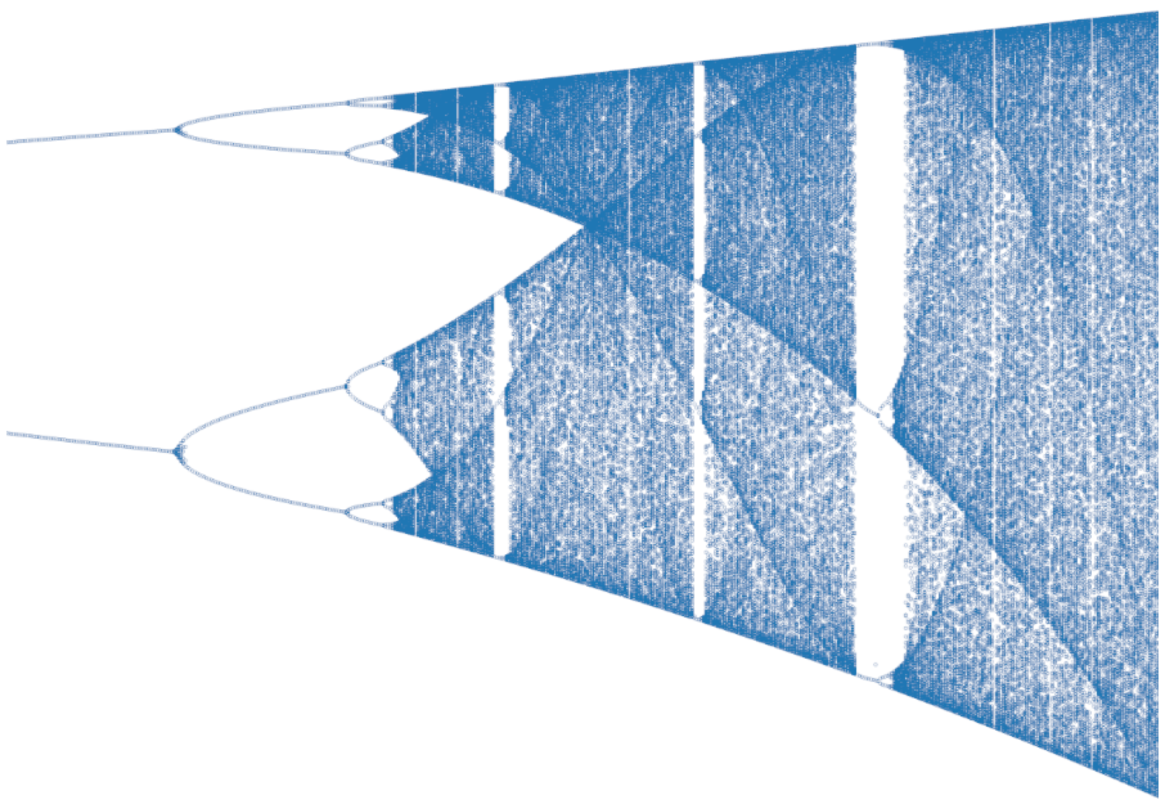
MATHEMATICS LESSON NOTES

BY

JOSEPH MAKOWSKI

MATH AND SCIENCE TUTORING



UPDATED

January 29, 2024

Preface

This is a collection of mathematics lesson notes that I have written for my tutoring business, MATH AND SCIENCE TUTORING. The text is self-contained and only assumes prior knowledge of basic arithmetic at approximately the elementary school level, including conceptual familiarity with exponents, rational numbers, and using variables in functions and equations. Generally, each new chapter assumes familiarity with the content presented in the previous chapters.

This is a "living document" which is added to and revised according to the needs of my current students; the most recent version of this book can be found online on my website. This document should be a useful supplement for most high school courses in Algebra, Geometry, Pre-Calculus/Trigonometry, or Calculus. However, I have also included additional topics from areas of mathematics beyond the scope of a typical high school course, providing students a bedrock for understanding mathematics in greater depth. In the Trigonometry chapter, for instance, I take the opportunity to introduce concepts like equivalence relations, injective and surjective functions, and connections of trigonometric functions to other areas of math, including geometric transformations and complex numbers. For similar reasons, I tend to prefer the inclusion of proofs, except where I feel they are not very instructive.

Tutoring and writing these notes are essentially hobbies of mine that I find time for outside of work; thus, I have not (yet) had the time to add a large number of examples or exercises to each section, and given the choice, I almost always prioritize the addition of new topics. Although it was not my explicit intention, I think a strength of this approach is that most concepts are described concisely without a lot of "fluff", allowing one to quickly grasp the core ideas and techniques that underpin a topic. Nonetheless, I encourage students to seek out exercises to practice these techniques with thier own hands, and future versions of these notes (or a companion document) will eventually include exercises for each section. Math is not a spectator sport!

Thanks to my friends and colleagues that continue to provide feedback and suggestions to improve the notes, and a big thanks to my students for their patience, curiosity, and enthusiasm.

JOSEPH A. MAKOWSKI
MATH AND SCIENCE TUTORING
SOLE PROPRIETOR

Content Status and Updates

- **Chapter 1: Algebra**. Re-writing for clarity, to include topics needed for some middle school entrance exams, and to include a greater amount of basic material on functions like $f(x) = mx + b$ or the absolute value. I also need to fill in various ways one can solve for the roots of polynomials.

- **Chapter 2: Basic Probability and Statistics**. I plan to write this such that algebra is the only pre-requisite to reflect "algebra-based" statistics classes in many schools (e.g., AP Stats)

- **Chapter 3: Euclidean Geometry**. In progress. I still have to typeset some pen and paper notes I have starting from Symmetry and Transformations

- **Chapter 4: Trigonometry**. No major additions are planned at the moment.

- **Chapter 5: Single Variable Calculus**. Early in the chapter, I need to fill in some details about limits of functions, continuity, and maybe some details about the standard metric topology over $\mathbb{R}$; a bit later, I need to go back and expand on some applications, series, maybe rewrite the section introducing the Riemann integral, and complete the remainder of the chapter starting from numerical methods. However, in terms of the content covered in a *typical* calculus course, this chapter is close to "complete".

- **Chapter 6: Abstract Linear Algebra**. In contrast with the "linear algebra" taught in many high school Algebra II courses, this will be more reflective of a typical "Advanced Linear Algebra" course in college.

- **Chapter 7: Multivariate Calculus and Vector Analysis**. Planned. I would like to put very short introductions to other math topics at the end so students have some perspective on "what comes next".

- **Chapter 8: Data Analysis and "Machine Learning"**. I may just make a separate document for this.

Table of Contents

Preface **iii**

1 Algebra **1**

1.1 Introduction . . . 1
1.1.1 Basic Set Theory and Types of Numbers . . . 1
1.1.2 Set Operations . . . 4
1.1.3 Functions . . . 7
1.1.4 Equations and Inequalities . . . 8
1.1.5 Real Number Algebra . . . 10
1.1.6 Complex Number Algebra . . . 14
1.1.7 Exponents and Radicals . . . 16
1.1.8 Table of Helpful Rules and Properties . . . 18
1.1.9 Inverse Functions and Composition . . . 19
1.1.10 Domain, Codomain, Image, Preimage, and Range . . . 21
1.2 Straight Line Functions . . . 22
1.2.1 Introduction . . . 22
1.2.2 Transforming a Line . . . 23
1.2.3 Absolute Value Function . . . 24
1.3 Polynomials . . . 25
1.3.1 Introduction to Polynomials . . . 25
1.3.2 Transforming and Graphing Polynomials . . . 26
1.3.3 Factoring Polynomials (aka "Reverse FOIL") . . . 27
1.3.4 The Quadratic Formula and "Completing the Square" . . . 28
1.3.5 Finding Roots with a Graphing Calculator . . . 29
1.4 Rational Functions . . . 30
1.4.1 Introduction to Rational Functions . . . 30
1.4.2 Asymptotes of Rational Functions . . . 31
1.5 Exponential and Logarithmic Functions . . . 33
1.5.1 Introduction to Exponential Functions . . . 33
1.5.2 Introduction to Logarithms . . . 35
1.5.3 Applications of Exponents and Logarithms . . . 37
1.5.4 Inequalities . . . 41
1.6 Basic Linear Algebra . . . 44
1.6.1 Linear Equations and Transformations . . . 44
1.6.2 Systems of Linear Equations . . . 45

1.6.3 Matrix Algebra . . . 47
1.6.4 What are Matrices? . . . 50
1.6.5 Applications of Linear Equations . . . 52

2 Basic Probability and Statistics 53
2.1 Introduction . . . 53
2.2 Summary Statistics . . . 53
2.3 Probability Theory . . . 53
2.4 The Normal Distribution . . . 53
2.5 Sampling and Estimation . . . 53
2.6 Hypothesis Testing . . . 53
2.7 Regression . . . 54

3 Euclidean Geometry 55
3.1 Introduction . . . 55
3.2 Logic, Reasoning, and Proof . . . 56
3.2.1 Introduction to Propositional Logic . . . 56
3.2.2 Logical Operators . . . 56
3.2.3 Rules of Inference . . . 56
3.2.4 Examples . . . 56
3.3 Geometry with Lines . . . 57
3.3.1 Distance Formula . . . 57
3.3.2 Midpoint Formula . . . 57
3.3.3 Parallel and Perpendicular Lines . . . 57
3.3.4 Transversal Intersecting Parallel Lines . . . 57
3.4 Symmetry and Transformations . . . 60
3.4.1 Introduction . . . 60
3.4.2 Basic Algebra Perspective . . . 60
3.4.3 "Linear Algebra" Perspective . . . 61
3.4.4 Abstract Algebra Perspective . . . 61
3.5 Triangles . . . 62
3.5.1 Triangle Relationships . . . 62
3.5.2 Congruence and Similarity . . . 62
3.5.3 Right Triangles . . . 62
3.6 Quadrilaterals . . . 63
3.7 Polygons . . . 64
3.7.1 Classifying Polygons . . . 64
3.7.2 Interior and Exterior Angles . . . 64
3.8 Circles . . . 65
3.9 Constructions . . . 66
3.10 Perimeter, Area, and Volume . . . 67
3.11 Conic Sections . . . 68
3.11.1 Ellipses . . . 68
3.11.2 Parabola . . . 68
3.11.3 Hyperbola . . . 68

4 Trigonometry **69**
4.1 The Unit Circle 69
4.1.1 Definition and Angles 69
4.1.2 Relations and Equivalence Relations 71
4.1.3 Unit Circle Reference Page 75
4.2 Trigonometric Functions 76
4.2.1 Sine and Cosine Functions 76
4.2.2 Injective, Surjective, and Bijective Functions 82
4.2.3 Tangent and Secant Functions 84
4.2.4 Cotangent and Cosecant Functions 88
4.2.5 Inverse Trigonometric Functions 89
4.3 Proving Trigonometric Idenities 92
4.3.1 Angle Addition and Subtraction 92
4.3.2 Double Angle and Half Angle 94
4.3.3 Product to Sum 95
4.3.4 Law of Cosines 96
4.3.5 Law of Sines 97
4.4 Applications of Trigonometry to Vector Algebra and Complex Numbers . . . 98
4.5 Table of Important Identities 105

5 Single Variable Calculus **107**
5.1 What are Real Numbers? 107
5.1.1 What are Natural Numbers? Integers? Rational Numbers? 107
5.1.2 Why do we need $\mathbb{R}$? 108
5.1.3 Properties of $\mathbb{R}$ 109
5.2 Sequences and Limits 111
5.2.1 Limit of a Sequence 111
5.2.2 Table of Common Sequences and Limits 115
5.2.3 Solving Limits of Sequences with the Algebraic Limit Theorem 115
5.2.4 Cauchy Sequences and Solving Limits of Sequences with a Computer 116
5.2.5 Limits of Functions $\mathbb{R} \to \mathbb{R}$ 118
5.2.6 Squeeze Theorem 120
5.3 Continuous Functions and Topology of the Reals 121
5.4 Definition(s) of the Derivative 122
5.5 Derivative Rules and Theorems 126
5.5.1 Linearity of the Derivative 126
5.5.2 Chain Rule 127
5.5.3 Product and Quotient Rules 127
5.5.4 L'Hôpital's rule 129
5.5.5 Derivatives of Exponential and Logarithmic Functions 130
5.5.6 Derivative of x^n and Polynomials 132
5.5.7 Derivatives of Trigonometric Functions 133
5.5.8 Table of Derivative Rules and Theorems 134
5.6 Applications of Derivatives 135
5.6.1 Optimization Problems 135

5.6.2 Related Rates . . . 135
5.6.3 Implicit Differentiation . . . 135
5.7 Introduction to Series . . . 137
5.8 Riemann Integral and Antiderivatives . . . 138
5.9 Integration Rules and Theorems . . . 141
5.9.1 Linearity of the Antiderivative . . . 141
5.9.2 First Fundamental Theorem of Calculus . . . 142
5.9.3 Second Fundamental Theorem of Calculus . . . 142
5.9.4 Integral Rules from Derivative Rules . . . 142
5.10 Table of Integration Rules and Theorems . . . 143
5.11 Integration Strategy: u Substitution . . . 144
5.12 Integration Strategy: Integration by Parts . . . 145
5.12.1 Rapid Repeated Integration By Parts . . . 146
5.13 Integration Strategy: Partial Fraction Decomposition . . . 147
5.14 Integration Strategy: Long Division . . . 150
5.15 Integration Strategy: Complete the Square . . . 151
5.16 Integration Strategy: Trigonometric Substitution . . . 153
5.17 Integration Strategy: Numerical and Approximative Approaches . . . 155
5.17.1 Add Rectangles . . . 155
5.17.2 Taylor and Maclaurin Series . . . 155
5.18 Improper Integrals . . . 156
5.19 Analytic Geometry . . . 156
5.20 Basic Probability Theory . . . 156
5.21 Introduction to Ordinary Differential Equations . . . 156
5.22 Applications of Ordinary Differential Equations . . . 157
5.22.1 Rocket Science . . . 157
5.22.2 "Romeo and Juliet" Dynamics of Love . . . 157
5.22.3 Law of Mass Action . . . 157

6 Abstract Linear Algebra **159**
6.1 Fields . . . 160
6.2 Vector Spaces . . . 160
6.2.1 Definition and Properties . . . 160
6.2.2 Subspaces . . . 160
6.2.3 Linear Independence . . . 160
6.2.4 Bases . . . 160
6.3 Linear Transformations . . . 160
6.3.1 Introduction . . . 160
6.3.2 Image and Kernel . . . 160
6.3.3 Rank and Nullity . . . 160
6.3.4 Invertible Transformations . . . 160
6.3.5 Composition of Transformations . . . 160
6.4 Normed Spaces and Inner Product Spaces . . . 160
6.4.1 Definitions . . . 160
6.4.2 Orthogonal Subspaces . . . 160

6.4.3 Orthogonal Projection 160
6.4.4 Orthonormal Bases 160
6.5 Determinants 160
6.5.1 Definition 160
6.5.2 Inverse by Determinant 160
6.5.3 Cramer's Rule 160
6.6 Eigenvalues and Eigenvectors 160
6.6.1 Definition 160
6.6.2 Characteristic polynomial 160
6.6.3 Diagonalization 160
6.6.4 Eigenspaces 160
6.6.5 Applications 160

7 Multivariate Calculus and Vector Analysis 161
7.1 Functions of Multiple Variables 161
7.2 Vector Valued Functions 161
7.3 Polar and Spherical Coordinates 161
7.4 Partial Derivatives 161
7.5 Multiple Integrals 161
7.6 Vector Analysis 161
7.6.1 Gradient Theorem 162
7.6.2 Divergence Theorem 162
7.6.3 Curl Theorem 162
7.6.4 Green's Theorem 162
7.7 Glimpse at Partial Differential Equations 162
7.8 Glimpse at Complex Analysis 162
7.8.1 Sequences and Series 162
7.8.2 Cauchy-Riemann Equations 162
7.8.3 Cauchy's Integral Formula 162
7.9 Glimpse at Calculus on Manifolds 163
7.9.1 Differential Forms 163
7.9.2 Generalized Stoke's Theorem 163

8 Data Analysis and "Machine Learning" 165
8.1 Introduction 165
8.2 Data Processing 165
8.3 Dimensionality Reduction 165
8.3.1 Principal Component Analysis (PCA) 166
8.3.2 Isomap 166
8.3.3 Uniform Manifold Approximation and Projection (UMAP) 166
8.4 Clustering 166
8.4.1 k-Nearest Neighbors (k-NN) 166
8.4.2 k-means 166
8.4.3 Spectral Clustering 166
8.5 Regression and Classification 166

Chapter 1

Algebra

1.1 Introduction

1.1.1 Basic Set Theory and Types of Numbers

You are likely familiar with the notion of "different types of numbers" or the need to measure different things in different ways. To begin these notes, we will describe different types of numbers that commonly appear throughout mathematics, as well as some of their **algebraic** properties.

Before we begin, it is necessary to define what a **set** is in mathematics. Loosely speaking, it is just a collection of things. If we wanted to write "the set containing the letters a, b, and c" in common mathematical notation, we would write

$$\{a, b, c\}$$

where curly brackets indicate that a list of things are part of a set together. Mathematicians tend to categorize different kinds of numbers by saying they belong to different sets; then they give the set a distinctive name. To illustrate, let's introduce the set of **Natural numbers** $\mathbb{N}$:

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$$

The set of natural numbers includes all counting numbers like 1, 10, 34, or one hundred billion. It is common to say that these numbers are **elements** of the set(s) they belong to. Symbollically, we write this as

$$10 \in \mathbb{N}$$

which means "10 is a natural numbers" or "10 is an element of $\mathbb{N}$." A similar set is the **strictly positive natural numbers** $\mathbb{N}_+$:

$$\mathbb{N}_+ = \{1, 2, 3, 4, \dots\}$$

The only difference between $\mathbb{N}_+$ and $\mathbb{N}$ is that $0 \in \mathbb{N}$ while $0 \notin \mathbb{N}_+$ (0 is not an element of the natural numbers). Since $\mathbb{N}_+$ has all the natural numbers inside it except zero, we say that the strictly positive natural numbers are a **subset** of the natural numbers. To be precise, a set A is a subset of another set B if all elements of A are also elements of B. Typically, mathematicians notate this

$$\mathbb{N} \subset \mathbb{W}$$

where the "$\subset$" symbol is like a "less than" symbol for sets.

If we add or multiply two natural numbers or whole numbers together, we always get another natural; however, if we subtract two natural numbers, it is possible to get a negative number, which is no longer in the set $\mathbb{N}$. In other words, we say that $\mathbb{N}$ is **closed** under addition and multiplication but not closed under subtraction. If we want to be closed under subtraction as well, we need a set that also contains the negative numbers. What we are describing is the set of **Integers** $\mathbb{Z}$:

$$\mathbb{Z} = \{\cdots - 4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$$

We can add, subtract, and multiply any integers together and still get an integer back out. However, we still cannot divide. This leads us to the set of **Rational numbers** $\mathbb{Q}$:

$$\mathbb{Q} = \left\{\frac{a}{b} \mid a, b \in \mathbb{Z},\ b \neq 0\right\}$$

What we wrote above is called **set builder notation**. Going off of the color code, it is saying that $\mathbb{Q}$ contains all fractions of the form a/b such that the numbers a and b are both integers, but b does not equal 0 (since dividing by zero is not defined). Set builder notation is very common in mathematics since most sets we are interested in follow some sort of pattern that can be described in closed form. For example, the set of **even numbers** $\mathbb{E}$ could be defined as

$$\mathbb{E} = \{2n \mid n \in \mathbb{Z}\}$$

which, ultimately, means "the set of all the integers with a factor of 2." We can also write the set of **odd numbers** in a few creative ways:

$$\{2n + 1 \mid n \in \mathbb{Z}\}$$

$$\{n + 1 \mid n \in \mathbb{E}\}$$

$$\{n \mid n \in \mathbb{Z} \text{ and } n \notin \mathbb{E}\}$$

A shorthand way of writing the last version ("the set of numbers n that are integers but not even numbers") is to write

$$\mathbb{Z} - \mathbb{E}$$

which means "the set of elements of $\mathbb{Z}$ that are not also in $\mathbb{E}$."

The sets of numbers we have described so far have led us to the rational numbers $\mathbb{Q}$, which are closed under addition, multiplication, subtraction, and division. Do we even need any more numbers? It turns out, the answer is "yes, we do." As far back as ancient Greece, mathematicians had some awareness of numbers which could not be expressed as a ratio of integers. For example, given the very simple and short equation

$$x^2 = 2$$

it is possible to prove that the solution is not a rational number. Additionally, given a circle, it turns out the ratio of its circumference to its diameter, a constant we call π, can also never be perfectly expressed as a ratio of integers. Thus, mathematicians sought to define sets which contain more numbers than $\mathbb{Q}$, allowing us to do more sophisticated algebra and geometry than what made sense with only rational numbers.

The set that "fills in the gaps" between rational numbers, adding numbers like $\sqrt{2}$ and π, is the **Real numbers** $\mathbb{R}$. This is the set that students of algebra work with most of the time due to various nice properties it has, several of which we discussed above. It is common to write $\mathbb{R}$ in **interval notation**, which looks like

$$\mathbb{R} = (-\infty, \infty)$$

which means "$\mathbb{R}$ contains all numbers between (but not including) $-\infty$ and ∞." As another example, consider the **unit interval** I in this notation:

$$I = [0, 1]$$

Notice that we use square brackets now. This means "the set of all numbers between 0 and 1, including 0 and 1." This means the same as

$$I = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$$

Finally, if the unit interval represents a **closed** interval because it contains its endpoints and $\mathbb{R}$ is an **open** interval because it does not, it is worth noting that we can have also have "half-open" intervals, like

$$(0, 5] = \{x \in \mathbb{R} \mid 0 < x \leq 5\}$$

All of the numbers in $\mathbb{R}$ that are not in $\mathbb{Q}$ (i.e., the set $\mathbb{R} - \mathbb{Q}$), is called the **irrational numbers** and contains numbers like π, Euler's constant e, $\sqrt{2}$, and many others that one may struggle to even write.

$\mathbb{R}$ has many attractive properties, but it still fails one last test: the square root of negative numbers is not defined. For example,

$$x^2 = -1$$

is another scenario where we have need for even more numbers. Indeed, it is actually very common in algebra to end up in a situation where you need to take the square root of a negative! Our solution is the **complex numbers** $\mathbb{C}$, defined as follows:

$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R},\ i^2 = -1\}$$

Every complex number has a "real part" (a) and an "imaginary part" (ib). For that reason, it is common to plot a single complex number on a 2D coordinate grid, as you will see later. Technically, all real numbers are also complex numbers, but the imaginary componenet is equal to zero. Despite the name "complex," $\mathbb{C}$ is perhaps the "nicest" set we have for algebra and has many attractive properties for applications in physics and broader mathematical study.

In summary, we have the following relation between the sets we defined:

$$\mathbb{N}_+ \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

When we discuss **functions** in later sections, it will become clear that the idea of "closure" of operations like addition, multiplication, division, "square roots," etc. is actually very meaningful for how we can do algebra; we want the correct set of numbers for the job!

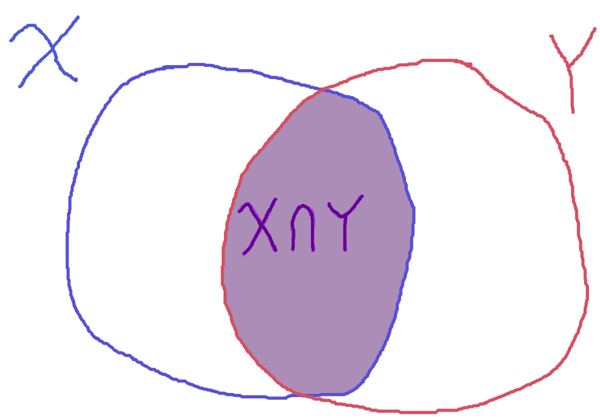
1.1.2 Set Operations

Let $A = \{a, b, c\}$, $B = \{x, y, z\}$, and $C = \{c, z\}$ be sets. Informally speaking, we can see that the sets A, B, and C are all different from each other. For one thing, A and B don't have any elements in common; and although C may have some elements in common with either A or B, it is a smaller set that only has two elements instead of three. In this section, we will study **set operations** that provide a mathematical framework for making similar comparisons.

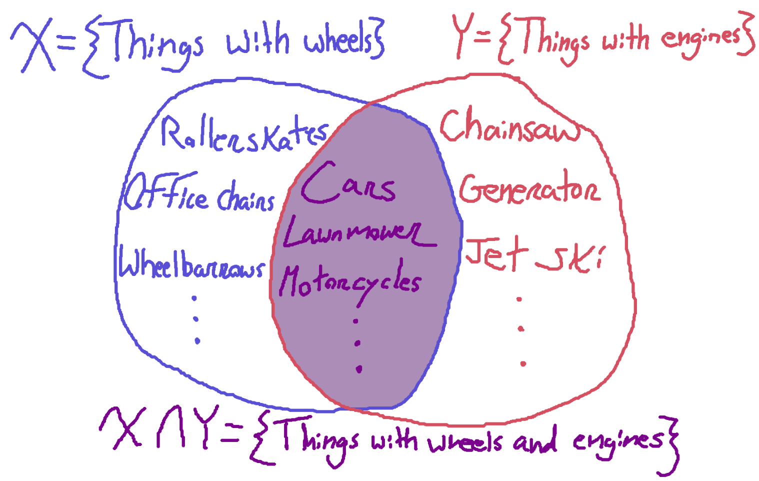
The **intersection** of the sets X and Y, denoted by $X \cap Y$, is a new set defined by the following rule:

$$X \cap Y = \{z \mid z \in X \text{ and } z \in Y\}$$

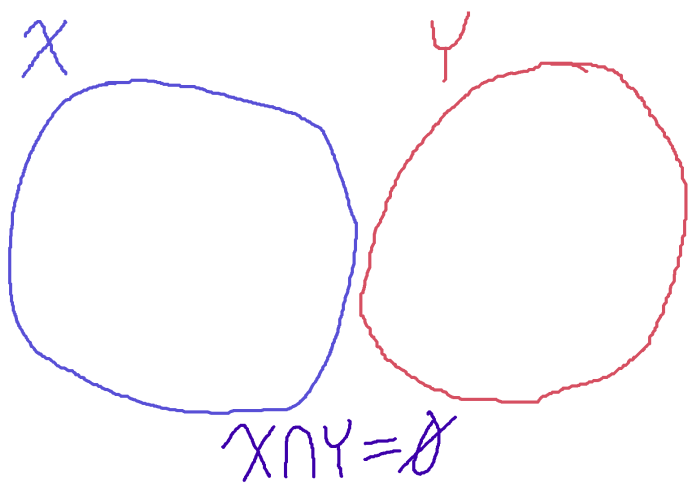
In other words, $A \cap B$ is the set of elements that A and B have in common. We can represent this graphically with a **Venn Diagram**:



Each circle represents a set, and the purple highlighted area where they overlap represents elements they have in common. For example, consider the sets of "things with wheels" and "things with engines":



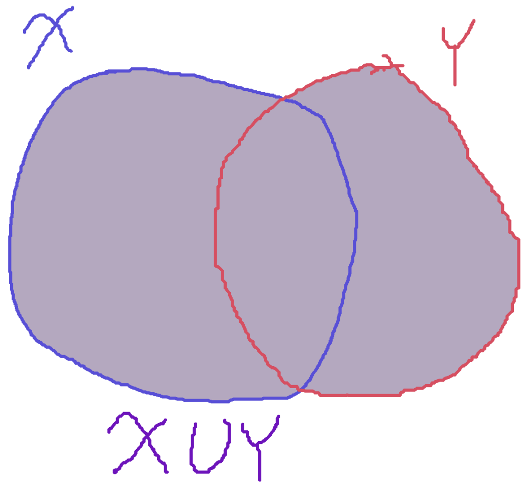
Consider the intersection of the sets A and B we defined earlier. We observed that they have **no elements in common**, so the intersection $A \cap B = \{\}$, an **empty set** with no elements in it. Typically, we denote the empty set by the symbol $\emptyset$, and if the intersection $X \cap Y = \emptyset$, then we call X and Y **disjoint**. For instance, the set of even numbers is disjoint from the odd numbers because it is impossible for a number to be both even and odd. As a Venn Diagram, we can depict disjoint sets as circles that never overlap with each other:



Another prominent set theoretical operation is called the **union**. We denote the union of two sets X and Y by $X \cup Y$, and it is defined as follows:

$$X \cup Y = \{z \mid z \in X \text{ or } z \in Y\}$$

In other words, as long as an element is in either X or Y (or both!), then it is inside $X \cup Y$. Graphically, the union is very straightforward:



The union is just everything from either set. Consider the union of the sets A and C from earlier:

$$A \cup C = \{a, b, c, z\}$$

Consider the following set:

$$Z = \left\{x \mid x \in \mathbb{R} \text{ and } \frac{1}{(x-1)(x+1)} \in \mathbb{R}\right\}$$

The presence of the word "and" possibly has you thinking that Z should be the intersection of two sets. You are correct! We can write

$$Z = \mathbb{R} \cap D$$

where, without overcomplicating things, we can think of D as a subset of the complex numbers $\mathbb{C}$, but we will worry about the details in later sections. By finding the intersection of $\mathbb{R}$ and D, however, we know that Z only has real numbers in it. Since division by zero is not defined (i.e., $\frac{1}{0} \notin \mathbb{R}$), to figure out what specific elements are in Z, we need to figure out which values of x make the expression $(x-1)(x+1) = 0$ true. We can verify that plugging in $x = 1$ or $x = -1$ satisfy the equation:

$$(1-1)(1+1) = 0 \cdot 2 = 0 \qquad \text{and} \qquad (-1-1)(-1+1) = (-2) \cdot 0 = 0$$

Since $x = 1$ or $x = -1$ produce $1/0$ for the given expression, which is not a real number, we know that $-1, 1 \notin Z$. Ultimately, this means we can write Z in the following way by using the union operator:

$$Z = (-\infty, -1) \cup (-1, 1) \cup (1, \infty) = \{x \in \mathbb{R} \mid x \neq -1 \text{ and } x \neq 1\}$$

Basically, the union and intersection operators give us flexibility to express what elements in are in a set a variety of different ways.

(planned - set difference, cardinality, cartesian product)

1.1.3 Functions

A **function** is a rule that matches one number to another. Typically, we say that a function maps a number from one set, the **domain**, to another, the **codomain**. If f is a function, X is the domain, and Y is the codomain, we express how the function maps one number to another by the following notation:

$$f\colon X \to Y$$

Most of the functions you are familiar take a real number x and somehow transform it into another real number y. We denote such a function

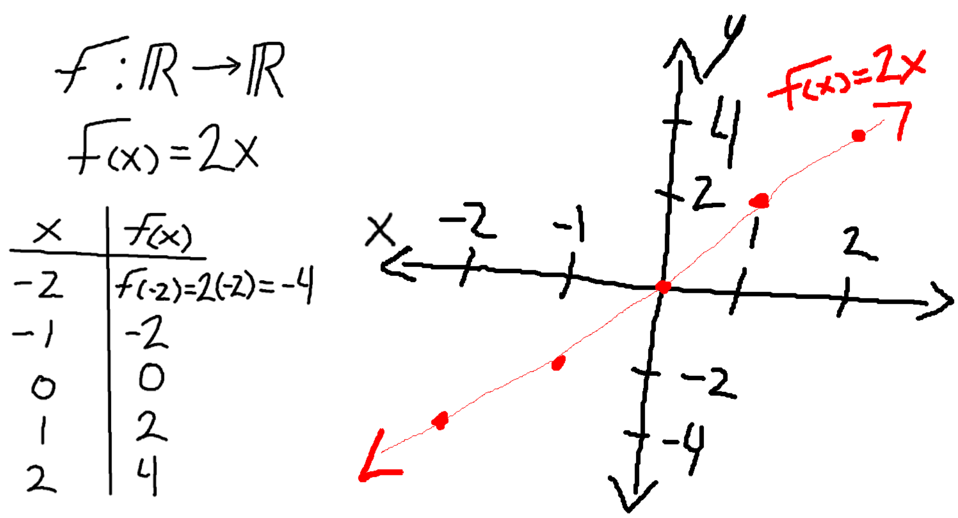
$$f\colon \mathbb{R} \to \mathbb{R}$$

which means "the function f maps a real number to another real number." Further, we say that $y = f(x)$ since, given a value of x, our function transforms it into y. Once we establish the domain and codomain of a function, it is common to specifically define it using common operations or other unambiguous rules of assignment. For instance, our function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x^2 + \pi$$

which squares the input x and adds the constant π to determine which element of the codomain x is mapped to. Functions need to be "well-defined," which simply means that it should only give you one output for every input, and it must have clear, consistent mappings from the domain to the codomain.

It is common to visualize the features of a function by **graphing** it. Study the figure below, depicting the process of graphing a function:



On the left hand side, notice that we create a table where we check the value of $f(x) = 2x$ for various values of x. A sample calculation is shown for $f(-2)$. Then, on the right hand side, we construct a set of **coordinate axes**. Notice that the **horizontal axis** is labaled x because it represents the domain of the function, while the **vertical axis** is labled y since it represents the codomain. For each value of x we tested in the table, we find it on the $x-$axis and then move up or down the $y-$axis until we reach the value of $f(x)$, where we place a point representing the value of the function. After plotting all five points from the table, we see that the general shape of the function is a straight line, so we "connect the dots" and trace out the shape of the function to show what $f(x)$ equals for numbers we didn't explicitly test in the table. Usually, the more points you compute by hand in a table, the more accurate your sketch of the function will be. When you need a very accurate graph of a function, using a computer is a good idea since it can quickly compute thousands of points along the graph of $f(x)$.

Not all functions are very easy to plot. In the example above, we just had one input variable x and one output y. But what about this function?

$$f\colon \mathbb{R}^3 \to \mathbb{R}^{100} \quad \text{defined by} \quad f(x,y,z) = (\text{100 different numbers depending on x, y, or z})$$

This may seem like a silly example, but it is not. We use the set $\mathbb{R}^3$ to represent 3D space, like positions of objects in the real world, so this function may be taking someone's location from a cell phone GPS and returning 100 different numbers to an app associated with the location, like the temperature at the location, information about the nearest stores, the average speed of nearby cars, and so on. That is, functions that are simply too big (or "too highly dimensional") to graph on one axis are very common in real life. Scenarios like this are also common in various fields of math, which may want to study functions from all kinds of sets, not just ones that are easy to plot.

In cases where we cannot use graphical representations of a function to understand its features, understanding mathematics becomes essential. No matter how big the domain and range get, many functions can still be studied and understood by using algebra or other math rules.

1.1.4 Equations and Inequalities

An **equation** simply refers to a mathematical argument with an equality sign (=). For example,

$$x^2 = 5$$

Not all equations represent functions. Some equations may have multiple solutions or no solutions at all, making them unsuitable for representing functions. Instead, many equations represent a **solution set** of numbers which make the equality true. In the case of $x^2 = 5$, the solution set is

$$X = \{x \in \mathbb{R} \mid x^2 = 5\} = \{\sqrt{5}, -\sqrt{5}\}$$

Similarly, an **inequality** is a relation which tells us how certain numbers are not the same, such as

$$x > 5$$

where the solution set represents all the values of x for which the inequality is true. On the other hand, equations written with two or more variables can often be interpreted as functions. For example, the equation

$$y = x - 5$$

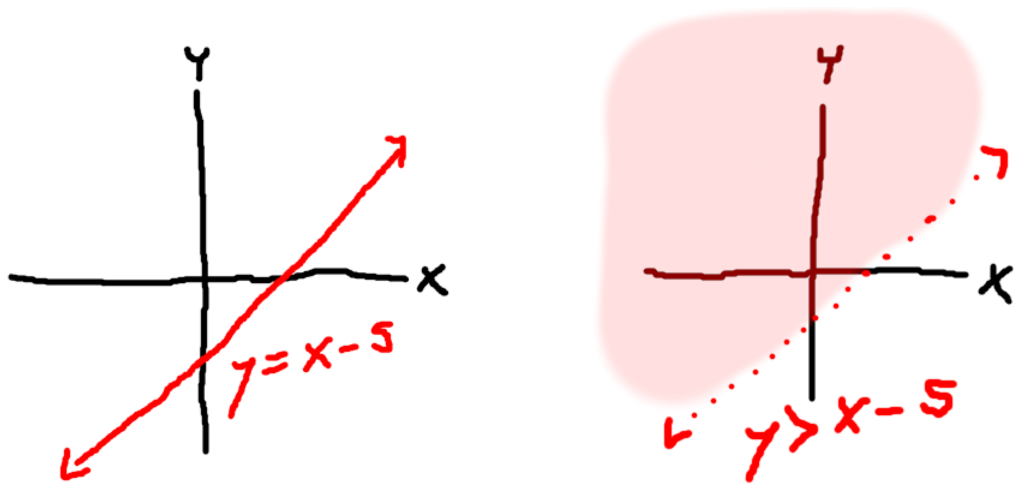
can be viewed as a function if we take $y = f(x)$. Otherwise, we can still conceptualize the equation as representing a solution set where we need to account for the presence of two unknowns. If we take $x, y \in \mathbb{R}$, then our solution set would be

$$X = \{(x, y) \in \mathbb{R}^2 \mid y = x - 5\}$$

Sets like the one above are called the **graph** of the function $y = f(x)$. Likewise, inequalities with two variables, such as

$$y > x - 5$$

commonly represent solution sets involving both variables. Graphically, if you plotted the solution set (or function) represented by the first equation, it would be a straight line with slope $m = 1$ and y-intercept $b = -5$. The plot of the inequality, on the other hand, would include many more values of x and y above the line. In fact, we would have to shade half the plane.



Notice that in the sketch above, dotted lines are used along the boundary of the shaded region. That is because, on the dotted line, y is **equal to** $x - 5$, NOT greater than $x - 5$. Even though it helps us establish the edge of the region where we have solutions, we can't include it.

1.1.5 Real Number Algebra

As you have likely come to appreciate from our discussion from earlier in the chapter, algebra does not necessarily work the same way on all sets of numbers. The fact that we have "real numbers" and "complex numbers" at all is largely driven by mathematicians' desire to have sets that accomodate more mathematical possibilities. With this in mind, it should make sense that mathematicians have come to characterize the rules of algebra in the real numbers by a specific list of rules. In this section, we will discuss each of these rules and, where possible, try to put into simple terms what they mean.

1. (Associative property of addition): $x + (y + z) = (x + y) + z$

This property says that the order you add real numbers doesn't matter. For example,

$$1 + (2 + 3) = 1 + 5 = 6$$

and

$$(1 + 2) + 3 = 3 + 3 = 6$$

Note that this also applies to negative numbers. For example:

$$(1 - 2) + 3 = 1 + (-2 + 3)$$

Be careful! Observe that we **do not** have the following equality:

$$(1 - 2) + 3 \neq 1 - (2 + 3)$$

Keep the negative sign "glued" to the correct number.

2. (Associative property of multiplication): $x(yz) = (xy)z$

This is similar to the previous rule: the order you multiply in doesn't matter. For example,

$$2(3 \cdot 4) = 2 \cdot 12 = 24$$

and

$$(2 \cdot 3) \cdot 4 = 6 \cdot 4 = 24$$

3. (Commutative property of addition): $x + y = y + x$

This property tells you that you are allowed to switch the order you add numbers. For example:

$$2 + 3 = 5 = 3 + 2$$

Similar to the associative property, this also applies to negative numbers:

$$2 - 3 = -1 = -3 + 2$$

Be careful! "Subtraction" itself is **not** commutative:

$$2 - 3 \neq 3 - 2$$

Again, make sure to keep track of which number the negative sign is stuck to and keep it there!

4. (Commutative property of multiplication): $xy = yx$

The order numbers are multiplied can be switched. For example:

$$5 \cdot 3 = 15 = 3 \cdot 5$$

With multiplication, you need to be much less careful about negative signs. This is because $-a = (-1) \cdot a$ and the factor of (-1) itself can commute around! For instance:

$$3 \cdot (-5) = (-5) \cdot 3 = (-1) \cdot (5 \cdot 3) = -(5 \cdot 3) = -(3 \cdot 5)$$

5. (Existence of additive identity): There exists $0 \in \mathbb{R}$ for which $0 + x = x$

This rule just gives the name of "additive identity" to 0 because it has the special property of not changing a number when added.

6. (Existence of multiplicative identity): There exists $1 \in \mathbb{R}$ for which $1 \cdot x = x$

This rule just gives the name of "additive identity" to 1 because it has the special property of not changing a number when multiplied.

7. (Existence of additive inverse): For each $x \in \mathbb{R}$, there exists $-x \in \mathbb{R}$ for which $x + (-x) = 0$

For every real number, there is a negative version, and when you take their sum, you get 0. That is,

$$5 - 5 = 0$$

Although this may seem obvious, remember that many sets (e.g., the natural numbers) do not have this property, so it is still worthwhile to note.

8. (Existence of multiplicative inverse): For each $x \in \mathbb{R}$, there exists $x^{-1} \in \mathbb{R}$ for which $x \cdot x^{-1} = 1$ $(x, x^{-1} \neq 0)$

In this case, we have that $x^{-1} = 1/x$, so this rule says that the reciprocol of every real number is still a real number (except 0), and the product of a real number with its reciprocol is 1. For example:

$$5 \cdot \frac{1}{5} = \frac{5}{5} = 1$$

9. (Distributive property of multiplication over addition): $x(y + z) = xy + xz$

One of the most important and useful rules. As we will see in the second half of this section, this rule has many practical applications. For example:

$$5 \cdot 3 + 5 \cdot 2 = 5(3 + 2)$$

We can use this rule to remove a common factor from two or more numbers added together or to distribute such a common factor over a sum.

10. (Distinct additive and multiplicative identities): $0 \neq 1$

This is another instance where the rule seems obvious, but mathematicians have discovered some sets where the additive identity ("0") and the multiplicative identity ("1") are actually the same number. Consider the set $\{0\}$ as an example. Suppose $x \in \{0\}$. Then $x + 0 = x$, so 0 is the additive identity. We also have $x \cdot 0 = 0$, but the only possibility in $\{0\}$ is $x = 0$. Since multiplying by 0 technically keeps the value of x the same, it must be the case that 0 is *also* the multiplicative identity. We make sure to state that $0 \neq 1$ to emphasize that the real numbers are different from unusual cases like these.

It is now time to consider some practical aspects of real number algebra that many people (unfortunately!) overlook.

- "Multiply by 1"

  As a motivating example, suppose your math teacher asks you to "rationalize" the following expression by taking the radical out of the denominator:

  $$\frac{1}{\sqrt{2}}$$

  This is an instance where "multiplying by 1" will serve as a helpful manipulation. Recall: if $a \in \mathbb{R}$ and $a \neq 0$, then

  $$\frac{a}{a} = 1$$

  but also that, if $b \in \mathbb{R}$, then

  $$b \cdot 1 = b$$

  Consequentially, if we multiply by a/a, which itself just equals 1, then our expression is **unchanged**:

  $$b \cdot \frac{a}{a} = b$$

  Let's return to our motivating example. Since we want to remove anything containing a radical from the denominator and we know that $(\sqrt{2})^2 = 2$, we can make the following observation to simplify:

  $$\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{2}}{\sqrt{2}} = \frac{\sqrt{2}}{\sqrt{2}^2} = \frac{\sqrt{2}}{2}$$

  We have successfully "rationalized" the expression. Explicit attention will be drawn to other uses of this manipulation in the following section of complex number algebra.

- "Add 0"

  This technique is very similar to "multiplying by 1". Recall that for any $a, (-a) \in \mathbb{R}$ that we have

  $$a + (-a) = 0$$

  and that, for some $b \in \mathbb{R}$, it is also true that

  $$b + 0 = b$$

so,

$$b + (a - a) = b$$

Deliberately "adding 0" enjoys many applications in manipulating **polynomials**, which we will discuss at length in future sections. For now, consider the following simpler example:

$$\frac{24}{5} = \frac{24 + (1-1)}{5} = \frac{25-1}{5} = \frac{5^2 - 1}{5} = 5 - \frac{1}{5}$$

We were able to manipulate the numerator to make it easier to divide by 5.

- "FOIL" (First Outer Inner Last) In algebra, it is very common to find expressions like

$$(a + b)(c + d)$$

By the distributive property, this is equal to

$$(a + b)c + (a + b)d$$

By the distributive property (again), we have

$$ac + bc + ad + bd$$

A popular neumonic for remembering this process is "FOIL", which stands for "first, outer, inner, last", referring to taking the product of the first two numbers in either quanitity (ac), the outer pair (ad), the inner pair (bc), and the last pair (bd). Suppose $a = c = x$. Then this problem becomes

$$(x + b)(x + d) = (x)(x) + (x)c + b(x) + bd = x^2 + (b + c)x + bd$$

and if $b = d = y$, then we can further simplify to

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

The latter two cases are common occurances when dealing with polynomials. Although the first case where none of the values are assumed be equal may seem complicated, using the distributive property in that manner can be a helpful heuristic for various problems.

Suppose you had to compute $18 \cdot 27$ without a calculator. One way to do it is to realize that

$$18 \cdot 27 = (10 + 8)(20 + 7)$$

Since you know the distributive property, you can now break one difficult multiplication problem into four easy ones:

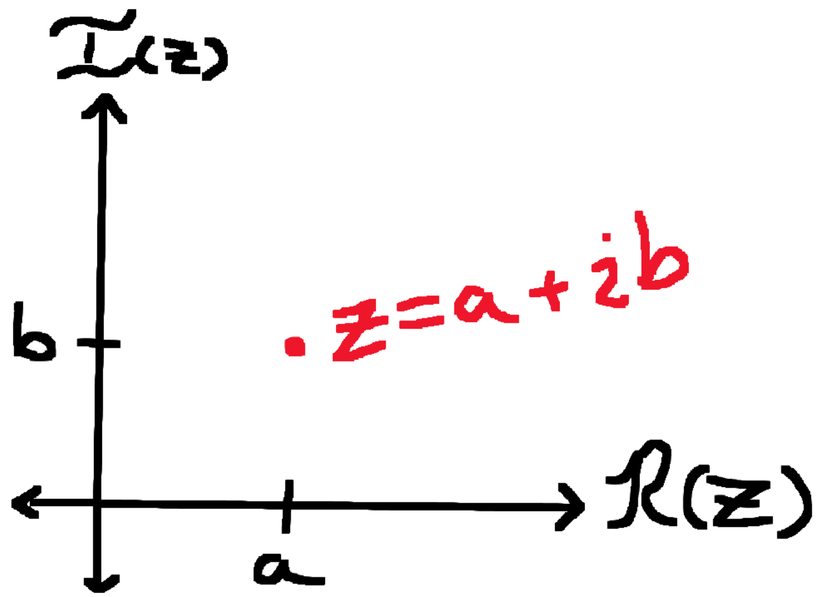
$$(10+8)(20+7) = 10 \cdot 20 + 8 \cdot 20 + 10 \cdot 7 + 8 \cdot 7 = 200 + 160 + 70 + 56 = 360 + 126 = 486$$

1.1.6 Complex Number Algebra

Recall that every complex number $z \in \mathbb{C}$ can be written

$$z = a + ib$$

where $a, b \in \mathbb{R}$ are real numbers, a is called the **real component** of z, and b is called the **imaginary component** of z. Instead of plotting individual complex numbers on a number line, we instead need to use an entire coordinate plane which breaks up the real and imaginary components.



On occasion, people will use the notation $\Re(z)$ and $\Im(z)$ to denote the real and imaginary parts of z, respectively.

It happens to be the case that the exact same properties we explored in the previous section for algebraic operations on $\mathbb{R}$ also apply to $\mathbb{C}$. In general, the practical differences in algebra arise from the presence of the imaginary component of complex numbers.

Let $z, w \in \mathbb{C}$ be any two complex numbers, such that $z = a + ib$ and $w = (c + id)$. In the case of addition, we have the following:

$$\begin{aligned} z + w &= (a + ib) + (c + id) \\ &= a + ib + c + id && \text{(Associative property)} \\ &= (a + c) + (ib + id) && \text{(Commutative property)} \\ &= (a + c) + i(b + d) && \text{(Distributive property)} \end{aligned}$$

In words, we add the real components and imaginary components of each complex number separately. The method for subtracting two complex numbers is almost identical to the way we add them; just subtract each component from each other.

Multiplication is a bit less straightforward:

$$\begin{aligned}
z \cdot w &= (a+ib)\cdot(c+id) \\
&= (a+ib)c + (a+ib)id && \text{(Distributive property)} \\
&= ac + (ib)c + a(id) + (ib)(id) && \text{(Distributive property)} \\
&= ac + ibc + iad + i^2bd && \text{(Commutative property)} \\
&= ac + ibc + iad - bd && (i^2 = -1) \\
&= (ac - bd) + (ibc + iad) && \text{(Commutative property)} \\
&= (ac - bd) + i(bc + ad) && \text{(Distributive property)}
\end{aligned}$$

Unlike addition, the fact that $i^2 = -1$ ends up being important, turning the product of two imaginary numbers into a real number.

How about division? When we want the quotient z/w, we get an expression like this:

$$\frac{z}{w} = \frac{a+ib}{c+id}$$

We want to be able to write any complex number as the sum of a real component and an imaginary component. To achieve this goal, we need to find a way to simplify. For that purpose, we will introduce the **complex conjugate**.

Let $z = a + ib$. Then the **complex conjugate** of z is the number $z^* = a - ib$, where the sign of the imaginary component is switched. Observe an important property of the complex conjugate:

$$z \cdot z^* = (a+ib)\cdot(a-ib) = a^2 - iab + iab + b^2 = a^2 + b^2$$

Multiplying a number by its complex conjugate produces a real number whose imaginary component goes to zero.

In our division problem, we can use the complex conjugate and the "Multiply by 1" trick discussed in the previous section to simplify. Observe:

$$\begin{aligned}
z/w &= \frac{a+ib}{c+id} \\
&= \frac{a+ib}{c+id}\cdot\frac{w^*}{w^*} && \text{(“Multiply by 1” trick)} \\
&= \frac{(a+ib)(c-id)}{c^2+d^2} && (w \cdot w^* = c^2 + d^2) \\
&= \frac{(ac+bd) + i(bc-ad)}{c^2+d^2} && \text{(Complex multiplication formula)} \\
&= \frac{ac+bd}{c^2+d^2} + i\left(\frac{bc-ad}{c^2+d^2}\right) && \text{(Distributive property)}
\end{aligned}$$

We use the quantity $z \cdot z^*$ to define the **modulus** $|z|$ of a complex number, defined by

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

The modulus represents the **distance** of the complex number z away from 0. We will discuss distance at length in future sections; for now, we introduce the modulus so we can simplify our formula for division a little bit. Notice that $c^2 + d^2 = |w|^2$. Applied to our formula above, we get the following:

$$\begin{aligned} z/w &= \frac{ac + bd}{c^2 + d^2} + i\left(\frac{bc - ad}{c^2 + d^2}\right) \\ &= \frac{ac + bd}{|w|^2} + i\left(\frac{bc - ad}{|w|^2}\right) \\ &= \frac{1}{|w|^2}\left((ac + bd) + i(bc - ad)\right) \quad \text{(Distributive property)} \end{aligned}$$

Finally, if we remember from earlier that $(ac+bd)+i(bc-ad) = z\cdot w^*$, then we can substitute to obtain

$$z/w = \frac{z \cdot w^*}{|w|^2}$$

While this is a lot "cleaner" (insofar as it users fewer symbols), our previous unsimplified formula will probably be more useful if you ever need to actually compute the quotient of complex numbers.

To conclude, let's briefly consider what $1/i$ equals. While we could use the big formula we just derived, there is a simpler way by "Multiplying by 1":

$$\frac{1}{i} = \frac{1}{i} \cdot \frac{i}{i} = \frac{i}{i^2} = \frac{i}{-1} = -i$$

We will revisit complex numbers in later sections to discuss how we can represent them using different kinds of functions using distances and angles.

1.1.7 Exponents and Radicals

When we first learn about exponents, they are usually presented in the context of repeated multiplication. For example,

$$2^3 = 2 \cdot 2 \cdot 2 = 8$$

This kind of exponentiation helps us discover the rule for "exponent addition," by the following reasoning:

$$2^5 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = (2 \cdot 2) \cdot (2 \cdot 2 \cdot 2) = 2^2 \cdot 2^3$$

Multiplying by 2 two times and also multiplying by 2 three times is the same as multiplying by 2 five times. So, in the example above, the exponents add $2 + 3 = 5$.

Mathematicians wondered what would happen if we took this rule and started to expand what it meant. One kind of equation they wanted to solve is

$$x^2 = 2$$

Using the rules of adding exponents, we can figure out

$$x^2 = x \cdot x = 2 \implies x = 2^{1/2}$$

because $\frac{1}{2} + \frac{1}{2} = 1$. However, we also know that $\sqrt{2} \cdot \sqrt{2} = 2$, so it stands to reason that $\sqrt{2} = 2^{1/2}$.

Next, mathematicians wanted to find a way to deal with division using exponents. If we think about equations like

$$\frac{x^2}{x} = \frac{x \cdot x}{x} = x$$

then we know that $1/x$ needs to be equal to x^{-1} because it matches the equation

$$x^2 \cdot x^{-1} = x^{2-1} = x$$

This is also how we know that anything to the power of 0 equals 1. See here:

$$1 = x \cdot x^{-1} = x^{1-1} = x^0$$

As such, we now get to work with all sorts of new exponents with rational numbers and negatives. Even though it is not always easy to write them out as repeated multiplication, mathematicians made sure that they at least follow the same general rules. These properties, as well as some of their consequences, are tabulated in the following section.

1.1.8 Table of Helpful Rules and Properties

Name	Property	Example
Order of Operations (PEMDAS)	Parentheses, exponents, multiplication/division, addition/subtraction	$5\left(\frac{\sqrt{4/2}}{3}\right)^2 = 5\left(\frac{\sqrt{2}}{3}\right)^2 = 5\left(\frac{2}{9}\right)$ $= 10/9$
Multiplicative inverses (or Negative exponents)	$\frac{x}{x} = x \cdot x^{-1} = 1$	$\frac{2}{2} = 2 \cdot 2^{-1} = 1$
Exponent Addition	$x^a \cdot x^b = x^{a+b}$	$2^{3/2} \cdot 2^{1/2} = 2^{\frac{3+1}{2}} = 2^2$
Exponent Multiplication	$(x^a)^b = x^{a \cdot b}$	$(2^2)^3 = 2^6$
Exponent "Distribution"	$(x \cdot y)^a = x^a \cdot y^a$	$\left(\frac{2}{3}\right)^2 = \frac{2^2}{3^2} = \frac{4}{9}$
"Double Negative" Exponents	$\frac{1}{x^{-1}} = (x^{-1})^{-1} = x$	$\frac{2}{1/2} = 2 \cdot (2^{-1})^{-1} = 2 \cdot 2 = 4$
Radicals	$\sqrt[n]{x} = x^{1/n}$	$\frac{1}{\sqrt{2}} = \frac{1}{2^{1/2}} = 2^{-1/2}$
Radical Exponentiation	$\left(\sqrt[n]{x}\right)^m = x^{m/n}$	$(\sqrt{2})^3 = 2^{3/2} = 2\sqrt{2}$
Radical Division	$\frac{\sqrt[n]{x}}{\sqrt[n]{y}} = \sqrt[n]{\frac{x}{y}}$	$\frac{\sqrt{6}}{\sqrt{3}} = \sqrt{\frac{6}{3}} = \sqrt{2}$

1.1.9 Inverse Functions and Composition

Consider some function $f\colon X \to Y$. How can we "undo" f? For that purpose, we introduce inverse functions.

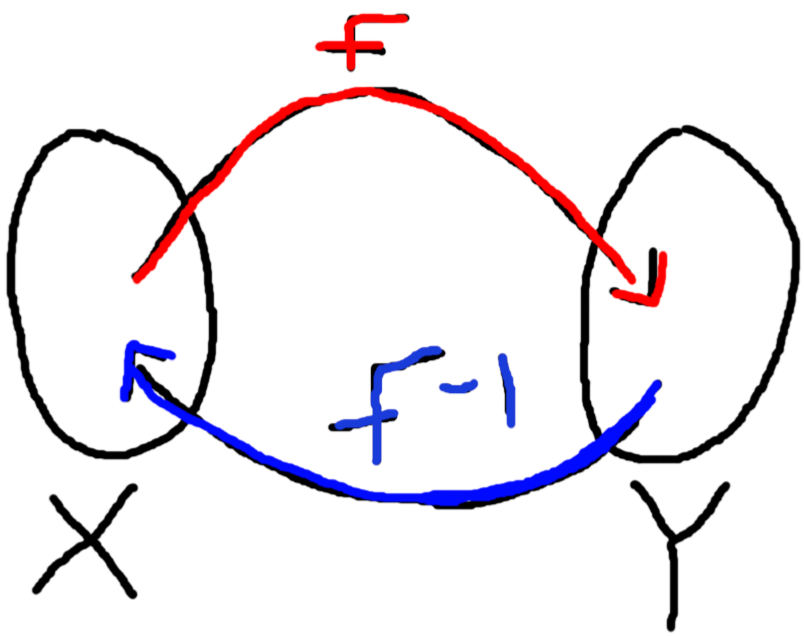
You are already familiar with multiple kinds of inverses. For every real number a, there is a unique additive inverse $(-a)$ such that

$$a + (-a) = 0$$

Likewise, for multiplication, we have the multiplicative inverse a^{-1}, for which

$$a \cdot a^{-1} = 1$$

We should expect inverse functions to behave similarly. The function f has a domain set X and a codomain set Y, so the inverse function f^{-1}, called "f inverse," which undoes f, should have the domain and codomain switched around, so $f^{-1}\colon Y \to X$.



For this section of the notes, we will write $x = f^{-1}(y)$ to emphasize how it is different from $y = f(x)$.

There is one last idea we need to know before starting with inverse functions, called **function composition**. All this means is you plug one function into another. Consider the two functions

$$f\colon \mathbb{N} \to \mathbb{R} \quad f(x) = x + \pi$$

$$g\colon \mathbb{R} \to \mathbb{C} \quad g(x) = ix$$

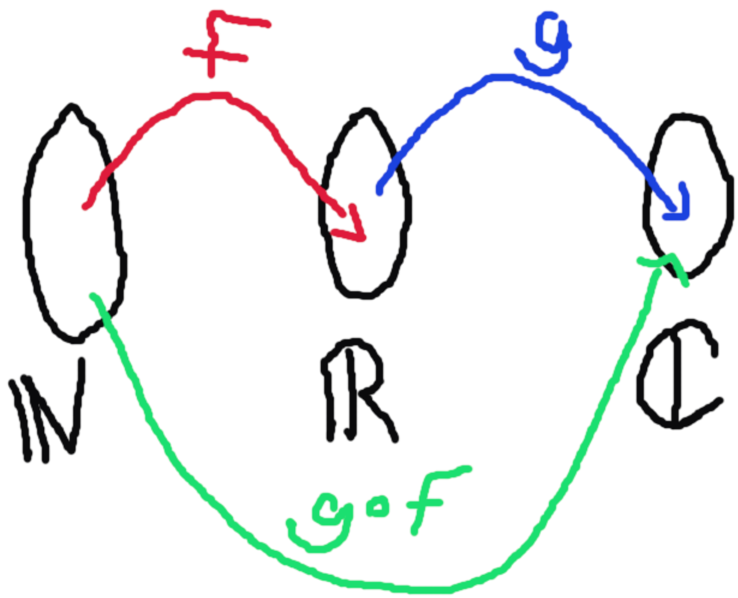
We can plug f into g like this:

$$g(f(x)) = i(x + \pi)$$

This can be thought of as a new function $g \circ f$, called "g of f." Since x orignally comes from the domain of f, the natural numbers $\mathbb{N}$, and eventually ends up in the codomain of g, the complex numbers $\mathbb{C}$, we can conclude that

$$g \circ f : \mathbb{N} \to \mathbb{C}$$

This is easier to understand with a picture:



Now we are prepared to study inverse functions. Let's work with something simple:

$$f(x) = 5x + 1$$

To find the inverse function, replace x with $f^{-1}(y)$ and $f(x)$ with y:

$$y = 5f^{-1}(y) + 1$$

Now solve for f^{-1}.

$$f^{-1}(y) = \frac{y-1}{5}$$

What happens when we take the composition $f^{-1} \circ f$?

$$f^{-1} \circ f(x) = f^{-1}(f(x)) = \frac{f(x)-1}{5} = \frac{5x+1-1}{5} = x$$

What happens when we take the composition $f \circ f^{-1}$?

$$f \circ f^{-1}(y) = f(f^{-1}(y)) = 5f^{-1}(y) + 1 = 5\left(\frac{y-1}{5}\right) + 1 = y$$

If we have that $f^{-1} \circ f(x) : X \to X$, then we call $f^{-1} \circ f(x) = \mathrm{id}_X(x) = x$ **the identity function** over X, or just the identity, because the function always equals exactly what you put in. This means that, algebraically, the inverse function acts just like an additive or multiplicative inverse!

1.1.10 Domain, Codomain, Image, Preimage, and Range

Let $f\colon \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = \sqrt{n}$. Based on what we learned from previous sections, we know from the definition of f that the **domain** is the natural numbers $\mathbb{N}$ and the **codomain** is the real numbers $\mathbb{R}$. We know this is true because the basic behavior of the function is to take a natural number and transform it into a real number by taking a square root. However, the domain and codomain are not the only sets mathematicians are interested in when studying functions. In the following section, we will define the **image**, **preimage**, and **range** of a function and describe how these sets are both similar and different from our familiar notion of domain and codomain.

Consider a general function $f\colon X \to Y$, and suppose $A \subseteq X$. Then the **image** of A, commonly denoted by $f(A)$ or $f[A]$, is the following set:

$$f(A) = \{f(a) \mid a \in A\}$$

That is, $f(A)$ is the set you get when every element of A is changed by f. Consider the following examples:

- If $C = \{1, 2, 3\}$ and $g(x) = x^2$, then $g(C) = \{1^2, 2^2, 3^2\} = \{1, 4, 9\}$.
- If $I = [0, 1]$ and $h(x) = x + 5$, then we get $h(I)$ by increasing every number in I by 5. The result is that $h(I) = [5, 6]$.

We will now use the concept of the image of a set to define a very important concept in mathematics. If $f\colon X \to Y$, then the **range** of f is the set $f(X) \subseteq Y$. That is, the range is the set of values in the codomain that are actually mapped to by the function. Although the range is sometimes equal to the codomain, it is important to realize that the range will often be a "smaller" set. Consider the following examples:

- Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Observe that if we plug a negative number into $f(x)$, then it becomes positive by cancellation of the negative signs:

  $$f(-a) = (-a)^2 = (-a)(-a) = a^2$$

  Consequentially, we find $f(\mathbb{R}) = [0, \infty)$ because it is impossible for $f(x)$ to equal a negative number when the domain is real numbers. In this case, the range is a proper subset of the codomain.

- Let $\text{id} : \mathbb{R} \to \mathbb{R}$ be defined by $\text{id}\,(x) = x$. Since the identity function $\text{id}\,(x)$ maps a number back to itself, we see that $\text{id}\,(\mathbb{R}) = \mathbb{R}$. In this case, the range and the codomain are equal.

The last type of set we will define in this section is the **preimage** of a function. Let $f\colon X \to Y$, and suppose $B \subseteq Y$. Then the **preimage** of B, denoted by $f^{-1}(B)$ or $f^{-1}[B]$, is the following set:

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}$$

Note that the preimage should not be confused with the inverse function; despite the similar notation, they are not exactly the same thing. Even if a function does not have a well-defined inverse, we can still find its preimage given some subset B of the codomain. Consider the following examples:

- Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$, and take $B = \{1, 4, 9\}$. Since $x^2 = (-x)^2$, there are up to **two** possible elements of the domain (x and $-x$) which can map to each element of the codomain. Thus, $f^{-1}(B) = \{-1, 1, -2, 2, -3, 3\}$.
- Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$, and take $B = \{-1, -2, 25\}$. It is impossible to find elements in $\mathbb{R}$ which square to a negative number, so this time we ignore everything except the positive number in B to obtain $f^{-1}(B) = \{-5, 5\}$.

Let's conclude by applying the concept of the preimage to better our understanding of the **domain**. Since a function needs to be well-defined, we need to make sure that the domain we specify doesn't contain values for which a function is undefined. This issue comes up often in algebra, and we will discuss it at length in later sections about **rational functions**.

Suppose we have a function $f\colon D \to \mathbb{R}$ defined by $f(x) = \frac{1}{x^2}$, and $D \subseteq \mathbb{R}$ is our domain. To find the biggest possible set D could possibly be, we simply need to find the preimage of our codomain. That is, we set $D = f^{-1}(\mathbb{R})$. It turns out that the only real number for which $f(x)$ is not defined is $x = 0$, so $D = (-\infty, 0) \cup (0, \infty)$.

1.2 Straight Line Functions

1.2.1 Introduction

The equation for a line is typically given by

$$y = mx + b$$

where m is the **slope** of a line and b is the y-intercept. This is called the **slope-intercept** form of a line. Alternatively, if we know one point (x_o, y_0) on the line and its slope, we can instead use

$$(y - y_0) = m(x - x_0)$$

which is called the **point-slope** form of a line. By algebraic manipulation, we find that

$$y = mx + (y_0 + mx_0)$$

By matching with the slope intercept form of a line, we find that it is possible to determine the y-intercept of a line from a point and its slope by the formula

$$b = y_0 + mx_0$$

Given two points $A(a, b)$ and $B(c, d)$, we can define a line which passes through both. First, use the slope formula:

$$m = \frac{\Delta y}{\Delta x} = \frac{d - b}{c - a}$$

Then use either one of the points in the point-slope formula, and we are done!

1.2.2 Transforming a Line

(planned - changing the slope, shifting left/right on horizontal axis, changing the y-intercept)

1.2.3 Absolute Value Function

(planned)

1.3 Polynomials

1.3.1 Introduction to Polynomials

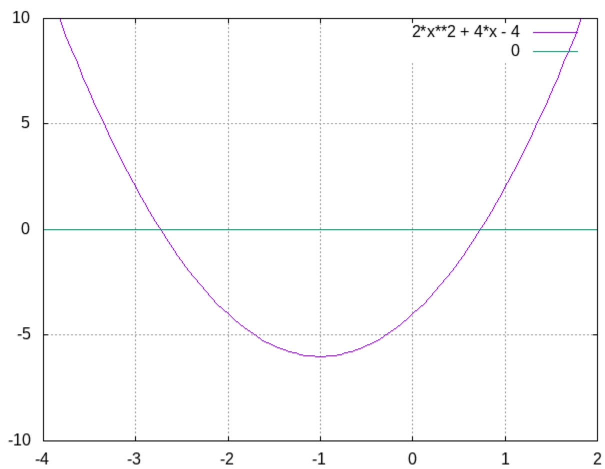
A **polynomial function** has the form

$$f(x) = c_n x^n + \cdots + c_3 x^3 + c_2 x^2 + c_1 x^1 + c_0 x^0$$

up to some **order** n, where each c is a constant (possibly equal to 0). This representation of polynomials is a little abstract, so we will work mostly with concrete examples. For instance,

$$g(x) = 2x^2 + 4x - 4$$

is a **2nd order** polynomial because the highest exponent is 2. Just like any other function, we can graph it and try to understand some of its special features:



The x-intercepts of polynomials are usually called the **roots** of the equation. The Fundamental Theorem of Algebra tells us that every polynomial has exactly the same number of roots as its order. In the case of the graph above, it is obvious that there are two roots because the graph has two x-intercepts. However, in some polynomials, the roots can be imaginary.

In the following sections, we will discuss two methods of finding the roots of polynomials.

1.3.2 Transforming and Graphing Polynomials

(planned)

1.3.3 Factoring Polynomials (aka "Reverse FOIL")

Consider the binomials given by

$$f(x) = x + 2$$

and

$$g(x) = x - 3$$

These are 1st order polynomials because the highest power is x^1. By setting each binomial equal to zero, is easy to see that the root of $f(x)$ is $x = -2$ and the root of $g(x)$ is $x = 3$.

If we multiply these two binomials together, we use the distributive property (FOIL) to get a new polynomial out, which we can call $h(x)$:

$$h(x) = (x + 2)(x - 3) = x^2 - x - 6$$

What are the roots of this polynomial? It is 2nd order, so we expect there to be two roots. We can **factor** a binomial out of the polynomial to make it more obvious which values of x make the equation equal to zero.

Pretending we don't know the factorization of $h(x)$ already, we can try the "reverse FOIL" method for finding factors.

First, check the x^2 term. This time, it is clearly just $x \cdot x$.

Next, check the last term, and find the factor tree. For -6, we have possible factors of -2 and 3 or -3 and 2.

Last, check the middle term. What two numbers do you add together to make the coefficient on x? In this example, the coefficient is -1, which we notice is the same as $-3 + 2$. So we can guess that

$$h(x) = (x + 2)(x - 3)$$

FOIL to check your work. We did everything correctly, so we can now figure out the roots of $h(x)$ from its factors.

If $(x + 2) = 0$, then the whole thing equals zero. Likewise, if $(x - 3) = 0$, then the whole thing equals zero. As such, we see that the roots are $x = -2$ and $x = 3$.

We can use this method to find polynomials with any roots we want. For example, if I wanted to find a polynomial with roots $x = 2$, $x = 3$, and $x = 5$, then we just need to write out the product of binomials

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

then FOIL.

1.3.4 The Quadratic Formula and "Completing the Square"

The quadratic formula lets us find the roots of 2nd order polynomials. In general, for 2nd order polynomails, we have

$$ax^2 + bx + c = 0$$

from which we can derive the quadratic formula. First, divide both sides of the equation by a to obtain the expression

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

Next, subtract c/a from both sides:

$$x^2 + \frac{b}{a}x = -\frac{c}{a}$$

Now, we want to use a technique called **completing the square**. Basically, the idea is to add a constant to both sides that would allow us to "reverse FOIL" the polynomial half into something easy to work with.

At this point, we observe that the middle x term could be broken into two halfs and written as the sum

$$\frac{1}{2}\frac{b}{a}x + \frac{1}{2}\frac{b}{a}x = \frac{b}{a}x$$

To complete the square, we now know that we need to add $(\frac{1}{2}b/a)^2$ to both sides:

$$x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 = \left(\frac{b}{2a}\right)^2 - \frac{c}{a}$$

Now we complete the square by "reverse FOILing" the polynomial side of the equation:

$$x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 = \left(x + \frac{b}{2a}\right)^2 = \left(\frac{b}{2a}\right)^2 - \frac{c}{a}$$

Get the right side of the equation into a common denominator:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}$$

Finally, take the square root of both sides and subtract $\frac{1}{2}b/a$ from both sides:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

This is the **quadratic formula**. As you can see, deriving it is a bit complicated, so most people just memorize this result. There are other equations which allow us to find the roots of 3rd and 4th order polynomials, but they are so long and complicated that nobody uses them.

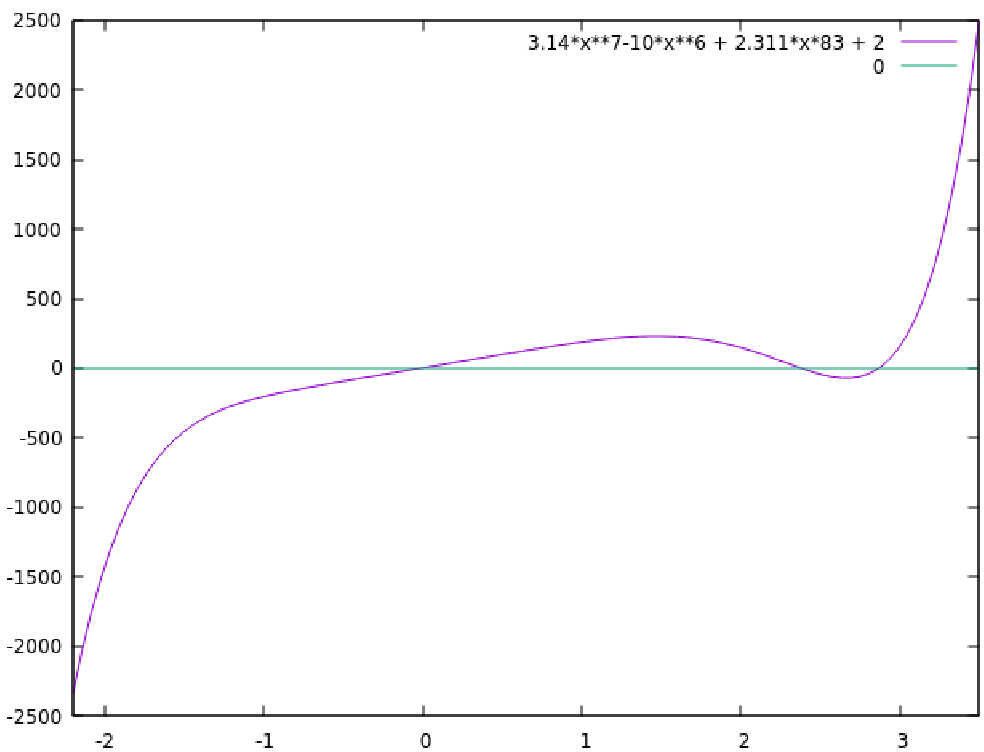
1.3.5 Finding Roots with a Graphing Calculator

Sometimes, it is too difficult to find roots by factoring, and it might be impossible to use the quadratic formula because the order of the polynomial is bigger than 2. In cases like these, we can still use a graphing calculator to obtain an approximate answer.

Consider the example polynomial

$$f(x) = \pi x^7 - 10x^6 + 2.311x \cdot 83 + 2$$

This is a 7th order polynomial, meaning there are a total of 7 roots. We can't use the quadratic formula, and it is too complicated to factor. In this case, one of our best options is to graph it on a calculator and visually identify where the roots are:



There appear to be three places where the graph crosses the x-axis: once near 0, and twice between 2 and 3. It is important to choose a viewing window that lets you see all the roots in clear detail. If the x-axis extended too far, the function would get so big that it would be impossible to tell what is happening with the roots. Trial and error will give you the best viewing window.

1.4 Rational Functions

1.4.1 Introduction to Rational Functions

Rational functions have the form
$$f(x) = \frac{p(x)}{q(x)}$$
where both p and q are functions of x. Despite how the name sounds, there is no requirement that any of the numbers involved in a rational function belong to the set of rational numbers $\mathbb{Q}$. For our purposes, rational functions will be real-valued.

When your rational function is built out of polynomials, such as
$$g(x) = \frac{x^2 + 2x + 1}{x + 1}$$
there are a couple strategies for attempting to simplify. First, try to "reverse FOIL" and divide out common factors:
$$g(x) = \frac{x^2 + 2x + 1}{x + 1} = \frac{(x + 1)^2}{x + 1} = x + 1$$
Sometimes, it may not be obvious how to do this, such as in the harder example:
$$h(x) = \frac{6x^2 + 7x - 20}{2x + 5}$$
We can try **polynomial long division**, demonstrated here:
$$\begin{array}{r|rrr} & 3x & -4 & \\ \hline 2x+5 & 6x^2 & +7x & -20 \\ & -6x^2 & -15x & \\ \hline & & -8x & -20 \\ & & 8x & +20 \\ \hline & & & 0 \end{array}$$
So our rational function becomes
$$h(x) = 3x - 4$$
Polynomials aside, we will conclude this section by introducing a helpful trick for simplifying rational expressions involving radicals: conjugates. Consider this function:
$$\psi(x) = \frac{x^2\sqrt{5}}{x - \sqrt{2}}$$
If we want to remove radicals from the denominator of a rational expression, then we must multiply by the conjugate:
$$\psi(x) = \frac{x^2\sqrt{5}}{x - \sqrt{2}} \cdot \frac{x + \sqrt{2}}{x + \sqrt{2}} = \frac{x^3\sqrt{5} + x^2\sqrt{10}}{x^2 - 2}$$
With conjugates, notice that the denominator will always use the **difference of two squares** rule.

1.4.2 Asymptotes of Rational Functions

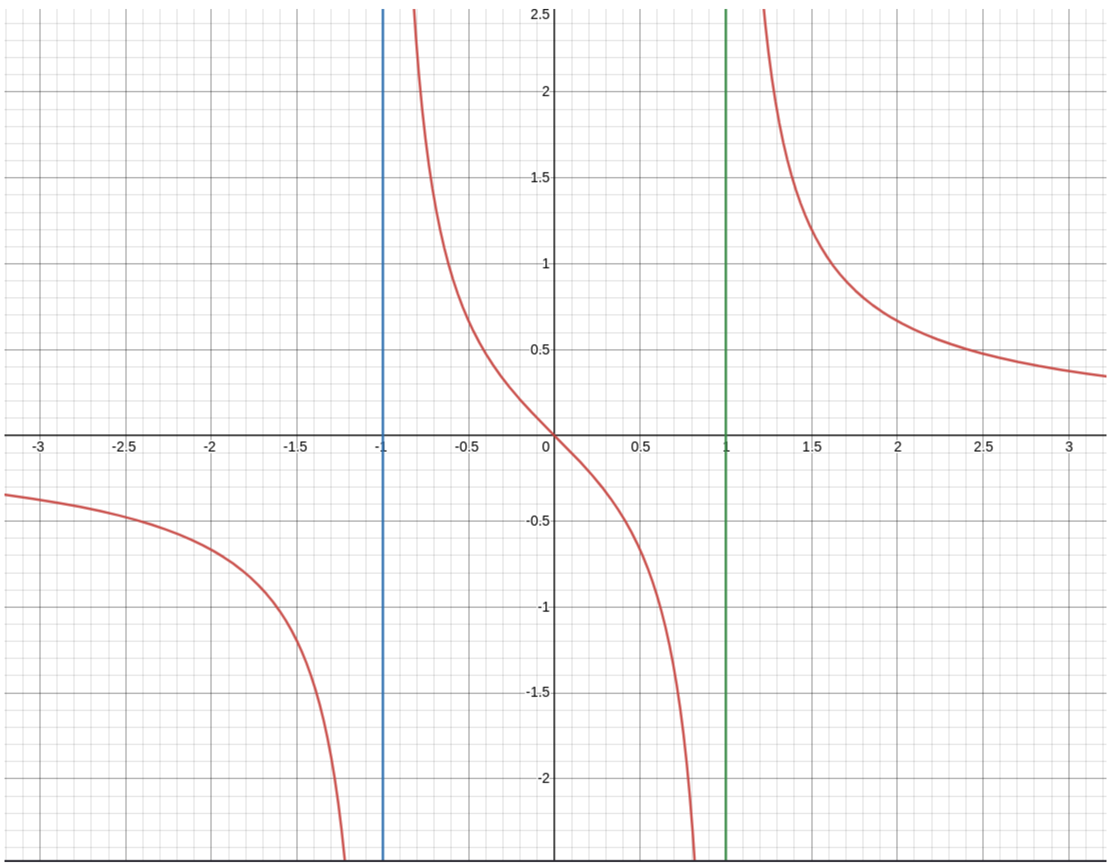
Since rational functions often have a denominator which changes as a function of x, we need to be mindful of **vertical asymptotes**. These are points where the function is **undefined** because the denominator is equal to 0. To illustrate, consider the rational function

$$f(x) = \frac{x}{x^2 - 1}$$

We need to know when the denominator equals zero so we can find the vertical asymptotes. In general, all we need to do is set the denominator equal to 0 and solve for x:

$$x^2 - 1 = 0 \implies x^2 = 1 \implies x = \pm 1$$

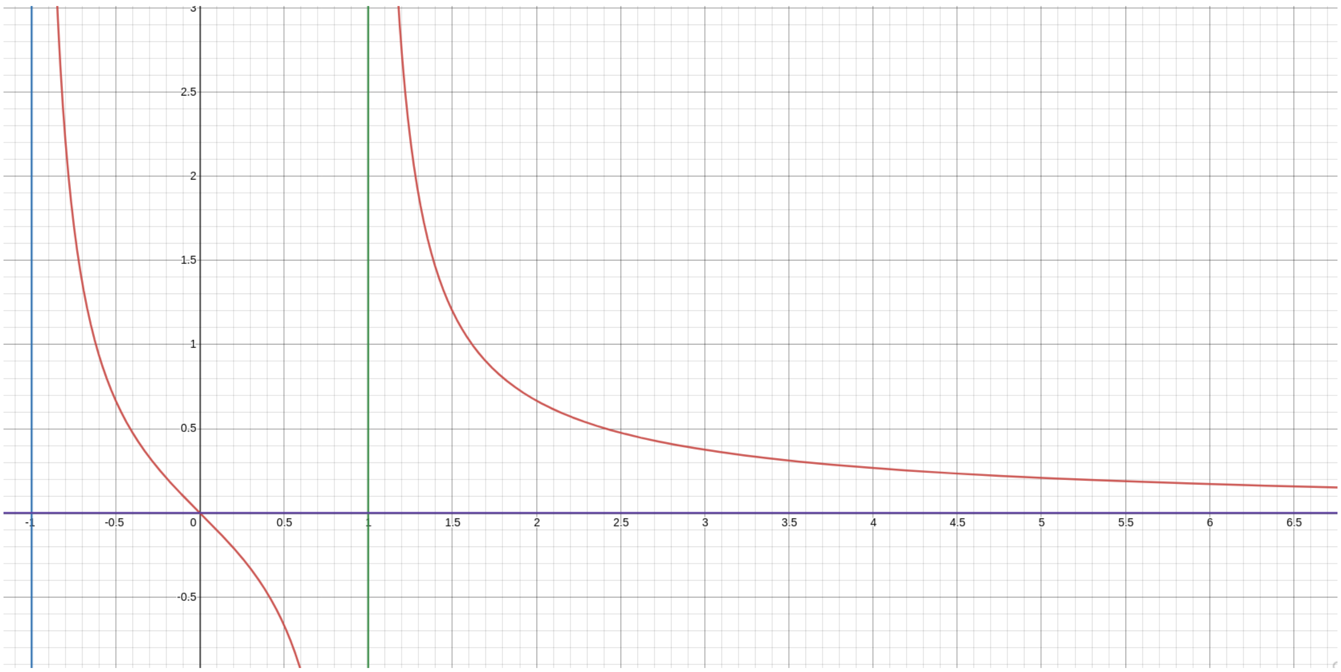
This means that the function is undefined whenever $x = 1$ or $x = -1$. If we graph $f(x)$, we can see what this looks like:



The function $f(x)$ is in red, while the vertical asymptotes are in blue and green. When we draw asymptotes by hand, we typically use dotted lines. The asymptote is like a "wall" the function cannot pass through; all it can do is get closer and closer to the wall, but never touch.

In the future, we will discuss the idea of "continuous" functions, as well as "discontinuities" on the graph, such as those caused by vertical asymptotes. For now, understand that because you need to "lift your pen off the paper" to draw the graph of $f(x)$ above, it has points where it is not continuous.

There is one more kind of asymptote we can identify on the graph of $f(x)$ called a **horizontal asymptote**:



Notice how the graph past the green vertical asymptote slowly approaches the x-axis, highlighted in purple. Since it keeps approaching, little by little, $f(x)$ will never actually equal the value of the horizontal asymptote; it will just get closer and closer. So, the graph tells us that we have a horizontal asymptote at $y = 0$.

In the future, you will learn about "limits" of functions. Horizontal asymptotes are a kind of limit. In calculus, we would write

$$\lim_{x \to \infty} f(x) = 0$$

which we read as "as x approaches infinity, the limit of $f(x)$ equals 0." All this means is that as x gets bigger and bigger, $f(x)$ grows closer and closer to $y = 0$.

We can learn a lot of useful information about different phenomena modeled by math by understanding these asymptotes. Vertical asymptotes represent "singularities" where, physically, we interpret a quantity as getting infinitely big (usually we try to avoid these because "infinity" of anything doesn't make a lot of sense in the real world). Horizontal asymptotes tell us about things like the carrying capacity of an animal population; as time goes on, the population will stabilize around the carrying capacity.

1.5 Exponential and Logarithmic Functions

1.5.1 Introduction to Exponential Functions

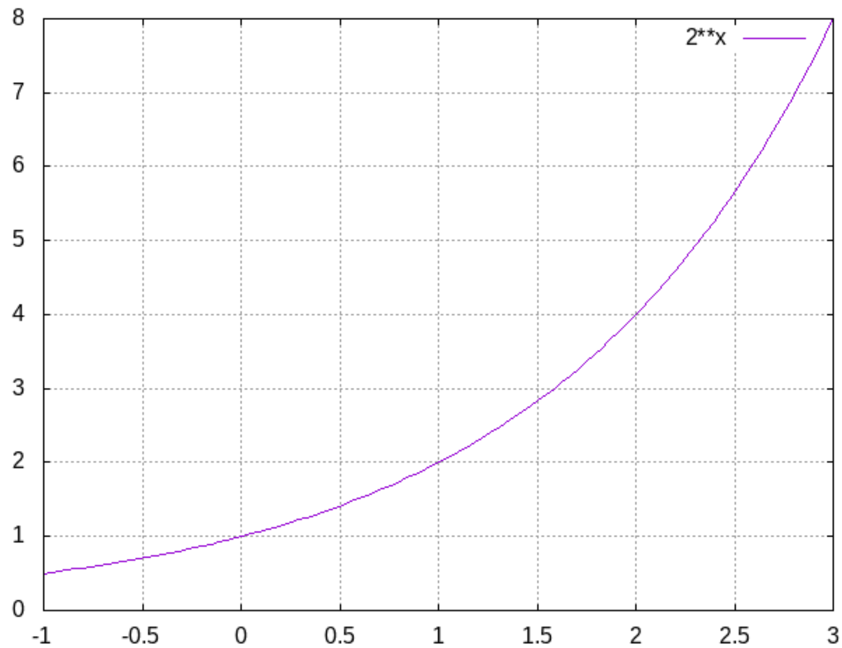
One of the simplest forms an exponential function can have is

$$f(x) = a^x$$

where a is some constant number called the **base**. A concrete example of this kind of function is

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

A typical exponential function plot looks like this:



The plot has a horizontal asymptote at $y = 0$. This makes sense because, ordinarily, it is impossible to make a positve number negative by exponentiating it. If this function were multiplied by -1, then the graph would be flipped upside down, and there would still be a horizontal asymptote at $y = 0$.

The horizontal asymptote tells us that exponential functions tend to get closer and closer to 0 whenever $x < 0$. Let's consider what happens when x gets bigger. We see that $f(0) = 2^0 = 1$; this is a feature that many exponential functions have in common. Then $f(1) = 2^1 = 2$, which is just the constant. Then, whenever x increases by 1, $f(x)$ doubles; for example, $f(1) = 2$, then $f(2) = 4$, and $f(3) = 8$. By this behavior, we see that exponential functions tend to grow very fast.

What if you have an exponential function like this?

$$g(x) = (0.5)^x$$

Let's use some exponent rules to find out. First, rewrite it as a ratio:

$$g(x) = \left(\frac{1}{2}\right)^x$$

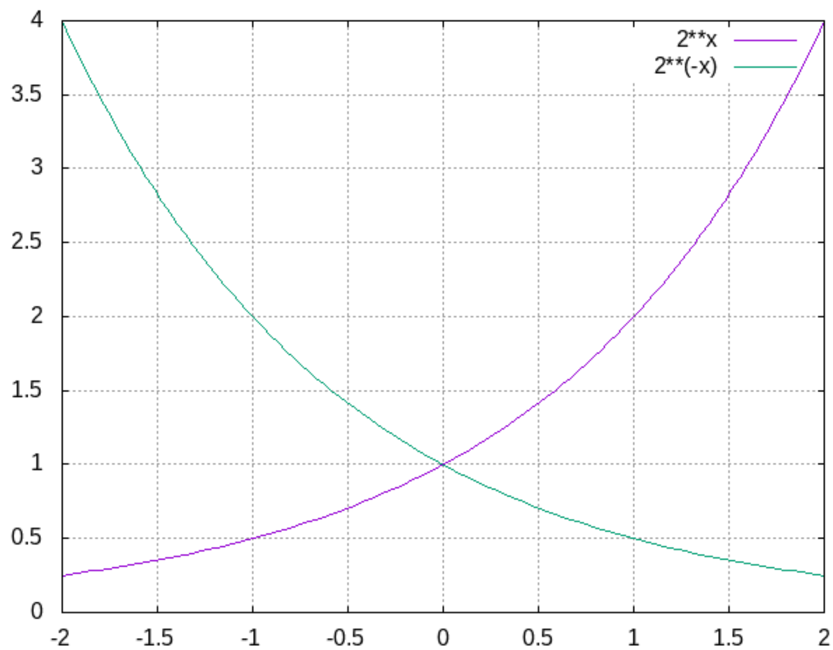
We know that 1/2 is the same thing as 2^{-1}, so

$$g(x) = (2^{-1})^x$$

Finally, use the "exponent multiplication" rule:

$$g(x) = 2^{-x}$$

The negative exponent means $g(x)$ has the opposite behavior of a normal exponential function. Whenver x gets big, $g(x)$ will approach 0; whenever x is a negative number, $g(x)$ will get large. To visualize, $g(x)$ is plotted in blue next to $f(x)$ below:



Exponential functions with a negative argument are a very common way of modeling things that decrease over time or distance.

An important constant for exponential functions is Euler's number e, which is approximately equal to 2.718, but it is actually an irrational number like π. A special function based on e is

$$\exp(x) = e^x$$

which has many special properties that make doing calculations with it much easier. The most famous is that the rate of change at every point of e^x is also equal to e^x, but there are also many subtle and interesting ways that the constant e appears throughout models of nature. In depth discussion of e is greatly aided with tools from calculus, so we will not get into that yet.

1.5.2 Introduction to Logarithms

What if we want to solve an equation like this?

$$2^x = 7$$

For this purpose, we introduce logarithms, the inverse function of exponentials. Out of all functions studied in an algebra or pre-calculus course, logarithms (usually shortened to "logs") are some of the most mysterious, so we will explore some examples first:

$$\log_2(8) = 3$$

$$\log_{10}(100) = 2$$

$$\log_3\left(\frac{1}{9}\right) = -2$$

Do you see the pattern? The logs above help us solve equations like these:

$$2^x = 8$$

$$10^x = 100$$

$$3^x = \frac{1}{9}$$

In words,

2 raised to what power gives me 8?

10 raised to what power gives me 100?

3 raised to what power gives me 1/9?

The **base** of the log tells you which number is being exponentiated. When we write $\log(x)$ without a base subscript, it is typically assumed to be base 10, the same as $\log_{10}(x)$. If the base of a log is Euler's number e, then we notate it $\ln(x)$, which stands for the "natural logarithm of x."

Let's try using logs to manipulate an equation we don't already know the answer to:

$$5^x = 91$$

First, we take the log of both sides:

$$\log(5^x) = \log(91)$$

One of the most helpful properties of logs is that $\log(a^x) = x\log(a)$. In other words, we can pull exponents out front. Doing that, we obtain

$$x\log(5) = \log(91)$$

Finally,

$$x = \frac{\log(91)}{\log(5)}$$

Unfortunately, it is not easy to calculate most logs by mental math or by hand; you need to use a calculator. Most calculators and computer algebra systems have built-in functions for calculating $\log_{10}(x)$ and $\ln(x)$; be careful if you are in another base.

We are now going to start over and solve for x a slightly different way. Instead of log, we will begin with $\log_5$ on both sides:

$$\log_5(5^x) = \log_5(91) \implies x\log_5(5) = \log_5(91)$$

We know how to solve $\log_5(5)$. It is asking "5 raised to what power gives me 5?" This is nothing but 1, which we can plug in to obtain

$$x(1) = x = \log_5(91)$$

Since $x = \dfrac{\log(91)}{\log(5)}$ and $x = \log_5(91)$, we can see that

$$\frac{\log(91)}{\log(5)} = \log_5(91)$$

This gives us a formula for converting between different log bases. In general,

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

Let's discover one more very helpful property of logs. Consider the following equations:

$$x = \log_c(a)$$

$$y = \log_c(b)$$

where $x, y \in \mathbb{R}$ and $c > 0$. Then we know how to re-write both equations:

$$c^x = a$$

$$c^y = b$$

Take the product of c^x and c^y to obtain

$$c^x c^y = ab$$

Use your rule for "exponent addition:"

$$c^{x+y} = ab$$

Take $\log_c$ of both sides:

$$\log_c(c^{x+y}) = \log_c(ab)$$

We know how to solve $\log_c(c^{x+y})$:

$$x + y = \log_c(ab)$$

Remember that $x = \log_c(a)$ and $y = \log_c(b)$. In general, for log of any base c, we now have the rule for log addition:

$$\log_c(a) + \log_c(b) = \log_c(a \cdot b)$$

The rule for log subtraction follows when either a or b above has a negative exponent:

$$\log_c(a) - \log_c(b) = \log_c\left(\frac{a}{b}\right)$$

1.5.3 Applications of Exponents and Logarithms

Exponential and logarithmic functions are commonly used to model growth and decay. For example, we can model the progression of chemical reactions like this

$$\mathrm{X} \longrightarrow \mathrm{Y}$$

where some chemical X gets converted into another chemical Y. If we want to know how much of chemical X is left ater a certain amount of time, we use the following equation:

$$\mathrm{X}(t) = \mathrm{X}_0 \cdot e^{-kt}$$

In this model, X_0 is the starting amount of X, t is the time since the reaction started, and k is called the **rate constant**. In this example, X_0 determines how big the graph of $\mathrm{X}(t)$ starts, while k changes how fast $\mathrm{X}(t)$ decreases.

We can use this equation to find the **half-life** $t_{1/2}$ of X, which is the amount of time it takes for X to decrease to half its starting amount. All we need to do is figure out what $t_{1/2}$ is when $\mathrm{X}(t_{1/2}) = \frac{1}{2}\mathrm{X}_0$. So,

$$\frac{1}{2}\mathrm{X}_0 = \mathrm{X}_0 e^{-kt_{1/2}}$$

Divide both sides by X_0 to get

$$\frac{1}{2} = e^{-kt_{1/2}}$$

Now take the natural log of both sides:

$$\ln(1/2) = -kt_{1/2}$$

Finally, divide to get $t_{1/2}$.

$$t_{1/2} = -\frac{\ln(1/2)}{k}$$

This is one version of a complete answer, but we can simplify in one last step using the log rule for subtraction:

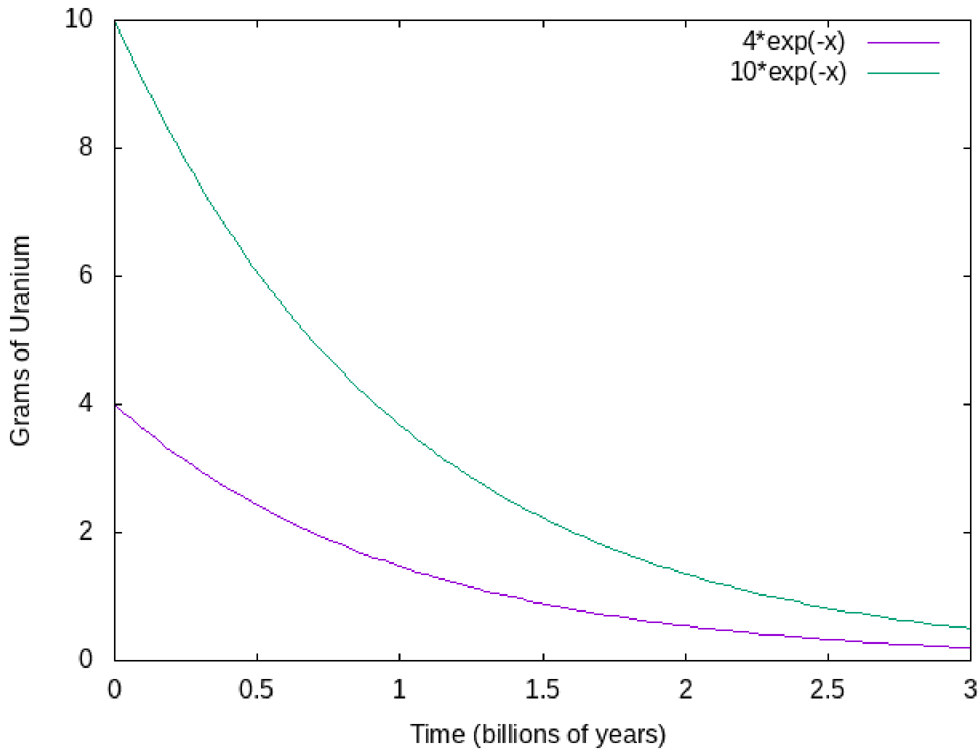
$$t_{1/2} = -\frac{\ln(1) - \ln(2)}{k}$$

We know $\ln(1) = 0$ because anything to the 0th power is equal to 1. So,

$$t_{1/2} = \frac{\ln(2)}{k}$$

Interestingly, this means that the half-life of something is the same no matter how much you start with. The only thing that changes it is the rate constant k. If uranium decayed in a nuclear reactor the way this mathematical model describes, then it would take the same amount of time for 100 g of uranium to decay to 50 g as it takes 2 g to decay to 1 g.

We can plot $X(t)$ to visualize the half life. On the graph below, there are two starting amounts: 4 g of uranium and 10 g of uranium. Notice that both of the curves hit half the starting amount at the same time.



In reality, the half lfie of uranium is around 4.6 billion years. We could use this to figure out what the rate constant k should be:

$$4.6 = \frac{\ln(2)}{k} \implies k = \frac{\ln(2)}{4.6} \approx 0.151$$

If this mathematical model accurately describes uranium decay, then our equation would be

$$U(t) = U_0 \cdot e^{-0.151t}$$

We can write most exponential functions in terms of logs if we want. Just take the log of both sides of the equation:

$$\ln(U(t)) = \ln(U_0 \cdot e^{-0.151t})$$

Now we use the log rule for addition to break it up:

$$\ln(U(t)) = \ln(U_0) + \ln(e^{-0.151t})$$

Since $\ln(x)$ is the inverse function of e^x, we conclude

$$\ln(U(t)) = \ln(U_0) - 0.151t$$

We can re-exponentiate everything by a similar process.

Exponents can be used to calculate compound interest by the formula

$$A(t) = P\left(1 + \frac{r}{n}\right)^{nt}$$

where $A(t)$ is the "account" (amount owed in a loan), P is the "principal" of the loan (how much you borrowed originally), r is the interest rate, and n is the number of times interest compounds per unit of time. To understand the equation, let's apply it to student loans for college.

In college, tuition costs \$1000 per credit hour on average, and most students typically take a "full-time" load of 16 credit hours per semester (32 per academic year). After scholarships and awards proportional to high school GPA and SAT scores, many students graduate around $\$30,000.00$ in debt (or more). For a given student, assume that all of that debt is subject to an interest rate of 3.73% APR (annual percentage rate).

A common misconception is that interest only compounds yearly; in fact, for student loans and credit cards, interest compounds daily, even if the rate is given as an APR. We can use our equation for compound interest to investigate the difference between daily and yearly compounding interest:

$$A_{yearly}(t) = (\$30,000)\left(1 + \frac{0.0373}{1}\right)^{(1)t} = \$30,000\,(1.0373)^t$$

$$A_{daily}(t) = (\$30,000)\left(1 + \frac{0.0373}{365}\right)^{365t} = \$30,000\,(1.000102)^{365t}$$

Observe: the number being exponentiated in A_{yearly} is much larger than in A_{daily}, basically indicating how much bigger the account will get every time the interest compounds. However, notice that A_{daily} will be compounded (exponentiated) 365 times more than A_{yearly}. So which one is more expensive? After $t = 1$ year, $A_{yearly}(1) = \$31,119$ while $A_{daily}(1) = \$31,140$, so it is more expensive to take a loan where the interest compounds daily. What about after 10 years?

$$A_{yearly}(10) = \$43,267$$

$$A_{daily}(10) = \$43,531$$

What would happen if you paid for your college education on a credit card with a 27.99% APR?

$$A_{CC}(t) = \$30,000\left(1 + \frac{0.2799}{365}\right)^{365t}$$

Evaluating this function for $t = 1$ and $t = 10$ years yields

$$A_{CC}(1) = \$39,685$$

$$A_{CC}(10) = \$492,588$$

In actuality, many credit cards compound daily but divide the APR monthly, so we might come up with an equation like this to describe the account balance:

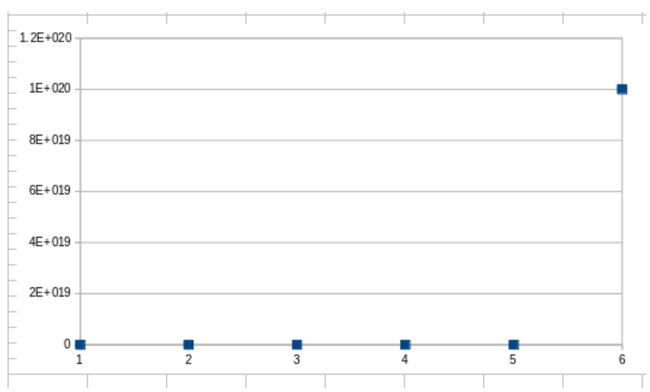
$$A(t) = P\left(1 + \frac{\text{APR}/12}{31}\right)^{31t}$$

where t is in units of months.

Logs are helpful in visualizing data that changes very quickly. Consider the following table of data:

Time (minutes)	Number of bacteria
1	1
2	1×10^2
3	1×10^{12}
4	1×10^{15}
5	1×10^{16}
6	1×10^{20}

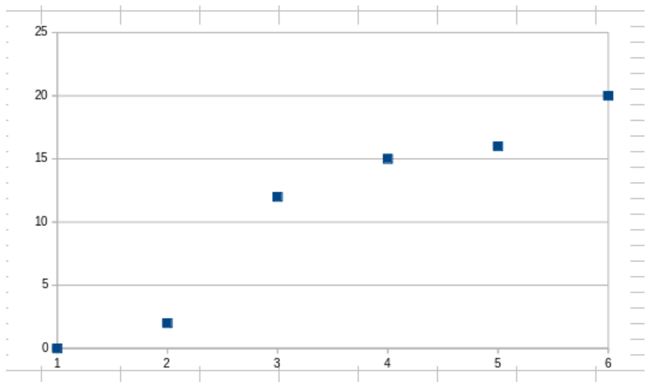
The number of bacteria is growing so quickly that it would be ineffective to plot. See here:



All the points are so small compared to the largest point that they all look the same. Let's try taking the $\log_{10}(x)$ of each large data point:

Time (minutes)	Log of Number of bacteria
1	0
2	2
3	12
4	15
5	16
6	20

This is called a "log scale." Now when we plot it,



We could fit a reasonably good linear regression line on this data, unlike the unmodified plot.

1.5.4 Inequalities

According to some mathematicians, “inequality” is one of the most fascinating concepts in the entire field. We are used to finding when things are equal; but what about when they are not the same? What makes two things different?

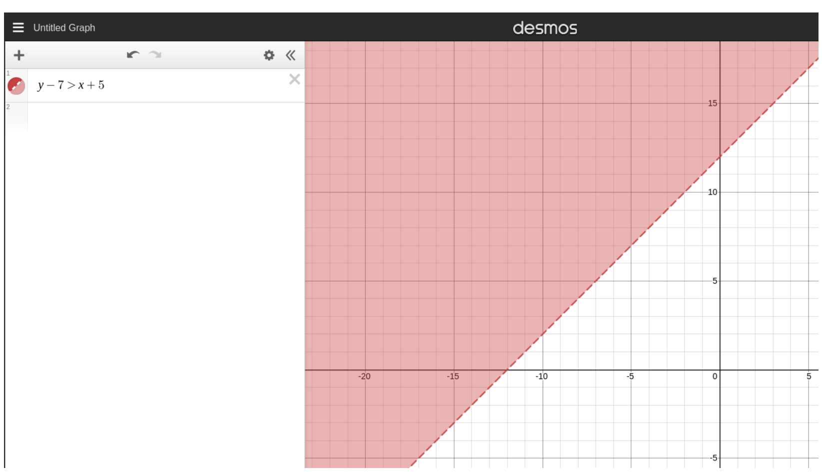
In algebra, we worry about when specific numbers or functions are not the same. To start, let’s consider the following inequality:

$$x + 5 < y - 7$$

You are most likely used to solving inequalities like these already. Just isolate either x or y on one side:

$$x + 5 < y - 7 \implies x + 12 < y$$

Recall that we would graph this by drawing a dotted line version of $y = x + 12$; then we would shade above that line to indicate all the possible solutions:



We can do the same thing for a broad range of functions. To illustrate, take $f(x) = x + 5$ and $g(y) = y - 7$. Then our first example becomes

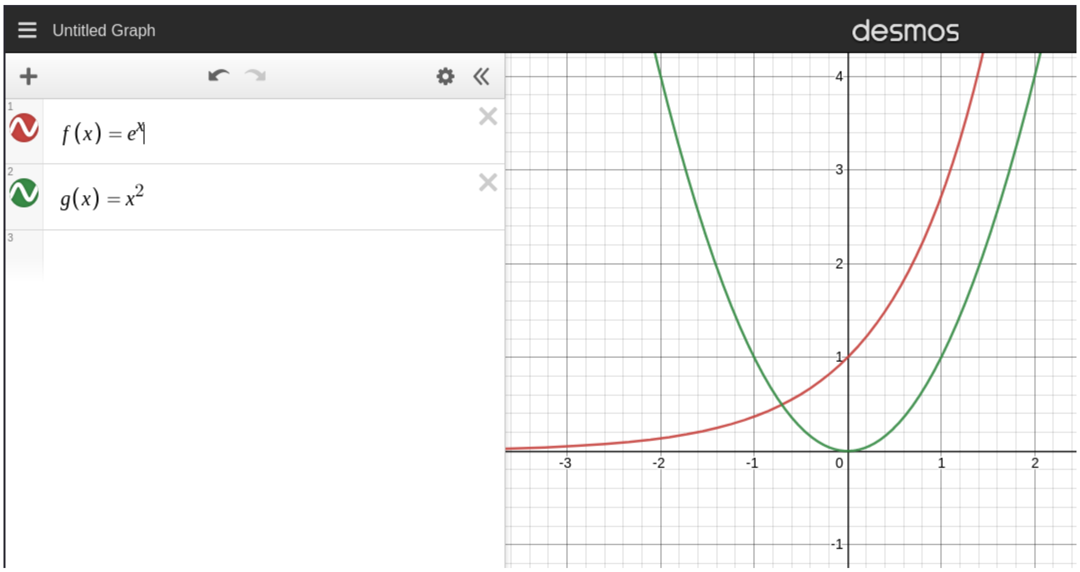
$$f(x) < g(y)$$

The same graph represents solutions to the inequality which make it true that $f(x)$ is less than $g(y)$. Let's apply this idea to some of the functions we studied previously.

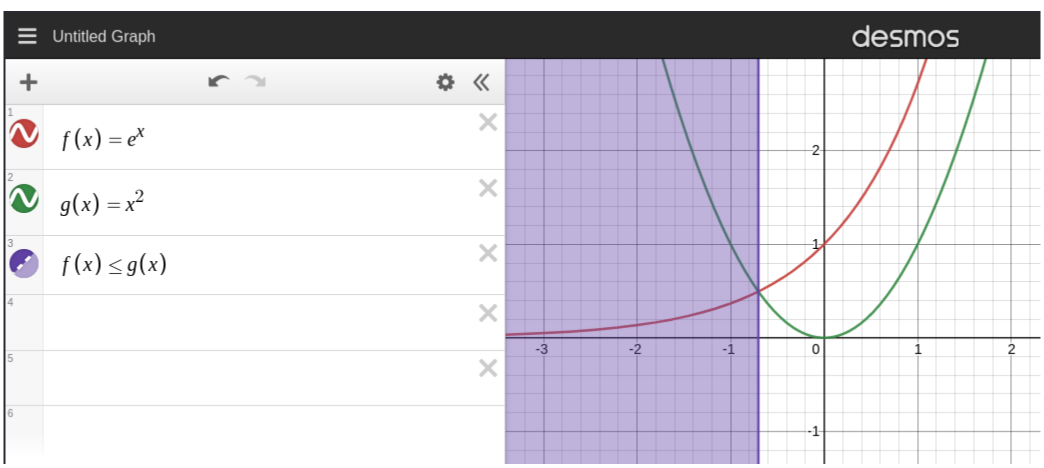
If $f(x) = e^x$ and $g(x) = x^2$, then we could set up an inequality

$$f(x) \leq g(x)$$

What does this mean, and how do we solve for it? First, let's plot both graphs together on the same coordinate plane:



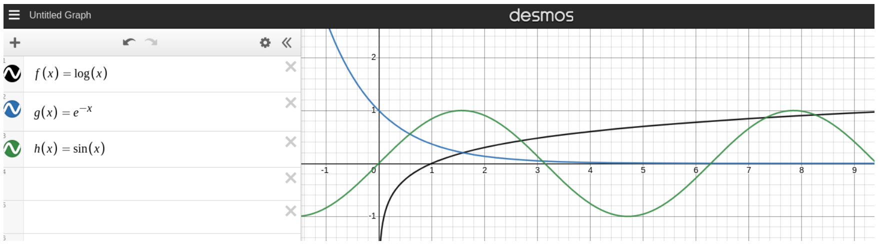
We can visually inspect where $f(x)$ is less than or equal to $g(x)$ this way. Observe that $f(x) = g(x)$ at $x = -1$ and all the $x < -1$ fall below the graph of $g(x)$. Thus, we get a plot like this for the inequality:



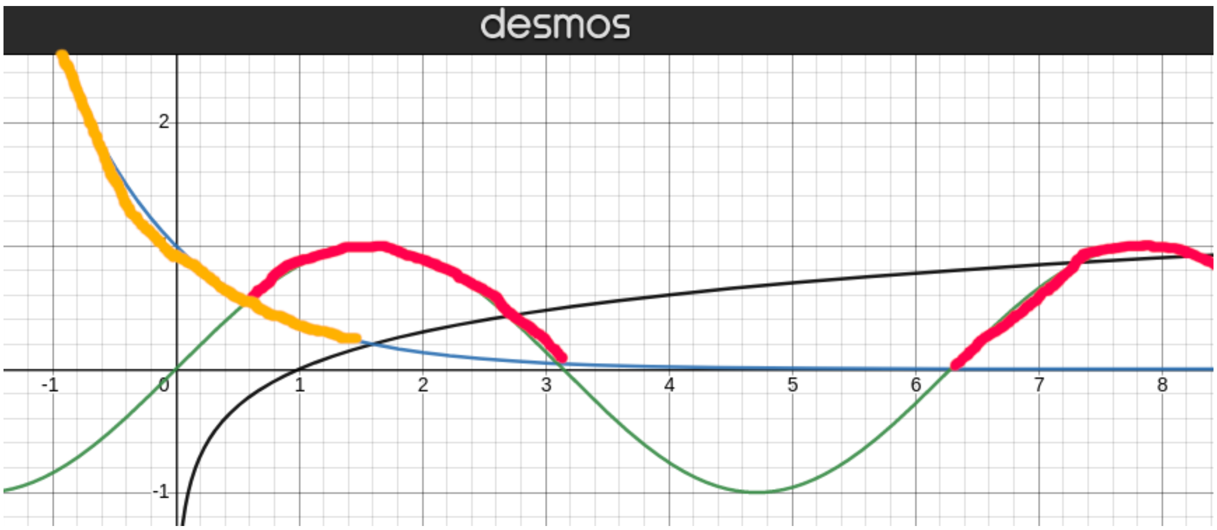
This can be extended to three or more functions. Let $f(x) = \log(x)$, $g(x) = e^{-x}$ and $h(x) = \sin(x)$. Our inequality will be

$$f(x) < g(x) < h(x)$$

As before, we should plot all three on the same graph:



We can tackle this in two parts: first, find where $f(x) < g(x)$, then where $g(x) < h(x)$. When we know both, we can find where the two inequalities overlap. This is as easy as tracing $h(x)$ where it is bgger than $g(x)$ and tracing where $g(x)$ is bigger than $f(x)$:



There is only a small interval where the graph is traced in both red and orange, which is approximately $[0.6, 1.6]$ in interval notation. All those values of x are our solutions.

While graphic solutions to inequalities are nice to look at and easy to understand, it is still nice to solve inequalities algebraically when possible. Any **separable** equation (i.e. an equation where either x or y can be written alone on one side) could have a nicely solvable inequality conterpart.

1.6 Basic Linear Algebra

1.6.1 Linear Equations and Transformations

Consider some function $f\colon \mathbb{R} \to \mathbb{R}$. We say that f is **linear** if it satisfies the following two properties:

Additive	$f(x+y) = f(x)+f(y)$
Homogenous	$f(ax) = af(x)$

Linear functions are sometimes called **linear transformations**. So, what are some examples of linear functions? The easiest example of a linear function is the identity:

$$f(x) = x$$

We can check this. Is it additive?

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

Yes, it is additive. Is it homogenous?

$$f(ax) = ax = af(x)$$

It is also homogenous. Therefore, we have proved that $f(x) = x$ is a linear transformation! What about the equation for a straight line? Is it linear?

$$f(x) = mx + b$$

Is it additive?

$$f(x+y) = m(x+y) + b = (mx+b) + y = f(x) + y$$

It is not additive. We know already that it can't possibly be linear, but let's check to see if it is at least homogenous:

$$f(ax) = m(ax) + b = a(mx) + b \neq a(mx+b)$$

It is not even homogenous. So, technically speaking, the equation for a straight line isn't linear if the y-intercept is anything but $b = 0$.

Aside from the example with the straight line, most of your intuition about what is "linear" and what isn't is still correct. For example, $f(x) = x^2 + 2$, $g(x) = \log(x)$, and $h(x) = \cos(x)$ are all nonlinear.

Linear equations, which do not need to be written as a function, are not so strict; rearrangement of $y = mx + b$ gives us a perfectly valid linear equation:

$$y - mx - b = 0$$

In general, a linear equation with two variables x and y will have the form

$$ay + bx + c = 0$$

The mathematical field of **linear algebra** studies the properties of linear functions and equations, the ways we can solve them, and the consequences of combining different kinds of linear functions.

1.6.2 Systems of Linear Equations

Sally has x quarters and y dimes and has a total of \$2.50. We also know that Sally has a total of 13 coins. We can set up two equations based on this information:

$$(\$0.25)x + (\$0.10)y = \$2.50$$

$$x \text{ quarters} + y \text{ dimes} = 13 \text{ coins}$$

This is called a system of linear equations. As you have most likely seen before, we can solve for either x or y and the plug what we get back into the other equation:

$$x = 13 - y \implies 0.25(13 - y) + 0.10y = 2.50 \implies y = 5$$

Now that we know y, we just need to solve for x:

$$x + (5) = 13 \implies x = 8$$

So Sally has 8 quarters and 5 dimes.

Linear equations can have as many variables as we want. However, if we want to solve for all the variables, we need the same number of equations to narrow it down to a unique answer. Here's a 3-D example:

$$3x + 2y + z = 0$$

$$5x - y + z = 2$$

$$x - 3y + 8z = 1$$

When we have so many variables, it is not easy to use the previous method of solving for one and then plugging into the others. Before doing that, we should try to add/subtract equations from each other.

Observe that if we take the 3rd equation and multiply both sides by -5, we get this:

$$-5(x - 3y + 8z) = -5(1) \implies -5x + 15y - 40z = -5$$

If we add this to the 2nd equation, something good happens:

$$(5x - y + z) + (-5x + 15y - 40z) = (2) + (-5) \implies 14y - 39z = -3$$

We have successfully eliminated one variable from the equation! Let's try to do this to some of the others:

$$3(3x + 2y + z) + 2(x - 3y + 8z) = 0 + 2(1) \implies 11x + 19z = 2$$

So far, we have managed to reduce our system of equations down to this:

$$11x + 19z = 2$$

$$14y - 39z = -3$$

$$x - 3y + 8z = 1$$

Before trying to solve for x, y, and z, we should try to eliminate x from the 3rd equation.

$$-11(x - 3y + 8z) + (11x + 19z) = -11(1) + 2 \implies 33y - 69z = -9$$

So we now have

$$11x + 19z = 2$$

$$14y - 39z = -3$$

$$33y - 69z = -9$$

In principal, this should be very straightforward to solve now: solve for y in terms of z, then plug in somewhere else (2nd equation); this will tell you what z is. Plug z into the top equation to get x, and plug z into the bottom equation to get y. In practice, this is a pain in the neck because it would involve lots of arithmetic with rational numbers.

Generally, a system of linear equations has the form

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = c_n$$

where each x_j is a different variable (like x, y, or z, except we distinguish them with little subscripts after we run out of letters of the alphabet) and each a_{ij} is a coefficient. One way to represent such a system of linear equations is with a **matrix**, which is (basically) a grid of numbers with n rows and m columns (in other words, an $n \times m$ matrix). For a general system of linear equations, we might write a matrix $\mathbf{A}$ as follows:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Our matrix $\mathbf{A}$ is an $n \times n$ matrix because it has n rows and n columns. Another way of saying this is that $\mathbf{A}$ is a *square matrix*. All of the values in the first column correspond to coefficients on the variable x_1; the second column is x_2; and so on. For reasons we will explore later, the variables are not written inside of $\mathbf{A}$.

In the next section, we will learn how to do basic matrix arithmetic and some of the algebraic properties of matrices. This will give us the tools we need to solve linear equations with matrices.

1.6.3 Matrix Algebra

Let **A** and **B** be matrices defined by

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix}$$

Algebraically, what can we do with them? First, we can add matrices:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 1+6 & 2+7 \\ 3+8 & 4+9 \end{bmatrix} = \begin{bmatrix} 7 & 9 \\ 11 & 13 \end{bmatrix}$$

This makes sense. If I have $x + 2y = 0$ and $6x + 7y = 0$ and add them together, I would get $7x + 9y = 0$, just like the matrix above. Subtraction works exactly the same way. Note that the two matrices need to have the same dimensions to be added together. This time, both **A** and **B** are 2×2 square matrices, so we are allowed to add them.

Let's define two new matrices **C** and **D**:

$$\mathbf{C} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} 6 & 7 \end{bmatrix}$$

Observe that **C** is a 3×1 matrix because it has 3 rows and 1 column; on the other hand, **D** is a 1×2 matrix because it has 1 row and 2 columns. Matrices that only have 1 row or column are usually called **vectors**. We can't add or subtract **C** and **D** since they have different dimensions. However, we are allowed to multiply them:

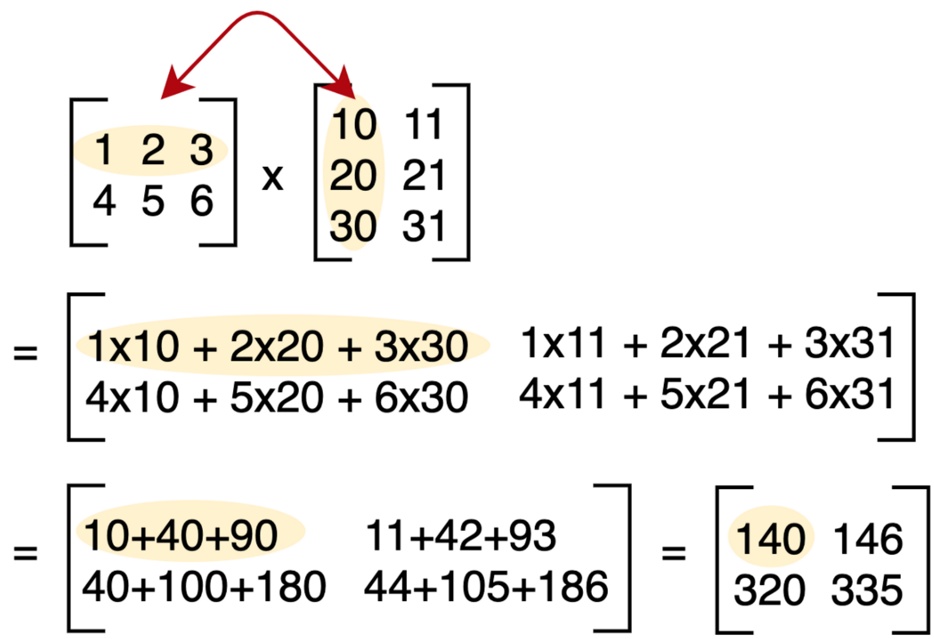
$$\mathbf{C} \cdot \mathbf{D} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 6 & 7 \end{bmatrix} = \begin{bmatrix} (1)(6) & (1)(7) \\ (3)(6) & (3)(7) \\ (0)(6) & (0)(7) \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 18 & 21 \\ 0 & 0 \end{bmatrix}$$

Before explaining the process of matrix multiplication, observe that something interesting happened: we started with the product of a 3×1 matrix with a 1×2 matrix and ended up with a 3×2 matrix. In general, to successfully multiply matrices, they need to "match" dimensions. That is, you can only multiply an $n \times k$ matrix with a $k \times m$ matrix, where n and m can be any two natural numbers but k has to be the same. The matrix you get from multiplying an $n \times k$ with a $k \times m$ will always be an $n \times m$ matrix.

The actual process of multiplying matrices, even though I demonstrated an example above, is likely a little mysterious. Mathematically speaking, suppose we have two matrices **A** and **B** which multiply together to make **C**. If we want to calculate c_{ij}, the number in the i-th row and j-th column of **C**, then we need to use the following formula:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

In words, we hold the row constant for **A** and change column; for **B**, we fix the column and go down to each row. Refer to the illustration on the next page.



For good measure, one more simple example:

$$\begin{bmatrix}1 & 0\\2 & 1\end{bmatrix}\cdot\begin{bmatrix}3 & 0\\4 & 1\end{bmatrix}=\begin{bmatrix}(1)(3)+(0)(4) & (1)(0)+(0)(1)\\(2)(3)+(1)(4) & (2)(0)+(1)(1)\end{bmatrix}=\begin{bmatrix}3 & 0\\10 & 1\end{bmatrix}$$

When we do normal multiplication, we typically enjoy the commutative property, which lets us say $a \cdot b = b \cdot a$. Do we have a commutative property for matrix multiplication? Let's reverse the order of the matrices being multiplied in the last example.

$$\begin{bmatrix}3 & 0\\4 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0\\2 & 1\end{bmatrix}=\begin{bmatrix}(3)(1)+(0)(2) & (3)(0)+(0)(1)\\(4)(1)+(1)(2) & (4)(0)+(1)(1)\end{bmatrix}=\begin{bmatrix}3 & 0\\6 & 1\end{bmatrix}$$

We didn't get the same result, which proves that matrix multiplication doesn't have to be commutative.

Matrices *do* have the distributive property of multiplication. So,

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$$

and

$$(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$$

Matrix multiplication is also associative:

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$

Proof of these properties follows from the equation for matrix multiplication, but that is beyond the scope of this section.

Normal numbers are called **scalars** in the context of matrices and vectors. When we do "scalar multiplication," that means we do something like this:

$$7 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 14 \end{bmatrix}$$

Just multiply the scalar by every number inside the matrix. Scalars distribute over matrix addition:

$$c\,(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$$

As long as we are working with real numbers, it will also generally be the case that

$$(c\mathbf{A})\mathbf{B} = c(\mathbf{AB}) = (\mathbf{AB})c = \mathbf{A}(\mathbf{B}c)$$

but this may not be true in more abstract settings.

Finally, we are allowed to do "elementary row operations" to matrices, which enables us to transform matrices without changing certain key properties about them. We can switch rows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

We can multiply rows by a scalar:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} 5a & 5b \\ c & d \end{bmatrix}$$

We can add and subtract rows from each other:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} a + c & b + d \\ c & d \end{bmatrix}$$

There are a number of "nice" matrices we might want to simplify down to. First, we have the identity matrix:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In actuality, there are many identity matrices; it just has to be $n \times n$ with ones on the diagonal and zeroes elsewhere. Anything you multiply by the correct identity matrix will be returned unchanged (just like multiplying by 1).

Otherwise, we have the **Row-Echelon Form** of matrices:

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

This form makes it very easy to do things like solve linear equations since the bottom row just represents something like $z =$ constant, which can be used to solve the entire system.

1.6.4 What are Matrices?

Consider a linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (2x, 3y)$. Basically, an ordered pair goes in, and another ordered pair comes out. Last section, we learned about the identity matrix I, which is like the number 1 for matrix multiplication. What happens if we plug $(1, 0)$, the first column of the 2×2 identity matrix, into $T(x, y)$?

$$T(1, 0) = (2, 0)$$

How about the second column of the identity matrix, $(0, 1)$?

$$T(0, 1) = (0, 3)$$

We can turn $(2, 0)$ and $(0, 3)$ into the columns of a new matrix:

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Observe what happens when we take the following matrix product:

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ 3y \end{bmatrix}$$

The product is the same as the regular function! We have discovered a way we can express the function $T(x, y)$ in terms of matrix multiplication. This means we can write

$$T(x, y) = \mathbf{Ax}$$

where $\mathbf{A}$ is the matrix we discovered above and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$.

Interestingly, every linear transformation can be written as a matrix product; every matrix also corresponds to a linear transformation! In fact, this is why matrix multiplication has to follow the (seemingly bizarre) rule that you can only multiply an $n \times k$ matrix with a $k \times m$ matrix. We will illustrate with an example.

Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ and $S\colon \mathbb{R}^2 \to \mathbb{R}$ be linear functions. We are allowed to make the composite function $S \circ T$ because the codomain of T is the same as the domain of S.

Let's define $T(x, y, z) = (x + y, z + y)$ and $S(a, b) = a + 2b$. Then the composite function $S \circ T$ is

$$S \circ T = (x + y) + 2(z + y)$$

What happens if we plug in $(1, 0, 0)$ to $S \circ T$?

$$S \circ T(1, 0, 0) = 1$$

How about $(0, 1, 0)$?

$$S \circ T(0, 1, 0) = 1 + 2 = 3$$

Finally, $(0,0,1)$?

$$S \circ T(0,0,1) = 2$$

We can turn this into a matrix:

$$S \circ T = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + 3y + 2z \end{bmatrix} = (x+y) + 2(z+y)$$

By multiplying a 1×3 and a 3×1 matrix together, we just get a 1×1 matrix (in other words, a normal number), and it matches what we found from doing normal function composition!

Now what happens if we make matrices out of regular $S(a,b)$ and $T(x,y,z)$? Skipping some of the work (all you do is plug in the columns of the identity matrix over and over), here is the result:

$$S(a,b) = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$T(x,y,z) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Observe what happens when we take the following matrix product:

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}$$

It is the same matrix that represents $S \circ T$! This shows concretely that multiplying matrices is intimately related to function composition. Just like how it would be nonsense to plug an imaginary number into a function that requires an integer, it is nonsense to multiply matrices unless the dimensions match.

1.6.5 Applications of Linear Equations

Sally has x quarters and y dimes and has a total of \$2.50. We also know that Sally has a total of 13 coins. How many quarters and dimes does Sally have?

We can set this problem up with matrices just like a system of linear equations. Generally, a system of linear equations will have the form

$$\mathbf{Ax} = \mathbf{b}$$

where **A** is the coefficient matrix, **x** contains the variables, and **b** are the solutions to the equations. So, for our problem,

$$\begin{bmatrix} 0.25 & 0.10 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2.50 \\ 13 \end{bmatrix}$$

If we multiplied this out, we would get

$$\begin{bmatrix} 0.25x & 0.10y \\ x & y \end{bmatrix} = \begin{bmatrix} 2.50 \\ 13 \end{bmatrix}$$

but we don't need to; all we need to do is perform elementary row operations on the coefficient matrix. However, we also need to "augment" it to include the total amount of money and the total number of coins in our consideration:

$$\left[\begin{array}{cc|c} 0.25 & 0.10 & 2.50 \\ 1 & 1 & 13 \end{array}\right]$$

The line is to remind you that \$2.50 and 13 are constants, while everything else is a coefficient on a variable. Now we just need to perform EROs to get the matrix into Row-Echelon form:

$$\left[\begin{array}{cc|c} 0.25 & 0.10 & 2.50 \\ 1 & 1 & 13 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0.40 & 10 \\ 1 & 1 & 13 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0.40 & 10 \\ 0 & 0.6 & 3 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0.40 & 10 \\ 0 & 6 & 30 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0.40 & 10 \\ 0 & 1 & 5 \end{array}\right]$$

This tells us that $y = 5$. If we wanted, we could stop here and just plug $y = 5$ into the other equation. However, we can also solve the entire problem with matrices by getting to **Reduced** Row-Echelon form:

$$\left[\begin{array}{cc|c} 1 & 0.40 & 10 \\ 0 & 1 & 5 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & 8 \\ 0 & 1 & 5 \end{array}\right]$$

The main chunk of the matrix looks like the identity matrix; this is the hallmark of the Reduced Row-Echelon form. We also have both solutions: $x = 8$ and $y = 5$.

Chapter 2

Basic Probability and Statistics

2.1 Introduction

(planned; statistics a "mathematical science", probability a field of mathematics; descriptive and inferential statistics)

2.2 Summary Statistics

(mean, median, mode, standard deviation, quartiles, percentiles)

2.3 Probability Theory

(planned; Bayes' theorem, counting principle, probability rules, conditional probability, probability mass and density functions)

2.4 The Normal Distribution

(planned; Z-scores, using Z-tables to find probabilities

2.5 Sampling and Estimation

Simple random sampling Point estimation and properties of estimators Confidence intervals Sample size determination

2.6 Hypothesis Testing

(planned; Null and alternative hypotheses, type I and Type II errors, test statistics and p-values, one-sample and two-sample tests)

2.7 Regression

(planned; Linear regression, multiple regression; quadratic, cubic, exponential, and logarithmic regression models)

Chapter 3

Euclidean Geometry

3.1 Introduction

Geometry is the study of shapes; specifically, ideas like distance, area, volume, position, and size might be of concern to someone studying geometry. We know from graphing functions that there are no more than three ways we can (easily) visualize shapes:

- In 1-D, the number line: limited to only points, line segments, rays, and just one line.
- In 2-D, the usual coordinate plane: circles, polygons, points given by an ordered pair (x, y), infinitely many lines, one plane, and more.
- In 3-D: spheres, cubes, infinitely many planes, infinitely many lines, and more.

In the following sections, careful definitions of concepts like a "line" will be given.

Euclidean geometry refers to the geometric ideas developed by Euclid, who developed his ideas around 300 B.C.E. in Greece. Euclid's geometry book *The Elements* has been used to teach and learn geometry for over 2000 years. The ideas of Euclidean geometry pertain to "plane" or "planar" geometry (done in 2-D, as if on a piece of paper) and "solid" geometry (3-D shapes).

Non-Euclidean geometry was developed much later to account for geometry on curved surfaces, such as the surface of a globe. In addition to applications in aviation, non-Euclidean geometry is applied in Einstein's theory of general relativity, which asserts that "spacetime" (a 4-D surface made from 3 spatial dimensions plus 1 time dimension) is actually curved, not flat. While interesting, this is beyond the scope of the exposition to follow.

Since the time of Euclid, various new ways have been discovered to reach the conclusions of planar and solid geometry. In contrast with Euclid, who relied much on "ruler and compass" constructions of shapes, it is now possible to reach various geometric conclusions, for instance, with calculus in an approach called Analytical Geometry. In the exposition to follow, we will use whatever method of proof lends itself best to understanding the idea.

3.2 Logic, Reasoning, and Proof

(redoing this entire section)

3.2.1 Introduction to Propositional Logic

A **proposition** can be thought of as a statement that has a **truth value**; that is, the proposition is either **true** or **false**. The field of **propostional logic** uses rules to determine truth values and draw new inferences from combinations of propositions. (planned - definition, statements, truth value, arguments)

3.2.2 Logical Operators

(and or not implication bicondictional)

3.2.3 Rules of Inference

(modus ponens, modus tolens, etc)

3.2.4 Examples

(planned: will likely include example proofs like "sum of two even numbers is an even number")

3.3 Geometry with Lines

3.3.1 Distance Formula

Suppose you have two points $A(a, b)$ and $B(c, d)$. If we draw a line passing through both of them, what is the distance on that line between A and B? That information is given by the **distance formula**

$$d(A, B) = \sqrt{(c-a)^2 + (d-b)^2}$$

For instance, how far is the point $X(5, 3)$ from the origin $O(0, 0)$? Simply evaluate directly with the distance formula:

$$d(X, O) = \sqrt{(5-0)^2 + (3-0)^2} = \sqrt{5^2 + 3^2} = \sqrt{25 + 9} = \sqrt{34} \approx 5.8$$

3.3.2 Midpoint Formula

Suppose you have two points $A(a, b)$ and $B(c, d)$. If we draw a line passing through both of them, what are the coordinates of the point C that lies direclty in between them? For that, we employ the **midpoint formula** given by

$$(x_{\text{mid}}, y_{\text{mid}}) = \left(\frac{c+a}{2}, \frac{d+b}{2}\right)$$

It is helpful to realize that we are just finding the average of the x- and y-coordinates.

3.3.3 Parallel and Perpendicular Lines

Two lines are **parallel** if they never intersect. In other words, two lines are parallel if they have the same slope but a different y-intercept. Mathematically, then, the form of two parallel lines is given by

$$y_1 = mx + b$$

$$y_2 = mx + c$$

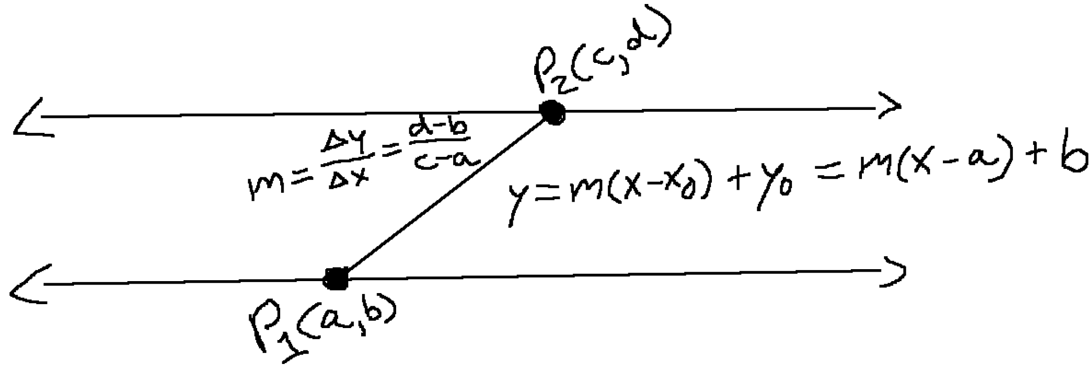
where $b \neq c$. On the other hand, **perpindicular lines** intersect and make a 90° angle with each other. Given a line $y = mx + b$, it is possible to find a family of perpendicular lines by the formula

$$y_\perp = -\frac{1}{m}x + d$$

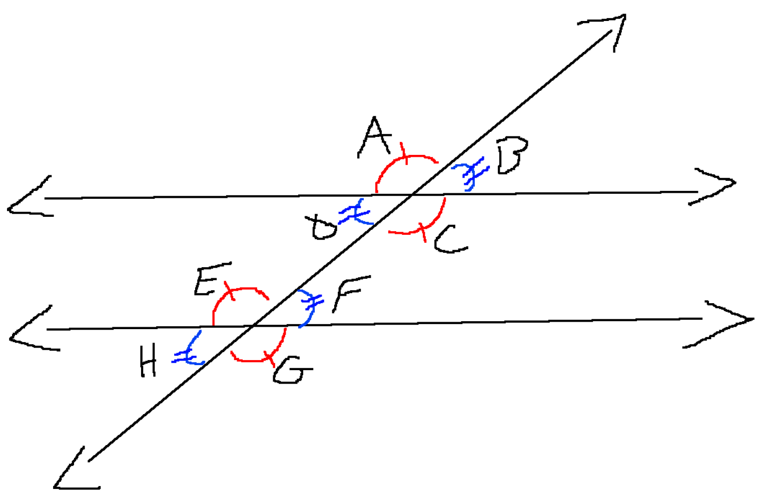
where d is any y-intercept; only the slope determines whether two lines are perpendicular.

3.3.4 Transversal Intersecting Parallel Lines

Suppose you have two parallel lines, and consider a pair of distinct points, $P_1(a, b)$ and $P_2(c, d)$, with one on each line. We can define a new line which passes through both of these points as follows:



We refer to lines that intersect two other lines in this manner as **transversals**. "It turns out" (details planned) that all of the red angles in the figure below are the same; likewise, all of the blue angles are also the same. Further, we have that $A + B = 180$ since they are **supplementary angles**.



The angles A, B, G, and H are **exterior angles**, while all the others are **interior angles**. We apply the adjectives **alternate** and **same-side** to specific interior and exterior angles to describe which side of the transversal another angle is. We illustrate by way of example:

- A and E are **corresponding angles**
- D and F are **alternate interior angles**
- B and H are **alternate exterior angles**
- C and F are **same side interior angles**
- G and B are **same side exterior angles**

The patterns suggested by color in the bullet list hold in general; corresponding angles, alternate interior angles, and alternate exterior angles are always the same. On the other hand, same side interior angles and same side exterior angles are supplementary and sum to $180°$.

3.4 Symmetry and Transformations

3.4.1 Introduction

Symmetry and transformations are fundamental concepts in Euclidean geometry that play a crucial role in understanding the underlying structure and properties of geometric objects. **Symmetry** refers to the property of an object that remains unchanged under certain **transformations**. In the context of Euclidean geometry, a transformation is a function that maps points from one location in the plane to another location. These transformations can include translations, rotations, reflections, dilations, and combinations of these operations. Each transformation has a specific effect on the geometric object being transformed, such as changing its position, orientation, size, or shape. These transformations not only preserve the shape and size of objects, but they also provide a powerful tool for solving geometric problems and exploring their properties. For instance, symmetry can be used to identify congruent figures, find the center of mass of an object, or prove geometric theorems. Therefore, the study of symmetry and transformations in Euclidean geometry is essential for developing a deeper understanding of the subject and its applications in various fields, including physics, engineering, and computer graphics.

Transformations that preserve distances between points are called **isometries**. Examples of isometries which we will regularly employ include the following:

- **Translation**: shift a figure up or down by a constant amount
- **Rotation**: move a figure counterclockwise relative to a point
- **Reflection**: flip a figure over a line

However, if you change the size of a shape (i.e., perform a **dilation** transformation), then you are not keeping the distance between points the same, and such a function is not an isometry. Instead, if such a function still preserves the overall shape of a figure, it is referred to as a **similarity transformations**. We will explore these nuances in greater detail later.

3.4.2 Basic Algebra Perspective

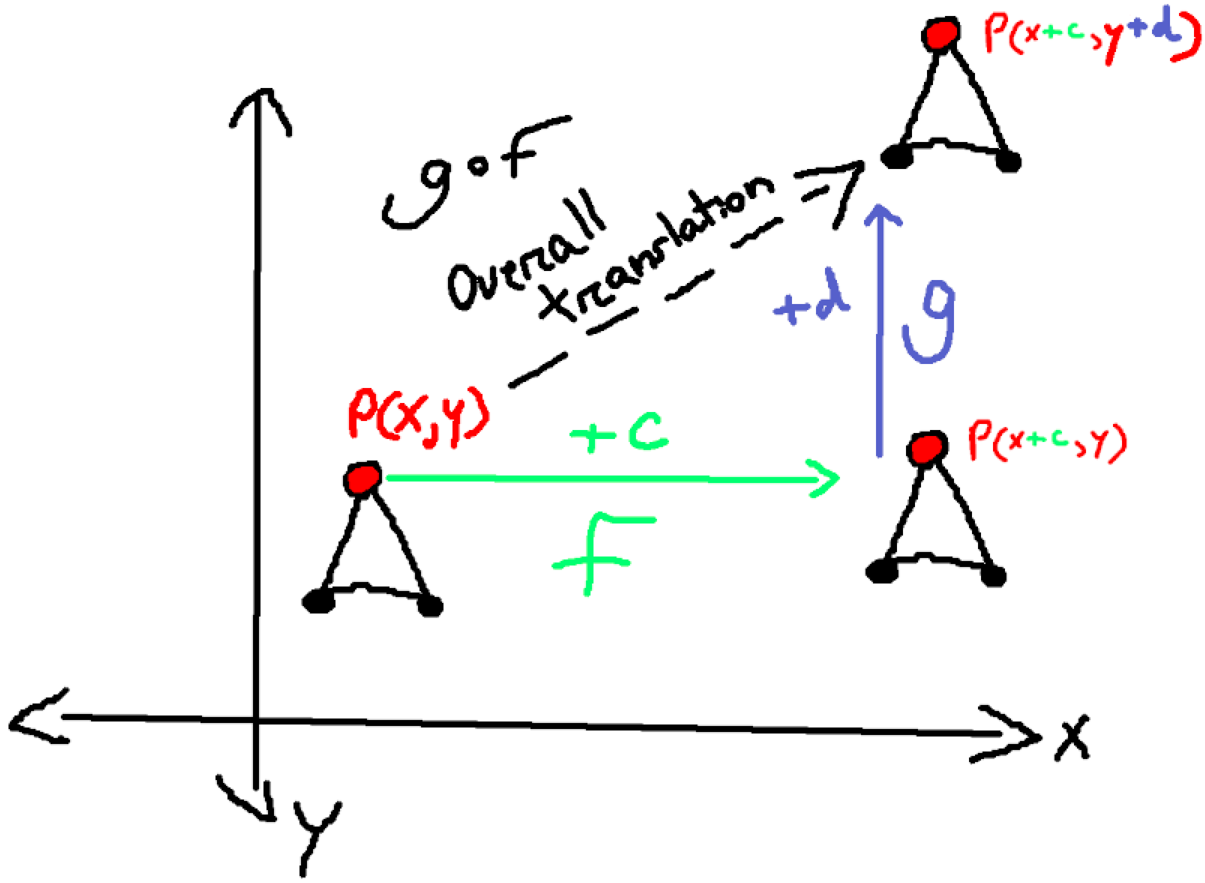
Let $P(x, y)$ be a point. Then a **translation** of P along the x-axis is given algebraically by

$$(x, y) \mapsto (x + c, y)$$

where c is some constant. Likewise, a translation along the y-axis is given by

$$(x, y) \mapsto (x, y + d)$$

where d is also a constant. If we perform a translation along the x-axis and immediately follow with a translation along the y-axis, the overall effect is the same as if you translated the figure along a line (or vector).



Rotations are more difficult to treat with basic algebra alone; therefore, we will reserve many details for the next section on the "linear algebra" treatment of transformations. However, there is still much we can deduce without advanced math. (to be continued)

3.4.3 "Linear Algebra" Perspective

(planned - given a vector **v** representing a point, we can use linear functions (matrices) to do all the transformations discussed previously)

3.4.4 Abstract Algebra Perspective

(planned - "group theory" is in large part the formal study of symmetry)

3.5 Triangles

3.5.1 Triangle Relationships

(planned)

3.5.2 Congruence and Similarity

(planned)

3.5.3 Right Triangles

(planned)

3.6 Quadrilaterals

(made of triangles)

3.7 Polygons

3.7.1 Classifying Polygons

(convex, concave, regular)

3.7.2 Interior and Exterior Angles

(planned)

3.8 Circles

(planned)

3.9 Constructions

(planned)

3.10 Perimeter, Area, and Volume

(planned)

3.11 Conic Sections

(planned)

3.11.1 Ellipses

(planned; center, vertices, co-vertices, foci, and eccentricity)

3.11.2 Parabola

(planned; vertex, focus, directrix, and eccentricity)

3.11.3 Hyperbola

(planned; center, vertices, foci, asymptotes, and eccentricity)

Chapter 4

Trigonometry

4.1 The Unit Circle

4.1.1 Definition and Angles

Before beginning any trigonometry, it is necessary to be acquainted with **the unit circle**. At first glance, there is nothing special about the unit circle: it is just the circle centered at the origin $O(0,0)$ with radius $r = 1$. In set builder notation, it is equal to

$$S = \{(x, y) \in \mathbb{R}^2 \mid \sqrt{x^2 + y^2} = 1\}$$

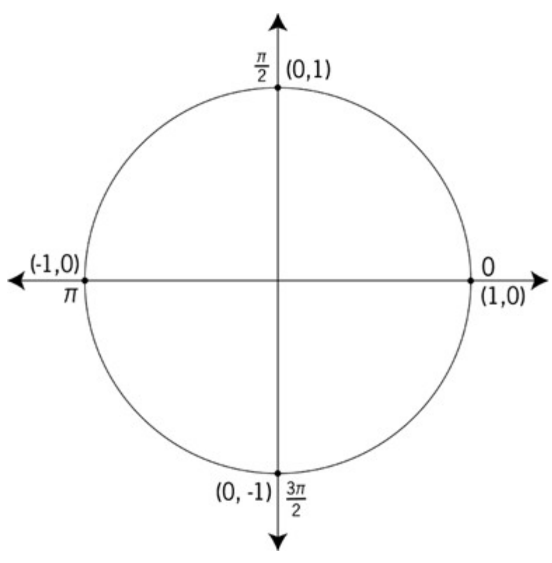
which is the set of all points (x, y) a distance 1 away from the origin. As an equation, we have

$$x^2 + y^2 = 1$$

which is the same as

$$y = \pm\sqrt{1 - x^2}$$

An illustration is provided below:

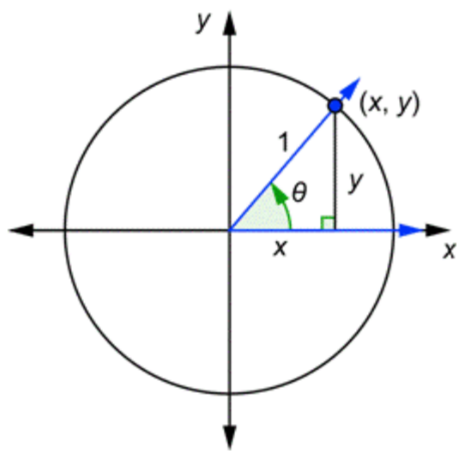


You may recall from geometry that the circumference of a circle is equal to $2\pi r$. Thus, it stands to reason that the circumference of the unit circle is equal to 2π.

If we start at the point $(1, 0)$ and travel a quarter of the way around the unit circle, we arrive at $(0, 1)$, and the distance we traveled is equal to $2\pi/4 = \pi/2$. When we are halfway around the unit circle, we arrive at $(-1, 0)$ and have traveled a distance of π. When we are three quarters of the way around, we land at $(0, -1)$ and have traveled a distance of $3\pi/2$. Finally, after traveling the final quarter, we arrive back where we started at $(1, 0)$, and we have traveled the full 2π distance around the circumference of the circle.

Since traveling a distance of 2π and 0 both correspond to the same point $(1, 0)$, we say that these two values are equivalent on the unit circle. In fact, no matter where you start, if you add 2π, you will always end up at the same spot.

Practically speaking, numbers in the interval $[0, 2\pi)$ help us define **angles**:



When we draw arrows pointing from the origin to where we started on the unit circle - $(1, 0)$ by convention - and our current position (x, y), observe that an angle θ is formed between them!

Often, when one writes an angle in the form of a number between 0 and 2π, we say the angle is in "units" of **radians** (rad). In truth, since "radians" naturally appear from the circumference of the unit circle, angles measured in radians do not have units. However, the idea of "radians" as a unit is useful when contrasting with other ways to measure angles, such as **degrees** (°). The formula for converting from radians to degrees is as follows:

$$\text{degrees} = 360 \cdot \frac{\text{radians}}{2\pi}$$

Always be aware of whether you are using degrees or radians!

Small but important note about angles: if your angle is negative, you travel in the **opposite direction** around the circle. For example, $-90°$ or $-\pi/2$ would be at the point

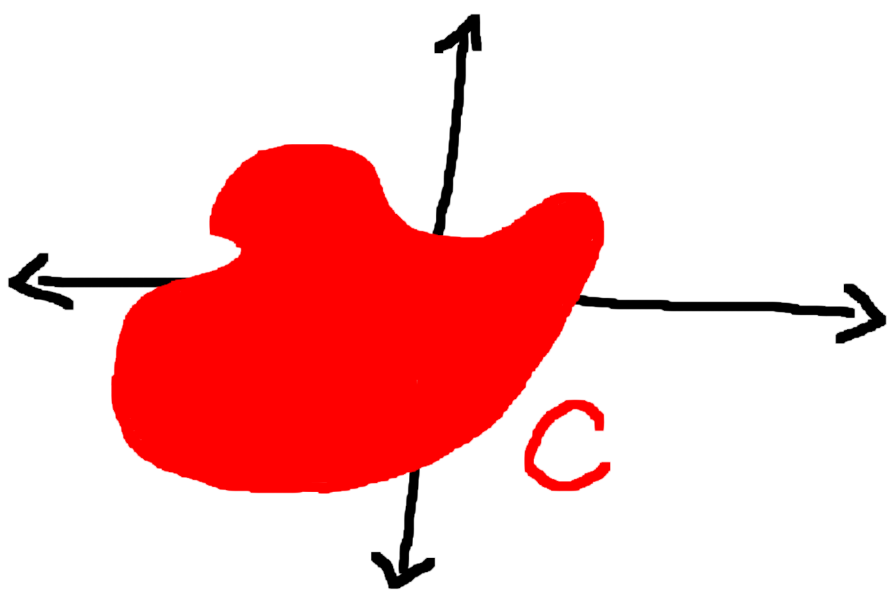
$(0, -1)$, just like $3\pi/2$.

Finally, see that we don't just get an angle at every point along the unit circle: we can also construct a full triangle. The hypotenuse is always equal to 1, but the other sides depend on the point on the unit circle.

4.1.2 Relations and Equivalence Relations

To better understand the unit circle and some of the properties of trigonometric functions, we will take a brief detour through **relations**.

Graphically, if we have a coordinate plane, such as $\mathbb{R}^2$ (sometimes denoted as $\mathbb{R} \times \mathbb{R}$, called a Cartesian product), then a **relation** is some subset C of the plane:



How does this make sense? What does a set have to do with relations? Here is how: if "x is related to y," then the ordered pair (x, y) appears in the set C. Some basic examples of "rules of assignment" that yield relations are given below:

- $x > y$
- $x < y$
- $x = y + 1$
- $y = \pm\sqrt{r^2 - x^2}$

All of these should be familiar to you. Inequalities are famous for having many solutions, requiring you to **shade in** the region of the graph where it is true (you're drawing the set C!). Equations can also make relations since there is also a distinct set of points in the coordinate plane which make it true. Indeed, the definition of a relation is very broad; many different expressions could be one.

Formally speaking, the relation C that is a subset of the plane $\mathbb{R} \times \mathbb{R}$ is said to be "on" the set $\mathbb{R}$ because we are making pairwise comparisons of elements from $\mathbb{R}$. In general, a relation on a set X is a subset C of the set $X \times X$.

An **equivalence relation** is a special kind of relation "on" a set X. Instead of symbols like $>$, $<$, or $=$, mathematicians often use $\sim$ to represent equivalence relations. If we write $x \sim y$, then we read it as "x is equivalent to y." Equivalence relations all have the following three properties:

1. (Reflexivity) $x \sim x$ for all $x \in X$

2. (Symmetry) If $x \sim y$ then $y \sim x$

3. (Transitivity) If $x \sim y$ and $y \sim z$ then $x \sim z$

If this seems a little abstract, don't worry. Let's try to understand better by investigating the very familiar "equals to" ($=$) relation on $\mathbb{R}$.

Theorem 4.1. *The "equals to" relation ($=$) is an equivalence relation on* $\mathbb{R}$*.*

Proof. We know that $x = x$ no matter what for every $x \in \mathbb{R}$. We also know that if $x = y$ then $y = x$. Finally, if $x = y$ and $y = z$, then it is definitely true that $x = z$. Therefore, "equals to" is an equivalence relation. Q.E.D.

So an equivalence relation is just some relation which "acts like" an equality sign ($=$) between two numbers. To see how this can be an interesting concept, let's use what we know about angles to come up with an equivalence relation.

Since we end up back at the same spot on the unit circle if we walk an extra 2π rad (or $360°$) along the boundary, and it doesn't matter how many times we do this since we will still end up back at the same spot, we obtain the following relation:

$$\theta \sim \phi \Longleftrightarrow \phi = \theta + 2\pi n$$

for which $n \in \mathbb{Z}$ is some integer.

Proposition 4.1. $\sim$ *is an equivalence relation on* $\mathbb{R}$*.*

Proof. Let $\theta, \phi, \psi \in \mathbb{R}$. We proceed by proving each of the three criteria of an equivalence relation.

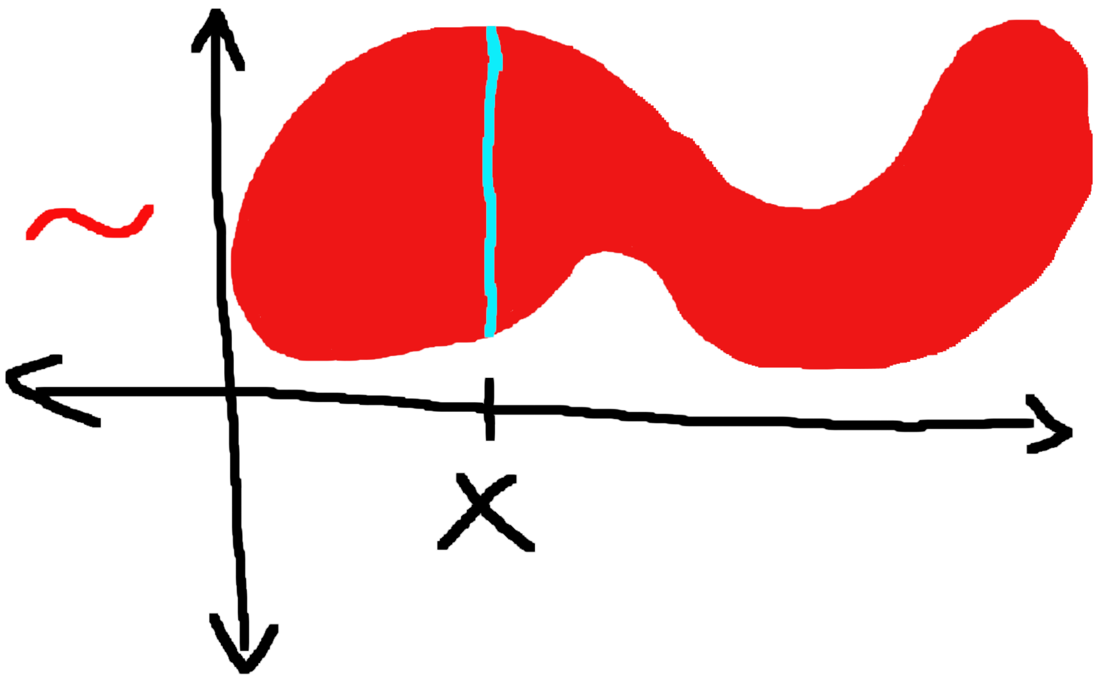
1. (Reflexivity) We know that $\theta = \theta + 2\pi(0)$, so $\theta \sim \theta$.

2. (Symmetry) If $\theta \sim \phi$, then $\phi = \theta + 2\pi n$. This means that $\theta = \phi - 2\pi n = \phi + 2\pi(-n)$. We know that $(-n)$ is also an integer, so $\phi \sim \theta$.

3. (Transitivity) If $\theta \sim \phi$ and $\phi \sim \psi$, then $\phi = \theta + 2\pi n$ and $\psi = \phi + 2\pi m$, for which $m, n \in \mathbb{Z}$. Rearrangment yields $\phi = \psi - 2\pi m$, which we can substitute into the other equation to get $\psi - 2\pi m = \theta + 2\pi n$. When we add $2\pi m$ to both sides, the equation becomes $\phi = \theta + 2\pi(n + m)$. Since we know that $n + m$ is also an integer, $\theta \sim \phi$.

All three properties hold. Therefore, $\sim$ is an equivalence relation on $\mathbb{R}$. Q.E.D.

With this equivalence relation, we can say any two angles θ and ϕ are equivalent if they are only different by an integer multiple of 2π. We can talk about all the angles equivalent to θ in $\sim$; the set of all such angles is called the **equivalence class** of θ.



In this graphic, the red blob represents the set $\sim$ which is some equivalence relation. The blue line beneath x is the equivalence class of x; it is the set containing all the ordered pairs of the form $(x, \text{something})$. The "somethings" are all the numbers considered to be "equivalent to" x.

Returning to our example concerning the equivalence relation between angles, we can write the equivalence class of some angle theta as the set

$$E_\theta = \{\phi \in \mathbb{R} \mid \phi = \theta + 2\pi n\}$$

Equivalence classes are subsets of the original set. The set E_θ above, for instance, contains many real numbers separated in intervals of 2π. It also turns out that any two equivalence classes, say E_θ and E_ϕ, are either equal (the same set) or **disjoint** (nothing in common between them). Graphically, this makes sense; if you imagine that the red blob above is made out of an infinite number of parallel light blue line segments, then you know from geometry that any two lines (aka, any two equivalence classes) will never cross unless they are the exact same line segment. This is great for intuition, but it is not a logically concrete proof; that is given here:

Theorem 4.2. *Let E_θ and E_ϕ be equivalence classes of $\sim$. If E_θ and E_ϕ are not disjoint, then they are the same.*

Proof. Suppose E_θ and E_ϕ have at least one element ψ in common (so they are not disjoint). That means $\psi \sim \theta$ and $\psi \sim \phi$. Because $\sim$ is symmetric, we know that $\phi \sim \psi$. Because $\sim$ is transitive, we know that because $\phi \sim \psi$ and $\psi \sim \theta$ then $\phi \sim \theta$, which is the same as $\theta \sim \phi$ by symmetry. Again, by transitivity, any angle equivalent to θ is also equivalent to ϕ; likewise, any angle equivlaent to ϕ must also be equivalent to θ. Therefore, since θ and ϕ are equivalent to every element of their equivalence classes, both equivalence classes E_θ and E_ϕ must be the same. Q.E.D.

When we know what an equivalence class of some equivalence relation $\sim$ looks like, it becomes possible to do something very interesting. Since we know all the points that are equivalent to each other, we can find the **quotient set**. In the quotient set, we "smash" every equivalence class down to only one point (a bit like dividing a number by itself to get 1, hence the term "quotient").

The quotient set of $\mathbb{R}$ with the equivalence relation $\sim$ is denoted by $\mathbb{R}/\sim$. To understand the result, let's write out one of the equivalence classes:

$$E_0 = \{\dots, -4\pi, -2\pi, 0, 2\pi, 4\pi, \dots\}$$

These are all the numbers that will be "smashed" down to just 0. Notice that the equivalence classes must stop before 2π because it is already equivalent to 0. In other words, we can tell equivalence classes apart from each other up until 2π. So our quotient set is

$$\mathbb{R}/\sim = [0, 2\pi)$$

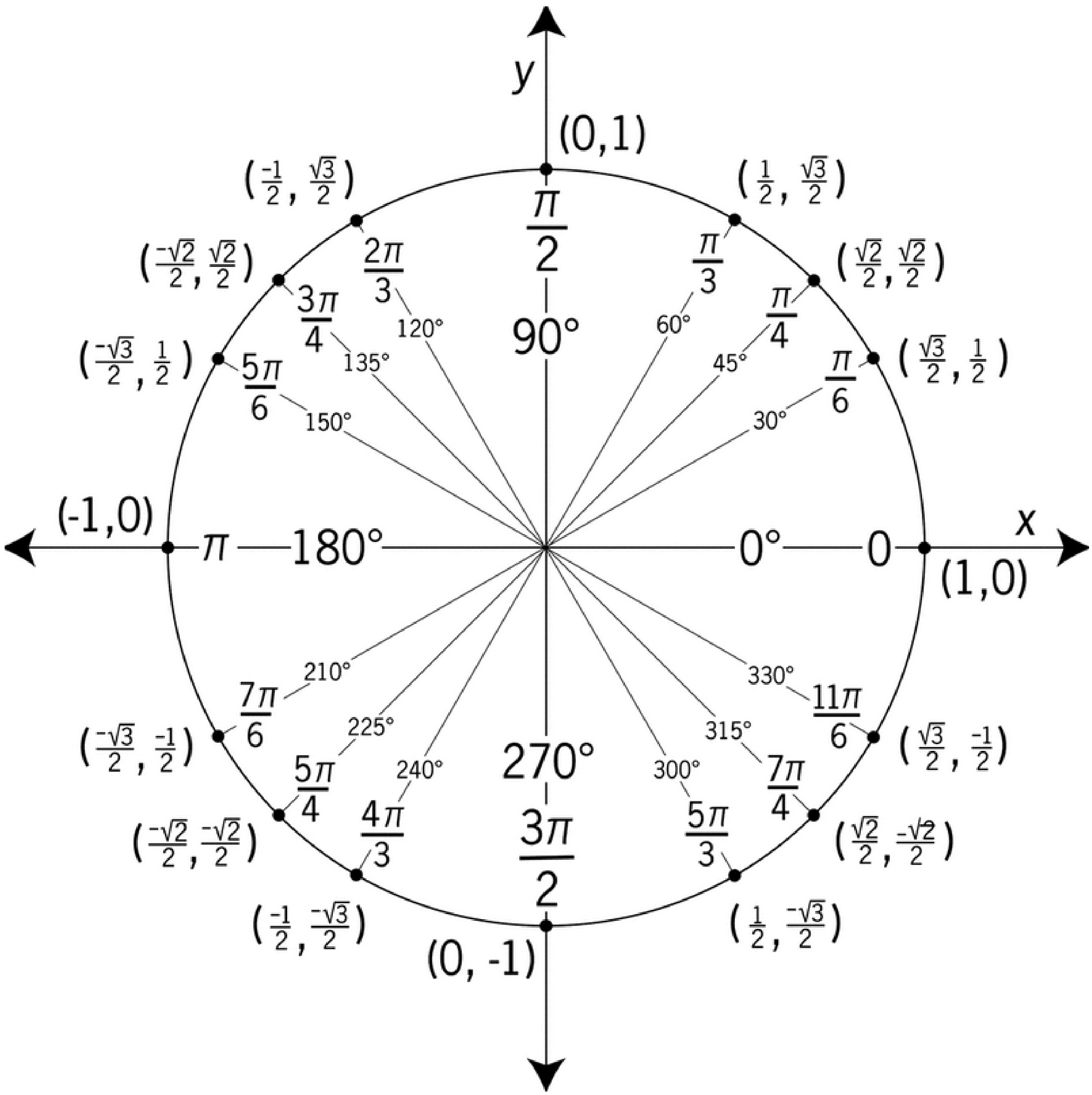
The quotient set for our equivalence relation is just the line segment the same length as the circumference of the unit circle.

Using a clever equivalence relation and a quotient, we can actually construct the unit circle from a straight line. I define a new equivalence relation $\sim$ on $[0, 2\pi]$ as follows:

$$\theta \sim \phi \Longleftrightarrow \begin{cases} \theta = \phi \\ \theta = 0 \text{ and } \phi = 2\pi \\ \theta = 2\pi \text{ and } \phi = 0 \end{cases}$$

This equivalence relation sets 0 and 2π eqivalent and makes every other point equal to itself. When we do the quotient $[0, 2\pi]/\sim$ and equip the set with a **topology**, we obtain a circle which, like silly putty or clay, can be continuously shaped into the unit circle. While interesting, these details go far beyond the scope of trigonometry.

4.1.3 Unit Circle Reference Page

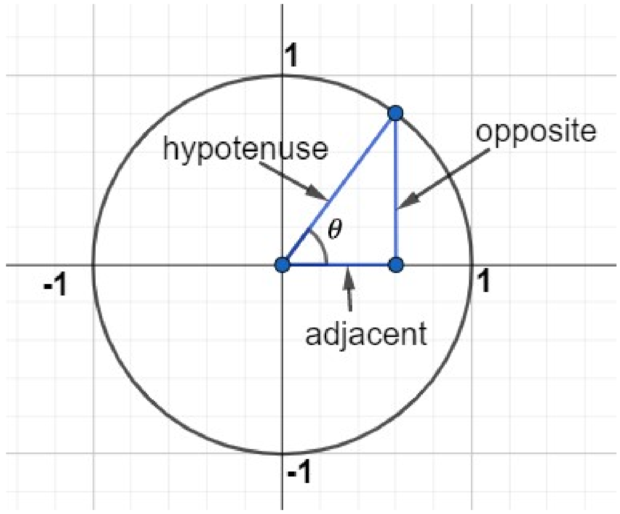


Inside the circle: Key angles to know in both radians and degrees.

Outside the circle: The points (x, y) on the circle at the specified angle. These are also the lengths of the sides of the triangle corresponding to that angle!

4.2 Trigonometric Functions

4.2.1 Sine and Cosine Functions



Recall that the real numbers $\mathbb{R}$ equipped with the equivalence relation $\theta \sim \phi \iff \phi = \theta + 2\pi n$ allows us to "smash" all possible angles down to just the interval $[0, 2\pi)$. For this reason, it is acceptable to write the domain of the following two trigonometric functions, sine and cosine, as either the entire real like $\mathbb{R}$ or the interval $[0, 2\pi)$. Since it is easier for me to type, the following functions will use $\mathbb{R}$ as the domain.

Let the sine function $\sin\colon \mathbb{R} \to [-1, 1]$ be defined by

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

Let the cosine function $\cos\colon \mathbb{R} \to [-1, 1]$ be defined by

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$$

These functions work on any triangle and return the same value as long as the angle is the same. If the triangles in question come from the unit circle, then the hypotenuse is always equal to 1. So, for the unit circle only,

$$\sin\theta = \text{opposite}$$

$$\cos\theta = \text{adjacent}$$

However, we ought to remember what the sides of the triangle actually mean: the adjacent side is exactly equal to the x value of the point on the unit circle, and the opposite side is the y. Thus,

$$\sin\theta = y$$

$$\cos\theta = x$$

If we are dealing with any circle, then

$$\sin\theta = \frac{y}{r}$$

$$\cos\theta = \frac{x}{r}$$

where r is the radius of the circle (and the hypotenuse of the triangle). By dividing by the hypotenuse, $\sin\theta$ and $\cos\theta$ have the exact same value for any circle as long as the angle θ is the same.

We can derive one of the most important identities from trigonometry with this information. First, we need to solve for x and y in terms of $\sin\theta$ and $\cos\theta$:

$$x = r\cos\theta$$

$$y = r\sin\theta$$

Now, take the Pythagorean identity,

$$x^2 + y^2 = r^2$$

and plug in x and y:

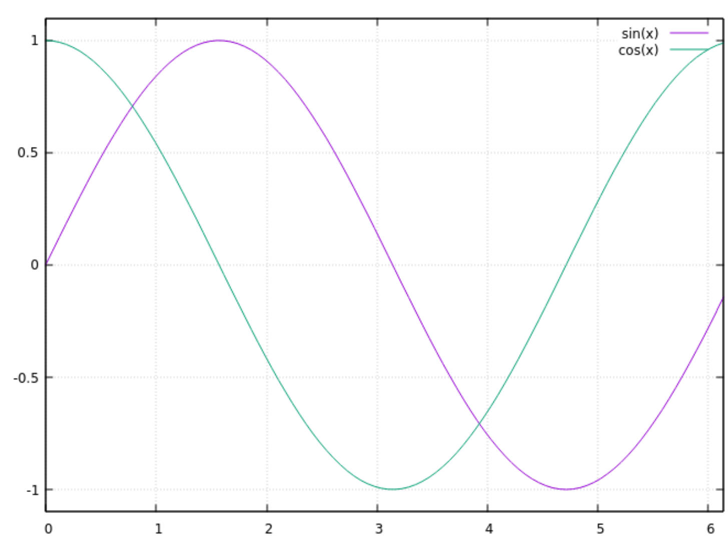
$$(r\sin\theta)^2 + (r\cos\theta)^2 = r^2$$

When we divide both sides by r^2, we obtain

$$\boxed{\sin^2\theta + \cos^2\theta = 1}$$

This is the Pythagorean theorem in trigonometric form. Remember this one! It comes in handy for manipulating all sorts of trig expressions.

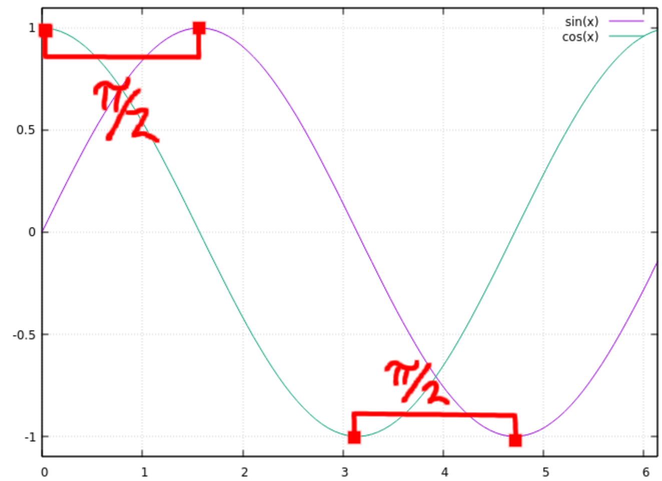
Abstract stuff aside for the moment, let's investigate some of the properties of the sine and cosine functions.



In the picture above, $\sin(x)$ in purple and $\cos(x)$ in blue are plotted from $x = 0$ to $x = 2\pi$.

It should come as little surprise that the range of sine and cosine is the interval $[-1, 1]$. Since $\sin\theta$ represents the y value and $\cos\theta$ the x value, the biggest either of them will ever be is when the angle is standing straight up and down ($90°$ and $270°$) or straight across the horizontal axis ($0°$ and $180°$).

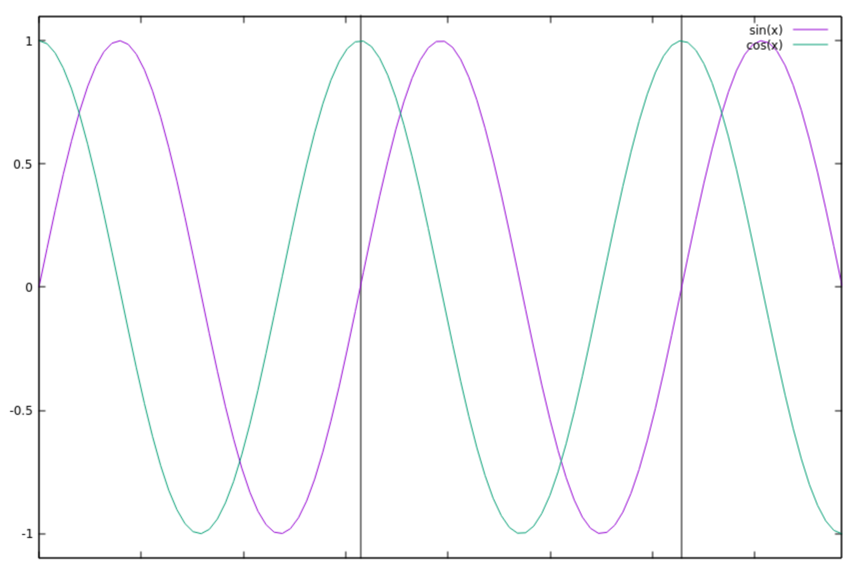
Notice: $\cos\theta$ has the same shape as $\sin\theta$, but it appears to be "lagging behind" on the graph. With waves, we call this being "out of phase."



If we compare the distance between any two analogous points (I chose the peaks since it was easiest), we see that they are separated by a "phase angle" of $\phi = \pi/2$. Later, we will have the tools to prove the following identity:

$$\sin\theta = \cos\left(\theta - \frac{\pi}{2}\right)$$

Another important feature of the graph is that it repeats, or is "periodic," every 2π:



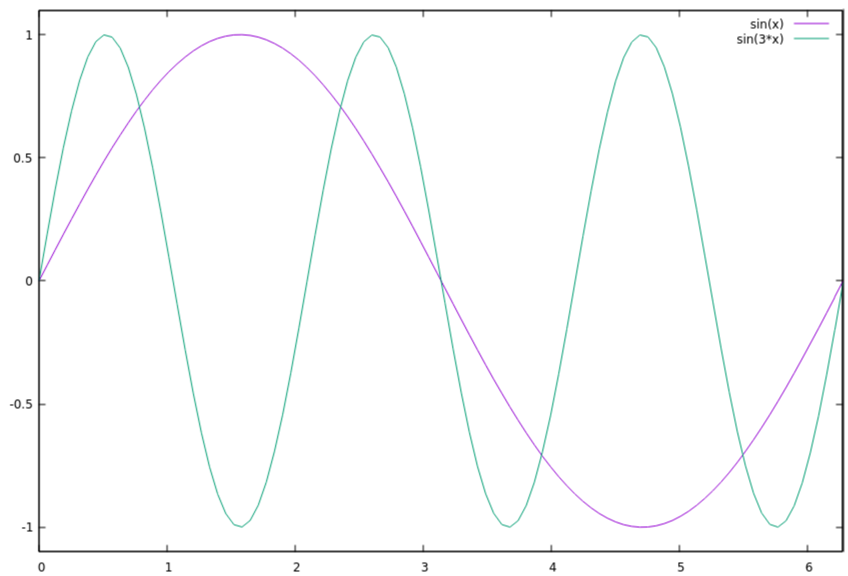
The sine and cosine functions can be transformed in several ways. In general, we have

$$f(x) = A\sin\left(\omega(x+\phi)\right) + B$$

We have already discussed the *phase angle* ϕ. This will shift the wave left or right along the $x-$axis. When you add a phase angle $\phi = 90°$ to sine, you obtain a function which behaves identically to regular cosine.

A is the *amplitude*. Since sine and cosine usually only go from -1 to 1, the amplitude factor will stretch the wave up or down from $-A$ to A.

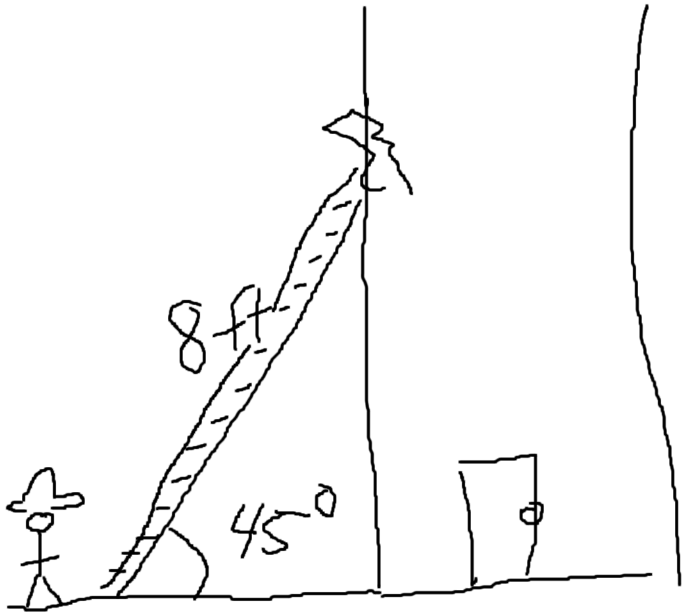
ω (lowercase "omega") is the *frequency* of the wave, which determines how quickly the wave *oscillates*. When $\omega = 1$ (like in the unmodified form $\sin\theta$ or $\cos\theta$), the frequency is 1 cycle per 2π radians. If I had $\omega = 3$, like in $\sin(3x)$, then the new frequency would be 3 cycles per 2π radians.



B simply shifts the entire wave up or down the $y-$axis.

The tops of a wave are called *peaks* while the bottoms are called *troughs*. The distance between peaks is called the *wavelength* and is typically represented by the Greek letter "lambda" (λ).

You can make some fun functions by replacing any of the constants above with another function. For example, the function $g(x) = x\sin x$ is a normal sine wave, but its amplitude keeps getting bigger. Sine and cosine themselves are very good for inducing "wiggliness" in an equation.



Oh no! The mayor dropped a candle, and Town Hall is on fire! Fortunately, Firefighter Frank is on the scene to save the day.

Frank's ladder is 8 ft long, and it is resting against the building at an angle of 45°. How many feet off the ground is the top of the ladder?

We know that it is possible to relate an angle, the hypotenuse, and a side of a triangle with trig functions. Specifically, to solve this problem, we want $\sin\theta$, which relates hypotenuse and the y coordinate. So,

$$\sin(45°) = \frac{y}{8 \text{ ft}} \implies y = (8 \text{ ft}) \sin(45°)$$

We can check the unit circle to see that $\sin(45°) = \sqrt{2}/2$, which allows us to conclude

$$y = 4\sqrt{2} \approx 5.6 \text{ ft}$$

The mayor's window is only 5.6 feet off the ground. If we also wanted to know the horizontal side of the triangle, we would use the Pythagorean theorem:

$$x^2 + y^2 = r^2 \implies x = \sqrt{(8)^2 - (5.6)^2} \approx 5.6 \text{ ft}$$

We will close off this section with a few more miscellaneous facts about sine and cosine.

- $\sin(x)$ is an odd function, meaning that $\sin(-x) = -\sin(x)$
- $\cos(x)$ is an even function, meaning that $\cos(-x) = \cos(x)$
- For reasons we will uncover when discussing series in calculus, for small values of θ (say, $\theta < \pi/12$ rad or $\theta < 15°$) we can use the approximation $\sin\theta \approx \theta$ and $\cos\theta \approx 1$
- The equations relating x and y to r and θ from earlier,

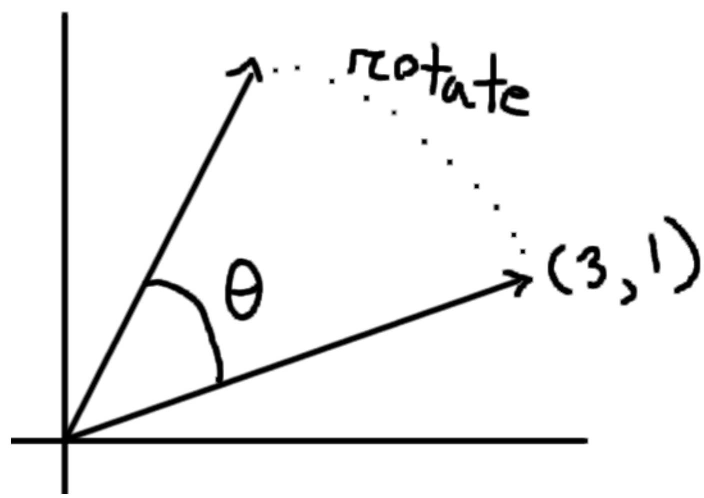
$$x = r\cos\theta$$

$$y = r\sin\theta$$

  are used to convert from Polar coordinates (r, θ) to regular Cartesian coordinates (x, y)

- The **rotation matrix** has entries using $\sin\theta$ and $\cos\theta$:

$$\mathbf{R}_{2\times 2}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



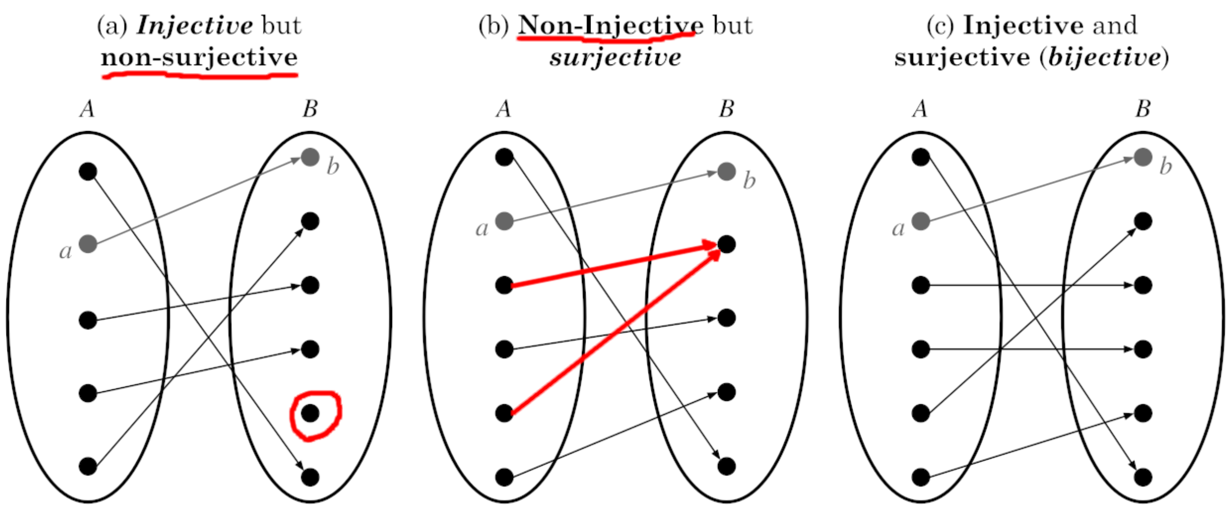
If we wanted to rotate a vector (i.e. a point in $\mathbb{R}^2$) by 60°, all we would have to do is multiply by the rotation matrix:

$$\begin{bmatrix} \cos 60° & -\sin 60° \\ \sin 60° & \cos 60° \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3\cos(60°) - \sin(60°) \\ 3\sin(60°) + \cos(60°) \end{bmatrix} = \begin{bmatrix} 0.63 \\ 3.1 \end{bmatrix}$$

We will soon have the tools needed to derive the rotation matrix ourselves.

4.2.2 Injective, Surjective, and Bijective Functions

Now that you are familiar with a wide array of functions, including polynomials, rational functions, exponents and logarithms, linear transformations, and basic trigonometric functions, it is worthwhile to take another small mathematical detour to compare the different ways a function may map elements from its domain to its range. In the exposition to follow, we will work with the generic function $f\colon D \to R$ where D is some domain and R is some range.



Logically speaking, a function is **injective** if "for all $x, y \in D$, if $f(x) = f(y)$, then $x = y$." This means that two different numbers from the domain cannot map to the same value in the range (unless the two numbers aren't different after all). For this reason, we sometimes call injective functions **one-to-one**.

An easy example of an injective function is the exponential function $\exp\colon \mathbb{R} \to \mathbb{R}$ defined by $\exp(x) = e^x$. Contrast with the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. We know that $f(x) = f(-x)$ because $(-x)^2 = x^2$, so f cannot be injective. On the other hand, we can show with logs that if $e^x = e^y$, then $x = y$, making the exponential function one-to-one.

A function is **surjective** if "for every $y \in R$ there exists an $x \in D$ for which $y = f(x)$." This means that the function maps to every element of the range at least once. If you imagine that you are covering the range in a blanket that covers the whole set, it will make sense that we sometimes call surjective functions **onto**.

Do you think the exponential function from our previous example is surjective? The answer is no: we defined the function as $\exp\colon \mathbb{R} \to \mathbb{R}$, but e^x is always greater than 0, skipping all negative real numbers! On the other hand, if we define the cosine function as $\cos\colon \mathbb{R} \to [-1, 1]$, then this function is onto because cosine hits every real number between -1 and 1 over and over.

A function is **bijective** if it is both injective and surjective. This means that every element of the domain maps to exactly one element of the range without missing any possible values. For this reason, bijective functions are commonly called a **one-to-one correspondence**.

We asserted that cosine is surjective, but it is not also injective because $\cos\theta = \cos(\theta + 2\pi)$. Therefore, cosine is not bijective. Even though $\exp\colon \mathbb{R} \to \mathbb{R}$ is not surjective, we can make it bijective by changing the range to make it surjective. All we need to do is remove the negative real numbers from the range, making it $\exp\colon \mathbb{R} \to (0, \infty)$. This version of $\exp(x)$ is bijective! We can also make $f(x) = x^2$ bijective by assigning the correct domain and range. If we say that $f\colon [0, \infty) \to [0, \infty)$, then f becomes injective because there are no negative numbers in the domain. It also becomes surjective because we removed all the negative numbers from the range, which would be impossible to get to with x^2.

Bijective functions ("bijections") have many attractive properties which make them very useful in mathematics. First, a bijection f is guaranteed to always have an inverse function f^{-1}. It is not difficult to convince yourself of this idea; simply imagine that all the bijective function arrows in picture (c) above switched directions: that would be your inverse function. If you tried to do that to (a) or (b), you would quickly run into problems where the function is either undefined or not single valued (fails vertical line test).

Another popular application of bijections is comparing the size of sets. Suppose you have two sets $\{1, 2, 3\}$ and $\{4, 5\}$. It is easy to see that a one-to-one correspondence isn't possible because the two sets have a different number of elements. The number of elements a set S has is called its **cardinality**, and we represent it by $|S|$ or $\text{card}(S)$, and the existence of a bijective function between two sets is how we know they have the same cardinality.

4.2.3 Tangent and Secant Functions

Let the tangent function $\tan\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{y}{x}$$

Recall that $x = r\cos\theta$ and $y = r\sin\theta$. Substituting these expressions into the tangent function yields

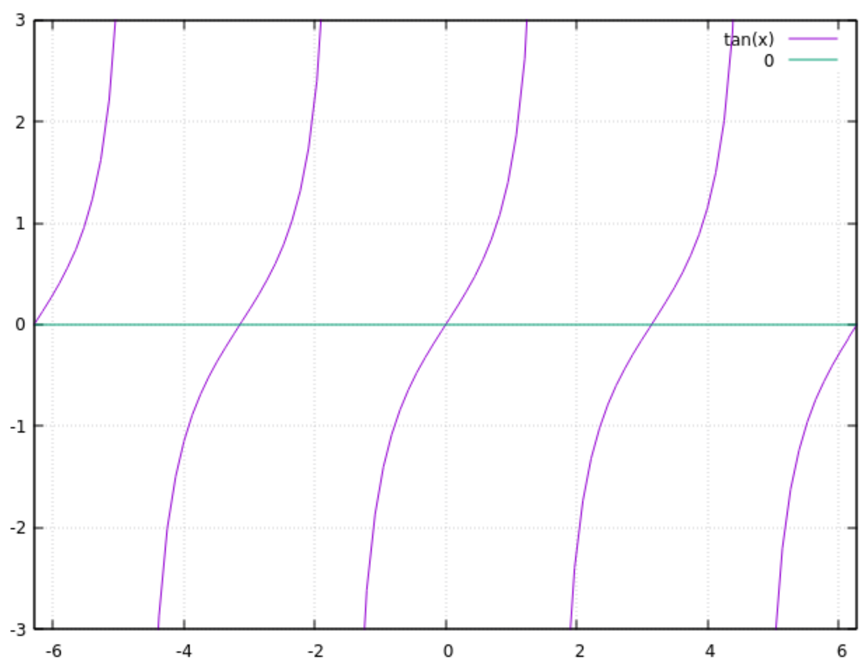
$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

Now that we know the definition of $\tan\theta$, it is time to learn one of the most popular neumonics for remembering the definitions of trigonometric functions:

SOH CAH TOA

Sine equals **O**pposite over **H**ypotenuse; **C**osine equals **A**djacent over **H**ypotenuse; and **T**angent equals **O**pposite over **A**djacent.

It is important to notice that $\tan\theta$ is a rational function that has vertical asymptotes whenever $\cos\theta$ equals zero. As a result, the tangent looks like this when plotted:



The easiest way to plot $\tan\theta$ by hand is to follow these steps:

1. Draw the vertical asymptotes of $\tan\theta$ as dotted lines ($x = \pi/2, 3\pi/2, 5\pi/2, \dots$)

2. Draw a dot where $\tan\theta$ equals zero ($x = 0, \pi, 2\pi, \dots$)

3. Draw a line with a steep, positive slope that slows down as it gets closer to crossing 0; then draw it rapidly increasing again after passing where $\tan\theta = 0$

Based on the graph, we see that $\tan\theta$ is a surjective function because it maps to every possible real number between any two asymptotes. We also know $\tan\theta$ is not injective because $\tan\theta = \tan(\theta + 2\pi)$. Just like sine or cosine, the tangent function is periodic, but due to the asymptotes, it is not continuous everywhere.

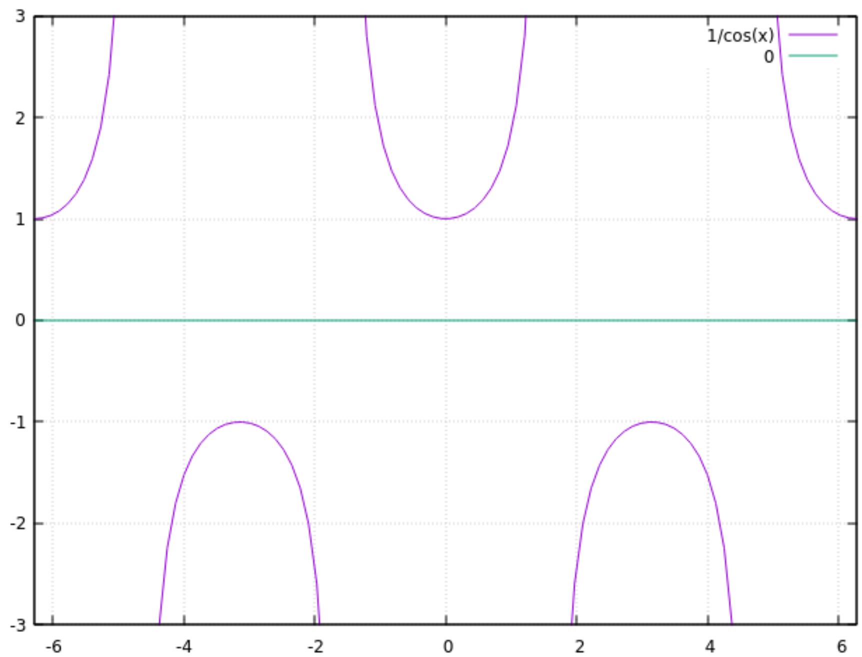
Define the secant function $\sec\colon \mathbb{R} \to \mathbb{R}$ by

$$\sec\theta = \frac{\text{hypotenuse}}{\text{adjacent}} = \frac{r}{x}$$

which is the same as

$$\sec\theta = \frac{1}{\cos\theta}$$

This is another rational function with many asyptotes. For secant, we would draw dotted vertical asymptotes everywhere $\cos\theta = 0$. When $|\cos\theta| < 1$, we would expect $|\sec\theta|$ to be bigger than 1, and when $|\cos\theta| = 1$, we accordingly get $|\sec\theta| = 1$. This being the case, the graph of the secant function looks like this:



Since it is impossible for $\sec\theta$ to equal any value in the interval $(-1, 1)$, we know that $\sec\colon \mathbb{R} \to \mathbb{R}$ is not a surjective function. Like other periodic, trigonometric functions, we also know that secant is not injective because $\sec\theta = \sec(\theta + 2\pi)$.

Recall that we were able to relate sine and cosine with the Pythagorean theorem because both functions divide by the hypotenuse r of a triangle. In analogy, observe that both tanget and secent divide by the adjacent side x of the triangle. From the definitions of both functions, we obtain

$$r = x\sec\theta$$

$$y = x\tan\theta$$

Now substitute into the Pythagorean theorem:

$$x^2 + (x\tan\theta)^2 = (x\sec\theta)^2$$

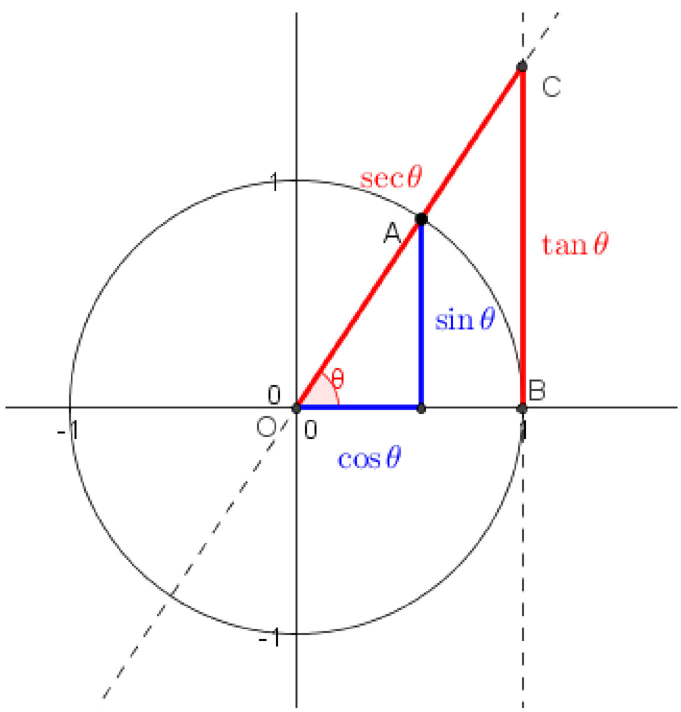
Divide both sides by x^2 to obtain

$$1 + \tan^2\theta = \sec^2\theta$$

We could also derive this version of the Pythagoren theorem starting from sine and cosine:

$$\cos^2\theta + \sin^2\theta = 1 \implies 1 + \frac{\sin^2\theta}{\cos^2\theta} = \frac{1}{\cos^2\theta} \implies 1 + \tan^2\theta = \sec^2\theta$$

Returning to the idea that $\tan\theta$ and $\sec\theta$ are also related to the sides of a triangle, consider the simple case where the adjacent side x is equal to 1. Then we can draw the triangle giving $\tan\theta$ and $\sec\theta$ from the unit circle like this:

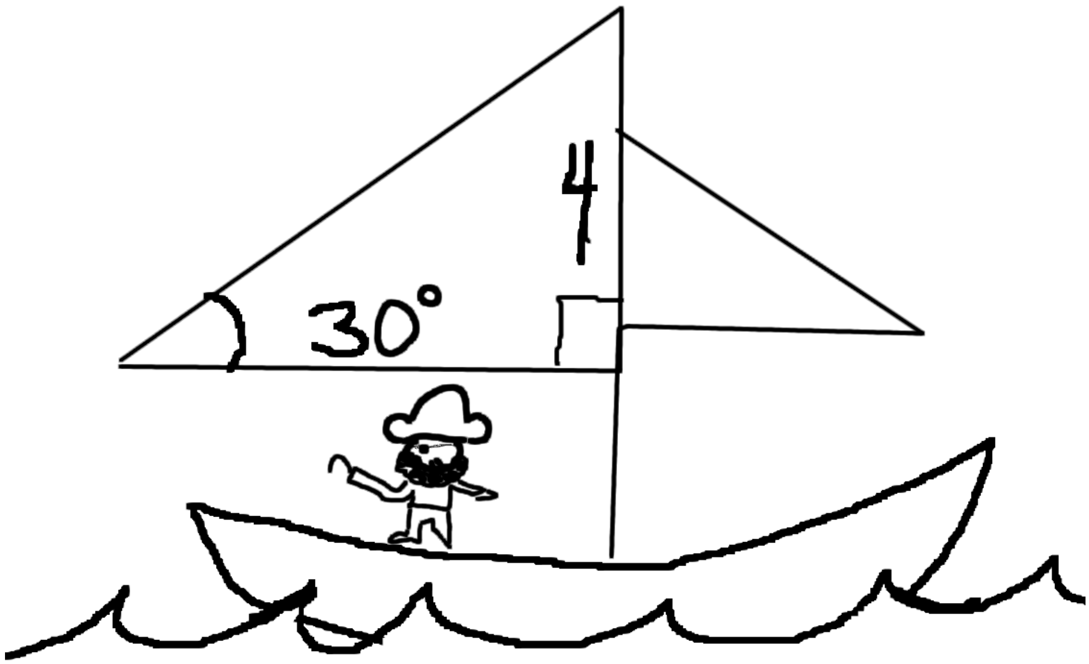


When we set $x = 1$ in the equations for tangent and secant, is important to remember that the resulting triangle (red) is different from the triangle relating sine and cosine (blue). But there is nothing stopping us from using the same triangle for both; we just need to make sure the base of the red triangle is $x = \cos\theta$. Observe what happens when we make the substitution:

$$r = x\sec\theta = (\cos\theta)\sec\theta = \frac{\cos\theta}{\cos\theta} = 1$$

$$y = x\tan\theta = (\cos\theta)\tan\theta = \frac{\cos\theta \cdot \sin\theta}{\cos\theta} = \sin\theta$$

This is telling us that the sides of both triangles (red and blue) match up perfectly!



Captain Pirate sailed to port to replace his 2-D ship sail. He knows the interior angles of the triangle and the length of one side, but the merchants demand that he provide them with at least two lengths! Determine the dimensions of Captain Pirate's sail.

We know that it is possible to relate the vertical and horizontal sides of a triangle with the angle by using the tangent function. So,

$$\tan(30°) = \frac{4}{x}$$

To find $\tan(30°)$, we need to write it in terms of sine and cosine, which we know how to evaluate by using the unit circle:

$$\frac{\sin(30°)}{\cos(30°)} = \frac{4}{x}$$

We check the unit circle to see that $\sin(30°) = 1/2$ and $\cos(30°) = \sqrt{3}/2$, so

$$\frac{1}{\sqrt{3}} = \frac{4}{x} \implies x = 4\sqrt{3} \approx 7$$

If necessary, the last side could be found with the Pythagorean identity.

4.2.4 Cotangent and Cosecant Functions

Let the cotangent function $\cot\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$\cot\theta = \frac{\text{adjacent}}{\text{opposite}} = \frac{x}{y}$$

Similar to $\tan\theta$ in the previous section, this is equivalent to

$$\cot\theta = \frac{\cos\theta}{\sin\theta}$$

We also make the important observation that

$$\tan\theta = \frac{1}{\cot\theta}$$

Let the cosecant function $\csc\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$\csc\theta = \frac{\text{hypotenuse}}{\text{opposite}} = \frac{r}{y}$$

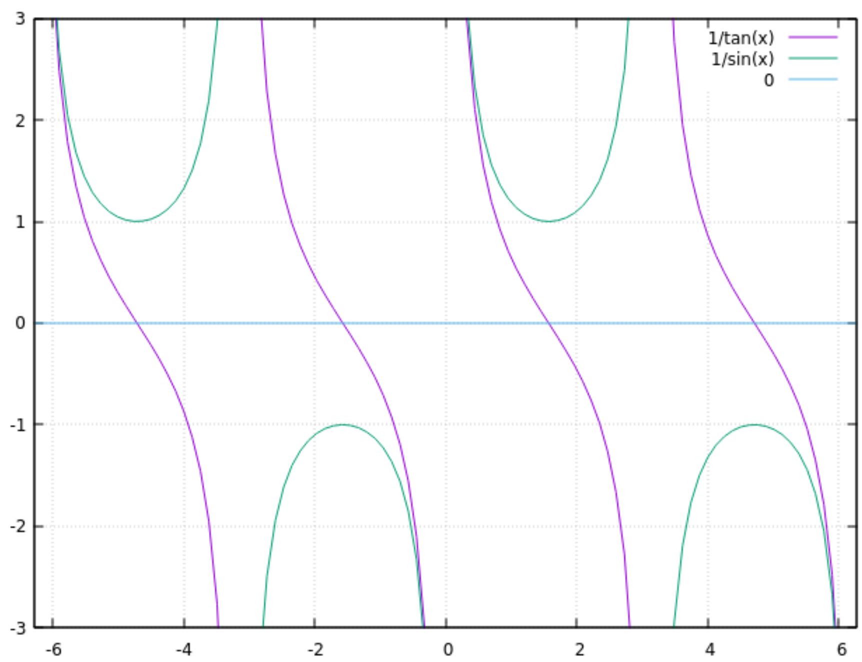
which is the same as

$$\csc\theta = \frac{1}{\sin\theta}$$

These functions behave almost identically to the tangent and secant functions we learned about in the previous section. We can even perform a similar triangle trick by setting $y = 1$ to obtain

$$\cot^2\theta + 1 = \csc^2\theta$$

We see that the graphs of $\csc\theta$ (blue) and $\cot\theta$ (purple) are strikingly similar to $\sec\theta$ and $\tan\theta$, respectively:



Specifically, the graphs are 90° or $\pi/2$ out of phase with their counterparts, and $\cot\theta$ has a negative slope.

4.2.5 Inverse Trigonometric Functions

Suppose you have a function f whose domain is a set D and range is a set R, for which we write

$$f\colon D \to R$$

Recall that the inverse function f^{-1} corresponding to f takes each $y \in R$ back to the corresponding $x \in D$, such that

$$f^{-1}\colon R \to D$$

For a function to have an inverse, it should be **bijective**. Recall that this means it is both injective and surjective, meaning

$$\text{(One-to-one)} \quad f(x) = f(y) \implies x = y$$

$$\text{(Onto)} \quad \text{For all } y \in R \text{ there exists an } x \in D \text{ for which } y = f(x)$$

However, when we define trigonometric functions, they are usually *not* bijective. For example,

$$\cos(90^\circ) = \cos(270^\circ) = 0, \quad \text{but} \quad 90^\circ \neq 270^\circ$$

Suppose we ignored this and tried to define the inverse function of cosine anyways (for now, let's call it "$\cos^{-1}$"). Then what is $\cos^{-1}(0)$? If both 90° and 270° are in the domain of the original function, then the inverse function should map to both of them. But then it fails the vertical line test, meaning $\cos^{-1}$ is not a proper function! This means we have to carefully define our trigonometric functions so they are a proper bijection, allowing us to have well-defined inverse functions.

It turns out that limiting the graph of sine and cosine to half a cycle is enough to make them bijective functions. Formally, we write

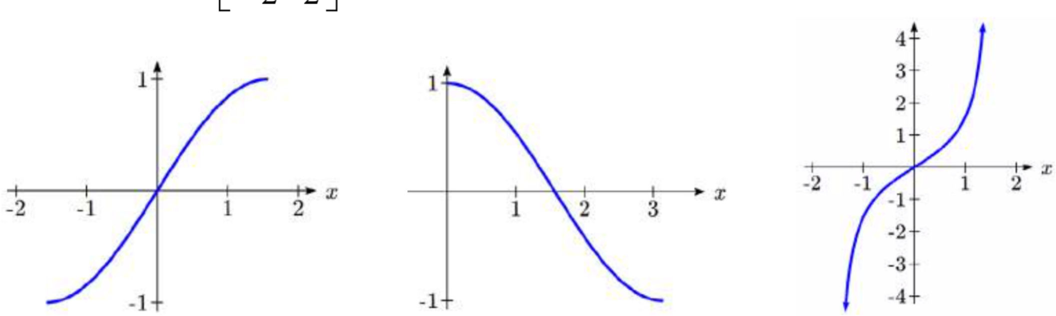
$$\cos\colon [0, \pi] \to [-1, 1]$$

$$\sin\colon [-\pi/2, \pi/2] \to [-1, 1]$$

We just need to limit tangent to one whole cycle (between two vertical asymptotes), so we write

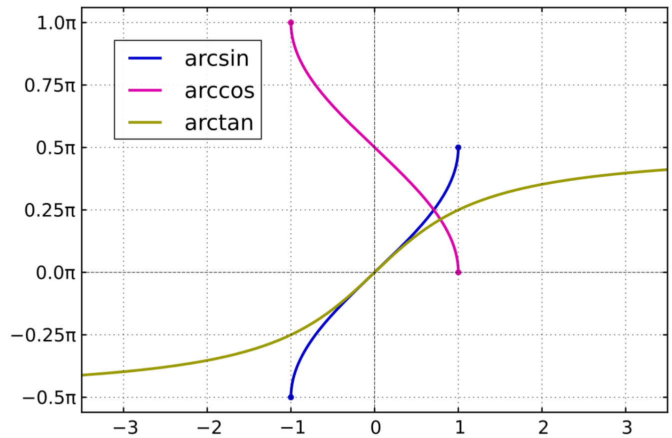
$$\tan\colon [-\pi/2, \pi/2] \to \mathbb{R}$$

Sine, limited to $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ Cosine, limited to $[0, \pi]$ Tangent, limited to $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$

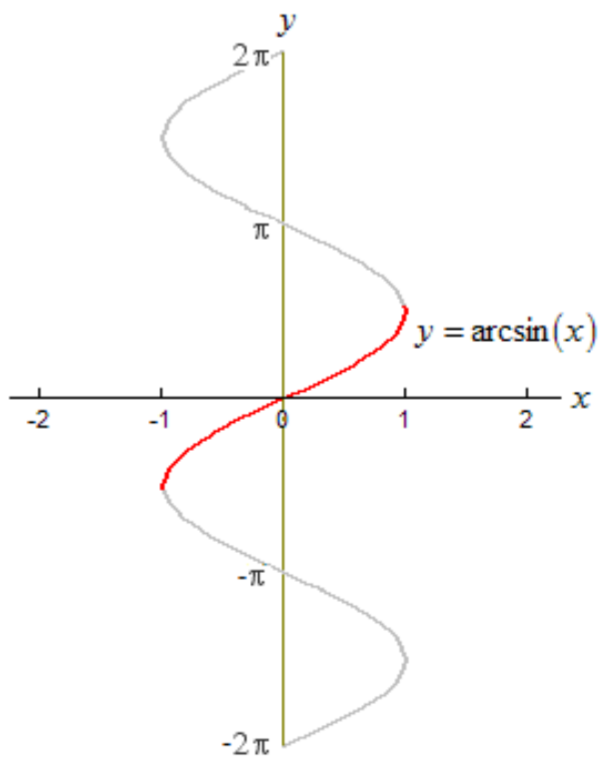


Finally, we can define the inverse functions of sine, cosine, and tangent by switching the domain and range as follows:

$$\arccos\colon [-1,1] \to [0,\pi]$$
$$\arcsin\colon [-1,1] \to [-\pi/2,\pi/2]$$
$$\arctan\colon \mathbb{R} \to [-\pi/2,\pi/2]$$



Observe that the graphs of the inverse functions look just like the original, except flipped on the side:



I again emphasize the importance of restricting the domains to make these functions bijective. As you can see above, arcsin(x) would fail the vertical line test if any more of the original sine wave was shaded red, meaning it would not be a well-defined function.

In practice, we use inverse trigonometric functions to find angles and other quantities in the argument of trig functions. for example, consider this equation:

$$A\sin(\omega(\theta+\phi))+5=2.7$$

If we know everything else, how do we determine what the angle θ is? First, isolate sin on one side of the equation:

$$\sin(\omega(\theta+\phi))=\frac{2.7-5}{A}$$

Now we take arcsin of both sides:

$$\arcsin(\sin(\omega(\theta+\phi)))=\arcsin\left(\frac{2.7-5}{A}\right)$$

Recall that any time you compose a function and its inverse you get the identity (i.e., the argument inside), so

$$\omega(\theta+\phi)=\arcsin\left(\frac{2.7-5}{A}\right)$$

Finally, we isolate θ on one side, allowing us to conclude

$$\theta=\frac{1}{\omega}\arcsin\left(\frac{2.7-5}{A}\right)-\phi$$

At this point, it will typically be necessary to use a calculator. If you get lucky, the answer could be on the unit circle. For instance,

$$\arcsin(\sqrt{2}/2)=45^\circ$$

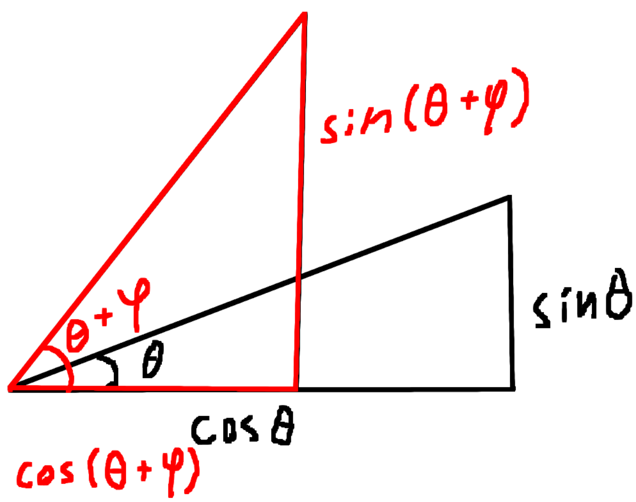
Inverse trigonometric functions are to regular trigonometric functions what logarithms are to exponents.

4.3 Proving Trigonometric Idenities

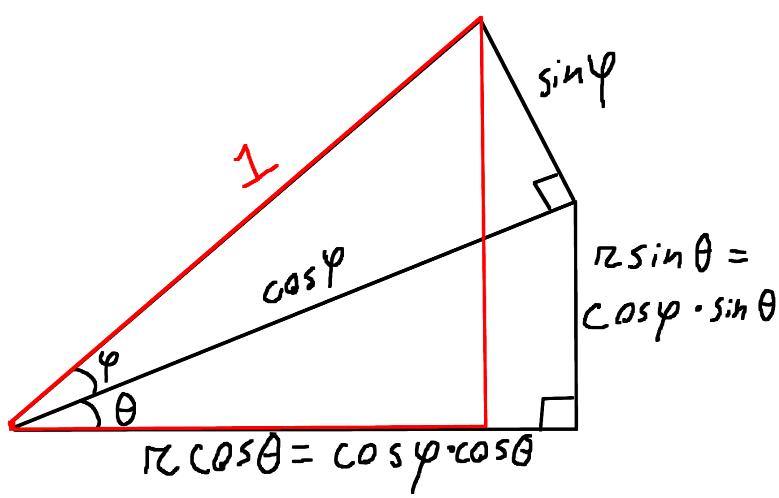
Working with trigonometric functions can be complicated, cumbersome, and confusing. Fortunately, there exist many **trigonometric identities** which allow us to simplify long trigonometric expressions, determine the value of trigonometric functions when the angle may not appear on the unit circle, and aid in our understanding of the geometry of many shapes and systems.

4.3.1 Angle Addition and Subtraction

The first identities provable to us are the **angle addition** identities. As the figure below would suggest, given two angles θ and φ, these identities allow us to determine the value of $\sin(\theta + \varphi)$ and others.



To determine an equation for $\sin(\theta + \varphi)$, we can use tools from geometry to find a formula in terms of triangles made from angles θ and φ on the unit cirlce. The first step is to draw two right triangles: one with angle θ and, on top of it, another with angle φ:



We let the triangle φ have hypotenuse $r = 1$ so we can superimpose the red triangle $\theta + \varphi$. But this means that the hypotenuse of the θ triangle is $r = \cos\varphi$. Accordingly, using our equations for the lengths of the triangles,

$$x = r\cos\theta$$

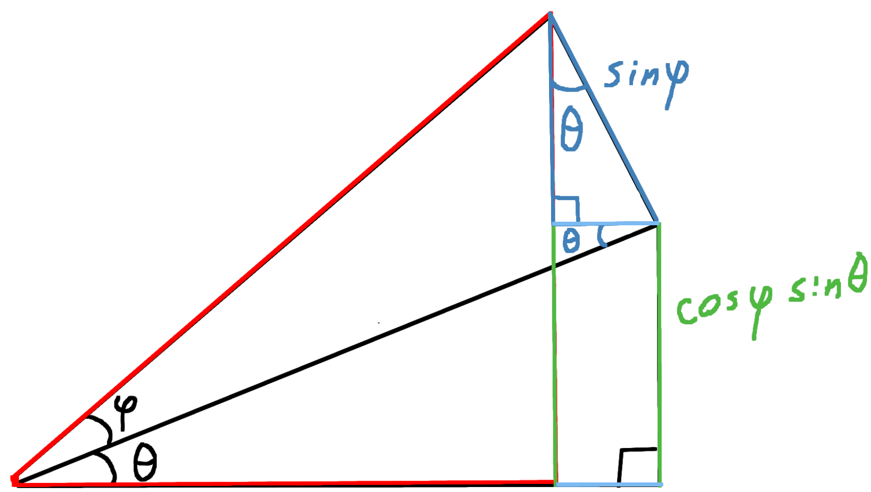
$$y = r\sin\theta$$

we plug in our hypotenuse to get

$$x = \cos\varphi\cos\theta$$

$$y = \cos\varphi\sin\theta$$

Now we need to make and solve a new triangle in the area outside the red $\theta + \varphi$ triangle to determine its sidelengths:



By making the blue triangle's base parallel with the base of the θ triangle, we accomplish some important things. First, we see that the green sections are equal length. Second, we use the Alternate Interior Angle Theorem from geometry to see that several blue angles are actually equivalent to θ. Last, we see that the sides in light blue should be the same length.

The final step is to find the sides of the blue triangle:

$$\cos\theta = \frac{x}{r} = \frac{x}{\sin\varphi}$$

Note that because θ has been rotated in the blue triangle, our "x" value here is actually the vertical side of the blue triangle. We conclude

$$x = \sin\varphi\cos\theta$$

Now we know the entire vertical side of the red $\theta + \varphi$ triangle. Putting it all together, we obtain the following angle addition identity:

$$\sin(\theta + \varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi$$

Now it is time to solve for the last side of the red triangle, which is given by $\cos(\theta + \varphi)$. Since we know the other two sides, you might be tempted to try the Pythagorean theorem, but there is an easier way. Simply solve for the light blue side now:

$$\sin\theta = \frac{y}{r} = \frac{y}{\sin\varphi}$$

Thus,

$$y = \sin\theta \sin\varphi$$

To get the length of the adjacent side of the red triangle, we simply need to subtract the light blue length from the base of the θ triangle, so we obtain

$$\cos(\theta + \varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi$$

Recall that $\sin\theta$ is an odd function, meaing that $\sin(-\theta) = -\sin\theta$. On the other hand, $\cos\theta$ is an even function, meaning $\cos(-\theta) = \cos\theta$. Using these facts, we can obtain the **angle subtraction** identities:

$$\sin(\theta - \varphi) = \sin\theta\cos(-\varphi) + \cos\theta\sin(-\varphi) = \sin\theta\cos\varphi - \cos\theta\sin\varphi$$

$$\cos(\theta - \varphi) = \cos\theta\cos(-\varphi) - \sin\theta\sin(-\varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi$$

4.3.2 Double Angle and Half Angle

The angle addition and subtraction identites are a great starting point for proving other helpful identities. To start, we will obtain the **double angle** identities:

$$\sin(2\theta) = \sin(\theta + \theta) = \sin\theta\cos\theta + \cos\theta\sin\theta = 2\sin\theta\cos\theta$$

$$\cos(2\theta) = \cos(\theta + \theta) = \cos\theta\cos\theta - \sin\theta\sin\theta = \cos^2\theta - \sin^2\theta$$

We can rewrite the result for $\cos(2\theta)$ using the Pythagorean theorem, replacing $\cos^2\theta$ with $1 - \sin^2\theta$. When we do, the result is

$$\cos(2\theta) = (1 - \sin^2\theta) - \sin^2\theta = 1 - 2\sin^2\theta$$

The double angle identities help us find the **half angle** identities. First, make the observation that

$$\cos\theta = \cos\left(2 \cdot \frac{\theta}{2}\right)$$

We can plug this into the double angle formula:

$$\cos\theta = 1 - 2\sin^2\left(\frac{\theta}{2}\right)$$

Now, simply solve for $\sin\frac{\theta}{2}$.

$$\sin\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{\cos\theta - 1}{2}}$$

To obtain the cosine half angle identity, start over and replace $\sin^2(\frac{\theta}{2})$ with $1-\cos^2(\frac{\theta}{2})$.

$$\cos\theta = 1 - 2\left(1 - \cos^2\left(\frac{\theta}{2}\right)\right) = 1 - 2 + 2\cos^2\left(\frac{\theta}{2}\right) = -1 + 2\cos^2\left(\frac{\theta}{2}\right)$$

Finally, rearrangment yields

$$\cos\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{1+\cos\theta}{2}}$$

A corollary of the half angle theorem is that $\sin\theta$ and $\cos\theta$ are separated by a 90° (i.e. $\pi/2$) phase angle:

$$\sin\left(\frac{\pi}{2} - \theta\right) = \sin\frac{\pi}{2}\cos\theta - \cos\frac{\pi}{2}\sin\theta = (1)\cos\theta - (0)\sin\theta = \cos\theta$$

$$\cos\left(\frac{\pi}{2} - \theta\right) = \cos\frac{\pi}{2}\cos\theta + \sin\frac{\pi}{2}\sin\theta = (0)\cos\theta + (1)\sin\theta = \sin\theta$$

4.3.3 Product to Sum

The angle addition identity also empowers us to learn the powerful **product to sum** identities:

$$\sin(\theta+\varphi) + \sin(\theta-\varphi) = (\sin\theta\cos\varphi + \cos\theta\sin\varphi) + (\sin\theta\cos\varphi - \cos\theta\sin\varphi) = 2\sin\theta\cos\varphi$$

Thus,

$$\sin\theta\cos\varphi = \frac{1}{2}\left[\sin(\theta+\varphi) + \sin(\theta-\varphi)\right]$$

The other identities of this kind follow analogously:

$$\sin(\theta+\varphi) - \sin(\theta-\varphi) = (\sin\theta\cos\varphi + \cos\theta\sin\varphi) - (\sin\theta\cos\varphi - \cos\theta\sin\varphi) = 2\cos\theta\sin\varphi$$

So,

$$\cos\theta\sin\varphi = \frac{1}{2}\left[\sin(\theta+\varphi) - \sin(\theta-\varphi)\right]$$

Likewise, we have

$$\cos\theta\cos\varphi = \frac{1}{2}\left[\cos(\theta+\varphi) + \cos(\theta-\varphi)\right]$$

and

$$\sin\theta\sin\varphi = \frac{1}{2}\left[\cos(\theta+\varphi) - \cos(\theta-\varphi)\right]$$

All of the identities above relating expressions with sine and cosine can easily be used to discover identities about tangent and other, less popular trigonometric functions. For example,

$$\cos\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{1+\cos\theta}{2}} \implies \frac{1}{\cos\left(\frac{\theta}{2}\right)} = \frac{1}{\pm\sqrt{\frac{1+\cos\theta}{2}}}$$

but we know that $1/\cos\theta = \sec\theta$, so

$$\sec\left(\frac{\theta}{2}\right) = \frac{1}{\pm\sqrt{\frac{1+\cos\theta}{2}}}$$

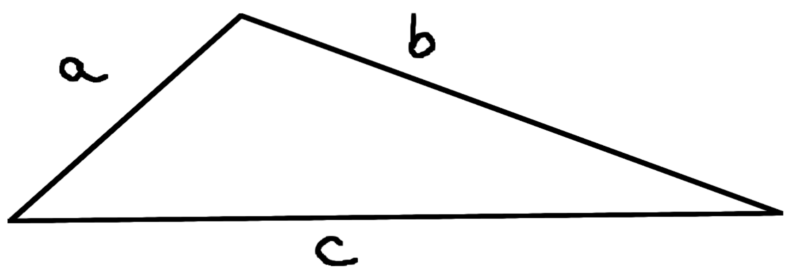
We can still simplify to make this a cleaner expression:

$$\sec\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{2}{1+\cos\theta}}$$

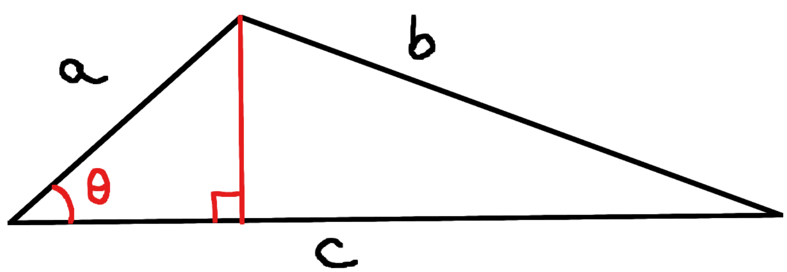
For now, we will supply most such identities in the reference table and leave their proofs as an exercise.

4.3.4 Law of Cosines

Suppose we have any triangle:



The Pythagorean theorem only works on right triangles. Let's try to find a way to relate the sides of any triangle. First, let's turn our triangle into two right triangles:



We don't know how side c was split up when we created two right triangles. To figure out, we need to determine how long the x side is on the triangle with hypotenuse a:

$$x = r\cos\theta$$

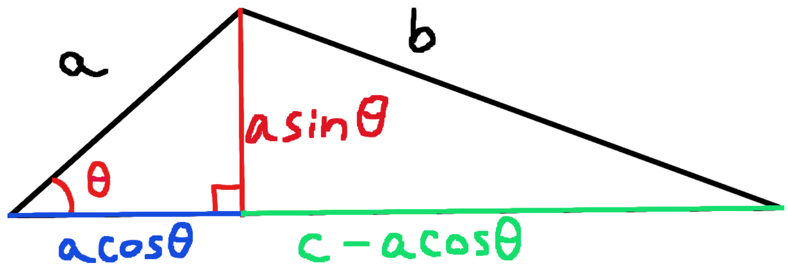
Since $r = a$, we know

$$x = a\cos\theta$$

Likewise, we know the red side is

$$y = a\sin\theta$$

Thus, we can relabel the sides like so:



We can make two equations using the Pythaogrean theorem:

$$a^2 \cos^2 \theta + a^2 \sin^2 \theta = a^2$$

$$(c - a \cos \theta)^2 + a^2 \sin^2 \theta = b^2$$

The first equation is just the normal Pythagorean theorem; it is not telling us anything we didn't know already. So, to make progress, we will focus on the second equation. Begin by "FOILing" out $(c - a \cos \theta)^2$.

$$(c^2 - 2ac \cos \theta + a^2 \cos^2 \theta) + a^2 \sin^2 \theta = b^2$$

Observe that we can regroup terms like this:

$$c^2 - 2ac \cos \theta + a^2(\cos^2 \theta + \sin^2 \theta) = b^2$$

By the Pythagorean theorem, $\cos^2 \theta + \sin^2 \theta = 1$, so

$$c^2 - 2ac \cos \theta + a^2 = b^2$$

This identity is what we call the **Law of Cosines**. In the special case where $\theta = 90°$, note that it reduces down to the regular Pythagorean theorem.

4.3.5 Law of Sines

(planned)

4.4 Applications of Trigonometry to Vector Algebra and Complex Numbers

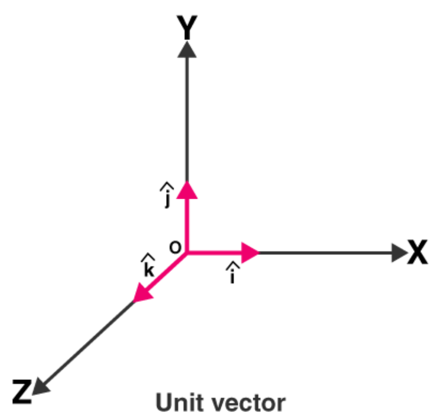
Recall that a **vector** is a $1 \times n$ or $n \times 1$ matrix (at least for our current purposes). As such, we can have vectors like

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{or} \quad \mathbf{w} = \begin{bmatrix} a & b & c \end{bmatrix}$$

where $\mathbf{v}$ is an example of a 3×1 **column vector** and $\mathbf{w}$ is an example of a 1×3 **row vector**. When we do not care as much about the specific orientation of a vector, we may instead write

$$\mathbf{v} = \langle x, y, z \rangle \quad \text{or} \quad \mathbf{w} = x\widehat{\mathbf{i}} + y\widehat{\mathbf{j}} + z\widehat{\mathbf{k}}$$

In this case, both $\mathbf{v}$ and $\mathbf{w}$ mean exactly the same thing. The notation for $\mathbf{v}$ is nothing very new; we just write the vector like a point, except we use pointed brackets to signify that it is a vector. In the case of $\mathbf{w}$, each direction is broken down to a single **unit vector**. These are just vectors of length 1 called $\widehat{\mathbf{i}}$, $\widehat{\mathbf{j}}$, and $\widehat{\mathbf{k}}$ ("i hat," "j hat," and "k hat") which point in the x, y, and z directions, respectively.



Specifically, in our new pointed bracket notation,

$$\widehat{\mathbf{i}} = \langle 1, 0, 0 \rangle$$

$$\widehat{\mathbf{j}} = \langle 0, 1, 0 \rangle$$

$$\widehat{\mathbf{k}} = \langle 0, 0, 1 \rangle$$

This means that when we have some scalar x multiplied by $\widehat{\mathbf{i}}$ we get

$$x\widehat{\mathbf{i}} = \langle x, 0, 0 \rangle$$

If we took the sum $x\widehat{\mathbf{i}} + y\widehat{\mathbf{j}}$ we would get

$$x\widehat{\mathbf{i}} + y\widehat{\mathbf{j}} = \langle x, 0, 0 \rangle + \langle 0, y, 0 \rangle = \langle x, y, 0 \rangle$$

The rule for adding vectors is fundamentally very similar to adding normal algebraic expressions. Observe:

$$(5x + 3y) + (2x + y) = 7x + 4y$$

$$(5\widehat{\mathbf{i}} + 3\widehat{\mathbf{j}}) + (2\widehat{\mathbf{i}} + \widehat{\mathbf{j}}) = 7\widehat{\mathbf{i}} + 4\widehat{\mathbf{j}}$$

So, if you like, you can think of the unit vectors as a "common factor" when performing vector addition and subtraction.

When we have a vector $\mathbf{v}$, it is often helpful to know its length, denoted by $\|\mathbf{v}\|$ and often called the **magnitude** of $\mathbf{v}$ or the **norm** of $\mathbf{v}$. If our vector $\mathbf{v}$ is given by

$$\mathbf{v} = \langle a, b, c \rangle$$

then we just need to use the distance formula to find $\|\mathbf{v}\|$, as follows:

$$\|\mathbf{v}\| = \sqrt{a^2 + b^2 + c^2}$$

Suppose we get a new vector $\widehat{\mathbf{v}}$ by dividing $\mathbf{v}$ by $\|\mathbf{v}\|$. What is the magnitude of $\widehat{\mathbf{v}}$?

$$\|\widehat{\mathbf{v}}\| = \sqrt{\frac{a^2}{\|\mathbf{v}\|^2} + \frac{b^2}{\|\mathbf{v}\|^2} + \frac{c^2}{\|\mathbf{v}\|^2}} = \sqrt{\frac{a^2 + b^2 + c^2}{\|\mathbf{v}\|^2}} = \sqrt{\frac{a^2 + b^2 + c^2}{a^2 + b^2 + c^2}} = 1$$

Anytime we divide a vector by its own length, we denote it $\widehat{\mathbf{v}}$. The vector $\widehat{\mathbf{v}}$ is a unit vector pointing in the same direction of $\mathbf{v}$ but has length 1.

We require one last tool before applying these facts to trigonometry. Recall that when we take the product of two matrices $\mathbf{A}_{n\times k}$ and $\mathbf{B}_{k\times m}$ we get a new matrix $\mathbf{C}_{n\times m}$. This is also true of vectors: if we take the matrix product of $\mathbf{v}_{n\times 1}$ and $\mathbf{w}_{1\times m}$ then we get a matrix $\mathbf{M}_{n\times m}$. However, there is another way to "multiply" vectors called the **dot product**, defined as follows:

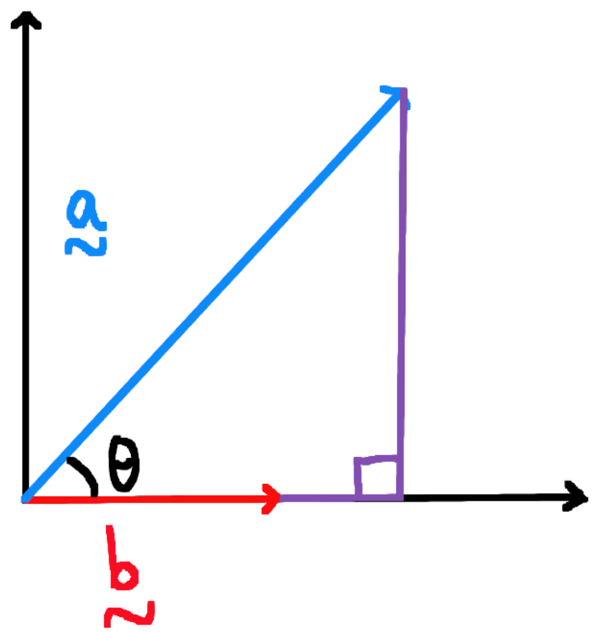
$$\mathbf{a} \cdot \mathbf{b} = \langle a_1, a_2 \rangle \cdot \langle b_1, b_2 \rangle = a_1 b_1 + a_2 b_2$$

For example, if $\mathbf{v} = \langle 1, 3, 5 \rangle$ and $\mathbf{w} = \langle 2, 4, -1 \rangle$, then

$$\mathbf{v} \cdot \mathbf{w} = \langle 1, 3, 5 \rangle \cdot \langle 2, 4, -1 \rangle = (1)(2) + (3)(4) + (5)(-1) = 2 + 12 - 5 = 9$$

The dot product is also called the **scalar product** because $\mathbf{a} \cdot \mathbf{b}$ is always a regular number, not a matrix.

Suppose we have two vectors **a** and **b**. Since vectors have length and direction, we can form a triangle from **a** and **b**:



Unfortunately, we seem to have a problem: the lengths of **a** and **b** may not always match up to allow us to make a right triangle. By going straight down from the tip of **a** in purple, we see that **b** is too short to be the base of a right triangle! What can we do? Is all hope lost?

No! There is hope yet. We can apply our knowledge of trigonometry to fix this problem. To start, if the vector **a** is our hypotenuse, then we know that $r = \|\mathbf{a}\|$. Further, we know that $x = r\cos\theta$. So,

$$x = \|\mathbf{a}\|\cos\theta$$

In this case, x is how long we *need* our base to be for us to make a right triangle. Since the base is drawn in the direction of **b**, if we find a new vector, it should be a "streched" version of **b**. First, find the unit vector $\widehat{\mathbf{b}}$ which points in the same direction as **b**:

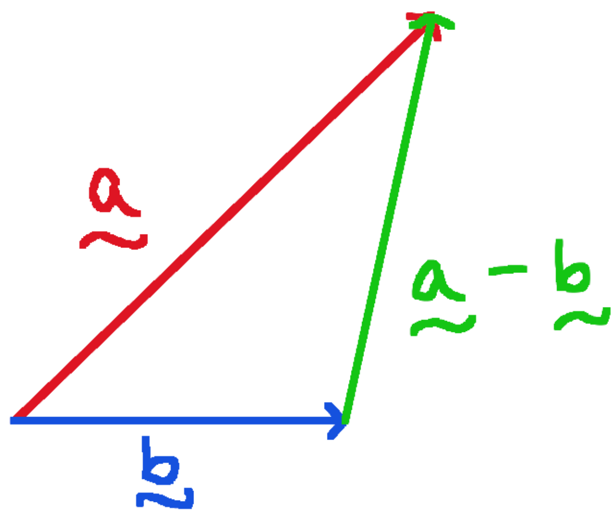
$$\widehat{\mathbf{b}} = \frac{\mathbf{b}}{\|\mathbf{b}\|}$$

Since $\|\widehat{\mathbf{b}}\| = 1$, we just need to multiply by our new desired length to get the full side of the triangle **d**:

$$\mathbf{d} = (\|\mathbf{a}\|\cos\theta)\,\widehat{\mathbf{b}} = \left(\frac{\|\mathbf{a}\|}{\|\mathbf{b}\|}\cos\theta\right)\mathbf{b}$$

Of course, at this point, we could solve for the length of the last side. We could also use the rotation matrix to find a vector **c** which makes a right angle with **b**. Then we could scale **c** in the exact same way we did to **b**: first, find $\widehat{\mathbf{c}}$ by finding $\mathbf{c}/\|\mathbf{c}\|$; then multiply $\widehat{\mathbf{c}}$ by whatever constant you want to be its new length.

What if we do not want a right triangle? Then we obtain something like this:



In this case, simply making the triangle is much easier. Suppose that **b** is pointing completely in the x-direction. Then when we subtract **b** from **a**, we only reduce the horizontal distance of **a**, not its height. This means that $\mathbf{a} - \mathbf{b}$ will always be the perfect size to form a triangle with **a** and **b**.

As before, we know that the length of the red side is $\|\mathbf{a}\|$, the length of the blue side is $\|\mathbf{b}\|$, and the green side is $\|\mathbf{a} - \mathbf{b}\|$. Given this information, we could plug our side lengths into the Law of Cosines.

$$c^2 - 2ac\cos\theta + a^2 = b^2 \implies \|\mathbf{a}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta + \|\mathbf{b}\|^2 = \|\mathbf{a} - \mathbf{b}\|^2$$

Since we know the formula for distance, we can plug in the components of **a** and **b** to this equation wherever we have $\|\mathbf{a}\|^2$ or $\|\mathbf{b}\|^2$. We will leave $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$ alone since there is not an easy way to get those out of the square root.

$$(a_x^2 + a_y^2) - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta + (b_x^2 + b_y^2) = \left[(a_x - b_x)^2 + (a_y - b_y)^2\right]$$

FOIL the right hand side.

$$(a_x^2 + a_y^2) - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta + (b_x^2 + b_y^2) = (a_x^2 - 2a_xb_x + b_x^2) + (a_y^2 - 2a_yb_y + b_y^2)$$

Regroup your terms on the right hand side like so:

$$(a_x^2 + a_y^2) - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta + (b_x^2 + b_y^2) = (a_x^2 + a_y^2) - 2(a_xb_x + a_yb_y) + (b_x^2 + b_y^2)$$

Notice that you have $(a_x^2 + a_y^2)$ and $(b_x^2 + b_y^2)$ on both sides. Simplify by subtracting from both sides.

$$-2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta = -2(a_xb_x + a_yb_y)$$

Divide by -2 and also make the observation that $a_x b_x + a_y b_y = \mathbf{a} \cdot \mathbf{b}$. We conclude

$$\|\mathbf{a}\| \|\mathbf{b}\| \cos\theta = \mathbf{a} \cdot \mathbf{b}$$

This is the **geometric definition of the dot product**. The θ in this expression is the angle between the two vectors. By manipulating a little more, we see that

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

So, if we ever need the angle θ between two vectors, it is given by

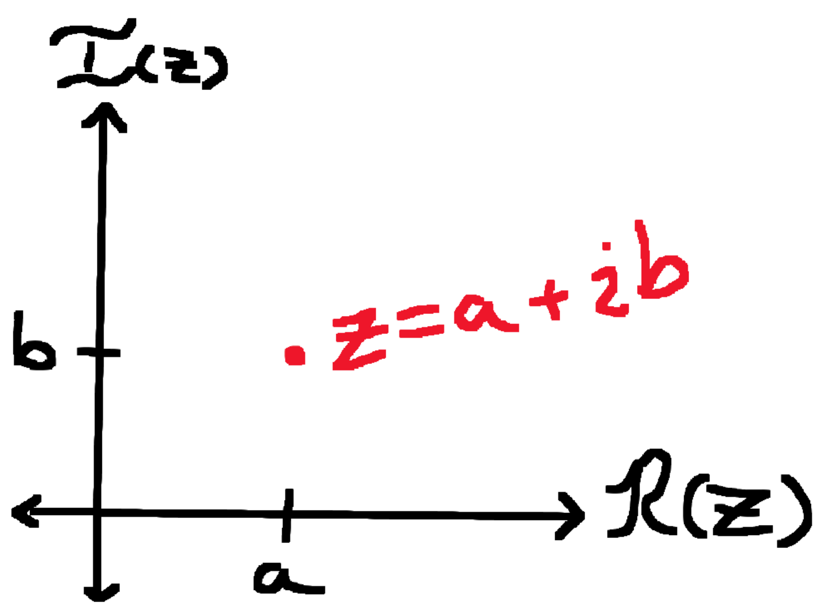
$$\theta = \arccos\left(\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}\right)$$

There exist other vector operations which aid in trigonometric calculations and proofs, but we will learn about those in greater detail later.

Moving on, recall that every complex number $z \in \mathbb{C}$ can be written

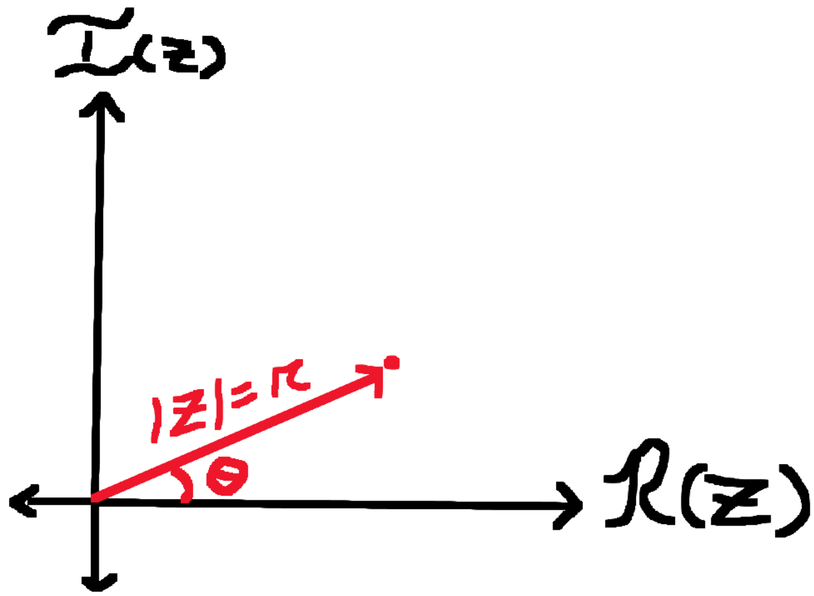
$$z = a + ib$$

where $a, b \in \mathbb{R}$ are real numbers, a is called the **real component** of z, and b is called the **imaginary component** of z. Instead of plotting individual complex numbers on a number line, we instead need to use an entire coordinate plane which breaks up the real and imaginary components.



On occasion, people will use the notation $\Re(z)$ and $\Im(z)$ to denote the real and imaginary parts of z, respectively.

It is sometimes more favorable to describe complex numbers using angles and distances instead of using the real and imaginary components.



This should remind you of the hypotenuse of a triangle! This time, we have

$$\Re(z) = r\cos\theta$$

$$\Im(z) = r\sin\theta$$

Putting it all together, we can write any complex number as

$$z = r(\cos\theta + i\sin\theta)$$

As shorthand for "**C**osine plus **i** times **S**ine," we sometimes write cis θ instead. Using this notation, we have

$$z = r \text{ cis } \theta$$

An extremely important fact about cis θ (which we are unfortunately unable to prove yet) is that

$$e^{i\theta} = \text{cis } \theta$$

That is,

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

This is called **Euler's equation**. When we plug in $\theta = \pi$,

$$e^{i\pi} = \cos\pi + i\sin\pi = -1 + 0$$

rearrangement gives the famous **Euler identity**:

$$e^{i\pi} + 1 = 0$$

Many people consider the Euler identity to be one of the most beautiful expressions in mathematics since it relates five seemingly separate constants from various fields of math: e, i, π, 1, and 0. Euler himself thought that such a simple relation of these five constants was so serendipitious that it proved the existence of God.

Given the Euler equation, we can now write any complex number as

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

This notation is far more popular than $\operatorname{cis}\theta$, and for good reason! This allows us to use all the rules we know from exponent algebra to derive new identities. For example,

$$z = e^{i\theta}$$

$$w = e^{i\phi}$$

$$z \cdot w = e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)} = \cos(\theta+\phi) + i\sin(\theta+\phi)$$

This next identity is of critical importance to trigonometry. First, we know

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

But when we plug in $-\theta$, we get

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

because $\cos\theta$ is even and $\sin\theta$ is odd. So, from the first equation,

$$\cos\theta = e^{i\theta} - i\sin\theta \qquad \sin\theta = \frac{e^{i\theta} - \cos\theta}{i} = -i(e^{i\theta} - \cos\theta)$$

and from the second equation,

$$\cos\theta = e^{-i\theta} + i\sin\theta \qquad \sin\theta = \frac{e^{-i\theta} - \cos\theta}{-i} = i(e^{-i\theta} - \cos\theta)$$

Now we combine them:

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

Thus,

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Using these definitions of the sine and cosine functions, you can prove trigonometric identities without ever drawing a triangle; instead, substitute the expressions above in to your equation, FOIL and apply exponent algebra, and you are done. The most tricky part is going back from the complex version to normal sine and cosine.

4.5 Table of Important Identities

Name	Property	Example
Pythagorean theorem	$a^2 + b^2 = c^2$	Given a right triangle with sides $a = 3$ and $b = 4$, we have $$3^2 + 4^2 = c^2 \implies c = 5$$
Law of cosines	Generalized Pythagorean theorem; works on all triangles. $$c^2 - 2ac\cos\theta + a^2 = b^2$$	If $\theta = 45°$ and $a = 2$, $b = 3$, then $$c^2 = 2ac\cos\theta + b^2 - a^2$$ $$= 2(2)(3)(\sqrt{2}/2) + (3)^2 - (2)^2$$ $$= 6\sqrt{2} + 5$$
Angle addition theorems	$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi$ $\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi$	$\sin 75° = \sin 30° \cos 45° + \cos 30° \cos 45°$
Angle subtraction theorems	$\sin(\theta-\varphi) = \sin\theta\cos\varphi - \cos\theta\sin\varphi$ $\cos(\theta-\varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi$	$\sin 15° = \sin 45° \cos 30° - \cos 45° \cos 30°$
Double angle theorems	$\sin(2\theta) = 2\sin\theta\cos\theta$ $\cos(2\theta) = \cos^2\theta - \sin^2\theta$ $\cos(2\theta) = 1 - 2\sin^2\theta$	$\cos 120° = 1 - 2\sin^2 60°$
Half angle theorems	$\cos\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{1+\cos\theta}{2}}$ $\sin\left(\frac{\theta}{2}\right) = \pm\sqrt{\frac{\cos\theta - 1}{2}}$	$\cos 15° = +\sqrt{\frac{1+\cos 30°}{2}}$

Name	Property	Example
Geometric definition of dot product	$\mathbf{a}\cdot\mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|\cos\theta$ $\cos\theta = \dfrac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$	$\mathbf{v}\cdot\mathbf{w} = (5)(4)\cos 80^\circ$
Angle betwen two vectors	$\theta = \arccos\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}\right)$	$\theta = \arccos\left(\dfrac{(1)(2)+(0)(0)}{(1)(2)}\right)$ $= \arccos(1) = 0$
Polar form of complex numbers	$z = r(\cos\theta + i\sin\theta) = r \text{ cis } \theta$	$z = 2(\cos 45^\circ + i\sin 45^\circ) = \sqrt{2} + i\sqrt{2}$
Euler's equation	$e^{i\theta} = \cos\theta + i\sin\theta = \text{cis } \theta$	$z = e^{i45^\circ} = \cos 45^\circ + i\sin 45^\circ$ $= \sqrt{2}/2 + i\sqrt{2}/2$
Euler's identity	$e^{i\pi} + 1 = 0$	Not actually very useful; mostly a fun trivia fact.
Complex definition of sine	$\sin\theta = \dfrac{e^{i\theta} - e^{-i\theta}}{2i}$	$i\sin 30^\circ = i\left(\dfrac{e^{i30^\circ} - e^{-i30^\circ}}{2i}\right)$ $= \dfrac{e^{i30^\circ} - e^{-i30^\circ}}{2}$
Complex definition of cosine	$\cos\theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2}$	$\cos^2\theta = \left(\dfrac{e^{i\theta} + e^{-i\theta}}{2}\right)^2$ $= \dfrac{e^{2i\theta} + 2 + e^{-2i\theta}}{4}$

Chapter 5

Single Variable Calculus

5.1 What are Real Numbers?

5.1.1 What are Natural Numbers? Integers? Rational Numbers?

Until now, we have used the set of real numbers $\mathbb{R}$ without raising any questions or concerns about the set's history, structure, or properties. In this section, we will begin to address some such ideas.

First, we should consider where sets of numbers come from in the first place. As we discussed in the Geometry chapter, mathematics is based on a number of **axioms**, which are statements assumed to be true within some logical framework, no proof required. That is, unless we make the explicit assumption that "sets" or "numbers" exist in mathematics, we must either (a) prove that they exist based off of other axioms, (b) discover no such thing actually exists with the axioms available, or (c) fail to discover an answer one way or another.

"It turns out" that assuming the existence of sets (among some other conditions) is sufficient for defining the natural numbers $\mathbb{N}$. First, we start out with the **empty set**:

$$\emptyset = \{\}$$

This is the set with nothing inside of it. We say that the number $0 = \emptyset$. The **successor** of 0 is then defined as

$$S(0) = 1 = \{0\}$$

This is the same as

$$1 = \{\emptyset\}$$

so the number 1 is the set containing the empty set. Then the successor of 1 is

$$S(1) = 2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$$

so the number 2 is the set containing both the empty set and the set cotaining the empty set. The important thing is that the set representing the number 2 actually has two things inside of it. Likewise, the set representing 0 is empty (zero things inside) and the set representing

1 has one element. This same pattern holds for any natural number $n \in \mathbb{N}$. It is empty sets all the way up!

Thanks to the **axiom of infinity**, we can assert that there exists a set X for which $S(y) \in X$ as long as $y \in X$. Suppose there is a maximum number $m \in X$ which is bigger than every other element of X. This is impossible because $S(m) = m + 1$ must also be in X, a contradiction! Therefore, X must be an infinite set.

We typically take X in the example above to be the natural numbers $\mathbb{N}$. Finally, we have an infinite set of natural numbers, our first step towards defining $\mathbb{R}$.

Next, we need the set of integers $\mathbb{Z}$. Omitting the details for now, we obtain integers by taking the Cartesian product $\mathbb{N} \times \mathbb{N}$ and defining a certain equivalence relation $(a, b) \sim (c, d)$ between ordered pairs of natural numbers. Each integer is identified with an equivalence class. The rational numbers $\mathbb{Q}$ are obtained similarly from defining an equivalence relation on $\mathbb{Z} \times \mathbb{Z}$. The end result, as you are no doubt aware, is the set

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z},\ b \neq 0 \right\}$$

5.1.2 Why do we need $\mathbb{R}$?

Why don't we stop here? What is so unsatisfactory about the rational numbers that we need to go even further to real numbers? We can approximate any decimal by a ratio of two integers as close as we want; how do real numbers help us with anything?

Theorem 5.1. *The solution to the equation $x^2 = 2$ is not a rational number.*

Proof. Suppose for contradiction that $\sqrt{2} \in \mathbb{Q}$. Then $\sqrt{2} = a/b$ for some pair $a, b \in \mathbb{Z}$, where a and b have no common factors except 1. Given this information, we have

$$2 = \frac{a^2}{b^2} \iff 2b^2 = a^2$$

This shows that a^2 is an even number. This implies that a is also an even number, because an odd number squared is always odd:

$$(2n+1)(2n+1) = 4n^2 + 4n + 1 = 2(2n^2 + 2n) + 1 = 2k + 1$$

where k is some integer. On the other hand, an even number squared is always even:

$$(2n)(2n) = 4n^2 = 2(2n^2) = 2k$$

where k is, again, some integer. Thus,

$$a = 2k$$

When we plug this back into a previous equation, we get

$$2b^2 = (2k)^2 = 4k^2 \iff b^2 = 2k^2$$

By the same reasoning as before, this implies that b is an even number. But this means that a and b have a common factor of 2, a contradiction!

Therefore, we conclude that $\sqrt{2}$ is not an element of $\mathbb{Q}$. Q.E.D.

The desire to solve equations like $x^2 = 2$ is thousands of years old, dating back to the Pythagoreans. However, for a long time, the idea of *irrational* numbers remained controversial. Since it is impossible to write an irrational number as the ratio of two integers, critics have dismissed irrational numbers for lacking the apparent elegance and structure of rational numbers. Some mathematicians, many of whom promote alternate axioms to form the foundations of math, continue to reject the existence of irrational numbers and, often, infinite sets entirely.

5.1.3 Properties of $\mathbb{R}$

We are finally prepared to talk a bit about the construction of the real numbers. In practice, obtaining the real numbers tends to be a bit more involved than merely adding all the irrational numbers to the rationals. These are typically covered in detail during an advanced course in Analysis (e.g., Real Analysis), but we may return to this point later after gaining some tools from sequences and limits.

Qualitatively speaking, real numbers have all the "gaps" between rational numbers filled with irrational numbers. It can be shown that the set of irrational numbers is a "bigger infinity" than the set of rational numbers; however, the rational numbers are still **dense** in the real numbers, meaning that for every $x \in \mathbb{R}$ and $\epsilon > 0$, the interval $(x-\epsilon, x+\epsilon)$ contains rational numbers, no matter how small ϵ gets.

Algebraically speaking, the real numbers equipped with addition and multiplication, denoted $(\mathbb{R}, +, \cdot)$, form a **field**, meaning they obey all the **field axioms**, as follows:

1. (Associative property of addition): $x + (y + z) = (x + y) + z$

2. (Associative propety of multiplication): $x(yz) = (xy)z$

3. (Commutative property of addition): $x + y = y + x$

4. (Commutative property of multiplication): $xy = yx$

5. (Existence of additive identity): There exists $0 \in \mathbb{R}$ for which $0 + x == x$

6. (Existence of multiplicative identity): There exists $1 \in \mathbb{R}$ for which $1 \cdot x = x$

7. (Existence of additive inverse): For each $x \in \mathbb{R}$, there exists $-x \in \mathbb{R}$ for which $x+(-x) = 0$

8. (Existence of multiplicative inverse): For each $x \in \mathbb{R}$, there exists $x^{-1} \in \mathbb{R}$ for which $x \cdot x^{-1} = 1$ $(x, x^{-1} \neq 0)$

9. (Distributive property of multiplication over addition): $x(y + z) = xy + xz$

10. (Distinct additive and multiplicative identities): $0 \neq 1$

When we say that a set is "equipped with" an operation like addition or multiplication, this implies some additional important properties. Without loss of generality, we shall consider the set $\mathbb{R}$ with the operation of addition $+$. Formally speaking, addition (and other algebraic operators) is a function which takes two (real) numbers in domain and maps to just one real number in the range. Thus, we can say addition is a function

$$\text{sum}\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

defined by $\text{sum}(a, b) = a + b$. This leads us to our operator properties:

1. (Closure of $\mathbb{R}$ under addition): If $a, b \in \mathbb{R}$, then $\text{sum}(a, b) \in \mathbb{R}$

2. (Well-defined): The function $\text{sum}(a, b)$ passes the vertical line test and is otherwis a singular valued function

3. (Defined for all $a, b \in \mathbb{R}$): The function $\text{sum}(a, b)$ is defined for any $a, b \in \mathbb{R}$; no vertical asymptotes!

The most helpful property among all of these is closure. It tells you that you will never leave the real numbers (or whatever set and operations you are working with) when using basic operations. It is obvious that addition works perfectly fine for all real numbers, but we would have trouble using a square root operator to make a field because it would return imaginary numbers when the argument is negative, violating closure.

The properties above are not unique to the real numbers; many other sets aside from the real numbers can form a field. The **completeness** of the real numbers is what truly sets it apart from other sets:

- (Least upper bound) Every nonempty set A of real numbers bounded above has a least upper bound.

The details of this point are beyond the scope of this section, but we can begin to understand what this means by contrasting with a field that lacks the least upper bound property, such as $\mathbb{Q}$.

First, we need a nonempty subset of $\mathbb{Q}$. For our current purposes, we will use

$$A = \{r \in \mathbb{Q} \mid r < \sqrt{2}\}$$

This set has many **upper bounds** - infinitely many, in fact. A number is an upper bound of a set if it is greater than or equal to every element of the set, so 3, 4, 5, 10/3, and many others are all upper bounds of the set A of rational numbers above.

A number is a **least upper bound** if it is the smallest upper bound possible for a set. So what is the least upper bound of A? It turns out, A doesn't have one; if we said A was a subset of $\mathbb{R}$ (which it is), then the least upper bound would be $\sqrt{2}$ since it is both (a) bigger than every element of A and (b) smaller than every other upper bound. However, we proved that $\sqrt{2}$ is not a rational number, so if we restrict ourselves to working just in $\mathbb{Q}$, then we can always pick smaller and smaller rational upper bounds that get closer and closer to $\sqrt{2}$ but never actually get there $(1, 1.4, 1.41, 1.414, \dots)$.

5.2 Sequences and Limits

5.2.1 Limit of a Sequence

A **sequence** is any function whose domain is $\mathbb{N}$. Typically, we write sequences as (a_n) instead of $f(n)$, but both can mean the same thing. The following is an example of a sequence:

$$(a_n)\colon \mathbb{N} \to \mathbb{R} \qquad a_n = \frac{1}{n}$$

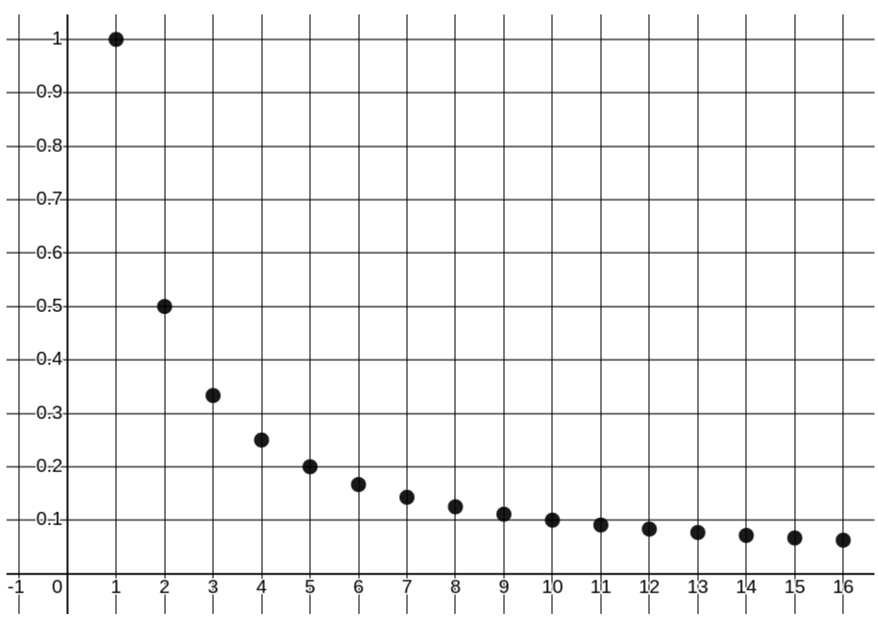
We use the parantheses notation for a sequence to signify that it is like an infinitely long list. In general, we can write

$$(a_1, a_2, a_3, \dots)$$

where the subscript represents the function argument ($f(1)$ and a_1 are analogous concepts). For the example sequence above, we can write

$$(a_n) = \left(1, \frac{1}{2}, \frac{1}{3}, \cdots\right)$$

We can plot sequences. This is our example from above:



We know from algebra that this function has a **horizontal asymptote** at $y = 0$, meaning the function will get closer and closer to $y = 0$ as n gets bigger, but it will never cross. In the language of calculus, this is the **limit** of the sequence (a_n), denoted by $\lim(a_n)$ or $\lim\limits_{n\to\infty}(a_n)$. The formal definition of a limit is as follows:

Definition 5.1 (Limit of a sequence)**.** *Let $\varepsilon > 0$ be arbitrary. If (a_n) is a sequence of real numbers, $L \in \mathbb{R}$ is the* **limit** *of (a_n) if there exists an $N \in \mathbb{N}$ for which $n > N$ implies*

$$|a_n - L| < \varepsilon$$

In simple words, this is telling us that the difference between a sequence and its limit, represented by $|a_n - L|$, can get as small as we want it to, represented by ε. We can make the function a_n as close to L as we want to, such that the difference between a_n and L can get as close to 0 as we want.

Based on what we learned from looking at the graph of (a_n), we can rigorously check that the limit is $L = 0$. First, we need a formula to find entries in the sequence that are smaller than any possible $\varepsilon > 0$. Specifically, we need to satisfy the following inequality:

$$|a_n - L| < \varepsilon$$

We are proposing that $L = 0$. We know that, in general, $a_n = 1/n$. Now we just need to plug them in and do some algebra.

$$|1/n - (0)| = |1/n| < \varepsilon$$

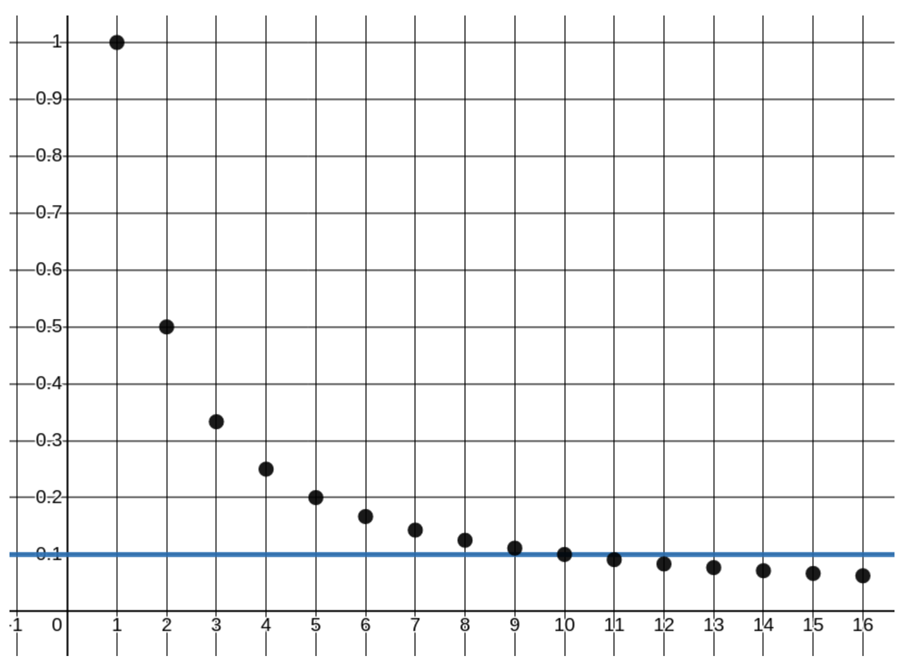
We know that $1/n$ is always positive, so we can get rid of the absolute value.

$$1/n < \varepsilon$$

Multiply both sides by n and divide both sides by ε.

$$1/\varepsilon < n$$

This inequality tells us how to find entries in the sequence that are smaller than whatever we choose for ε. For example, suppose $\varepsilon = 0.1$. Then we can graph it as a horizontal line on the scatterplot like so:



Notice that $\varepsilon = 0.1 = a_{10}$. This matches our inequality, which says

$$1/\varepsilon = 1/0.1 = 10 < n$$

This is saying that the sequence will be less than 0.1 distance away from the limit (in this case, $L = 0$) anytime the function argument n is bigger than 10, just like what we see on the graph. The last step is using the formula we found to prove $L = 0$. That is done like so:

I propose that $\lim_{n\to\infty}(a_n) = 0$.

Proof. Let $\varepsilon > 0$ be arbitrary, and let $N = 1/\varepsilon$. If $n > N$, then

$$1/\varepsilon < n \implies 1/n < \varepsilon$$

With some manipulation, we can show

$$1/n < \varepsilon \implies |1/n - 0| < \varepsilon \implies |a_n - L| < \varepsilon$$

Thus, by the definition of a limit, it must be the case that $\lim_{n\to\infty}(a_n) = 0$. Q.E.D.

After doing enough prep work, we always write something extremely similar to the boxed text to formally prove a limit.

A sequence is called **convergent** or is said to **converge** if it has a limit inside of its domain. On the other hand, a sequence is **divergent** if it does not converge. Two common ways sequences of real numbers can **diverge** is if the function approaches $\pm\infty$ or if the sequence never "settles" on a single limit. We will explore these ideas with a couple examples.

Example. Let (b_n) be a sequence defined by $b_n = n$. When we write this out list-wise, we see that the sequence is constantly increasing:

$$(b_n) = (1, 2, 3, 4, \dots)$$

This function can be any possible natural number, and natural numbers can get as big as we want. Therefore, we say that this sequence "approaches" ∞. This is not a real number, so the sequence (b_n) diverges.

This same idea applies to many functions we are familiar with. For instance, if $a_n = n^2$, then the function would grow even faster towards ∞ and also diverge.

Example. Let (c_n) be a sequence defined by $c_n = (-1)^n$. When we write this out list-wise, we see an alternating pattern:

$$(c_n) = (-1, 1, -1, 1, -1, 1, \dots)$$

This sequence diverges because it **alternates** between -1 and 1.

(c_n) is not the only alternating sequence. For example, $(0, 2, 0, 2, 0, 2, \dots)$ is another alternating sequence. Loosely speaking, since these sequences bounce between values and do not "settle down" like an asymptote, it is impossible to find a limit. This principle extends to a lot of other sequences, but alternating sequences are the easiest example.

The idea of an alternating sequence like (c_n) above motivates the following important theorem.

Theorem 5.2. *The limit of a sequence is unique.*

Many sequences diverge and have no limit, but no sequence has more than one limit.

5.2.2 Table of Common Sequences and Limits

Sequence	Limit as $n \to \infty$
$a_n = \frac{1}{n}$	$\lim 1/n = 0$
$a_n = c$ where c is a constant	$\lim c = c$
$a_n = (-1)^n$ (an alternating sequence)	No limit (diverges); applies to any alternating sequence
$a_n = n$	∞ (diverges)
$a_n = e^{-n}$	$\lim e^{-n} = 0$ (same applies for a lot of exponential functions, not just base e)

5.2.3 Solving Limits of Sequences with the Algebraic Limit Theorem

The formal approach of finding limits is somewhat difficult, especially when we want to prove a more complicated sequence has a certain limit. Fortunately, we can break big limits into smaller ones that tend to be either (a) easier to solve or (b) the same as a limit we already know. This is thanks to the following theorem:

Theorem 5.3 (Algebraic Limit Theorem)**.** *Let (a_n) and (b_n) be convergent sequences, such that $\lim(a_n) = a$ and $\lim(b_n) = b$ for some $a, b \in \mathbb{R}$. Also let $c \in \mathbb{R}$ be a constant. Then the following are true:*

- $\lim(c \cdot a_n) = c \cdot \lim(a_n) = c \cdot a$
- $\lim(a_n + b_n) = \lim(a_n) + \lim(b_n) = a + b$
- $\lim(a_n \cdot b_n) = \lim(a_n) \cdot \lim(b_n) = a \cdot b$
- $\lim(a_n/b_n) = \lim(a_n)/\lim(b_n) = a/b$ *(as long as $b \neq 0$)*

To demonstrate the power of the algebraic limit theorem, we will consider an example:

Let (a_n) be a sequence defined by $a_n = \frac{2n+2}{n}$. What is $\lim(a_n)$?

Example. Before attempting to figure out what the limit is, we should algebraically simplify the expression:

$$\lim_{n\to\infty} \frac{2n+2}{n} = \lim_{n\to\infty} 2 + \frac{2}{n} = \lim_{n\to\infty} 2(1+1/n)$$

By the algebraic limit theorem, we can take the factor of 2 out front of the limit:

$$\lim_{n\to\infty} 2(1+1/n) = 2 \cdot \lim_{n\to\infty} (1+1/n)$$

By the algebraic limit theorem, we can break this problem up into two easier ones:

$$2 \cdot \lim_{n\to\infty} (1+1/n) = 2\left(\lim_{n\to\infty} 1/n\right) + 2\left(\lim_{n\to\infty} 1\right)$$

We know what both of these limits are from the Table of Common Sequences and Limits, so we can just plug in the correct answers.

$$2\left(\lim_{n\to\infty} 1/n\right) + 2\left(\lim_{n\to\infty} 1\right) = 2(0) + 2(1) = 2$$

Therefore, the limit is 2.

5.2.4 Cauchy Sequences and Solving Limits of Sequences with a Computer

This section, our interest is focused on **Cauchy** sequences.

Definition 5.2 (Cauchy sequences)**.** *Let $\varepsilon > 0$ be arbitrary. A sequence (a_n) is called* **Cauchy** *if there exists an $N \in \mathbb{N}$ for which $n, m > N$ implies*

$$|a_n - a_m| < \varepsilon$$

In plain English, a sequence is Cauchy if its terms can get as close as possible to each other. The **completeness** of $\mathbb{R}$ alluded to in the first section gives Cauchy sequences a very helpful property:

Theorem 5.4. *A sequence in $\mathbb{R}$ is convergent if and only if it is Cauchy.*

The fact that any Cauchy sequence of real numbers has a limit is extremely helpful in computer programming. There are various instances where we need to approximate a limit on a computer; we can choose to stop when the terms of the sequence get "close enough" to each other. In this manner, we set ε as some number (usually very small, like 0.000001) and see if the difference between entries in a sequence is ever smaller than ε.

Our motivating example below is using a computer program to approximate the value of Euler's number e, which is analytically defined as

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \equiv e$$

```
#This code approximates the decimal value of Euler's number,
#an irrational number

begin program
#Define variables
integer :: max, n, count
real :: epsilon, a, b

#User specifices the maximum number of entries of the sequence
#the computer should try to generate. It is impossible to actually
#go to infinity, so we need to stop somewhere.
read n

#User specifies the ``convergence criterion," AKA
#how close two entries in the sequence need to be before
#it is considered ``converged" by the computer
read epsilon

#Generate the sequence. Generally, a stands for the n-th term of the
#sequence and b stands for the n+1-th term of the sequence.
count = 0
for n = 1, max:
count = count + 1
a = ( 1 + (1/n) )**n
b = ( 1 + (1/(n+1)) )**(n+1)

if |a - b| < epsilon, then break

end for

#Write out the decimal approximation of e
print b

#Write out how many steps it actually took
print count

end program
```

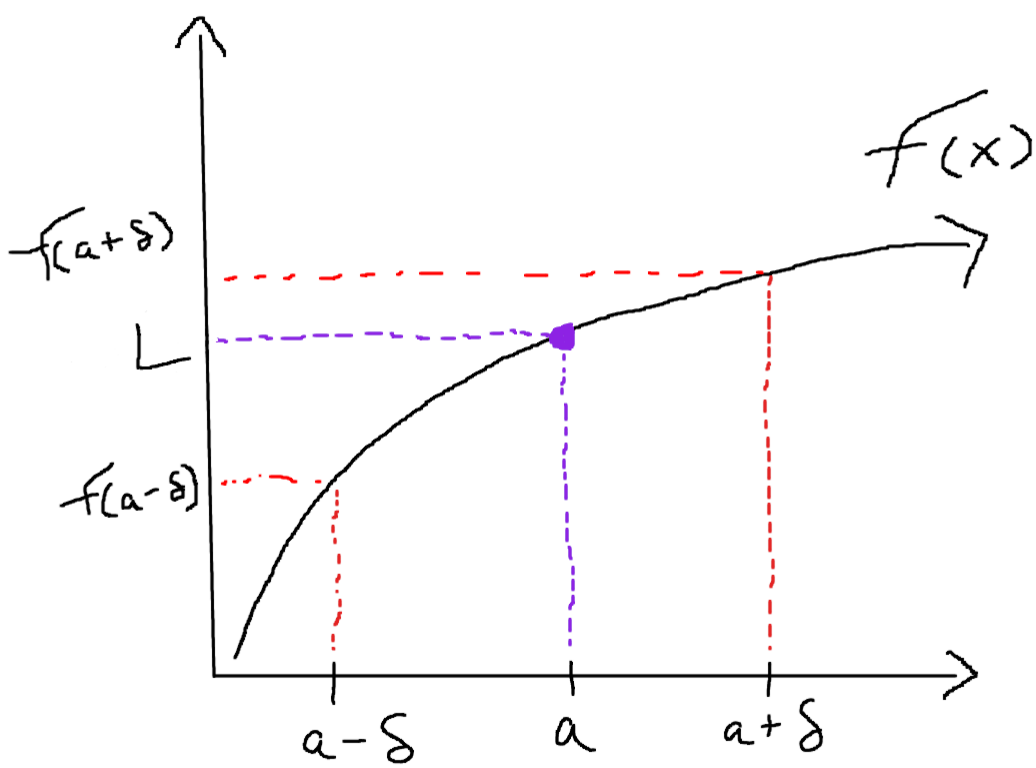
5.2.5 Limits of Functions $\mathbb{R} \to \mathbb{R}$

So far, we have only considered the limits of sequences, which, by definition, take a natural number and map it to any real number. As such, the limits we have taken so far have not been of "smooth" functions, and we have only been able to take limits as $x \to \infty$. While most of the machinery that we are used to remains the same, we formally change the way the limit is taken, as described below:

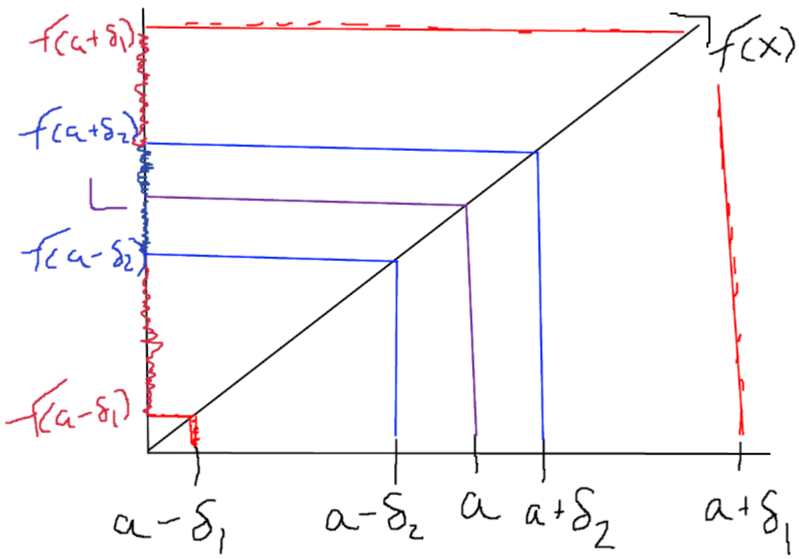
Definition 5.3 (Limit of a function $\mathbb{R} \to \mathbb{R}$)**.** *If $f(x)$ is a function defined on some open interval containing a point a, $L \in \mathbb{R}$ is the* **limit** *of $f(x)$ as $x \to a$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ for which*

$$|x - a| < \delta \implies |f(x) - L| < \varepsilon$$

Since we want to be able to take any limit as $x \to a$ we want, not just as $x \to \infty$, it is necessary to add the stipulation that our x values need to be within some distance δ of a. If they are, we are okay to take the limit as usual. A graphic portraying this is shown below:



Notice that as δ gets smaller (the closer to the actual value of a you get), the closer we generally expect $f(x)$ to get to L (which is often just $f(a)$, as before):



Here's an example proof that $\lim_{x\to 2} x = 2$:

Proof. Let $\varepsilon > 0$. Then there exists a $\delta = \varepsilon > 0$ for which $|x - 2| < \delta = \varepsilon$, so the limit as $x \to 2$ of $f(x) = x$ is 2. Q.E.D.

These proofs are generally a bit harder. After letting $\varepsilon > 0$, as usual, we need to choose some value of δ so we can manipulate $|x - a| < \delta$ into looking like $|f(x) - L| < \varepsilon$. In the example above, all we had to do was say $\delta = \varepsilon$, but it is usually not that easy. Since most of the important results about limits involving real functions are practically the same as sequences, we will conclude our discussion here.

5.2.6 Squeeze Theorem

(planned - I'll add exposition and probably put this somewhere else later)

Theorem 5.5 (Squeeze Theorem)**.** *Let $f, g,$ and h satisfy $f(x) \leq g(x) \leq h(x)$ for all x in some common domain A. If $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} h(x) = L$ at some limit point c of A, then $\lim_{x\to c} g(x) = L$ as well.*

Proof. Let $\varepsilon > 0$. By definition of the limit for functions, since $\lim_{x\to c} f(x) = L$, there exists a δ_1 for which $|x - c| < \delta_1$ implies $|f(x) - L| < \varepsilon$.

Likewise, because $\lim_{x\to c} h(x) = L$, there exists a δ_2 for which $|x-c| < \delta_2$ implies $|h(x)-L| < \varepsilon$.

Choose $\delta = \max\{\delta_1, \delta_2\}$, for which $|x - c| < \delta$ implies $|f(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$.

We apply the triangle inequality to obtain $|f(x)| < \varepsilon - |L|$ and $h(x) < \varepsilon - |L|$.

Because $f(x) \leq g(x) \leq h(x)$ for all $x \in A$, we have $|g(x)| < \varepsilon - |L|$. Thus, $|g(x) - L| < \varepsilon$, proving the Squeeze theorem. Q.E.D.

5.3 Continuous Functions and Topology of the Reals

(planned - extremely important topic)

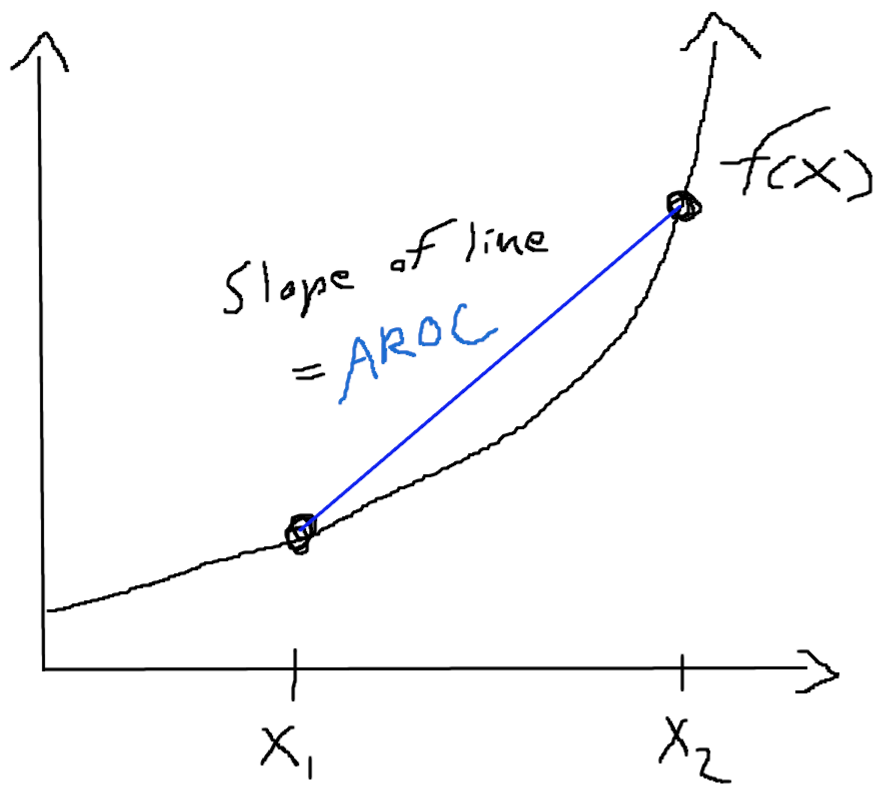
5.4 Definition(s) of the Derivative

Recall the definition of the slope of a line:

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

If we want the average rate of change ("AROC") of a function over some interval (x_1, x_2), we say

$$\text{AROC} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$



The notion of an average rate of change is a bit "coarse." What if we just wanted the **instantaneous** rate of change ("IROC") at a single point?

Algebraically, we struggle. If we just plug in the same point for x_1 and x_2, then we obtain

$$\text{IROC}(x) = \frac{f(x) - f(x)}{x - x} = \frac{0}{0}$$

Clearly, the best we can do with mere algebra is approximate the IROC by choosing two values of x that are very close to each other (say, x and $x + \varepsilon$, where $\varepsilon > 0$). Then we get

$$\text{IROC}(x) \approx \frac{f(x+\varepsilon) - f(x)}{x + \varepsilon - x} = \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

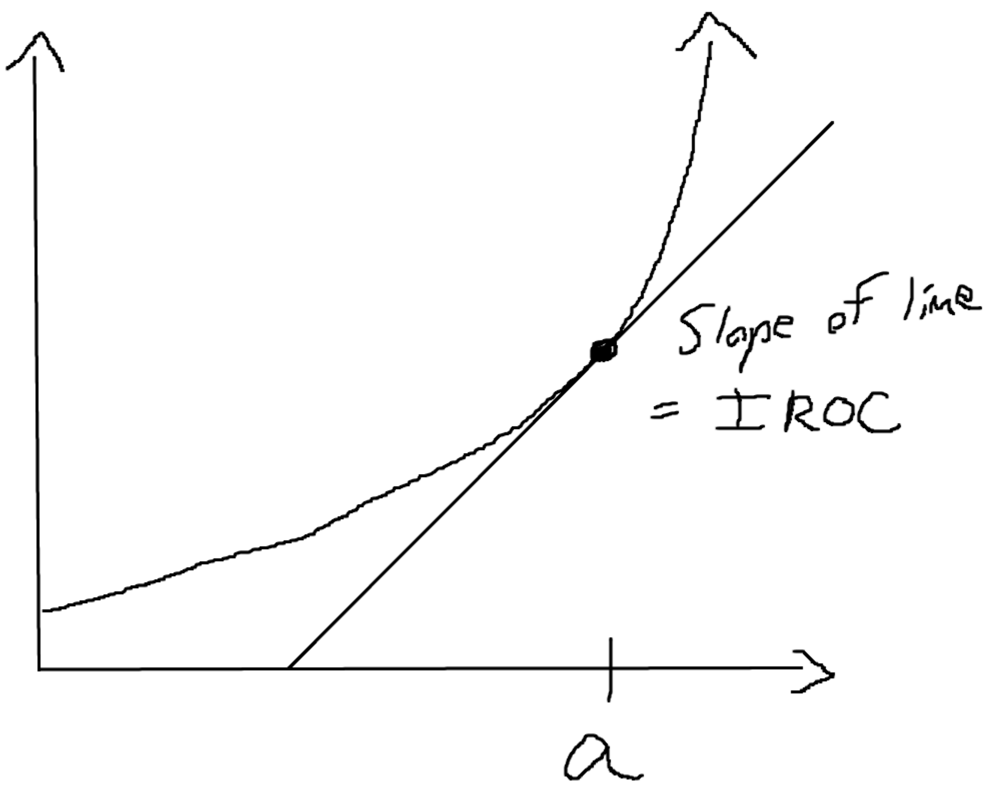
With our tools from calculus, we don't have to stop here. Although it is not defined algebraically, the limit as $\varepsilon \to 0$ may exist. So,

$$\text{IROC}(x) = \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

If we change the notation slightly,

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

we obtain the most common definition of the **derivative**. As emhpasized above, we interpret the derivative as being the slope at a single point of a function. The notation dy/dx is supposed to be reminiscent of $\Delta y/\Delta x$, except the roman "d" character (loosely speaking) represents a very, very tiny infinitesimal change in a quantity.



This is not the only way to write the derivative, however. Since the important thing is that $\Delta x \to 0$, we can instead say

$$x_2 - x_1 \to 0 \implies x_1 \to x_2$$

So another perfectly good way to write the derivative at a point a is

$$\frac{dy}{dx} = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

Feel free to use whichever one you prefer, though the version employing h may have some slight advantages when computing the derivative since it involves fewer terms.

Here is an example of computing the derivative of the function $f(x) = 5x + 2$:

Example.

$$\begin{aligned}\frac{dy}{dx} &= \lim_{h\to 0} \frac{[5(x+h)+2] - (5x+2)}{h} \\ &= \lim_{h\to 0} \frac{5x - 5x + 2 - 2 + 5h}{h} \\ &= \lim_{h\to 0} \frac{5h}{h} = \lim_{h\to 0} 5 = 5\end{aligned}$$

This makes sense since we know the slope of the line is $m = 5$. Observe that the process of calculating the derivative usually involves finding a way to divide by h **before** taking the limit. Otherwise, by plugging in $h = 0$, we obtain some quantity divided by zero, which is undefined.

Most functions *do not* have a constant slope; generally, it will depend on x. Consider the example $g(x) = x^2$:

Example.

$$\begin{aligned}\frac{dy}{dx} &= \lim_{h\to 0} \frac{(x+h)^2 - x^2)}{h} \\ &= \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h\to 0} \frac{2xh + h^2}{h} = \lim_{h\to 0} 2x + h = 2x\end{aligned}$$

The slope of x^2 doubles every time you increase x by 1.

Before ending this section, it is important to discuss the many different ways people write derivatives. All of the following mean the exact same thing:

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

Finally, note that taking the derivative *multiple times* results in a **higher derivative**. For example, the 2nd derivative of $f(x)$ is given by

$$\frac{d}{dx}\left(\frac{d}{dx}f(x)\right)$$

and we notate the 2nd derivative any of the following ways:

$$\frac{d^2y}{dx^2} = \frac{d^2}{dx^2}f(x) = \frac{d^2f}{dx^2}(x) = f''(x)$$

If one derivative (the "1st derivative") represents the speed of a function, then the 2nd derivative represents the acceleration of the function.

5.5 Derivative Rules and Theorems

Let's get right into it!

5.5.1 Linearity of the Derivative

Theorem 5.6. *The derivative is a linear transformation.*

Proof. By the definition of linearity, we must show that the derivative is both additive

$$\frac{d}{dx}(f(x) + g(x)) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$$

and homogenous

$$\frac{d}{dx}(a \cdot f(x)) = a \cdot \frac{d}{dx}f(x)$$

for any constant $a \in \mathbb{R}$. For additivity, observe:

$$\begin{aligned}
\frac{d}{dx}(f(x) + g(x)) &= \lim_{h\to 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \\
&= \frac{d}{dx}f(x) + \frac{d}{dx}g(x)
\end{aligned}$$

We conclude that the derivative is additive. For homogeneity:

$$\begin{aligned}
\frac{d}{dx}(a \cdot f(x)) &= \lim_{h\to 0} \frac{a \cdot f(x+h) - a \cdot f(x)}{h} \\
&= a \cdot \left(\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\right) \\
&= a \cdot \frac{d}{dx}f(x)
\end{aligned}$$

We conclude that the derivative is homogenous. Since it is both additive and homogenous, we know that the derivative is a linear transformation. Q.E.D.

This is a critical piece of information for using the derivative in practical scenarios. If we are taking the sum or difference of a lot of different functions, we are allowed to take the derivative one term at a time! Addtionally, if a constant can be factored out of the original function, then it can also be pulled out of the derivative, and it otherwise does not affect the calculation.

5.5.2 Chain Rule

This rule is related to function composition. Given some function $g \circ f = g(f(x))$, is there a systematic way to find its derivative? Yes!

Theorem 5.7 (Chain Rule)**.**

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

Proof. Provided f and g are each differentiable and $g(x) \neq g(a)$ in a neighborhood of x near a,

$$\begin{aligned}
\frac{d}{dx}g(f(x)) &= \lim_{x\to a} \frac{g(f(x)) - g(f(a))}{x - a} \\
&= \lim_{x\to a} \frac{g(f(x)) - g(f(a))}{x - a} \cdot \frac{f(x) - f(a)}{f(x) - f(a)} \\
&= \lim_{x\to a} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \cdot \frac{f(x) - f(a)}{x - a} \\
&= g'(f(x)) \cdot f'(x)
\end{aligned}$$

Q.E.D.

The power of this method may not be apparent now, but prior knowledge of the derivatives of common parent functions (e.g., e^x, x^2, $\cos x$, ...), this rule enables determining derivatives of otherwise very complicated composite functions.

5.5.3 Product and Quotient Rules

What if we are taking a product or quotient of two functions?

Theorem 5.8 (Product rule)**.**

$$\frac{d}{dx}f(x) \cdot g(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Proof.

$$\begin{aligned}
\frac{d}{dx}f(x) \cdot g(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) + (f(x)g(x+h) - f(x)g(x+h)) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x)(g(x+h) - g(x)) + (f(x+h) - f(x))g(x+h)}{h} \\
&= f(x) \cdot \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} + \lim_{h\to 0} g(x+h) \cdot \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= f(x)g'(x) + f'(x)g(x)
\end{aligned}$$

Q.E.D.

To prove the quotient rule, we will need a helpful lemma:

Lemma 5.1 (Reciprocol rule)**.**

$$\frac{d}{dx}\left(\frac{1}{f(x)}\right) = -\frac{f'(x)}{(f(x))^2}$$

Proof. Proof from first principles is possible by direct computation. Alternatively, if we assume $1/f(x)$ is differentiable, then by the product rule,

$$\frac{d}{dx}\left(f(x)\cdot\frac{1}{f(x)}\right) = f'(x)\frac{1}{f(x)} + f(x)\left(\frac{1}{f(x)}\right)' = \frac{d}{dx}1 = 0$$

Rearrangment yields the identity:

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

Q.E.D.

Now we can prove the quotient rule with relative ease.

Theorem 5.9 (Quotient rule)**.**

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

Proof. By the product rule,

$$\frac{d}{dx}\frac{f(x)}{g(x)} = f(x)\frac{d}{dx}\left(\frac{1}{g(x)}\right) + \frac{1}{g(x)}\frac{d}{dx}f(x)$$

We know the derivative of $1/g(x)$ from the reciprocol rule:

$$f(x)\frac{d}{dx}\left(\frac{1}{g(x)}\right) + \frac{1}{g(x)}\frac{d}{dx}f(x) = -f(x)\cdot\frac{g'(x)}{(g(x))^2} + \frac{f'(x)}{g(x)}$$

If we put everything into a common denominator of $(g(x))^2$, we conclude with the desired identity and conclude

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

Q.E.D.

For all our work, it turns out that the power rule is generally more straightforward to use than the quotient rule, even for quotients. That's because $a/b = a\cdot(1/b)$.

5.5.4 L'Hôpital's rule

(planned - you can take derivatives to make computing limits a lot easier; see below)

$$\lim \frac{f(x)}{h(x)} = \lim \frac{f'(x)}{h'(x)}$$

5.5.5 Derivatives of Exponential and Logarithmic Functions

Since it will be convenient for proving some other derivatives, we start with exponential and logarithmic functions.

Theorem 5.10 (Derivative of $\ln x$)**.**

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

Proof. Let $y = \ln x$. By the definition of the derivative, we have

$$\frac{dy}{dx} = \lim_{h\to 0}\frac{\ln(x+h)-\ln(x)}{h}$$

We employ the rule for log subtraction which says $\log a - \log b = \log(a/b)$ to obtain

$$\lim_{h\to 0}\frac{\ln(x+h)-\ln(x)}{h} = \lim_{h\to 0}\frac{\ln\frac{x+h}{x}}{h}$$

We can simplify the argument of the natural log.

$$\lim_{h\to 0}\frac{\ln\frac{x+h}{x}}{h} = \lim_{h\to 0}\frac{\ln\left(1+\frac{h}{x}\right)}{h}$$

A clever substitution is needed to compute the limit. Let $t = \frac{h}{x}$ so that $h = xt$. This is the result:

$$\lim_{h\to 0}\frac{\ln\left(1+\frac{h}{x}\right)}{h} = \lim_{t\to 0}\frac{\ln\left(1+t\right)}{xt}$$

We can use the log rule which says $a\log x = \log x^a$ to obtain

$$\lim_{t\to 0}\frac{\ln\left(1+t\right)}{xt} = \frac{1}{x}\cdot\lim_{t\to 0}\ln\left[(1+t)^{1/t}\right]$$

It turns out that the limit is one definition of the constant e, so we conclude

$$\frac{1}{x}\cdot\lim_{t\to 0}\ln\left[(1+t)^{1/t}\right] = \frac{1}{x}\ln e = \frac{1}{x}$$

Thus,

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

Q.E.D.

Next, we will do a^x. However, we will need to know the derivative of the identity function first.

Recall that the identity function is $f(x) = x$, which equals whatever value of x you put in. This is also the equation of a straight line with slope $m = 1$ and y-intercept $b = 0$, so we expect that the derivative will equal 1. This turns out to be true.

Lemma 5.2 (Derivative of the identity function)**.**

$$\frac{d}{dx}x = 1$$

Proof. Let $y = x$. Then

$$\frac{dy}{dx} = \lim_{h\to 0}\frac{x+h-x}{h} = \lim_{h\to 0}\frac{h}{h} = \lim_{h\to 0} 1 = 1$$

Q.E.D.

With this small piece of extra information, we are ready to determine the derivative of a^x.

Theorem 5.11.

$$\frac{d}{dx}a^x = a^x \cdot \ln a$$

Proof. Let $y = a^x$. Then

$$\ln y = x \ln a$$

Using the chain rule, we can find the derivative of both sides of the equation.

$$\frac{d}{dx}\ln y = \frac{1}{y}\cdot\frac{dy}{dx}$$

$$\frac{d}{dx}x\ln a = \ln a$$

So,

$$\frac{1}{y}\cdot\frac{dy}{dx} = \ln a$$

which implies

$$\frac{dy}{dx} = y\ln a$$

Since $y = a^x$, we conclude

$$\frac{d}{dx}a^x = a^x \cdot \ln a$$

Q.E.D.

Since we use the exponential function e^x so often, we will highlight the case where $a = e$ specifically:

$$\frac{d}{dx}e^x = e^x \cdot \ln e = e^x$$

The function e^x has the unique property that it is its own derivative at every point $x \in \mathbb{R}$.

5.5.6 Derivative of x^n and Polynomials

Polynomial functions are basically a sum of powers of x raised to natural numbers. Since the derivative is linear, we just need to figure out how to differentiate x^n for some constant n, enabling us to find the derivative to any polynomial function. There is a wide world of exponents beyond $0, 1, 2, 3, \ldots$, including negative numbers, rational numbers (in the case of roots, for example), and any other real numbers. Fortunately, the same identity works for any real number $n \in \mathbb{R}$, as we shall see below.

Theorem 5.12.

$$\frac{d}{dx}x^n = nx^{n-1}$$

Proof. Let $y = x^n$. Since $e^{\ln x} = x$, we can rewrite this expression as

$$y = e^{\ln x^n} = e^{n \ln x}$$

We can use the chain rule to find dy/dx:

$$\frac{dy}{dx} = e^{n \ln x} \cdot \frac{n}{x} = e^{\ln x^n} \cdot \frac{n}{x}$$

Now we replace $e^{\ln x^n}$ with x^n and conclude

$$\frac{dy}{dx} = \frac{nx^n}{x} = nx^{n-1}$$

Q.E.D.

If you care to explore alternative proofs, you will find that the approach above manages to circumvent a proof by **mathematical induction** (the details of which we will not discuss here) while also proving a stronger result!

5.5.7 Derivatives of Trigonometric Functions

Thanks to the product rule and/or quotient rule, knowing the derivatives of $\sin x$ and $\cos x$ is enough to figure out the derivatives of all other trigonometric functions without resorting to the definition of the derivative.

(remainder is planned - details depend on the Squeeze theorem; for now, results are pasted below for reference)

Theorem 5.13.

$$\frac{d}{dx}\sin x = \cos x$$
$$\frac{d}{dx}\cos x = -\sin x$$
$$\frac{d}{dx}\tan x = \sec^2 x$$

5.5.8 Table of Derivative Rules and Theorems

Name	Property	Example
Linearity	$\frac{d}{dx}a \cdot f(x) = a \cdot \frac{d}{dx}f(x)$ $\frac{d}{dx}(f(x)+g(x)) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$	$\frac{d}{dx}(4x^2+2x) = 4\frac{d}{dx}x^2 + 2\frac{d}{dx}x = 8x+2$
Chain rule	$\frac{d}{dx}f(g(x)) = f'(g(x)) \cdot g'(x)$	$\frac{d}{dx}\sqrt{\cos x} = \frac{1}{\sqrt{\cos x}} \cdot (-\sin x)$
Product rule	$\frac{d}{dx}f(x)g(x) = f'(x)g(x) + f(x)g'(x)$	$\frac{d}{dx}x\sin x = (1)\sin x + x\cos x$
Inverse rule	$\frac{d}{dx}\left(\frac{1}{f(x)}\right) = -\frac{f'(x)}{(f(x))^2}$	$\frac{d}{dx}\csc x = \frac{d}{dx}\frac{1}{\sin x} = -\frac{\cos x}{\sin^2 x} = -\cot x \csc x$
Quotient rule	$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{f(x)g'(x) - f'(x)g(x)}{(g(x))^2}$	$\frac{d}{dx}\tan x = \frac{d}{dx}\frac{\sin x}{\cos x} = \frac{-\sin^2 x - \cos^2 x}{\cos^2 x} = -\sec^2 x$
Power rule	$\frac{d}{dx}x^n = nx^{n-1}$	$\frac{d}{dx}5x^\pi = 5\pi x^{\pi-1}$
Natural log	$\frac{d}{dx}\ln x = \frac{1}{x}$	$\frac{d}{dx}\ln 5x = \frac{1}{5x} \cdot 5 = \frac{1}{x}$
Exponential	$\frac{d}{dx}a^x = a^x \cdot \ln a$	$\frac{d}{dx}e^{-x^2} = e^{-x^2} \cdot (-2x)$
Sine	$\frac{d}{dx}\sin x = \cos x$	$\frac{d}{dx}A\sin(\omega(x+\phi)) + B = A\cos(\omega(x+\phi)) \cdot \omega$
Cosine	$\frac{d}{dx}\cos x = -\sin x$	$\frac{d}{dx}\cos x^2 = -\sin x^2 \cdot 2x$

5.6 Applications of Derivatives

5.6.1 Optimization Problems

(planned) Punchline: if you want to find the maximum or minimum (a.k.a. vertices or "turning points") of a function, take its derivative and set it equal to zero. This has many applications since people will try to minimize cost, energy, and distance traveled or otherwise try to maximize a desirable quantity.

5.6.2 Related Rates

(planned)

5.6.3 Implicit Differentiation

(planned) Key idea: you can find potentially useful **differential equations** from difficult equations by taking the derivative of both sides. For example,

$$x^2 + y^2 = r^2$$

Derivative of both sides:

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}r^2$$

We know the derivative of x^2 equals $2x$, and we know the derivative of a constant like r^2 equals 0. What is $\frac{d}{dx}y^2$? There are multiple ways to find out. With the chain rule:

$$\frac{d}{dx}(y)^2 = 2y \cdot \frac{dy}{dx}$$

With the product rule:

$$\frac{d}{dx}y \cdot y = y \cdot y' + y' \cdot y = 2y\frac{dy}{dx}$$

Thus, when we take the derivative of both sides of the original equation, we get

$$2x + 2y\frac{dy}{dx} = 0$$

If we solve for dy/dx, we get

$$\frac{dy}{dx} = -\frac{x}{y}$$

This is a **differential equation**, which we *could* solve attempt to solve using techniques from integral calculus. This time, the solution is easier: just solve the original equation for y.

$$x^2 + y^2 = r^2 \implies y = \sqrt{r^2 - x^2}$$

We need to drop the $\pm$ before the square root to make it a well-defined function that passes the vertical line test (only half a circle). Then we plug in and get

$$\frac{dy}{dx} = -\frac{x}{\sqrt{r^2 - x^2}}$$

which is our final answer. However, not every such problem can be resolved so simply. Consider the following example:

$$xy^2 + x^2y = 1$$

This equation is **inseparable** since x and y cannot be algebraically manipulated onto separate sides. Nonetheless, we can try to find dy/dx like we did before.

$$\frac{d}{dx}(xy^2 + x^2y) = \frac{d}{dx}1$$

We use the product rule and obtain

$$(1)y^2 + 2xy\frac{dy}{dx} + 2xy + x^2\frac{dy}{dx} = 0$$

As before, this can be solved for dy/dx.

(to be continued)

5.7 Introduction to Series

In mathematics, it is sometimes the case that we need to take sums of many numbers. If a summation follows some predictable pattern, we can use a shorthand notation to write an entire expression with fewer characters. Specifically, we use the notation

$$\sum_{k=1}^{n} a_k$$

to denote the summation of n terms together, where $\sum$ (Greek "Sigma") represents a "big sum," k is an indexing number, and n represents the number at which the sum stops. To illustrate, we will examine a few concrete examples. First, we can write an n-th degree polynomial in terms of this summation notation:

$$\sum_{k=0}^{n} a_k x^k = a_0x^0 + a_1x^1 + a_2x^2 + \cdots + a_nx^n$$

where, as usual, each a_k is some real number coefficient. We can also use this summation notation to truncate large sums of constants:

$$\sum_{k=1}^{4} 1 = 1 + 1 + 1 + 1$$

$$\sum_{k=1}^{100} k = 1 + 2 + 3 + 4 + \cdots + 100$$

Ultimately, this "big sum" notation is no different from addition in elementary algebra. It only provides us a structured, shorthand notation for writing large sums.

Recall that a **sequence** of real numbers is a function $\mathbb{N} \to \mathbb{R}$, basically enumerating a list of real numbers. A sequence of **partial sums** is a sequence where incrementally more numbers are added in every entry. We can once again illustrate with an example. If $b_n = \sum_{k=1}^{n} k$, then the first three entries of the sequence (b_n) are

$$b_1 = \sum_{k=1}^{1} k = 1 \qquad b_2 = \sum_{k=1}^{2} k = 1 + 2 \qquad b_3 = \sum_{k=1}^{3} k = 1 + 2 + 3$$

A **series** or **infinite series** is the limit as $n \to \infty$ of a sequence of partial sums:

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k$$

To save space, it is common to truncate this to

$$\sum_{k=1}^{\infty} a_k$$

even though, technically, we are finding a limit and not actually adding infinite numbers. In the example of (b_n) above, if we took the limit as we add an infinite quantity of natural numbers together, it clearly diverges to infinity since the sums keep getting larger and larger. On the other hand, it is not difficult to find a series that converges to some finite number, despite (loosely speaking) adding infinitely many numbers together. For example,

$$\sum_{k=0}^{\infty} 10^{-k} = 1 + 0.1 + 0.01 + 0.001 + \cdots = 1/9$$

Series are foundational to mathematical analysis. We will explore how to use them in greater depth in later sections.

5.8 Riemann Integral and Antiderivatives

This section introduces a very important piece of machinery for calculus: the antiderivative. If a derivative is something that transforms one function f into a different one f', then the antiderivative is the inverse, which takes f' and returns it to f. Presented without context, however, the antiderivative may seem very mysterious in how it is defined and how it works. Thus, we will begin by presenting some motivating examples from algebra.

Recall that when we sought to define the derivative, we started wtih the algebraic concept of a slope

$$m = \frac{\Delta y}{\Delta x}$$

which represents the average rate of change of a function over some interval or, in the case of a straight line, its constant slope everywhere. Our motivating question (in terms of algebra) is as follows: if we know the slope m of a straight line, how do we find out what the line is?

First, we multiply both sides by Δx to obtain

$$\Delta y = m\Delta x$$

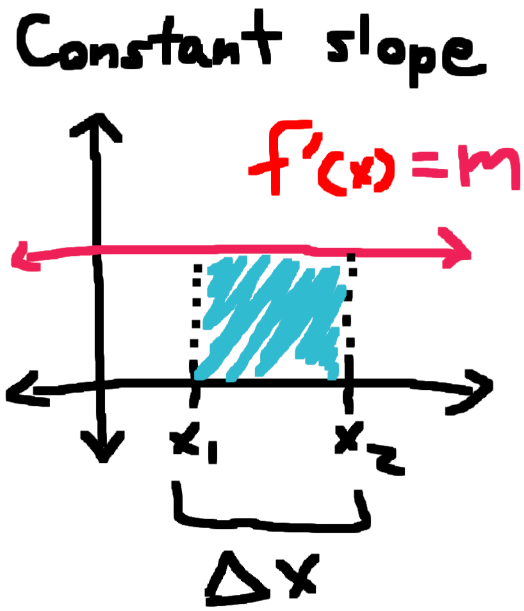
which, rewritten, is just the point-slope form of a line:

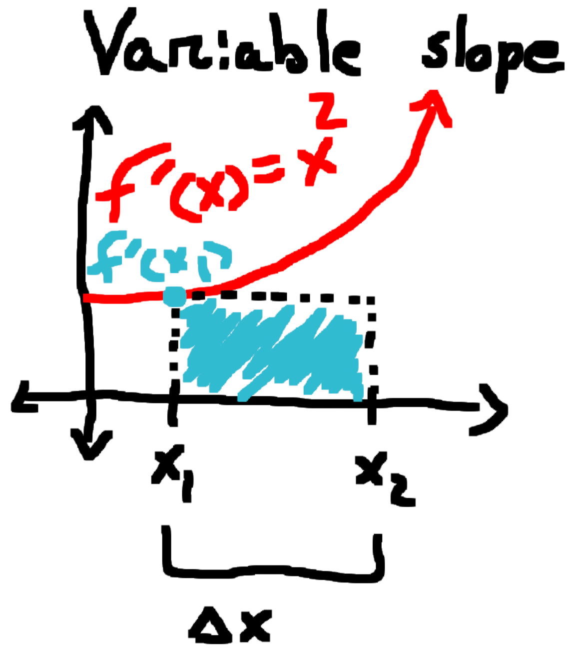
$$y - y_0 = m(x - x_0)$$

We can rewrite this in terms of the slope-intercept form of a line by adding y_0 to both sides of the equation:

$$y = mx + (mx_0 + y_0) = mx + b$$

We can geometrically interpret the expression $\Delta y = m\Delta x$ by realizing that $m\Delta x$ is like the area of a rectangle, where the length is given by how big the slope m is and the width by the size of Δx.

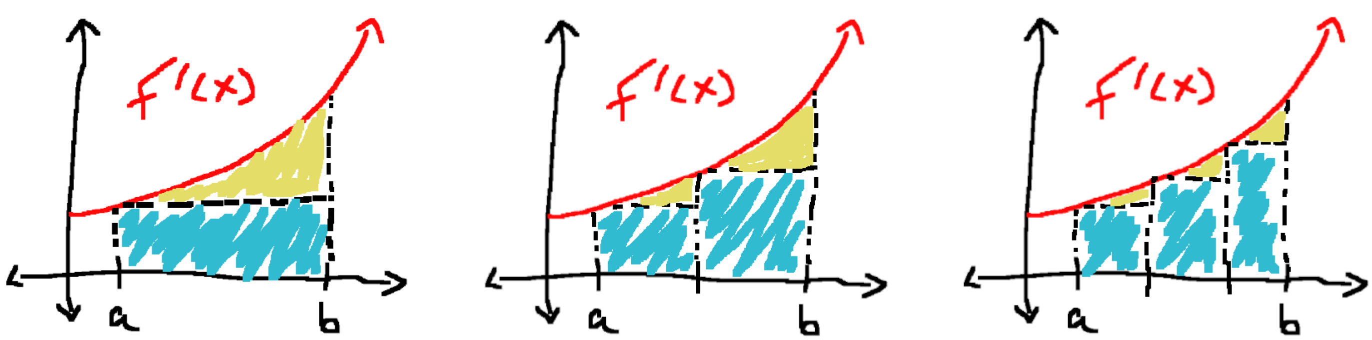




Basically, if we know the slope of a straight line (i.e., its derivative) and a Δx, we can find out what Δy is by finding the area of the rectangle they make, which tells us what y itself is equal to. However, when we try to apply this same trick to functions with variable slopes, we run into a problem: the area between the curve and the x-axis is no longer a rectangle, and saying

$$f(x) = f'(x_1)\Delta x$$

is *at best* a very crude approximation of $f(x)$, and we want *exact* answers. So how do we more accurately find the area under the curve and improve our approximation? The answer is to add more rectangles.



In the graphic above, the blue area represents the total measured by rectangles, while the yellow represents how much we fail to account for. Notice that the yellow area gets smaller and smaller the more rectangles we add, meaning that our blue area is getting closer and closer to representing the true value of $f(x)$.

We can represent adding more and more rectangles together by a sequence of partial sums

$$S_n = \sum_{k=1}^{n} f(x_k)\Delta x$$

each S_n called a **Riemann sum**. Here, Δx represents the width of each individual rectangle. In the cartoon above, our total interval is from a to b, so the total length is $b - a$. Then we can figure out how wide each rectangle has to be by dividing by the number of rectangles n. So $\Delta x = \frac{b-a}{n}$. This lets us assign values to each x_k. For example, $x_1 = a$ since that's the start of the interval; then $x_2 = a + \Delta x$, and we can repeat this n times.

To make this approximation exact, we now need to introduce limits. Instead of adding a finite number n rectangles under the curve, we want to add (loosely speaking) "infinite rectangles" by turning this into a series, giving us

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x$$

In the limit of $n \to \infty$, our Δx becomes very small, allowing us to add a very large number of very thin rectangles together. This process, called **integrating** $f(x)$, yields the **antiderivative** of $f(x)$, also called the **primitive** of $f(x)$ or the **indefinite integral** in certain contexts. Similar to the derivative, we replace the Greek characters with notation derived from Roman characters, giving us

$$\int f(x)\, dx \equiv \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x$$

When we want to find the antiderivative along a specific interval $[a, b]$, then we notate it

$$\int_a^b f(x)\, dx$$

and say we are finding the **definite integral** of $f(x)$ from a to b, representing the specific area under the curve on $[a, b]$.

We will formally prove in the next section the following two important facts:

$$\int f'(x)\, dx = f(x) + C$$

$$\frac{d}{dx} \int_a^x f(t)\, dt = f(x)$$

That is, the integral can undo derivatives, and derivatives undo integrals.

5.9 Integration Rules and Theorems

Let's begin with the basics.

5.9.1 Linearity of the Antiderivative

Theorem 5.14. *The antiderivative is a linear transformation.*

Proof. By the definition of linearity, we must show that the derivative is both additive

$$\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx$$

and homogenous

$$\int a \cdot f(x)dx = a \int f(x)dx$$

for any constant $a \in \mathbb{R}$. For the condition of homogeneity, we will approach by the definition of the antiderivative:

$$\int a \cdot f(x)dx = \lim_{n\to\infty} \sum_{k=1}^{n} a \cdot f(x_k)\Delta x$$

We are allowed to factor a out of the summation since it is a constant on every term, giving us

$$\lim_{n\to\infty} \sum_{k=1}^{n} a \cdot f(x_k)\Delta x = \lim_{n\to\infty} a \sum_{k=1}^{n} f(x_k)\Delta x$$

By the Algebraic Limit Theorem, we are allowed to factor the constant outside of the limit,

$$\lim_{n\to\infty} a \sum_{k=1}^{n} f(x_k)\Delta x = a \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x = a \int f(x)dx$$

and conclude that the antiderititive is homogenous. To show the antiderivative is additive, we also proceed from the definition:

$$\begin{aligned}
\int (f(x) + g(x))dx &= \lim_{n\to\infty} \sum_{k=1}^{n} (f(x) + g(x))\Delta x \\
&= \lim_{n\to\infty} \left(\sum_{k=1}^{n} f(x)\Delta x + \sum_{k=1}^{n} g(x)\Delta x \right) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} f(x)\Delta x + \lim_{n\to\infty} \sum_{k=1}^{n} g(x)\Delta x \\
&= \int f(x)dx + \int g(x)dx
\end{aligned}$$

We conclude that the antiderivative is additive. Since it is both additive and homogenous, we know that the integral is a linear transformation. Q.E.D.

Like derivatives, this is an important theorem for practical scenarios. This tells us we can "distribute" the antiderivative over whatever individual functions we are taking the sum or difference of. Further, constants can be pulled out of the integral and dealt with later without affecting the computation.

5.9.2 First Fundamental Theorem of Calculus

Theorem 5.15.

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

Proof. Planned. For the moment, the important thing to know here is that $F(x)$ just represents "whatever the antiderivative of $f(x)$ is." Q.E.D.

5.9.3 Second Fundamental Theorem of Calculus

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

Proof. Planned. A really quick sketch of some of the details:

$$\frac{d}{dx}\int_a^x f(t)\,dt = \frac{d}{dx}\left(F(x) - F(a)\right) = \frac{d}{dx}F(x) - \frac{d}{dx}F(a)$$

$F(x)$ is the antiderivative of $f(x)$, which gets undone by derivatives, and $F(a)$ is a constant because it is a function evaluated at a single point. So,

$$\frac{d}{dx}F(x) - \frac{d}{dx}F(a) = f(x) - 0 = f(x)$$

Q.E.D.

5.9.4 Integral Rules from Derivative Rules

Fortunately, we can skip a very large number of tricky proofs by just cleverly "undoing" our derivative rules. For example,

$$\frac{d}{dx}e^x = e^x \implies \int e^x\,dx = e^x + C$$

$$\frac{d}{dx}\sin x = \cos x \implies \int \cos x\,dx = \sin x + C$$

$$\frac{d}{dx}\cos x = -\sin x \implies \int \sin x\,dx = -\cos x + C$$

$$\frac{d}{dx}\frac{x^{n+1}}{n+1} = x^n \implies \int x^n\,dx = \frac{x^{n+1}}{n+1} + C \qquad (n \neq -1)$$

$$\frac{d}{dx}\ln x = \frac{1}{x} \implies \int \frac{1}{x}\,dx = \ln x + C$$

(planned: other integrals that end up being helpful, many of them based on other trig functions)

5.10 Table of Integration Rules and Theorems

Name	Property	Example
First Fundamental Theorem	$$\int_a^b f(x)\,dx = F(b) - F(a)$$	$\int_0^1 x^2\,dx = \left(\frac{x^3}{3}\right)_0^1 = \frac{1}{3} - \frac{0}{3} = 1/3$
Second Fundamental Theorem	$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$$	$\frac{d}{dx}\int f(x)\,dx = f(x)$
Linearity	$$\int (f(x)+g(x))dx = \int f(x)dx + \int g(x)dx$$ $$\int a \cdot f(x)dx = a\int f(x)dx$$	$\int (4x^2 + 2x)\,dx = 4\int x^2\,dx + 2\int x\,dx = \frac{4x^3}{3} + x^2$
Integration by parts	$$\int u\,dv = uv - \int v\,du$$	$\int xe^{2x}\,dx = \frac{xe^{2x}}{2} - \int \frac{e^{2x}}{2}dx$
Power rule	$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1)$$	$\int 5x^\pi\,dx = 5x^{\pi+1}/(\pi+1) + C$
$1/x$	$$\int \frac{1}{x}\,dx = \ln x + C$$	$\int 1/5x\,dx = \ln 5x + C$
Natural log	$$\int \ln x\,dx = x\ln x - x + C$$	$\int \ln 5x\,dx = 5x\ln 5x - 5x + C$
Exponential	$$\int e^{ax}\,dx = e^{ax}/a + C$$	$\int e^{-2x}\,dx = -e^{-2x}/2 + C$
Sine	$$\int \sin ax\,dx = -\frac{\cos ax}{a} + C$$	$\int \sin x\,dx = -\cos x + C$
Cosine	$$\int \cos ax\,dx = \frac{\sin ax}{a} + C$$	$\int \cos 3x\,dx = \frac{\sin 3x}{3} + C$

s

5.11 Integration Strategy: u Substitution

One of the most basic tools for solving integrals is u **substitution**. Consider the following example:

$$\int xe^{x^2}\,dx$$

We do not know the antiderivative of the integrand from common identities; it is too complicated. However, a clever change in variables makes the problem much easier to solve. To accomplish this, we simply let some variable u equal the most difficult part of our expression. In this case, we let $u = x^2$. To figure out how to do calculus with this new variable u, we need to take a derivative:

$$\frac{du}{dx} = \frac{d}{dx}x^2 = 2x$$

If we "multiply both sides" by dx, then

$$du = 2x\,dx$$

In other words,

$$dx = \frac{du}{2x}$$

Now we can make the substitutions $x^2 = u$ and $dx = du/2x$ in our original integral:

$$\int xe^{x^2}\,dx = \int \frac{xe^u}{2x}du = \frac{1}{2}\int e^u\,du = \frac{1}{2}e^u + C$$

We were able to solve the integral in the variable u because it transformed the expression into something simpler. Now, we just plug in $u = x^2$ to revert back to our original coordinates.

$$\int xe^{x^2}\,dx = \frac{1}{2}e^u + C = \frac{1}{2}e^{x^2} + C$$

(planned - "why does this work?" loosely speaking, it is a consequence of function composition, product rule, etc., but you don't need to worry about the technicalities to apply it to solving problems!)

5.12 Integration Strategy: Integration by Parts

Let $u(x)$ and $v(x)$ be differentiable functions. Then, by the product rule of derivatives,

$$\frac{d}{dx}uv = uv' + vu'$$

We can take the antiderivative of both sides to obtain

$$uv = \int (uv' + vu')\, dx$$

By linearity of the integral, we have

$$uv = \int uv'\, dx + \int vu'\, dx$$

If we subtract $\int vu'\, dx$ from both sides, we get

$$\int uv'\, dx = uv - \int vu'\, dx$$

Let $du = u'\, dx$ and $dv = v'\, dx$ and we conclude

$$\int u\, dv = uv - \int v\, du$$

This is the most common form of the **integration by parts** rule, which is commonly regarded as the integral rule most similar to the derivative product rule. It is a very powerful identity which provides a framework for breaking up and simplifying any integral into parts that are potentially easier to solve. This is most easily understood through examples.

Consider the following integral:

$$\int xe^x\, dx$$

This is the product of two functions that we know how to integrate from common identities. We can attempt to apply integration by parts. Let $u = x$ and $dv = e^x\, dx$. Then,

$$\frac{du}{dx} = \frac{d}{dx}x = 1 \implies du = dx$$

$$v = \int dv = \int e^x\, dx = e^x$$

Now we can plug our values for u, v, du, and dv into our equation for integration by parts:

$$\int xe^x\, dx = xe^x - \int e^x\, dx$$

We know how to solve $\int e^x\, dx$, so we conclude

$$\int xe^x\, dx = xe^x - e^x + C$$

In general, you can employ integration by parts as follows: express your integrand as the product of two functions, let the **easier one to integrate** equal dv and let the **easier one to differentiate** equal u. The best choice for u is a function that will get simpler the more you take its derivative, like $u = x$ above. The best choice for dv is an integrable function whose antiderivative doesn't become much more complicated, like $dv = e^x\, dx$ above.

5.12.1 Rapid Repeated Integration By Parts

Consider the following integral:

$$\int x^4 e^x \, dx$$

We learned how to solve integrals like this in the previous section. Since it appears to be the product of two functions for which we know the antiderivative, we can apply integration by parts. However, there is a problem: if we try to apply integration by parts to this integral, it will turn into a very long and tedious problem. I will illustrate by beginning the solution:

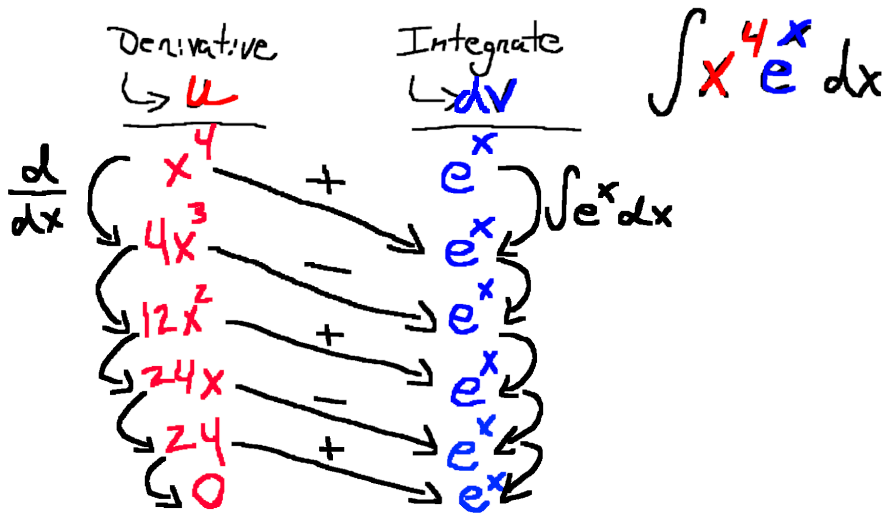
Let $u = x^4$ and $dv = e^x \, dx$. Then $du = 4x^3 \, dx$ and $v = e^x$. So,

$$\int x^4 e^x \, dx = x^4 e^x - \int 4x^3 e^x \, dx$$

We need to apply integration by parts again. Let $u = 4x^3$ and $dv = e^x \, dx$. Then $du = 12x^2 \, dx$ and $v = e^x$. So,

$$\int x^4 e^x \, dx = x^4 e^x - \left(4x^3 e^x - \int 12x^2 e^x \, dx\right)$$

We need to apply integration by parts yet again. In fact, to reach the final solution, we would need to do integration by parts another two times. Luckily for us, the Rapid Repeated Integration By Parts algorithm, illustrated below, allows us to skip a lot of work.



As usual, first set the factor easiest to integrate to dv and set the factor that is easy to differentiate to u. Then, take the derivative of u over and over until it equals zero (works best on polynomials). Then, integrate dv over and over until you have done it as many times as you have differentiated u. Finally, draw a diagonal arrow pointing from the first u to the second entry in the dv column with a plus sign over it. Repeat the pattern until you reach the bottom of the dv column, and alternate the sign between $+$ and $-$. For our example, this gives us the following final answer:

$$\int x^4 e^x \, dx = x^4 e^x - 4x^3 e^x + 12x^2 e^x - 24x e^x + 24 e^x + C$$

5.13 Integration Strategy: Partial Fraction Decomposition

From u substitution and our integral rules, it is generally very easy to integrate rational functions of the form

$$f(x) = \frac{a}{x+b}$$

To demonstrate,

$$\int f(x)\,dx = \int \frac{a}{x+b}\,dx = a\int \frac{1}{x+b}\,dx$$

Let $u = x+b$. Then $du/dx = 1$, which means $du = dx$. If we plug this into our integral,

$$a\int \frac{1}{x+b}\,dx = a\int \frac{1}{u}\,du = a\ln u + C = a\ln(x+b) + C$$

So if we have a degree 1 monomial or binomial in the denominator, we know the antiderivative is just equal to the natural log of the denominator times any constants. But what about more complicated rational functions? For example,

$$f(x) = \frac{1}{x^2+2x}$$

is too complicated to integrate using conventional rules. What we want is to turn this difficult problem into a combination of easier ones, like the one we just solved above. This is accomplished through **partial fraction decomposition**.

First, let's factor the denominator of $f(x)$ to make it look simpler:

$$f(x) = \frac{1}{x^2+2x} = \frac{1}{x(x+2)}$$

Then, to make integrating easy, our goal is to "decompose" $f(x)$ so the denominator is the product of monomials and binomials, giving us

$$f(x) = \frac{1}{x(x+2)} = \frac{A}{x} + \frac{B}{x+2}$$

where $A, B \in \mathbb{R}$ are constants. This makes sense as long as we are able to find specific values for the constants A and B which make the equality true for all possible values of x. "It turns out" that, as written above (and in more complicated cases, discussed below), this is **always** possible. A complete proof that partial fraction decomposition always works depends on the Fundamental Theorm of Algebra (since we are working with rational functions of polynomials) and shall be omitted for now.

In any case, our task is to determine the values of A and B. First, we start with the equality

$$\frac{1}{x(x+2)} = \frac{A}{x} + \frac{B}{x+2}$$

and multiply both sides by $x(x+2)$ to obtain

$$1 = A(x+2) + Bx$$

We can rewrite the right hand side as a polynomial, where terms with similar factors of x are grouped together:

$$1 = A(x+2) + Bx \implies 1 = x(A+B) + 2A$$

Observe that the left hand side is constant. This is equivalent to having any variable term multiplied by zero:

$$1 = 0 \cdot x + 1$$

This tells us that $(A+B)$ on the right hand side must equal zero to make sure equality between the two sides holds. Likewise, we know that $2A$ must equal 1. This gives us a system of linear equations

$$\begin{aligned} A + B &= 0 \\ 2B &= 1 \end{aligned}$$

We solve this the usual way. First, we see that $B = 1/2$ and plug it into the top equation, telling us that $A = -1/2$. Thus, we can rewrite our original function as

$$f(x) = \frac{-1/2}{x} + \frac{1/2}{x+2}$$

or in simplified form as

$$f(x) = \frac{1}{2(x+2)} - \frac{1}{2x}$$

This is the sum of functions that are easy to integrate. Using the same approach as the first example in this section on each fraction, we conclude

$$\int f(x)\,dx = \int \frac{1/2}{x+2}\,dx - \int \frac{1/2}{x}\,dx = (1/2)\ln(x+2) - (1/2)\ln x + C = \frac{\ln(x+2) - \ln x}{2} + C$$

The example given above is one of the simplest. To deal with rational functions of polynomials in general, we need to be aware of a few more technicalities.

What if you have a situation like this?

$$f(x) = \frac{1}{x^2 + 2x + 1} = \frac{1}{(x+1)^2}$$

Since the factor $(x+1)$ shows up two times in the factorization of the polynomial, we say that $(x+1)$ has **multiplicity** $m = 2$. When any irreducible factor in a polynomial has multiplicity greater than 1, we make the following change to our partial fraction:

$$f(x) = \frac{1}{(x+1)^2} = \frac{A}{x+1} + \frac{B}{(x+1)^2}$$

In general, if you had

$$f(x) = \frac{1}{(x+a)^m}$$

then the decomposition would look like

$$\frac{1}{(x+a)^m} = \frac{A_1}{x+a} + \frac{A_2}{(x+a)^2} + \cdots + \frac{A_m}{(x+a)^m}$$

where each A_k is a constant.

Sometimes, we end up with "irreducable" polynomial factors that only have complex roots. For example, in the case of

$$f(x) = \frac{1}{x(x^2+x+1)}$$

you can check by substitution or by the quadratic formula that the only roots of the degree 2 polynomial in the denominator are $x = (-1 \pm i\sqrt{3})/2$. In partial fraction decomposition, we leave such polynomials alone and do not decompose into monomial and binomial factors. In cases like these, our decomposition looks like this:

$$f(x) = \frac{1}{\color{red}{x}\color{black}{(}\color{blue}{x^2+x+1}\color{black}{)}} = \frac{A}{\color{red}{x}} + \frac{Bx+C}{\color{blue}{x^2+x+1}}$$

Since our denominator is a degree 2 polynomial, we put a degree 1 polynomial in the numerator. In general, if your denominator has an irreducible degree n polynomial, you assume the numerator is degree $n-1$ in the decomposition.

An example combining polynomial degree with multiplicity is as follows:

$$f(x) = \frac{1}{(x^2+x+7)^3} = \frac{Ax+B}{x^2+x+7} + \frac{Cx+D}{(x^2+x+7)^2} + \frac{Ex+F}{(x^2+x+7)^3}$$

Although the approach for solving for these coefficients is fundamentally no different from the "easy" example we worked out earlier, it is certainly more tedious to solve for six unknowns from six linear equations rather than just two. In cases like these, it may be expedient to use matrices or a computer algebra system (free: SageMath, Wolfram Alpha, GNU Octave; also free to make your own computer program!) to help you find the constants. Integrating some of these examples may also prove to be somewhat challenging and will be addressed in another section.

5.14 Integration Strategy: Long Division

In the preceding section, we learned how to use partial fraction decomposition to integrate many rational functions of polynomials. However, in every case we explored previously, the numerator was a polynomial of smaller degree than the denominator. What if we have an integral like this instead?

$$\int \frac{x^5 + 3x^3 - 5x + 2}{x^2 + 2x + 1}\, dx$$

We need to do some extra work to get the integrand into a form where we can employ partial fraction decomposition. In cases like these, simply employ polynomial long division, as discussed in Algebra:

$$
\begin{array}{r|rrrrrr}
 & & & x^3 & -\,2x^2 & +\,6x & -\,10 \\ \hline
x^2+2x+1 & x^5 & & +\,3x^3 & & -\,5x & +\,2 \\
 & -\,x^5 & -\,2x^4 & -\,x^3 & & & \\ \cline{2-4}
 & & -\,2x^4 & +\,2x^3 & & & \\
 & & 2x^4 & +\,4x^3 & +\,2x^2 & & \\ \cline{3-5}
 & & & 6x^3 & +\,2x^2 & -\,5x & \\
 & & & -\,6x^3 & -\,12x^2 & -\,6x & \\ \cline{4-6}
 & & & & -\,10x^2 & -\,11x & +\,2 \\
 & & & & 10x^2 & +\,20x & +\,10 \\ \cline{5-7}
 & & & & & 9x & +\,12
\end{array}
$$

Based on this result, we can rewrite our integral as follows:

$$\int \frac{x^5 + 3x^3 - 5x + 2}{x^2 + 2x + 1}\, dx = \int \left(x^3 - 2x^2 + 6x - 10\right)\, dx + \int \frac{9x + 12}{x^2 + 2x + 1}\, dx$$

In this case, we managed to divide out a polynomial, which is easy to integrate; our only remaining task is to perform partial fraction decomposition on the remainder term.

5.15 Integration Strategy: Complete the Square

Before we begin with this section, we will need a helpful Lemma:

Lemma 5.3 (Derivative of arctan)**.**

$$\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$$

Proof. Let $y = \arctan x$. Since arctan is the inverse of tan, we have

$$\tan y = x$$

We know the derivative of the tangent function:

$$\frac{d}{dx}\tan x = \sec^2 x$$

So, taking the derivative of both sides,

$$\frac{d}{dy}\tan y = \frac{dx}{dy} \implies \sec^2 y = \frac{dx}{dy}$$

We know from trigonometry that $\sec^2 y = 1 + \tan^2 y$, so

$$\frac{dx}{dy} = 1 + \tan^2 y$$

Since $y = \arctan x$ and $\tan(\arctan x) = x$, we have

$$\frac{dx}{dy} = 1 + x^2$$

Finally,

$$\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$$

Q.E.D.

A helpful corollary of this lemma is that

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C$$

which is clear from just "undoing" the derivative. This fact is helpful for working through many different kinds of integrals, especially when used in conjunction with the rule for **completing the square**.

Suppose we want to find the following indefinite integral:

$$\int \frac{1}{x^2+8x+99}\,dx$$

Assuming this polynomial is irreducible, we do not already know many options which we could use to deal with it. One plausible approach is to simplify the denominator of the rational function, which, by far, appears to be the most tricky and complicated part. Since we know how to integrate functions which look like $1/(x^2+1)$, we should try to manipulate it into that form. This is accomplished by completing the square. Recall that to complete the square, we need to add and subtract $(b/2a)^2$. In this case, $a = 1$ and $b = 8$, so

$$\frac{1}{x^2+8x+99} = \frac{1}{x^2+8x+(16-16)+99}$$

If we move our parentheses around a little bit, as permitted by the associative property of addition, then we get

$$\frac{1}{(x^2+8x+16)-16+99}$$

We can re-write $x^2+8x+16$ as $(x+4)^2$ since we "completed the square" and replace $-16+99$ by 83. Thus,

$$\int \frac{1}{x^2+8x+99}\,dx = \int \frac{1}{(x+4)^2+83}\,dx$$

It is time to bring ourselves closer to the arctan integral. Let $u = (x+4)$. Then $du/dx = 1$, so $du = dx$. This gives us

$$\int \frac{1}{(x+4)^2+83}\,dx = \int \frac{1}{u^2+83}\,du$$

We need some way to turn the 83 into a 1. It would be possible if we could factor 83 out of the denominator (and, by extension, out of the entire integral). To accomplish this, we will do yet another subtitution: let $u = w\sqrt{83}$. Then $du = dw\sqrt{83}$. This gives us

$$\int \frac{1}{u^2+83}\,du = \int \frac{\sqrt{83}}{83w^2+83}\,dw = \frac{\sqrt{83}}{83}\int \frac{1}{w^2+1}\,dw = \frac{\sqrt{83}}{83}\arctan w + C$$

Our work is almost done. Now we need to go back from w to u, then from u to x.

$$\frac{\sqrt{83}}{83}\arctan w = \frac{\sqrt{83}}{83}\arctan(u/\sqrt{83}) = \frac{\sqrt{83}}{83}\arctan((x+4)/\sqrt{83})$$

So our final answer is

$$\int \frac{1}{x^2+8x+99}\,dx = \frac{\sqrt{83}}{83}\arctan((x+4)/\sqrt{83}) + C$$

5.16 Integration Strategy: Trigonometric Substitution

Until now, a recurring theme in many of our integration strategies has been using techniques we learned in algebra to turn impossibly difficult integals into something we know how to solve. In analogy, we now aim to put our collection of trigonometric identities to good use in simplifying integrals.

Consider the following:

$$\int \frac{1}{\sqrt{4-x^2}}\,dx$$

The root over the denominator makes this difficult to solve. However, one might observe that $4-x^2$ looks a little bit like $r^2 - r^2\sin^2\theta$, which equals $r^2\cos^2\theta$ by the Pythagorean theorem. In an effort to simplify our integrand, we therefore make the substiution of $x = 2\sin\theta$, which implies $dx = 2\cos\theta\,d\theta$. So,

$$\int \frac{1}{\sqrt{4-x^2}}\,dx = \int \frac{2\cos\theta}{\sqrt{4-4\sin^2\theta}}\,d\theta$$

By the Pythagorean theorem, we have

$$\int \frac{2\cos\theta}{\sqrt{4-4\sin^2\theta}}\,d\theta = \int \frac{2\cos\theta}{\sqrt{4\cos^2\theta}}\,d\theta = \int \frac{2\cos\theta}{2\cos\theta}\,d\theta = \int d\theta = \theta + C$$

Now, we only need to go back from θ coordinates to x. To do this, let's analyze our original substitution:

$$x = 2\sin\theta$$

By rearrangement, we have

$$\sin\theta = \frac{x}{2}$$

In trigonometric substitution, it is often the case that we need to determine all the side lengths of the triangle. Since sine equals opposite divided by hypotenuse, that tells us the opposite side equals x and the hypotenuse equals 2. If we needed, it would be possible to find the adjacent side using the Pythagorean theorem again. This allows us to find the value of any other trigonometric functions that may appear in the solution by expressing them in terms of x. However, our path forward is a bit simpler this time: since we only have the variable θ in our solution, we just need to isolate it in our equation:

$$\sin\theta = \frac{x}{2} \implies \theta = \arcsin\frac{x}{2}$$

So our final solution is

$$\int \frac{1}{\sqrt{4-x^2}}\,dx = \arcsin\frac{x}{2} + C$$

Generally speaking, trigonometric substitution is typically employed as a tool to integrate functions with square roots and other radicals. Consequentially, the most helpful trigonometric identities to remember are

$$\sin^2\theta + \cos^2\theta = 1$$

and

$$\tan^2\theta - \sec^2\theta = 1$$

since they allow you to replace multiple terms in a radical with a single square. In particular, there are three common cases to look out for:

Pattern	Substitution	Result
$\sqrt{a^2 - x^2}$	$x = a\sin\theta$	$\sqrt{a^2 - a^2\sin^2\theta} = a\cos\theta$
$\sqrt{a^2 + x^2}$	$x = a\tan\theta$	$\sqrt{a^2 + a^2\tan^2\theta} = a\sec\theta$
$\sqrt{x^2 - a^2}$	$x = a\sec\theta$	$\sqrt{a^2\sec^2\theta - a^2} = a\tan\theta$

Note, as a technicality, that we need to assume the correct domain for θ to avoid adding an absolute value to e.g. $a\cos\theta$, which could otherwise be either positive or negative and still satisfy equality.

5.17 Integration Strategy: Numerical and Approximative Approaches

5.17.1 Add Rectangles

(planned)

5.17.2 Taylor and Maclaurin Series

(planned - could talk about these earlier, but I think these are interesting to talk about in the context of making difficult problems easier)

5.18 Improper Integrals

(planned - when you want to do integrals over e.g. the entire real number line from $(-\infty, \infty)$, you need to take a limit after finding the antiderivative)

5.19 Analytic Geometry

(planned - an application of integration to finding surface area and volume of unusual shapes bounded by different functions)

5.20 Basic Probability Theory

(planned - certain functions tell you the probability of something happening at a particular value; basically, it is helpful to integrate these functions to find the overall probability of something happening in a certain range)

5.21 Introduction to Ordinary Differential Equations

(planned - extremely important for modelling, majority of theoretical physics)

5.22 Applications of Ordinary Differential Equations

(planned - there are so many applications in science, physics, and engineering that I can't even come close to showing all of them. additionally, there are an abundance of techniques people use to deal with different kinds of differential equations that I can't show. I'll try to highlight a few interesting cases that aren't talked about in physics or math classes as often as other examples)

5.22.1 Rocket Science

(planned - so you can say you know rocket science)

5.22.2 "Romeo and Juliet" Dynamics of Love

(planned - model of social dynamics)

5.22.3 Law of Mass Action

(planned - a favorite of mine and many other chemists; one concept allows you to understand tons of different dynamic equilbria)

Chapter 6

Abstract Linear Algebra

6.1 Fields

6.2 Vector Spaces

6.2.1 Definition and Properties

6.2.2 Subspaces

6.2.3 Linear Independence

6.2.4 Bases

6.3 Linear Transformations

6.3.1 Introduction

6.3.2 Image and Kernel

6.3.3 Rank and Nullity

6.3.4 Invertible Transformations

6.3.5 Composition of Transformations

6.4 Normed Spaces and Inner Product Spaces

6.4.1 Definitions

6.4.2 Orthogonal Subspaces

6.4.3 Orthogonal Projection

6.4.4 Orthonormal Bases

6.5 Determinants

6.5.1 Definition

6.5.2 Inverse by Determinants

Chapter 7

Multivariate Calculus and Vector Analysis

7.1 Functions of Multiple Variables

(planned)

7.2 Vector Valued Functions

(planned)

7.3 Polar and Spherical Coordinates

(planned)

7.4 Partial Derivatives

(planned)

7.5 Multiple Integrals

(planned)

7.6 Vector Analysis

(planned)

7.6.1 Gradient Theorem

(planned)

$$\int_{L\subset\mathbb{R}^n} \nabla f \cdot d\mathbf{r} = f(\mathbf{b}) - f(\mathbf{a})$$

for a curve connecting $\mathbf{a}$ to $\mathbf{b}$

7.6.2 Divergence Theorem

(planned)

$$\int \cdots \int_{V\subset\mathbb{R}^n} (\nabla \cdot \mathbf{F}) dV = \oint \cdots \oint_{\partial V} \mathbf{F} \cdot d\mathbf{S}$$

7.6.3 Curl Theorem

(planned)

$$\int \int_{S\subset\mathbb{R}^3} (\nabla \times \mathbf{F}) \cdot dS = \oint_{\partial S} \mathbf{F} \cdot d\mathbf{r}$$

7.6.4 Green's Theorem

$$\oint_C (f\;dx + g\;dy) = \int \int_D \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx\;dy$$

7.7 Glimpse at Partial Differential Equations

(planned)

7.8 Glimpse at Complex Analysis

(planned)

7.8.1 Sequences and Series

(planned)

7.8.2 Cauchy-Riemann Equations

(planned)

7.8.3 Cauchy's Integral Formula

(planned)

7.9 Glimpse at Calculus on Manifolds

7.9.1 Differential Forms

(planned)

7.9.2 Generalized Stoke's Theorem

(planned)

Chapter 8

Data Analysis and "Machine Learning"

8.1 Introduction

(planned; "supervised" versus "unsupervised" learning, what kinds of questions each generally answers)

8.2 Data Processing

(standardization, variable encoding, strategies)

8.3 Dimensionality Reduction

(just a sample; sought to pick out some that are interesting or commonly used)

8.3.1 Principal Component Analysis (PCA)

8.3.2 Isomap

8.3.3 Uniform Manifold Approximation and Projection (UMAP)

8.4 Clustering

8.4.1 k-Nearest Neighbors (k-NN)

8.4.2 k-means

8.4.3 Spectral Clustering

8.5 Regression and Classification

(planned)