depends on a different input variable, each side must be equal to a constant. So this gives us

$$\frac{g''(t)}{g(t)} = \frac{f''(x)}{f(x)} = -\lambda^2,$$

where $-\lambda^2$ is an undetermined constant but was chosen to make the next steps easier. We then get two ODEs

$$\begin{aligned} f''(x) &= -\lambda^2 f(x), \\ g''(t) &= -\lambda^2 g(t), \end{aligned}$$

which are both harmonic oscillator equations. Thus, since we know the solutions to the harmonic oscillator equation, we have

$$\begin{aligned} f(x) &= C_1 \sin(\lambda x) + C_2 \cos(\lambda x), \\ g(t) &= C_3 \sin(\lambda t) + C_4 \cos(\lambda t). \end{aligned}$$

It follows that our solution is thus

$$u(x, t) = f(x)g(t) = (C_1 \sin(\lambda x) + C_2 \cos(\lambda x))(C_3 \sin(\lambda t) + C_4 \cos(\lambda t)).$$

With a general solution to the wave equation written down. We can work to solve a particular case of the wave equation. Let's see this.

Example 14.6.2: Particular Solution to the 1D Wave Equation

We found that the general solution to the 1D wave equation is

$$u(x, t) = (C_1 \sin(\lambda x) + C_2 \cos(\lambda x))(C_3 \sin(\lambda t) + C_4 \cos(\lambda t)).$$

Let us multiply this out and re-collect the constants to get

$$u(x, t) = C_1 \sin(\lambda x) \sin(\lambda t) + C_2 \sin(\lambda x) \cos(\lambda t) + C_3 \cos(\lambda x) \sin(\lambda t) + C_4 \cos(\lambda x) \cos(\lambda t).$$

In order to specify these constants, we provide the following:

- Domain: Let $\Omega = (0, 1)$.
- Initial Conditions: We let $u(x, 0) = \sin(\pi x)$ and $\frac{\partial u}{\partial t}(x, 0) = 0$.
- Boundary Conditions: Take $u(0) = u(1) = 0$.

Note the need for both initial position $u(x, 0)$ and initial velocity $\frac{\partial u}{\partial t}(x, 0)$.

Now, we find the particular solution by first applying our boundary conditions. Specifically, we have

$$0 = u(0, t) = C_1 \sin(0) \sin(\lambda t) + C_2 \sin(0) \cos(\lambda t) + C_3 \cos(0) \sin(\lambda t) + C_4 \cos(0) \cos(\lambda t)$$

which reduces to

$$0 = C_3 \sin(\lambda t) + C_4 \cos(\lambda t).$$

The only way this can be equal to zero for all t is if $C_3 = C_4 = 0$. Thus, we now have

$$u(x, t) = C_1 \sin(\lambda x) \sin(\lambda t) + C_2 \sin(\lambda x) \cos(\lambda t).$$

Applying the next boundary condition

$$0 = u(1, t) = C_1 \sin(\lambda) \sin(\lambda t) + C_2 \sin(\lambda) \cos(\lambda t)$$

gives us that $\lambda = n\pi$ for any integer n since in this case $\sin(n\pi) = 0$. And so our solution is now

$$u(x, t) = C_1 \sin(n\pi x) \sin(n\pi t) + C_2 \sin(n\pi x) \cos(n\pi t).$$

We then apply the initial conditions. Specifically, we required that

$$\sin(\pi x) = u(x, 0) = C_1 \sin(n\pi x) \sin(0) + C_2 \sin(n\pi x) \cos(0)$$

which reduces to

$$\sin(\pi x) = C_2 \sin(n\pi x).$$

Thus we have that $n = 1$ and $C_2 = 1$. Our solution is now

$$u(x, t) = C_1 \sin(\pi x) \sin(\pi t) + \sin(\pi x) \cos(\pi t).$$

Now, we also required that

$$0 = \frac{\partial u}{\partial t}(x, 0) = C_1 \pi \sin(\pi x) \cos(0) - \pi \sin(\pi x) \sin(0)$$

which reduces to

$$0 = C_1 \pi \sin(\pi x)$$

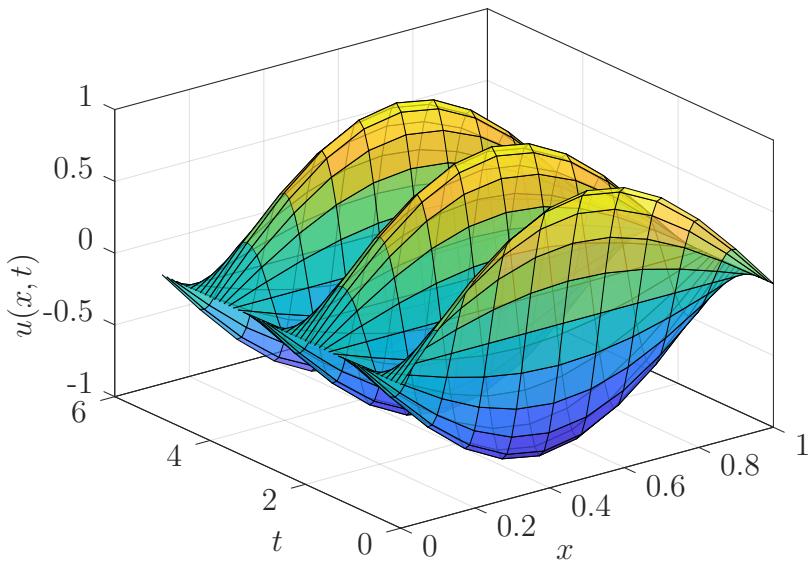
which means that

$$C_1 = 0.$$

Thus, we now have the particular solution

$$u(x, t) = \sin(\pi x) \cos(\pi t).$$

We can plot a graph of this solution with z representing the height of the function, the x -axis giving our position in the domain Ω and the t -axis moving perpendicularly to x and z . We get



15

Separation of variables

Part VI

Further Topics in Linear Algebra

16

Complex functions and transformations

16.1 Introduction

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

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

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

16.2 Complex Functions

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

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

$$z = x + iy,$$

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

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

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

16.2.1 Cartesian and Polar Representations

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

$$f(z) = w,$$

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

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

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

Of course, we can also write

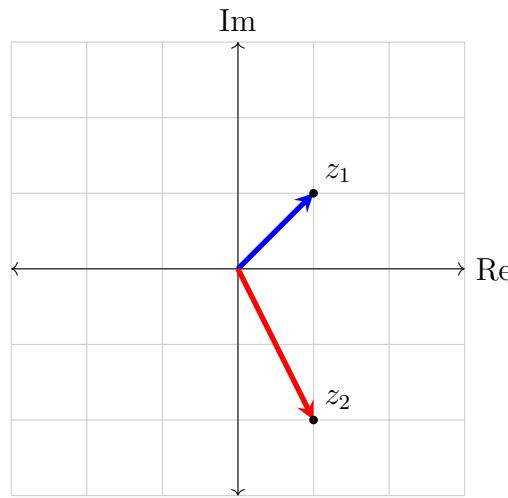
$$f(z) = r(z)e^{i\theta(z)},$$

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

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

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

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



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

Example 16.2.1: Plotting a Complex Function

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

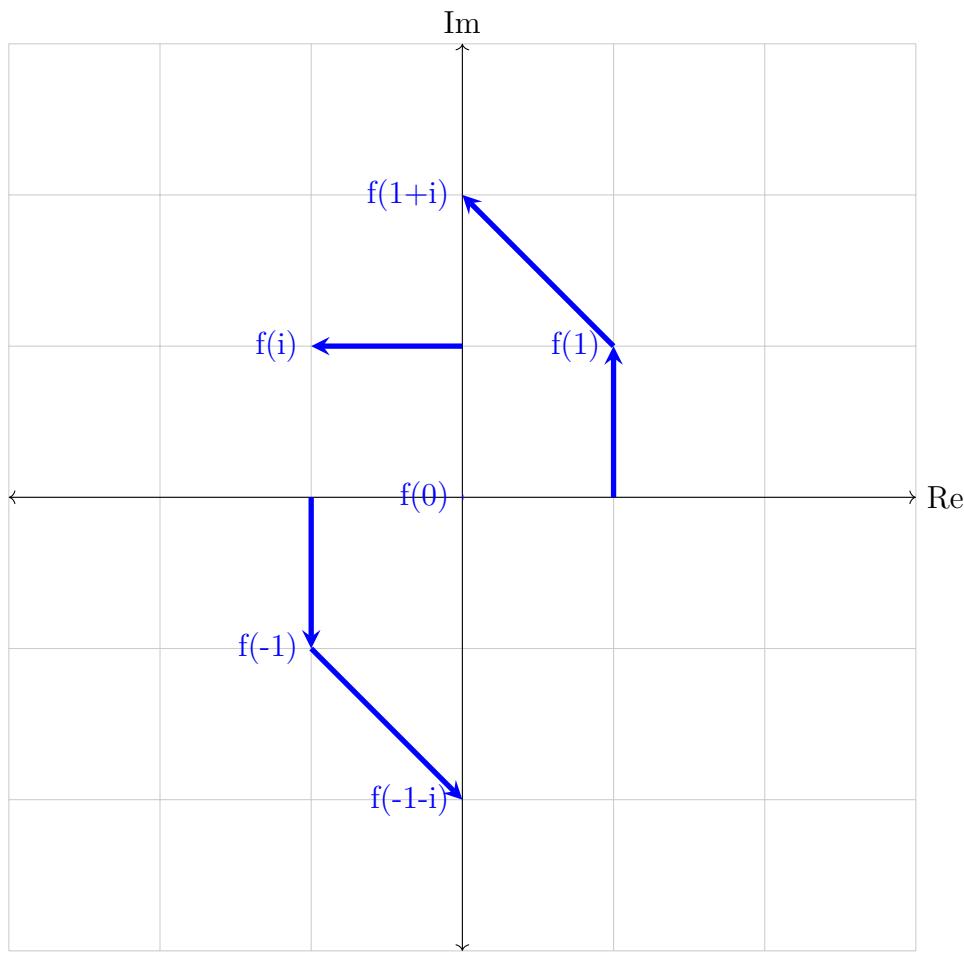
$$f(z) = iz = -\text{Im}(z) + i\text{Re}(z).$$

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

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

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

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



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

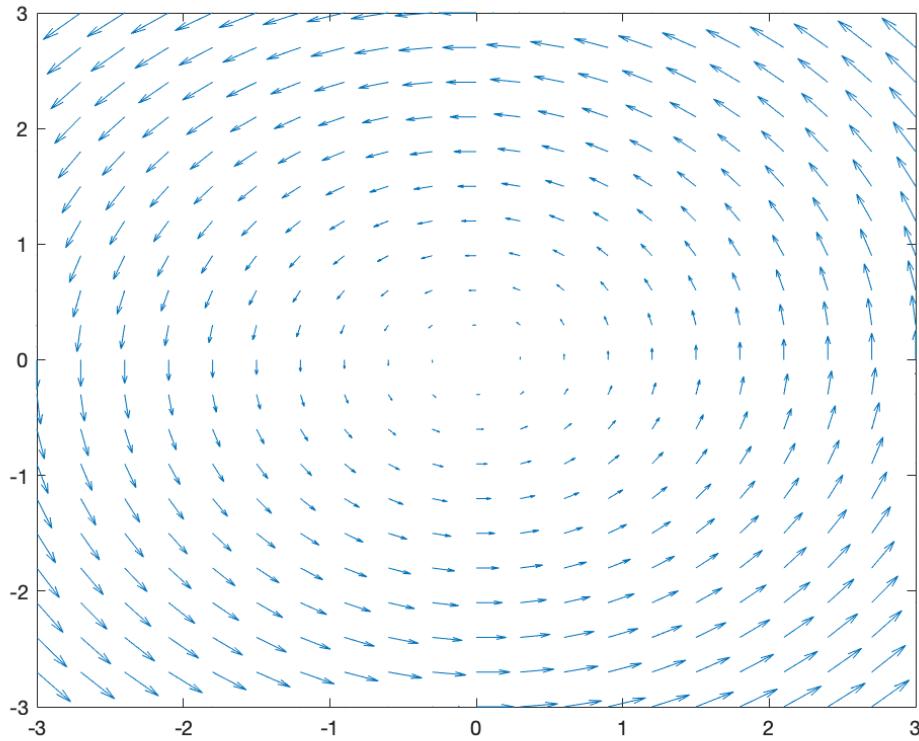


Figure 16.1: A plot of various different outputs for their corresponding inputs.

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

16.2.2 Complex Valued Functions

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

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

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

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

or the polar decomposition

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

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

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

and in the polar case we have

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

which follows from the chain rule.

Exercise 16.2.1. Verify the polar derivative above is correct.

Integration follows the fundamental theorem of calculus and hence we have

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

So, for example, in the cartesian representation we have

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

Remark 16.2.1. Complex functions (i.e., functions with complex valued inputs) have different behavior with integration and differentiation which we will not discuss at all. The closest we will get to this structure is calculus in \mathbb{R}^2 .

Example 16.2.2: A Line in \mathbb{C}

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

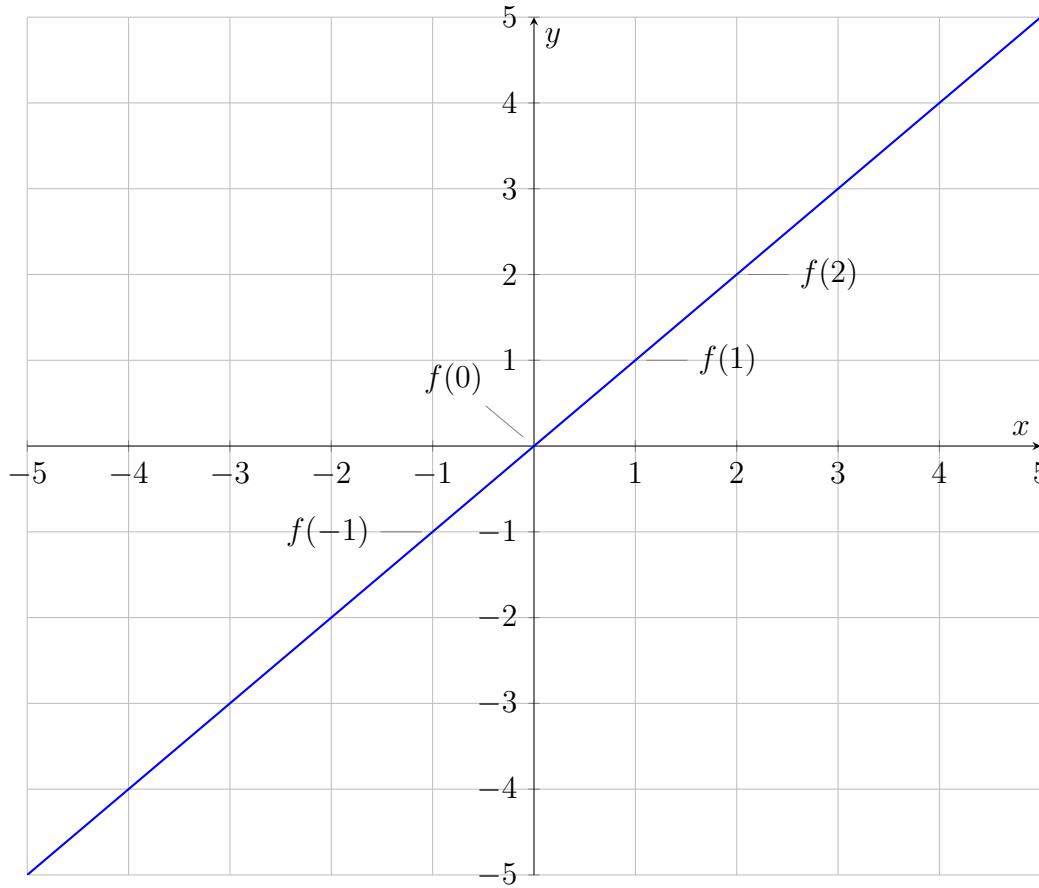
$$f(x) = x + ix.$$

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

$$\begin{array}{ll} f(0) = 0 & f(1) = 1 + i \\ f(-1) = -1 - i & f(2) = 2 + 2i. \end{array}$$

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

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



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

Example 16.2.3: Wavefunctions in the Box

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

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

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

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

Though these states are real valued, there is no physics that requires this. Similarly, the coefficients a_n are also not constrained to be real valued constants either. In the broadest generality, Ψ can be a complex valued function and the coefficients a_n can be complex as well.

Fundamentally, this is due to the physical understanding of the solutions to Schrödinger's equation. When we are looking for physically meaningful interpre-

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

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

where we have the pointwise modulus of the complex valued function

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

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

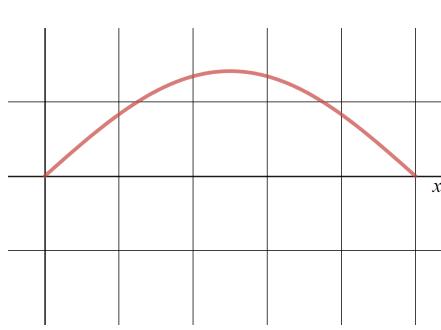
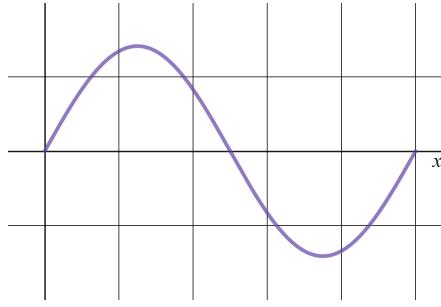
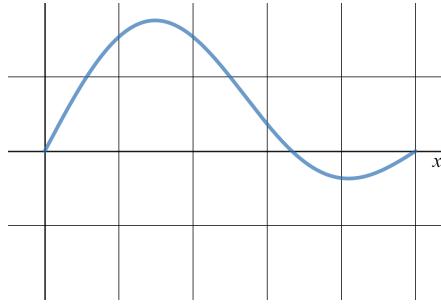
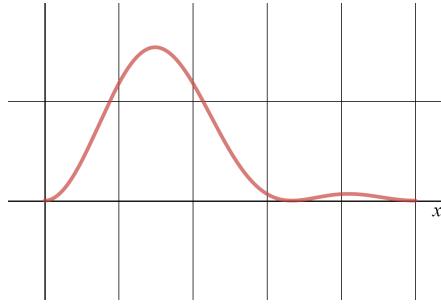
$$\|\Psi(x)\|^2 = u^2(x) + v^2(x).$$

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

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

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

We can plot the functions used above to see why this is the case.

(a) Normalized state $\psi_1(x)$.(b) Normalized state $\psi_2(x)$.(c) The wavefunction $\Psi(x)$.(d) The probability function $\|\Psi(x)\|^2$.

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

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

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

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

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

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

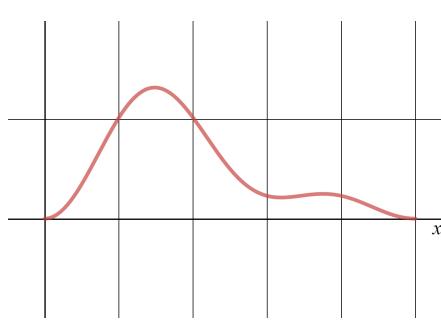
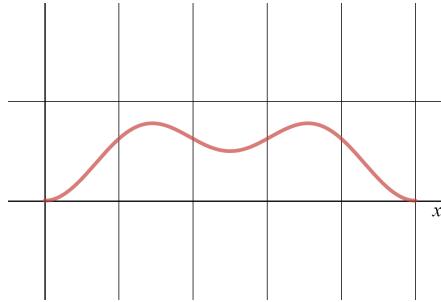
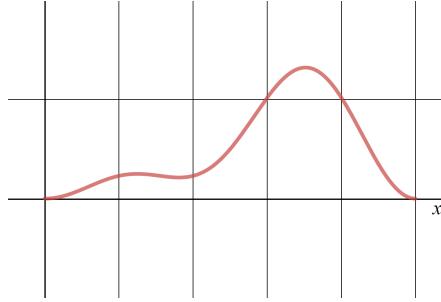
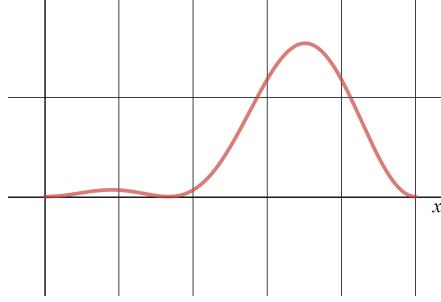
This is now slightly different! However, we can note that

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

and hence we have

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

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

(a) Phase difference $\delta = 0$.(b) Phase difference $\delta = \frac{\pi}{3}$.(c) Phase difference $\delta = \frac{2\pi}{3}$.(d) Phase difference $\delta = \pi$.

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

There are two important remarks to note here.

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

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

Remark 16.2.2. In a sense, the integral defined above $P_{[a,b]}(\Psi)$ is a real valued function with a function as an input. Though we have not noted this until now, it becomes important in the future.

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



Inner product spaces

17.1 Introduction

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

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

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

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

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

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

Underlying much of the theory of space was this structure.

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

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

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

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

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

and computing

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

Likewise, the length of a complex vector is given by

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

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

17.1.1 Infinite Dimensions

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

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

17.2 Function Spaces

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

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

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

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

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

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

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

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

17.3 Inner Products and Norms

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

Exercise 17.3.1. Find the definition in the previous text and review it.

Definition 17.3.1: Inner Product

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

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

that satisfies

- (Nondegenerate) For a $\vec{a} \in V$ we have that $\langle \vec{a}, \vec{a} \rangle = 0$ if and only if $\vec{a} = \vec{0}$;
- (Positive definite) For any nonzero $\vec{a} \in V$ we have that $\langle \vec{a}, \vec{a} \rangle > 0$;
- (Symmetric) For any $\vec{a}, \vec{b} \in V$ we have that $\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle$. If the vector space is complex, then we have conjugate symmetry $\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle^*$.

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

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

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

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

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

The first position is simply linear.

Exercise 17.3.2. Verify that the dot product for \mathbb{R}^n and the Hermitian inner product for \mathbb{C}^n are indeed inner products.

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

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

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

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

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

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

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

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

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

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

Exercise 17.3.3. Show that V is a vector space.

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

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

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

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

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

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

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

and if we have that

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

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

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

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

Hence, the inner product is positive definite.

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

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

and

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

which means that we have

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

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

17.3.1 Norms

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

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

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

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

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

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

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

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

Remark 17.3.1. From here on out we shall use the notation $\langle \cdot, \cdot \rangle$ to refer to the inner product and $\|\cdot\|$ to refer to the induced norm from that inner product. The choice of inner product should be noted or clear from the choice of vector space.

17.4 Inner Product Spaces

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

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

Exercise 17.4.1. We defined a Cauchy sequence in the prequel. Find the definition.

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

Example 17.4.1: A Cauchy Sequence of Functions

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

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

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

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

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

We noted that as we increased N , the function $A_N(x)$ became closer and closer to the function $f(x)$. This was entirely reasonable as if the contrary were true, at some point a large N would provide us a worse approximation to $f(x)$.

The completeness assumption for a Hilbert space will give us this ability. It will allow one to properly approximate quantities such as infinite sums of functions in a way that makes intuitive sense.

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

17.5 Symmetries

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

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

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

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

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

Exercise 17.5.1. Recall the definition of the matrix group $O(2)$.

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

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

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

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

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

we have

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

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

$$\begin{aligned} \langle [\text{Ref}]_\theta \vec{u}, [\text{Ref}]_\theta \vec{v} \rangle &= (\cos(2\theta)u_1 + \sin(2\theta)u_2)(\cos(2\theta)v_1 + \sin(2\theta)v_2) \\ &\quad + (\sin(2\theta)u_1 - \cos(2\theta)u_2)(\sin(2\theta)v_1 - \cos(2\theta)v_2) \\ &\quad \vdots \\ &= u_1v_1 + u_2v_2. \end{aligned}$$

Exercise 17.5.2. Show the remaining steps in the above computation.

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

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

17.6 Bases

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

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

so that any vector

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

could be written as

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

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

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

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

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

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

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

17.6.1 Orthonormal Bases

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

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

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

Exercise 17.6.1. Show the vectors \vec{u} , \vec{v} , and \vec{w} above form a basis. Then, to see that there is some overlap, you can find the components of a vector (of your choice) in the directions \hat{u} , \hat{v} , and \hat{w} . Hint: to see the overlap, determine the length of the vector you chose, and compare it to the length computed from components in the \hat{u} , \hat{v} , and \hat{w} .

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

Remark 17.6.1. Orthogonal bases in finite dimensions also appear when we find eigenvalues and eigenvectors for Hermitian matrices. Specifically, the set of eigenvectors corresponding to different eigenvalues for a Hermitian matrix are always orthogonal!

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

Recall the equation we wished to solve was

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence, any solution to that problem can be written as

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

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

17.6.2 Series and Integrals as Linear Combinations

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

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

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

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

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

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

When \mathcal{H} is infinite dimensional, we can no longer easily manipulate vectors as column vectors (it just doesn't always make sense)! So, we our hand is forced and we must write our vectors as some form of an infinite sum. Of course, this is just an infinite version of a linear combination! For now, let us assume H is sufficiently nice¹ so that we can write any vector $\Psi \in \mathcal{H}$ by

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

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

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

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

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

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

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

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

Remark 17.6.2. There is much going on here. Much if it is coming down to the sizes of infinities and underlying spaces we are working with. The point is to see that we are solely generalizing the notion of writing a vector in a basis. Some spaces just have a very large amount of basis elements!

17.6.3 Projection

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

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

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

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

This is a surprisingly helpful fact.

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

Example 17.6.1: Triangle Wave

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

You should verify the integral above.

What this means is that we can now put

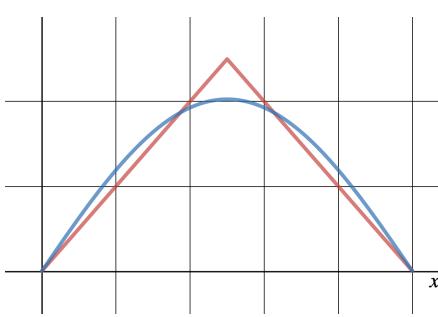
$$f(x) = \sum_{n=1}^{\infty} \sqrt{2} \left(\frac{L \sin\left(\frac{\pi n}{2}\right) - \frac{1}{2}\pi L n \cos\left(\frac{\pi n}{2}\right)}{\pi^2 n^2} + \frac{L \left(2 \sin\left(\frac{\pi n}{2}\right) + \pi n \cos\left(\frac{\pi n}{2}\right)\right)}{2\pi^2 n^2} \right) \sqrt{2} \sin\left(\frac{n\pi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

The question now is whether this is believable at all. Because, let's face it, this seems pretty odd. So, let's examine what these functions look like.

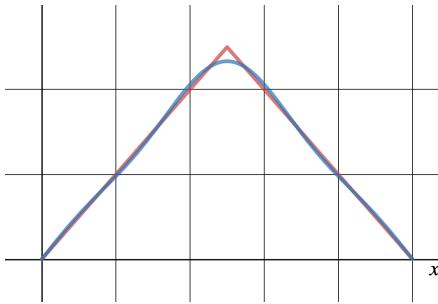
There is no nice way of evaluating the above infinite sum, but we can take finite sum approximations from this. For example, we can choose some positive integer N and take

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

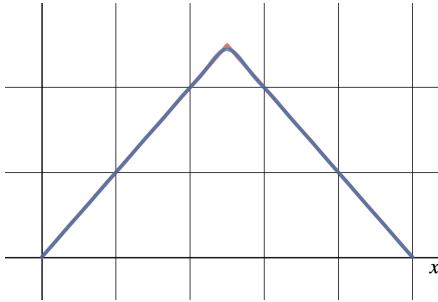
Let us plot various degrees of approximations with the original function $f(x)$.



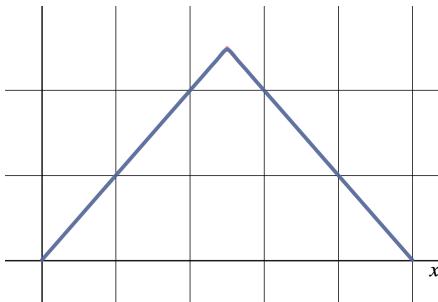
(a) Approximation of $f(x)$ with $n = 1$ basis elements.



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



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



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

Figure 17.1: Approximation of $f(x)$ with $n = 50$ basis elements.

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

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



Linear Operators

18.1 Matrices

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

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

which transformed space. We wrote

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

Recall that a linear transformation is one that satisfies

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

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

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

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

Linear Equations

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

$$[A]\vec{x} = \vec{b}, \quad (18.1.1)$$

for the vector \vec{x} . This problem was solved by row reduction. However, we also found that we could (in general) invert the matrix $[A]$ to produce $[A]^{-1}$. This method was far more powerful since we could quickly solve the 18.1.1 with any given vector. Specifically, we have that

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

In fact, the eigenvalue problem is extremely related to this idea as well. Given the problem in 18.1.1, we can isolate the matrix $[A]$ and decompose the action of the matrix into scaling on individual vectors. That is, we solve the eigen-equation

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

Scaling is an easy process to invert since we have

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

Exercise 18.1.1. Can you prove the above statement?

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

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

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

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

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

Exercise 18.1.2. Can you show the work between the above steps?

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

Adjoints

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

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

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

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

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

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

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

18.2 Linear Operators

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

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

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

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

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

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

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

also converges on $[0, L]$.

Exercise 18.2.1. Can you argue why this must be true? *Hint: use the ratio test.*

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

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

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

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

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

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

18.2.1 Adjoint

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

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

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

Example 18.2.1: Matrix Adjoint

Let

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

Then let us take the inner (dot) product

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

First, we compute

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

and then

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

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

$$\langle \vec{u}, [A]^\dagger \vec{v} \rangle = u_1 v_1 + (u_1 + u_2) v_2.$$

Let us put

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

then

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

Then we also require that

$$u_1 v_1 + (u_1 + u_2) v_2 = \langle \vec{u}, [A]^\dagger \vec{v} \rangle = u_1(av_1 + bv_2) + u_2(cv_1 + dv_2).$$

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

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

is just the transpose of $[A]$.

Exercise 18.2.2. Can you prove that $[A]^\dagger = ([A]^*)^T$ when $[A]$ is an arbitrary complex 2×2 matrix? Hint: use the same steps as above but start with an arbitrary matrix and use the Hermitian inner product.

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

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

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

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

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

which changes the phase of a wavefunction.

Example 18.2.2: Phase Operator

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

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

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

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

This leads us to the equation

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

Thus, it must be that

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

Then, taking the complex conjugate of both sides we have

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

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

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

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

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

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

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

18.2.2 Hermitian Operators

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

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

In other words, we have

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

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

Theorem 18.2.1: Hermitian Matrix Eigenvalues and Eigenvectors

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

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

Then we have

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

but also

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

Hence it must be that $\lambda = \lambda^*$ and thus λ must be real.

To prove the second statement, we take two eigenvectors \vec{e}_j and \vec{e}_k corresponding to distinct eigenvalues λ_j and λ_k and note

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

as well as

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

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

$$\lambda_j = \lambda_k$$

which contradicts our original supposition or that

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

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

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

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

means that L is Hermitian.

Theorem 18.2.2: Hermitian Operators Eigenvalues and Eigenfunctions

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

Proof. The proof is analogous to the previous. \square

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

Example 18.2.3: Laplace Operator is Self Adjoint

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

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

Dividing both sides by the constants on both sides yields

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

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

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

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

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

which correspond to the normalized eigenfunctions

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

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

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

when $n \neq m$.

Exercise 18.2.3. Fill in the missing steps in the integration by parts above.

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

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

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

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

$$Lf = g,$$

which exactly parallels the equation

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

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

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

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

18.3 Differential and Integral Operators

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

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

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

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

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

$$Lf = g,$$

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

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

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

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

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

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

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

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

so that

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

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

Later on in this course we visit new forms of differential operators as we increase the spatial dimension. For example, in 3-dimensional space, we have a notion of rotation of a vector field. The example in 16.2.1 shows us a field that rotates. To compute how much rotation the field has, we need the differential operator

$$\nabla \times,$$

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

$$\nabla \cdot .$$

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

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

18.3.1 Integral Operators

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

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

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

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

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

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

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

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

18.4 Spectra

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

Earlier in this chapter, we compared the equations

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

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

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

Exercise 18.4.1. Revisit the solution for the free particle in the 1-dimensional box. The equation given was

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

In this case, our linear operator is

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

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

Example 18.4.1: Spectrum for the Derivative on \mathbb{R}

In the prequel, we began studying ODEs with the equation

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

Notice, that we can take

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

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

We can note that this equation is separable, and the particular solution is

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

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

One may ask how this helps us solve the equation

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

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

Exercise 18.4.2. Review how to solve the above separable equation.

Remark 18.4.1. The story of this methodology for solving differential equations is a long one. There is a great history of this method beginning in the early 1800's with Joseph Fourier. As of now, there is still active research in differential equations where these types of methods are generalized and applied to more problems. We will seek to cover the linear theory with a few specific examples. Just know that many details are being glossed over for the sake of promoting understanding as opposed to rigor.



Fourier Theory

19.1 Fourier Series

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

$$\mathcal{L}f = g, \quad (19.1.1)$$

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

For the sake of example, we will take the operator $\mathcal{L} = -\frac{d^2}{dx^2}$ which is referred to as the 1-dimensional **Laplace operator** or **Laplacian**. To follow the methodology outlined in the previous chapter, before we approach 19.1.1, we must find the spectrum of \mathcal{L} . Hence, we must solve the eigenvalue equation

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

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

$$-\frac{d^2}{dx^2} f(x) = \omega^2 f(x) \quad \text{with } f(0) = f(L). \quad (19.1.2)$$

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

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

When we solved the matrix eigenvalue problem

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

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

In our equation 19.1.2, we know the general solution is

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

Exercise 19.1.1. Determine the above general solution using either the characteristic polynomial (which in reality comes from a determinant) or by using a power series.

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

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

and

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

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

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

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

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

We can then solve for ω to find

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

Exercise 19.1.2. Review the analysis above for determining ω .

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

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

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

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

19.1.1 Solving Equations

The equation 19.1.1 before is linear, and hence any sum of solutions is also a solution. Thus, we can write a solution $f(x)$ as a series by putting

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

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

Let us define the following inner product

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

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

Though we won't prove it here, the eigenfunctions found above serve as a basis for the solutions for the equation 19.1.1. To simplify future work, we will normalize our eigenfunctions. Thus, we must solve

$$\langle c_1, c_1 \rangle = 1, \quad \left\langle c_o \sin\left(\frac{2n\pi x}{L}\right), c_o \sin\left(\frac{2n\pi x}{L}\right) \right\rangle = 1, \quad \left\langle c_e \cos\left(\frac{2n\pi x}{L}\right), c_e \cos\left(\frac{2n\pi x}{L}\right) \right\rangle = 1$$

Solving for the constants, we get

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

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

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

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

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

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

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

as well as the terms a_n and b_n by

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

Example 19.1.1: Square Wave

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

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

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

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

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

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

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

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

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

compute the constants b_n by

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

Thus, we have that

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

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

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

for various values of N

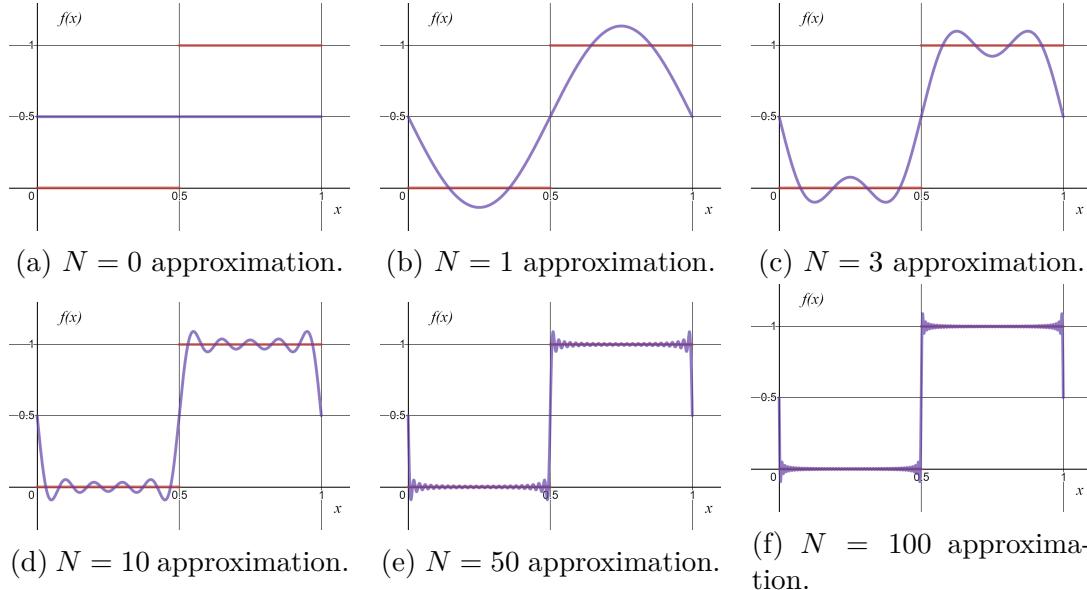


Figure 19.1: True $f(x)$ in red. Approximations in purple.

19.2 Fourier Transform

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 19.2.1: Square Wave Fourier Transform

If we consider the square wave function given by

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

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

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

Thus, we compute the coefficients in the same by

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

Hence, the Fourier transform of f is given by

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

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

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

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

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

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

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

19.2.1 Fourier Transform Operator

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

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

Later, we will introduce the inverse operation in that

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

These operators are fundamental in solving differential equations!

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

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

Now, the operator is unitary since we also have that

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

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

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

19.3 The Dirac Delta Function

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

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

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

$$\int_a^b f(x)\delta(x)dx = f(0).$$

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

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

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

Remark 19.3.1. $\delta(x)$ is in fact not a function at all. It is something more general known as a distribution. But this fact is not entirely relevant. We will continue to refer to $\delta(x)$ as a function despite this slight misnomer.

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

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

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

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

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

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

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

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

Remark 19.3.2. Computing these integrals requires tools from complex analysis that we have not seen. Thus, we will have to blackbox how these integrals are computed in exchange for their usefulness.

19.4 Computing Fourier Transforms

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

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

Example 19.4.1: Fourier Transform of Cosine

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

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

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

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

and

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

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

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

Example 19.4.2: Fourier Transform of Sine

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

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

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

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

Example 19.4.3: Fourier Transform of Complex Exponential

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

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

and thus we can compute the Fourier transform by

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

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

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

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

19.5 The Inverse Fourier Transform

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

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

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

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

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

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

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

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

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

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

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

If one revisits example 18.2.2, then we can see that the adjoint to multiplication by a phase is to multiply by the negative phase. That is, the adjoint to $e^{i\theta}$ is $e^{-i\theta}$. That is the fact that lets us define the inverse Fourier transform above.

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

19.6 Solving Differential Equations

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

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

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

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

Exercise 19.6.1. Using integration by parts, show that the work above is correct.

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

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

We also have the following as well

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

This means that the Fourier transform turns multiplication into differentiation.

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

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

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

Example 19.6.1: Fourier Series Solution of a Second Order Equation

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

We can model this equation by

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

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

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

which has the following graph.

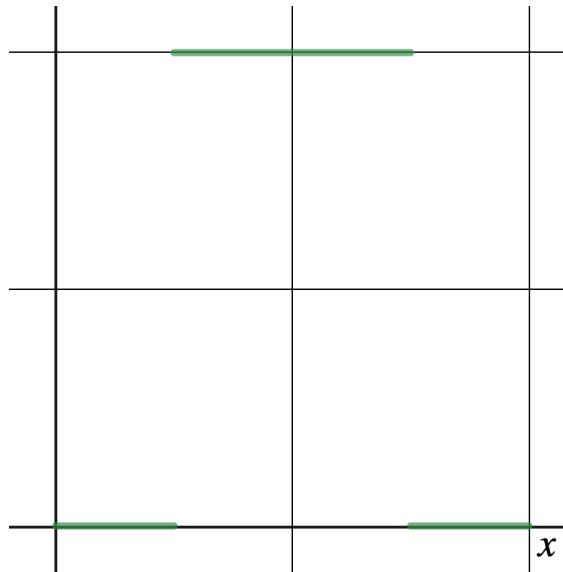


Figure 19.2: A plot of the external force $g(x)$.

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

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

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

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

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

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

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

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

Thus we have the general solution

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

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

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

Evaluating the above infinite series gives us

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

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

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

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

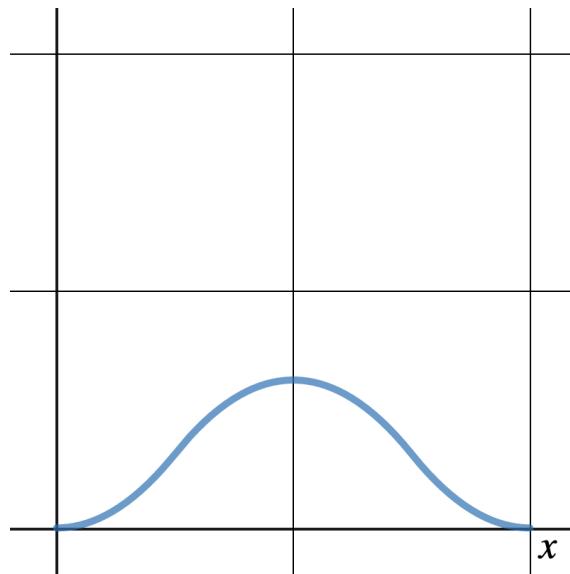


Figure 19.3: An approximation to the solution $f(x)$ with $N = 100$.

Exercise 19.6.2. Verify that the Fourier series for $g(x)$ above is correct.

Example 19.6.2: Fourier Transform Solution of a Second Order Equation

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

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

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

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

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

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

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

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

Hence, the Fourier transform of the whole differential equation reads

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

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

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

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

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

Remark 19.6.1. Though arguably less intuitive, the Fourier transform approach is a bit cleaner to work with. If one were to plot approximations for our function $f(x)$, we would receive the same answer as in the prior example.

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

(technically, not even function) right hand side.

Example 19.6.3: Fourier Transform for a Second Order Equation on \mathbb{R}

Consider the following equation

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

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

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

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

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

Thus, we can then find

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

Computing this inverse Fourier transform yields

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

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

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

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

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

which garners the solution

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

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

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