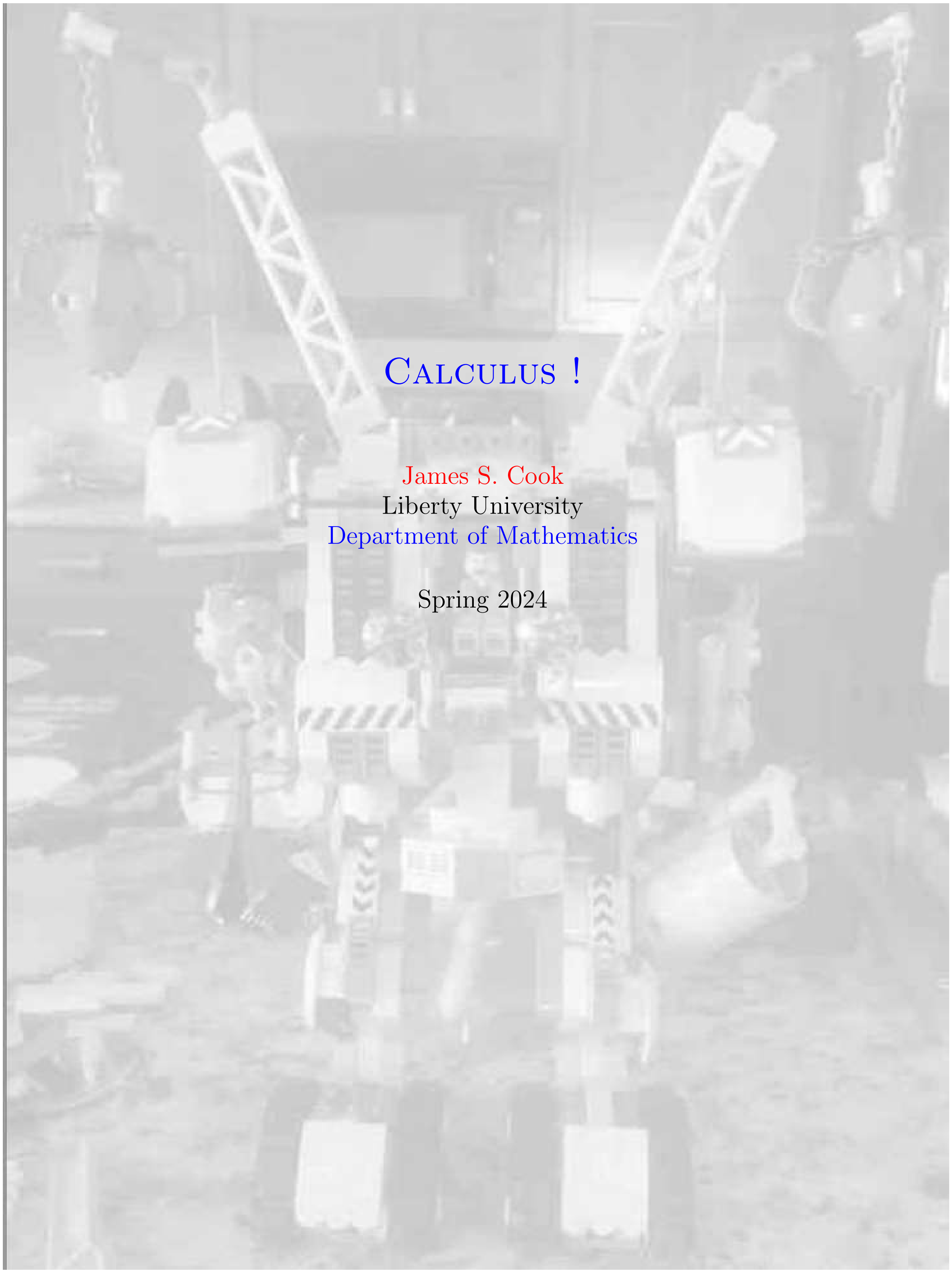


Calculus !

James S. Cook
Liberty University
Department of Mathematics

Spring 2024

preface

how to succeed in calculus

I do use the textbook, however, I follow these notes. You should use both. From past experience I can tell you that the students who excelled in my course were those students who both studied my notes and read the text. They also came to every class and paid attention. I recommend the following course of study:

(1.) Submit yourself to learn, keep a positive attitude. This course is a lot of work. Yes, probably more than 3 others for most people. Most people have a lot of work to do in getting up to speed on real mathematical thinking. There is no substitute for time and effort. If you're complaining in your mind about the workload etc... then you're wasting your time.

(2.) Study my notes.

(3.) Come to class, take notes, **think**.

(4.) Attempt the homework, you will likely find forming a study group is essential for success here.

practical philosophy for this course

Let's begin with several questions:

(1.) what is math?

(2.) how should we understand math?

(3.) how should we do math?

I'll begin with 1, if you listen to the general public you'll get the idea that math is about numbers. For this reason people are often puzzled when they hear about people who are "mathematicians". Can you really make a living just from studying numbers? Well, yes. However, most practicing mathematicians study more abstract aspects of mathematics. We'll just scratch the surface of modern math in the calculus sequence. At this time in history you could spend your whole life studying nothing but math and you would still be missing large portions of mathematics. In a typical math major you'd take courses in: calculus I, II and III, differential equations, complex variables, probability and statistics, discrete math, proofs and logic, linear algebra, abstract algebra and real analysis. In addition, if you're a bit more ambitious you might like to study manifold theory, measure theory, fiber bundles, Lie algebras, Lie groups, topos theory, point-set topology, algebraic topology, homology, category theory, quivers, algebraic geometry, noncommuative geometry, Riemannian geometry, semi-groups, complex analysis, vertex operator algebras, tropical geometry, set theory, modules, inverse problems, variational calculus, differential Galois theory, Galois theory, number theory, combinatorics, partial differential equations, symmetry methods in DEqns, tensor calculus, gauge theory, poisson algebras, homotopy, nonstandard analysis, ... If you searched online you could add to my list. My point? This list is just a tiny subset of the topics which mathematicians continue to actively study. Math is not done. Math is much more than numbers. I'll not attempt a definition of math here, however the concept of definition is probably the most crucial distinguishing feature of math from other fields of study. Mathematical definitions cut much more finely than other fields of study. To know math is to know definitions.

That brings us to item 2, my last sentence needs clarification. Knowledge and understanding are not necessarily the same thing. Many people have knowledge of Christ, few people understand who He is in their heart. Knowledge is necessary but it is not sufficient. How then should we understand mathematics? What process is needed? There is no one answer to this question. Answers include: analyzing historical story which led to the current definition, consistency with other mathematics, seeing how math is applied in the real world, working out examples of a general definition in specific contexts, intuition or creative leaps,... to summarize: all these suggestions boil down to spending time to get to know the math.

Finally we get to the real point here. I suspect you think of math primarily as item 3. Nothing wrong with that based on your experience thus far in math. I'd be surprised if you had a teacher before who emphasized the "why" rather than the "how" of mathematics. This is perhaps the primary distinguishing feature of university calculus: we aspire to calculate with maximal understanding. We ought not use a theorem unless we have an idea of how to prove it. This is our goal. In all the courses I teach in mathematics I attempt to provide proofs for those theorems and propositions which I claim to be true. Granted, there is not always enough time, but we should be ready to give a defense for those truths which we hold dear.

Humility is required from the outset. Some things we cannot understand completely with the tools which are currently at our disposal. Calculus is built with real numbers. I will not attempt to **construct** real numbers from first principles. Instead, our starting point is to assume that real numbers exist, replete with their standard properties. From those rules we will build the calculus.

format of my notes

Please be warned these notes are a work in progress. I look forward to your input on how they can be improved, corrected and supplemented. I prepared them with LaTeXwhich is the standard format for modern mathematical literature. These notes were prepared with LaTeX. You'll notice a number of standard conventions in my notes:

1. definitions are in green.
2. remarks are in red.
3. theorems, propositions, lemmas and corollaries are in blue.
4. proofs start with a **Proof:** and are concluded with a $\square$.
5. often figures in these notes were prepared with **Graph**, a simple and free math graphing program.

Notation	Meaning of Notation
§	section
$\exists$	there exists
$\nexists$	there does not exist
w.r.t.	with respect to
l.h.s.	left hand side
r.h.s.	right hand side
$x \in B$	the element x is inside the set B
$A \implies B$	A implies B
$A \iff B$	A and B are equivalent statements
$\therefore$	therefore
$\forall$	for all
$\equiv$	definition
$\approx$	approximately
$\text{eq}^{\underline{n}}$	equation
$\text{sol}^{\underline{n}}$	solution
$\mathbb{N}$	natural numbers; 1,2,3,...
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers
$\mathbb{Z}$	integers
$\mathbb{R}^2$	the Cartesian plane

I have relegated some proofs to an Appendix, but that doesn't mean they are not important. I do cover some proofs in lecture and such proofs are important to the course either in exhibiting the importance of a given definition or in illustrating a proof technique or calculational method. The **In-Class-Examples** are what they sound like, I've left some spots where I can write in examples which are tailored to the particular conversation sparked in lecture this semester. You can take notes separate from these notes and then write in examples into the whitespace later if that helps you learn better. I don't force note taking in any particular format.

Keep in mind as you read a printed version of the notes that there are hyperlinks to various items throughout this document ranging from interactive websites with adjustable examples courtesy of www.desmos.com to pdf-handouts I've written and posted for a deeper dive into some particular topic or course.

You are free to read whatever you wish about calculus, but keep in mind that this current version of notes is closest to my expectations of argument and logic for this course.

James S. Cook,

November 20, 2023

Contents

Chapter 1

foundations and background

This chapter is unlike the later portions of this course in that there are no proofs. We have three major aims. First, explain the mathematical prehistory which led to calculus. Second, to review and/or introduce terminology. Third, to review calculational techniques which the reader ought to know from previous study of mathematics. I have included textboxes to write examples where they are given in lecture.

1.1 brief history of math

In this section I detail some history which predates the discovery of calculus. The history of math which was discovered and created after calculus is far more vast.

1.1.1 geometry

The ancient Chinese, Greeks, Egyptians and Babylonians all had some understanding of numbers and geometry. Apparently the pythagorean theorem $a^2 + b^2 = c^2$ was known to Babylonians as early as 1700 BC[1]. Pythagorus was one of the earliest Greek mathematicians (572-497 BC) and his followers the pythagoreans were an interesting bunch. They elevated math to a form of mysticism. Their creed was that numbers were the substance of all things. Calculations were tied to music to make the mystic connection between numbers and reality and they used special geometric patterns to aid arithmetic calculations. Plato(429-348 BC) and Aristotle(387-322 BC) advanced the cause of axiomatic reasoning. For mathematics this probably was a good thing. For physics, not so much. Aristotle's flawed physical ideas were so philosophically appealing that we were unable to escape them for over a milennia. Of course, all physical ideas are flawed at some level, Aristotle's physics did explain much, but the explanations were hardly what we could call mathematical. That said, the axiomatic approach did inspire Euclid to make his book of elements at a level of rigor which was valuable to many future generations of mathematicians. Geometry is the perhaps the earliest example of an accurate mathematical model of reality. In fact, for about 2000 years no

[1]see pg. 30 of Katz' *History of Mathematics* second ed., page 45 has nice summary of different societies respective mathematical achievements

one could convincingly imagine any other idea of geometry. The study of physics for things which don't move is called *statics*. The architecture of ancient societies speak to the fact that mathematics were known to at least some in those societies. Probably much has been lost. The history of mathematics is full of multiple discoveries of mathematical theorems, it is common for different mathematicians to find the same theorems even though they never met, or perhaps even lived in the same time.

1.1.2 numbers

What about numbers? The ancients certainly knew about whole numbers and fractions. The phythagoreans took it a step further and realized that there must be more than just numbers of that type. They proved that the hypotenuse of a triangle had a length that need not be a fraction. For example, if you consider a right triangle with side lengths 1 and 1 then the hypotenuse must have length $\sqrt{2}$. They actually proved that $\sqrt{2}$ could not have the form p/q for a pair of whole numbers p, q. One way to understand the development of numbers is to understand the questions which prompted their discovery:

(1.) enumeration or counting leads us to natural numbers and zero.

(2.) accounting leads us to negative numbers since you can either make $$ or lose it.

(3.) fractions come from commerce or manufacture; take a pie and cut it into fractions.

(4.) analytic two-dimensional geometry leads us to irrational numbers; triangles can have irrational side-lengths.

(5.) algebra leads us to complex numbers. The solution to the cubic equation necessitates complex numbers even in the case that the solutions are real.

(6.) three dimensional geometry leads us to quaternions. Hamilton showed how to use quaternions to describe motion in three dimensions.

(7.) quantum mechanics for fields leads us to super numbers. Berezin invoked mathematics which demanded the variables anticommute. Such variables can be thought of as taking values in the super numbers.

There are dozens if not hundreds of other types of numbers. This list is merely reflects my interest in physics. In almost every case when a new type of number was discovered it would be relegated to a lesser status than those earlier known numbers. There was a time when mathematicians would not count negative solutions because they weren't "real" solutions. Later, Kronecker and his followers eschewed use of non-rational numbers. To them the worth of transcendental numbers was in doubt. In my experience students rarely doubt the validity of real numbers. The decimal expansion is quite convincing and we have machines which say it's true so it must be, right? Those same machines will sometimes closemindedly say that $x^2 + 1 = 0$ has no solution. But, $x^2 + 1 = 0$ does have solution. It's just an imaginary solution. Gauss proved that imaginary numbers exist in about 1800. Of course, mathematicians had used complex numbers in one way another for about 200 years before Gauss. This course is primarily focused on real numbers however I will spend some time discussing complex numbers from time to time.

1.1.3 algebra and physics

The connection between physics and algebra is profound. It is this connection that allowed Galileo and Newton to push past the "common sense" of Aristotle. Galileo(1554-1642) studied Archimedes and Aristotle, but he found the later to be illogical. His reaction to his doubt is what changed things, rather than being content to make purely philosophical objections he took it a step further and investigated through experiments to deduce what the correct rules were. For example, through the study of balls rolling down inclines he was able to deduce the formula $y = \frac{1}{2}gt^2$, the height dropped is proportional to the square of the time, *independent of weight*. Galileo's work helped provide a back-drop for Newton and others who were able to explain how Galileo's equations arose from basic physics.

Kepler(1571-1630) also used math to study astronomical data collected by Tycho Brahe over several decades. Upon Brahe's death Kepler tried to fit the data to show the planets traveled in circles around the sun (the heliocentric circular model was proposed to Europe by Nicolaus Copernicus(1473-1543)). However, the data forced Kepler to admit that the planets actually travel in ellipse according to what we now call it Kepler's Laws. In a nutshell, Kepler observed the planets orbit in ellipses while sweeping out equal areas in equal times such that the square of the semi-major axis was proportional to the cube of the period. Obviously, to understand these statements you need to have the idea of Cartesian coordinates. Interestingly, Kepler actually was not so happy about the data's seeming departure from the supposed perfect symmetry of circles. He spend a large amount of his later years trying to fit the solar system into his system of platonic solids. Platonic solids are regular polyhedra which can be inscribed in a sphere: these are associated to the four basic elementals of the ancient greeks: the cube of earth, fire of the tetrahedron, air of the octahedron, water of the icosahedron and over them all the universe of the dodecahedron. Kepler wanted to somehow use the platonic solids to model space. It didn't work. All of this laid the foundation for the discovery of calculus.

I suppose there were two major changes that were in motion at the time just before and including Newton. First, flat earth or earth-centered cosmology was being more and more doubted as evidence mounted for Copernican heliocentric models. The observations of Galileo of moons orbiting Jupiter made the possibility of orbital motion undeniable. Second, the idea that math should be used to phrase physical ideals was encouraged by the methodology of Galileo, Kepler and others. The physical question that would lead Newton to calculus was prompted by all of these events.

1.1.4 discovery of calculus

The term "calculus" apparently originates from the ancient Romans practice of using tiny pebbles to calculate. A calculus was one such pebble. The greeks, chinese and probably others discovered portions of calculus, but none of them possessed a notation which made the ideas accessible to anyone except experts. In contrast, we ordinary mortals can understand calculus without making it our life's work (although, you may feel that way at certain points this semester). Archimedes(287-212 BC) made arguments that very much mirror arguments we have only formalized in the 19-th century. His argument to determine the value for π shows he had an idea much like we will formalize with limits. Beyond limits calculus is largely motivated by two problems:

(1.) what is the tangent line to a given shape?

(2.) what is the area of some shape?

Both of these will be solved carefully this semester by applying appropriate limiting processes. The ancients had no formal method for limits, but they did have some intuitive grasp of limits. The idea of dividing an area into smaller pieces to add together to find the net-area is hardly new to Newton's time. Solutions to various tangent problems also existed before calculus. Isaac Barrow was Newton's teacher before his great discoveries and Barrow did important work on the tangent problem. In fact, Barrow had some understanding of the fundamental theorem of calculus. He understood something about the connection between tangents and areas, however he did not appreciate the importance enough to push the theory forward.

Sir Isaac Newton(1642-1727) was the first to see clearly the connection between these seemingly disparate problems of areas, tangents and physics. In physics, Newton insisted his answers were mathematically phrased. He took Galileo's ideal to a whole new level. He was also unkind to those who refused to follow this route, apparently Hooke said he solved some of the problems Newton solved concerning gravitation. However, Hooke's solution lacked mathematical clarity so Newton rejected his ideas and went so far as to eliminate mention of Hooke in his *Principia*. Newton insisted physical law must be mathematical.

Let me say just a bit more about what distinguished Newton's historical period from that of say Galileo(1554-1642). The representation of irrational numbers by decimal expansions was apparently due to work by the French mathematician Viete (1540-1603), the Dutch mathematician Stevin(1548-1620) and the Scottish mathematician John Napier(1550-1617)[2]. Modern symbolism for algebra was not known to the ancients as far as we know. The compact notation we use today was arrived at through a progression of steps. See Katz' text for details. In a nutshell our notation is due to Viete(1540-1603), Descartes(1596-1650) and Fermat(1601-1665). Descartes' master work set forth a framework in which Newton was free to conduct concrete geometric experiments while the number system put forth by Stevin gave a notation to think about numbers of arbitrarily small magnitude. Basically, the mathematics needed to make calculus happen only arose in the 50 years or so before Newton made his great advances. By the time Newton came of age the ideas of analytic geometry and unending decimal expansions of numbers were taught in the university. In retrospect, Descartes and Fermat were close to the discovery, but they were missing the fundamental theorem of calculus. They understood parts of the puzzle, but Newton and Leibniz grasped the big picture.

Despite the great success of Newton's version of calculus, it was not entirely rigorous. His arguments involved the use of *fluxions* which were strange quantities which were not zero but were really really small. How small you ask? Well, if you divided one by a fluxion then you'd obtain ∞.

[2] page 418 of Katz' text

What were these fluxions and what is ∞? It was easy to set aside these worries because the list of problems that Newton solved grew ever larger as his discoveries came to light in the 17-th century. After postulating his laws of mechanics he was able to derive the formula found by Galileo. Then, prompted by Edmund Halley, he proved Kepler's Laws follow from his universal law of gravitation. Anyway, we could go on about Newton for many pages. Even after all this success there were those mathematicians who were unhappy because at the base of it all these fluxions seemed ad-hoc and not so well-posed mathematically.

Gottfried Wilhelm Leibniz(1646-1716) independently discovered calculus after Newton but published it before him. Leibniz also lacked formal rigor at the base of his theory, but his notation was superior to Newton's and for that reason we still use many notations first introduced by Leibniz.

1.2 set theory

A set is a collection of objects called **elements**. We denote the sentence "x is an element of S" by the short-hand symbolic sentence; "$x \in S$". The sentence "$x \in S$" can also be read "x is in S". For example, $\mathbb{R}$ is the set of all real numbers so to say $x \in \mathbb{R}$ simply means that x is a real number.

Definition 1.2.1. *set equality*

We say $S = T$ when S and T have the all the same elements.

A common notation to characterize the elements of a set is simply to list the elements: for example, $S = \{A, B, C\}$ means that S is a set which contains the objects A, B and C. The ordering of the elements is not special for a general set, this means $S = \{B, A, C\} = \{C, A, B\}$ etc... Often it is difficult or impossible to list all the elements is a set. In such a case we may be able to use **set-builder notation** to define a set:

$$S = \{x \mid x \text{ has property P }\}. \tag{1.1}$$

The set with no elements is called the **empty set** and it is denoted by $\{\}$ or $\emptyset$.

Definition 1.2.2. *subset*

We say S is a subset of T and denote $S \subseteq T$ if for each $s \in S$ we can show $s \in T$.

Notice that set-equality can be conveniently characterized by the concept of a subset. Think about it: $S = T$ means that $S \subseteq T$ and $T \subseteq S$.

Definition 1.2.3. *union, intersection and difference of sets*

Let S and T be sets,

$$S \cup T = \{x \mid x \in S \text{ or } x \in T\} \quad S \cap T = \{x \mid x \in S \text{ and } x \in T\} \quad S - T = \{s \in S \mid s \notin T\}$$

I should introduce some standard notations:

Definition 1.2.4.

- **Natural numbers**; $\mathbb{N} = \{1, 2, 3, \dots\}$.
- **Integers**; $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$.
- **Rational numbers**; $\mathbb{Q} = \left\{\frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0\right\}$.

The notation $\mathbb{Z}_{>0} = \mathbb{N}$ and $\mathbb{Z}_{\geq 0} = \mathbb{N} \cup \{0\}$ and $\mathbb{Z}_{<0} = -\mathbb{N} = \{-x \mid x \in \mathbb{N}\}$ are also encouraged if the reader finds them natural.

In-Class Example 1.2.5. *Let $S = \{1, 2, 3\}$ and $T = \{-1, 0, 1\}$. Find $S \cup T$ and $S \cap T$. Determine which of $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{Q}$ contains $S \cup T$ as a subset.*

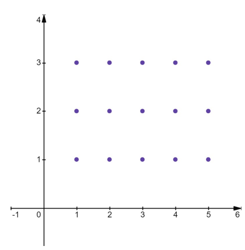
Given a pair of elements we can form an *ordered pair*. A key property of the ordered pair (x, y) is that $(x, y) = (a, b)$ if and only if $x = a$ and $y = b$.

Definition 1.2.6. *Cartesian Product:*

If A, B are sets then $A \times B = \{(x, y) \mid x \in A, y \in B\}$.

We use the notation $A \times A = A^2$ for the set of ordered pairs from A. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is the set of ordered pairs of real numbers. We naturally associate $\mathbb{R}^2$ with the (x, y)-plane.

Example 1.2.7. *Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 3\}$ then $A \times B$ has 15 elements and can be visualized as a grid of points. Click here to see the Desmos code for creating the diagram below*



In-Class Example 1.2.8. *Using A, B as above, is $(1, 5) \in A \times B$? Is $(5, 1) \in A \times B$?*

1.3 real numbers

Real numbers can be constructed from set theory and about a semester of mathematics. We will accept the following as **axioms**[3]

Definition 1.3.1. *The set of real numbers is denoted $\mathbb{R}$ and is defined by the following axioms:*

(A1) addition commutes; $a + b = b + a$ for all $a, b \in \mathbb{R}$.

(A2) addition is associative; $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{R}$.

(A3) zero is additive identity; $a + 0 = 0 + a = a$ for all $a \in \mathbb{R}$.

(A4) additive inverses; for each $a \in \mathbb{R}$ there exists $-a \in \mathbb{R}$ and $a + (-a) = 0$.

(A5) multiplication commutes; $ab = ba$ for all $a, b \in \mathbb{R}$.

(A6) multiplication is associative; $(ab)c = a(bc)$ for all $a, b, c \in \mathbb{R}$.

(A7) one is multiplicative identity; $a1 = a$ for all $a \in \mathbb{R}$.

*(A8) multiplicative inverses for nonzero elements;
for each $a \neq 0 \in \mathbb{R}$ there exists $\frac{1}{a} \in \mathbb{R}$ and $a\frac{1}{a} = 1$.*

(A9) distributive properties; $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all $a, b, c \in \mathbb{R}$.

(A10) totally ordered field; for $a, b \in \mathbb{R}$:

- *(i) antisymmetry; if $a \leq b$ and $b \leq a$ then $a = b$.*
- *(ii) transitivity; if $a \leq b$ and $b \leq c$ then $a \leq c$.*
- *(iii) totality; $a \leq b$ or $b \leq a$*

(A11) least upper bound property: every nonempty subset of $\mathbb{R}$ that has an upper bound, has a least upper bound. This makes the real numbers ***complete****.*

Modulo A11 and some math jargon this should all be old news. An **upper bound** for a set $S \subseteq \mathbb{R}$ is a number $M \in \mathbb{R}$ such that $M > s$ for all $s \in S$. Similarly a lower bound on S is a number $m \in \mathbb{R}$ such that $m < s$ for all $s \in S$. If a set S is bounded above and below then the set is said to be **bounded**. The **intervals** of $\mathbb{R}$ are defined next:

Definition 1.3.2. *Intervals of $\mathbb{R}$ include $\mathbb{R} = (-\infty, \infty)$ as well subsets of $\mathbb{R}$ of the form:*

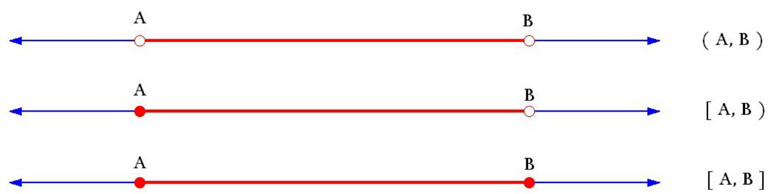
- *open interval from a to b; $(a, b) = \{x | a < x < b\}$.*
- *half-open interval; $(a, b] = \{x \mid a < x \leq b\}$.*
- *half-open interval; $[a, b) = \{x \mid a \leq x < b\}$.*
- *closed interval; $[a, b] = \{x \mid a \leq x \leq b\}$.*
- *closed ray from a to ∞; $[a, \infty) = \{x \mid x \geq a\}$.*
- *closed ray from $-\infty$ to a; $(-\infty, a] = \{x \mid x \leq a\}$.*

[3] an axiom is a basic belief which cannot be further reduced in the conversation at hand. If you'd like to see a construction of the real numbers from other math, see Ramanujan and Thomas' *Intermediate Analysis* which has the construction both from the so-called Dedekind cut technique and the Cauchy-class construction.

- *open ray from a to ∞; $(a, \infty) = \{x \mid x > a\}$.*
- *open ray from $-\infty$ to a; $(-\infty, a) = \{x \mid x < a\}$.*

Example 1.3.3. $[a, b], (a, b), [a, b), (a, b]$ *all have a as a lower bound and b as an upper bound. The set $(-\infty, a)$ has least upper bound a, but is unbounded below. Likewise, (a, ∞) has greatest lower bound a, but is unbounded above.*

Another ubiqitous concept is the **number line**[4]. In the diagram below I picture some of the standard intervals. We use solid bold dots to indicate the point is included in the set, whereas an open dot indicates that point is excluded.

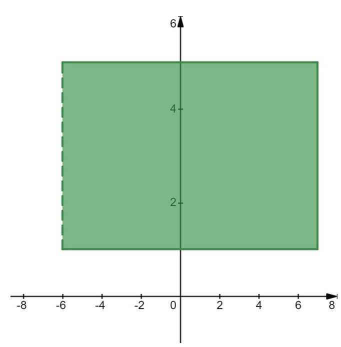


The cartesian product of $\mathbb{R}$ and $\mathbb{R}$ gives us $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$. In this context (x, y) is called an **ordered pair** of real numbers. Notice that the notation (a, b) could refer to a point in $\mathbb{R}^2$ or it could refer to a open interval. These are very different objects yet we use the same notation for both. The point $(a, b) \in \mathbb{R}^2$ whereas the interval $(a, b) \subseteq \mathbb{R}$.

In-Class Example 1.3.4. *Question: is $(4, 3)$ a point or an open interval?*

We can also appreciate the distinction between intervals and points by picturing the cartesian product of intervals as a subset of $\mathbb{R}^2$:

Example 1.3.5. *I created (via Desmos) the picture below to illustrate the subset* $(-6, 7] \times [1, 5]$



[4]Rene' Descartes popularized this concept in the early 17^{th} century; the number line is the foundation of analytical geometry. The fundamental idea in analytic geometry is that there is a 1-1 correspondence between lines and the real numbers.

The real numbers and rational numbers are examples of **fields**. A field is a set which satisfies axioms A1-A9. In fact, both $\mathbb{Q}$ and $\mathbb{R}$ are ordered fields which means follow axioms A1-A10. However, the rational numbers are not complete, they do not satisfy A11. For example, $\pi \notin \mathbb{Q}$ and $\sqrt{2} \notin \mathbb{Q}$. If you take the set of truncated decimal expansions of $\sqrt{2}$; $S = \{1, 1.4, 1.41, \dots\}$ then $S \subseteq \mathbb{Q}$ yet the least upper bound of S is $\sqrt{2} \notin \mathbb{Q}$. The rational numbers are missing something we need for making analytical arguments. Fortunately, the completion[5] of the rational numbers gives us all of $\mathbb{R}$ which does satisfy A11. Certainly A11 is the most technical of all the axioms of $\mathbb{R}$ and it is also the property which is crucial to many central theorems of calculus[6].
It is useful to catalogue the following properties of inequalities:

Theorem 1.3.6. *properties of inequalities:*

Let $a, b, c, d \in \mathbb{R}$,

(1.) $a^2 \geq 0$ and $a^2 = 0$ if and only if $a = 0$,

(2.) If $a < b$ then $a + c < b + c$ and $a - c < b - c$,

(3.) If $a < b$ and $c < d$ then $a + c < b + d$,

(4.) If $a < b$ and $b < c$ then $a < c$,

(5.) if $c > 0$ then $a < b$ implies $ca < cb$.

(6.) if $c < 0$ then $a < b$ implies $ca > cb$.

You should not be surprised to hear that a similar theorem also holds if we replace $<$ with $>$ or $\leq$ with $\geq$ as appropriate throughout.

Definition 1.3.7. *absolute value*

The **absolute value** of a real number x is denoted $|x|$ and is defined $|x| = \sqrt{x^2}$. Notice this formula is equivalent to the case-wise formula:

$$|x| = \begin{cases} x & if\ x \geq 0 \\ -x & if\ x \leq 0 \end{cases}$$

This **distance** between $a, b \in \mathbb{R}$ is denoted $d(a, b)$ and it is defined by $d = |b - a|$.

[5]if you're curious, and this comment is certainly from beyond the proper scope of this course, completion of the rational numbers is formed by adjoining the limit points of all Cauchy sequences of rational numbers. Moreover, the completion of $\mathbb{Q}$ can be done in a rather different way to give the pyadics, much of real analysis also can be done in that context. Ask me if interested, I can recommend some reading.

[6]I should mention, there are all sorts of nonstandard ideas about calculus on sets other than $\mathbb{R}$, there is fractional calculus, discrete calculus, and for the past decade or so I've tinkered in something called $\mathcal{A}$-calculus where $\mathcal{A}$ is an algebra which could be complex numbers or other stranger things

Notice that $|x|$ is the **distance from the origin to** x; $d(0,x) = |x-0| = |x|$. I should also point out our custom is that the square root function is by definition the positive root; $\sqrt{x} \geq 0$. Notice we can characterize a nonzero **positive** number by the equation $x = |x|$ whereas a nonzero **negative** number x has $|x| = -x$.

Theorem 1.3.8. *properties of absolute value:*

Let $a, b, \varepsilon \in \mathbb{R}$ with $\varepsilon > 0$,

(1.) absolute value is non-negative; $|a| \geq 0$,

(2.) absolute value is zero only if number is zero; $|a| = 0$ iff $a = 0$,

(3.) absolute value of product is product of absolute values; $|ab| = |a||b|$,

(4.) bounded absolute value unwraps as compound inequality; $|a| < \varepsilon \Leftrightarrow -\varepsilon < a < \varepsilon$,

(5.) unbounded absolute value unwraps into two disjoint cases;

$$|a| > \varepsilon \Leftrightarrow a < -\varepsilon \;\text{ or }\; a > \varepsilon.$$

(6.) triangle inequalities ;

$$(i.)\ |a+b| \leq |a| + |b| \qquad (ii.)\ |a-b| \geq |a| - |b| \qquad (iii.)\ \big||a| - |b|\big| \leq |a-b|$$

It is probably useful to study the geometric significance of the theorem on absolute values. Note that $|x| = |x-0| = d(x,0)$**; absolute value gives the distance to the origin**. This makes (4.) and (5.) easy to understand.

In-Class Example 1.3.9. *Let* $A = \{x \in \mathbb{R} \mid |3x+12| < 3\}$. *Express* A *as an interval. Also, let* $B = \{x \in \mathbb{R} \mid |3x+12| \geq 3\}$ *and express* B *as the union of two intervals. Find* $A \cup B$*, and discuss.*

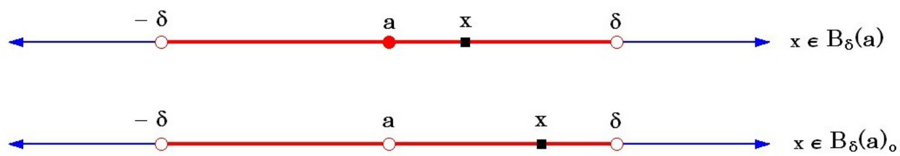
Definition 1.3.10. *neighborhoods*

An **open neighborhood** centered at a with radius $\delta > 0$ is denoted $B_\delta(a)$ where

$$B_\delta(a) = \{x \in \mathbb{R} \mid d(a,x) < \delta\} = (a - \delta, a + \delta).$$

An **deleted open neighborhood** centered at a with radius $\delta > 0$ is denoted $B_\delta(a)_o$ where

$$B_\delta(a)_o = \{x \in \mathbb{R} \mid 0 < d(a,x) < \delta\} = (a - \delta, a) \cup (a, a + \delta).$$



The concept of a deleted neighborhood will be central to the study of limits. [7]

In-Class Example 1.3.11. *Find a and δ for which $B_\delta(a) = (3, 11)$. Hint: a is the midpoint of $B_\delta(a)$ and the length of $B_\delta(a)$ on the number line is twice the radius δ.*

We would sometimes like to insist that a give set of real numbers has no holes. In other words, you can draw the set as a connected line-segment or ray on the number line.

Definition 1.3.12. *connected subsets of real numbers.*

We say $U \subseteq \mathbb{R}$ is connected if and only if

$$U \in \{\mathbb{R}, (-\infty, a), (a, \infty), (-\infty, a], [a, \infty), [a, b], (a, b], [a, b), (a, b), \{a\}\}$$

for some $a, b \in \mathbb{R}$ where $a < b$.

The definition I gave above is rather clumsy, but I believe it should be readily understood by calculus students[8]. Next, we sometimes need the concept of a boundary point. In a nut-shell a boundary point is a point on the edge of a set.

---

[7] I'm using a B for neighborhood because it matches a notation I'll use later for studies of higher dimensional open sets: generally, $B_\delta(a) = \{x \in \mathbb{R}^n \mid ||x - a|| < \delta\}$ is an open-ball of radius δ in n-dimensional space. Also, be warned that the concept of a neighborhood varies from text to text.

[8] Equivalently, you could say U is connected iff there do not exist U_1, U_2 such that $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = U$. A pair of sets like U_1, U_2 is called a **separation** of U.

Definition 1.3.13. *boundary points.*

We say $p \in U \subseteq \mathbb{R}$ is a **boundary point** of U if and only if every open neighborhood centered at p intersects points in $\mathbb{R} - U$ and U. In other words, boundary points of U are positioned so that they are close to points both inside and outside U. We denote the boundary of U by $bd(U)$. If $a \in bd(U)$ and there exists $\varepsilon > 0$ such that

- $(a - \varepsilon, a) \subset U$ then a is a **right boundary point** of U,
- $(a, a + \varepsilon) \subset U$ then a is a **left boundary point** of U.

Notice that a boundary point of U need not be in U; for example $U = (0, 1]$ has $bd(U) = \{0, 1\}$ and $0 \notin U$. On the other hand, it is possible for the whole set to be comprised of boundary points: $bd(\mathbb{N}) = \mathbb{N}$. We can break down any set of real numbers into two types of points:

(1.) boundary points

(2.) interior points

For example, $[0, 1) = \{0\} \cup (0, 1)$. We have $bd[0, 1) = \{0, 1\}$ whereas $int([0, 1) = (0, 1)$.

Definition 1.3.14. *interior points.*

Suppose $U \subset \mathbb{R}$ then we say $p \in U$ is an **interior point** of U if there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. The set of all interior points of U is denoted $int(U)$.

Note $int(\mathbb{N}) = \emptyset$ whereas $int(0, 1) = (0, 1)$. In contrast, $bd(\mathbb{N}) = \mathbb{N}$ and $bd(0, 1) \cap (0, 1) = \emptyset$. Finally, we have all the terminology necessary to carefully define an open set:

Definition 1.3.15. *open sets, closed sets.*

We say $U \subseteq \mathbb{R}$ is an **open set** if and only if each point in U is an interior point. Likewise, we say U is a **closed set** if and only if $U = U \cup \partial U$.

A closed set contains all its boundary points whereas an open set contains only interior points.

In-Class Example 1.3.16.

1.4 functions

In this section we review terminology for functions.

Definition 1.4.1.

Let $A, B \subseteq \mathbb{R}$. We say $f : A \to B$ is a **function** if for each $x \in A$ the rule for f assigns a single-element $f(a) \in B$. In particular, if $f(a) = b_1$ and $f(a) = b_2$ then $b_1 = b_2$.

- The **domain** is the set of inputs for f; $\text{dom}(f) = A$.
- The **range** is the set of outputs for f; $\text{range}(f) = \{f(x) \mid x \in A\}$.
- The **graph** of f is given by $\text{graph}(f) = \{(x, y) \mid x \in \text{dom}(f), y = f(x)\}$.

The *rule* for f could be given by a formula, a given graph, diagram or even a table of values. When f is defined by a formula $f(x)$ our *convention* is to let the domain for f be the set of all x for which $f(x)$ is a real number. The graph of a function must satisfy the **vertical line test**; the intersection of $\text{graph}(f)$ and $x = a$ has at most one point.

In-Class Example 1.4.2. *Let* $f_1(x) = \sqrt{3 - x}$*,* $f_2(x) = \frac{1}{x^2 - 4}$ *and* $f_3(x) = 7$*.*
Find the natural domains for f_1, f_2 *and* f_3*.*

In-Class Example 1.4.3. *Consider the two graphs given in lecture. Which is the graph of a function ? For the function graph, find the domain and range.*

Definition 1.4.4.

If $f, g : U \to \mathbb{R}$ are functions then we say $f = g$ if $f(x) = g(x)$ for all $x \in U$. Otherwise, we say the functions are not equal and write $f \neq g$.

In other words, two functions are equal if they *pointwise* agree. We are very picky about the "for all" part above. If there is even one point of disagreement then the functions are not equal.

In-Class Example 1.4.5. *Let $f(x) = x$ and $g(x) = \frac{x^2}{x}$. Explain why $f \neq g$.*

Given functions f, g we define new functions $f + g, f - g, fg, \frac{f}{g}, f \circ g$ by the rules:

$$(f \pm g)(x) = f(x) \pm g(x), \quad (fg)(x) = f(x)g(x), \quad \frac{f}{g}(x) = \frac{f(x)}{g(x)}, \quad (f \circ g)(x) = f(g(x)). \tag{1.2}$$

We should notice the product of functions commutes; $fg = gf$. However, the composition of functions need not commute: $f \circ g \neq g \circ f$. The next example illustrates these observations.

In-Class Example 1.4.6. *Let $f(x) = 3x + 2$ and $g(x) = 7 - 4x$. Calculate formulas for $(fg)(x), (f \circ g)(x)$ and $(g \circ f)(x)$.*

Definition 1.4.7.

- **image of S under f** is given by $f(S) = \{f(x) \mid x \in S\}$.
- **inverse image of T under f** is given by $f^{-1}(T) = \{x \text{ dom}(f) | f(x) \in T\}$.

Graphically, the image of $S \subseteq \text{dom}(f)$ is the set of y-values attained from S-inputs. In contrast, the inverse image of T is the set of all x-values in the domain for which the y-values fall in T.

In-Class Example 1.4.8. *Forward and inverse image from a given graph.*

Example 1.4.9. *Suppose $f(x) = 3x + 2$. Observe that:*

$$f([0, 2]) = \{3x + 2 \mid x \in [0, 2]\} = [2, 8].$$

On the other hand, $f^{-1}([0, 2])$ is the set of $x \in \mathbb{R}$ such that

$$3x + 2 \in [0, 2] \quad \Rightarrow \quad 0 \leq 3x + 2 \leq 2 \quad \Rightarrow \quad -2 \leq 3x \leq 0 \quad \Rightarrow \quad -2/3 \leq x \leq 0.$$

We find $f^{-1}([0, 2]) = [-2/3, 0]$.

Example 1.4.10. *Let $f(x) = x^2$ then*

$$f([0, 2]) = \{x^2 \mid x \in [0, 2]\} = [0, 4] \qquad \& \qquad f^{-1}([0, 4]) = [-2, 2].$$

Definition 1.4.11. *Let $A, B \subseteq \mathbb{R}$ and $f : A \to B$ be a function, then f is*

- **onto** *if range$(f) = f(A) = B$.*
- **one-to-one** *if for all $x_1, x_2 \in A$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.*

The term **surjective** is sometimes used in place of onto and the term **injective** is synonomous with one-to-one. Notice that a function which is one-to-one must pass the **horizontal line test**;

Horizontal line test: no horizontal line intersects the graph of f at two or more points.

We call $f^{-1}\{y\} = \{x \in \text{dom}(f) \mid f(x) = y\}$ the **fiber** of f over y, so another way to characterize injectivity is to say all the nonempty fibers are singletons[9]. One-to-one means that each output value is attained by just one input value.

Example 1.4.12. *Let $f(x) = mx + b$ where $m \neq 0$. If $f(x_1) = f(x_2)$ then $mx_1 + b = mx_2 + b$ implies $mx_1 = mx_2$ implies $x_1 = x_2$ thus f is one-to-one. Indeed, the graph $y = mx + b$ is a non-horizontal line and it is graphically clear this passes the horizontal line test.*

Definition 1.4.13. *Let $A, B \subseteq \mathbb{R}$ and $f : A \to B$ be a one-to-one and onto function, then $f^{-1} : B \to A$ is the* **inverse function** *defined by $f^{-1}(y) = x$ if and only if $f(x) = y$.*

A given function and its inverse are related by the following equations:

$$f^{-1}(f(x)) = x \qquad \& \qquad f(f^{-1}(y)) = y \tag{1.3}$$

for all $x \in \text{dom}(f) = \text{range}(f^{-1})$ and $y \in \text{range}(f) = \text{dom}(f^{-1})$. These equations are important since they allow us to effectively cancel f or f^{-1} by composition with f^{-1} or f as appropriate.

[9] a singleton is a set which contains just one point

Example 1.4.14. *If $f(x) = 10^x$ then the inverse function is known as the the* **common log** *or* **log base 10** *and we denote $f^{-1}(y) = \log(y)$. If $10^x = 10^{3-x}$ then we may solve by taking the common log of the equation; $\log(10^x) = \log(10^{3-x})$ implies $x = 3 - x$ thus $2x = 3$ and we find $x = 3/2$.*

A good problem to stretch your algebra muscles is found in calculating the inverse function for a given function. **Algorthim:** to find $f^{-1}(y)$ solve $y = f(x)$ for x.

In-Class Example 1.4.15. *Let $f(x) = \dfrac{3x+1}{2x-7}$. Find $f^{-1}(y)$. Also, find domain and range for f.*

Example 1.4.16. *Let $f(x) = x^2 - 4$ then $y = x^2 - 4$ yields $x^2 = y + 4$ and thus $x = \pm\sqrt{y+4}$. Thus f is not invertible. Indeed, f is not one-to-one since $y = x^2 - 4$ fails the horizontal line test.*

Example 1.4.17. *Let $f(x) = x^n$ for some $n \in \mathbb{N}$. If n is even then $y = f(x)$ fails horizontal line test. If n is odd then $y = x^n$ has unique solution $x = y^{1/n}$ thus $f^{-1}(y) = y^{1/n}$.*

I should mention, $(x^n)^{1/n} = |x| = \pm x$ and we sometimes use the radical notation $x^{1/n} = \sqrt[n]{x}$. It is often useful to narrow the domain of a given function to make it one-to-one.

Definition 1.4.18. *Let $A, B \subset \mathbb{R}$ and $f : A \to B$ be a function and $S \subseteq A$ then we say that g is the* **restriction of f to S** *if $g : S \to B$ and $g(x) = f(x)$ for all $x \in S$. We write $g = f|_S$ to indicate that g is the restriction of f to the set S. If g is a restriction of f then we say that f is an* **extension** *of g.*

We can use restriction to carefully define the concept of a **local inverse**.

Definition 1.4.19. *Let $A, B \subset \mathbb{R}$ and $f : A \to B$ be a function and $S \subseteq A$ such that $f|_S$ is a one-to-one function. If $f(S) = T$ then $g : T \to S$ is defined by $g(y) = x$ if and only if $f(x) = y$. We say g is a* **local inverse** *of f with respect to S and write $g = f|_S^{-1}$.*

Notice the condition that $f|_S$ is one-to-one implies the equation $f(x) = y$ has a unique solution x.

In-Class Example 1.4.20. *Let $f(x) = x^2 - 4$. Find an S for which the local inverse $f|_S^{-1}$ exists.*

Example 1.4.21. *The trigonometric functions $f(x) = \sin x$, $g(x) = \cos x$ and $h(x) = \tan x$ are not one-to-one on their domains. When we talk about inverse functions for sine, cosine and tangent we're actually talking about local inverses. In particular,*

$$\sin^{-1} = f\Big|^{-1}_{[-\pi/2,\pi/2]}, \qquad \& \qquad \cos^{-1} = g\Big|^{-1}_{[0,\pi]} \qquad \& \qquad \tan^{-1} = h\Big|^{-1}_{(-\pi/2,\pi/2)}.$$

If you're calculator is in radian mode, you will observe range(sin^{-1}) $= [-\pi/2, \pi/2]$ whereas range($\cos^{-1}$) $= [0, \pi]$ and range(tan^{-1}) $= (-\pi/2, \pi/2)$.

Appendix 7.4 has pictures which explain the example above in some depth. The local inverses of the trigonometric functions are especially useful because of the periodicity of the trigonometric functions. For example:

In-Class Example 1.4.22. *Find the solution set of $\sin x = 1/2$. Note, $\sin^{-1}(1/2) = \pi/6$. Picture $y = \sin x$ to aid in finding the solution set.*

Let me give some precise language to describe graphs going up or down:

Definition 1.4.23. *Given function $f : U \subseteq \mathbb{R} \to \mathbb{R}$ and $I \subseteq U$ then f is*

- **increasing** *on $I \subseteq U$ if $a, b \in I$ with $a < b$ implies $f(a) \leq f(b)$,*
- **strictly increasing** *on $I \subseteq U$ if $a, b \in I$ with $a < b$ implies $f(a) < f(b)$,*
- **decreasing** *on $I \subseteq U$ if $a, b \in I$ with $a < b$ implies $f(a) \geq f(b)$,*
- **strictly decreasing** *on $I \subseteq U$ if $a, b \in I$ with $a < b$ implies $f(a) > f(b)$,*

If f is either increasing, strictly increasing, decreasing or strictly decreasing on $I = dom(f)$ then we say that f is respectively increasing, strictly increasing, decreasing or strictly decreasing. If f is either increasing or decreasing then we say f is **monotonic***.*

If $f(x) = c$ for all $x \in I$ then we say f is **constant** on I. In view of our definition a constant function is both increasing and decreasing.

Definition 1.4.24. *Suppose f is a function such that $x \in dom(f)$ implies $-x \in dom(f)$ then*

- *we say f is* **even** *if $f(-x) = f(x)$ for all $x \in dom(f)$,*
- *we say f is***odd** *if $f(-x) = -f(x)$ for all $x \in dom(f)$.*

Notice the graph of an even function f has $(x, y) \in \text{graph}(f)$ only if $(-x, y)$ is also in the graph. Such graphs are said to be **symmetric with respect to the y-axis**. In contrast, the graph of an odd function is said to be **symmetric with respect to the origin**.

In-Class Example 1.4.25. *Let $n \in \mathbb{N}$ and suppose $f(x) = x^{2n-1}$, $g(x) = x^{2n}$, $h(x) = 1 + x$. Determine if the given functions are even, odd or neither.*

If $x, -x \in dom(f)$ for each $x \in dom(f)$) then we can express f as the sum of an even and odd function since $f(x) = \underbrace{\frac{1}{2}[f(x) + f(-x)]}_{\text{even}} + \underbrace{\frac{1}{2}[f(x) - f(-x)]}_{odd}$. For example:

Example 1.4.26. *Even and odd parts of e^x are the* **hyperbolic cosine** *and* **hyperbolic sine**

$$e^x = \underbrace{\tfrac{1}{2}[e^x + e^{-x}]}_{\cosh x} + \underbrace{\tfrac{1}{2}[e^x - e^{-x}]}_{\sinh x}$$

In-Class Example 1.4.27. *Show that* $(\cosh x)^2 - (\sinh x)^2 = 1$.

I'll conclude by reminding you how we can create new graphs from old by standard transformations:

Theorem 1.4.28.

Let f be a function and suppose $c > 0$ then

- if $g(x) = f(x) + c$ then graph(g) is a **upward translation** of graph(f) by c-units.
- if $g(x) = f(x) - c$ then graph(g) is a **downward translation** of graph(f) by c-units.
- if $g(x) = f(x - c)$ then graph(g) is a **right translation** of graph(f) by c-units.
- if $g(x) = f(x + c)$ then graph(g) is a **left translation** of graph(f) by c-units.
- if $g(x) = -f(x)$ then graph(g) is the **relection across the x-axis** of graph(f).
- if $g(x) = f(-x)$ then graph(g) is the **relection across the y-axis** of graph(f).
- if $g(x) = cf(x)$ then graph(g) is the **vertical dilation** of graph(f) by a factor of c. If $0 < c < 1$ then the graph is compressed whereas if $c > 1$ then the graph is stretched.
- if $g(x) = f(x/c)$ then graph(g) is the **horizontal dilation** of graph(f) by a factor of c. If $0 < c < 1$ then the graph is compressed whereas if $c > 1$ then the graph is stretched.

In-Class Example 1.4.29. *Consider $y = 3 - \sqrt{x - 1}$. Use transformations to obtain this graph from the known $y = \sqrt{x}$ graph.*

1.5 catalogue of elementary functions

The functions discussed in this section are central to our study of calculus.

1.5.1 power functions

We say f is a **power function** if $f(x) = x^a$ where a is a fixed constant. There are a few special cases with added labels,

(1.) $a = n \in \mathbb{N}$ then $f(x) = x^n$ is a homogeneous polynomial.

(2.) $a = \frac{1}{n}$ with $n \in \mathbb{N}$ then $f(x) = x^{\frac{1}{n}} \equiv \sqrt[n]{x}$ is the n^{th}-root function.

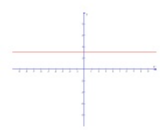
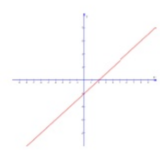
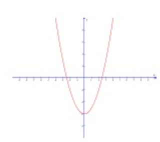
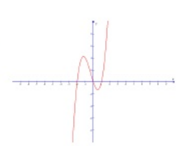
(3.) $a = -1$ then $f(x) = \frac{1}{x}$ is the **reciprocal function**.

1.5.2 polynomial functions

We say p is a **polynomial function** of degree n if it has the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots a_1 x + a_o \tag{1.4}$$

where $a_n \neq 0$ and we call $a_n, a_{n-1}, \dots, a_o \in \mathbb{R}$ the **coefficients** of the polynomial. w1here and we call the coefficients of the polynomial. The set of all polynomials in the variable x is denoted $\mathbb{R}[x]$. To say $p(x) \in \mathbb{R}[x]$ is to say $p(x)$ is a polynomial.

Formula	Name	Zeros	Graph of p
$p(x) = c$	constant function	None, unless $c = 0$ in which case there are infinitely many.	
$p(x) = mx + b$	linear function	$x = -\frac{b}{m}$ we assume $m \neq 0$.	
$p(x) = ax^2 + bx + c$	quadratic function	$x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$ if $b^2 - 4ac \geq 0$, $a \neq 0$.	
$p(x) = ax^3 + bx^2 + cx + d$	cubic function	No simple formula. There is always one zero. In some cases there are 3 zeros.	

1.5.3 rational functions

We say that f is a **rational function** if it has the form $f(x) = p(x)/q(x)$ for a pair of polynomial functions p and q. The zeros of f occur at the zeros of p if anywhere. However, it is possible that a zero of p is also a zero of q in which case the point could be a zero, a hole in the graph or a vertical asymptote. The domain of a rational function is all the points where we avoid division by zero;

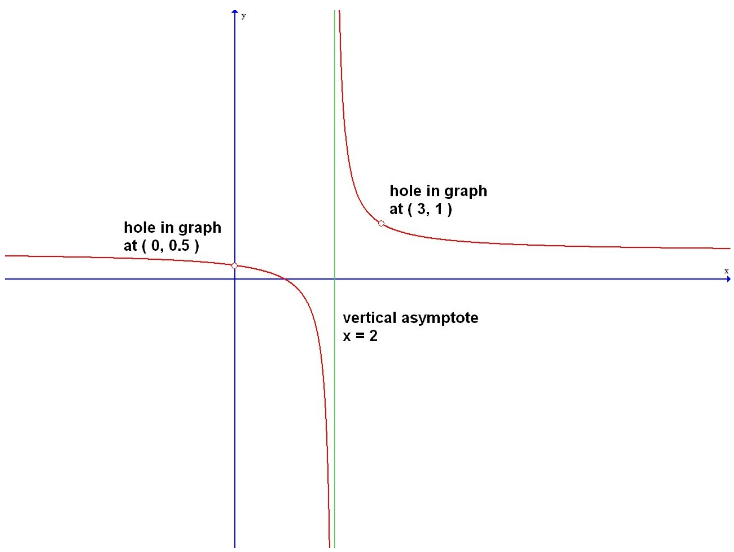
$$dom\left(\frac{p}{q}\right) = \{x \in \mathbb{R} \mid q(x) \neq 0\}.$$

The reciprocal function is a rational function.

Example 1.5.1. *A typical example of a rational function is*

$$f(x) = \frac{x(x-1)(x-3)}{x(x^2-5x+6)}$$

this function has a hole in the graph at zero and three. It has a vertical asymptote at $x = 2$. It has a zero at $(1, 0)$.

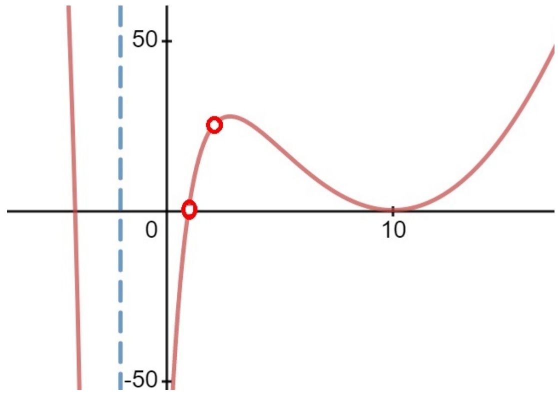


Note, $dom(f) = (-\infty, 0) \cup (0, 2) \cup (2, 3) \cup (3, \infty) = \mathbb{R} - \{0, 2, 3\}$.

Example 1.5.2. *Consider $f(x) = \dfrac{(x-4)(x-1)^2(x-10)^2(x-2)}{(x+2)^2(x-1)(x-2)}$ this function has* **algebraic critical numbers** $-4, -2, 1, 2, 10$. *There are holes in its graph at $(1, 0)$ and $(2, 24)$. It also has a vertical asymptote at $x = -2$. The reduced function in this case is obtained from cancelling the $(x-1)$ and $(x-2)$ factors in the denominator to obtain:*

$$f_{red}(x) = \frac{(x-4)(x-1)^2(x-10)^2(x-2)}{(x+2)^2(x-1)(x-2)}$$

which has no holes in the graph, however, $x = 2$ is still a veritcal asymptote for $y = f_{red}(x)$. I made a graph in Desmos, you can explore it here.



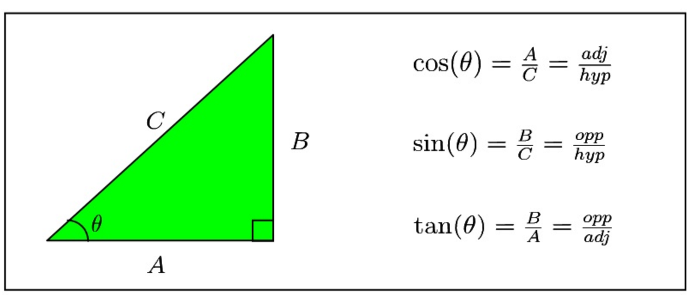
1.5.4 algebraic functions

We say that f is an **algebraic function** if it has a formula which is comprised of finitely many algebraic operations. By *algebraic* we mean you may add, subtract, multiply, divide and raise to powers or take roots. This category of functions includes power functions, polynomial functions and rational functions and a host of more complicated functions built from root-based formulas. For example, $f(x) = \sqrt{x^2} = |x|$ defines an algebraic function.

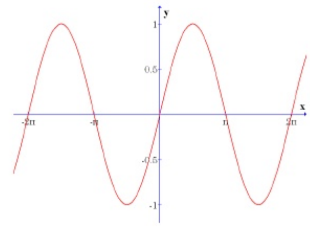
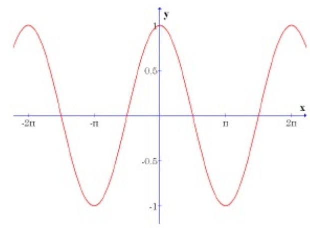
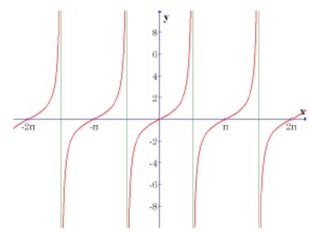
In-Class Example 1.5.3. *Casewise defined functions can be algebraic functions in disguise. Find an algebraic formula for the* $f(x) = \begin{cases} x^2 & x \geq 0 \\ -x^2 & x < 0 \end{cases}$

1.5.5 trigonometric functions

Trigonometric functions such as sine, cosine and tangent are based on the geometry of triangles. Recall a right triangle is one for which an angle measures 90 degrees (or radians, or 100 grads, etc...).



In the picture above we assume that $A, B, C > 0$ and we have drawn the triangle so that $0 < \theta < \pi/2$, it is an *acute* angle. You may recall that the side A is **adjacent** to the angle θ while the side B is **opposite** the angle θ. The longest side C is called the **hypotenuse**. We know the **Pythagorean Theorem** states $A^2 + B^2 = C^2$.

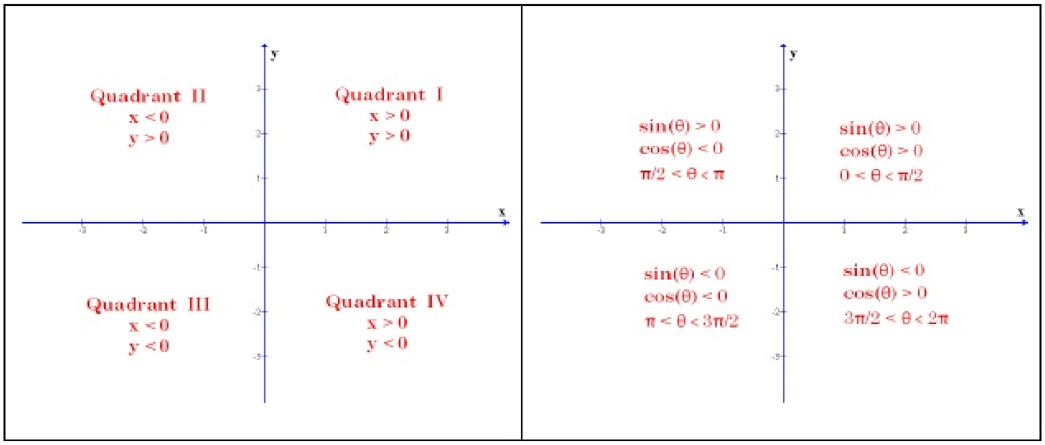
notation	name	Zeros	graph
$\sin(x)$	sine	$x = 0, \pm\pi, \pm 2\pi, \ldots$ Equivalently, $x = n\pi,\ n \in \mathbb{Z}$	
$\cos(x)$	cosine	$x = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \ldots$ Equivalently, $x = n\pi + \frac{\pi}{2},\ n \in \mathbb{Z}$	
$\tan(x)$	tangent	Same as sine. The green lines are the vertical asymptotes which happen where cosine is zero.	

These functions extend the quadrant I geometric quantities to the other three quadrants. The definitions also make **polar coordinates** work. The polar coordinates of $P = (x, y)$ are r, θ where

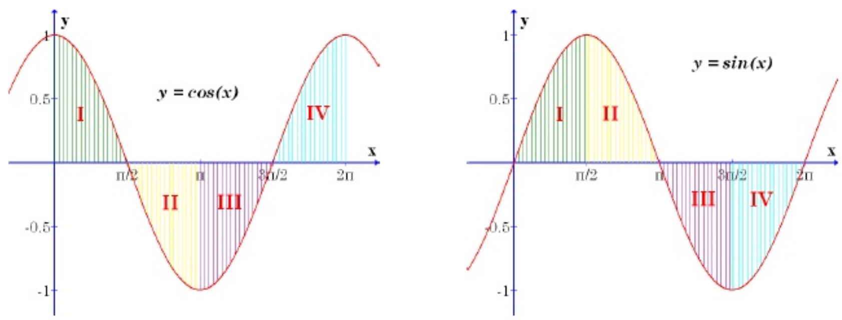
$$\boxed{x = r\cos(\theta) \qquad y = r\sin(\theta), \qquad r^2 = x^2 + y^2, \qquad \tan(\theta) = \frac{y}{x}}$$

and we call r the **radial coordinate** and θ is the **standard angle**. There are a number of conventions as to what particular values the polar coordinates should be allowed to take. We usually[10] insist that $r \geq 0$ but make no particular restriction on θ, this means that $r = \sqrt{x^2 + y^2}$ however θ is not uniquely defined for a given point because we can always add a integer multiple of 2π and still get the same point. The xy-plane is divided into four quadrants. See below how the sine and cosine of the standard angle θ matches the signs of $\sin(\theta)$ and $\cos(\theta)$.

[10] admittedly , we do allow $r < 0$ when we discuss polar graphing later in the calculus sequence. There is some ambiguity about what is meant by polar coordinates, I simply made a choice here.



Or perhaps the following diagrams make more sense to you,



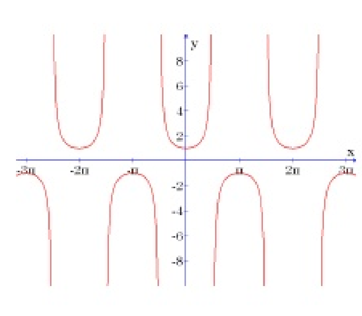
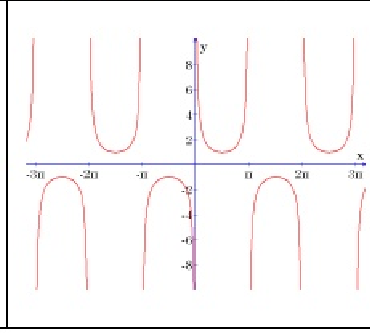
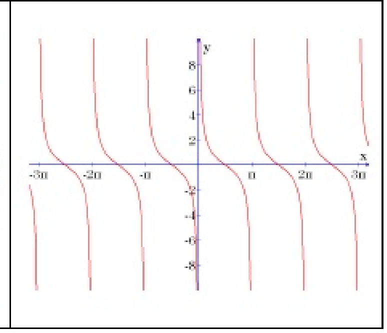
Since $r = \sqrt{x^2 + y^2} \geq 0$ we see that the formulas $x = r\cos(\theta)$ and $y = r\sin(\theta)$ reproduce the correct signs for the Cartesian coordinates x and y. My point here is simply that sine and cosine not only include basic geometric ratios about triangles, they also encode the signs of the Cartesian coordinates in all four quadrants.

1.5.6 reciprocal trigonometric functions

Reciprocal trigonometric functions: these appear quite often in difficult integrations. **Secant, cosecant** and **cotangent** are defined to be one over the functions cosine, sine and tangent respectively. We use the notation,

$$\boxed{\sec(\theta) = \frac{1}{\cos(\theta)} \qquad \csc(\theta) = \frac{1}{\sin(\theta)} \qquad \cot(\theta) = \frac{\cos(\theta)}{\sin(\theta)}}$$

The graphs of these functions are given below:

Graph of $y = \sec(x)$	Graph of $y = \csc(x)$	Graph of $y = \cot(x)$
		

In-Class Example 1.5.4. *We know* $\cos^2\theta + \sin^2\theta = 1$. *Find identities for the recipocal trig functions by dividing by* $\cos^2\theta$ *and also* $\sin^2\theta$.

1.5.7 inverse trigonometric functions

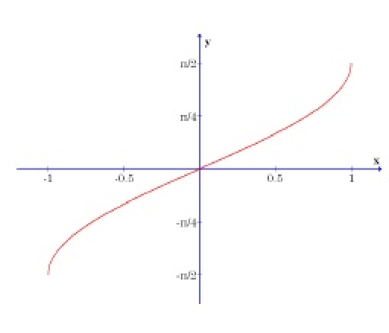
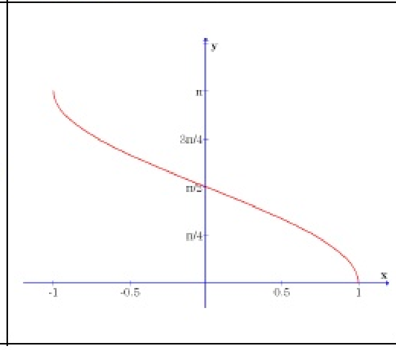
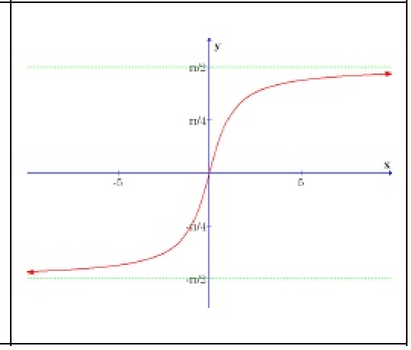
Inverse trigonometric functions: we should be careful to distinguish the inverse trigonometric functions from the reciprocal trig functions. The inverse trig functions are denoted by $\sin^{-1}, \cos^{-1}$ and $\tan^{-1}$ which I refer to as **inverse sine, inverse cosine** and **inverse tangent** respectively. They satisfy the equations,

$$\boxed{\sin^{-1}(\sin(x)) = x \qquad \cos^{-1}(\cos(y)) = y \qquad \tan^{-1}(\tan(z)) = z}$$

for $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $y \in [0, \pi]$ and $z \in (-\frac{\pi}{2}, \frac{\pi}{2})$

$$\boxed{\sin(\sin^{-1}(x)) = x \qquad \cos(\cos^{-1}(y)) = y \qquad \tan(\tan^{-1}(z)) = z}$$

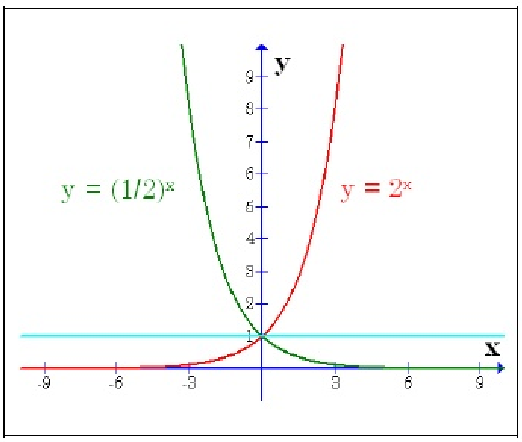
For $x \in [-1, 1]$, $y \in [-1, 1]$ and $z \in \mathbb{R}$. Let us collect the graphs of the inverse trig functions for future reference.

Graph of $y = \sin^{-1}(x)$	Graph of $y = \cos^{-1}(x)$	Graph of $y = \tan^{-1}(x)$
		

The green lines illustrate horizontal asymptotes of inverse tangent. The occur at $y = \pi/2$ and $y = -\pi/2$. These are all local inverses, this is the reason the "inverse tangent" failed to provide us the correct angle outside quadrants I and IV. The inverse tangent function is only truly the inverse of tangent in quadrants I and IV for $-\pi/2 < \theta < \pi/2$.

1.5.8 exponential functions

Exponential functions: let $a > 0$ then we say that is an **exponential function** if $f(x) = a^x$ for each $x \in \mathbb{R}$. The fixed number a is called the **base** of the exponential function. Exponential functions are nonzero everywhere. The graph below shows the three shapes an exponential function may take.



If $a > 1$ then $f(x) = a^x$ gives us exponential growth. If $0 < a < 1$ then $f(x) = a^x$ gives us exponential decay. The graph appears to get to zero, but this is not the case, exponential functions never reach zero. We see that if $a \neq 1$,

$$\boxed{dom(a^x) = (-\infty, \infty) \qquad range(a^x) = (0, \infty)}$$

If $f(x) = e^x$ then this is **the exponential function**, more often than not we will work with this particular base, the number $e \approx 2.71\ldots$ is called **Euler's number** in honor of the famous mathematician Euler. It is a *transcendental number* which means it is defined by an equation which transcends simple algebra. We will discuss e^x is some depth in later chapters.

1.5.9 logarithmic functions

Logarithmic functions: these are the inverse functions of the exponential functions. Suppose $a > 1$, we say that $f(x) = \log_a(x)$ is a logarithmic function, and that the **log base a of x** (this is how we verbalize the formula when we're talking out the math) satisfies the following equations,

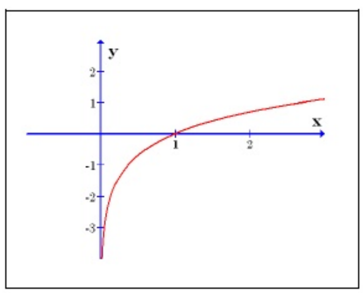
$$\boxed{\log_a(a^x) = x \qquad a^{\log_a(x)} = x}$$

In this sense the logarithm and exponential functions cancel. An equivalent way to define the logarithm is to say that if $y = a^x$ then $\log_a(y) = x$. Notice that the input of the logarithm must be positive since $a^{\log_a(x)}$ is positive; $dom(\log_a(x)) = 0, \infty)$.

$$\boxed{dom(\log_a(x)) = (0, \infty) \qquad range(\log_a(x)) = (-\infty, \infty)}$$

The **natural log function** is denoted $\ln(x)$, this the logarithmic function with base $e = 2.71\ldots$ that simply means $\log_e(x) = \ln(x)$. This particular logarithmic function is so important that it gets its own notation. We will encounter it frequently in later chapters.

The graph of $y = \ln(x)$ shows that the natural log has one zero at $x = 1$.



We can see that $dom(\ln(x)) = (0, \infty)$ and the $range(\ln(x)) = (-\infty, \infty)$.

The following table has useful identities:

PROPERTIES OF EXPONENTIALS AND LOGARITHMS:

We assume that $a, b > 0$ in the equations that follow. I assume that you know these formulas and how to use them. Technically there is no need for the equations in the bottom two squares since they are the same as the top two once we set $a = e$ .For your convenience I include them.

$a^{x+y} = a^x a^y.$	$\log_a(xy) = \log_a(x) + \log_a(y)$
$(a^x)^y = a^{xy}$	$\log_a(x^c) = c\log_a(x)$
$a^{-x} = \frac{1}{a^x}$	$\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$
$a^{x-y} = \frac{a^x}{a^y}$	$\log_a(a) = 1$
$(ab)^x = a^x b^x$	$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$
$e^{x+y} = e^x e^y.$	$\ln(xy) = \ln(x) + \ln(y)$
$(e^x)^y = e^{xy}$	$\ln(x^c) = c\ln(x)$
$e^{-x} = \frac{1}{e^x}$	$\ln(\frac{x}{y}) = \ln(x) - \ln(y)$
$e^{x-y} = \frac{e^x}{e^y}$	$\ln(e^x) = x$
$e^{\ln(x)} = x$	$\log_a(x) = \frac{\ln(x)}{\ln(a)}$

In-Class Example 1.5.5. *Express* a^x *and* $\log_a(x)$ *in terms of the exponential and log base* e.

1.5.10 hyperbolic functions

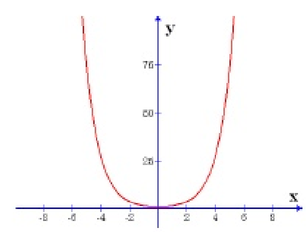
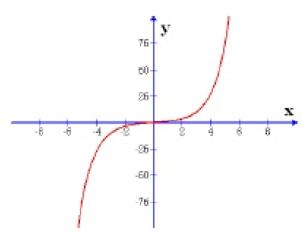
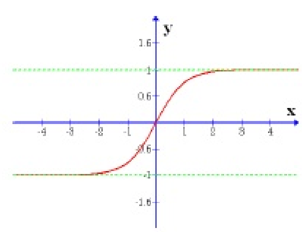
Hyperbolic functions: these are little less common then some of the other functions we have discussed so far, however they are useful both for certain questions of integration and also Einstein's special relativity.

1. **hyperbolic cosine**: $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$,
2. **hyperbolic sine**: $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$,
3. **hyperbolic tangent**: $\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$.

At first glance it is a little strange to call these *trigonometric*, that label comes from an understanding of cosine and sine in terms of imaginary exponentials e^{ix} where $i = \sqrt{-1}$. We will discuss imaginary exponentials in due time. For now just observe that

$$\boxed{\cosh^2(x) - \sinh^2(x) = 1.}$$

This is clearly similar to the corresponding identity $\cos^2(x) + \sin^2(x) = 1$. We also note that $\cosh(0) = 1$ and $\sinh(0) = 0$, these identities make hyperbolic cosine and sine a better choice of notation than e^x and e^{-x} for certain questions.

graph of $y = \cosh(x)$	graph of $y = \sinh(x)$	graph of $y = \tanh(x)$
		

The **inverse hyperbolic functions** are $\cosh^{-1}(x), \sinh^{-1}(x)$ and $\tanh^{-1}(x)$. These satisfy the formulas,

$$\boxed{\cosh(\cosh^{-1}(x)) = x \qquad \sinh(\sinh^{-1}(y)) = y \qquad \tanh(\tanh^{-1}(z)) = z}$$

for $x \in [1, \infty)$, $y \in \mathbb{R}$ and $z \in (-1, 1)$ and,

$$\boxed{\cosh^{-1}(\cosh(x)) = x \qquad \sinh^{-1}(\sinh(y)) = y \qquad \tanh^{-1}(\tanh(z)) = z}$$

for $x \in [0, \infty)$, $y \in \mathbb{R}$ and $z \in \mathbb{R}$. The hyperbolic sine and tangent functions are injective so they have a global inverse. In contrast, the hyperbolic cosine is not injective and it is customary to let

$cosh^{-1}(x)$ denote the local inverse for hyperbolic cosine restricted to $[0, \infty)$.

The inverse hyperbolic functions can be expressed with logs of algebraic functions:

$$\cosh^{-1}(x) = \ln(x + \sqrt{x^2 - 1}) \qquad \text{for } x \geq 1.$$

$$\sinh^{-1}(x) = \ln(x + \sqrt{x^2 + 1}) \qquad \text{for } x \in \mathbb{R}.$$

$$\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \qquad \text{for } |x| < 1.$$

In-Class Example 1.5.6. *Derive one or more of the above identities.*

1.6 algebra

I'll focus primarily on problems of polynomial algebra in this section. Let us begin with the basic terminology, a **polynomial in standard form** is an **expression** of the form

$$p(x) = a_n x^n + \cdots + a_1 x + a_0 \tag{1.5}$$

where $a_0, a_1, \dots, a_n \in \mathbb{R}$ are called **coefficients** and if the leading coefficient $a_n \neq 0$ then we write $\deg(p) = n$ to indicate the **degree** of $p(x)$ is n. If the leading coefficient $a_n = 1$ then we call $p(x)$ a **monic polynomial**. Some special cases of note:

- for $m \neq 0$ the polynomial $p(x) = mx + b$ is a **linear polynomial**
- for $a \neq 0$ the polynomial $p(x) = ax^2 + bx + c$ is a **quadratic polynomial**
- for $a \neq 0$ the polynomial $p(x) = ax^3 + bx^2 + cx + d$ is a **cubic polynomial**

If there exist polynomials $p_1(x), p_2(x), \dots, p_r(x)$ with $\deg(p_j) \geq 1$ such that

$$p(x) = p_1(x)p_2(x)\cdots p_r(x) \tag{1.6}$$

then the expression above is called a **factorization** of $p(x)$ with **factors** $p_1(x), p_2(x), \dots, p_r(x)$. Up to this point I have only discussed **expressions**, let us now turn our attention to **equations**. A **polynomial equation of degree** n is an equation of the form

$$a_n x^n + \cdots + a_1 x + a_0 = 0 \tag{1.7}$$

where $a_n, \dots, a_1, a_0 \in \mathbb{R}$ and $a_n \neq 0$. Solving such an equation is in general very challenging and may even defy closed form solution[11]. Solving the polynomial equation $p(x) = 0$ is greatly simplified if we have a factorization of $p(x)$. Observe

$$p(x) = p_1(x)p_2(x)\cdots p_r(x) = 0 \tag{1.8}$$

is true if there is some $p_j(x) = 0$. Therefore, to solve $p(x) = 0$ we can instead solve:

$$p_1(x) = 0, \quad p_2(x) = 0, \quad \cdots \quad p_r(x) = 0. \tag{1.9}$$

In-Class Example 1.6.1. *Solve $x^3 - 2x^2 - 5x + 6 = 0$.*
Hint: $x^3 - 2x^2 - 5x + 6 = (x-1)(x+2)(x-3)$.

[11] only for degree $n = 1, 2, 3, 4$ are their general solutions, for $n \geq 5$ there does not exist a general formula in terms of the coefficients via elementary algebra. This was proved by Abel in about 1819 then greatly clarified by the work of Galois around 1830.

Theorem 1.6.2. *Polynomial long division.*

If polynomials $p(x), q(x)$ have $\deg(p) > \deg(q)$ then there exist polynomials $f(x)$ and $r(x)$ such that

$$p(x) = f(x)q(x) + r(x)$$

and either $r(x) = 0$ or $\deg(r) < \deg(q)$. We call $r(x)$ the **remainder** and $f(x)$ the **polynomial part** of $p(x)/q(x)$.

There is a standard algorithm to calculate $f(x)$ and $r(x)$ as in the theorem above. For example:

Example 1.6.3.

$$\begin{array}{r|l}
 & \quad\quad\quad\quad\quad\quad x \quad +3 \\
\hline
x^3-3 & x^4 + 3x^3 + x^2 + 2x \quad +6 \\
 & -x^4 \quad\quad\quad\quad\quad +3x \\
\hline
 & \quad\quad 3x^3 + x^2 + 5x \quad +6 \\
 & \quad -3x^3 \quad\quad\quad\quad\quad +9 \\
\hline
 & \quad\quad\quad\quad x^2 + 5x + 15
\end{array}$$

This reveals that $x^4 + 3x^3 + x^2 + 2x + 6 = (x + 3)(x^3 - 3) + x^2 + 5x + 15$. *Note the remainder* $r(x) = x^2 + 5x + 15$ *whereas the polynomial part is* $x + 3$. *In addition,*

$$\frac{x^4 + 3x^3 + x^2 + 2x + 6}{x^3 - 3} = x + 3 + \frac{x^2 + 5x + 15}{x^3 - 3}. \tag{1.10}$$

Once we know polynomial long division, the remainder theorem naturally follows:

Theorem 1.6.4. *remainder theorem*

Let $c \in \mathbb{R}$ and $p(x)$ be a polynomial. Then, $p(x) = (x - c)f(x) + r$ and $r = p(c)$.

Example 1.6.5.

$$\begin{array}{r|l}
 & \quad\quad\quad x^2 + 2x \\
\hline
x+1 & x^3 + 3x^2 + 2x + 7 \\
 & -x^3 \; - x^2 \\
\hline
 & \quad\quad 2x^2 + 2x \\
 & \quad -2x^2 - 2x \\
\hline
 & \quad\quad\quad\quad\quad\quad\quad 7
\end{array}$$

Shows $p(x) = x^3 + 3x^2 + 2x + 7 = (x + 1)(x^2 + 2x) + 7$. *Note* $p(-1) = -1 + 3 - 2 + 7 = 7$.

Theorem 1.6.6. *factor theorem*

Let $c \in \mathbb{R}$ and $p(x)$ be a polynomial. Then, $p(x) = (x - c)f(x)$ if and only if $p(c) = 0$.

Proof: Suppose $p(x) = (x - c)f(x)$ then $p(c) = (c - c)f(c) = 0$. Conversely, if $p(c) = 0$ then by the remainder theorem $p(x) = (x - c)f(x)$. $\square$

In-Class Example 1.6.7. *Let* $p(x) = x^4 - 5x^3 + 5x^2 + 5x - 6$. *Factor* $p(x)$.
Hint: consider $-1, 1, 2, 3$ *as potential zeros for* $p(x)$.

In-Class Example 1.6.8. *Let* $p(x) = x^3 + 3x^2 + x - 5$. *Factor* $p(x)$.
Hint: $p(1) = 0$ *and* $x^2 + 4x + 5$ *is a factor of* $p(x)$.

It turns out that $x^2 + 4x + 5$ cannot be factored using real polynomials. In general, a quadratic polynomial can be factored via the technique of completing the square, but it is always fastest to guess and check or to use $A^2 - B^2 = (A - B)(A + B)$ or $(A + B)^2 = A^2 + 2AB + B^2$.

In-Class Example 1.6.9. *Factor* $p(x) = x^2 + 3x + 2$.

In-Class Example 1.6.10. *Factor* $p(x) = 16x^2 - 9$.

In-Class Example 1.6.11. *Factor* $p(x) = x^2 + 6x + 9$.

To complete the story of factoring over $\mathbb{R}$ we need to introduce complex numbers.

Definition 1.6.12. *complex numbers and terminology*

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

We say $z = x + iy$ where $x, y \in \mathbb{R}$ is in **cartesian form** where $Re(z) = x$ is the **real part** of z and $Im(z) = y$ is the **imaginary part** of z. The **complex conjugate** of z is given by $\bar{z} = x - iy$. The **length** of $z = x + iy$ is denoted $|z| = \sqrt{x^2 + y^2}$ and since $z\bar{z} = x^2 + y^2$ we have $|z| = \sqrt{z\bar{z}}$. If $z \neq 0$ then $\dfrac{1}{z} = \dfrac{\bar{z}}{z\bar{z}} = \dfrac{x - iy}{x^2 + y^2}$.

Complex numbers share all the same *algebraic properties* as real numbers; Axioms 1-9 from Definition 1.3.1 hold for $\mathbb{C}$ provided we multiply complex numbers as follows:

$$(a + ib)(c + id) = ac - bd + i(ad + bc). \tag{1.11}$$

There is much more to say about $\mathbb{C}$ and you can read Appendix 7.3 if you are curious.

In-Class Example 1.6.13. *Let $z = 3 + 2i$. Find the cartesian form of $\frac{1}{z} + iz$.*

Note that since $-i^2 = 1$ we may express a sum of squares as a difference of squares;

$$A^2 + B^2 = A^2 - i^2B^2 = A^2 - (iB)^2 = (A - iB)(A + iB) \tag{1.12}$$

We should keep the identity above in mind as we study how to factor a quadratic.

In-Class Example 1.6.14. *Factor $x^2 + 9$.*

In-Class Example 1.6.15. *Factor $9(x + 1)^2 + 16$.*

Theorem 1.6.16. *completing the square*

The quadratic polynomial $p(x) = ax^2 + bx + c$ can be expressed as $p(x) = a(x-h)^2 + k$ for $h = -b/2a$ and $k = p(h)$.

The technique of completing the square has basically two steps:

- factor out a: $p(x) = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)$,
- group terms with x:

$$p(x) = ax^2 + bx + c = a\left[\left(x + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a}\right],$$

In the examples below, the polynomials have $a = 1$ so the first step is not needed.

In-Class Example 1.6.17. *Complete the square for $p(x) = x^2 + 2x + 5$.*

In-Class Example 1.6.18. *Complete the square for $p(x) = x^2 + 10x + 25$.*

Example 1.6.19. *Since $A^2 - B^2 = (A-B)(A+B)$ we can factor after completing the square:*

$$\begin{aligned} p(x) &= x^2 - 6x - 1 \\ &= (x-3)^2 - 9 - 1 \\ &= (x-3)^2 - (\sqrt{10})^2 \\ &= (x - 3 - \sqrt{10})(x - 3 + \sqrt{10}). \end{aligned}$$

In-Class Example 1.6.20. *Complete the square for $p(x) = x^2 - 4x + 6$ and use the concept from Equation 1.12 to factor $p(x)$.*

We observe three possible patterns following from the mechanics of completing the square; either we obtain constant term which is negative, zero or positive. Correspondingly, the factorization of the quadratic features distinct real factors, a repeated real factor or a pair of complex factors.

Theorem 1.6.21. *factoring $ax^2 + bx + c$ and solving $ax^2 + bx + c = 0$.*

If $a, b, c \in \mathbb{R}$ and $a \neq 0$ then $p(x) = a(x - r_+)(x - r_-)$ where

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

- **distinct real factors** if $b^2 - 4ac < 0$ then $p(x) = a(x - r_+)(x - r_-)$
- **repeated real factor** if $b^2 - 4ac = 0$ then $p(x) = a(x - r)^2$
- **irreducible over $\mathbb{R}$, conjugate complex factors** if $b^2 - 4ac > 0$ then

$$p(x) = a((x - \alpha)^2 + \beta^2) = a(x - \alpha - i\beta)(x - \alpha + i\beta).$$

Likewise, to solve $ax^2 + bx + c = 0$ we find solutions $x = r_\pm$. If $b^2 - 4ac \geq 0$ then a real solution exists, however if $b^2 - 4ac < 0$ then there is no real solution. In the case $b^2 - 4ac < 0$ the solutions form a conjugate pair: $\overline{r_+} = r_-$.

Pragmatically, you could just use the formula for $r_\pm$ to factor $ax^2 + bx + c$ as indicated. That said, completing the square on specific examples is almost always faster... unless you can see how to factor by guessing or some other special form.

In-Class Example 1.6.22. *Factor $p(x) = x^2 + 3x + 2$ and solve $p(x) = 0$.*

Theorem 1.6.23. *complex solutions come in conjugate pairs and correspond to irreducible factors.*

If $p(x) = a_n x^n + \cdots + a_1 x + a_0$ is a polynomial with coefficients $a_n, \ldots, a_1, a_0 \in \mathbb{R}$ and $p(\alpha + i\beta) = 0$ where $\alpha, \beta \in \mathbb{R}$ then $p(\alpha - i\beta) = 0$ and $(x - \alpha)^2 + \beta^2$ is a factor of $p(x)$.

In-Class Example 1.6.24. *Factor $p(x) = x^4 - 4x^3 + 6x^2 - 4x + 5$ over $\mathbb{R}$ given that $p(2 + i) = 0$. Hint: the theorem above indicates that $(x - 2)^2 + 1 = x^2 - 4x + 5$ is a factor of $p(x)$.*

In-Class Example 1.6.25. *Reverse example time. Write a polynomial $p(x)$ of least degree for which $p(1) = 0$, $p(-3) = 0$ and $p(3 + 2i) = 0$ such that $p(0) = 7$. Leave in factored form!*

Theorem 1.6.26. *fundamental theorem of algebra*

> If $p(x)$ is a polynomial with real coefficients then it can be factored into a product of linear and irreducible quadratic factors. Each irreducible quadratic factor can be split into a pair of linear factors corresponding to a conjugate pair of complex zeros to the equation $p(x) = 0$.

Example 1.6.27.

$$p(x) = x^4 - 1 = (x^2+1)(x^2-1) = \underbrace{(x^2+1)}_{\textit{irred. quad.}}\underbrace{(x+1)(x-1)}_{\textit{linear}} = \underbrace{(x+i)(x-i)(x+1)(x-1)}_{\textit{complex linear factorization}}$$

Example 1.6.28.

$$p(x) = x^2 + 4x + 13 = \underbrace{(x+2)^2 + 9}_{\textit{irred. quad}} = \underbrace{(x+2)^2 - (3i)^2}_{\textit{difference of squares}} = \underbrace{(x+2-3i)(x+2+3i)}_{\textit{conjugate factors}}$$

Example 1.6.29.

$$p(x) = x^4 + 4x^3 + 3x^2 = x^2(x^2 + 4x + 3) = x^2(x+1)(x+3)$$

Finally, there are occasions where Pascal's Triangle and the special form identities for $(A+B)^n$ are useful. I'll share the first few of these for reference:

$$\begin{aligned}
(A+B)^2 &= A^2 + 2AB + B^2 \qquad (1.13)\\
(A+B)^3 &= A^3 + 3A^2B + 3AB^2 + B^3\\
(A+B)^4 &= A^4 + 4A^3B + 6A^2B^2 + 4AB^3 + B^4\\
(A+B)^5 &= A^5 + 5A^4B + 10A^3B^2 + 10A^2B^3 + 5AB^4 + B^5.
\end{aligned}$$

In-Class Example 1.6.30. *Copy Pascal's triangle from the board. Also, factor $x^3 - 6x^2 + 12x - 8$.*

1.7 sign-chart method to solve inequalities

The logical justification for the techniques used in this section is provided later in this course when we study continuity. It turns out that a theorem due to a 19-th century Jesuit priest named Bolzano justifies carefully how a function may change signs from positive to negative. Long story short, if we are dealing with a polynomial or a rational function then the sign changes can only occur at vertical asymptotes, holes in the graph or simply a zero of the function. We call numbers where the function is either zero or undefined *algebraic critical numbers.*

Definition 1.7.1. *algebraic critical numbers.*

Suppose $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}$ is a function then we say $c \in dom(f) \cup bd(dom(f))$ is an **algebraic critical number** iff either $c \notin dom(f)$ or $f(c) = 0$.

I have added the qualifier "*algebraic*" to distinguish this concept from a later technical meaning we ascribe to the term critical point[12].

The guiding principle of this section is that a function can only change signs at algebraic critical numbers. Therefore, if we draw a number line with the algebraic critical points labeled and draw little $\pm$'s to indicate the sign of the function then we can roughly sketch the function and also quickly read solutions to inequalities. That's the big idea, let's see how it is implemented.

Example 1.7.2. *Suppose $f(x) = x^2 + x - 6$. Find solution of $x^2 + x - 6 \geq 0$. Notice that we can factor $f(x) = (x+3)(x-2)$ thus $f(-3) = 0$ and $f(2) = 0$. Pick tests points to the left and right of each algebraic critical number and evaluate the function. In this case, easy* **choices** *are*

$$f(-4) = (-1)(-6) = 6, \qquad f(0) = -6, \qquad f(3) = (6)(1) = 6$$

hence the following sign chart is derived:

We find $x^2 + x - 6 \geq 0$ if $x \in (-\infty, -3] \cup [2, \infty)$. As an additional application of this sign chart, suppose you were asked to find the domain of $g(x)$ which is defined implicitly by the following formula:

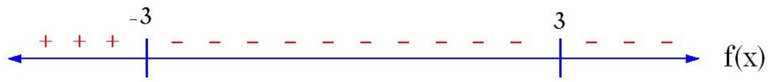
$$g(x) = \frac{1}{\sqrt{6 - x - x^2}}.$$

We would require $x \in dom(g)$ if and only if $6 - x - x^2 > 0$. But, this is the same as stating $x \in dom(g)$ if and only if $x^2 + x - 6 < 0$ hence, by the sign chart, $dom(g) = (-3, 2)$.

[12] critical numbers are algebraic critical numbers of the derivative function, but you're not allowed to know that just yet... oops.

The other way to attack such problems is to tackle the nonlinear inequalities one case at a time until the possibilities are exhausted. For some of you who are gifted in that vein of thought I do not discourage your line of thinking. However, I believe the sign-chart will aid understanding for many. In particular, it helps me sort things out when the expression is less than trivial. Notice that we don't even have to graph the function. The sign chart captures all the data we need for the solution of inequalities.

Example 1.7.3. *Find the domain of* $g(x) = \sqrt{-(x+3)(x-3)^2}$. *Note that we need* $-(x+3)(x-3)^2 \geq 0$. *Define* $f(x) = -(x+3)(x-3)^2$ *and observe* $c = -3, 3$ *are algebraic critical numbers. Observe that* $f(-4) = 1 > 0$, $f(0) = -27 < 0$ *and* $f(4) = -7 < 0$ *hence the sign chart for* f *is:*



We find that $-(x+3)(x-3)^2 \geq 0$ *for* $x \in (-\infty, -3] \cup \{3\}$. *Therefore,* $dom(g) = (-\infty, -3] \cup \{3\}$.

In-Class Example 1.7.4. *Solve the following inequality:*

$$\frac{(x^2+3x)(x^2+4x+5)}{x^2-2x} \leq 0.$$

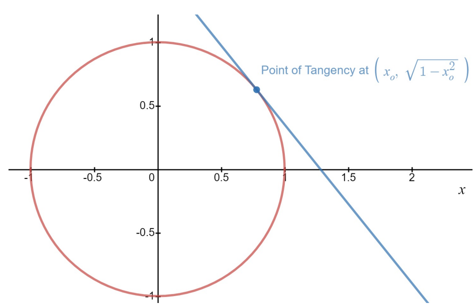
Chapter 2

limits

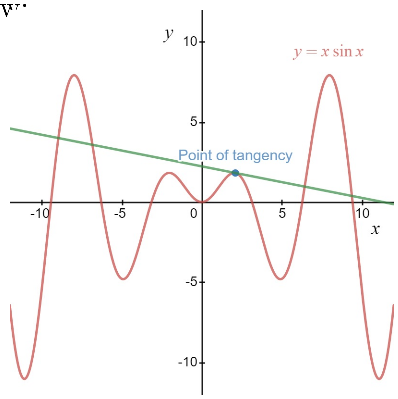
In this chapter we begin by studying the tangent problem and why it necessarily requires the invention of a limiting process. Then, in Section 2.2 we give a careful definition of the limit and we introduce related concepts of left and right limits as well as divergence of limits and finally limits at $\pm\infty$. This toolset of limiting concepts allows us to give careful descriptions of geometric features such as *finite-jump discontinuities*, *hole-in-the-graph*, *vertical asymptote* and finally *horizontal asymptote*. Careful limit proofs from the $\varepsilon\delta$-definition are offered in Subsection 7.5.1. Section 7.6 contains statements and proofs of the *limit laws* as well as the basic limits of elementary functions such as polynomial, rational, algebraic, exponential, hyperbolic and trigonometruc functions. Application of the limit laws to calculate a variety of limits is given in Section 2.3. The Squeeze Theorem is proved and illustrated in Section 7.7. Continuity is defined and the continuity of elementary functions is detailed in Section 2.5. Finally, in Section 2.6 we prove the Intermediate Value Theorem (IVT) and use it to show that the inverse of a continuous function is likewise continuous.

2.1 the tangent problem: why we need limits

The tangent line to a given point on a circle is a line which intersects the circle at the point of tangency and nowhere else. Geometrically is intuitively clear that such a tangent line must be perpendicular to the diameter of the circle which also goes through the point of tangency. For example, the unit circle given by $x^2 + y^2 = 1$ has point $(x_o, \sqrt{1 - x_o^2})$ with tangent line as pictured below, you can move the line around for fun at Desmos demonstration page:



Properties of tangents to circles have been studied since the time of the ancients. For example, there are numerous results in Euclid's *Elements* which are centered around the study of these tangents. Naturally we would like to extend the concept of a tangent line to other curves. For instance, the **graph** of a function $y = f(x)$. Intuitively we want the tangent line to touch the curve at the point of tangency in such a way that it resembles the given graph as closely as possible. In other words, the tangent line should be the *best linear approximation* of the graph near the point of tangency. For example, see the picture below:



you can click here to try moving the tangent line around. Let's pause to review equations of lines before we go further into our discussion of tangent lines.

In-Class Example 2.1.1. *Write the equation of a line containing the point* $(-3, 8)$ *with slope* 3

In-Class Example 2.1.2. *Find the equation of the line from* $(a, f(a))$ *to* $(b, f(b))$ *where* $a \neq b$.

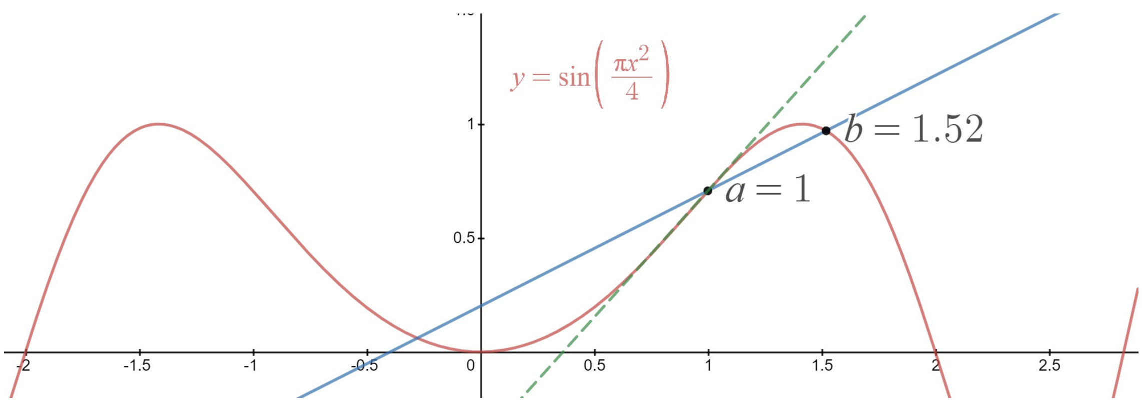
How should we define the tangent line to $y = f(x)$ at $x = a$ carefully ? If we denote the function whose graph is the tangent line by L_f^a then since $(a, f(a))$ is the point of tangency it follows from the point-slope equation of a line that:

$$y = L_f^a(x) = f(a) + m(x - a)$$

for appropriate slope m. How do we select the slope m for the tangent line ? One natural approach is to think of the tangent line has having the slope of secant lines which are very close to the point of tangency. A secant line from $x = a$ to $x = b$ on the graph $y = f(x)$ is just the line which connects the dots $(a, f(a))$ and $(b, f(b))$. The equation of the secant line is:

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \tag{2.1}$$

To be clear, we **cannot** have $a = b$ in the equation above as that causes division by zero. **Yet**, that is precisely where we want to be. When $a \approx b$ the secant line *approaches* the tangent line. I have graphed the tangent line as a green-dotted line and graphed the secant line based at $a = 1$ in the picture below:



I would encourage you to play with this Desmos demonstration page to appreciate how the secant lines get closer and closer to the tangent line as we take values of b which are closer and closer to a. In particular, intuitively we see the slope m for the tangent line should be obtained by calculating:

$$m = \lim_{b \to a} \frac{f(b) - f(a)}{b - a} \tag{2.2}$$

where this **limiting process** which we denote by $\lim_{b \to a}$ should be understood in terms of taking values for b which are arbitrarily close to a. However, this limiting process cannot actually allow $a = b$ since that would cause division by zero for our intended application of the limit. In the next section we'll set down a careful definition which does not require vague appeals to intuition. I hope this section serves to convince you we need the limit to find the slope of the tangent line. Algebra alone does not have the technology to play the game we need to play; we cannot just plug in $b = a$.

2.2 definition of limit

Before we can define the limit we must define the sort of points where we can reasonably take limits for functions on $\mathbb{R}$.

Definition 2.2.1. *limit point*

> Let $U \subseteq \mathbb{R}$. We say p is a **limit point** of U if for every $\delta > 0$ we have $B_\delta(p)_o \cap U \neq \emptyset$. We say p is a limit point of a function f on $\mathbb{R}$ if and only if p is a limit point of $dom(f)$. If there exists $\eta > 0$ for which $B_\eta(p)_o \subseteq dom(f)$ then p is an **interior limit point** of f.

Recall $B_\delta(p)_o = (p-\delta, p) \cup (p, p+\delta)$ thus the definition above simply reduces to the condition that for p to be a limit point of U there must be at least one other point in U near p no matter how small we make δ. In fact, we see a limit point p for a function f is a point at which there are infinitely many points in $dom(f)$ near p. Not every function has a limit point:

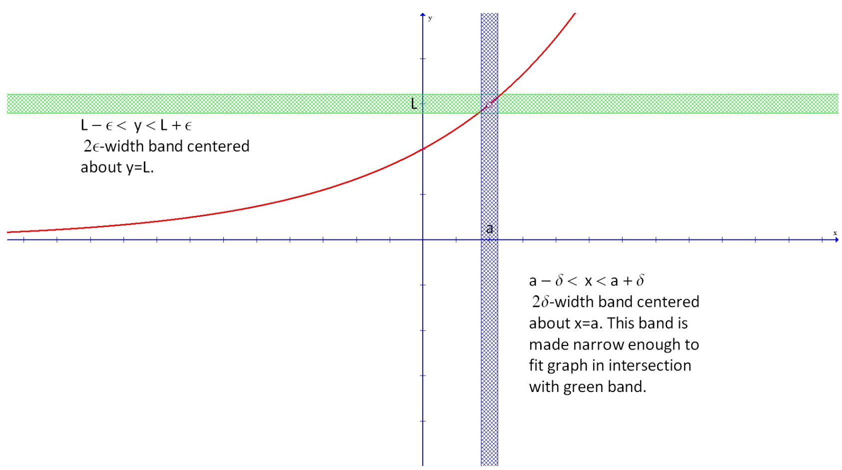
Example 2.2.2. *Sequences on $\mathbb{R}$ are functions $a : \mathbb{N} \to \mathbb{R}$ where we usually denote $a(n) = a_n$. In this case, no point in $\mathbb{N}$ is a limit point since $(n - 1/2, n + 1/2) \cap \mathbb{N} = \{n\}$ hence $B_{1/2}(n)_o \cap \mathbb{N} = \emptyset$. Every point in the domain of a sequence is called an* **isolated point***.*

Sequences are important, we will see them again when we study the area problem later in this course. For now though, we will focus on functions where there are more than just isolated points:

Definition 2.2.3. *limit or double-sided limit*

> Let f be a function with interior limit point a and suppose $L \in \mathbb{R}$. We say that $f(x) \to L$ as $x \to a$ if and only if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$ it follows $|f(x) - L| < \varepsilon$. In the case that the condition above is met we say that the limit exists and denote this by $\lim\limits_{x \to a} f(x) = L$.

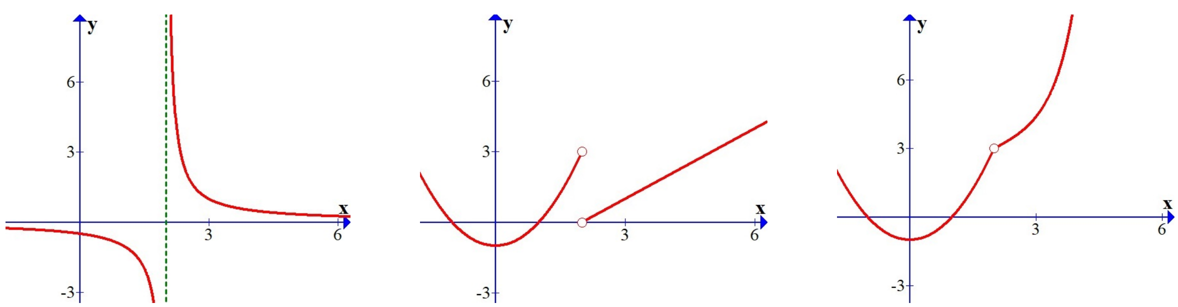
Basically the idea is just that if we zoom in on an ϵ-band centered about L then the limit exists if we can find a δ-band centered about a such that the box made from the intersection of these bands captures the graph of the function for all the values in $(a - \delta, a) \cup (a, a + \delta)$



In-Class Example 2.2.4. *Prove* $\lim_{x \to a} (x) = a$.

In-Class Example 2.2.5. *Prove* $\lim_{x \to a} (x^2) = a^2$.

Example 2.2.6. *In each case below,* $a = 2$ *is an interior limit point of the function graphed.*



However, only the rightmost has a double-sided limit which exists; $\lim_{x \to 2} f(x) = 2$.

It is useful to have langauge to distinguish between the behavior of the graphs above. The concept of left and right limits helps us in refining such language. Using the terminology introduced by the definition below we see the middle graph above has $\lim_{x \to 2^-} f(x) = 3$ and $\lim_{x \to 2^+} f(x) = 0$.

Definition 2.2.7. *one-sided limits*

If f is a function with limit point a.

1.) Assume there exists $\eta > 0$ for which $(a-\eta, a) \cap dom(f) \neq \emptyset$. If for each $\varepsilon > 0$ there exists $\delta > 0$ for which $x \in \mathbb{R}$ with $a < x < a + \delta$ implies $|f(x) - L| < \varepsilon$ then we write $\lim_{x \to a^+} f(x) = L$ and call this the **right-limit** at $x = a$ of f.

2.) Assume there exists $\eta > 0$ for which $(a, a+\eta) \cap dom(f) \neq \emptyset$. If for each $\varepsilon > 0$ there exists $\delta > 0$ for which $x \in \mathbb{R}$ with $a - \delta < x < a$ implies $|f(x) - L| < \varepsilon$ then we write $\lim_{x \to a^-} f(x) = L$ and call this the **left-limit** at $x = a$ of f.

If $(a-\eta, a) \cap dom(f)$ and $(a, a+\eta) \cap dom(f)$ then $B_\eta(a)_o \subseteq dom(f)$ and a is an interior limit point. Interior limit points allow for the calculation of left, right and double-sided limits. In contrast, if the domain of f was $[0,1)$ then I would not consider $\lim_{x \to 0^-} f(x)$ or $\lim_{x \to 1^+} f(x)$ to be a well-defined limits. However, $\lim_{x \to 0^+} f(x)$ or $\lim_{x \to 1^-} f(x)$ would be well-defined limits (which may or may not exist). Quibbles aside, there is a simple relation between left, right and double-sided limits:

Theorem 2.2.8. *two-sided limit holds if and only if both left and right limits hold.*

Let f be a function with interior limit point a. Let $L \in \mathbb{R}$,

$$\lim_{x \to a} f(x) = L \qquad \Leftrightarrow \qquad \left\{ \lim_{x \to a^+} f(x) = L \quad \text{and} \quad \lim_{x \to a^-} f(x) = L \right\}$$

Proof: see Appendix section on limit proofs. $\square$.

In-Class Example 2.2.9.

2.2.1 divergent limits and limits at $\pm\infty$

The leftmost graph in Example 2.2.6 had a vertical asymptote at the limit point $a = 2$. We should provide a careful definition to cover such cases.

Definition 2.2.10. *limits which diverge to* ∞

Let f be a function and $a \in \mathbb{R}$.

- We say that $f(x) \to \infty$ as $x \to a$ if and only if for each $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ whenever $0 < |x - a| < \delta$. In the case that the condition above is met we say that the limit **diverges to** ∞ and denote this by $\lim_{x \to a} f(x) = \infty$.
- If for each $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ whenever $a < x < a + \delta$ then we say $f(x) \to \infty$ as $x \to a^+$ and write $\lim_{x \to a^+} f(x) = \infty$
- Likewise, if for each $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ whenever $a - \delta < x < a$ then we say $f(x) \to \infty$ as $x \to a^-$ and write $\lim_{x \to a^-} f(x) = \infty$.

The definitions of $f(x) \to -\infty$ as $x \to a$ or $x \to a^{\pm}$ are very similar we just replace the condition $f(x) > M$ with $f(x) < N$ for $N < 0$. It is also interesting that the proposition given in the last section also applies in this context:

Proposition 2.2.11. *two-sided limit diverges to $\pm\infty$ iff both left and right limits diverge to $\pm\infty$.*

$$\lim_{x \to a} f(x) = \pm\infty \qquad \Leftrightarrow \qquad \left\{ \lim_{x \to a^+} f(x) = \pm\infty \quad \text{and} \quad \lim_{x \to a^-} f(x) = \pm\infty \right\}$$

In-Class Example 2.2.12.

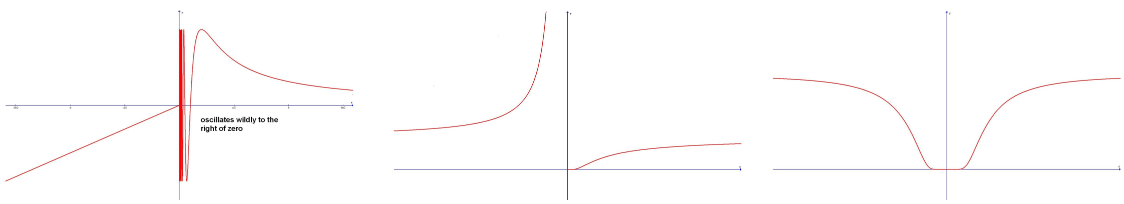
One satisfying aspect of carefully defining divergent limits is that we can give a concrete definition of a vertical asymptote. In fact, we should pause and note that we now have a non-graphical method of distinguishing between vertical asymptotes, holes in the graph and jump-discontinuities of a function. All three can arise from formulas which fail if evaluated at the point in question. The concept of a limit helps us to carefully distinguish what algebra alone cannot hope to detect.

Definition 2.2.13. *vertical asymptotes (VA), holes and jumps.*

Let f be a function and $a \in \mathbb{R}$.

1. We say that f has a **vertical asymptote** $x = a$ if and only if either of the left or right limits diverge to $\pm\infty$. That is, $x = a$ is a VA if and only if $\lim_{x \to a^{\pm}} f(x) = \pm\infty$.
2. We say that f has a **hole in the graph** at (a, L) iff $a \notin dom(f)$ and $\lim_{x \to a} f(x) = L$
3. We say that f has a **finite jump-discontinuity** at $x = a$ if and only if both the left and right limits of $f(x)$ exist in $\mathbb{R}$ and do not agree; $\lim_{x \to a^+} f(x) = L_+ \in \mathbb{R}$ and $\lim_{x \to a^-} f(x) = L_-$ and $L_+ \neq L_-$.

Example 2.2.14. *For the following three graphs only the rightmost graph has a double-sided limit which exists as $x \to 0$; in fact, $f(x) \to 0$ and $x \to 0$ for the rightmost graph.*



The left graph has $f(x) \to 0$ as $x \to 0^-$ whereas the right limit fails to exist due to oscillation as $x \to 0^+$. The middle graph has a vertical asymptote at $x = 0$ since $f(x) \to \infty$ as $x \to 0^-$.

In-Class Example 2.2.15.

2.2.2 limits at $\pm\infty$, a brief look

The behavior a function for $x >> 0$ or for $x << 0$ is captured by the limit of the function at $\pm\infty$,

Definition 2.2.16. *limits at ∞ or $-\infty$.*

We say $\lim\limits_{x\to\infty} f(x) = L$ if and only if for each $\varepsilon > 0$ there exists $N \in \mathbb{R}$ with $N > 0$ such that if $x > N$ then $|f(x) - L| < \varepsilon$. Likewise, $\lim\limits_{x\to-\infty} f(x) = L$ if and only if for each $\varepsilon > 0$ there exists $M \in \mathbb{R}$ with $M < 0$ such that if $x < M$ then $|f(x) - L| < \varepsilon$.

In-Class Example 2.2.17.

Definition 2.2.18. *horizontal asymptotes.*

If $\lim_{x\to\infty} f(x) = L$ then the function f is said to have a **horizontal asymptote of $y = L$ at ∞**. If $\lim_{x\to-\infty} f(x) = L$ then the function f is said to have a **horizontal asymptote of $y = L$ at $-\infty$**.

Example 2.2.19. *Let $f(x) = \tan^{-1}(x)$. We saw in the preliminaries chapter that the inverse tangent function had horizontal asymptotes of $y = \frac{\pi}{2}$ for $x >> 0$ and $y = -\frac{\pi}{2}$ for $x << 0$. Therefore,*

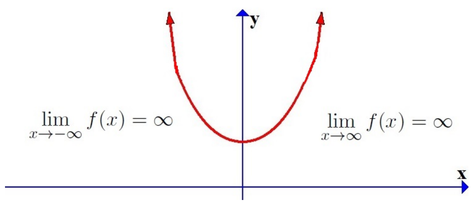
$$\lim_{x\to\infty} \tan^{-1}(x) = \frac{\pi}{2} \qquad \lim_{x\to-\infty} \tan^{-1}(x) = -\frac{\pi}{2}.$$

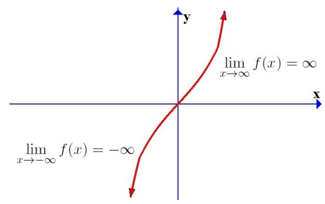
Vertical asymptotes of the function correspond to horizontal asymptotes for the inverse function. We can also discuss limits which go to infinity at infinity. It's just the natural merger of both definitions but I state it here for completeness.

Definition 2.2.20. *infinite limits at infinity.*

The limit at ∞ for a function f is ∞ iff for each $M > 0$ there exists $N > 0$ such that for $x > N$ we find $f(x) > M$. We denote $\lim_{x\to\infty} f(x) = \infty$ in this case. Likewise, the limit at $-\infty$ for a function f is ∞ iff for each $M > 0$ there exists $N < 0$ such that if $x < N$ then $f(x) > M$. We denote this by $\lim_{x\to-\infty} f(x) = \infty$. Similarly, if for each $M < 0$ there exists $N > 0$ such that $x > N$ implies $f(x) < M$ we say $\lim_{x\to-\infty} f(x) = -\infty$. Finally, if for each $M < 0$ there exists $N < 0$ such that $x < N$ implies $f(x) < M$ we say $\lim_{x\to-\infty} f(x) = -\infty$.

Example 2.2.21. *I offer a pair of illustrations to conclude here:*





In-Class Example 2.2.22.

If time permits, we will return to this topic later in the course. See Section 5.4 for many examples and much advice. I don't think we need all that jazz at the moment, so I defer it.

2.3 limit calculation

We assume $a \in \mathbb{R}$ and f, g are functions with limit point a throughout this section unless otherwise explicitly stated. I've written careful proofs for the claims in this section in Appendix 7.6, most of those proofs are omitted here.

Theorem 2.3.1. *Limit Laws*

(1.) If $\lim_{x \to a} f(x) = L_1$ and $\lim_{x \to a} f(x) = L_1$ then $L_1 = L_2$.

(2.) If $\lim_{x \to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \to a} g(x) = L_g \in \mathbb{R}$ then $\lim_{x \to a} [f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$.

(3.) Suppose $c \in \mathbb{R}$ and $\lim_{x \to a} f(x) = L \in \mathbb{R}$ then $\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x)$.

(4.) Suppose $a \in \mathbb{R}$ and $f_i(x) \to L_i \in \mathbb{R}$ as $x \to a$ for $i = 1, 2, \dots, n$. Then,

$$\lim_{x \to a} \left(c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x)\right) = c_1 \lim_{x \to a} f_1(x) + c_2 \lim_{x \to a} f_2(x) + \cdots + c_n \lim_{x \to a} f_n(x).$$

(5.) If $\lim_{x \to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \to a} g(x) = L_g \in \mathbb{R}$ then $\lim_{x \to a} [f(x)g(x)] = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} g(x)\right)$.

(6.) If $\lim_{x \to a} f(x) = L_1$ and $\lim_{y \to L_1} g(y) = L_2$ then $\lim_{x \to a} g(f(x)) = L_2$.

(7.) If $\lim_{x \to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \to a} g(x) = L_g \in \mathbb{R}$ with $L_g \neq 0$ then $\lim_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$.

(8.) If either one of the following limits exist then so does the other and $\lim_{x \to a} f(x) = \lim_{h \to 0} f(a + h)$.

Proof: (I will probably prove (2.))

Theorem 2.3.2. *Limits of Elementary Functions*

(1.) $\lim\limits_{x\to a} x = a$.

(2.) $\lim\limits_{x\to a} c = c$.

(3.) Let $a \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$, $\lim\limits_{x\to a} x^n = a^n$.

(4.) Suppose $c_n, \dots, c_1, c_0 \in \mathbb{R}$ and $p(x) = c_n x^n + \cdots + c_1 x + c_0$ then $\lim\limits_{x\to a} p(x) = p(a)$.

(5.) If $a \neq 0$ then $\lim_{x\to a} \frac{1}{x} = \frac{1}{a}$

(6.) If $a > 0$ then $\lim_{x\to a} \sqrt{x} = \sqrt{a}$. In addition, $\lim_{x\to 0^+} \sqrt{x} = 0$.

(7.) Let $a \in \mathbb{R}$ with $a > 0$ and $n \in \mathbb{N}$, $\lim_{x\to a} x^{\frac{m}{n}} = a^{\frac{m}{n}}$.

(8.) Let $f(x)$ be defined by a finite number of algebraic operations (possibly including addition, multiplication, division, taking integer or fractional roots) then $\lim\limits_{x\to a} f(x) = f(a)$

(9.) Let $a \in \mathbb{R}$, $\lim_{x\to a} \sin(x) = \sin(a)$ and $\lim_{x\to a} \cos(x) = \cos(a)$.

(10.) Let $b > 0$, $\lim_{x\to a} b^x = b^a$.

Example 2.3.3. *In each of the limits below the limit point is on the interior of the domain of the elementary function so we can just evaluate to calculate the limit.*

$$i.)\ \lim_{x\to 3}(\sin(x)) = \sin(3)$$

$$ii.)\ \lim_{x\to -2}\left(\frac{\sqrt{x^2-3}}{x+5}\right) = \frac{\sqrt{4-3}}{-2+5} = \frac{1}{3}$$

$$iii.)\ \lim_{h\to 0}(\sin^{-1}(h)) = \sin^{-1}(0) = 0$$

$$iv.)\ \lim_{x\to a}(x^3 + 3x^2 - x + 3) = a^3 + 3a^2 - a + 3.$$

We did not even need to look at a graph to calculate these limits. Of course it is also possible to evaluate most limits via a graph or a table of values, but those methods are less reliable.

The example below illustrates the table of values idea.

Example 2.3.4. *The following table of values indicates that* $\lim_{x\to 0^+} \frac{\sin(x)}{x} = 1$

x	sin(x)/x
0.5	0.958851
0.2	0.993347
0.1	0.998334
0.01	0.999983
0.001	0.999999

Now the limit considered in Example 2.3.4 is not nearly as obvious as the limits in Example 2.3.3. I should mention that the limit has indeterminant form of type $0/0$ since both $\sin(x)$ and x tend to zero as x goes to zero. One of main goals in this part of the course is to learn how to analyze indeterminant forms. Thus far we have only encountered case (1.) of the definition below. The reason these are called "indeterminant forms" is simply that the value of the limit with an indeterminant form is not known without further analysis. Limits with these forms might diverge to infinity, simply not exist or even converge to any number of finite values.

Definition 2.3.5. *indeterminant forms.*

- we say $\lim \frac{f}{g}$ is of **"type $\frac{0}{0}$"** iff $\lim f = 0$ and $\lim g = 0$
- we say $\lim \frac{f}{g}$ is of **"type $\frac{\infty}{\infty}$"** iff $\lim f = \pm\infty$ and $\lim g = \pm\infty$
- we say $\lim fg$ is of **"type 0∞"** iff $\lim f = 0$ and $\lim g = \pm\infty$
- we say $\lim f - g$ is of **"type $\infty - \infty$"** iff $\lim f = \infty$ and $\lim g = \infty$

Now it is time for us to test our algebraic might. The examples given in this section illustrate all the basic algebra tricks to unravel undetermined limits. I like to say we do algebra to determine the limit. The limits are not just decoration, many times an expression with the limit is correct while the same expression without the limit is incorrect. On the other hand we should not write the limit if we do not need it in the end. How do we know when and when not? We practice.

In-Class Example 2.3.6.

In-Class Example 2.3.7.

In-Class Example 2.3.8.

In-Class Example 2.3.9.

In-Class Example 2.3.10.

In-Class Example 2.3.11.

In-Class Example 2.3.12.

Example 2.3.13. *Piecewise defined functions can require a bit more care.* $\lim_{x\to 0}\left(\dfrac{|x|}{x}\right) = ?$. *Recall* $|x| = \begin{cases} -x & : x < 0. \\ x & : x \geq 0. \end{cases}$. *In the left limit* $x \to 0^-$ *we have* $x < 0$ *so* $|x| = -x$ *thus,*

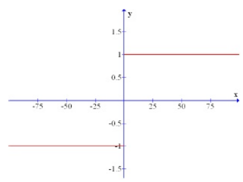
$$\lim_{x\to 0^-}\left[\frac{|x|}{x}\right] = \lim_{x\to 0^-}\left[\frac{-x}{x}\right] = \lim_{x\to 0^-}\left[\frac{-1}{1}\right] = -1.$$

In the right limit $x \to 0^+$ *we have* $x > 0$ *so* $|x| = x$ *thus,*

$$\lim_{x\to 0^-}\left[\frac{|x|}{x}\right] = \lim_{x\to 0^-}\left[\frac{x}{x}\right] = \lim_{x\to 0^-}\left[\frac{1}{1}\right] = 1.$$

Consequently we find that the left and right limits disagree hence $\lim_{x\to 0}\left[\frac{|x|}{x}\right] = d.n.e.$.

The function we just looked at in preceding is a *step function*. They are very important to engineering since they model switching. The graph $y = |x|/x$ looks like a single stair step,



Example 2.3.14. *This limit below is not indeterminant, the type* $\infty/0$ *will diverge. The question is merely how does it diverge? It becomes clear this limit is positive after we simplify,*

$$\lim_{x\to 0}\left(\frac{\cot(x)}{\tan(x)}\right) = \lim_{x\to 0}\left(\frac{1}{\cot^2(x)}\right) = \infty.$$

Example 2.3.15. *This limit below is not indeterminant, the type* $\infty/0$ *will diverge. The question is merely how does it diverge?*

$$\lim_{x\to 0^-}\left(\frac{e^x + 3}{\sin(x)}\right) = -\infty.$$

I knew it diverged to $-\infty$ *since the values of the function are negative for inputs just a little to the left of zero and I know the sine function has negative values for such inputs.*

Remark 2.3.16. *Intuition is very important. One of the main reasons to do a lot of homework is that it refines and sharpens your intuition. Whenever a person with experience is faced with a limit problem the usual first step we make is to decide what we think the answer ought to be. Then we supply algebra to confirm our suspicion. If the function is complicated I often plug in points really close to the limit point to get a feel for the problem. This approach will fail for a certain class* *of sarcastically crafted pathological problems but it is successful for almost all problems assigned in this introductory course. My point? You can figure out what the answer is often even when you can't show your work. This will earn you some partial credit, but the idea here is not just to find an answer. The steps showing how the answer is deduced are important. At a minimum you ought to show how indeterminancy is removed for a given problem. I did that in every example in this section.*

2.4 squeeze theorem

There are limits not easily solved through algebraic trickery. Sometimes the "Squeeze" or "Sandwich" Theorem allows us to calculate the limit. The proof is given in Appendix 7.7

Proposition 2.4.1. *squeeze theorem*[1].

Let $f(x) \leq g(x) \leq h(x)$ for all x near a then we find that the limits at a follow the same ordering,

$$\lim_{x \to a} f(x) \leq \lim_{x \to a} g(x) \leq \lim_{x \to a} h(x).$$

Moreover, if $\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L \in \mathbb{R}$ then $\lim_{x \to a} f(x) = L$.

We can think of $h(x)$ as the top slice of the sandwich and $f(x)$ as the bottom slice. The function $g(x)$ provides the BBQ or peanut butter or whatever you want to put in there.

Example 2.4.2. *Use the squeeze theorem to calculate* $\lim_{x \to 0}(\ x^2 \sin(\frac{1}{x}) \ )$. *Notice that the following inequality is suggested by the definition or graph of sine*

$$-1 \leq \sin(\theta) \leq 1$$

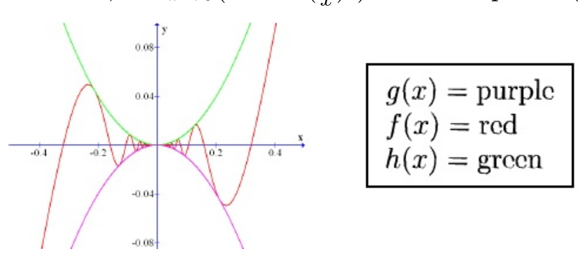
Substitute $\theta = 1/x$ *and multiply by* x^2 *which is positive if* $x \neq 0$ *so the inequality is maintained,*

$$-x^2 \leq x^2 \sin\left(\frac{1}{x} \right) \leq x^2$$

We identify that $f(x) = -x^2$ *and* $h(x) = x^2$ *sandwich the function* $g(x) = x^2 \sin(\frac{1}{x})$ *near* $x = 0$. *Moreover, it is clear that*

$$\lim_{x \to 0}(\ x^2 \ ) = 0 \qquad \lim_{x \to 0}(\ -x^2 \ ) = 0.$$

Therefore, by the squeeze theorem, $\lim_{x \to 0}(\ x^2 \sin(\frac{1}{x}) \ ) = 0$. *Graphically we can see why this works,*



[1]this theorem can also be stated for $x \to \pm\infty$ or $x \to a^{\pm}$ provided the inequality holds for appropriate values where the limit is taken

Perhaps, you're wondering why we could not just use the limit of product proposition $\lim fg = \lim f \lim g$. The problem is that since the limit of $\sin\left(\frac{1}{x}\right)$ at zero does not exist due to wild oscillation at zero. Therefore, we have no right to apply the limit proposition.

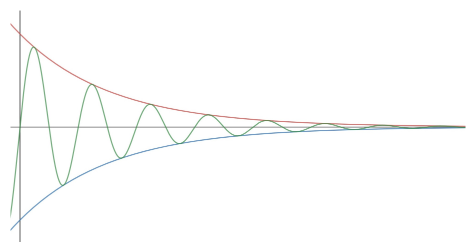
Example 2.4.3. *Suppose that all we know about the function $f(x)$ is that it is sandwiched by $1 \leq f(x) \leq x^2 + 2x + 2$ for all x. Can we calculate the limit of $f(x)$ as $x \to -1$? Well, notice that*

$$\lim_{x \to -1}(1) = 1 \qquad \lim_{x \to -1}(x^2 + 2x + 2) = 1.$$

Therefore, by the Squeeze Theorem, $\lim_{x \to -1} f(x) = 1$.

The Squeeze Theorem applies to other types of limits with appropriate modification.

Example 2.4.4. *Observe $-1 \leq \sin(10t) \leq 1$ implies $-e^{-t} \leq e^{-t}\sin(10t) \leq e^{-t}$. Note $\pm e^{-t} \to 0$ as $t \to \infty$ thus the Squeeze Theorem implies $\lim\limits_{t \to \infty} e^{-t}\sin(10t) = 0$. The graph below has $y = \pm e^{-t}$ as the red and blue curves which envelop the graph of $y = e^{-t}\sin(10t)$*



In-Class Example 2.4.5.

2.5 continuity of functions

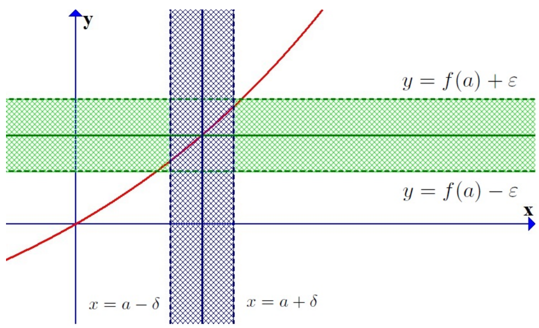
We've seen that in many circumstances $\lim_{x\to a} f(x) = f(a)$. A function which satisfies such a condition at each point in its domain is called **continuous**. It is convenient to give a definition here in terms of inequalities since it simultaneous includes endpoints[2]

Definition 2.5.1. *continuity.*

Let f be a function and $a \in dom(f)$ then f is **continuous at** a if and only if for each $\varepsilon > 0$ there exists $\delta > 0$ for which $|x - a| < \delta$ implies $|f(x) - f(a)| < \varepsilon$. If f is continuous for each $x \in U$ then f is **continuous on** U. If f is continuous on its domain then f is **continuous**.

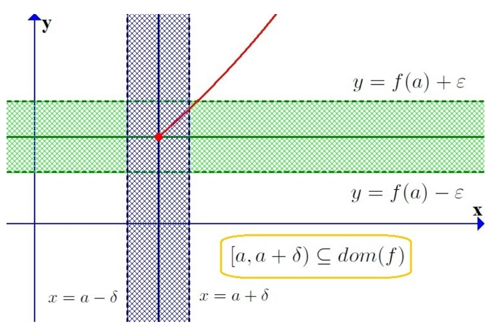
Let us picture the different cases which the above definition captures.

1. If there exists $\eta > 0$ for which $B_\eta(a) \subseteq dom(f)$ then continuity of f at a implies for each $\pm\varepsilon$-band centered about $y = f(a)$ we can select the blue $\pm\delta$-band centered at $x = a$ for which the outputs of f fit within the pictured green band:



We should recognize that $\lim\limits_{x\to a} f(x) = f(a)$.

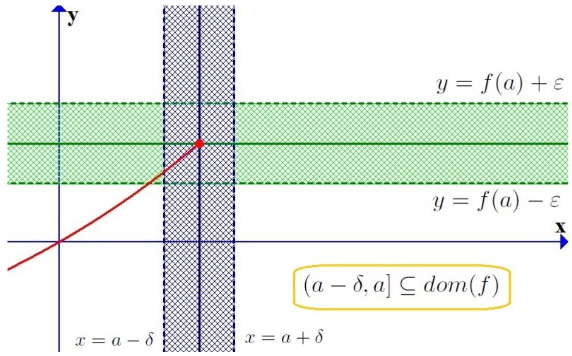
2. If $a \in dom(f)$ is a **left boundary point** of $dom(f)$ the continuity of f at $x = a$ indicates for each $\pm\varepsilon$-band about $y = f(a)$ there exists a blue $\pm\delta$-band about $x = a$ whose intersection with $dom(f)$ returns values for $f(x)$ which fit within the green band:



In this case we have $\lim\limits_{x\to a^-} f(x) = f(a)$.

[2]this definition also makes sequences continuous functions, a fact which probably does not matter to the current course

3. If $a \in dom(f)$ is a **right boundary point** of $dom(f)$ the continuity of f at $x = a$ indicates for each $\pm\varepsilon$-band about $y = f(a)$ there exists a blue $\pm\delta$-band about $x = a$ whose intersection with $dom(f)$ returns values for $f(x)$ which fit within the green band:



In this case we have $\lim_{x \to a^+} f(x) = f(a)$.

In summary, a function is continuous at $x = a$ if and only if the limit of the function as x approaches a within $dom(f)$ is given by evaluating f at $x = a$.

Theorem 2.5.2. *Characterizing continuity via limits: f is continuous at $a \in dom(f)$ provided*

1. $a \in int(dom(f))$ *and* $\lim_{x \to a} f(x) = f(a)$.

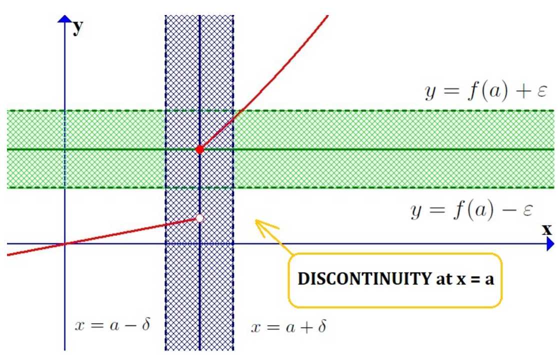
2. *a is a left boundary point and* $\lim_{x \to a^+} f(x) = f(a)$.

3. *a is a right boundary point and* $\lim_{x \to a^-} f(x) = f(a)$.

Proof: I leave the proof to the reader, it is not especially difficult. □

Another useful visualization is given below.

4. If we find a single $\pm\varepsilon$-band centered about $y = f(a)$ fow which it is impossible to contain the values $y = f(x)$ for $x \in (a - \delta, a + \delta)$ then this shows that $\lim_{x \to a} f(x)$ does not exist. For example:



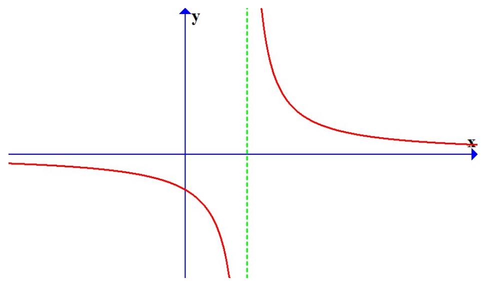
Discontinuity can also arise from just a single point moving off the graph. The key idea is that continuous functions have graphs where the values adhere to one another locally.

I should caution the reader who remembers the definition of continuity from their previous course work. Often in precalculus the description of a continuous function is give by the following slogan:

A continuous function is one whose graph is drawn without lifting your pen.

Unfortunately, this slogan does not work unless the domain of the function is **connected**. If the domain is not connected then graph of the function is likewise disconnected.

For example, consider $f(x) = \frac{1}{x}$ the graph $y = \frac{1}{x}$ cannot be drawn unless we lift our pen at the vertical asymptote $x = 0$:



Notice that $dom(f) = (-\infty, 0) \cup (0, \infty)$ and for $a \neq 0$ we have $\lim_{x \to a} \frac{1}{x} = \frac{1}{a}$. Thus f is continuous at each point in its domain. Hence f is a continuous function.

In-Class Example 2.5.3. Problem: *choose a value for c which makes f continuous on* $\mathbb{R}$ *given that*

$$f(x) = \begin{cases} cx + 3. & x < -2 \\ x^2 + 3, & x \geq -2 \end{cases}.$$

We now return to Section 7.6 and apply the limit laws to our study of continuous functions. To begin we catalog how we can construct new continuous functions from old:

Theorem 2.5.4. *Let f and g be continuous at a and $c \in \mathbb{R}$,*

1.) *$f + g$ is continuous at a,*

2.) *cf is continuous at a,*

3.) *fg is continuous at a,*

4.) *given $g(a) \neq 0$, $\frac{f}{g}$ is continuous at a.*

Moreover, if f and g are continuous functions then $f + g, cf$ and fg are likewise continuous.

Proof: from Proposition 7.6.4 and continuity of f and g at a we find:

$$\lim_{x \to a}(f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x) = f(a) + g(a) \tag{2.3}$$

Therefore, $f + g$ is continuous a for each $a \in dom(f) \cap dom(g)$. We find $f + g$ is continuous on $dom(f) \cap dom(g)$. Likewise, if $c \in \mathbb{R}$ then Proposition 7.6.5 and continuity of f at a yields

$$\lim_{x \to a}(cf(x)) = c \lim_{x \to a} f(x) = cf(a). \tag{2.4}$$

Therefore, cf is continuous at a and it follows cf is continuous on $dom(f)$ since the argument above holds for all $a \in dom(f)$. Finally, Proposition 7.6.7 and continuity of f and g at a provides $f(x)g(x) \to f(a)g(a)$ for each $a \in dom(f) \cap dom(g)$. Thus fg is a continuous function. Finally, apply Proposition 7.6.12 to prove (4.). $\square$

We also have the composite of continuous functions is continuous:

Theorem 2.5.5. *If f is continuous at a and g is continuous at $f(a)$ then $g \circ f$ is continuous at a.*

Proof: If f is continuous at a and g is continuous at $f(a)$ then Proposition 7.6.10 provides $\lim_{x \to a} g(f(x)) = \lim_{y \to f(a)} g(y) = g(f(a))$. Therefore, $g \circ f$ is continuous at a. $\square$

Continuity allows us to pull limits in and out of function evaluation: if $f(x) \to L_f \in \mathbb{R}$ as $x \to a$ and g is continuous at L_f then:

$$\lim_{x \to a} g(f(x)) = g\left(\lim_{x \to a} f(x)\right). \tag{2.5}$$

Elementary functions include polynomials, sine, cosine, algebraic functions, exponentials, hyperbolic functions as well as the inverse functions of all of these. See Section 1.5 for graphs and discussion of some of the quirks and features of these functions[3]. Elementary functions are continuous because we the limit laws allow us to calculate the limit of an elementary function by evaluation. The points at which the limit of an elementary function does not exist are points outside the domain of the function.

[3]I blame Doug Demuro's You Tube Channel for this quirk and feature of my writing.

In-Class Example 2.5.6.

In-Class Example 2.5.7.

Let us record the continuity of elementary functions in a theorem for future reference:

Theorem 2.5.8. *Each function below is continuous on its domain:*

1.) *Polynomial functions,*

2.) *Rational functions,*

3.) *Algebraic functions,*

4.) *Trigonmetric functions and their reciprocal functions.*

5.) *Exponential functions,*

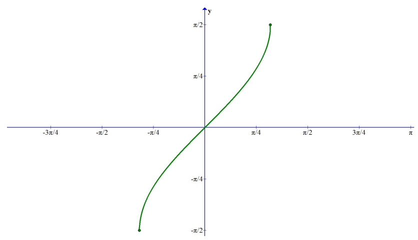
6.) *Hyperbolic trigonmetric functions and their reciprocal functions.*

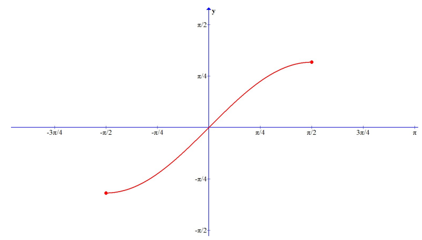
The proof is given in the Appendix at 7.8.1. Essentially all these claims follow from the limit laws. It turns out the inverse functions require greater sophistication as proved in the Appendix at 7.10.2.

Theorem 2.5.9. *invertible continuous function have continuous inverses.*

Suppose $S, T \subseteq \mathbb{R}$ and S is connected. If $f : S \to T$ is continuous with inverse $f^{-1} : T \to S$ then f^{-1} is continuous.

Remark 2.5.10. *If f is a one-to-one function on S then $f(x) = y$ can be used to define $f^{-1}(y) = x$. Moreover, $(a, b) \in graph(f)$ if and only if $(b, a) \in graph(f^{-1})$. This means we can draw the graph of the inverse function by drawing the same pattern as the function graph by simply exchanging the vertical and horizontal directions in the graphing in the xy-plane. For example, here are the graphs of sine and its inverse:*





If f is continuous then there is no jump in the graph, since f^{-1} is drawn in the same way (just diagonally flip the paper) it stands to reason that f^{-1} also has no jump in its graph. That is to say, continuity of the inverse function for a continuous function with connected domain is completely unsurprising.

Proposition 2.5.11. *continuity of power function for arbitrary power.*

Let $p \in \mathbb{R}$ and $a > 0$ then $\lim_{x \to a} x^p = a^p$.

Proof:

Given the inverse function theorem and the results already given in this section it should be clear that all the functions from Section 1.5 are continuous.

Theorem 2.5.12. *most elementary functions are continuous on the interior of their domain.*

Polynomial, rational, power, trigonometric, hyperbolic as well as their respective local inverse functions are continuous on the interior of their respective domains.

Proof: Theorem 7.8.1 provides continuity of the polynomial, rational, algebraic, trigonometric, exponential and hyperbolic functions. Each inverse function is given a connected domain hence Theorem 7.10.2 applies to provide the desired continuity for the inverse function. $\square$

In-Class Example 2.5.13.

In-Class Example 2.5.14.

In-Class Example 2.5.15.

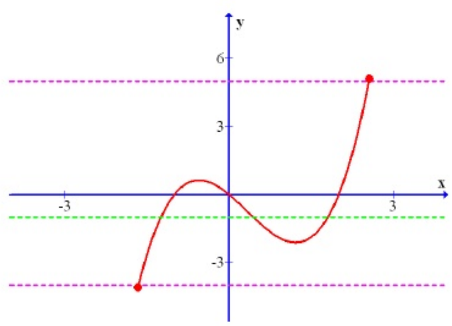
2.6 intermediate value theorem

The proof of the intermediate value theorem is given at the conclusion of this section.

Theorem 2.6.1. *intermediate value theorem (IVT).*

> Suppose that f is continuous on an interval $[a, b]$ with $f(a) \neq f(b)$ and let N be a number such that N is between $f(a)$ and $f(b)$ then there exists $c \in (a, b)$ such that $f(c) = N$.

Notice that this theorem only tells us that there exists a number c, it does not actually tell us how to find that number. This theorem is quite believable if you think about it graphically. Essentially it says that if you draw a horizontal line $y = N$ between the lines $y = f(a)$ and $y = f(b)$ then since the function is continuous we must cross the line $y = N$ at some point. Remember that the graph of a continuous function has no jumps in it so we cannot possibly avoid the line $y = N$. Let me draw the situation for the case $f(a) < f(b)$,



Green line is $y = N$. Purple lines are $y = f(a)$ and $y = f(b)$. In this example there is more than one point c such that $f(c) = N$. There must be ***at least*** one such point provide that the function is continuous.

The IVT can be used for an indirect manner to locate the zeros of continuous functions. The theorem motivates an iterative process of divide and conquer to find a zero of the function. Essentially the point is this, if a continuous function changes from positive to negative or vice-versa on some interval then it must be zero at least one place on that interval. This observation suggests we should guess where the function is zero and then look for smaller and smaller intervals where the function has a sign change. We can just keep zooming in further and further and getting closer and closer to the zero. Perhaps you have already used the IVT without realizing it when you looked for an intersection point on your graphing calculator.

Example 2.6.2. *Show that there exists a zero of the polynomial* $P(x) = 4x^3 - 6x^2 + 3x - 2$ *on the interval* $[1, 2]$. *Observe that,*

$$P(1) = 4 - 6 + 3 - 2 = -1 < 0$$
$$P(2) = 32 - 24 + 6 - 2 = 12 > 0$$

We know that P *is continuous everywhere and clearly* $P(1) < 0 < P(2)$ *so by the IVT we find there exists some point* $c \in (1, 2)$ *such that* $P(c) = 0$. *To find the precise value of* c *would require more work.*

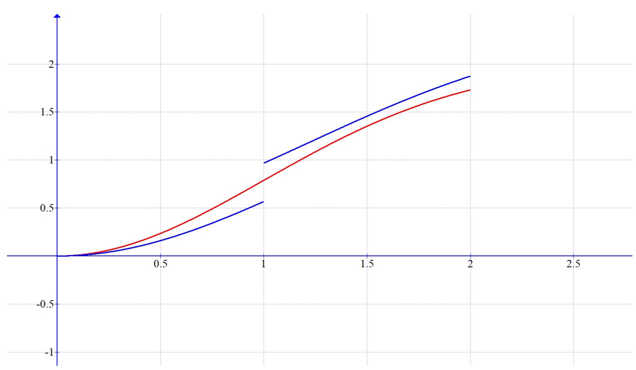
In-Class Example 2.6.3. *Show there exists a solution of* $x^5 + 4x^2 = 100$

Example 2.6.4. *Does* $\tan^{-1}(x) = -\cos(x)$ *for some* $x \in (-2, 2)$ *? Let's rephrase the question. Does* $f(x) = \tan^{-1}(x) + \cos(x) = 0$ *for some* $x \in (-2, 2)$*? This is the same question, but now we can use the IVT plus the sign change idea. Observe,*

$$f(-2) = \tan^{-1}(-2) + \cos(-2) = -1.52$$
$$f(2) = \tan^{-1}(2) + \cos(2) = 0.691.$$

Obviously $f(-2) < 0 < f(2)$ *and both* $\tan^{-1}(x)$ *and* $\cos(x)$ *are continuous everywhere so by the IVT there is some* $c \in (-2, 2)$ *such that* $f(c) = 0$*. Clearly* c *has* $\tan^{-1}(c) = -\cos(c)$*. If you examine the graphs of* $y = \tan^{-1}(x)$ *and* $y = -\cos(x)$ *you will find that they intersect at* $c = -0.82$ *(approximately).*

Example 2.6.5. *Consider two increasing functions which model some physical process as a function of time* t*. Suppose further* $f(t) < g(t)$ *for* $0 \leq t < 1$ *and* $f(t) > g(t)$ *for* $t > 1$*. We might think it must be the case that* $f(1) = g(1)$*. How else can the values of* $f(t)$ *overtake the values of* $g(t)$ *? Well, consider the following:*



If we define $h(t) = f(t) - g(t)$ *then notice* $h(t) < 0$ *for* $t < 1$ *yet* $h(t) > 0$ *for* $t > 1$*. If we carelessly applied the IVT then we would conclude* $h(1) = 0$*. However,* f *is* **discontinuous** *hence* h *is not continuous and we cannot apply the IVT. Indeed, there is no* $t > 0$ *for which* $f(t) = g(t)$*. The graph* $y = f(t)$ *jumps over* $y = g(t)$ *somewhat magically at* $t = 1$*.*

This sort of discontinuity is often due to a switch which is activated by some entity. In contrast, macroscopic motion of large processes tend to follow more gradual patterns with variation which follows a typical bound throughout the process. If I ran an experiment and I observed there was a jump like that in the graph I would assume much investigation was needed as to the mechanics which caused the jump. It makes one suspect the rules governing the rest of the processes were violated at the jump.

Remark 2.6.6. *root finder for continuous functions.*

Let me take a moment to write an algorithm to find roots. Suppose we are given a continuous function f, we wish to find c such that $f(c) = 0$.

1. Guess that f is zero on some interval (a_o, b_o).
2. Calculate $f(a_o)$ and $f(b_o)$ if they have opposite signs go on to 3.) otherwise return to 1.) and guess differently.
3. Pick $c_1 \in (a_o, b_o)$ and calculate $f(c_1)$.
4. If the sign of $f(c_1)$ matches $f(a_o)$ then say $a_1 = c_1$ and let $b_1 = b_o$. If the sign of $f(c_1)$ matches $f(b_o)$ then say $b_1 = c_1$ and let $a_1 = a_o$
5. Pick $c_2 \in (a_1, b_1)$ and calculate $f(c_2)$.
6. If the sign of $f(c_2)$ matches $f(a_1)$ then say $a_2 = c_2$ and let $b_2 = b_1$. If the sign of $f(c_2)$ matches $f(b_1)$ then say $b_2 = c_2$ and let $a_2 = a_1$

and so on... If we ever found $f(c_k) = 0$ then we would stop there. Otherwise, we can repeat this process until the subinterval (a_k, b_k) is so small that we know the zero to some desired accuracy. Say you wanted to know 2 decimals with certainty, if you did the iteration until the length of the interval (a_k, b_k) was 0.001 then you would be more than certain. Of course, a careful analysis of this algorithm and its limitations would also need to consider rounding errors and the inherent limitations of machine arithmetic. Beware the machine ε.

In-Class Example 2.6.7. *Find solution of* $x^5 + 4x^2 = 100$ *on* $[0, 3]$ *by the method of bisection.*

Chapter 3

differential calculus

We will define the derivative of a function in this chapter. The need for a derivative arises naturally within the study of the motion of physical bodies.

You are probably already familiar with the average velocity of a body. For example, if a car travels 100 miles in two hours then it has an average velocity of 50 mph. That same care may not have traveled the same velocity the whole time though, sometimes it might have gone 70mph at the bottom of a hill, or perhaps 0mph at a stoplight. Well, this concept I just employed used the idea of instantaneous velocity. It is the velocity measured with respect to an instant of time.

How small is an "instant"? Well, it's pretty small. You might imagine that this "instant" is some agreed small unit of time. That is not the case, there is no natural standard for all processes. I suppose you could argue with the policeman that your average rate of speed to school was 30mph (taking the "instant" to be 10 minutes for me) but I bet all he'll care about is the 40mph you did through the 20mph school zone. The "velocity" of a car as measured by radar is essentially the instantaneous velocity. It is the time rate change in distance for an arbitrarily small increment of time. It seems intuitive to want such a description of motion, I have a hard time thinking about how we would describe motion without instantaneous velocity. But, then I have (we all have) grown up under the influence of Isaac Newton's ideas about motion. Certainly he was not alone in the development of these ideas, Galileo, Kepler and a host of others also pioneered these concepts which we take for granted these days. Long story short, differential calculus was first motivated by the study of motion. Our goal in this chapter is to give a precise meaning to such nebulous phrases as "instant" of time. The limits of the previous chapter will aid us in this description.

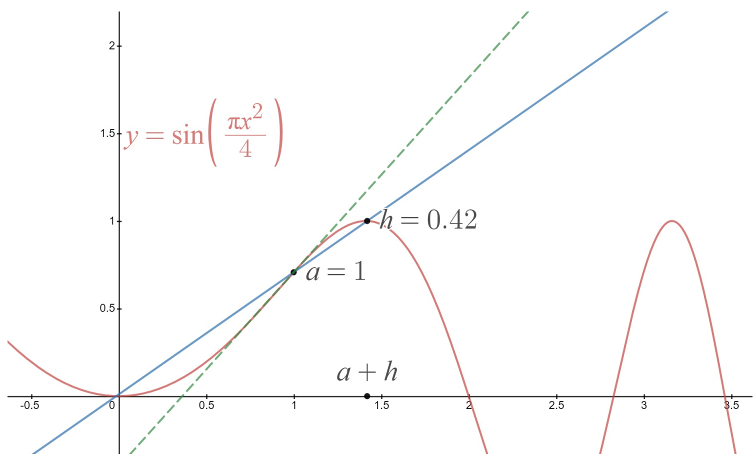
Generally, the derivative of a function describes how the function changes with respect to its independent variable. When the independent variable is time then it is a time-rate of change. But, that need not always be the case. I believe that Newton first thought of things changing with respect to time, he had physics on the brain. In contrast, Leibniz considered more abstract rates of change and the modern approach probably is closer to his work. We typically use Leibniz' notation.

Let me briefly describe the content of this chapter. We begin by defining tangent lines and infinitesimal rates of change. Then the derivative as a function is defined and several examples exhibiting the tangent line construction are given. Next, linearity and the power rule are developed. Breaking from logical minimalism for the sake of pedagogical efficiency we then find derivatives of exponential, sine and cosine functions. Inclusion of that material at that point allows us to integrate those important transcendental function in the later sections of the chapter. Finally, we conclude the chapter by working out the major rules of differential calculus: the product, quotient and chain rules and their beautiful applications in the techniques of logarithmic and implicit differentiation.

Finally, I cannot overstate the importance of this chapter. The derivative forms the core of the calculus sequence. And it describes much more than velocity, that is just one application. Basically, if something changes then a derivative can be used to model it. It's ubiquitous.

3.1 differentiability at a point

Let a be a fixed number throughout this discussion. Let h be an number which we allow to vary. Then a **secant line** at $(a, f(a))$ is simply a line which connects $(a, f(a))$ to another point $(a+h, f(a+h))$ which is also on the graph of the function. For example:



You can imagine that as h increases or decreases we will get a different secant line. In fact, there are infinitely many secant lines. See this Desmos graph to adjust the value of h, or to animate $-1 < h < 1$. Notice that the slope of the pictured secant line is just the rise over the run, that is

$$m = \frac{\triangle y}{\triangle x} = \frac{f(a+h) - f(a)}{a + h - a} = \frac{f(a+h) - f(a)}{h}. \tag{3.1}$$

This may look familiar to you. it is the so-called "*difference quotient*" some of you may have seen in your precalculus course. We should also realize the slope of the secant line gives the **average rate of change of y with respect to x**. In the limit $h \to 0$ we obtain the **tangent line** whose slope can be interpreted as the **instantaneous rate of change of y with respect to x**.

Definition 3.1.1. *slope of function, derivative at point, tangent line.*

If the limit below exists then we say f is **differentiable** at $x = a$ and define

$$f'(a) = \lim_{h \to 0}\left(\frac{f(a+h) - f(a)}{h} \right)$$

For f differentiable at $x = a$, the equation of the tangent line is $\boxed{y = f(a) + f'(a)(x-a)}$ and we say the **slope** of f at a is $f'(a)$.

The tangent line is unique when it exists because limits are unique when they exist. However, there is more than one method to formulate $f'(a)$. Substitute $x = a + h$. As $h \to 0$ note $x = a + h \to a$. Thus, by the substition law for limits,

$$f'(a) = \lim_{x \to a}\left(\frac{f(x) - f(a)}{x - a} \right) \tag{3.2}$$

The formulation of $f'(a)$ above is sometimes useful.

Example 3.1.2. *If we consider y as position and t as time then* **average velocity** *from t to $t + \triangle t$ is given by $\frac{\triangle y}{\triangle t} = \frac{y(t+\triangle t) - y(t)}{\triangle t}$. The* **velocity** *at time t for position y is thus defined,*

$$v(t) = y'(t) = \lim_{\triangle t \to 0}\left(\frac{y(t+\triangle t) - y(t)}{\triangle t} \right). \tag{3.3}$$

No qualifier is placed on $v(t)$ because it is understood from here on out that unless qualified the "velocity" is the "instantaneous velocity". The necessity of this concept led Newton and others interested in the physics of motion to the mathematics of calculus.

We can question the necessity of limits in the formulation of the tangent line. Indeed, when Newton and Leibniz formulated calculus initially there was no well formulated concept of limit. Newton thought of the derivative as a quotient of *fluxions* and apparently Leibniz had some similar idea. A fluxion is alternatively called an *infinitesimal*. These are very very small quantities that have properties which ultimately forbid them being understood as real numbers.

Nonstandard analysis introduces some algebraic formalism to implement infinitesimals along side ordinary real numbers in an extended number system. In such a formalism we can literally claim the slope is formed by a ratio. Limits are replaced with the trouble of introducing infinitesimals carefully. As a matter of intuition, it is sometimes convenient to conceptualize $\frac{dy}{dx}$ as a fraction[1] You can read more in a number of places. For example, see this discussion on the Math Educator Stack Exchange website.

[1] It is a useful heuristic to guide the construction of models. For example, pressure $P = \frac{dF}{dA}$ where dF is the little pressure applied over the little area dA. We can think of $dF = PdA$ in such a context. Calculations such as that fall under a general family of calculations I call the *infinitesimal method*. If time permits we will discuss how the infinitesimal method aids applications of integral calculus. Of course, this is rather off topic at the moment

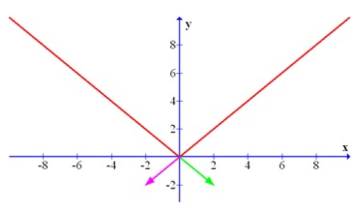
In-Class Example 3.1.3. *Find the slope of the function $y = x^2$ at $x = 1$ and graph both the function and its tangent line.*

Example 3.1.4. *The absolute value function is $f(x) = |x|$. Observe difference quotient has different left and right limits at zero.*

$$\lim_{h\to 0^-} \frac{|0+h| - |0|}{h} = \lim_{h\to 0^-} \frac{|h|}{h} = \lim_{h\to 0^-} \frac{-h}{h} = \lim_{h\to 0^-} (-1) = -1.$$

$$\lim_{h\to 0^+} \frac{|0+h| - |0|}{h} = \lim_{h\to 0^+} \frac{|h|}{h} = \lim_{h\to 0^+} \frac{h}{h} = \lim_{h\to 0^+} (1) = 1.$$

Therefore $f'(0) = \lim_{h\to 0} \frac{f(0+h) - f(0)}{h} \lim_{h\to 0} \frac{|h|}{h} = d.n.e.$. Geometrically this is evidenced in our inability to pick a unique tangent line at the origin. Which should we choose, the positive (purple) or the negative (green) sloped tangent line?



Another way the derivative at a point can fail to exist is for the function to have a vertical tangent. Vertical lines do not have a well-defined slope[2].

In-Class Example 3.1.5. *The graph of* $y = \sqrt[3]{x}$ *has a vertical tangent at* $(0, 0)$*. Show* $f'(0) = \infty$

We saw in the previous example that a function can be continuous at a point yet fail to be differentiable at that same point. In contrast, if a function is differentiable at a point it must be continuous at that point. (I defer proof to the end of this section)

Theorem 3.1.6. *If* f *is differentiable at* a *then* f *is continuous at* a*.*

If $f'(a)$ exists for a function f then $\lim_{x \to a} f(x) = f(a)$.

In-Class Example 3.1.7. *Identify the points in the graph given in lecture for which* $f'(a)$ *d.n.e.*

[2]that is more a deficiency of our current formalism than anything else. If we adopt a **parametric** viewpoint then the difference between horizontal and vertical tangents is washed away and much more general curves are easily described. We defer discussion of parametric curves until later in the calculus sequence. For now we focus on the special case of functions and graphs.

In-Class Example 3.1.8. *Let* $f(x) = 3$*. Calculate* $f'(a)$*.*

In-Class Example 3.1.9. *Let* $f(x) = mx + b$*. Calculate* $f'(a)$*.*

In-Class Example 3.1.10. *Let* $f(x) = x^3 + x - 2$*. Calculate* $f'(2)$*.*

3.1.1 Caratheodory's characterization of differentiability

This section was inspired in large part from Bartle and Sherbert's third edition of *Introduction to Real Analysis*. The central point is Caratheodory's Theorem which gives us an exact method to relate the function and its tangent line approximation (linearization). It is so simple it is clever. Consider a function f defined near $x = a$, we can write for $x \neq a$

$$f(x) - f(a) = \left[\frac{f(x) - f(a)}{x - a}\right](x - a).$$

If f is differentiable at a then as $x \to a$ the difference quotient $\frac{f(x)-f(a)}{x-a}$ tends to $f'(a)$ and we arrive at the approximation $f(x) - f(a) \approx f'(a)(x - a)$.

Theorem 3.1.11. *Caratheodory's Theorem.*

Let f be a function whose domain includes the interval I and let $a \in I$. Then f is differentiable at a iff there exists a function $\phi : I \to \mathbb{R}$ with the following two properties:

(**1**.) ϕ is continuous at a, (**2**.) $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in I$

Proof:($\Rightarrow$) Suppose f is differentiable at a. Define $\phi(a) = f'(a)$ and set $\phi(x) = \frac{f(x)-f(a)}{x-a}$ for $x \neq a$. Differentiability of f at a yields:

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = f'(a) \quad \Rightarrow \quad \lim_{x \to a} \phi(x) = \phi(a).$$

thus (1.) is true. Finally, note if $x = a$ then $f(x) - f(a) = \phi(x)(x - a)$ as $0 = 0$. If $x \neq a$ then $\phi(x) = \frac{f(x)-f(a)}{x-a}$ multiplied by $(x - a)$ gives $f(x) - f(a) = \phi(x)(x - a)$. Hence (2.) is true.

($\Leftarrow$) Conversely, suppose there exists $\phi : I \to \mathbb{R}$ with properties (1.) and (2.). Note (2.) implies $\phi(x) = \frac{f(x)-f(a)}{x-a}$ for $x \neq a$ hence $\lim_{x \to a} \frac{f(x)-f(a)}{x-a} = \lim_{x \to a} \phi(x)$. However, ϕ is continuous at a thus $\lim_{x \to a} \phi(x) = \phi(a)$. We find f is differentiable at a and $f'(a) = \phi(a)$. $\square$

Let us prove Theorem 3.1. Suppose f is differentiable at a then there exist ϕ continuous at $x = a$ for which $f(x) = f(a) + \phi(x)(x - a)$ for x near a. Then, by the usual limit laws,

$$\lim_{x \to a} f(x) = \lim_{x \to a} [f(a) + \phi(x)(x - a)] = f(a) + \phi(a)(a - a) = f(a).$$

Thus f is continuous at $x = a$. $\square$.

Caratheodory's result allows for direct verification of many rules of differential calculus. I'll offer proofs for the product rule and chain-rule using Caratheodory.

3.2 definition of the derivative function

The derivative of a function f is simply the function f' which is defined point-wise by the slope of the tangent line to the function f at the given point.

Definition 3.2.1. *derivative as a function.*

If a function f is differentiable at each point in $U \subseteq \mathbb{R}$ then we define a new function denoted f' which is called the **derivative of** f. It is defined point-wise by,

$$f'(x) = \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right)$$

We also may use the notation $f' = df/dx = \frac{df}{dx}$. Let $U \subseteq \mathbb{R}$. When a function is has a derivative f' which is continuous on U we say that $f \in C^1(U)$. If the derivative has a continuous derivative f'' on U then we say $f \in C^2(U)$. If we can take arbitrarily many derivatives which are continuous on U then we say that f is a **smooth function** and we denote this by $f \in C^\infty(U)$.

The notation $\frac{df}{dx}$ gives one the idea of taking the infinitesimal change dy and dividing by the infinitesimal change dx. There are times when it is quite useful to think of dy/dx as the quotient of infinitesimals but that time is not now. For now the symbol dy/dx is simply a notation to implicit the limiting process we just defined. Geometrically, it is clear that df/dx should give us a function whose values are the slope of f at each point where such slope is well-defined. The symbol $C^1(U)$ represents a set of functions, each function in this set is said to be **continuously differentiable**. There are functions which are differentiable but not continuously differentiable at a given point.

In-Class Example 3.2.2. *Suppose $f(x) = \sqrt{x}$. Calculate $f'(x)$ directly from the definition.*

Example 3.2.3. *Suppose $f(x) = \frac{1}{x^2}$. Calculate $f'(x)$ directly from the definition, assume $x \neq 0$. By definition,*

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0}\left(\frac{f(x+h)-f(x)}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{\frac{1}{x+h}-\frac{1}{x}}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{\frac{x-(x+h)}{x(x+h)}}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{-h}{hx(x+h)} \right) \\
&= \lim_{h\to 0}\left(\frac{-1}{x(x+h)} \right) \\
&= \frac{-1}{x^2}.
\end{aligned}
$$

In other notation,

$$\frac{df}{dx} = \frac{-1}{x^2} \qquad or \qquad \frac{d}{dx}\left[\frac{1}{x}\right] = \frac{-1}{x^2}.$$

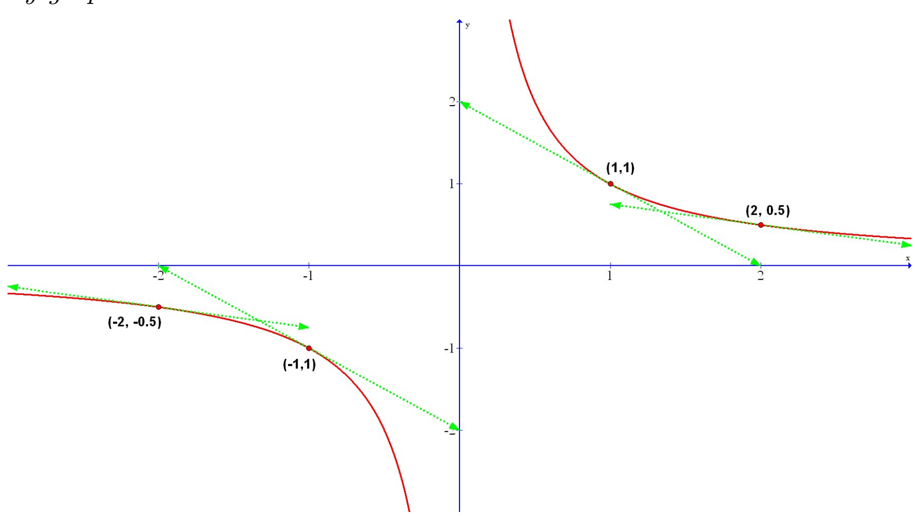
Let's take a moment to appreciate that the formula above allows us to set-up many different tangent lines for the graph $y = \frac{1}{x}$. For example,

$$f'(-2) = -1/9 \quad f'(-1) = -1 \quad f'(1) = -1 \quad f'(2) = -1/4$$

Tell us the slopes of the tangent lines at $(-2, -1/2), (-1, -1), (1, 1)$ and $(2, 1/2)$ respective. We find tangent lines:

$$y = -\frac{1}{2} - \frac{1}{9}(x+2), \quad y = -1-(x+1), \quad y = 1-(x-1), \quad y = \frac{1}{2} - \frac{1}{4}(x-2)$$

Here's how they graph:



3.3 linearity of the derivative and the power rule

These properties are crucial. Happily they're also way easier than our previous methods! I begin with linearity, we then work out the power rule for natural number powers.

Proposition 3.3.1.

The derivative d/dx is a *linear operator*. If $c \in \mathbb{R}$ and the functions f and g are differentiable then

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

We also can write $f'(x) = \frac{df}{dx}$ and

$$(cf)'(x) = cf'(x) \qquad (f+g)'(x) = f'(x) + g'(x).$$

Proof: follows easily from the definition of the derivative. Additivity:

$$\begin{aligned}
(f+g)'(x) &= \lim_{h\to 0}\left(\frac{(f+g)(x+h)-(f+g)(x)}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{f(x+h)+g(x+h)-f(x)-g(x)}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{f(x+h)-f(x)}{h} \right) + \lim_{h\to 0}\left(\frac{g(x+h)-g(x)}{h} \right) \\
&= f'(x)+g'(x).
\end{aligned}$$

Likewise, homogeneity:

$$\begin{aligned}
(cf)'(x) &= \lim_{h\to 0}\left(\frac{(cf)(x+h)-(cf)(x)}{h} \right) \\
&= \lim_{h\to 0}\left(\frac{cf(x+h)-cf(x)}{h} \right) \\
&= c\lim_{h\to 0}\left(\frac{f(x+h)-f(x)}{h} \right) \\
&= cf'(x).
\end{aligned}$$

While proofs may not excite you, I hope you can see that these are really very simple proofs. We didn't do anything except apply the properties of the limit itself (namely $\lim(f+g) = \lim f + \lim g$ and $\lim(cf) = c\lim f$) to the definition of the derivative for the functions f and g respective. □

Rather than stating the power rule from the outset we will examine a number of cases to suggest the rule. This will help us get more practice with the definition and perhaps a deeper appreciation for the power rule itself. In each case I will again emphasize the utility of the d/dx notation.

3.3.1 derivative of a constant

Suppose $f(x) = c$ for all $x \in \mathbb{R}$ then calculate,

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right) \\ &= \lim_{h \to 0} \left(\frac{c - c}{h} \right) \\ &= \lim_{h \to 0} (0) \\ &= 0. \end{aligned}$$

In *operator notation* we may write this result as follows:

$$\boxed{\frac{d}{dx}(\ c\ ) = 0}$$

Here we think of the operator $\frac{d}{dx}$ acting on a constant function to return the zero function.

3.3.2 derivative of identity function

Let $f(x) = x$ for all $x \in \mathbb{R}$,

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \left(\ \frac{f(x+h) - f(x)}{h} \ \right) \\ &= \lim_{h \to 0} \left(\ \frac{x + h - x}{h} \ \right) \\ &= \lim_{h \to 0} \left(\ \frac{h}{h} \ \right) \\ &= \lim_{h \to 0} (\ 1\ ) \\ &= 1. \end{aligned}$$

In *operator notation* we may write this result as follows:

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

Which also show you that $\frac{dx}{dx} = 1$ which helps reinforce my claim that thinking of dx as a tiny increment of x is not totally off base. We ought to have dx cancelling dx. Beware, this sort of thinking is not without peril.

3.3.3 derivative of quadratic function

Let $f(x) = x^2$ for all $x \in \mathbb{R}$,

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \left(\frac{f(x+h) - f(x)}{h} \right) \\ &= \lim_{h \to 0} \left(\frac{(x+h)^2 - x^2}{h} \right) \\ &= \lim_{h \to 0} \left(\frac{x^2 + 2xh + h^2 - x^2}{h} \right) \\ &= \lim_{h \to 0} (\ 2x + h\ ) \\ &= 2x. \end{aligned}$$

In *operator notation* we may write this result as follows:

$$\boxed{\frac{d}{dx}(\ x^2\ ) = 2x}$$

3.3.4 derivative of cubic function

Let $f(x) = x^2$ for all $x \in \mathbb{R}$,

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \left(\frac{(x+h)^3 - x^3}{h} \right) \\ &= \lim_{h \to 0} \left(\frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^3}{h} \right) \\ &= \lim_{h \to 0} (\ 3x^2 + 3xh + h^2\ ) \\ &= 3x^2. \end{aligned}$$

In *operator notation* we may write this result as follows:

$$\boxed{\frac{d}{dx}(\ x^3\ ) = 3x^2}$$

3.3.5 power rule

We should start to notice a pattern here: the derivative always returns a function with one less power than we put into the derivative. Let's list them to ponder the pattern,

(1.) $\frac{d}{dx}(1) = \frac{d}{dx}(x^0) = 0x^{0-1} = 0.$

(2.) $\frac{d}{dx}(x) = \frac{d}{dx}(x^1) = 1x^{1-1} = x.$

(3.) $\frac{d}{dx}(x^2) = 2x^{2-1} = 2x^1 = 2x.$

(4.) $\frac{d}{dx}(x^3) = 3x^{3-1} = 3x^2$.

I bet most of you could guess that $\frac{d}{dx}(x^4) = 4x^3$ (and you would be correct). We can summarize:

Proposition 3.3.2. *power rule*

Suppose $n \in \mathbb{R}$ then,

$$\boxed{\frac{d}{dx}\big(\ x^n\ \big) = nx^{n-1}.}$$

The proof I give below is for the case that $n \in \mathbb{N}$ meaning $n = 1, 2, 3, \dots$ (we already proved $n = 0, 1/2$ and -1 in previous arguments). We begin by recalling the *binomial theorem*,

$$(x+h)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}h^k = x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + h^n.$$

The symbol $\binom{n}{k} \equiv \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)\cdots 3\cdot 2\cdot 1}$ is read "n choose k" due to its application and interpretation in basic counting theory. They are also called the "binomial coefficients". There is a neat construction called Pascal's triangle which allows you to calculate the binomial coefficients without use of the formula just stated.

Proof: of power rule for $n \in \mathbb{N}$ follows from definition and binomial theorem:

$$\begin{aligned}
\frac{d}{dx}\big(\ x^n\ \big) &= \lim_{h\to 0}\left(\ \frac{(x+h)^n - x^n}{h}\ \right) \\
&= \lim_{h\to 0}\left(\ \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-1}h^2 + \cdots + h^n - x^n}{h}\ \right) \\
&= \lim_{h\to 0}\big(\ nx^{n-1} + \frac{n(n-1)}{2}x^{n-1}h + \cdots + h^{n-1}\ \big) \\
&= nx^{n-1}. \qquad \square
\end{aligned}$$

This proof is no good if $n = 1/2$ since we have no binomial theorem in that case[3]. However, we proved in Example 3.2.2 that $\frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}}$. In other words, $\frac{d}{dx}(x^{\frac{1}{2}}) = \frac{1}{2}x^{1-\frac{1}{2}}$ (power rule works). You should also note we also proved the case $n = -1$ in Example 3.2.3. In fact, the power rule is still true in the case that $n \in \mathbb{R} - \mathbb{N}$ [4], we just need another method of proof. I will give the general proof towards the end of this chapter.

In-Class Example 3.3.3. *Using the power rule correctly mostly boils down to you having a good grasp of laws of exponents.*

$$\frac{d}{dx}(xx^4) =$$

[3] We will learn in a later calculus course that the binomial expansion has infinitely many terms when $n \notin \mathbb{N}$.

[4] for example, $\frac{d}{dy}(y^{\pi+2}) = (\pi+2)y^{\pi+1} \approx 5.142y^{4.142}$

In-Class Example 3.3.4. *We can use linearity in conjunction with the power rule for added fun,*

$$\frac{d}{dx}\left[\frac{3x^3}{x} + \sqrt{4x}\right] =$$

Example 3.3.5. *Sometimes the independent variable is not "x", rather t, or c or even μ*

$$\frac{d}{dt}(\ ttt\ ) = \frac{d}{dt}(t^3) = 3t^2 \qquad \& \qquad \frac{d}{dc}(c) = 1 \qquad \& \qquad \frac{d}{d\mu}(\mu^k) = k\mu^{k-1}.$$

Proposition 3.3.6. *extended linearity.*

If functions $f_1, f_2, \dots, f_n$ are differentiable and $c_1, c_2, \dots c_n$ are constant then

$$\frac{d}{dx}\left[\ c_1 f_1 + c_2 f_2 + \cdots + c_n f_n\ \right] = c_1 \frac{df_1}{dx} + c_2 \frac{df_2}{dx} + \cdots + c_n \frac{df_n}{dx}$$

Or, using summation notation,

$$\frac{d}{dx}\left[\ \sum_{k=1}^{n} c_k f_k\ \right] = \sum_{k=1}^{n} c_k \frac{df_k}{dx}.$$

Proof: by induction. Left to the curious reader as an exercise.

In-Class Example 3.3.7.

$$\frac{d}{dx}(x + x^2 + 3) =$$

Or, suppose $a, b, c \in \mathbb{R}$ then

$$\frac{d}{dx}(\ ax^2 + \frac{b}{3}x^3 - \frac{1}{x} + c^3\ ) =$$

Example 3.3.8. *We will find other ways to do this one later, but now algebra is our only hope.*

$$\frac{d}{dx}\left[\ \frac{1}{\sqrt{x}}(\ x - \sqrt{x^3}\ ) + x^7\ \right] = \frac{d}{dx}(\ \sqrt{x} - x\ ) + 7x^6 = \frac{1}{2\sqrt{x}} - 1 + 7x^6.$$

In-Class Example 3.3.9. *What is the slope of the line $y = mx + b$ at the point $(x_o, mx_o + b)$?*

In-Class Example 3.3.10. *What is the slope of* $y = f(x) = ax^2 + bx + c$ *at the point* $(x_o, f(x_o))$*? What is the significance of the point where* $f'(x_o) = 0$ *?*

In-Class Example 3.3.11. *Let* $f(x) = 3x^7 + 2x^2 + 5$*. Find the equation of the tangent line to* $y = f(x)$ *at* $x = 1$*.*

3.4 the exponential function

Transcendental numbers cannot be defined in terms of a solution to an algebraic equation. In contrast, you could say that $\sqrt{2}$ is not a transcendental number since it is a solution to $x^2 = 2$ (it turns out $\sqrt{2}$ has a finite expansion in terms of *continued fractions*, it is a *quadratic irrational*). Mathematicians have shown that there exist infinitely many transcendental numbers, but there are precious few that are familiar to us. Probably $\pi = 3.1415\dots$ is the most famous. Next in popularity to π we find the number e named in honor of Euler. I can think of at least four seemingly distinct ways of defining $e = 2.718\dots$. We choose a definition which has the advantage of not using any mathematics beyond what we have so far discussed.

Let $f(x) = a^x$ for some $a > 0$, $a \neq 1$. Lets calculate the derivative of this exponential function, we'll use this calculation to define e in a somewhat indirect manner.

$$
\begin{aligned}
\frac{d}{dx}(\ a^x\ ) &= \lim_{h \to 0}\left(\ \frac{a^{x+h} - a^x}{h}\ \right) \\
&= \lim_{h \to 0}\left(\ \frac{a^x a^h - a^x}{h}\ \right) \\
&= \lim_{h \to 0}\left(\ \frac{a^x(a^h - 1)}{h}\ \right) \\
&= a^x \lim_{h \to 0}\left(\ \frac{a^h - 1}{h}\ \right)
\end{aligned}
$$

We will learn that this limit is finite for any $a > 0$. Thus the derivative of an exponential function is proportional to the function itself. We can define $a = e$ to be the case where the derivative is equal to the function.

Definition 3.4.1. *Euler's number; e.*

> The number e is the real number such that
> $$\lim_{h \to 0}\left(\ \frac{e^h - 1}{h}\ \right) = 1.$$

It is not at all obvious how to calculate that $e = 2.718\dots$ directly from this definition. This definition *implicitly* defines the number e. Notice that the calculation preceding the definition simplifies for this very special base; if $a = e$ then

$$\boxed{\frac{d}{dx}(\ e^x\ ) = e^x.}$$

The exponential function $f(x) = e^x = f'(x)$ is a very special function, it has the unique property that its output is the same as the slope of its tangent line at that point.

I have pictured a few representative tangents along with $y = e^x$.

By the way, I sometimes use the alternate notation $e^x = exp(x)$.

Remark 3.4.2.

In case you are curious and impatient I include a list of all the ways to define the exponential function and the number e in turn:

1. Define e^x to be the function such that $\frac{d}{dx}(e^x) = e^x$ then the number e would be defined by the function: $e^x|_{x=1} = e^1 = e$. This is nearly what we did in this section.

2. The following limit is a more direct description of what the value of e is,

$$\boxed{e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n}$$

notice that this limit is type 1^∞ and we have yet to discuss the tools to deal with such limits. Many folks take this as the definition of e, so be warned. It turns out that l'Hopital's Rule connects this definition and our definition. This definition arises naturally in the study of repeated multiplication in continuously compounded interest.

3. The natural logarithm $f(x) = \ln(x)$ arises in the study of integration in a very special role. You could define $f^{-1}(x) = e^x$ and then $e = f^{-1}(1)$.

4. The exponential could be defined by $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \cdots$ and again we could just set $e = 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \cdots$, perhaps this is the easiest to find e since with just the terms listed we get $e = 1 + 1 + 0.5 + 0.1\bar{6} + \cdots \approx 2.66$ not too far off the real value $e = 2.71\ldots$. This definition probably raises more questions than it answers so we'll just leave it at that until we discuss Taylor series.

By the way, the $\lim_{h\to 0}\frac{a^h-1}{h}$ is not easily calculated with the methods so far at our disposal. If you could show me how to calculate this limit by using the definition of e given in this section then I would probably award you some bonus points.

3.5 derivatives of sine and cosine

There are a few basic nontrivial limits which we need to derive in order to calculate the derivatives of sine and cosine. To begin we must establish the following for the *radian*-based sine function:

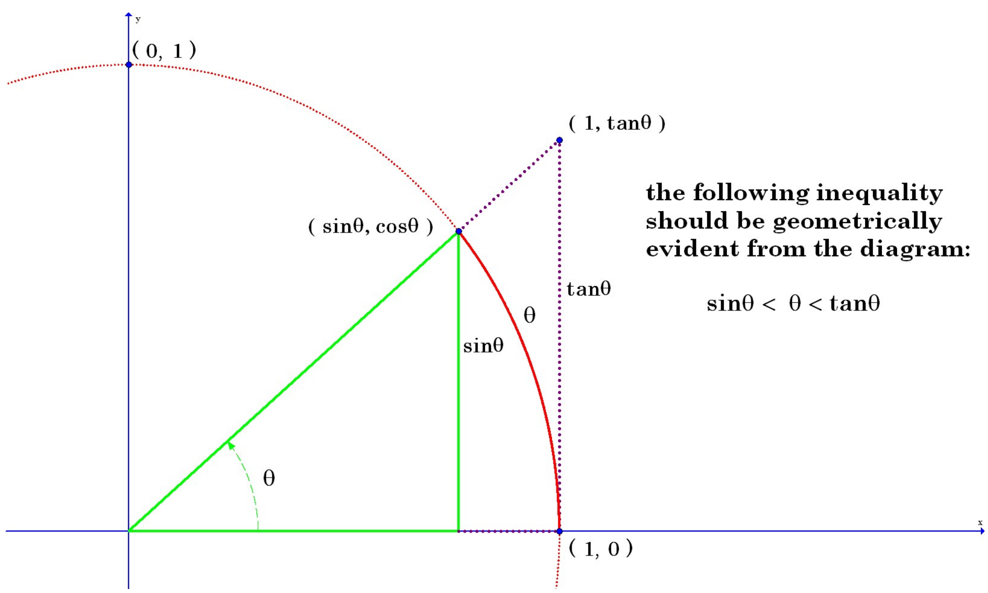
$$\boxed{\lim_{x\to 0}\left(\frac{\sin(x)}{x}\right) = 1.}$$

Observe that if we can prove $\lim_{x\to 0^+}\frac{\sin(x)}{x} = 1$ then the double sided limit follows naturally since sine is an odd function and

$$\lim_{x\to 0^-}\frac{\sin(x)}{x} = \lim_{x\to 0^-}\frac{\sin(-x)}{-x} = \lim_{y\to 0^+}\frac{\sin(y)}{y}$$

where in the last step we made the substitution $y = -x$ which naturally changes the left-limit of $x \to 0^-$ to the right limit $y \to 0^+$.

Proof: in the diagram below we consider a triangle inscribed in the unit circle (dotted-red) with angle $\theta > 0$ as pictured. The arclength subtended is given by $s = r\theta = \theta$ (bold red). Then the larger triangle has adjacent side-length of one unit thus $\tan(\theta) = \frac{opp}{adj}$ solves to yield $opp = \tan(\theta)$.



Continuing, notice that $\sin(\theta) < \theta < \tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)} \ \Rightarrow\ 1 < \frac{\theta}{\sin(\theta)} < \frac{1}{\cos(\theta)} \ \Rightarrow\ \cos(\theta) < \frac{\sin(\theta)}{\theta} < 1$. We proved previously that $\lim_{\theta\to 0^+}\cos(\theta) = 1$ and $\lim_{\theta\to 0^+} 1 = 1$ hence be the squeeze theorem it follows that $\lim_{\theta\to 0^+}\frac{\sin(\theta)}{\theta} = 1$. $\square$

I learned the argument above from Dr. Honore Mavinga and it is found in many calculus texts.

Next we show that, $\boxed{\lim_{x\to 0}\left(\dfrac{\cos(x)-1}{x}\right)=0}$. Observe,

$$\begin{aligned}
\lim_{x\to 0}\left(\frac{\cos(x)-1}{x}\right) &= \lim_{x\to 0}\left(\frac{\cos(x)-1}{x}\cdot\frac{\cos(x)+1}{\cos(x)+1}\right)\\
&= \lim_{x\to 0}\left(\frac{\cos^2(x)-1}{x(\cos(x)+1)}\right)\\
&= \lim_{x\to 0}\left(\frac{-\sin^2(x)}{x(\cos(x)+1)}\right)\\
&= \lim_{x\to 0}\left(\frac{\sin(x)}{x}\right)\cdot\lim_{x\to 0}\left(\frac{-\sin(x)}{\cos(x)+1}\right) = 1\cdot\frac{-\sin(0)}{\cos(0)+1} = 0.
\end{aligned}$$

We now have all the tools we need to derive the derivatives of sine and cosine. I should mention that I assume you know the "adding angles" formulas for sine and cosine:

$$\sin(a\pm b)=\sin(a)\cos(b)\pm\sin(b)\cos(a) \qquad \& \qquad \cos(a\pm b)=\cos(a)\cos(b)\mp\sin(a)\sin(b).$$

In-Class Example 3.5.1. *Show* $\dfrac{d}{dx}(\ \sin x\ )=\cos x$ *and* $\dfrac{d}{dx}(\ \cos x\ )=-\sin x$

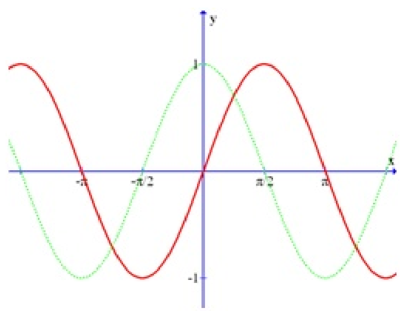
To summarize this section so far it's pretty simple,

Proposition 3.5.2. *derivatives of (radian-based) sine and cosine.*

$$\boxed{\frac{d}{dx}(\ \sin(x)\ ) = \cos(x) \qquad \frac{d}{dx}(\ \cos(x)\ ) = -\sin(x)}$$

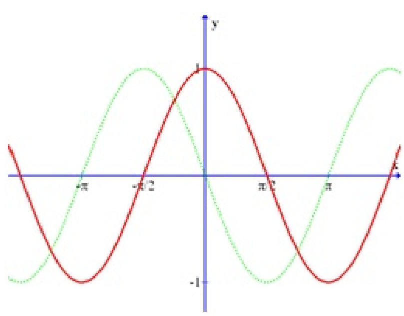
The function called "sine" for degree measure of angles is not the same function as the "sine" for radian-measured angle. We can relate them by a simple conversion: $\sin(\theta) = \sin_{degrees}(\frac{180\theta}{\pi})$. For example, $\sin(\pi/2) = \sin_{degrees}(90)$. Even your calculator knows these are different functions, that is why you have to change modes to clarify if you are using radians or degrees. Let it be understood that in calculus we always use radian-based sine and cosine.

Let's examine how this plays out graphically,



I have graphed in red $y = f(x) = \sin(x)$ and in green $y = f'(x) = \cos(x)$. Can you see that where the sine has a horizontal tangent the cosine function is zero? On the other hand whenever sine crosses the x-axis the cosine function is at either one or minus one. Question, what is the quickest that sine can possibly change? Notice that the slope of the sine function characterizes how quickly the sine function is changing.

The graph below has $y = g(x) = \cos(x)$ in red and $y = g'(x) = -\sin(x)$ in green.



I hope you see how the derivative and the function are related.

3.6 product rule

It is often claimed by certain students that $\frac{d}{dx}(fg) = \frac{df}{dx}\frac{dg}{dx}$ but this is almost never the case. Instead, you should use the **product rule**.

Proposition 3.6.1. *product rule.*

Let f and g be differentiable functions then

$$\boxed{\frac{d}{dx}(\ fg\ ) = \frac{df}{dx}g + f\frac{dg}{dx}}$$

which can also be written $(fg)' = f'g + fg'$.

Proof I: start with the definition of the derivative,

$$\begin{aligned}
(fg)'(x) &= \lim_{h \to 0}\left(\frac{(fg)(x+h) - (fg)(x)}{h} \right) \\
&= \lim_{h \to 0}\left(\frac{f(x+h)g(x+h) - f(x)g(x)}{h} \right) \\
&= \lim_{h \to 0}\left(\frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h} \right) \\
&= \lim_{h \to 0}\left(\left[\frac{f(x+h) - f(x)}{h}\right] \cdot g(x+h) \ + \ f(x) \cdot \left[\frac{g(x+h) - g(x)}{h}\right] \right) \\
&= \lim_{h \to 0}\left[\frac{f(x+h) - f(x)}{h}\right] \cdot \lim_{h \to 0}(\ g(x+h)\ ) \ + \ f(x) \cdot \lim_{h \to 0}\left[\frac{g(x+h) - g(x)}{h}\right] \\
&= f'(x)g(x) + f(x)g'(x).
\end{aligned}$$

In the very last step I used that $\lim_{h \to 0} g(x+h) = g(x)$ which is true since g is a differnentiable and thus continuous at x. $\square$

Proof II: Suppose f, g are differentiable at a then there exist ϕ_f, ϕ_g continuous at $x = a$ such that

$$f(x) = f(a) + \phi_f(x)(x-a) \qquad \& \qquad g(x) = g(a) + \phi_g(x)(x-a)$$

for all x near a and $\lim_{x \to a} \phi_f(x) = f'(a)$ and $\lim_{x \to a} \phi_g(x) = g'(a)$. Calculate,

$$\begin{aligned}
f(x)g(x) &= [f(a) + \phi_f(x)(x-a)]\,[g(a) + \phi_g(x)(x-a)] \\
&= f(a)g(a) + [\phi_f(x)g(a) + f(a)\phi_g(x) + \phi_f(x)\phi_g(x)(x-a)]\,(x-a)
\end{aligned}$$

Identify $\phi_{fg}(x) = \phi_f(x)g(a) + f(a)\phi_g(x) + \phi_f(x)\phi_g(x)(x-a)$ is a continuous function at $x = a$ for which $(fg)(x) - (fg)(a) = \phi_{fg}(x)(x-a)$. Thus fg is differentiable at a by Caratheodory. Moreover,

$$\lim_{x \to a} \phi_{fg}(x) = \lim_{x \to a} [\phi_f(x)g(a) + f(a)\phi_g(x) + \phi_f(x)\phi_g(x)(x-a)] = f'(a)g(a) + f(a)g'(a). \qquad \square$$

Example 3.6.2. *Lets derive the derivative of x^2 a new way,*

$$\frac{d}{dx}(x^2) = \frac{d}{dx}(xx) = \frac{dx}{dx}x + x\frac{dx}{dx} = 2x.$$

We derived this fact from the definition before, I think this way is easier. Anyway, I always recommend knowing more than one way to understand, it helps when doubt ensues.

In-Class Example 3.6.3. *Identify $f(x) = x$ and $g(x) = e^x$ in applying the product rule,*

$$\frac{d}{dx}(xe^x) =$$

In-Class Example 3.6.4. *Identify $f(x) = \sin(x)$ and $g(x) = \cos(x)$ for the product rule,*

$$\frac{d}{dx}(\sin(x)\cos(x)) =$$

The product rule for three factors can be derived the usual product rule applied twice:

$$\begin{aligned}
\frac{d}{dx}(fgh) &= \frac{d(fg)}{dx}h + fg\frac{dh}{dx} \\
&= \left(\frac{df}{dx}g + f\frac{dg}{dx}\right)h + fg\frac{dh}{dx} \\
&= \frac{df}{dx}gh + f\frac{dg}{dx}h + fg\frac{dh}{dx}
\end{aligned}$$

so the rule for products of three functions follows from the product rule for two functions. You could likewise derive that $(fghj)' = f'ghj + fg'hj + fgh'j + fghj'$ by the same logic.

Example 3.6.5.

$$\begin{aligned}
\frac{d}{dx}(x^2\sin(x)e^x) &= \frac{d(x^2)}{dx}\sin(x)e^x + x^2\frac{d(\sin(x))}{dx}e^x + x^2\sin(x)\frac{d(e^x)}{dx} \\
&= 2x\sin(x)e^x + x^2\cos(x)e^x + x^2\sin(x)e^x.
\end{aligned}$$

In-Class Example 3.6.6. *You can combine the product rule with linearity,*

$$\frac{d}{dx}(\sqrt{x} + 3x^3e^x) =$$

3.7 quotient rule

Let $Q(x) = f(x)/g(x)$ then and suppose $g(x) \neq 0$. Assume for the sake of discussion that $Q'(x)$ exists. Since $f(x) = Q(x)g(x)$ the product rule provides: $f' = (Qg)' = Q'g + Qg'$. Solve for Q',

$$Q' = \frac{f' - Qg'}{g} = \frac{f' - \frac{f}{g}g'}{g} = \frac{f'g - fg'}{g^2}. \tag{3.4}$$

The formula above is known as the **quotient rule**.

Proposition 3.7.1. *quotient rule.*

Let f and g be differentiable functions with $g \neq 0$,

$$\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{\frac{df}{dx}g - f\frac{dg}{dx}}{g^2}$$

this is called the **quotient rule**. In the prime notation, $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

Proof: if $Q(x) = f(x)/g(x)$ then we must show $Q'(x)$ exists. Consider the difference quotient

$$\begin{aligned} \frac{Q(a+h) - Q(a)}{h} &= \frac{1}{h}\left[\frac{f(a+h)}{g(a+h)} - \frac{f(a)}{g(a)}\right] \qquad (3.5) \\ &= \frac{1}{h}\left[\frac{f(a+h)}{g(a+h)} - \frac{f(a)}{g(a+h)} + \frac{f(a)}{g(a+h)} - \frac{f(a)}{g(a)}\right] \\ &= \frac{1}{h}\left[\frac{(f(a+h) - f(a))g(a)}{g(a+h)g(a)} + \frac{f(a)(g(a+h) - g(a))}{g(a)g(a+h)}\right] \\ &= \left[\frac{f(a+h) - f(a)}{h}\right]\frac{g(a)}{g(a+h)g(a)} + \frac{f(a)}{g(a)g(a+h)}\left[\frac{g(a+h) - g(a)}{h}\right] \end{aligned}$$

Notice as $h \to 0$ we have $\frac{f(a+h)-f(a)}{h} \to f'(a)$ and $\frac{g(a+h)-g(a)}{h} \to g'(a)$ and $f(a+h) \to f(a)$ and $g(a+h) \to g(a)$. In view of the difference quotient above we find $Q'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}$. $\square$

In-Class Example 3.7.2. *Show* $\frac{d}{dx}\tan x = \sec^2 x$.

Example 3.7.3.

$$\frac{d}{dx}\left(\frac{x^3}{x^2+7}\right) = \frac{3x^2(x^2+7) - x^3(2x)}{(x^2+7)^2} = \frac{x^4+21x^2}{(x^2+7)^2}.$$

In-Class Example 3.7.4.

$$\frac{d}{dx}\left(\frac{1}{3x+5}\right) =$$

Example 3.7.5.

$$\frac{d}{dx}(\sec(x)) = \frac{d}{dx}\left(\frac{1}{\cos(x)}\right) = \frac{\frac{d}{dx}(1)\cos(x) - 1\frac{d}{dx}(\cos(x))}{\cos^2(x)} = \frac{\sin(x)}{\cos^2(x)} = \sec(x)\tan(x).$$

In-Class Example 3.7.6. *Show* $\frac{d}{dx}\csc x = -\csc x \cot x$.

Example 3.7.7. *the quotient rule is used in conjunction with other rules sometimes, here I use linearity to start,*

$$\begin{aligned}
\frac{d}{dx}\left(e^x + \frac{x+x^2}{3-x}\right) &= \frac{d}{dx}(e^x) + \frac{d}{dx}\left(\frac{x+x^2}{3-x}\right) \\
&= e^x + \frac{\frac{d}{dx}(x+x^2)(3-x) - (x+x^2)\frac{d}{dx}(3-x)}{(3-x)^2} \\
&= e^x + \frac{(1+2x)(3-x) - (x+x^2)(-1)}{(3-x)^2} \\
&= e^x + \frac{3+6x-x^2}{x^2-6x+9}.
\end{aligned}$$

The last step was just algebra to make the answer pretty.

3.8 chain rule

If I were to pick a name for this rule it would be the *composite function rule* because the "chain rule" actually just tells us how to differentiate a composite function. Of all the rules so far this one probably requires the most practice. So be warned. Also, let me warn you about notation.

$$\boxed{f'(x) = \frac{df}{dx} = \frac{df}{dx}(x) = \left.\frac{df}{dx}\right|_x}$$

We have suppressed the (x) up to this point, reason being that it was always the same so we'd get tired of writing the (x) everywhere. Now we will find that we need to evaluate the derivative at things other than just (x). For example suppose that $f(x) = x^2$ so we have $f'(x) = 2x$ then

$$\frac{df}{dx}(x^3+7) = \left.\frac{df}{dx}\right|_{(x^3+7)} = 2(x^3+7)$$

We substituted $x^3 + 7$ in the place of x. I sometimes avoid the notation $\frac{df}{dx}(x)$ because it might be confused with multiplication by x. The difference should be clear from the context of the equation. Sometimes the substitution could be more abstract, again suppose $f(x) = x^2$ so we have $f'(x) = 2x$ then

$$\frac{df}{dx}(u) = \left.\frac{df}{dx}\right|_u = 2u$$

Proposition 3.8.1. *chain rule.*

The **Chain Rule** states that if $h = f \circ u$ is a composite function such that f is differentiable at $u(x)$ and u is differentiable at x then

$$\begin{aligned}\frac{d}{dx}(f \circ u) = (f \circ u)'(x) &= f'(u(x))u'(x) \\ &= \frac{df}{dx}(u(x))\frac{du}{dx} \\ &= \left.\frac{df}{dx}\right|_u \frac{du}{dx} \\ &= \frac{df}{du}\frac{du}{dx}.\end{aligned}$$

In words, the derivative of a composite function is the product of the derivative of the outside function (f) evaluated at the inside function (u) with the derivative of the inside function.

Please don't worry too much about all the notation, you are free to just use one that you like (provided it is correct of course). Anyway, let's look at an example or two before I give a proof.

Example 3.8.2. *Consider $h(x) = (3x+7)^5$ we can identify that this is a composite function with inside function $u(x) = 3x+7$ and outside function $f(x) = x^5$.*

$$\begin{aligned} \frac{d}{dx}(3x+7)^5 &= \frac{df}{dx}\bigg|_{3x+7} \frac{d}{dx}(3x+7) \\ &= 5x^4\bigg|_{3x+7} \cdot 3 \\ &= 15(3x+7)^4 \end{aligned}$$

I could also have written my work in the last example as follows,

$$\frac{d}{dx}(3x+7)^5 = \frac{d}{dx}(u^5) = 5u^4\frac{du}{dx} = 5(3x+7)^4 \cdot 3 = 15(3x+7)^4.$$

Or you could even suppress the u notation all together and just write

$$\frac{d}{dx}(3x+7)^5 = 5(3x+7)^4\frac{d}{dx}(3x+7) = 15(3x+7)^4.$$

I just recommend writing at least one middle step, if you try to do it all at once in your head you are likely to miss something generally speaking.

In-Class Example 3.8.3. *Identify u and use the chain-rule to differentiate:*

$$\frac{d}{dx}(\sin(x^2)) =$$

In-Class Example 3.8.4. *Identify u and use the chain-rule to differentiate:*

$$\frac{d}{dx}(exp(3x^2+x)) =$$

Proof of the Chain Rule: The proof I give here relies on approximating the function by its tangent line, this is called the *linearization of the function.* Observe that $u'(x) = \lim_{h\to 0}(\frac{u(x+h)-u(x)}{h})$ and we can rewrite the l.h.s. in terms of a matching limit $u'(x) = \lim_{h\to 0}(\frac{hu'(x)}{h})$. Thus

$$\lim_{h\to 0}\left(\frac{u'(x)h}{h}\right) = \lim_{h\to 0}\left(\frac{u(x+h)-u(x)}{h}\right).$$

This shows that if $h \to 0$ then $u'(x)h \approx u(x+h) - u(x)$ which says that $u(x+h) \approx u(x) + u'(x)h$. We can make the same argument to show that $f(u+\delta) \approx f(u) + f'(u)\delta$ for small δ (*the* $\delta = u'(x)h$ *which is small in the argument below since* $u'(x)$ *is finite and* $h \to 0$). Consider then,

$$\begin{aligned}
\frac{d}{dx}(f\circ u) &= \lim_{h\to 0}\left(\frac{(f\circ u)(x+h)-(f\circ u)(x)}{h}\right)\\
&= \lim_{h\to 0}\left(\frac{f(u(x+h))-f(u(x))}{h}\right)\\
&= \lim_{h\to 0}\left(\frac{f(u(x)+u'(x)h))-f(u(x))}{h}\right)\\
&= \lim_{h\to 0}\left(\frac{f(u(x))+u'(x)hf'(u(x))-f(u(x))}{h}\right)\\
&= \lim_{h\to 0}\big(u'(x)f'(u(x))\big)\\
&= f'(u(x))u'(x).
\end{aligned}$$

So the proof of the chain rule relies on approximating both the inside and outside function by their tangent line. I give another statement of the chain-rule as well as a careful proof via Caratheodory at the conclusion of this section. Let's get back to the examples.

In-Class Example 3.8.5.

$$\frac{d}{dx}(e^{\sqrt{x}}) =$$

In-Class Example 3.8.6. *Let a be a constant,*

$$\frac{d}{dx}(\sin(ax)) =$$

Example 3.8.7. *Let a be a constant,*

$$\frac{d}{dx}(e^{ax}) = \frac{d}{dx}(e^u) = e^u\frac{du}{dx} = e^{ax}\frac{d}{dx}(ax) = ae^{ax}$$

Example 3.8.8. *Let a be a constant,*

$$\frac{d}{dx}(f(ax)) = \frac{d}{dx}(f(u)) = f'(u)\frac{du}{dx} = f'(ax)\frac{d}{dx}(ax) = af'(ax).$$

I let the function f be arbitrary just to point out the past two examples can be generalized to any expression of this type. We must have a function which is differentiable at ax in order for the calculation to hold true.

We will neglect the extra u notation past this point unless it is helpful,

In-Class Example 3.8.9. *Let a, b, c be constants,*

$$\frac{d}{dx}\left(\sqrt{ax^2 + bx + c} \right) =$$

I admit that all the examples up to this point have been fairly mild. The remainder of the section I give examples which combine the chain rule with itself and the product or quotient rules.

In-Class Example 3.8.10.

$$\frac{d}{dx}\left(\sqrt{x^2 + \sqrt{x^2 + 3}} \right) =$$

Example 3.8.11. *Let a, b, c be constants,*

$$\begin{aligned}\frac{d}{dx}\Big(\cos(a\sin(bx+c))\Big) &= -\sin(a\sin(bx+c))\cdot\frac{d}{dx}\big(a\sin(bx+c)\big)\\ &= -\sin(a\sin(bx+c))\cdot a\cos(bx+c)\frac{d}{dx}(bx+c)\\ &= -ab\sin(a\sin(bx+c))\cos(bx+c).\end{aligned}$$

We have to work outside in, one step at a time. Both of these examples followed the pattern $(f\circ g\circ h)(x) = f(g(h(x)))$ which has the derivative $(f\circ g\circ h)'(x) = f'(g(h(x)))g'(h(x))h'(x)$. Of course, in practice I do not try to remember that formula, I just apply the chain rule repeatedly until the problem boils down to basic derivatives.

In-Class Example 3.8.12.

$$\frac{d}{dx}(x^3e^{2x}\cos(x^2)) =$$

Example 3.8.13.

$$\begin{aligned}\frac{d}{dx}(e^xx^2)^3 &= 3(e^xx^2)^2\frac{d}{dx}\big(e^xx^2\big)\\ &= 3(e^xx^2)^2\left(\frac{d(e^x)}{dx}x^2 + e^x\frac{d(x^2)}{dx}\right)\\ &= 3(e^xx^2)^2\big(x^2e^x + 2xe^x\big).\end{aligned}$$

The better way to think about this one is that $(e^xx^2)^3 = e^{3x}x^6$ then the differentiation is prettier in my opinion

$$\begin{aligned}\frac{d}{dx}(e^{3x}x^6) &= \frac{d(e^{3x})}{dx}x^6 + e^{3x}\frac{d(x^6)}{dx}\\ &= 3e^{3x}x^6 + 6x^5e^{3x}.\end{aligned}$$

Example 3.8.14.

$$\begin{aligned}\frac{d}{d\theta}\left(\frac{\sin(3\theta)}{\sqrt{\theta+4}}\right) &= \frac{3\cos(3\theta)\sqrt{\theta+4} - \sin(3\theta)\frac{1}{2\sqrt{\theta+4}}}{(\sqrt{\theta+4})^2} \\ &= \frac{3\cos(3\theta)\sqrt{\theta+4}\sqrt{\theta+4} - \sin(3\theta)\frac{\sqrt{\theta+4}}{2\sqrt{\theta+4}}}{(\sqrt{\theta+4})^3} \\ &= \frac{6(\theta+4)\cos(3\theta) - \sin(3\theta)}{2(\theta+4)^{\frac{3}{2}}}.\end{aligned}$$

Example 3.8.15. *Observe we can derive the power rule from the product rule.*

$$\begin{aligned}\frac{d}{dx}(x^n) = \frac{d}{dx}(xx\cdots x) &= \frac{dx}{dx}x^{n-1} + x\frac{dx}{dx}x^{n-2} + \cdots + x^{n-1}\frac{dx}{dx} \\ &= x^{n-1} + x^{n-1} + \cdots + x^{n-1} \\ &= nx^{n-1}.\end{aligned}$$

In-Class Example 3.8.16.

$$\frac{d}{dt}\big(\sin(\sqrt{2t-1})\big) =$$

In-Class Example 3.8.17.

$$\frac{d}{dt}(t^2\cos(\sin(t)) =$$

In most of the examples we have been able to reduce the answer into some expression involving no derivatives. This is generally not the case. As the next couple of examples illustrate, we can have expressions that once differentiated yield a new expressions which still contain derivatives.

Example 3.8.18. *Suppose that c and f are functions of t then,*

$$\frac{d}{dt}(cf) = \frac{dc}{dt}f + c\frac{df}{dt}$$

Notice that if c is a constant then $\frac{dc}{dt} = 0$ so in that case we have that $\frac{d}{dt}(cf) = c\frac{df}{dt}$.

Example 3.8.19. *Suppose that a particle travels on a circle of radius R centered at the origin. The particle has coordinates (x, y) that satisfy the equation of a circle; $x^2 + y^2 = R^2$. Moreover, both x and y are functions of time t. What can we say about dx/dt and dy/dt ?*

$$\frac{d}{dt}(x^2 + y^2) = 2x\frac{dx}{dt} + 2y\frac{dy}{dt}$$

Notice since the radius R is constant it follows that R^2 is also constant thus $\frac{d}{dt}(R^2) = 0$. Apparently the derivatives dx/dt and dy/dt must satisfy

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0$$

Now this says that $\frac{dx}{dt} = \frac{-y}{x}\frac{dy}{dt}$ (for points with $x \neq 0$).

3.8.1 Caratheodory's Theorem and the chain rule

I have long been disatisfied with the earlier proof of the chain rule in this section from an analysis perspective.

Theorem 3.8.20. *Chain Rule.*

Suppose f, g are functions and I, J are intervals such that $I \subseteq dom(f)$ and $f(I) \subseteq J \subseteq dom(g)$. If $a \in I$ and f is differentiable at a and g is differentiable at $f(a)$ then $g \circ f$ is differentiable at a and $(g \circ f)'(a) = g'(f(a))f'(a)$.

Proof: apply Caratheodory's Theorem twice. Since f is differentiable at a we know there exists ϕ such that $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in I$ and $\phi(a) = f'(a)$. Since g is differentiable at $f(a)$ we know these exists β such that $g(y) - g(f(a)) = \beta(y)(y - f(a))$ for all $y \in J$ where $\beta(f(a)) = g'(f(a))$. Suppose $x \neq a$ and calculate:

$$\frac{(g(f(x)) - g(f(a))}{x - a} = \frac{\beta(f(x))(f(x) - f(a))}{x - a} = \frac{\beta(f(x))\phi(x)(x - a)}{x - a} = \beta(f(x))\phi(x).$$

By Caratheodory's Theorem we know $\lim_{x \to a} \beta(f(x)) = g'(f(a))$ and $\lim_{x \to a} \phi(x) = f'(a)$. Therefore,

$$\lim_{x \to a} \frac{(g(f(x)) - g(f(a))}{x - a} = \lim_{x \to a} \beta(f(x))\phi(x) = \lim_{x \to a} \beta(f(x)) \cdot \lim_{x \to a} \phi(x) = g'(f(a))f'(a). \quad \square$$

3.9 higher derivatives

Higher derivatives are defined iteratively.

Definition 3.9.1. *the n-th derivative of a function.*

Suppose $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}$ and $U \subseteq dom(f)$. We define $f^{(0)}(x) = f(x)$ and $f^{(1)}(x) = \frac{df}{dx}$ for all such $x \in dom(f)$ that $f'(x) \in \mathbb{R}$. Furthermore, for each $n \in \mathbb{N}$ we define $f^{(n+1)}(x) = \frac{d}{dx}[f^{(n)}(x)]$ for all such $x \in dom(f)$ that $f^{(n+1)}(x) \in \mathbb{R}$. If f has continuous derivatives $f', f'', \dots, f^{(k)}$ on $U \subseteq dom(f)$ then $f \in C^k(U)$. If we can take arbitrarily many derivatives of f and those derivatives are continuous on $U \subseteq dom(f)$ then we say f is **smooth**. The set of all smooth functions on $U \subseteq \mathbb{R}$ is denoted $C^\infty(U)$.

Many elementary functions are smooth over large subsets of $\mathbb{R}$.

In-Class Example 3.9.2. *Suppose $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$. Derivatives of all orders for $f(x)$.*

Geometrically the second derivative of a function is connected to the curvature of the graph. The third, fourth and higher derivatives also contain geometric information about a function. If we are given all derivatives of a smooth function it is often possible to recreate the function everywhere with a formula built using those derivatives. Using the last example, you might notice that

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \frac{1}{6}f'''(0)x^3 + \frac{1}{24}f^{(4)}(0)x^4 + \frac{1}{120}f^{(5)}(0)x^5.$$

Knowledge of the derivatives at zero gives global information about f in the equation above. This is an interesting pattern which we will explore in more depth later.

Physically the higher derivatives are also of interest.

Example 3.9.3. *Suppose $s : \mathbb{R} \to \mathbb{R}$ is the position of some particle as a function of time t. The* **velocity** *at time t is defined to be (the dot-notation is still prevalent in modern classical mechanics courses, it dates back to Newton whereas the d/dx notation is due to Leibniz)*

$$\boxed{v(t) = \frac{ds}{dt} = \dot{s}.}$$

The second derivative with respect to time is called the **acceleration** *at time t and it is defined by*

$$\boxed{a(t) = \frac{d^2 s}{dt^2} = \ddot{s}.}$$

Notice we can equivalently state $a(t) = \frac{dv}{dt} = \dot{v}$. If the particle has mass m then Newton's Second Law states that $F_{net} = ma$ where F_{net} is the total force placed on the mass m. Beyond acceleration we have the **jerk** *which is the instantaneous rate of change of the acceleration $j(t) = \frac{da}{dt} = \frac{d^3 s}{dt^3}$.*

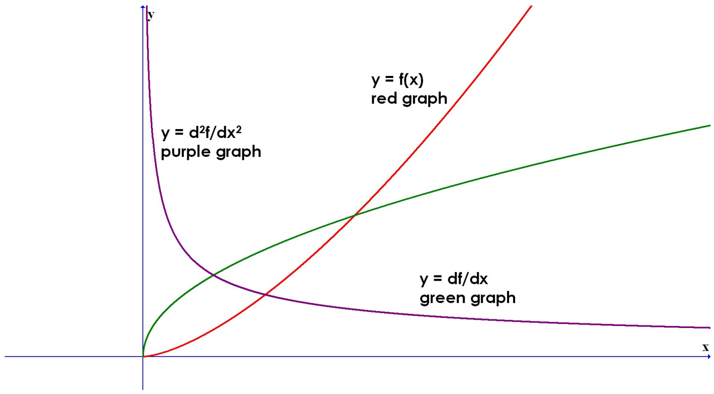
In-Class Example 3.9.4. *Let $s(t) = 3t^2 + t^3$ be the position at time t. Calculate the velocity, acceleration and jerk for the given position.*

Example 3.9.5. *How many times is $f(x) = x^{\frac{3}{2}}$ differentiable at zero[5]? Calculate,*

$$f'(x) = \frac{3}{2}x^{\frac{1}{2}}, \qquad f''(x) = \frac{3}{4}x^{\frac{-1}{2}}$$

Notice that $f''(0) = \frac{3}{4\sqrt{0}} \notin \mathbb{R}$. The second derivative of f is not defined at zero. We say that f is differentiable at zero, but f is not twice differentiable at zero. The source of this difficulty is that f' has a vertical tangent at zero.

[5] In such a case we ask only that the right limit of the difference quotient exists. We define that $f'(0) = \lim_{h \to 0^+} \frac{f(0+h)-f(0)}{h}$ in the case $0 \in \partial(dom(f))$



On the other hand it is not hard to see that $f \in C^\infty(0,\infty)$ since differentiating n-times we'll find $f^{(n)}(x) = kx^{\frac{3}{2}-n}$ for some constant k. The formula for $f^{(n)}(x)$ is clearly well-defined for $x > 0$.

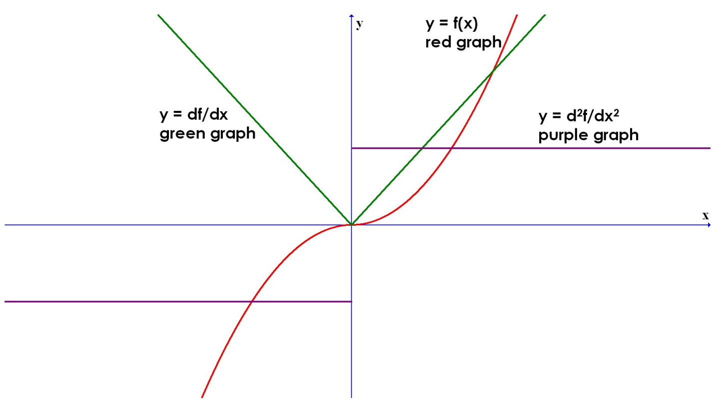
Example 3.9.6. *Another interesting function which fails to be smooth is $f(x) = x|x|$. The graph resembles a cubic function but it is actually a pair of half-parabolas glued at the origin. For $x > 0$ we have $f(x) = x^2$ and for $x < 0$ we have $f(x) = -x^2$. It follows that*

$$f'(x) = \begin{cases} 2x & x \geq 0 \\ -2x & x \leq 0 \end{cases}$$

In this case $f'(0) = 0$ since $\lim_{h\to 0^-} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0^+} \frac{f(0+h)-f(0)}{h} = 0$. Consider the second derivative,

$$f''(x) = \begin{cases} 2 & x > 0 \\ -2 & x < 0 \end{cases}.$$

In this case $f''(0)$ does not exist since $\lim_{h\to 0^-} \frac{f(0+h)-f(0)}{h} = -2$ whereas $\lim_{h\to 0^+} \frac{f(0+h)-f(0)}{h} = 2$. The source of this difficulty is the kink in the graph of f' at zero.



If you want a function which is just k-times differentiable at zero you could use $f(x) = x^k|x|$. Notice that in all the examples I've given thus far if the function was differentiable on some interval then the derivative function was also continuous. In other words, you might wonder if the distinction between *differentiable* and *continuously differentiable* is a meaningful distinction. Since I'm posing this question by now you probably know the answer is yes.

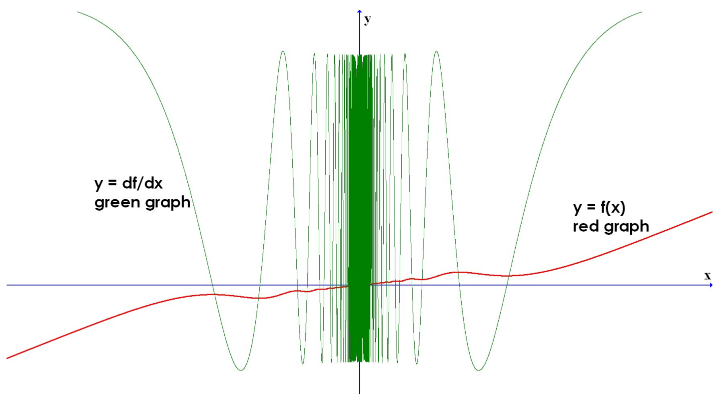
Example 3.9.7. *I found this example in Hubbard's advanced calculus text(see Ex. 1.9.4, pg. 123). It is a source of endless odd examples, notation and bizarre quotes. Let* $f(x) = 0$ *and*

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

for all $x \neq 0$*. I can be shown that the derivative* $f'(0) = 1/2$ *(hard to see from the green graph !). Moreover, we can show that* $f'(x)$ *exists for all* $x \neq 0$*, we can calculate:*

$$f'(x) = \frac{1}{2} + 2x\sin\frac{1}{x} - \cos\frac{1}{x}$$

Notice that $dom(f') = \mathbb{R}$*. Note then that the tangent line at* $(0, 0)$ *is* $y = x/2$*. You might be tempted to say then that this function is increasing at a rate of 1/2 for* x *near zero. But this claim would be false since you can see that* $f'(x)$ *oscillates wildly without end near zero.*



We have a tangent line at $(0, 0)$ *with positive slope for a function which is not increasing at* $(0, 0)$ *(recall that increasing is a concept we must define in a open interval to be careful). This function has infinitely many critical points in a nbhd. of zero. You couldn't even draw a sign-chart for the derivative if you wanted. Continuity of the derivative helps eliminate pathological examples.*

This sort of example is likely to occur to mathematicians but not so likely to occur to anyone else. Usually if a function is differentiable at a point is also continuously differentiable. For functions of several variables the story is much more involved [6]

[6] Continuity of the derivative function is later replaced with the requirement that the partial derivatives of a multivariate function are continuously differentiable. It is a fortunate accident of one-dimensional mathematics that the tangent line is well-defined in the case the derivative is not continuous (and yet exists). For functions of several variables existence of partial derivatives need not indicate the existence of a tangent space. However, it is still true that continuous differentiability signals that the tangent plane both exists and well-approximates the mapping near the point of tangency. I discuss this in more depth in calculus III or advanced calculus.

3.10 implicit differentiation and derivatives of inverse functions

Up to this point we have primarily dealt with expressions where it is convenient to just differentiate what we are given directly. We just wrote down our $f(x)$ and proceeded with the tools at our disposal, namely linearity, the product, quotient and chain rules. For the most part this direct approach will work, but there are problems which are best met with a slightly indirect approach. We typically call the thing we want to find y then we'll differentiate some equation which characterizes y and usually we get an equation which implicitly yields $\frac{dy}{dx}$. This technique will reward us with the formulas for the derivatives of all sorts of inverse functions. Before we get to the inverse functions let's start with a few typical implicit derivatives.

Example 3.10.1. *Observe that the equation $x^2 + y^3 = e^y$ implicitly defines y as a function of x (locally). Let's find $\frac{dy}{dx}$. Differentiate the given equation on both sides.*

$$\frac{d}{dx}(x^2 + y^3) = \frac{d}{dx}(e^y)$$

now differentiate and use the chain rule where appropriate,

$$2x + 3y^2\frac{dy}{dx} = e^y\frac{dy}{dx}$$

Now solve for $\frac{dy}{dx}$,

$$(e^y - 3y^2)\frac{dy}{dx} = 2x \;\Rightarrow\; \boxed{\frac{dy}{dx} = \frac{2x}{e^y - 3y^2}}$$

Notice that this equation is a little unusual in that the derivative involves both x and y.

In-Class Example 3.10.2. *Given $xy + \sin(x) = e^{xy}$. Calculate $\frac{dy}{dx}$.*

Proposition 3.10.3. *Method to Calculate Derivative of Inverse Function*

To calculate $\frac{d}{dx}f^{-1}(x)$ given that we already know the derivative of f we can:

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

(2.) solve for $x = f(y)$ and differentiate with respect to x to obtain $1 = \frac{df}{dy}\frac{dy}{dx}$.

(3.) solve for $\frac{dy}{dx}$ and eliminate y as appropriate to the example.

Example 3.10.4. *Let* $y = \cos^{-1}(x)$ *we wish to derive* $\frac{d}{dx}(\cos^{-1}(x))$*. To begin we take the cosine of both sides of* $y = \cos^{-1}(x)$ *to obtain*

$$\cos(y) = \cos(\cos^{-1}(x)) = x$$

Now differentiate with respect to x *and solve for* $\frac{dy}{dx}$

$$-\sin(y)\frac{dy}{dx} = 1 \quad \Rightarrow \frac{dy}{dx} = \frac{-1}{\sin(y)}$$

Now $\sin^2(y) + \cos^2(y) = 1$ *thus* $\sin(y) = \sqrt{1-\cos^2(y)}$ *but remember that we found* $\cos(y) = x$ *so* $\sin(y) = \sqrt{1-x^2}$ *thus we find*

$$\boxed{\frac{d}{dx}\big(\ \cos^{-1}(x)\ \big) = \frac{-1}{\sqrt{1-x^2}}.}$$

In-Class Example 3.10.5. *Show* $\dfrac{d}{dx}\big(\ \ln(x)\ \big) = \dfrac{1}{x}$.

In-Class Example 3.10.6. *Show* $\dfrac{d}{dx}\big(\ \sin^{-1}(x)\ \big) = \dfrac{1}{\sqrt{1-x^2}}$

In-Class Example 3.10.7. *Show* $\dfrac{d}{dx}(\tan^{-1}(x)) = \dfrac{1}{1+x^2}$.

Example 3.10.8. *Let $y = \sec^{-1}(x)$ we wish to derive $\frac{d}{dx}(\sec^{-1}(x))$. To begin we take the secant of both sides of $y = \sec^{-1}(x)$ to obtain*

$$\sec(y) = \sec(\sec^{-1}(x)) = x$$

Now differentiate with respect to x and solve for $\frac{dy}{dx}$

$$\sec(y)\tan(y)\frac{dy}{dx} = 1 \quad \Rightarrow \frac{dy}{dx} = \frac{1}{\sec(y)\tan(y)}$$

Now $\tan^2(y) + 1 = \sec^2(y)$ tells us that $\tan(y) = \sqrt{\sec^2(y) - 1}$. But we know that in this example $\sec(y) = x$ hence $\tan(y) = \sqrt{x^2 - 1}$. Thus,

$$\boxed{\frac{d}{dx}\big(\ \sec^{-1}(x)\ \big) = \frac{1}{x\sqrt{x^2 - 1}}.}$$

I hope you can see the pattern in the last five examples. To find the derivative of an inverse function we simply need to know the derivative of the function plus a little algebra. The same technique would allow us to derive the derivatives of $\cosh^{-1}(x), \sinh^{-1}(x), \tanh^{-1}(x), \csc^{-1}(x), \cot^{-1}(x)$. I have not included those in these notes because we have yet to calculate the derivatives of $\cosh(x), \sinh(x), \tanh(x), \csc(x), \cot(x)$. Rest assured these functions can be dealt with by the same techniques we thus far exhibited in these notes. The next examples follow the same general idea, but the pattern differs a bit.

Example 3.10.9. *Suppose that $y = a^x$ we have yet to calculate the derivative of this for arbitrary $a > 0$ except the one case $a = e$. Turns out that this one case will dictate what the rest follow. Take the natural log of both sides to obtain $\ln(y) = \ln(a^x) = x\ln(a)$. Now differentiate, by Example 3.10.5,*

$$\frac{d}{dx}\big(\ln(y)\big) = \frac{1}{y}\frac{dy}{dx} = \frac{d}{dx}(x\ln(a)) = \ln(a).$$

Now solve for $\frac{dy}{dx}$,

$$\frac{dy}{dx} = \ln(a)y = \ln(a)a^x \quad \Longrightarrow \quad \boxed{\frac{d}{dx}(a^x) = \ln(a)a^x.}$$

I should mention that I know another method to derive the boxed equation. In fact I prefer the following method which is based on a useful purely algebraic trick: $a^x = exp(x\ln(a))$ so we can just calculate

$$\frac{d}{dx}(a^x) = \frac{d}{dx}(e^{x\ln(a)}) = e^{x\ln(a)}\frac{d(x\ln(a))}{dx} = e^{x\ln(a)}\ln(a) = \ln(a)a^x.$$

but beware the sneaky step, how did I know to insert the $exp \circ ln$? I just did.

Example 3.10.10. *Suppose that $y = x^x$. This is not a function we have encountered before. It is neither a power nor an exponential function, it's sort of both. I'll admit the only place I've seen them is on calculus tests. Anyway to begin we take the natural log of both sides;* $\ln(y) = \ln(x^x) = x\ln(x)$. *Differentiate w.r.t x,*

$$\frac{1}{y}\frac{dy}{dx} = \ln(x) + x\frac{1}{x} \implies \frac{dy}{dx} = y(\ln(x)+1) \implies \boxed{\frac{d}{dx}(x^x) = (\ln(x)+1)x^x.}$$

In-Class Example 3.10.11. *Calculate* $\frac{d}{dx}x^{x\sin x}$

If you have a problem with an unpleasant exponent it sometimes pays off take the logarithm. It may change the problem to something you can deal with. The process of morphing an unsolvable problem to one which is solvable through known methods is most of what we do in calculus. We learn a few basic tools then we spend most of our time trying to twist other problems back to those simple cases. I have one more basic derivative to address in this section.

Example 3.10.12. *Let $y = \log_a(x)$ we can exponentiate both sides w.r.t. base a which cancels the $\log_a$ in the sense $a^{\log_a(x)} = x$,*

$$a^y = x \implies \ln(a)a^y\frac{dy}{dx} = 1 \implies \frac{dy}{dx} = \frac{1}{\ln(a)a^y}$$

But then since $a^y = x$ therefore we conclude,

$$\boxed{\frac{d}{dx}\big(\log_a(x)\big) = \frac{1}{\ln(a)x}}$$

Notice in the case $a = e$ we have $\log_e(x) = \ln(x)$ and $\ln(e) = 1$. Therefore, this result agrees with Example 3.10.5.

At this point I have derived almost every elementary function's derivative. Those which I have not calculated so far can certainly be calculated using nothing more than the strategies and methods advertised thus far.

3.11 logarithmic differentiation

The idea of logarithmic differentiation is fairly simple. When confronted with a product of bunch of things one can take the logarithm to convert it to a sum of things. Then you get to differentiate a sum rather than a product. This is a labor saving device. We pause to note how the chain rule works for the natural log:

$$\frac{d}{dx}\ln(u) = \frac{1}{u}\frac{du}{dx} = \frac{du/dx}{u}$$

we make use of the identity above throughout what follows.

Example 3.11.1. *Find $\frac{dy}{dx}$ via logarithmic differentiation. Let.*

$$y = \left(\frac{1}{2-x}\right)(x+32)^{\frac{1}{4}}(x^2-3)^4$$

Take the natural log to begin,

$$\begin{aligned}\ln(y) &= \ln(2-x)^{-1} + \ln(x+32)^{\frac{1}{4}} + \ln(x^2-3)^4 \\ &= -\ln(2-x) + \frac{1}{4}\ln(x+32) + 4\ln(x^2-3).\end{aligned}$$

We used the properties of the natural log to simplify as best we could before going on to the next step: differentiate w.r.t. x

$$\frac{1}{y}\frac{dy}{dx} = \frac{1}{2-x} + \frac{1}{4(x+32)} + \frac{4(2x)}{x^2-3}$$

$$\Rightarrow \boxed{\frac{dy}{dx} = \left(\frac{1}{2-x}\right)(x+32)^{\frac{1}{4}}(x^2-3)^4\left(\frac{1}{2-x} + \frac{1}{4(x+32)} + \frac{8x}{x^2-3}\right)}.$$

This is much easier than the 3-term product rule for this problem

In-Class Example 3.11.2. *Find the derivative of $y = xe^{x^2+9}\sqrt{3x+7}$ using log. differentiation.*

In-Class Example 3.11.3. *Let a, b, c be constants. Differentiate* $y = \left(\frac{1}{x-a}\right)\left(\frac{1}{x-b}\right)^2\left(\frac{1}{x-c}\right)^3$ *via the technique of logarithmic differentiation*

Example 3.11.4. *Differentiate y.*

$$y = (x^2+1)(x-3)^2(x^3+x)^3(x-1)^4$$

Take the natural log to begin,

$$\ln(y) = \ln(x^2+1) + 2\ln(x-3) + 3\ln(x^3+x) + 4\ln(x-1)$$

$$\Rightarrow \ \frac{1}{y}\frac{dy}{dx} = \frac{2x}{x^2+1} + \frac{2}{x-3} + \frac{3(3x^2+1)}{x^3+x} + \frac{4}{x-1}$$

$$\Rightarrow \ \boxed{\frac{dy}{dx} = y\left(\frac{2x}{x^2+1} + \frac{2}{x-3} + \frac{3(3x^2+1)}{x^3+x} + \frac{4}{x-1}\right)}.$$

Example 3.11.5. *Sometimes we might have a to start with, but the same algebraic wisdom applies, simplify products to sums then differentiate. Find $\frac{dy}{dx}$ for $y = \ln\left(\frac{\sin(x)\sqrt{x}}{x^2+3x-2}\right)$.*

$$y = \ln\left(\frac{\sin(x)\sqrt{x}}{x^2+3x-2}\right) = \ln(\sin(x)) + \frac{1}{2}\ln(x) - \ln(x^2+3x-2).$$

Now differentiate w.r.t. x and we're done.

$$\frac{dy}{dx} = \frac{\cos(x)}{\sin(x)} + \frac{1}{2x} - \frac{2x+3}{x^2+3x-2}.$$

Example 3.11.6. *What about*

$$y = \ln((x+1)^{30} + 2)$$

We **cannot** *simplify this one because we do not have a product inside the natural log. Just differentiate w.r.t x*

$$\frac{dy}{dx} = \frac{1}{(x+1)^{30}+2}\frac{d}{dx}\Big((x+1)^{30}+2\Big) = \frac{30(x+1)^{29}}{(x+1)^{30}+2}.$$

Knowing what you cannot do is sometimes the more important thing.

I wish there was some nice simple formula to break apart $\ln(A+B)$ but as far as I know $\ln(A+B) =$?, by this I simply mean that there is no simple formula to split it up. On the other hand we have used $\ln(AB) = \ln(A) + \ln(B)$ together with $\ln(A^c) = c\ln(A)$.

3.11.1 proof of power rule

Finally we return to the power rule. As we mentioned from the start the power rule $\frac{d}{dx}(x^n) = nx^{n-1}$ holds for all $n \in \mathbb{R}$. Now we have the tools to prove it.

Proof: Let $y = x^n$ and take the natural log to obtain $\ln(y) = \ln(x^n) = n\ln(x)$. Differentiate,

$$\frac{1}{y}\frac{dy}{dx} = \frac{n}{x} \;\Rightarrow\; \frac{dy}{dx} = \frac{ny}{x} = \frac{nx^n}{x} = nx^{n-1}.$$

This proof (in contrast to our earlier proof) works in the case that $n \notin \mathbb{N}$. Somehow these curious little logarithms have circumvented the whole binomial theorem. We conclude that for any $n \in \mathbb{R}$

$$\boxed{\frac{d}{dx}(x^n) = nx^{n-1}.}$$

Note, if $n < 0$ and $n \in \mathbb{Z}$ then $f(x) = x^n = \frac{1}{x^{-n}}$ is a function which has domain $\mathbb{R} - \{0\}$. The proof offered above fails for $x < 0$ since $\ln(x)$ is not real in such case. However, for cases of interest such as $n = -2, -3, \dots$ the argument can be modified. I leave this as an exercise for the reader. □

Another method to derive rules such as $\frac{d}{dx}(\frac{1}{x^n}) = \frac{-n}{x^{n+1}}$ is apply the product rule n-times for the reciprocal function for which we have already shown $\frac{d}{dx}(\frac{1}{x}) = \frac{-1}{x^2}$.

Example 3.11.7.

$$\frac{d}{dx}(\ \sqrt[3]{x}\ ) = \frac{d}{dx}(\ x^{1/3}\ ) = (1/3)x^{-2/3} = \frac{1}{3x^{2/3}}.$$

Example 3.11.8.

$$\frac{d}{dy}(y^{\pi+2}) = (\pi+2)y^{\pi+1} \approx 5.142y^{4.142}$$

3.12 summary of basic derivatives

I collect all the basic derivatives for future reference.

$f(x)$	$\frac{df}{dx}$	Comments about $f(x)$	Formulas I use
c	0	constant function	
x	1	line $y = x$ has slope 1	
x^2	$2x$		
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$		
x^n	nx^{n-1}	power rule	
e^x	e^x	the exponential	
5^x	$\ln(5)5^x$		
a^x	$\ln(a)a^x$	an exponential	
$\ln(x)$	$\frac{1}{x}$	the natural log	$\ln(e^x) = x$, $e^{\ln(x)} = x$
$\log x$	$\frac{1}{\ln(10)x}$	log base 10	
$\log_a(x)$	$\frac{1}{\ln(a)x}$	log base a	$\log_a(a^x) = x$, $a^{\log_a(x)} = x$
$\sin(x)$	$\cos(x)$		$\sin^2(x) + \cos^2(x) = 1$
$\cos(x)$	$-\sin(x)$		
$\tan(x)$	$\sec^2(x)$		$\tan^2(x) + 1 = \sec^2(x)$
$\sec(x)$	$\sec(x)\tan(x)$	reciprocal of cosine	$\sec(x) = 1/\cos(x)$
$\cot(x)$	$-\csc^2(x)$	reciprocal of tangent	$\cot(x) = \cos(x)/\sin(x)$
$\csc(x)$	$-\csc(x)\cot(x)$	reciprocal of sine	$\csc(x) = 1/\sin(x)$
$\sin^{-1}(x)$	$\frac{1}{\sqrt{1-x^2}}$	inverse sine	$\sin(\sin^{-1}(x)) = x$
$\cos^{-1}(x)$	$\frac{-1}{\sqrt{1-x^2}}$	inverse cosine	$\cos^{-1}(\cos(x)) = x$
$\tan^{-1}(x)$	$\frac{1}{x^2+1}$	inverse tangent	
$\sinh(x)$	$\cosh(x)$	hyperbolic sine	$\cosh(x) = \frac{1}{2}(e^x + e^{-x})$
$\cosh(x)$	$\sinh(x)$	hyperbolic cosine	$\sinh(x) = \frac{1}{2}(e^x - e^{-x})$
$\tanh(x)$	$sech^2(x)$	hyperbolic tangent	
$\sinh^{-1}(x)$	$\frac{1}{\sqrt{1+x^2}}$	inverse sinh	
$\cosh^{-1}(x)$	$\frac{1}{\sqrt{x^2-1}}$	inverse cosh	
$\tanh^{-1}(x)$	$\frac{1}{1-x^2}$	inverse tanh	

Finally, let us conclude this chapter with a list of useful rules of differentiation. These in conjunction with the basic derivatives we listed earlier in this section will allow us to differentiate almost anything you can imagine. (*this is quite a contrast to integration as we shall shortly discover*)

name of property	operator notation	prime notation
Linearity	$\frac{d}{dx}[f+g] = \frac{d}{dx}[f] + \frac{d}{dx}[g]$ $\frac{d}{dx}[cf] = c\frac{d}{dx}[f]$	$(f+g)' = f' + g'$ $(cf)' = cf'$
Product Rule	$\frac{d}{dx}[fg] = \frac{df}{dx}g + f\frac{dg}{dx}$	$(fg)' = f'g + fg'$
Quotient Rule	$\frac{d}{dx}\left[\frac{f}{g}\right] = \frac{\frac{df}{dx}g - f\frac{dg}{dx}}{g^2}$	$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$
Chain Rule	$\frac{d}{dx}[f \circ u] = \frac{df}{du}\frac{du}{dx}$	$(f \circ u)'(x) = f'(u(x))u'(x)$

Beyond these basic properties we have seen in this chapter that the technique of implicit differentiation helps extend these simple rules to cover the inverse functions. It all goes back to the definition logically speaking, but it is comforting to see that once we have established the derivatives of the basic functions and these properties we have little need of applying the definition directly. I would argue this is part of what separates modern (say the last 400 years) mathematics from ancient mathematics. We have no need to calculate limits by some exhaustive numerical method. Instead, for a wealth of examples, we can find tangents through what are essentially algebraic calculations. This is an amazing simplification. However, more recent times have shown computers can model problems which defy algebraic description. A student of mathematics would be wise to study computer aided solutions. Not so much for the purpose of gaining ease with homework, but rather to gain skills which many employers seek and need.

3.13 related rates

Given some algebraic relation that connects different dynamical quantities we can differentiate implicitly. This relates the rates of change for the various quantities involved. Such problems are called "related rates problems".

Example 3.13.1. *Suppose that we know the radius of a spherical hot air balloon is expanding at a rate of 1 meter per minute due to an inflating fan. At what rate is the volume increasing if the radius R is at 10 meters ? To begin we need to recall that the volume V is related to the radius R by the equation $V = \frac{4\pi}{3}R^3$ for the sphere. Then,*

$$\frac{dV}{dt} = \frac{d}{dt}\left(\frac{4\pi}{3}R^3\right) = 4\pi R^2 \frac{dR}{dt} = 4\pi(10m)^2 \frac{m}{min} \approx 1200 \frac{m^3}{min}.$$

In-Class Example 3.13.2. *Imagine a circular oil slick which grows uniformly as oil is added by the EPA. If the EPA adds oil at a rate such that 3 square meters is added every second then how quickly is the radius of the oil slick increasing when $r = 10m$?*

Solution:

Example 3.13.3. *Suppose you add water to a rectangular bathtub at a rate of 5 cubic feet per minute. If the dimensions of the tub are $5ft$ by $3ft$ then how quickly does the water rise?*

Solution: *We should define the variables; call the volume of water in the tub V and the area of the base A and the height of water in the tub h. Since the tub is rectangular we have $V = Ah$ where $A = 15ft^2$. We can relate the time-rate of change of V and h:*

$$\frac{dV}{dt} = A\frac{dh}{dt} \qquad \Rightarrow \qquad \frac{dh}{dt} = \frac{1}{A}\frac{dV}{dt}.$$

We were given $\frac{dV}{dt} = 5ft^3/min$ thus $\boxed{\frac{dh}{dt} = \frac{1}{3}\frac{ft}{min}.}$

In-Class Example 3.13.4. *Suppose you add water to a triangular water trough built such that it has equilateral triangular ends with side length 2ft and a length of 4ft. If the water is added at a rate of 5 cubic feet per minute then how quickly does the water level rise if the water is at a height of $1ft$ from the base? You may assume the trough is set-up on level ground such that the water level is parallel to the base of the trough.*

Solution:

Example 3.13.5. Problem: *If the sun travels across the sky over a period of 12 hrs and the distance to the sun is known to be 93 million miles then how fast is the sun going? Suppose that the earth is fixed and the sun is traveling in a circle at a constant rate. Also, for convenience you may neglect the size of the earth relative to your observation.*

Solution: *The equation relating arclenth to angle θ subtended is $s = R\theta$. The sun goes from $\theta = 0$ to $\theta = \pi$ in the course of the given $12hr$ day. Since the rate at which the sun travels is constant the instantaneous rate of change matches the average rate of change:*

$$\frac{d\theta}{dt} = \frac{\Delta\theta}{\Delta t} = \frac{\pi}{12hrs}.$$

Differentiating the arclength relation we find $\frac{ds}{dt} = \frac{d}{dt}(R\theta) = R\frac{d\theta}{dt}$. We were given $R = 93,000,000$ miles thus $\frac{ds}{dt} = (93,000,000miles)\frac{\pi}{12hrs} \approx \boxed{24,300,000mph.}$

We know that the perception of the sun travelling across the sky is actually due to the earth spinning. The speed at which the earth rotates relative to its center is roughly $v = \frac{2\pi(4000miles}{24hrs} \approx 1050mph$ (at the equator). The circumference at the equator is about 25,000 miles. In contrast, the land at the North of South poles rotates at a much slower tangential speed. For this reason the Earth is actually an *oblate spheroid* [7] because the equator is spun further away from the center due to the centripetal force. If you consider the last example you can see why it was easy to give up on the idea of the earth being at the center and everything else rotating around us. Stars further away than the sun would have to go even faster. You might wonder how it can be determined the sun is 93 million miles away. The answer is trigonometry. I'll leave it at that for here.

[7] which is basically just a squished sphere

In-Class Example 3.13.6. *If a $10ft$ ladder slides down a wall without slipping such that the top of the ladder slides down the wall at $3ft/s$ then how fast is the base of the ladder sliding away from the wall when the ladder is $4ft$ from the wall ?*

Solution:

Example 3.13.7. Problem: *If a $10ft$ ladder slides down a parabolic wall (with equation $y = 6 - x^2$) without slipping such that the top of the ladder slides down the wall at $1ft/s$ (dy/dt) then how fast is the base of the ladder sliding away from the wall when the ladder is at $x = 2ft$?*

Solution: *the problem statement tells us $y = 6 - x^2$ thus $\frac{dy}{dt} = -2x\frac{dx}{dt}$ thus we may solve for dx/dt: (I omit units, we agree to work in ft and s)*

$$\frac{dx}{dt} = \frac{-1}{2x}\frac{dy}{dt} = \frac{-1}{4}(-1) = 0.25$$

Thus, bringing back the units, $\boxed{\dfrac{dx}{dt} = 0.25ft/s.}$

Remark 3.13.8.

Units are important however, writing explicit units is not always the best approach. A common technique is to state from the outset which units you intend to use then you may add them back at the end of the calculation. The answer should have units. To be honest, the last example is not even well-posed if you are a stickler for units. I cannot write such an equation as $y = 6 - x^2$ unless I assume that x and y are dimensionless. To be careful I'd need to write something like $y = 6ft - (\frac{1}{ft})x^2$ if both x, y are written in terms of ft. But, that equation is uglier than $y = 6 - x^2$ so we prefer to write less and just be careful to put given numerical data into our set of chosen units. We chose $[x] = [y] = ft$ and $[t] = s$ in the last example (I use $[f]$ to denote the customary units of f).

Example 3.13.9. Problem: *Imagine two cars begin traveling from a point which we label as the origin. If car* **A** *travels at 30mph along the direction* $\theta_A = \pi/6$ *and if car* **B** *travels at 40mph along the direction* $\theta_B = 5\pi/4$ *then how quickly is the distance* s *between them increasing at time* t*? What is* ds/dt *at* $t = 1hr$ *?*

Solution: *we should imagine a triangle at time* t*. One vertex is at the origin and the other two are at cars* **A** *and* **B** *respective. The angle at the origin is calculated to be* $\beta = \theta_B - \theta_A = 5\pi/4 - \pi/6 = (30-4)\pi/24 = 26\pi/24$*. Notice that as the cars travel along the straight lines the triangle gets bigger and the angles at* **A** *and* **B** *are changing whereas the angle* β *is independent of time. We can write the law of cosines for the angle* β*, note the opposite side is the distance between* **A** *and* **B** *which we labeled* s*:*

$$s^2 = s_A^2 + s_B^2 - 2s_A s_B \cos(\beta).$$

Where s_A, s_B *are the distances from the origin to cars* **A** *and* **B** *respective. Since the cars travel at constant speed we can relate the distance to the time by the equations* $s_A = 30t$ *and* $s_B = 40t$*. Thus,*

$$s^2 = 900t^2 + 1600t^2 - 2400t^2 \cos(\beta).$$

Differentiate with respect to time,

$$2s\frac{ds}{dt} = 1800t + 3200t - 4800t \cos(\beta).$$

Note $s = \sqrt{900t^2 + 1600t^2 - 2400t^2 \cos(\beta)}$ *hence*

$$\boxed{\frac{ds}{dt} = \frac{1}{2\sqrt{900t^2 + 1600t^2 - 2400t^2 \cos(26\pi/24)}}\left[1800t + 3200t - 4800t \cos(26\pi/24)\right].}$$

To calculate $s'(1)$ *we need only evaluate the expression above at* $t = 1$*; hence* $\boxed{s'(1) = 69.41mph}$

In-Class Example 3.13.10. *Imagine two cars traveling from a point which we label as the origin. If car* **A** *travels at 30mph due east at beginning at noon and if car* **B** *travels at 40mph due North beginning at 1:00 PM then how quickly is the distance* s *between them increasing at 2:00 PM? ?*

Solution:

There are many examples of related rates. It is likely your homework has additional types of problems. Keep in mind the general advice:

- identify variables to describe the problem,
- assess what is given and what you are trying to find,
- draw a picture of the generic situation at time t,
- write a model equation which relates the variables,
- differentiate implicitly the model equation and then plug in data given as appropriate and solve for the requested rate.

Sometimes it is possible to write on of the variables as an **explicit function** of time, othertimes we are not given enough information so we must find general relations of which we only know data at one given instant. The beauty of the method is even when we have incomplete data about the entire time evolution of a variable there are still strict constraints on the relation of the rates following from the model equation.

Chapter 4

derivatives and linear approximations

Linearization of a function is the process of approximating a function by a line near some point. The tangent line is the graph of the linearization. The differential is closely connected with the linearization. In short, the difference between the concepts is as follows:

1. the linearization is an approximates the function near a given point.

2. the differential approximates the change in the function at a given point.

We examine how to apply linearizations to approximate nonlinear functions. We also consider how the differential is useful in the analysis of error propagation. Finally, we use derivatives in the formulation of Newton's method. This iterative method allows us to use the power of calculus to find approximate solutions to algebraic or even transcendental equations.

4.1 linearizations

We have already found the linearization of a function a number of times. The idea is to replace the function by its tangent line at some point. This usually[1] provides a good approximation if we are near to the point. The linearization of a function f at a point $a \in dom(f)$ is denoted by L_f^a or simply L_f in this course,

$$\boxed{L_f^a(x) \equiv f(a) + f'(a)(x-a)}$$

The graph of L_f^a is the tangent line to $y = f(x)$ at $(a, f(a))$.

Example 4.1.1. *Suppose the singularity has occurred and the robot holocaust has cast doubt on the service of all machines. You need to calculate a squareroot but you can't trust your calculator. What to do? Let $f(x) = \sqrt{x}$ and use the linearization. Take the number you wish to find the root for and pick the closest easy root you can find center the linearization. Then the linearization of the number will give a close estimate of the root you wish to find. For example, $\sqrt{4.01}$. Notice,*

[1]we discuss limitations of the tangent line approximation at the conclusion of this section

$4.01 = 4 + 0.01$ and we know $\sqrt{4} = 2$ thus we use $a = 4$ as the center of the approximation. Calculate that $f'(x) = \frac{1}{2\sqrt{x}}$ hence,

$$f(x) \approx L_f^4(x) = f(4) + f'(4)(x-4) = 2 + \frac{1}{4}(x-4).$$

Therefore,

$$\sqrt{4.01} \approx 2 + \frac{4.01-4}{4} = 2 + 0.0025 = 2.0025.$$

Since Wolfram-alpha is still free and fairly benevolent I believe that in truth

$$\sqrt{4.01} = 2.00249843945007857276972121483226054214864513129159...$$

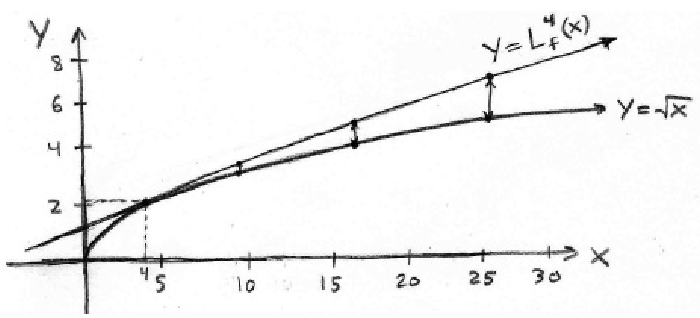
As you can easily see we did very well considering the crudeness of our method (in fact the error is only about 0.0001%). Is a line a parabola? Certainly not. But that is the heart of what I just did. I said you can replace a curve with a line locally and get good approximations. But, what is "local" how far does this linearization give "good" results? Just for perspective I list a few less accurate results from this linearization:

$$\sqrt{9} \approx 2 + \frac{1}{4}(9-4) = 3.25 \text{ (8.33\% error)}$$

$$\sqrt{16} \approx 2 + \frac{1}{4}(16-4) = 5 \text{ (25\% error)}$$

$$\sqrt{25} \approx 2 + \frac{1}{4}(25-4) = 7.25 \text{ (45\% error)}$$

Here's a picture of what just happened.



Many authors would replace x with $4+h$ and use $g(h) = \sqrt{4+h}$ in which case the center of the approximation is naturally taken to be zero thus $\sqrt{4+h} \approx 2 + \frac{h}{4}$. It's just a matter of notation. In the same sense your text has more to say about the "differential", however if you examine the mathematics closely you'll learn that the differential and the linearization are being used to accomplish the same goal. I will discuss both just to be safe. In a nutshell, the differential approximates the *change* in the function near some base point whereas the linearization approximates the function itself near the base point. By "base point" I simply mean the point at which the approximation is based. In the last example we had base point $a = 4$.

In-Class Example 4.1.2. *Find the approximate value of* $\ln(3)$ *by calculating* $L_f^a(3)$ *for of* $f(x) = \ln x$ *with base point* $a = e \cong 2.71\ldots.$ *Note,* $\ln e = 1$.

4.2 differentials and error

The change in the function between a and $a+h$ is denoted $\Delta f = f(a+h) - f(a)$ and if $y = f(x)$ then we may likewise state $\Delta y = f(a+h) - f(a)$. Likewise, the change in x with respect to these two points is $\Delta x = a + h - a = h$. The linearization based at a for f is given by $L_f^a(x) = f(a) + f'(a)(x-a)$. If we substitute $x = a + h$ into the formula for the linearization we find $L_f^a(a+h) = f(a) + f'(a)h$ which gives that $L_f^a(a+h) - f(a) = f'(a)h$. If $h \approx 0$ then we expect $L_f^a(a+h) \approx f(a+h)$ thus it follows that

$$\Delta f \approx f'(a)h.$$

The notation is deceptively simple here: $\Delta f = \Delta y$, $f'(a) = \frac{df}{dx}(a)$ and $h = \Delta x$. This gives:

$$\Delta y \approx \frac{df}{dx}(a)\Delta x.$$

Definition 4.2.1.

Suppose $f : U \subseteq \mathbb{R} \to \mathbb{R}$ then if f has a derivative at a then it also has a **differential** $df_a : \mathbb{R} \to \mathbb{R}$ at a which is a function defined by $df_a(h) = hf'(a)$.

Notice that the derivative at a point $(f'(a))$ is a number whereas the differential at a point (df_a) is a linear function. The linearization (L_f^a) of the function at $(a, f(a))$ is actually an *affine function* which just means it has a graph which is a line with a possibly nonzero y-intercept.

Example 4.2.2. *Estimate the uncertainty in the volume of a cubical box if you measure the length of the side to be $20in \pm 0.2in$. Let x denote the length of the side and V the volume of the box then*

$$V = x^3$$

Thus $\frac{dV}{dx} = 3x^2$. We find,

$$\Delta V \approx \frac{dV}{dx}(a)\Delta x.$$

We are given $a = 20in$ and $\Delta x = \pm 0.2in$ thus,

$$\Delta V \approx 3(20)^2(0.2)in^3 = 240in^3.$$

Thus the uncertainty in the volume of the cubical box is approximately $\boxed{\pm 240in^3}$

In-Class Example 4.2.3. *Estimate the uncertainty in the area enclosed by a circular fence if you measure the radius of the area to be $3miles \pm 50ft$. Note: one mile has 5280 ft.*

It should be mentioned that the total and correct analysis of error propagation is more involved that this section indicates. If you want to see more you might look at *Data Reduction and Error Analysis for the Physical Sciences* by Bevington and Robinson.

4.3 Newton's method

In this section we use linearizations to find roots of equations. The idea is actually very simple: we wish to solve $f(x) = 0$ for a given differentiable function f

(1.) guess a solution x_o and calculate $f(x_o)$ and if it is close enough to zero then stop.

(2.) construct $L_o(x) = f(x_o) + f'(x_o)(x - x_o)$ and solve $L_o(x_1) = 0$ to find solution

$$x_1 = x_o - f(x_o)/f'(x_o).$$

(3.) calculate $f(x_1)$ and if it is close enough to zero then stop.

(4.) construct $L_1(x) = f(x_1) + f'(x_1)(x - x_1)$ and solve $L_1(x_2) = 0$ to find solution

$$x_2 = x_1 - f(x_1)/f'(x_1).$$

(5.) calculate $f(x_2)$ and if it is close enough to zero then stop.

(6.) construct $L_2(x) = f(x_2) + f'(x_2)(x - x_2)$ and solve $L_1(x_3) = 0$ to find solution

$$x_3 = x_2 - f(x_2)/f'(x_2).$$

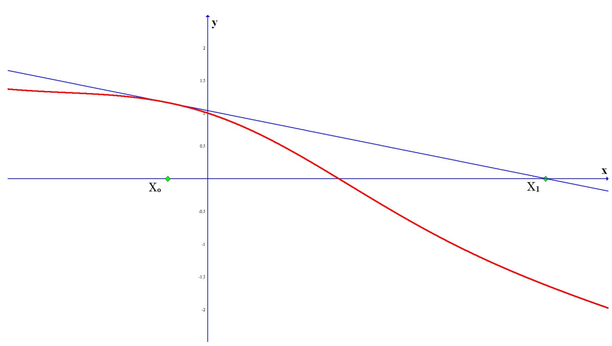
(7.) calculate $f(x_3)$ and so forth and so on until you get close enough to consider it a solution for the purposes of your application.

To summarize: we wish to solve $f(x) = 0$ then we guess x_o to begin then calculate iteratively by the rule

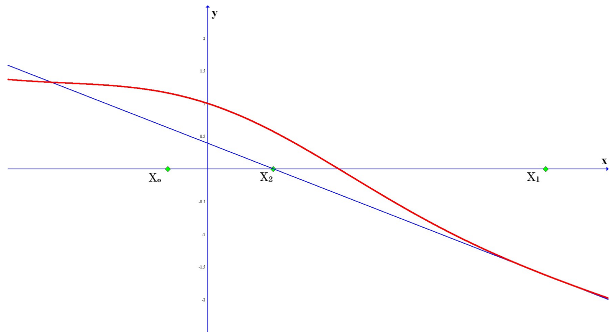
$$\boxed{x_{n+1} = x_n - f(x_n)/f'(x_n)}$$

until $|f(x_n)| < \varepsilon$ where ε is an upper bound on the error you allow for the approximate solution.

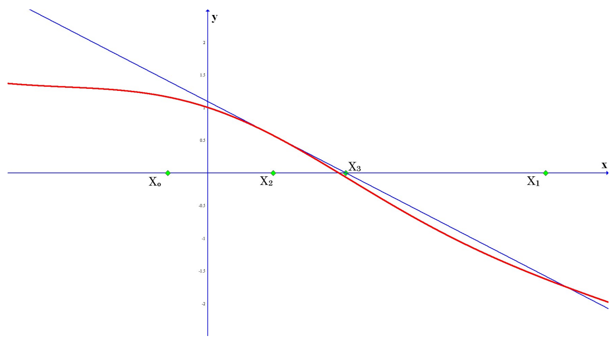
Example 4.3.1. *Let's see how to solve the equation $e^{-x^2} = x$ to within ± 0.01. First construct $f(x) = e^{-x^2} - x$ and note that the problem becomes solving $f(x) = 0$. Calculate $f'(x) = -2xe^{-x^2} - 1$. To begin we guess $x_o = -0.2$. Note $f(-0.2) \approxeq 1.16$. Calculate $x_1 = x_o - f(x_o)/f'(x_o) = 1.69$. I have pictured the initial guess x_o as well as the first iterate x_1 with green diamonds on the $x-axis$:*



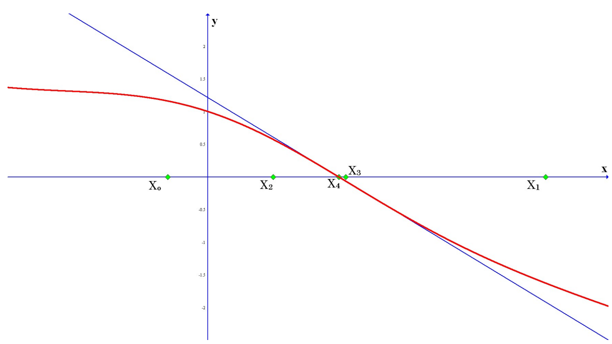
You can see that x_1 is the x-intercept of the tangent line from x_o. Next, we can calculate $x_2 = x_1 - f(x_1)/f'(x_1) = 0.324$.



You can see that x_2 is the x-intercept of the tangent line from x_1. Next, we can calculate $x_3 = x_2 - f(x_2)/f'(x_2) = 0.686$.



You can see that x_3 is the x-intercept of the tangent line from x_2. Next, we can calculate $x_4 = x_3 - f(x_3)/f'(x_3) = 0.653$.



At this point the tangent line so closely follows the function it is difficult to see where the tangent line based at $(x_3, f(x_3))$ crosses the x-axis. We calculate

$$f(0.653) = e^{-0.653^2} - 0.653 = -0.00015.$$

Therefore, to a good approximation, the solution of $e^{-x^2} = x$ is $x = 0.653$.

I included the pictures in the preceding example to emphasize the idea of the method. In practice the graphs are not necessary for the calculation. However, looking at a graph is a good method to select the initial guess of x_o.

Example 4.3.2. *Calculate $\sqrt[3]{20}$. Use your imagination, if $x = \sqrt[3]{20}$ then $x^3 = 20$. We need to solve the equation $x^3 - 20 = 0$ in other words, find the zero of $f(x) = x^3 - 20$. We'll use Newton's method with an initial guess of $x_o = 2.5$ since we know that our answer must be somewhere between 2 and 3 since $2^3 = 8$ and $3^3 = 27$. Note $f'(x) = 3x^2$.*

$$x_1 = x_o - f(x_o)/f'(x_o) = 2.5 + 4.375/18.75 = 2.733$$

$$x_2 = x_1 - f(x_1)/f'(x_1) = 2.733 - 0.4136/22.41 = 2.715$$

$$x_3 = x_2 - f(x_2)/f'(x_2) = 2.715 - 0.013/22.11 = 2.714$$

We calculate that $2.714^3 = 19.991$ thus $\sqrt[3]{20} \approx 2.714$.

What if you wanted to calculate $\log_3(7)$ via Newton's method? I leave this as an exercise for the reader. The ideas presented in this section are used by calculators with great success. There are examples for which Newton's method fails to find a root, but it's not hard to modify the naive algorithm in this section to capture most roots. We've seen in our examples that even in less than 10 iterations the method zoomed in on the root. How quickly the method converges to the answer is important for applications because it determines the number of computer operations we will have to perform to execute the method. There are a number of useful estimates on the error of a particular iterate however they are beyond the scope of this course. You might read pages 160-165 of Edwards *Advanced Calculus* if you're interested in the pure mathematics of the topic. The goal of Edwards section is to prove the following theorem:

Proposition 4.3.3. *inverse function theorem for one-variable function.*

> Let f be a continuously differentiable functions on a nbhd of a then if $f'(a) \neq 0$ there exists a $\delta > 0$ such that f is invertible when restricted to $(a-\delta, a+\delta)$. In other words, if $f'(a) \neq 0$ and $f'(x)$ exists for x near a then f is locally invertible near a.

Proof Sketch: since $f'(a) \neq 0$ and f' is continuous it follows $f'(x) \neq 0$ on some nbhd. of a. Thus $f'(x)$ is either positive or negative on this nbhd and thus f is strictly monotonic and may therefore by inverted. □

The proof in Edwards is fascinating and **constructive**. He shows how to find a sequence of *functions* which converges to the inverse function. This means he shows how to construct an inverse function even in cases where you cannot implement the precalculus algorithm to find the inverse [2]. Incidentally, I don't mean to indicate that this idea is unique to Edward's text. These ideas are older and can be found in dozens, if not hundreds, of modern texts on numerical methods.

[2]you remember, $y = f(x)$, solve for x then say $x = f^{-1}(y)$. Don't switch x and y, it's better to use different letters for the domain and range of the function since they may well have different physical interpretations and/or be different sets. We can think of functions of y or functions of x. We are not slaves to notation!

Chapter 5

geometry and differential calculus

In this modern age it is tempting to neglect a careful study of graphing since we have so much technological assistance. However, by doing such we would rob ourselves of basic geometric intuition. In my view there is no substitute for seeing the nuts and bolts of calculus and their application to graphing. Moreover, the application of this analysis to word problems answers many nontrivial questions. Given a mathematical model we might wish to know which values of the variable make it the fastest, tallest, shortest, coolest, cheapest, etc... these sort of questions are easily answered by the analysis in this chapter.

In Chapter 4 we learned how to differentiate. In Chapter 5 we learned that the basic interpretation of the derivative at a point is as a linear approximation. In this chapter we learn what a derivative as a function means. We also analyze the geometric significance of higher derivatives. To complete the story of graphing we analyze limits at $\pm\infty$. We also apply such limits to analyze the asymptotic behavior of a function thus generalizing the idea of a horizontal asymptote. The asymptotic behavior of a model is sometimes the most interesting case. l'Hopital's rule is introduced and justified. Finally, breaking from some calculus-orthodoxy I discuss Taylor's Theorem[1] with Lagrange's form of the remainder. It is my opinion that the power of the theorem warrants some discussion at this time. Taylor's theorem elucidates and expands the second derivative test. Moreover, the idea of polynomial approximation is a very important idea to many applications. I show how polynomial approximations play a special role in physics.

5.1 graphing with derivatives

We would like to develop a strategy to locate where a given function takes its largest positive or negative values. In an application this tells us the boundaries of what is possible for a given model. For example, the motion of a spring oscillates between two positions. In other words, we can *bound* the motion between those two positions. In contrast, we might study a bridge over which a wind

[1] note we do not discuss series at this juncture, the totality of that topic waits until calculus II, I simply include some discussion here in the interest of deeper geometric insight. As a side consequence I also hope this inclusion strengthens the student for calculus II's travails

blows with a certain frequency. If the frequency of the wind matches the resonant frequency of the structure then the oscillation or waving motion of the bridge could build without bound. In that case a good mathematical model[2] of the bridge would reveal motion which is *unbounded*. The idea of bounded motion is closely connected with the following:

Definition 5.1.1. *absolute extrema.*

- f has an **absolute maximum** of $f(c)$ at c if $f(c) \geq f(x)$ for all $x \in dom(f)$.
- f has an **absolute minimum** of $f(c)$ at c if $f(c) \leq f(x)$ for all $x \in dom(f)$.
- absolute maximum and minimum values are called the **global extrema** of f.

In-Class Example 5.1.2. *Illustrate the existence or lack thereof of global extrema for several example functions.*

[2] a good mathematical model is the sort which anticipates these sort of problems before they occur

Suppose we are given a model plus some additional condition so that we know the model must have variables whose values are near some given data point. In a case such as that it is interesting to know what the largest positive or negative values the function takes near the given data point. Mathematically this is encapsulated by the idea of a local extreme value:

Definition 5.1.3. *local extrema.*

- A function f has a **local maximum** of $f(c)$ at c if there exists a connected set J with $c \in J$ and $J \subseteq dom(f)$ such that $f(c) \geq f(x)$ for all $x \in J$.
- A function f has a **local minimum** of $f(c)$ at c if there exists a connected set J with $c \in J$ and $J \subseteq dom(f)$ such that $f(c) \leq f(x)$ for all $x \in J$.
- If $f(c)$ is either a local maximum or a local minimum then we say $f(c)$ is a **local extrema** at c.

In-Class Example 5.1.4. *illustrate the definition with several graphs.*

The following theorem is at the heart of most everything that follows in this chapter.

Proposition 5.1.5. *Extreme value theorem.*

Suppose that f is a function which is continuous on $[a, b]$ then f attains its absolute maximum $f(c)$ on $[a, b]$ and its absolute minimum $f(d)$ on $[a, b]$ for some $c, d \in [a, b]$.

It's easy to see why the requirement of continuity is essential. If the function had a vertical asymptote on $[a, b]$ then the function gets arbitrarily large or negative so there is no biggest or most negative value the function takes on the closed interval. Of course, if we had a vertical asymptote then the function is not continuous at the asymptote. The proof of this theorem is technical and beyond the scope of this course. See Apostol pages 150-151 for a nice proof.

Notice the extreme value theorem does not really tell us how to find extrema. It merely states they exist somewhere if the given function is continuous. Naturally we would like a way to locate such points. Given our earlier work with tangent lines it would seem intuitively natural to suppose those extrema should be found at points where there is either a horizontal tangent or a jump or kink in the graph. Those graphical features will either make the derivative at the point to be zero or undefined. We wish to prove this intuition valid. Begin by defining the points of interest:

Definition 5.1.6. *critical numbers.*

We say $c \in \mathbb{R}$ is a **critical number** of a function f if either $f'(c) = 0$ or $f'(c)$ does not exist. If $c \in dom(f)$ is a critical number then $(c, f(c))$ is a **critical point** of f.

In-Class Example 5.1.7. *Find critical numbers for* $f(x) = e^{-x^3}$

Notice that a critical number need not be in the domain of a given function. For example, $f(x) = 1/x$ has $f'(x) = -1/x^2$ and thus $c = 0$ is a critical numbers as $f'(0)$ does not exist in $\mathbb{R}$. Clearly $0 \notin dom(f)$ either. It is usually the case that a vertical asymptote of the function will likewise be a vertical asymptote of the derivative function.

Proposition 5.1.8. *Fermat's theorem.*

If f has a local extreme value of $f(c)$ and $f'(c)$ exists then $f'(c) = 0$.

Proof: suppose $f(c)$ is a local maximum. Then there exists $\delta_1 > 0$ such that $f(c+h) \leq f(c)$ for all $h \in B_{\delta_1}(0)$. Furthermore, since $f'(c) \in \mathbb{R}$ we have $\lim_{h\to 0} \frac{f(c+h)-f(c)}{h} = f'(c) \in \mathbb{R}$. If $h > 0$ and $h \in B_{\delta_1}(0)$ then $f(c+h) - f(c) \leq 0$ hence $\frac{f(c+h)-f(c)}{h} \leq 0$. Using the squeeze theorem we find $f'(c) = \lim_{h\to 0^+} \frac{f(c+h)-f(c)}{h} \leq \lim_{h\to 0}(0) = 0$. Likewise, if $h < 0$ and $h \in B_{\delta_1}(0)$ then $f(c+h) - f(c) \leq 0$ hence $\frac{f(c+h)-f(c)}{h} \geq 0$. Hence, by the squeeze theorem, $f'(c) = \lim_{h\to 0^-} \frac{f(c+h)-f(c)}{h} \geq \lim_{h\to 0}(0) = 0$. Consequently, $f'(c) \leq 0$ and $f'(c) \geq 0$ therefore $f'(c) = 0$. The proof in the case that $f(c)$ is a local minimum is similar. $\square$

Remember, if $f'(c)$ does not exist then c is a critical point by definition. Therefore, if $f(c)$ is a local extrema then c must be a critical point for one of two general reasons:

(1.) $f'(c)$ exists so Fermat's theorem proves $f'(c) = 0$ so c is a critical point.

(2.) $f'(c)$ does not exist so by definition c is a critical point.

Sometimes Fermat's Theorem is simply stated as "local extrema happen at critical points". That said, you canno reverse the sentence. In fact, not every critical point gives a local extreme.

In-Class Example 5.1.9. *Discuss* $f(x) = x^3$ *and* $g(x) = \sqrt[3]{x}$.

Proposition 5.1.10. *Rolle's theorem.*

Suppose that f is a function such that

1. f is continuous on $[a, b]$,
2. f is differentiable on (a, b),
3. $f(a) = f(b)$.

Then there exists $c \in (a, b)$ such that $f'(c) = 0$.

Proof: If f is a constant function then $f'(x) = 0$ for all $x \in (a, b)$ so Rolle's Theorem is true. Otherwise, suppose f is nonconstant and use the Extreme Value Theorem and Fermat's Theorem to prove there exists $c \in (a, b)$ such that $f'(c) = 0$.

Let's think about Rolle's theorem as it applies to the physics of projectile motion. If the height of a cat is $y(t)$ and it represents a cat thrown up into the air for 3 seconds meaning $y(0) = y(3) = 0$. Then $v = dy/dt$ must be zero at some point during the flight of the cat. What goes up must come down, and before it comes down it has to stop going up.

In-Class Example 5.1.11. *verify Rolle's theorem for $f(x) = \sin x$ on $0 \leq x \leq \pi$.*

Proposition 5.1.12. *Mean Value Theorem (MVT).*

Suppose that f is a function such that

1. f is continuous on $[a, b]$,
2. f is differentiable on (a, b),

Then there exists $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. That is, there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.

Proof: Suppose f is continuous on $[a, b]$ and differentiable on (a, b). Define function h by the difference of the secant line from $(a, f(a)$ to $(b, f(b))$ and f,

$$h(x) = f(x) - s(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Observe that $h(a) = h(b) = 0$ and h is clearly continuous on $[a, b]$ because f is continuous and besides that the function is constructed from a sum of a polynomial with f. Additionally, it is clear that h is differentiable on (a, b) since polynomials are differentiable everywhere and f was assumed to be differentiable on (a, b). Thus Rolle's Theorem applies to h so there exists a $c \in (a, b)$ such that $h'(c) = 0$ which yields

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \quad \Longrightarrow \quad \boxed{f'(c) = \frac{f(b) - f(a)}{b - a}}. \qquad \square$$

Physical Significance of the Mean Value Theorem: The term "mean" could be changed to "average". Apply the MVT to the case that the independent variable is time t and the dependent variable is position y and we get the simple observation that the average velocity over some time interval is equal to the instantaneous velocity at some time during that interval of time. For example, if you go 60 miles in one hour then your average velocity is clearly 60mph. The MVT tells us that some time during that hour your instantaneous velocity was also 60mph.

In-Class Example 5.1.13. *Verify MVT for $f(x) = x^3 + 1$ on $[-1, 2]$.*

Proposition 5.1.14. *sign of the derivative function f' indicates strict increase or decrease of f.*

Suppose that f is a function and J is a connected subset of $dom(f)$

1. if $f'(x) > 0$ for all $x \in J$ then f is strictly increasing on J
2. if $f'(x) < 0$ for all $x \in J$ then f is strictly decreasing on J.

Proof: suppose $f'(x) > 0$ for all $x \in J$. Let $[a, b] \subseteq J$ and note f is continuous on $[a, b]$ since it is given to be differentiable on a superset of $[a, b]$. The MVT applied to f with respect to $[a, b]$ implies there exists $c \in [a, b]$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Notice that $f(b) - f(a) = (b - a)f'(c)$ but $b - a > 0$ and $f'(c) > 0$ hence $f(b) - f(a) > 0$. Therefore, for each pair $a, b \in J$ with $a < b$ we find $f(a) < f(b)$ which means f is strictly increasing on J. Likewise, if $f'(c) < 0$ then almost the same argument applies to show $a < b$ implies $f(a) > f(b)$. $\square$

Theorem 5.1.15. *derivative zero implies constant function.*

If $f'(x) = 0$ for each $x \in (a, b)$ then f is a constant function on (a, b).

Proof: apply the Mean Value Theorem. We know we can because the derivative exists at each point on the interval and this implies the function is continuous on the open interval, so it is continuous on any closed subinterval of (a,b). Let us denote this closed subinterval by $J = [a_o, b_o] \subset (a, b)$. We have to apply the Mean Value Theorem to $J = [a_o, b_o]$ because we do not know for certain that the function is continuous on the endpoints. We find,

$$0 = \frac{f(b_o) - f(a_o)}{b_o - a_o} \quad \Longrightarrow \quad f(b_o) = f(a_o)$$

But this is for an arbitrary closed subinterval hence the function is constant on (a,b). $\square$

Caution: we cannot say the function is constant beyond the interval (a, b). It could do many different things beyond the interval in consideration. Piecewise continuous functions are such examples, they can be constant on the pieces yet at the points of discontinuity the function can jump from one constant to another.

In-Class Example 5.1.16. *Let* $f(x) = \dfrac{\sqrt{x^2}}{x}$ *show* $f'(x) = 0$ *for all* $x \neq 0$*. Is* f *constant ?*

Theorem 5.1.17. *if derivatives of two functions agree then the functions have same shaped graph.*

If $f'(x) = g'(x)$ for each $x \in (a, b)$ then $f(x) = g(x) + c$ for some constant $c \in \mathbb{R}$.

Proof: Apply Proposition 5.1.15 to $h(x) = f(x) - g(x)$. Notice $h'(x) = f'(x) - g'(x) = 0$ hence $h(x) = c$ and thus $c = f(x) - g(x)$. The proposition follows. $\square$

Notice that the assumption is that they are equal on an open interval. If we had that the derivatives of two functions were equal over some set which consisted of disconnected pieces then we could apply Theorem 5.1.17 to each piece separately then we would need to check that those constants from different components matched up. (for example if $\frac{df}{dx} = \frac{dg}{dx}$ on $(0, 1) \cup (2, 3)$ then we could have that $f(x) = g(x) + 1$ on (0,1) whereas $f(x) = g(x) + 2$ on $(2, 3)$).

Proposition 5.1.18. *sign-charts for derivatives reveal increase and decrease of function.*

> If f has finitely many critical numbers and f then the intervals of increase and decrease for f can be determined through the use of a sign-chart for $f'(x)$. In particular, one draws a number line with all critical points then labels either $(+)$ or $(-)$ on each subinterval based on a test point for $f'(x)$ in the subinterval. The function is either increasing or decreasing on each subinterval bounded by the critical points.

Proof: since there are finitely many critical points we may partition the real line into a finite number of disjoint open intervals which are joined at critical numbers. Then we apply Proposition 5.1.14 to each open interval to determine strict increase or decrease. The sign-chart is simply a number line indicating this analysis in a nice organized fashion. See the next subsection for examples. □

The sign-chart also applies to the case of countably many critical points which are separated by finite open intervals. For example $f(x) = \cos(\pi x)$ has $f'(x) = -\pi \sin(\pi x)$ and we have infinitely many critical numbers of the form $c = n$ for $n \in \mathbb{Z}$. The concept of the sign-chart does just fine for an example like $f(x) = \cos(\pi x)$. However, the sign-chart is not helpful for functions which have dense accumulations of critical points in some nbhd. (see Example 3.9.7 for this bad behavior).

5.1.1 first derivative test

The following theorem naturally follows from the sign-test theorem.

Theorem 5.1.19. *sign-charts for derivatives reveal increase and decrease of function.*

> Suppose f is continuous on an open interval containing a critical number c then
>
> 1. if $f'(x)$ changes signs from positive to negative at c then $f(c)$ is a local maximum.
> 2. if $f'(x)$ changes signs from negative to positive at c then $f(c)$ is a local minimum.
> 3. if $f'(x)$ does not change signs at c then $f(c)$ is not a local extrema.

In each of the examples that follow in this section we aim to use calculus to analyze the graph of the function. In particular, we are interesting in locating any local extrema and the intervals of increase and decrease for the given functions.

Example 5.1.20. *Let $f(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 6x$. Find all critical numbers and classify the critical points as local maximums, minimums or neither. Observe,*

$$f'(x) = x^2 + x - 6 = (x-2)(x+3).$$

We have two critical numbers, $c = 2$ and $c = -3$. Therefore, we set-up the sign-chart as follows:

Then we test a point somewhere in the interior of each region,

$$f'(-4) = (-4-2)(-2+3) = 8 > 0$$
$$f'(0) = (-2)(3) = -6 < 0$$
$$f'(3) = (3-2)(3+3) = 6 > 0$$

Hence the completed sign-chart,

By the First Derivative Test we conclude,

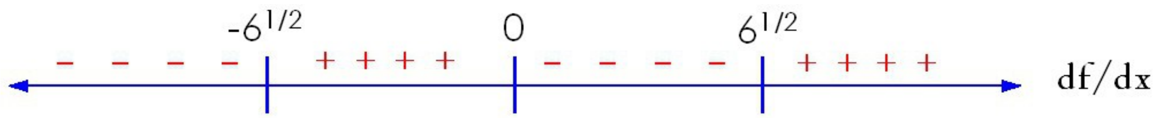
1. $f(-3) = -27/3 + 9/2 - 6(-3) = 27/2$ *is a local maximum.*

2. $f(2) = 8/3 + 4/2 - 6(2) = -22/3$ *is a local minimum.*

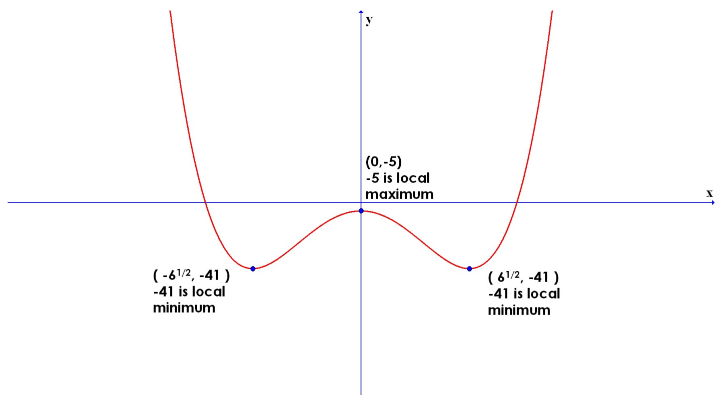
Example 5.1.21. *Let* $f(x) = e^x + x$. *Note that* $f'(x) = e^x + 1$. *This function has no critical points since the equation* $e^x + 1 = 0$ *has no solutions. It follows that* $y = e^x + 1$ *has no local extrema. However, we can deduce that* $f(x)$ *is increasing on* $\mathbb{R}$ *since* $f'(x) = e^x + 1 \geq 2$ *for all* $x \in \mathbb{R}$.

In-Class Example 5.1.22. *Let* $f(x) = x^4 - 4x^3 - 12x^2 + 32x$. *Find intervals of increase and decrease. Classify each critical point. Sketch graph. Hint:* $f'(1) = 0$.

Example 5.1.23. *Let $f(x) = x^4 - 12x^2 - 5$. Calculate $f'(x) = 4x^3 - 24x = 4x(x^2 - 6) = 3x(x + \sqrt{6})(x - \sqrt{6})$ hence we find critical numbers $c = 0, \pm\sqrt{6}$. In invite the reader to confirm that the test points $-3, -1, 1, 2$ reside between the critical points and $f'(-3) < 0, f'(-1) > 0, f'(1) < 0$ and $f'(3) > 0$ therefore the sign-chart for the derivative function is as follows:*



We identify that f is increasing on $(-\sqrt{6}, 0) \cup (\sqrt{6}, \infty)$ and it f is decreasing on $(-\infty, -\sqrt{6}) \cup (0, \sqrt{6})$. By the first derivative test we observe that $f(-\sqrt{6}) = 36 - 72 - 5 = -41$ and $f(\sqrt{6}) = 36 - 72 - 5 = -41$ are local minima whereas $f(0) = -5$ is a local maximum. The graph can be deduced from these facts.

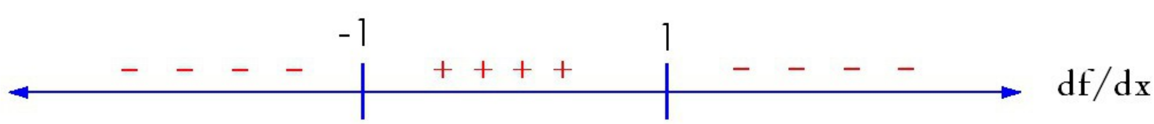


Notice I did not even need to find the zeros of the graph to make a good sketch of the curve.

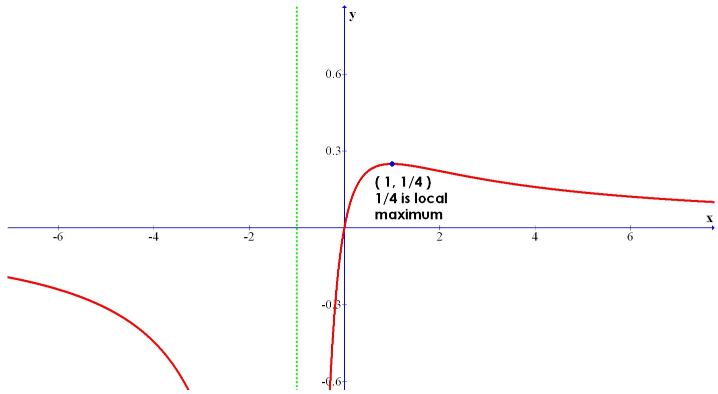
Example 5.1.24. *Let $f(x) = \frac{x}{(1+x)^2}$. By quotient rule*

$$\frac{d}{dx}\frac{x}{(1+x)^2} = \frac{1(1+x)^2 - 2(1+x)x}{(1+x)^4} = \frac{1-x}{(1+x)^3}$$

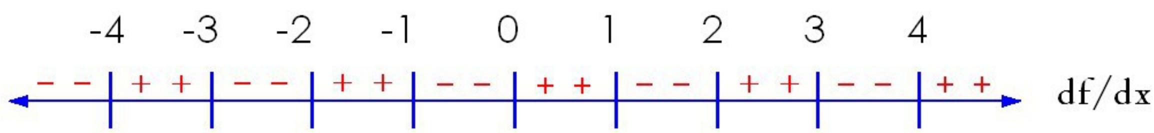
Thus the critical points are $c = 1$ and $c = -1$. The sign-chart is



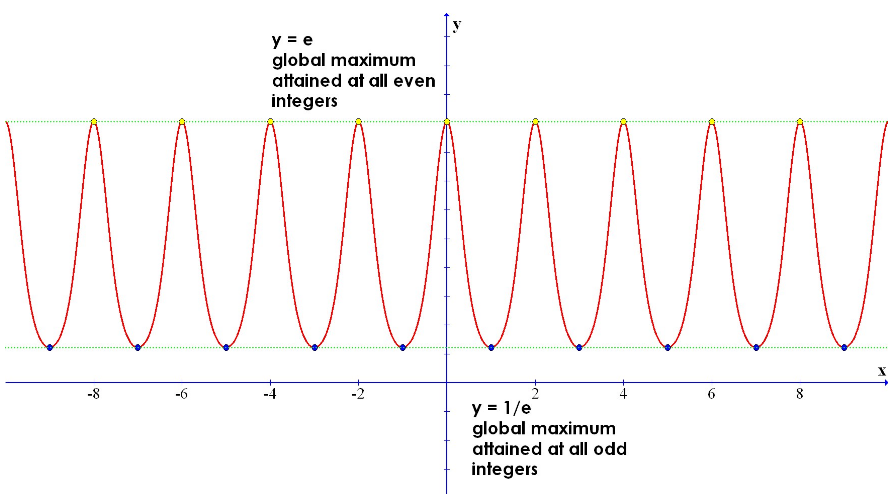
Observe that $x = -1$ is a VA and by the first derivative test $f(1) = 1/4$ is a local maximum. The function is increasing on $(-1, 1)$ and it is decreasing on $(-\infty, 1) \cup (1, \infty)$



Example 5.1.25. *Suppose $f(x) = e^{\cos(\pi x)}$. We calculate by chain rule $f'(x) = -\pi \sin(\pi x)e^{\cos(\pi x)}$. Note that the exponential function is nonzero thus $f'(x) = 0$ implies $\sin(\pi x) = 0$, but we recall from our study of trigonometry that the set of solutions are precisely those $x \in \mathbb{R}$ such that $\pi x = n\pi$ for some $n \in \mathbb{Z}$. In this example we find infinitely many critical points. In particular, $c_n = n$ implies $f'(c_n) = 0$. The sign-chart is*

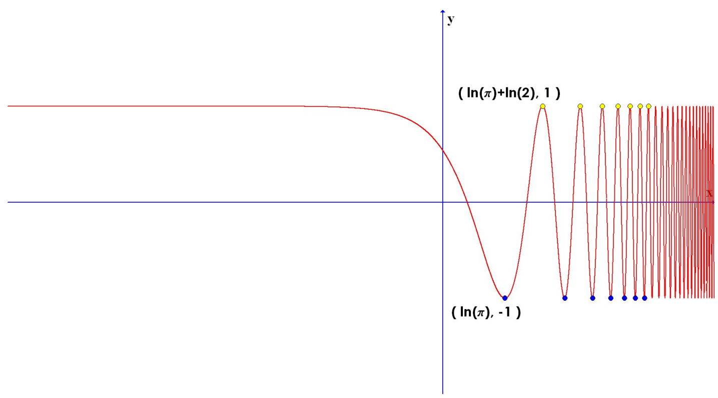


For each even integer $2n$ we apply first derivative test to find $f(2n) = e$ is the global maximum of f and for each odd integer $2n+1$ we apply first derivative test to find $f(2n+1) = 1/e$ is the global minimum of f. The graph is sort of like an cosine graph, although it is bounded by $1/e \leq e^{\cos(\pi x)} \leq e$ and you can see the shape not the same as cosine.



I have pointed out a few maxima $(2n, e)$ with yellow dots and minima $(2n-1, \frac{1}{e})$ with blue dots in the picture above.

Example 5.1.26. *Suppose* $f(x) = \cos(e^x)$. *The chain rule provides* $f'(x) = -e^x \sin(e^x)$. *We will find infinitely many solutions for the critical number criteria* $f'(x) = -e^x \sin(e^x) = 0$. *Note* $e^x \neq 0$ *for all* $x \in \mathbb{R}$ *hence we must have* $\sin(e^x) = 0$. *Consequently we find solutions described implicitly by* $e^x = n\pi$ *for* $n \in \mathbb{Z}$. *Since* $e^x > 0$ *we have no solutions with* $n \leq 0$. *If* $n > 0$ *then we can solve for* $x = \ln(n\pi) = \ln(n) + \ln(\pi)$. *Define* $c_n = \ln(n) + \ln(\pi)$, *then clearly* $f'(c_n) = 0$ *for each* $n \in \mathbb{N}$. *The critical numbers* $c_1, c_2, \dots$ *are not evenly spaced. Instead, as* n *increases we know the* $\ln(n)$ *grows slower and slower which means the critical numbers are closer and closer as* $x \to \infty$. *Note that* $-e^x \sin(e^x)$ *changes from* $+$ *to* $-$ *if* $e^x = 2n\pi$ *whereas* $-e^x \sin(e^x)$ *changes from* $-$ *to* $+$ *if we cross* $e^x = (2n-1)\pi$. *Therefore, by first derivative test,* $f(c_{2n}) = 1$ *is the global maximum which is attained at* $x = c_{2n}$ *for* $n \in \mathbb{N}$ *and* $f(c_{2n-1}) = -1$ *is the global minimum which is attained at* $x = c_{2n-1}$ *for* $n \in \mathbb{N}$.



The yellow dots are at $(c_{2n}, 1)$ *and the blue dots are at* $(c_{2n-1}, 1)$ *for* $n = 1, 2, 3, 4, 5$.

In-Class Example 5.1.27. *Let* $f(x) = xe^{-x}$. *Find intervals of increase and decrease. Classify any critical points. Sketch graph.*

Example 5.1.28. *Let $f(x) = \sqrt{(x-1)^2} - \sqrt{(x-2)^2}$. You should recognize*[3] *these are formulas for the absolute value function $y = |x|$ shifted either one or two units right. We expect there will be two critical points. Let us verify my conjecture,*

$$\begin{aligned} f'(x) &= \frac{d}{dx}\left[\sqrt{(x-1)^2} - \sqrt{(x-2)^2}\right] \\ &= \frac{2(x-1)}{2\sqrt{(x-1)^2}} - \frac{2(x-2)}{2\sqrt{(x-2)^2}} \\ &= \frac{(x-1)\sqrt{(x-2)^2} - (x-2)\sqrt{(x-1)^2}}{\sqrt{(x-1)^2}\sqrt{(x-2)^2}}. \end{aligned}$$

You might be tempted to just cancel terms in the numerator and conclude $f'(x) = 0$ for all $x \in \mathbb{R}$. However, this is not correct. In fact, $f'(1)$ and $f'(2)$ do not exist and $f'(x) = 2$ for $1 < x < 2$. Let us change notation a bit so the problem becomes clearer, the trouble with this problem is that we really need to break it into cases to see clearly:

$$\sqrt{(x-1)^2} = |x-1| = \begin{cases} x-1 & if\ x > 1 \\ 1-x & if\ x \leq 1 \end{cases} \qquad \sqrt{(x-2)^2} = |x-2| = \begin{cases} x-2 & if\ x > 2 \\ 2-x & if\ x \leq 2 \end{cases}$$

Therefore,

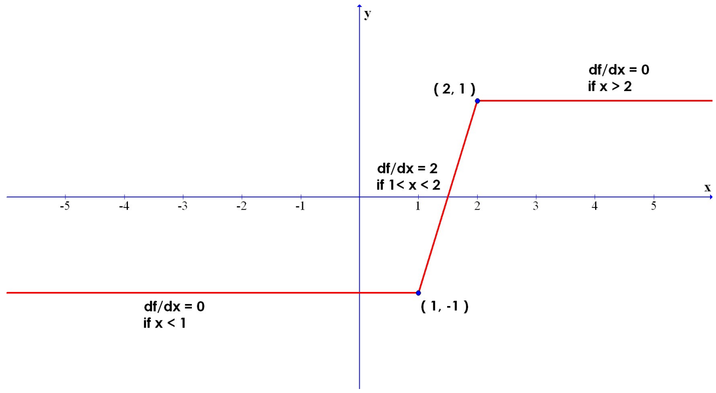
$$f(x) = |x-1| - |x-2| = \begin{cases} -1 & if\ x \leq 1 \\ 2x-1 & if\ 1 \leq x \leq 2 \\ 1 & if\ x \geq 2 \end{cases}.$$

It follows that

$$f'(x) = \begin{cases} 0 & if\ x < 1 \\ 2 & if\ 1 < x < 2 \\ 0 & if\ x > 2 \end{cases}.$$

we can show that f is continuous on $\mathbb{R}$ however the derivative f' is discontinuous at $x = 1$ and $x = 2$. In fact, $dom(f') = \mathbb{R} - \{1, 2\}$. The first derivative test does not apply to this example. Notice that the set of critical points for f is $(-\infty, 1] \cup [2, \infty)$. Since the derivative is constant on $(-\infty, 1]$ and $[2, \infty)$ we find the function is constant on those intervals. (we already found this but I point out that the differential calculus and our previous propositions on constant functions and derivatives do apply to this case even though the first derivative test is non-applicable.)

[3] we could write $\sqrt{(x-1)^2} = \pm(x-1)$, however I hope you realize that it is not correct to simply write $\sqrt{(x-1)^2} = x-1$ for generic x. This mistake made many students miss this problem on a previous semester's test



5.1.2 concavity and the second derivative test

A function is concave-up on an interval J if the function has the shape of a bowl which is right-side up over that interval J. A function is concave down on an interval J if the function has the shape of a bowl which is up-side down over that interval J. In other words, a concave up function stays below the secant line but a concave down function stays above the secant line.

Definition 5.1.29.

Let f be a function which is twice differentiable on some connected set J,

1. f is **concave up** if the derivative of f is decreasing on J (abbreviated CU on J)
2. f is **concave down** if the derivative of f is increasing on J (abbreviated CD on J)
3. if $p \in dom(f)$ is a point such that there exists $\delta > 0$ such that f is concave up(down) on $(p-\delta, p)$ and concave down(up) on $(p, p+\delta)$ then we say p or $(p, f(p))$ is an **inflection point**. An inflection point is a point where the concavity changes.

In-Class Example 5.1.30. *Illustrate CU and CD and label points of inflection for an example.*

In-Class Example 5.1.31. *What simple family of graphs is both CU and CD everywhere ?*

Remark 5.1.32.

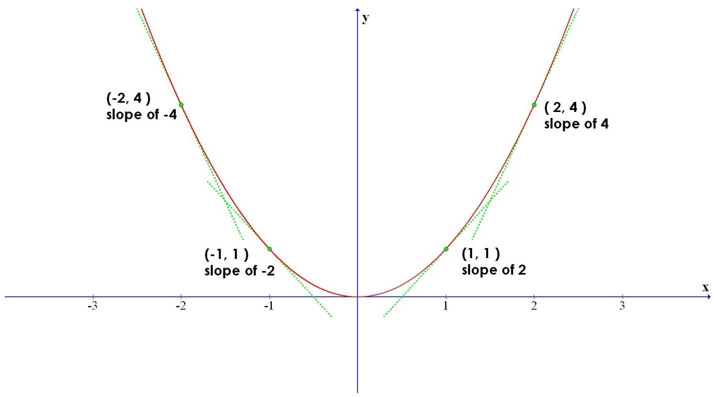
One easy way to remember which is up and which is down is the following slogan:

concave up: is locally like a **u**   concave down: is locally like a **n**.

This slogan is useful to help create graphs if you already know the concavity.

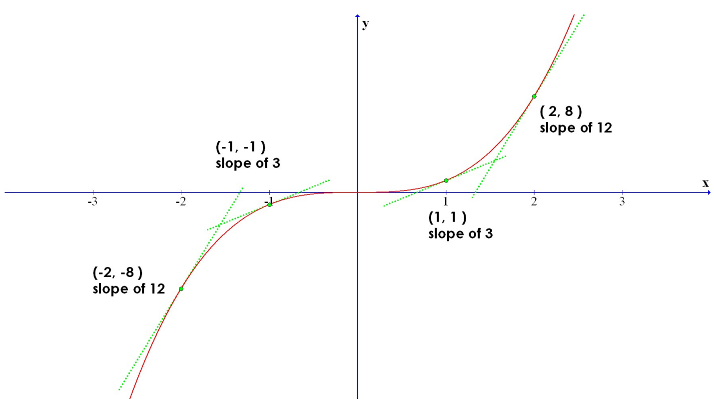
Incidentally, the term "convex" was historically used for concave down and this term is still used in physics particularly in the study of optics.

Example 5.1.33. *If $f(x) = x^2$ then $f'(x) = 2x$. Notice that $f''(x) = 2 > 0$ therefore f' is an increasing function on $\mathbb{R}$. It follows that $y = x^2$ is concave up on $\mathbb{R}$.*



Notice that the tangents (in green) are under the graph since the function is CU everywhere.

Example 5.1.34. *If $f(x) = x^3$ then $f'(x) = 3x^2$. Notice that $f''(x) = 6x$ is positive for $x > 0$ whereas $x < 0$ implies $f''(x) < 0$. Therefore, f' is increasing on $(0, \infty)$ and f' is decreasing on $(-\infty, 0)$. It follows that $y = x^3$ is CU on $(0, \infty)$ and CD on $(-\infty, 0)$. Thus $(0, 0)$ is an inflection point.*



Notice that the tangents (in green) are over the graph where it is CD $(x < 0)$ whereas the tangents are under the graph where the function is CU $(x > 0)$.

Theorem 5.1.35. *sign-charts for derivatives reveal increase and decrease of function.*

> Suppose f has continuous f'' on (a, b) and $f'(c) = 0$ for $c \in (a, b)$,
>
> 1. if $f''(x) > 0$ for all $x \in (a, b)$ then f is concave up on (a, b).
> 2. if $f''(x) < 0$ for all $x \in (a, b)$ then f is concave down on (a, b).
> 3. if $f''(c) = 0$ then this test is inconclusive.

I emphasize that when the second derivative is zero we might find an inflection point, but it doesn't have to be the case. The same is true for critical points. When a critical point is not at a local max or min it could be an inflection point or it might be something else.

Theorem 5.1.36. *Second Derivative Test.*

> Suppose f has a critical number c such that $f'(c) = 0$ and $f''(x)$ is exists for $x \in B_\delta(c)$ for some $\delta > 0$ then
>
> 1. if $f''(c) > 0$ then $f(c)$ is a local minimum at c.
> 2. if $f''(c) < 0$ then $f(c)$ is a local maximum at c.
> 3. if $f''(c) = 0$ then this test is inconclusive.

Proof: suppose $f''(c) > 0$ and $f'(c) = 0$. Notice that f' is continuous on $B_\delta(c)$ for some $\delta > 0$ since f'' is defined on that set and differentiability of f' implies continuity of f'. Furthermore, notice that f' is strictly increasing on $B_\delta(c)$ therefore f' is an injection on $B_\delta(c)$. Since $f'(c) = 0$ and f' is strictly increasing it follows that $f'(x) < 0$ for $x \in B_\delta(c)$ with $x < c$ and $f'(x) > 0$ for $x \in B_\delta(c)$ with $x > c$ therefore by the First Derivative Test we conclude $f(c)$ is a local minimum. A similar argument applies to case 2. $\square$

We will discover another proof for the second derivative test when we discuss Taylor's Theorem later in this chapter.

In-Class Example 5.1.37. *Let $f(x) = x^2 - 2bx + c$ where b, c are constants. Show $f(b)$ is a local minimum for this function.*

In-Class Example 5.1.38. *Suppose $f(x) = e^{-x^2}$. Find intervals of concavity and increase and decrease. Find extremal points and inflection points. Sketch the graph.*

In-Class Example 5.1.39. *Suppose* $f(x) = 3x^5 - 20x^4 + 40x^3$. *Find intervals of concavity and increase and decrease. Find extremal points and inflection points. Sketch the graph.*

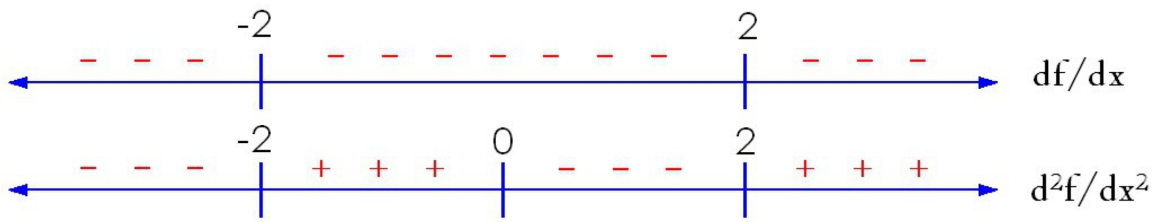
Example 5.1.40. *Suppose $f(x) = \frac{1}{x+2} + \frac{1}{x-2}$. If we make a common denominator we find $f(x) = \frac{2x}{x^2-4}$. We differentiate (the original given formula),*

$$f'(x) = \frac{-1}{(x+2)^2} - \frac{1}{(x-2)^2} = -\frac{2x^2+8}{(x+2)^2(x-2)^2}$$

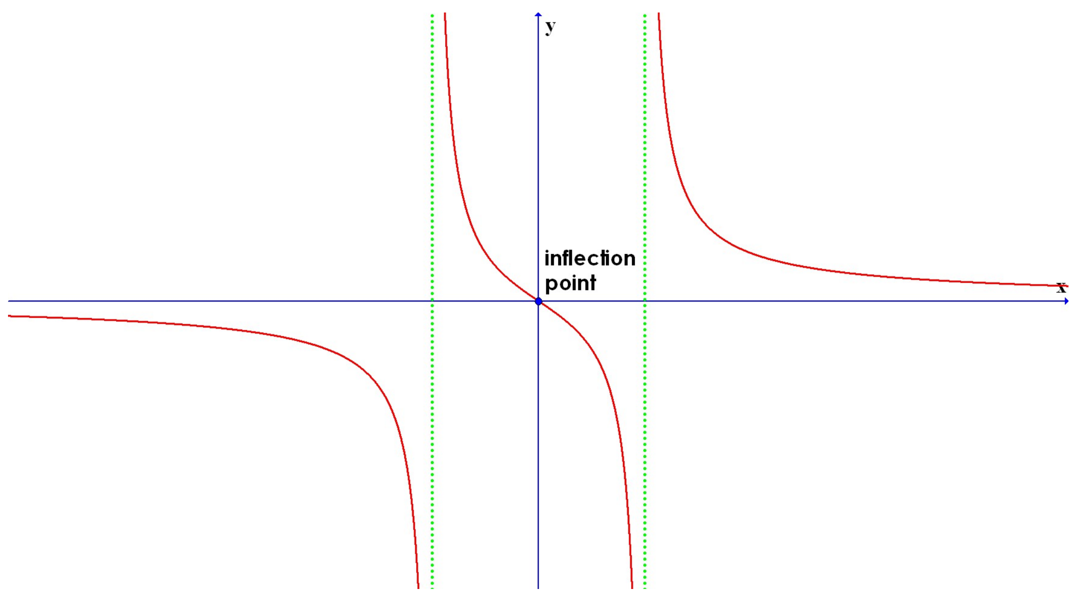
then differentiate again (using the unsimplified $f'(x)$ as starting point),

$$f'(x) = \frac{2}{(x+2)^3} + \frac{2}{(x-2)^3} = 2\frac{2x^3+48x}{(x+2)^3(x-2)^3} = \frac{4x(x^2+24)}{(x+2)^3(x-2)^3}.$$

We find critical points $c = -2, 2$ and points of possible inflection at $-2, 0, 2$.



We find the function is decreasing on $\mathbb{R}$ and of the three possible inflection points only $(0,0)$ is a point of inflection, the concavity also changes at $x = \pm 2$ but those are VA so we shouldn't say those are points of inflection. This rational function has a graph that is CU on $(-2,0)$ and $(2,\infty)$ and it is CD on $(-\infty,-2)$ and $(0,2)$.

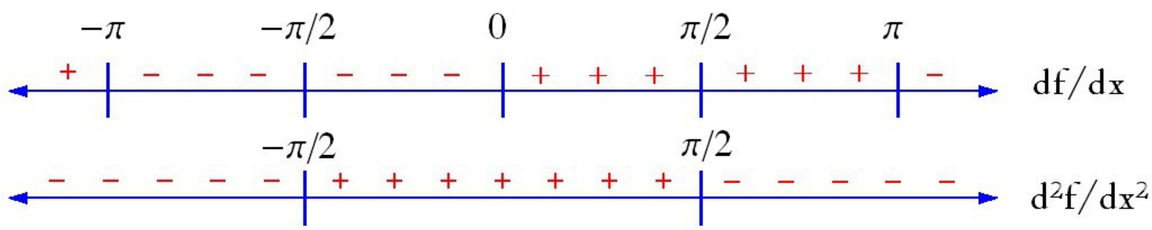


One physical interpretation mathematics found in the previous example is that $y = f(x)$ could be a graph of the electric potential along the x-axis for two positive point charges placed at $x = -2$ and $x = 2$. A divergence in the potential signals the presence of localized charge.

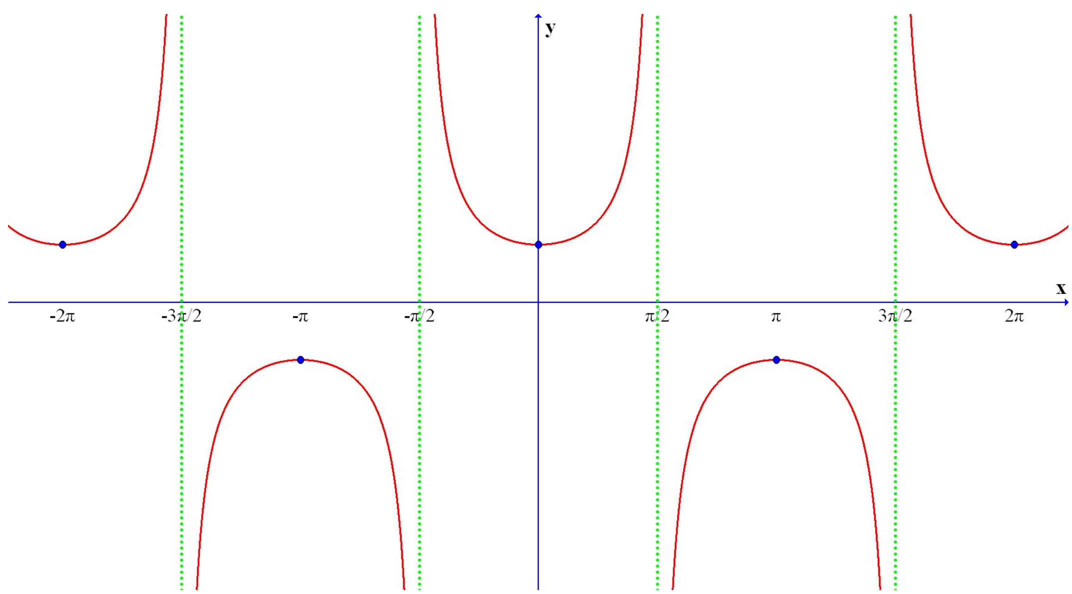
Example 5.1.41. *Suppose $f(x) = \sec(x)$ then $f'(x) = \sec(x)\tan(x)$ and $f''(x) = \sec^2(x)\tan(x) + \sec^3(x)$ by the product rule. Let us write these in terms of sine and cosine since we have a complete and working knowledge of all the zeros for sine and cosine.*

$$f(x) = \frac{1}{\cos(x)} \qquad \frac{df}{dx} = \frac{\sin(x)}{\cos^2(x)} \qquad \frac{d^2f}{dx^2} = \frac{\sin(x)+1}{\cos^3(x)}$$

It follows that critical points arise from where $\sin(x) = 0$ or where $f'(x)$ does not exist because $\cos(x) = 0$; that $c_n = \frac{\pi}{2}n$ for $n \in \mathbb{Z}$. We also see that the odd-integer critical points are also locations of possible concavity change since a vanishing cosine makes $f''(x)$ undefined. Note that $\sin(x) = -1$ has solutions $x_j = \frac{\pi}{2}4j - 1$ for $j \in \mathbb{Z}$. For example, $j = 0$ gives $\sin(-\frac{\pi}{2}) = -1$ and $j = 1$ gives $\sin(\frac{3\pi}{2}) = -1$. These points are included already as a subset of the zeros of cosine. The concavity can only change at a zero of cosine.



Notice that the local maximum of 1 is attained at $x = 2n\pi$ for $n \in \mathbb{Z}$ whereas a local minimum of 1 is attained at $x = (2n-1)\pi$ for $n \in \mathbb{Z}$. The fact these are respectively local maximums and minimums is verified by the second derivative test since $f''(2n\pi) = 1 > 0$ and $f''((2n-1)\pi) = -1 < 0$ for all $n \in \mathbb{Z}$. Naturally the first derivative test agrees. Both tests are evident from the sign-chart given above.



Remark 5.1.42. *concerning on case 3 of 2nd Derivative Test*

Maybe you are wondering, what is an example of a function which falls into case 3 of the derivative test? One simple example is a line $y = f(x) = mx + b$ which has $f'(x) = m$. Clearly f and f' are continuous everywhere. Notice $f''(x) = 0$ for each $x \in \mathbb{R}$. There are two cases:

1. $m = 0$, thus $f(x) = b$ and $y = b$ is the maximum and minimum value of the function at all points.

2. $m \neq 0$, then $f(x) = mx + b$ and the function has no extrema with respect to $\mathbb{R}$.

Notice also that $g(x) = x^4 + 1$ and $h(x) = x^5 + 4$ both have critical number $c = 0$ and $g''(0) = h''(0) = 0$ however $(0, g(0))$ is a local minimum whereas $(0, h(0))$ is an inflection point. The second derivative is too clumsy to detect the difference. Later in this chapter we'll discover that Taylor's polynomial approximation theorem covers cases like g or h.

5.2 closed interval method

The following theorem details how to actually find the extrema the Extreme Value Theorem indicated exist. If f is continuous on $[a, b]$ then the Extreme Value Theorem says there exist global extrema with respect to $[a, b]$. If an extrema are in the interior then it must also be a local extrema thus by Fermat's theorem it will occur at a critical number. Otherwise, the extrema are at the endpoints. Therefore, if we check endpoints and critical points we will find the extrema.

Theorem 5.2.1. *closed interval method.*

If we are given function f which is continuous on a closed interval $[a, b]$ the we can find the absolute minimum and maximum values of the function over the interval $[a, b]$ as follows:

1. Locate all critical numbers $x = c$ in (a, b) and calculate $f(c)$.

2. Calculate $f(a)$ and $f(b)$.

3. Compare values from steps 1. and 2. the largest of these values is the absolute maximum, the smallest (or largest negative) value is the absolute minimum of f on $[a, b]$.

Example 5.2.2. *Let* $f(x) = \sin(x)$ *find absolute extrema of* f *relative to interval* $0 \leq x \leq 2\pi$. *Note* $f'(x) = \cos(x)$ *and* $\cos(x) = 0$ *has solutions* $x = \frac{\pi}{2}, \frac{\pi}{2} \in [0, 2\pi]$.

$$f(0) = \sin(0) = 0, \quad f(\tfrac{\pi}{2}) = \sin(\tfrac{\pi}{2}) = 1, \quad f(\frac{3\pi}{2}) = \sin(\tfrac{3\pi}{2}) = -1, \quad \sin(2\pi) = 0$$

Therefore, by closed interval method $f(\frac{\pi}{2}) = 1$ *is the maximum and* $f(\frac{3\pi}{2}) = -1$ *is the minimum of* $f(x) = \sin(x)$ *on the interval* $[0, 2\pi]$.

In-Class Example 5.2.3. *Let* $f(x) = (x-3)(x-4)$ *find absolute extrema of* f *on* $[0,1]$.

In-Class Example 5.2.4. *Let* $f(x) = x^4 - 2x^2 + 3$ *find absolute extrema of* f *on* $[0,2]$.

Example 5.2.5. *Let $f(x) = e^{-x}\sin(x)$. Find the extreme values of f on $[0,4]$.*

$$f'(x) = -\sin(x)e^{-x} + \cos(x)e^{-x} = (\cos(x) - \sin(x))e^{-x}.$$

Solutions of $\cos(x) = \sin(x)$ are critical points. If you picture the graphs of sine and cosine on the same plot then the solutions are given from the points of intersection. In particular, $c = \frac{\pi}{4} + n\pi$ for $n \in \mathbb{Z}$. The critical points in $[0,4]$ are $\frac{\pi}{4}$ and $\frac{5\pi}{4} \approx 3.93$. Calculate,

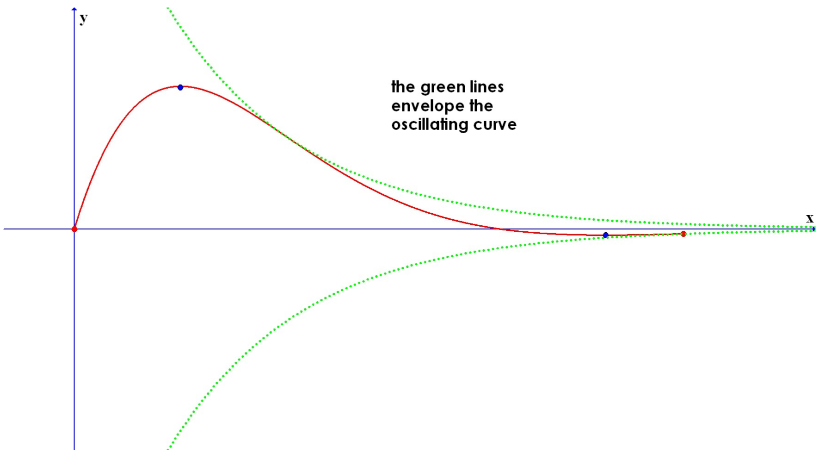
$$f(\tfrac{\pi}{4}) = e^{-\frac{\pi}{4}}\sin(\tfrac{\pi}{4}) \approx 0.32$$

$$f(\tfrac{5\pi}{4}) = e^{-\frac{5\pi}{4}}\sin(\tfrac{5\pi}{4}) \approx -0.0139$$

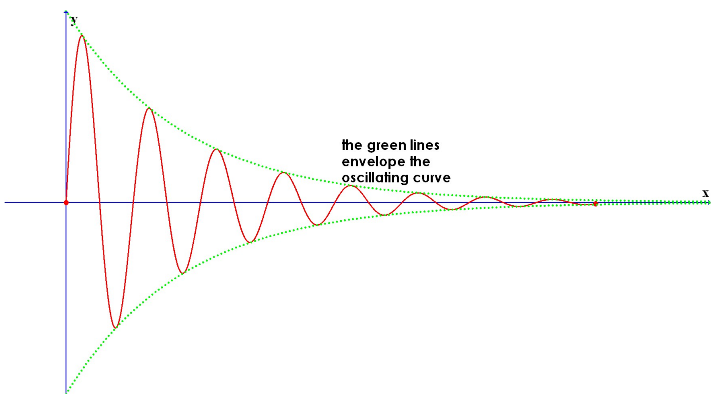
$$f(0) = e^{0}\sin(0) = 0$$

$$f(4) = e^{-4}\sin(4) = -0.138$$

We find $f(\frac{5\pi}{4}) = -0.0139$ is the minimum and $f(\frac{\pi}{4}) = 0.32$ is the maximum of f on the interval $[0,4]$. The graph has blue dots to illustrate the extrema.



I suppose we ought to be happy the last example wasn't $f(x) = e^{-x}\sin(11x)$. That would have required more calculation.



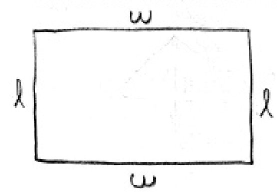
Physically these are very interesting functions. You should see it again when you study springs with friction or RLC circuits.

5.3 optimization

I have preserved the format of these examples from an earlier edition of my notes.

Example 5.3.1.

$\boxed{\text{E1}}$. Given 400ft of fencing what dimensions should you give a rectangular pen as to maximize the area?



$$2\ell + 2w = 400$$

$$A = \ell w$$

Notice that $\ell = 200 - w$

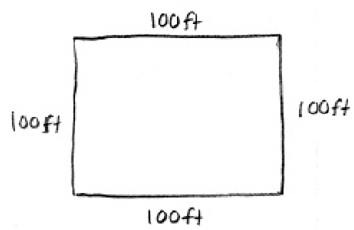
$$\Rightarrow \boxed{A = (200-w)w = 200w - w^2}$$

Maximize A as a function of w then,

$$\frac{dA}{dw} = 200 - 2w \Rightarrow \frac{dA}{dw} = 0 \text{ when } w = 100$$

Then note $\frac{d^2A}{dw^2} = -2 < 0 \Rightarrow A(100)$ is a maximum.

Thus by 2^{nd} derivative test $w = 100$ gives max. area. Thus



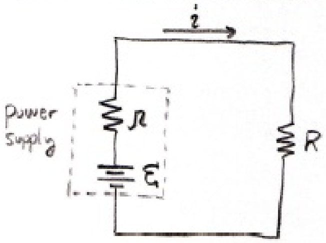
$$\ell = 200 - 100 = 100$$

These dimensions maximize the area of the pen.

In-Class Example 5.3.2. *Given 400 ft of fencing to build rectangular pen next to lava river. Find dimensions to maximize area. Note: cows are afraid of lava, so no fence needed on one side.*

Example 5.3.3.

$\boxed{E2}$ Given a power supply with fixed voltage $\mathcal{E}$ and internal resistance r what load R will maximize the power delivered to the load?



$\mathcal{E} = ir + iR$: Kirchoff's Voltage Law

$P = i^2 R$: Power delivered to R by current i

Solve to find $i = \dfrac{\mathcal{E}}{r+R}$ then substitute into $P = i^2 R = \dfrac{\mathcal{E}^2 R}{(r+R)^2}$

Now do calculus on P as a function of R,

$$
\begin{aligned}
\frac{dP}{dR} &= \mathcal{E}^2 \frac{d}{dR}\left(\frac{R}{(r+R)^2}\right) \\
&= \mathcal{E}^2 \left[\frac{(r+R)^2 - 2(r+R)R}{(r+R)^4}\right] \\
&= \frac{\mathcal{E}^2}{(r+R)^4}\left[r^2 + \cancel{2rR} + R^2 - \cancel{2rR} - 2R^2\right] \\
&= \frac{\mathcal{E}^2}{(r+R)^4}\left[r^2 - R^2\right] \\
&= \frac{\mathcal{E}^2}{(r+R)^2}\left[(r-R)(r+R)\right]
\end{aligned}
$$

: $r, R > 0$ for physical reasons so $R = r$ is only interesting critical point.

$$
\underset{0}{|}\; + + + + + + + + + + + \;\underset{r}{|}\; - - - - - - - - - - \rightarrow \frac{dP}{dR}
$$

Thus by 1^{st} derivative test the max power is delivered to R when $R = r$. This is a simple case of the

"Max Power Transfer Th^{m}"

Notice it implies that the Max Efficiency of a simple power supply is 50%. Electrical engineers can build more efficient power supplies called "switching" power supplies, in those supplies the output is automatically taylored as to be best for the load.

In-Class Example 5.3.4. *The range of a projectile fired on a level field is given by* $R = \dfrac{2v_o^2 \sin\theta \cos\theta}{g}$ *where* v_o, g *are constants. Find the angle which maximizes the range.*

In-Class Example 5.3.5. *If a projectile is fired with speed* v_o *at angle of inclination* θ *from initial height* y_o *then the height of the projectile is given by* $y = y_o + v_o \sin thetat - \frac{1}{2}gt^2$*. Find the maximum height of the projectile.*

In-Class Example 5.3.6. *Find the point on* $y = 2 - x$ *which is closest to* $(0, 0)$.

In-Class Example 5.3.7. *Find the point on* $y = 3 - x^2$ *which is closest to the point* $(1, 1)$. *Hint:* $4x^3 - 6x - 2 = (x + 1)(2x^2 - 2x - 1) = 0$.

5.4 to $\pm\infty$ and beyond

The behavior a function for $x >> 0$ or for $x << 0$ is captured by the limit[4] of the function at $\pm\infty$,

Definition 5.4.1. *limits at ∞ or $-\infty$.*

The limit at ∞ for a function f is $L \in \mathbb{R}$ if the values $f(x)$ can be made arbitrarily close to L for inputs x sufficiently large. We write

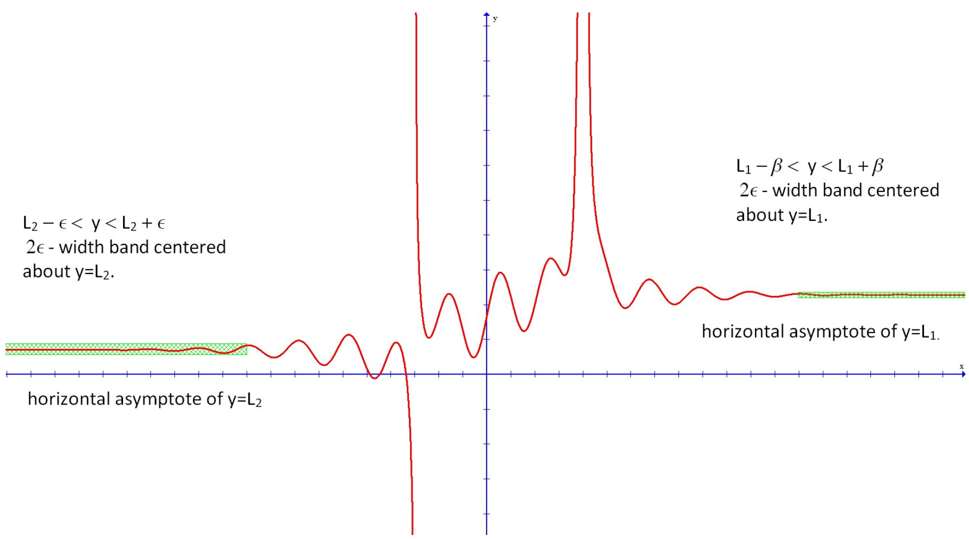
$$\lim_{x\to\infty} f(x) = L$$

in this case. To be more precise we should say that $\lim_{x\to\infty} f(x) = L$ iff for each $\varepsilon > 0$ there exists $N \in \mathbb{R}$ with $N > 0$ such that if $x > N$ then $|f(x) - L| < \varepsilon$. Likewise,

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

iff for each $\varepsilon > 0$ there exists $M \in \mathbb{R}$ with $M < 0$ such that if $x < M$ then $|f(x) - L| < \varepsilon$.

Geometrically this definition essentially says that if we pick a band of width 2ε about the line $y = L$ then for points to the right(or left) of N (or M) the graph $y = f(x)$ fits inside the band. In the picture below you can see that for any $\varepsilon > 0$ or $\beta > 0$ we can find a band about the limiting value in which the tail of the graph can be fit.



Given the graph above we expect $\lim_{x\to\infty} f(x) = L_1$ and $\lim_{x\to-\infty} f(x) = L_2$.

---

[4]we did discuss the values of the function tending to arbitrarily large positive or negative values with respect to some finite limit point. I would say those are limits which go to ∞ whereas this section is about limits which are taken at $\pm\infty$. These concepts are not mutually exclusive; $\lim_{x\to\infty} e^x = \infty$.

Example 5.4.2. *Let $f(x) = \frac{1}{x}$. Calculate the limit of $f(x)$ at ∞. Observe that,*

$$f(10) = 0.1, \qquad f(100) = 0.01, \qquad f(1000) = 0.001.$$

We see that the values of the function are getting closer and closer to zero as x gets larger and larger. This leads us to suspect,

$$\lim_{x\to\infty} \frac{1}{x} = 0.$$

In other words, if we divide something nonzero by a very big number then we get something very small. This sort of limit is not ambiguous, to determine the answer intuitively we either need to think about a table of values or perhaps a graph.

Or if you want to be rigorous you can argue as follows: Let $\varepsilon > 0$ choose $N = 1/\varepsilon$ and observe that for $x > N = 1/\varepsilon$ it follows that $1/x < \varepsilon$. Consequently, $x > N$ implies $|f(x) - 0| = |\frac{1}{x}| = \frac{1}{x} < \varepsilon$. Hence by the precise definition $\lim_{x\to\infty} \frac{1}{x} = 0$. $\square$

The limits at $-\infty$ are much the same.

Example 5.4.3. *Let $f(x) = \frac{1}{x}$. Calculate the limit of $f(x)$ at $-\infty$. Observe that,*

$$f(-10) = -0.1, \qquad f(-100) = -0.01, \qquad f(-1000) = -0.001.$$

We see that the values of the function are getting closer and closer to zero as x gets larger and negative. This leads us to suspect,

$$\lim_{x\to-\infty} \frac{1}{x} = 0.$$

In other words, if we divide something nonzero by a very big negative number then we get something very small and negative. This sort of limit is not ambiguous, to determine the answer intuitively we either need to think about a table of values or perhaps a graph.

Or if you want to be rigorous you can argue as follows: Let $\varepsilon > 0$ choose $M = -1/\varepsilon$ and observe that for $x < M = -1/\varepsilon$ it follows that $-1/x < \varepsilon$. Consequently, $x < N$ implies $|f(x) - 0| = |\frac{1}{x}| = -\frac{1}{x} < \varepsilon$. Hence by the precise definition $\lim_{x\to-\infty} \frac{1}{x} = 0$. $\square$

Clearly we'd prefer to avoid the picky ε-type arguments if possible. Towards that end I'm offering proofs for a number of standard results and theorems so that we have justification for later algebraic or intuitive arguments to solve limits at $\pm\infty$. As always it is still important we remember at the definition is actually precise even if we sometimes allow some amount of intuitive argumentation.

Example 5.4.4. *Let $f(x) = 1/x^n$ where $n > 0$. Calculate the limit of $f(x)$ at ∞. Observe that,*

$$f(10) = 1/10^n, \qquad f(100) = 1/100^n, \qquad f(1000) = 1/1000^n.$$

We see that the values of the function are getting closer and closer to zero as x gets larger and larger. This leads us to suspect,

$$\lim_{x\to\infty} \frac{1}{x^n} = 0.$$

Let $\varepsilon > 0$ *choose* $N = 1/\varepsilon^{\frac{1}{n}}$. *Suppose* $x > N = 1/\epsilon^{\frac{1}{n}}$ *thus* $1/x < \epsilon^{\frac{1}{n}}$ *which implies* $1/x^n < (\epsilon^{\frac{1}{n}})^n = \epsilon$. *Consider then, if* $x > N$ *then*

$$|f(x) - 0| = |1/x^n| = 1/x^n < \epsilon.$$

Therefore by the precise definition for limits at infinity, $\lim_{x\to\infty} \frac{1}{x^n} = 0$. $\square$

The graphical significance of all three examples thus far considered is that the function has a *horizontal asymptote* of $y = 0$ as $x \to \pm\infty$.

Definition 5.4.5. *horizontal asymptotes.*

If $\lim_{x\to\infty} f(x) = L$ then the function f is said to have a **horizontal asymptote of** $y = L$ **at** ∞. If $\lim_{x\to-\infty} f(x) = L$ then the function f is said to have a **horizontal asymptote of** $y = L$ **at** $-\infty$.

In-Class Example 5.4.6. *Find horizontal asymptotes of* $f(x) = \tan^{-1}(x)$.

Vertical asymptotes of the function correspond to horizontal asymptotes for the inverse function[5] We can also discuss limits which go to infinity at infinity. It's just the natural merger of both definitions but I state it here for completeness.

Definition 5.4.7. *infinite limits at infinity.*

[5]Challenge: what are the horizontal asymptotes of $y = \tan^{-1}(3x)$?

The limit at ∞ for a function f is ∞ iff for each $M > 0$ there exists $N > 0$ such that for $x > N$ we find $f(x) > M$. We denote

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

in this case. Likewise, the limit at $-\infty$ for a function f is ∞ iff for each $M > 0$ there exists $N < 0$ such that if $x < N$ then $f(x) > M$. We denote this by

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

Similarly, if for each $M < 0$ there exists $N > 0$ such that $x > N$ implies $f(x) < M$ we say $\lim_{x\to-\infty} f(x) = -\infty$. Finally, if for each $M < 0$ there exists $N < 0$ such that $x < N$ implies $f(x) < M$ we say $\lim_{x\to-\infty} f(x) = -\infty$.

Example 5.4.8. *I would say that the limit below are not indeterminant. Their values can be deduced by straightforward analysis from the definition. The formal proof of these claims I leave to the reader.*

$$\begin{array}{cccccc}
\displaystyle\lim_{x\to\infty}\frac{1}{x}=0 & \displaystyle\lim_{x\to-\infty}\frac{1}{x}=0 & \displaystyle\lim_{x\to0^+}\frac{1}{x}=\infty & \displaystyle\lim_{x\to0^-}\frac{1}{x}=-\infty & \displaystyle\lim_{x\to0}\frac{1}{x}=dne \\
\displaystyle\lim_{x\to\infty}\frac{1}{x^2}=0 & \displaystyle\lim_{x\to-\infty}\frac{1}{x^2}=0 & \displaystyle\lim_{x\to0^+}\frac{1}{x^2}=\infty & \displaystyle\lim_{x\to0^-}\frac{1}{x^2}=\infty & \displaystyle\lim_{x\to0}\frac{1}{x^2}=\infty \\
\displaystyle\lim_{x\to\infty}\frac{1}{\sqrt{x}}=0 & \displaystyle\lim_{x\to-\infty}\frac{1}{\sqrt{x}}=? & \displaystyle\lim_{x\to0^+}\frac{1}{\sqrt{x}}=\infty & \displaystyle\lim_{x\to0^-}\frac{1}{\sqrt{x}}=? & \displaystyle\lim_{x\to0}\frac{1}{\sqrt{x}}=dne \\
\displaystyle\lim_{x\to\infty}\sqrt{x}=\infty & \displaystyle\lim_{x\to-\infty}\sqrt{x}=? & \displaystyle\lim_{x\to0^+}\sqrt{x}=0 & \displaystyle\lim_{x\to0^-}\sqrt{x}=? & \displaystyle\lim_{x\to0}\sqrt{x}=dne \\
\displaystyle\lim_{x\to\infty}x^2=\infty & \displaystyle\lim_{x\to-\infty}x^2=\infty & \displaystyle\lim_{x\to0^+}x^2=0 & \displaystyle\lim_{x\to0^-}x^2=0 & \displaystyle\lim_{x\to0}x^2=0 \\
\displaystyle\lim_{x\to\infty}x^3=\infty & \displaystyle\lim_{x\to-\infty}x^3=-\infty & \displaystyle\lim_{x\to0^+}x^3=0 & \displaystyle\lim_{x\to0^-}x^3=0 & \displaystyle\lim_{x\to0}x^3=0.
\end{array}$$

I have used "?" instead of d.n.e. in a few places just to make it fit. Those limits are taken at a limit point which is not in the domain of the function, in some cases not even on the boundary of the function. If we can't take values close to the limit point then by default the limit is said to not exist, in which case we use "d.n.e." or "dne" as a shorthand.

We can also have limits which fail to exist at plus or minus infinity due to oscillation. All of the functions in the next example fall into that category.

Example 5.4.9. *the following limits all involve cyclic functions. They never settle down to one* *value for large positive or negative input values so the limits d.n.e.*

$$\begin{array}{rrr}
\lim_{x\to\infty} \sin(x) = d.n.e. & \lim_{x\to\infty} \cos(x) = d.n.e. & \lim_{x\to\infty} \tan(x) = d.n.e. \\
\lim_{x\to-\infty} \sin(x) = d.n.e. & \lim_{x\to-\infty} \cos(x) = d.n.e. & \lim_{x\to-\infty} \tan(x) = d.n.e. \\
\lim_{x\to\infty} \csc(x) = d.n.e. & \lim_{x\to\infty} \sec(x) = d.n.e. & \lim_{x\to\infty} \sec(x) = d.n.e. \\
\lim_{x\to-\infty} \csc(x) = d.n.e. & \lim_{x\to-\infty} \sec(x) = d.n.e. & \lim_{x\to-\infty} \sec(x) = d.n.e.
\end{array}$$

Example 5.4.10. *The interplay between a function and its inverse is especially enlightening for* $\ln(x), \sin^{-1}(x), \cos^{-1}(x)$. *I refer the reader to the earlier chapter on preliminary material if it is forgotten by now.*

$$\begin{array}{rrr}
\lim_{x\to\infty} \sin^{-1}(x) = d.n.e. & \lim_{x\to\infty} \cos^{-1}(x) = d.n.e. & \lim_{x\to\infty} \tan^{-1}(x) = \pi/2 \\
\lim_{x\to-\infty} \sin^{-1}(x) = d.n.e. & \lim_{x\to-\infty} \cos^{-1}(x) = d.n.e. & \lim_{x\to-\infty} \tan^{-1}(x) = -\pi/2 \\
\lim_{x\to\infty} e^{x} = \infty & \lim_{x\to\infty} e^{-x} = 0 & \lim_{x\to\infty} (1/2)^{x} = 0 \\
\lim_{x\to-\infty} e^{x} = 0 & \lim_{x\to-\infty} e^{-x} = \infty & \lim_{x\to-\infty} (1/2)^{x} = \infty
\end{array}$$

The domain of $\sin^{-1}(x)$ *and* $\cos^{-1}(x)$ *will be the range of sine and cosine respectively; that is* $dom(\sin^{-1}(x)) = [-1,1]$ *and* $dom(\cos^{-1}(x)) = [-1,1]$ *so clearly the limits at plus and minus infinity are not sensible as inverse sine and cosine are not even defined at* $\pm\infty$. *In contrast the range of the exponential function is all positive real numbers and* $\ln(x)$ *is the inverse function of* e^x *thus*

$$\lim_{x\to-\infty} \ln(x) = d.n.e. \qquad \lim_{x\to 0^+} \ln(x) = -\infty \qquad \lim_{x\to\infty} \ln(x) = \infty$$

For $x < 0$ *the* $\ln(x)$ *is not real, the middle limit you should have thought about in the earlier discussion of limits. The last one is true although an uncritical appraisal of the graph* $y = \ln(x)$ *gives the appearance of a horizontal asymptote, but appearances can be deceiving.*

The following lemma connects limits at $\pm\infty$ with one-sided limits at zero.

Lemma 5.4.11.

$$\lim_{x\to\infty} f(x) = \lim_{t\to 0^+} f(1/t) \qquad \text{and} \qquad \lim_{x\to-\infty} f(x) = \lim_{t\to 0^-} f(1/t).$$

The equalities above apply to the case that the limit exists as well as the cases where the limits do not exist. We mean for the equality to denote that both limits diverge in the same manner.

Proof: Let's begin with the case that $\lim_{x\to\infty} f(x) = L \in \mathbb{R}$. Let $\epsilon > 0$ and note the following inequalities are equivalent:

$$0 < M < x \qquad \Leftrightarrow \qquad 0 < \frac{1}{x} < \frac{1}{M}$$

Therefore, $0 < \frac{1}{x} < \frac{1}{m}$ implies $|f(x) - L| < \epsilon$ which indicates that

$$\lim_{\frac{1}{x}\to 0^+} f(x) = L \qquad \text{hence using } t = 1/x \text{ we find} \qquad \lim_{t\to 0^+} f(1/t) = L.$$

The proof $\lim_{x\to-\infty} f(x) = \lim_{t\to 0^-} f(1/t) \in \mathbb{R}$ is similar.

Suppose $\lim_{x\to-\infty} f(x) = \infty$. It follows that for each $N > 0$ there exists $M < 0$ such that $x < M$ implies $f(x) > N$. Note that $\frac{1}{M} < \frac{1}{x}$ is equivalent with $x < M$ thus $\frac{1}{M} < \frac{1}{x} < 0$ implies $f(x) > N$. But the last string of inequalities yields that

$$\lim_{\frac{1}{x}\to 0^-} f(x) = \infty \qquad \text{hence using } t = 1/x \text{ we find} \qquad \lim_{t\to 0^-} f(1/t) = \infty.$$

Proof for other cases are similar and left to the reader. The basic point is that if $x \to \pm\infty$ then $t = \frac{1}{x} \to 0^{\pm}$. $\square$

With the little lemma above in mind we see that all the limit theorems transfer over to limits at $\pm\infty$ since each such limit is in 1-1 correspondence with a one-sided limit at zero and we already proved the limit laws for limits at zero. Rather than restating all the limit laws again I will illustrate by example. In fact, let's get straight to the fun part: indeterminant limits.

5.4.1 algebraic techniques for calculating limits at $\pm\infty$

Up to this point I have attempted to catalogue the basic results. I'm sure I forgot something important, but I hope these examples give you enough of a basis to do those limits which are unambiguous at plus or minus infinity. There is another category of problems where the limits which are given are not obvious, there is some form of indeterminancy. All the same indeterminant forms (see defn. 2.3.5) arise again and most of the algebraic techniques we used back in section 2.3 will arise again here although perhaps in a slightly altered form.

The good news is that limits at infinity enjoy all the same properties as limits which are taken at a finite limit point, at least in as much as the properties make sense. Of course we can only apply the limit properties when the values of the limit are finite. For example,

$$\lim_{x\to\infty}(x - 2x) = \lim_{x\to\infty}(x) + \lim_{x\to\infty}(-2x) = \infty - \infty$$

is not valid because you might be tempted to cancel and find $\lim_{x\to\infty}(x-2x) = 0$ yet $\lim_{x\to\infty}(x - 2x) = \lim_{x\to\infty}(-x) = -\infty$ is the correct result. So we should only split limits by the limit laws when the subsequent limits are finite. That said, I do admit there are certain cases it doesn't hurt to apply the limit laws even though the limits are infinite. In particular, suppose $c \neq 0$, if $\lim f = \infty$ then $\lim cf = c\lim f = c\infty$ provided we agree to understand that $c\infty = \infty$ for $c > 0$ whereas $c\infty = -\infty$ if $c < 0$. Such statements are dangerous because the reader may be tempted to apply laws of arithmetic to expressions involving ∞ and it's just not that simple. We should always remember that ∞ is just a notation for a particular limiting process in calculus[6]

[6]in complex variables one can actually add the point at infinity and use the extended complex numbers. In fact,

Example 5.4.12. *this one is type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}
\lim_{x\to\infty}\left(\frac{x+3}{x-2}\right) &= \lim_{x\to\infty}\left(\frac{\frac{x}{x}+\frac{3}{x}}{\frac{x}{x}-\frac{2}{x}}\right) && \textit{divided top and bottom by } x \\
&= \lim_{x\to\infty}\left(\frac{1+0}{1-0}\right) && c/x \to 0 \textit{ as } x \to \infty \\
&= 1.
\end{aligned}$$

Example 5.4.13. *this one is also of type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}
\lim_{x\to\infty}\left(\frac{x^3+3x-2}{x^4-2x+1}\right) &= \lim_{x\to\infty}\left(\frac{\frac{1}{x}+\frac{3}{x^3}-\frac{2}{x^4}}{1-\frac{2}{x^3}+\frac{1}{x^4}}\right) && \textit{divided top and bottom by } x^4 \\
&= \lim_{x\to\infty}\left(\frac{0+0-0}{1-0+0}\right) && \textit{for } n=1,2,4,\ c/x^n \to 0 \textit{ as } x \to \infty \\
&= 0.
\end{aligned}$$

Example 5.4.14. *again, type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}
\lim_{x\to-\infty}\left(\frac{x^3+3x-2}{x^2-x+7}\right) &= \lim_{x\to-\infty}\left(\frac{x+\frac{3}{x}-\frac{2}{x^2}}{1-\frac{2}{x}+\frac{7}{x^2}}\right) && \textit{divided top and bottom by } x^4 \\
&= \lim_{x\to-\infty}\left(\frac{x}{1}\right) && \textit{for } n=1,2,\ c/x^n \to 0 \textit{ as } x \to -\infty \\
&= -\infty.
\end{aligned}$$

Another way of thinking about this one is to put in very big negative values of x. For example, when $x=-1000$ we find

$$\frac{x^3+3x-2}{x^2-x+7} = \frac{-1000^3-3000-2}{1000^2+1000-2} \approx \frac{-1000^3}{1000^2} = -1000 = x$$

This sort of reasoning is a good method to try if you are lost as to what algebraic step to apply. There are problems which no amount of algebra will fix, sometimes considering numerical evidence is the best way to figure out a limit. However, for some functions -1000 is not big enough, take $f(x)=\frac{1}{2x-1000}$ we find $f(-1000)=-1/3$. But, you can show $f(x)\to 0$ as $x\to-\infty$. To be safer you should experiment with more than one number, or better yet THINK.

---

some authors use a similar idea for calculus, they introduce the so-called extended real numbers or the "really long line" of $\mathbb{R}\cup\{\infty\}\cup\{-\infty\}$. If this sort of thing seems interesting to you then perhaps you ought to read the text *Elementary Calculus: An Infinitesimal Approach* by H. Jerome Keisler

Example 5.4.15. *you guessed it, type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}
\lim_{x\to\infty}\left(\frac{\sqrt{2x^4+3x-2}}{x^2-x+7} \right) &= \lim_{x\to\infty}\left(\frac{\frac{1}{x^2}\sqrt{2x^4+3x-2}}{\frac{1}{x^2}\left(x^2-x+7\right)} \right) \\
&= \lim_{x\to\infty}\left(\frac{\sqrt{\frac{2x^4+3x-2}{x^4}}}{1-\frac{1}{x}+\frac{7}{x^2}} \right) \\
&= \lim_{x\to\infty}\left(\frac{\sqrt{2+\frac{3}{x^3}-\frac{2}{x^4}}}{1-\frac{1}{x}+\frac{7}{x^2}} \right) \\
&= \sqrt{2}.
\end{aligned}$$

Example 5.4.16. *this has type $0\cdot\infty$ to begin.*

$$\begin{aligned}
\lim_{x\to\infty}\left(e^{-x}2^x \right) &= \lim_{x\to\infty}\left(e^{\ln(2^x)}e^{-x}\right) \qquad \textit{sneaky step} \\
&= \lim_{x\to\infty}\left(e^{x\ln(2)}e^{-x}\right) \\
&= \lim_{x\to\infty}\left(e^{x(\ln(2)-1)}\right) \\
&= 0.
\end{aligned}$$

In the last step I noticed $\ln(2)-1\approx 0.692-1<0$ thus the limit amounts to the exponential function evaluated at ever increasing large negative values which indicates the limit is zero. This example really belongs in the section with l'Hopital's Rule, I include it now for novelty only.

We find that limits of type ∞/∞ can result in many different final answers depending on how the indeterminancy is resolved. The next example is more general, I think it is healthy to think about something a little more abstract from time to time. The strategy used is essentially identical to the strategy employed in several of the preceding examples.

In-Class Example 5.4.17. *let P be a polynomial of degree p and let Q be a polynomial of degree q. This means there exist real coefficients $a_p, a_{p-1},\dots,a_1,a_0$ and $b_q,b_{q-1},\dots,b_1,b_0$ such that $a_p\neq 0$ and $b_q\neq 0$ where*

$$P(x)=a_px^p+\cdots+a_1x+a_0 \qquad Q(x)=b_qx^q+\cdots+b_1x+b_0$$

Determine the possible values of $\displaystyle\lim_{x\to\infty}\frac{P(x)}{Q(x)}$.

In case you forgot, a function f is said to be **bounded** if there exist $m, M \in \mathbb{R}$ such that $m < f(x) < M$ for all $x \in dom(f)$.

Example 5.4.18. *we can throw away a bounded function in a sum when the other function in the sum is unbounded, here are two examples of this idea in action:*

$$\lim_{x\to\infty} (\sin(x) + e^x) = \lim_{x\to\infty} (e^x) = \infty$$
$$\lim_{x\to-\infty} (x + 2) = \lim_{x\to-\infty} (x) = -\infty.$$

Example 5.4.19. *if we take a function $f(x)$ with a known limit of $L \in \mathbb{R}$ or $\pm\infty$ as $x \to \pm\infty$ then the limit of $f(x + a)$ for $a \in \mathbb{R}$ is the same for $x \to \pm\infty$. For example,*

$$\lim_{x\to\infty} (e^x) = \infty \quad \Longrightarrow \quad \lim_{x\to\infty} (e^{x+3}) = \infty$$
$$\lim_{x\to-\infty} \left(\frac{1}{x^2}\right) = 0 \quad \Longrightarrow \quad \lim_{x\to-\infty} \left(\frac{1}{(x-7)^2}\right) = 0$$
$$\lim_{x\to\infty} (\tan^{-1}(x)) = \frac{\pi}{2} \quad \Longrightarrow \quad \lim_{x\to\infty} (\tan^{-1}(x+2)) = \frac{\pi}{2}.$$

Example 5.4.20. *The Squeeze Theorem applies to limits at $\pm\infty$. Suppose we are given a function f such that*

$$\frac{2}{\pi} \tan^{-1}(x) \leq f(x) \leq \frac{\sqrt{4x^2+1}}{x-3}$$

for all $x \geq 14,000,000,000,000$ (national debt 2010). We can calculate the limit at ∞ via the Squeeze Theorem. Observe that

$$\lim_{x\to\infty} \left(\frac{2}{\pi} \tan^{-1}(x) \right) = \frac{2}{\pi} \cdot \frac{\pi}{2} = 1$$
$$\lim_{x\to\infty} \left(\frac{\sqrt{4x^2+1}}{2x-3} \right) = \lim_{x\to\infty} \left(\frac{\sqrt{4+1/x^2}}{2-3/x} \right) = \sqrt{4}/2 = 1.$$

Therefore, by the Squeeze Theorem, $\lim_{x\to\infty} f(x) = 1$.

5.5 l'Hopital's rule

In earlier sections we were able to resolve many indeterminant limits with purely algebraic arguments. You might have noticed we have not really tried to use calculus to help us solve limits better. In our viewpoint, limits were just something we needed to do in order to carefully define the derivative. However, we were certainly happy enough once those limits vanished and were replaced by a few essentially algebraic rules. Linearity, product, quotient and chain rules all involve a limiting argument if we consider the technical details. The fact that we can do calculus without dwelling on those details is in my view why calculus is so beautifully simple.

In this section we will learn about l'Hopital's Rule which allows us to use calculus to resolve limits which are indeterminant. We need to have limits of type ∞/∞ or $0/0$ in order to apply the rule. Often we will need to rewrite the given expression in order to change it to either type ∞/∞ or $0/0$. We will see that $\infty - \infty, 1^{\infty}, \infty^0, 0^0$ can all be resolved with the help of l'Hopital's Rule.

l'Hopital's Rule says that the limit of an indeterminant quotient of functions should be the same as the limit of the quotient of the derivatives of those functions. Essentially the idea is to compare how the numerator changes verses the how the denominator changes. This can be done at a finite limit point or with limits at $\pm\infty$.

Theorem 5.5.1. *l'Hopital's Rule*

Suppose that $\lim \frac{f}{g}$ is of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$ then

$$\lim\left(\frac{f}{g}\right) = \lim\left(\frac{f'}{g'}\right).$$

where equality includes all cases including those divergent cases. Note lim is meant to denote both left, right and double-sided limits at a finite point and also limits at $\pm\infty$.

Example 5.5.2. *Notice* $\lim_{x\to 0}\frac{\sin(x)}{x}$ *is type* $\frac{0}{0}$. *Observe that*

$$\begin{aligned} \lim_{x\to 0}\frac{\sin(x)}{x} &= \lim_{x\to 0}\frac{\cos(x)}{1} \qquad \textit{L'Hopital with } \frac{0}{0} \\ &= 1. \end{aligned}$$

Remark 5.5.3. *notation for l'Hopital's rule*

At the present time I have not found a way to adequately translate my notation for applying l'Hopital's rule into LaTeX. You should notice my notation in lecture is less cumbersome.

Example 5.5.4. *In this example we'll apply l'Hopital's rule twice to remove the indeterminancy.*

$$\begin{aligned} \lim_{x\to\infty}\frac{x^2+x-2}{x^2+3} &= \lim_{x\to\infty}\frac{2x+1}{2x} \qquad \textit{L'Hopital on type } \frac{\infty}{\infty} \\ &= \lim_{x\to\infty}\frac{2}{2} \qquad \textit{L'Hopital on type } \frac{\infty}{\infty} \\ &= 1. \end{aligned}$$

In-Class Example 5.5.5. *Calculate* $\lim\limits_{x\to\infty}\dfrac{\ln(x^2)}{\sqrt[3]{x}}$.

Example 5.5.6.

$$\begin{aligned}
\lim_{x\to 0^+} xe^{\frac{1}{x}} &= \lim_{x\to 0^+} \frac{e^{\frac{1}{x}}}{\frac{1}{x}} \qquad \textit{L'Hopital on type } \frac{\infty}{\infty} \\
&= \lim_{x\to 0^+} \frac{e^{\frac{1}{x}}\left(\frac{-1}{x^2}\right)}{\frac{-1}{x^2}} \\
&= \lim_{x\to 0^+} e^{\frac{1}{x}} \\
&= \infty.
\end{aligned}$$

Remark 5.5.7. *notation for l'Hopital's rule*

In the preceding example it was not initially possible to apply l'Hopital's rule. This is a common trouble in these problems. Often we are faced with type $0 \cdot \infty$ in which case we can either reformulate the quotient to be type $0/0$ or type ∞/∞. Which choice is best is exposed via trial, error and ultimately experience born from mathematical experimentation.

In-Class Example 5.5.8. *Calculate* $\lim_{x\to\infty} e^{-x}x^2$.

Example 5.5.9.

$$\begin{aligned}
\lim_{x\to\infty} \frac{x^2}{\ln(x)} &= \lim_{x\to\infty} \frac{2x}{\frac{1}{x}} \qquad \textit{L'Hopital on type } \frac{\infty}{\infty} \\
&= \lim_{x\to\infty} (2x^2) \\
&= \infty.
\end{aligned}$$

Apparently the natural logarithm grows slower than a quadratic function.

In-Class Example 5.5.10. *Calculate* $\lim_{\theta \to 0^+} (\csc(\theta) - \cot(\theta))$.

In-Class Example 5.5.11. *Calculate* $\lim_{x \to 1} \left(\frac{1}{\ln(x)} - \frac{1}{x-1} \right)$.

5.5.1 indeterminant powers

We have discussed indeterminant forms of type $0/0$, $\infty/infty$, $0 \cdot \infty$ and $\infty - \infty$ in some depth. There are three more forms to consider.

Definition 5.5.12. *forms of indeterminant power.*

1. we say $\lim f^g$ is of **"type 0^0"** iff $\lim f = 0$ and $\lim g = 0$
2. we say $\lim f^g$ is of **"type ∞^0"** iff $\lim f = \infty$ and $\lim g = 0$
3. we say $\lim f^g$ is of **"type 1^∞"** iff $\lim f = 1$ and $\lim g = \infty$

We will discover shortly that these forms largely reduce to the problems we previously considered once we understand a little lemma.

Lemma 5.5.13. *the power lemma.*

Suppose that $f(x) > 0$ for points considered in limit,

$$\lim \big[f(x)\big]^{g(x)} = exp\big(\lim g(x)\ln(f(x))\big)$$

where equality includes all cases including those divergent cases. In particular,

1. if $\lim[g(x)\ln(f(x))] = c \in \mathbb{R}$ then $\lim\big[f(x)\big]^{g(x)} = e^c$.
2. if $\lim[g(x)\ln(f(x))] = \infty$ then $\lim\big[f(x)\big]^{g(x)} = \infty$.
3. if $\lim[g(x)\ln(f(x))] = -\infty$ then $\lim\big[f(x)\big]^{g(x)} = 0$.

Note lim is meant to denote both left, right and double-sided limits at a finite point and also limits at $\pm\infty$.

Proof: follows from properties of natural logarithm as well as the continuity of the exponential function on $\mathbb{R}$:

$$\lim\big[f(x)\big]^{g(x)} = \lim\Big[exp\big(\ln\big[f(x)\big]^{g(x)}\big)\Big] = exp\big(\lim\ln\big[f(x)\big]^{g(x)}\big) = exp\big(\lim g(x)\ln(f(x))\big).$$

I leave the proof of the divergent cases for the reader. $\square$

Example 5.5.14. *Calculate* $\lim_{x\to 0^+} x^x$. *We use the power lemma, consider*

$$\lim_{x\to 0^+} x^x = exp(\underbrace{\lim_{x\to 0^+}(x\ln(x))}_{\star})$$

We focus on $\star$, notice it is of type $0 \cdot \infty$ so we use the standard technique to rewrite it as ∞/∞ and apply l'Hopital's rule

$$\star = \lim_{x\to 0^+} (x\ln(x)) = \lim_{x\to 0^+} \frac{\ln(x)}{\frac{1}{x}} = \lim_{x\to 0^+} \frac{\frac{1}{x}}{\frac{-1}{x^2}} = \lim_{x\to 0^+} (-x) = 0$$

Hence, we find $\star = 0$ and returning to our original limit,

$$\lim_{x\to 0^+} x^x = exp(0) = 1.$$

Example 5.5.15. *Calculate $\lim_{x\to\infty} x^{\frac{1}{x}}$. We use the power lemma,*

$$\lim_{x\to\infty} x^{\frac{1}{x}} = exp\Bigg[\underbrace{\lim_{x\to\infty} \frac{1}{x}\ln(x)}_{\star}\Bigg]$$

We focus on $\star$, notice it is of type $0 \cdot \infty$ so we use the standard technique to rewrite it as ∞/∞ and apply l'Hopital's rule

$$\star = \lim_{x\to\infty} \frac{1}{x}\ln(x) = \lim_{x\to\infty} \frac{\ln(x)}{x} = \lim_{x\to\infty} \frac{\frac{1}{x}}{1} = 0.$$

Hence, we find $\star = 0$ and returning to our original limit,

$$\lim_{x\to 0^+} x^{\frac{1}{x}} = exp(0) = 1.$$

In-Class Example 5.5.16. *Calculate* $\displaystyle\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{nx}$.

5.6 Taylor's Theorem about polynomial approximation

The idea of a Taylor polynomial is that if we are given a set of initial data $f(a), f'(a), f''(a), \dots, f^{(n)}(a)$ for some function $f(x)$ then we can approximate the function with an n^{th}-order polynomial which fits all the given data.

In-Class Example 5.6.1. *Let $c_0, c_1, \dots, c_n \in \mathbb{R}$ be coefficients of*

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n$$

where $a \in \mathbb{R}$. Find how the coefficients and the derivatives $f^{(k)}(x)$ are linked. Note, $f^{(0)} = f$.

The above example leads us to define:

Definition 5.6.2. *Suppose f has k-derivatives defined at a,*

We define $T_k(x)$ the **k-th order Taylor approximation of $f(x)$ centered at $x = a$** by

$$T_k(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots + \frac{1}{k!}f^{(k)}(a)(x-a)^k.$$

When $f(x)$ is itself a polynomial[7] then any Taylor polynomial of degree $k \geq deg(f)$ is literally the same function as f; $f(x) = T_k(x)$. However, generally $T_k(x)$ is just an approximation of $f(x)$. If $x \cong a$ then $f(x) \cong T_k(x)$, and we can make the approximation better by either using a larger order polynomial approximation, or by staying closer to the center of the Taylor expansion. I will soon illustrate this claim with several examples and Taylor's Theorem (proved later in this section) gives a precise bound on the error between $f(x)$ and $T_k(x)$.

[7] Example 3.9.2 foreshadows this comment

Suppose f is a function which is smooth at $x = a$, then:

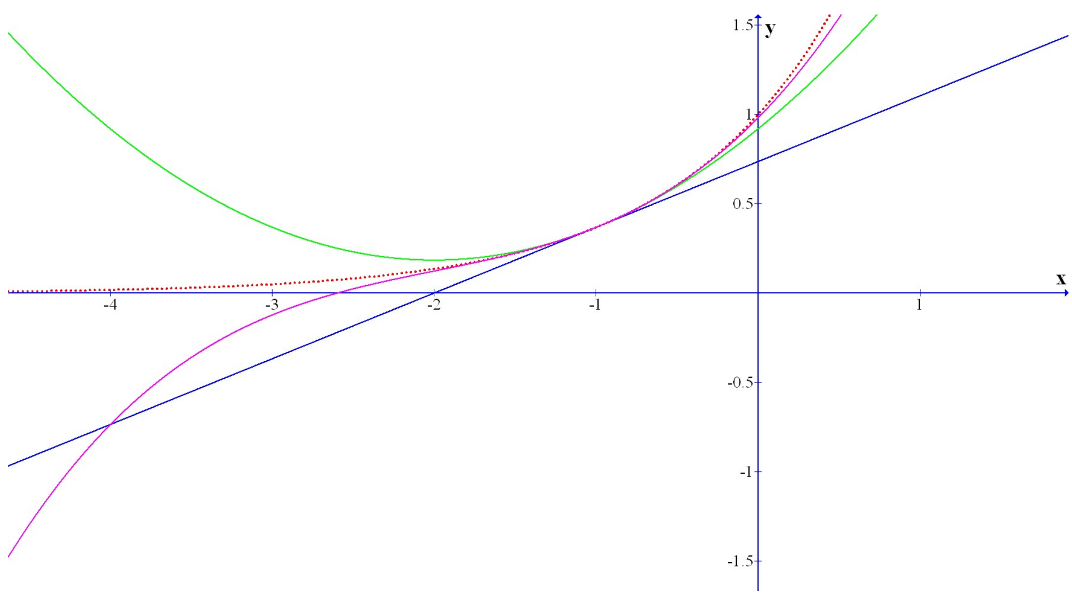
- **constant approximation:** $T_0(x) = f(a)$
- **linear approximation:** $T_1(x) = f(a) + f'(a)(x-a)$
- **quadratic approximation:** $T_2(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$
- **cubic approximation:** $T_3(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{6}f'''(a)(x-a)^3$.

Notice the *linear approximation* is simply the tangent line approximation. Quadratic and cubic approximations are not as familar, but the concept remains the same: take a complicated function and replace it by a simple function.

Example 5.6.3. *Suppose $f(x) = e^x$. Calculate the first few Taylor polynomials centered at $a = -1$. Calculate $f'(x) = f''(x) = f'''(x) = e^x$ and $f(-1) = e^{-1} = \frac{1}{e}$. Hence,*

$$\begin{aligned}
T_o(x) &= \frac{1}{e} \\
T_1(x) &= \frac{1}{e} + \frac{1}{e}(x+1) \\
T_2(x) &= \frac{1}{e} + \frac{1}{e}(x+1) + \frac{1}{2e}(x+1)^2 \\
T_3(x) &= \frac{1}{e} + \frac{1}{e}(x+1) + \frac{1}{2e}(x+1)^2 + \frac{1}{6e}(x+1)^3.
\end{aligned}$$

The graph below shows $y = e^x$ as the dotted red graph, $y = T_1(x)$ is the blue line, $y = T_2(x)$ is the green quadratic and $y = T_3(x)$ is the purple graph of a cubic. The cubic is the best approximation.



Example 5.6.4. *Suppose $f(x) = \frac{1}{x-2} + 1$. Calculate the first few Taylor polynomials centered at $a = 1$. Observe*

$$f(x) = \frac{1}{x-2} + 1, \quad f'(x) = \frac{-1}{(x-2)^2}, \quad f''(x) = \frac{2}{(x-2)^3}, \quad f'''(x) = \frac{-6}{(x-2)^4}$$

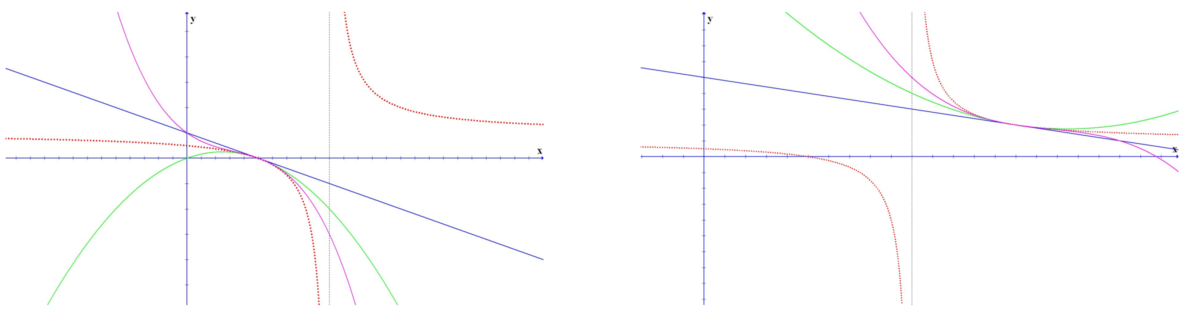
thus $f(1) = 0, f'(1) = -1, f''(1) = -2$ and $f'''(1) = -6$. Hence,

$$\begin{aligned} T_{1,a=1}(x) &= -(x-1) \\ T_{2,a=1}(x) &= -(x-1) + (x-1)^2 \\ T_{3,a=1}(x) &= -(x-1) + (x-1)^2 - (x-1)^3 \end{aligned}$$

Alternatively, for $a = 3$ calculate $f(3) = 2, f'(3) = -1, f''(3) = 2$ and $f'''(1) = -6$. Hence,

$$\begin{aligned} T_{1,a=3}(x) &= 2 - (x-3)) \\ T_{2,a=3}(x) &= 2 - (x-3) + (x-3)^2 \\ T_{3,a=3}(x) &= 2 - (x-3) + (x-3)^2 - (x-3)^3. \end{aligned}$$

The graph below shows $y = \frac{1}{x-2} + 1$ as the dotted red graph, $y = T_1(x)$ are the blue lines, $y = T_2(x)$ are the green quadratics and $y = T_3(x)$ are the purple graphs. You can see that the cubic is the best approximation in both cases. Also, you can see that $T_{k,a=1}$ will not give a good approximation to $f(x)$ to the right of the VA at $x = 2$ and $T_{k,a=3}$ do not well-approximate $f(x)$ to the left of the VA.



In-Class Example 5.6.5. *Find the $T_6(x)$ with $a = 0$ for $f(x) = \cosh x$ centered at $a = 0$.*

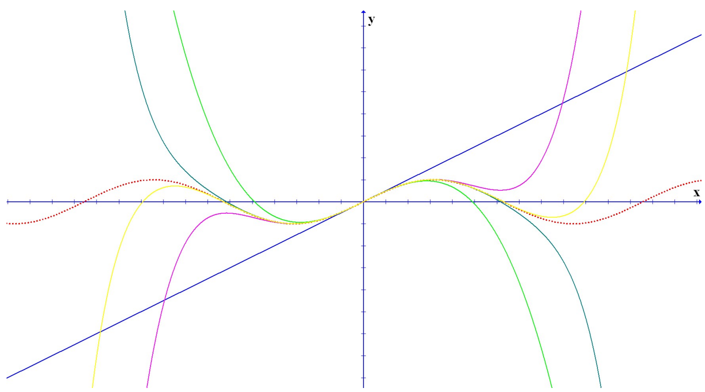
Example 5.6.6. *Let $f(x) = \sin(x)$. It follows that*

$$f'(x) = \cos(x), f''(x) = -\sin(x), f'''(x) = -\cos(x), f^{(4)}(x) = \sin(x), f^{(5)}(x) = \cos(x)$$

Hence, $f(0) = 0, f'(0) = 1, f''(0) = 0, f'''(0) = -1, f^{(4)}(0) = 0, f^{(5)}(0) = 1$. *Therefore the Taylor polynomials of orders* $1, 3, 5$ *are*

$$\begin{aligned} T_1(x) &= x && \textit{blue graph} \\ T_3(x) &= x - \frac{1}{6}x^3 && \textit{green graph} \\ T_5(x) &= x - \frac{1}{6}x^3 + \frac{1}{120}x^5 && \textit{purple graph} \end{aligned}$$

The graph below shows the Taylor polynomials calculated above and the next couple orders above. You can see how each higher order covers more and more of the graph of the sine function.



Taylor polynomials can be generated for a given *smooth*[8] function through a certain linear combination of its derivatives. The idea is that we can approximate a function by a polynomial[9], at least locally. We discussed the tangent line approximation to a function. We found that the linearization of a function gives a good approximation for points close to the point of tangency. If we calculate second derivatives we can similarly find a quadratic approximation for the function. Third derivatives go to finding a cubic approximation about some point. I should emphasize from the outset that a Taylor polynomial is just a polynomial, it will not be able to exactly represent a function which is not a polynomial. In order to exactly represent an analytic function we'll need to take infinitely many terms, we'll need a power series. We discuss those carefully in calculus II. Finally, let me show you an example of how Taylor polynomials can be of fundamental importance in physics.

[8] for $p \in \mathbb{R}$ the notation $f \in C^\infty(p)$ means there exists a nbhd. of $p \in \mathbb{R}$ on which f has infinitely many continuous derivatives.

[9] there do exist pathological examples for which all Taylor polynomials at a point vanish even though the function is nonzero near the point; $f(x) = exp(-1/x^2)$ for $x \neq 0$ and $f(0) = 0$

Example 5.6.7. *The relativistic energy E of a free particle of rest mass m_o is a function of its velocity v:*

$$E(v) = \frac{m_o c^2}{\sqrt{1 - v^2/c^2}}$$

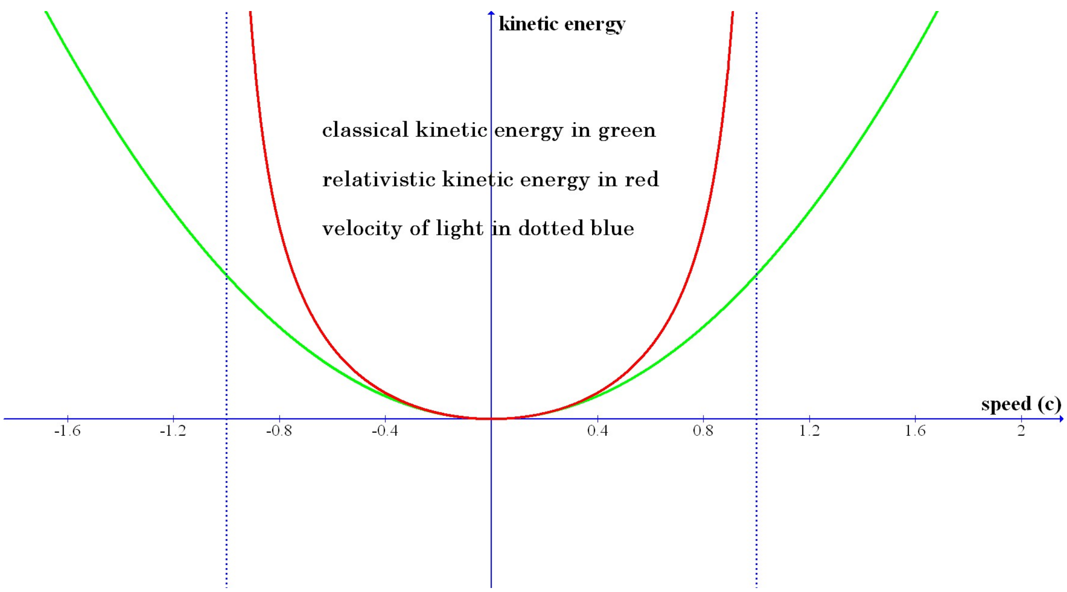
for $-c < v < c$ where c is the speed of light in the space. We calculate,

$$\frac{dE}{dv} = \frac{m_o v}{(1 - v^2/c^2)^{\frac{3}{2}}}$$

thus $v = 0$ is a critical number of the energy. Moreover, after a little calculation you can show the 4-th order Taylor polynomial in velocity v for energy E is

$$E(v) \approx m_o c^2 + \frac{1}{2} m_o v^2 + \frac{3m_o}{8c^2} v^4$$

The constant term is the source of the famous equation $E = m_o c^2$ and the quadratic term is precisely the classical kinetic energy. The last term is very small if $v \approx 0$. As $|v| \to c$ the values of the last term become more significant and they signal a departure from classical physics. I have graphed the relativistic kinetic energy $K = E - m_o c^2$ (red) as well as the classical kinetic energy $K_{Newtonian} = \frac{m_o}{2} v^2$ (green) on a common axis below:



The blue-dotted lines represent $v = \pm c$ and if $|v| > c$ the relativistic kinetic energy is not even defined. However, for $v \approx 0$ you can see they are in very good agreement. We have to get past 10% of light speed to even begin to see a difference. In every day physics most speeds are so small that we cannot see that Newtonian physics fails to correctly model dynamics. I may have assigned a homework based on the error analysis of the next section which puts a quantitative edge on the last couple sentences.

One of the great mysteries of modern science is this fascinating feature of *decoupling*. How is it that we are so fortunate that the part of physics which touches one aspect of our existence is so successfully described. Why isn't it the case that we need to understand relativity before we can pose solutions to the problems presented to Newtonian mechanics? Why is physics so nicely segmented that we can understand just one piece at a time? This is part of the curiosity which leads physicists to state that the existence of physical law itself is bizarre. If the universe is randomly generated as is life then how is it that we humble accidents can so aptly describe what surrounds us. What right have we to understand what we do of nature?

5.6.1 error in Taylor approximations

We've seen a few examples of how Taylor's polynomials will locally mimic a function. Now we turn to the question of extrema. Think about this, if a function is locally modeled by a Taylor polynomial centered at a critical point then what does that say about the nature of a critical point? To be precise we need to know some measure of how far off a given Taylor polynomial is from the function. This is what Taylor's theorem tells us. There are many different formulations of Taylor's theorem[10], the one below is partially due to Lagrange.

Theorem 5.6.8. *Taylor's theorem with Lagrange's form of the remainder.*

If f has k derivatives on a closed interval I with $\partial I = \{a, b\}$ then

$$f(b) = T_k(b) + R_k(b) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(b-a)^j + R_k(b)$$

where $R_k(b) = f(b) - T_k(b)$ is the k-th remainder. Moreover, there exists $c \in int(I)$ such that

$$R_k(b) = \frac{f^{(k+1)}(c)}{(k+1)!}(b-a)^{k+1}.$$

We have essentially proved the first portion of this theorem. It's straightforward calculation to show that $T_k(x)$ has the same value, slope, concavity etc... as the function at the point $x = a$. What is deep about this theorem is the existence of c. This is a generalization of the mean value theorem. Suppose that $a < b$, if we apply the theorem to

$$f(x) = T_o(x) + R_1(x)$$

we find Taylor's theorem proclaims there exists $c \in (a, b)$ such that $R_1(b) = f'(c)(b-a)$ and since $T_o(x) = f(a)$ we have $f(b) - f(a) = f'(c)(b-a)$ which is the conclusion of the MVT applied to $[a, b]$.

Proof of Taylor's Theorem: the proof I give here I found in *Real Variables with Basic Metric Space Topology* by Robert B. Ash. Proofs found in other texts are similar but I thought his was

[10]Chapter 7 of Apostol or Chapter II.6 of Edwards would be good additional readings if you wish to understand this material in added depth.

particularly lucid.

Since the k-th derivative is given to exist on I it follows that $f^{(j)}$ is continuous for each $j = 1, 2, \dots, k-1$ (we are not garaunteed the continuity of the k-th derivative, however it is not needed in what follows anyway). Assume $a < b$ and define M implicitly by the equation below:

$$f(b) = f(a) + f'(a)(b-a) + \cdots + \frac{f^{(k-1)}(a)}{(k-1)!}(b-a)^{(k-1)} + \frac{M(b-a)^k}{k!}.$$

Our goal is to produce $c \in (a, b)$ such that $f^{(k)}(c) = M$. Ash suggests replacing a with a variable t in the equation that defined M. Define g by

$$g(t) = -f(b) + f(t) + f'(t)(b-t) + \cdots + \frac{f^{(k-1)}(t)}{(k-1)!}(b-t)^{(k-1)} + \frac{M(b-t)^k}{k!}$$

for $t \in [a, b]$. Note that g is differentiable on (a, b) and continuous on $[a, b]$ since it is the sum and difference of likewise differentiable and continuous functions. Moreover, observe

$$g(b) = -f(b) + f(b) + f'(b)(b-b) + \cdots + \frac{f^{(k-1)}(t)}{(k-1)!}(b-b)^{(k-1)} + \frac{M(b-b)^k}{k!} = 0.$$

On the other hand, the definition of M implies $g(a) = 0$. Therefore, Rolle's theorem applies to g, this means there exists $c \in (a, b)$ such that $g'(c) = 0$. Calculate the derivative of g, the minus signs stem from the chain rule applied to the $b - t$ terms,

$$\begin{aligned}
g'(t) &= \frac{d}{dt}\big[-f(b) + f(t)\big] \ + \ \frac{d}{dt}\big[f'(t)(b-t)\big] \ + \cdots + \\
&\quad + \frac{d}{dt}\Big[\frac{f^{(k-1)}(t)}{(k-1)!}(b-t)^{(k-1)}\Big] \ + \ \frac{d}{dt}\Big[\frac{M(b-t)^k}{k!}\Big] \\
&= f'(t) - f'(t) + f''(t)(b-t) - \frac{1}{2}f''(t)2(b-t) \ + \cdots + \\
&\quad + \frac{f^{(k)}(t)}{(k-1)!}(b-t)^{(k-1)} - \frac{f^{(k-1)}(t)}{(k-1)!}k(b-t)^{(k-2)} - \frac{Mk(b-t)^{k-1}}{k!} \\
&= \frac{f^{(k)}(t)}{(k-1)!}(b-t)^{(k-1)} - \frac{Mk(b-t)^{k-1}}{k!} \\
&= \frac{(b-t)^{(k-1)}}{(k-1)!}\Big[f^{(k)}(t) - M\Big]
\end{aligned}$$

where we used that $\frac{k}{k!} = \frac{k}{k(k-1)!} = \frac{1}{(k-1)!}$ in the last step. Note that $c \in (a, b)$ therefore $c \neq b$ hence $(b-t) \neq 0$ hence $(b-t)^{(k-1)} \neq 0$ hence $\frac{(b-t)^{(k-1)}}{(k-1)!} \neq 0$. It follows that $g'(c) = 0$ implies $f^{(k)}(c) - M = 0$ which shows $M = f^{(k)}(c)$ for some $c \in (a, b)$. The proof for the case $b > a$ is similar. $\square$

In total, we see that Taylor's theorem is more or less a simple consequence of Rolle's theorem. In fact, the proof above is not much different than the proof we gave previously for the MVT.

Corollary 5.6.9. *error bound for $T_k(x)$.*

If a function f has $(k+1)$-continuous derivatives on a closed interval $[p,q]$ with length $l = q - p$ and $|f^{(k+1)}(x)| \leq M$ for all $x \in (p,q)$ then for each $a \in (p,q)$

$$|R_k^a(x)| \leq \frac{Ml^{k+1}}{(k+1)!}$$

where $f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j + R_k^a(x)$.

Proof: At each point a we can either look at $[a,x]$ or $[x,a]$ and apply Taylor's theorem to obtain $c_a \in \mathbb{R}$ such that $f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j + R_k^a(x)$ where $R_k^a(x) = \frac{f^{(k+1)}(c_a)}{(k+1)!}(x-a)^{k+1}$. Then we note $|f^{(k+1)}(c_a)| \leq M$ and the corollary follows. $\square$

Consider the criteria for the Second Derivative test. We required that $f'(c) = 0$ and $f''(c) \neq 0$ for a definite conclusion. If f'' is continuous at c with $f''(c) \neq 0$ then it is nonzero on some closed interval $I = [c-\delta, c+\delta]$ where $\delta > 0$. If we also are given that f''' is continuous on I then it follows there exists $M > 0$ such that $|f'''(x)| \leq M$ for all $x \in I$. Observe that

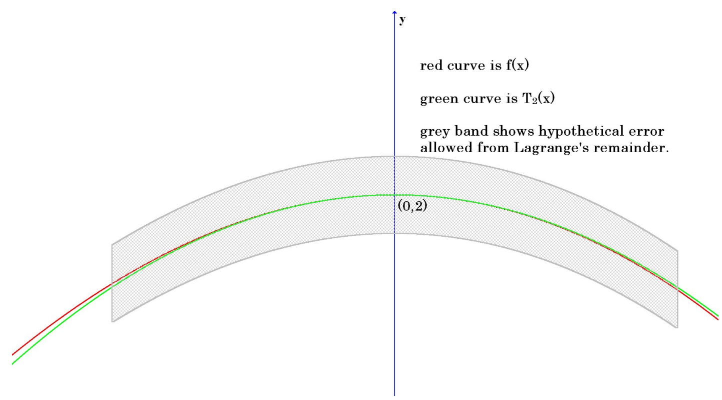
$$|f(x) - f(c) - \frac{1}{2}f''(c)(x-c)^2| = |\frac{1}{6}f'''(\zeta_x)(x-c)^3| \leq \frac{4M\delta^3}{3}$$

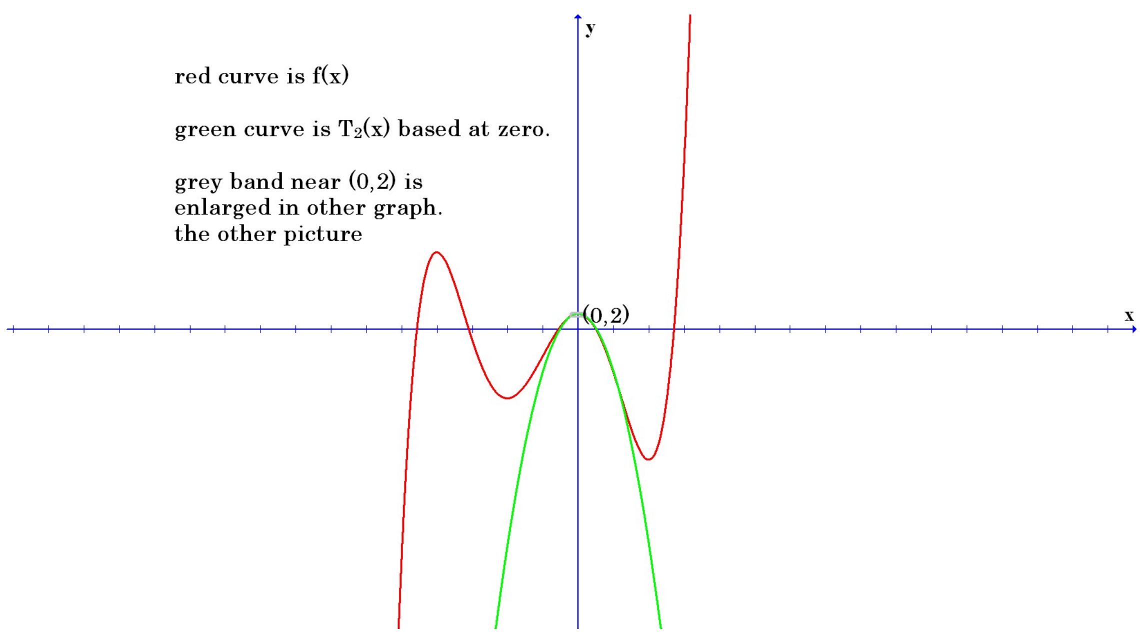
for all $x \in [c-\delta, c+\delta]$. This inequality reveals that we have $f(x) \approx f(c) + \frac{1}{2}f''(c)(x-c)^2$ as $\delta \to 0$. Therefore, locally the graph of the function resembles a parabola which either opens up or down at the critical point. If it opens up ($f''(c) > 0$) then $f(c)$ is the local minimum value of f. If it opens down ($f''(c) < 0$) then $f(c)$ is the local maximum value of f. Of course this is no surprise. However, notice that we may now quantify the error $E_2(x) = |f(x) - T_2(x)| \leq \frac{8M\delta^3}{3}$. If we can choose a bound for $f'''(x)$ independent of x then the error is simply bounded just in terms of the distance from the critical point which we can choose $\delta = |x - c|$ and the resulting error is just $\frac{4M\delta^3}{3}$. Usually, M will depend on the distance from c so the choice of δ to limit error is a bit more subtle. Let me illustrate how this analysis works in an example.

Example 5.6.10. *Suppose* $f(x) = 6x^5 + 15x^4 - 10x^3 - 30x^2 + 2$*. We can calculate that* $f'(x) = 30x^4 + 60x^3 - 30x^2 - 60x$ *therefore clearly* $(0, 2)$ *is a critical point of* f*. Moreover,* $f''(x) = 120x^3 + 180x^2 - 60x - 60$ *shows* $f''(0) = -60$*. I aim to show how the quadratic Taylor polynomial* $T_2(x) = f(2) + f'(2)x + \frac{1}{2}f''(2)x^2 = 2 - 30x^2$ *gives a good approximation for* $f(x)$ *in the sense that the maximum error is essentially bounded by the size of Lagrange's term. Note that*

$$f'''(x) = 360x^2 + 360x - 60 \qquad \textit{and} \qquad f^{(4)}(x) = 720x + 360$$

Suppose we seek to approximate on $-0.1 < x < 0.1$ *then for such* x *we may verify that* $f^{(4)}(x) > 0$ *which means* f''' *is increasing on* $[-0.1, 0.1]$ *thus* $f'''(-0.1) < f'''(x) < f'''(0.1)$ *which gives* $3.6 - 36 - 60 < f'''(x) < 3.6 + 36 - 60$ *thus* $-92.4 < f'''(x) < -20.4$*. Therefore, if* $|x| < 0.1$ *then* $|f'''(x)| < 92.4$*. Using* $\delta = 0.1$ *we should expect a bound on the error of* $\frac{4M\delta^3}{3} = 4(92.4)/3000 = 0.123$*. I have illustrated the global and local qualities of the Taylor Polynomial centered at zero. Notice that the error bound was quite generous in this example.*





Example 5.6.11. *Here we examine Taylor polynomials for $f(x) = \sin(x)$ on the interval $(-1, 1)$ and second on $(-2, 2)$. In each case we use sufficiently many terms to guarantee an error of less than $\epsilon = 0.1$. Notice that $f^{(2k-1)}(x) = \pm\sin(x)$ whereas $f^{(2k-2)}(x) = \pm\cos(x)$ for all $k \in \mathbb{N}$ therefore $|f^{(n)}(x)| \leq 1$ for all $x \in \mathbb{R}$.*

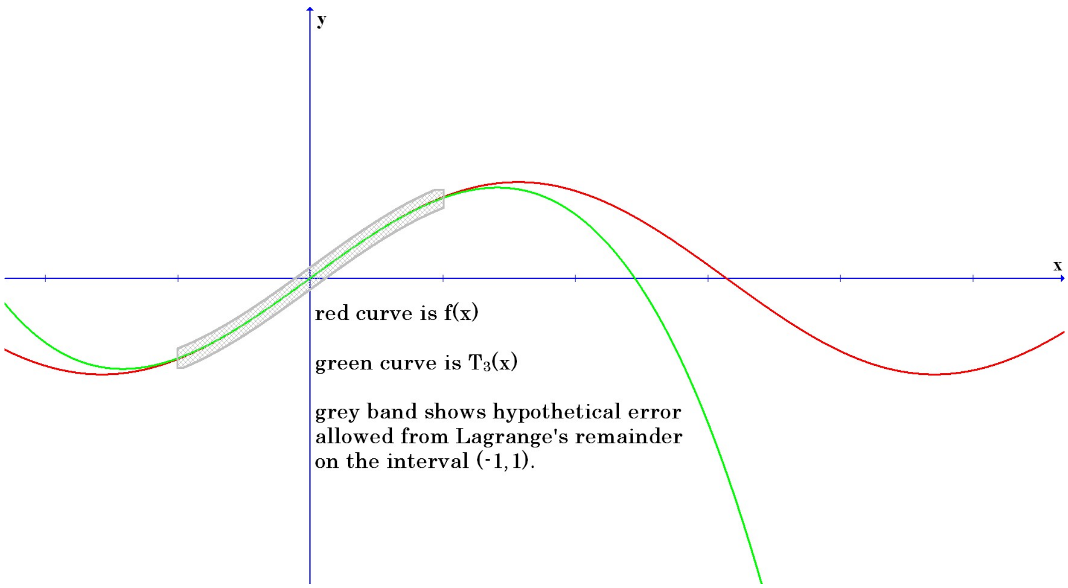
If we wish to bound the error to 0.1 on $-1 < x < 1$ then we to bound the remainder term as follows: (note $-1 < x < 1$ implies $l = 2$ and we just argued $M = 1$ is a good bound for any k)

$$|f(x) - T_k(x)| \leq \frac{Ml^{k+1}}{(k+1)!} = \frac{2^{k+1}}{(k+1)!} = E_k \leq 0.1$$

At this point I just start plugging various values of k until I find a value smaller than the desired bound. For this case,

$$E_1 = \frac{2^2}{2!} = 2, \;\; E_2 = \frac{2^3}{3!} = \frac{4}{3}, \;\; E_3 = \frac{2^4}{4!} = \frac{2}{3}, \;\; E_4 = \frac{2^5}{5!} = \frac{32}{120} \approx 0.25, \;\; E_5 = \frac{2^6}{6!} = \frac{64}{720} \approx 0.1$$

This shows that $T_4(x)$ will provide the desired accuracy. But, it just so happens that $T_3 = T_4$ in this case so we find $T_3(x) = x - \frac{1}{6}x^3$ will suffice. In fact, it fits the ± 0.1 tolerance band quite nicely:



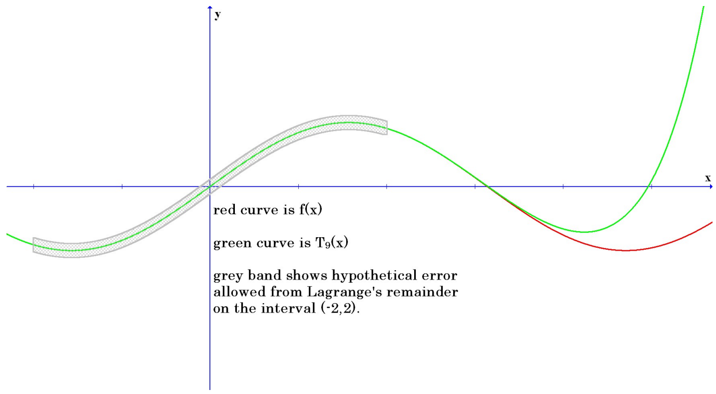
If we wish to bound the error to 0.1 on $-2 < x < 2$ then we to bound the remainder term as follows: (note $-2 < x < 2$ implies $l = 4$)

$$|f(x) - T_k(x)| \leq \frac{Ml^{k+1}}{(k+1)!} = \frac{4^{k+1}}{(k+1)!} = E_k \leq 0.1$$

At this point I just start plugging various values of k until I find a value smaller than the desired bound. For this case,

$$E_7 = \frac{4^8}{8!} \approx 1.6, \;\; E_9 = \frac{4^{10}}{10!} \approx 0.3, \;\; E_{11} = \frac{2^{12}}{12!} \approx 0.035$$

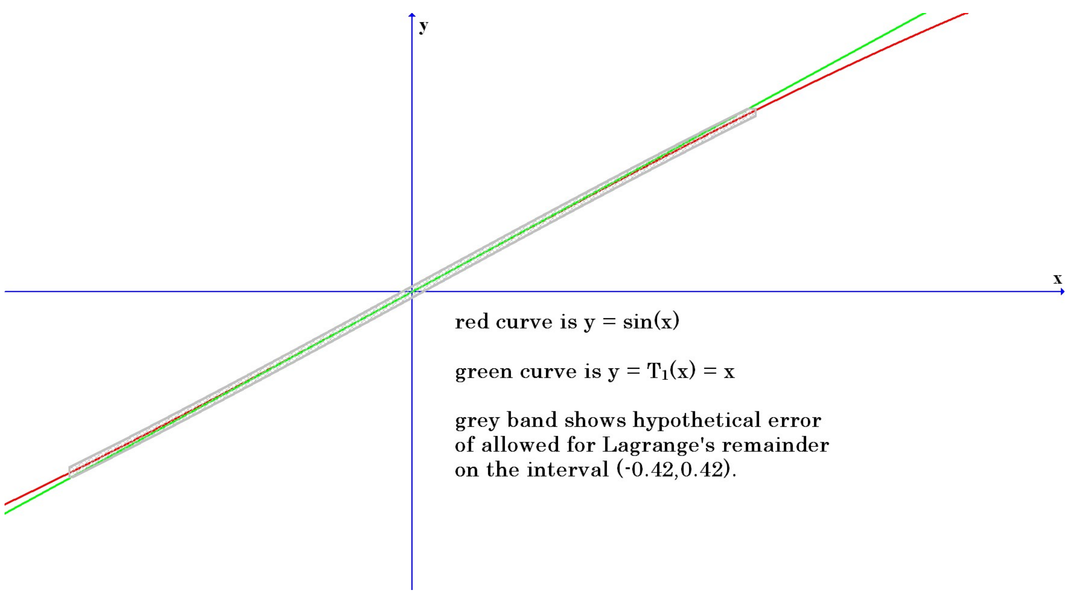
This shows that $T_{10}(x)$ will provide the desired accuracy. But, it just so happens that $T_9 = T_{10}$ in this case so we find $T_9(x) = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \frac{1}{5040}x^7 + \frac{1}{362880}x^9$ will suffice. In fact, as you can see below it fits the ± 0.1 tolerance band quite nicely well beyond the target interval of $-2 < x < 2$:



Example 5.6.12. *Let's think about $f(x) = \sin(x)$ again. This time, answer the following question: for what domain $-\delta < x < \delta$ will $f(x) \approx x$ to within ± 0.01 ? We can use $M = 1$ and $l = 2\delta$. Furthermore, $T_1(x) = T_2(x) = x$ therefore we want*

$$|f(x) - x| \leq \frac{(2\delta)^3}{(3!} = \frac{4\delta^3}{3} \leq 0.1$$

to hold true for our choice of δ. Hence $\delta^3 \leq 0.075$ which suggests $\delta \leq 0.42$. Taylor's theorem thus shows $\sin(x) \approx x$ to within ± 0.01 provided $-0.42 < x < 0.42$. (0.42 radians translates into about 24 degrees). Here's a picture of $f(x) = \sin(x)$ (in red) and $T_1(x) = x$ (in green) as well as the tolerance band (in grey). You should recognize $y = T_1(x)$ as the tangent line.



Example 5.6.13. *Suppose we are faced with the task of calculating $\sqrt{4.03}$ to an accuracy of 5-decimals. For the purposes of this example assume all calculators are evil. It's after the robot holocaust so they can't be trusted. What to do? We use the Taylor polynomial up to quadratic order: we have $f(x) = \sqrt{x}$ and $f'(x) = \frac{1}{2\sqrt{x}}$ and $f''(x) = \frac{-1}{4(\sqrt{x})^3}$. Apply Taylor's theorem,*

$$\begin{aligned}
\sqrt{4.03} &= f(4) + f'(4)(4.03-4) + \frac{1}{2}f''(4)(4.03-4)^2 + R \\
&= 2 + \frac{1}{4}\frac{3}{100} - \frac{1}{64}\frac{9}{10000} + R \\
&= 2 + 0.0075 - 0.000014062 + R \\
&= 2.\underline{007485}938 + R
\end{aligned}$$

If we bound $f'''(x) = \frac{3}{8(\sqrt{x})^5}$ by M on $[4, 4.03]$ then $|R| \leq \frac{M(0.03)^3}{6}$. Clearly $f''''(x) = \frac{-15}{16(\sqrt{x})^7} < 0$ for $x \in [4, 4.03]$ therefore, f''' is decreasing on $[4, 4.03]$. It follows $f'''(4) \geq f'''(x) \geq f'''(4.03)$. Choose $M = f'''(4) = \frac{3}{8(32)} = \frac{3}{256}$ thus

$$|R| \leq \frac{(0.03)^3}{6}\frac{3}{256} = \frac{27}{256}\frac{1}{10000} \approx \frac{1}{100000} = 0.000001.$$

Therefore, $\boxed{\sqrt{4.03} = 2.007486 \pm 0.000001}$. As far as I know my TI-89 is still benevolent so we can check our answer; the calculator says $\sqrt{4.03} = 2.\underline{007485}98999$.

In the last example, we again find that we actually are a whole digit closer to the answer than the error bound suggests. This seems to be typical. In calculus II we'll find a better error bound in the study of power series.

Example 5.6.14. *Newton postulated that the gravitational force between masses m and M separated by a distance of r is*

$$\vec{F} = -\frac{GmM}{r^2}\hat{r}$$

where r is the distance from the center of mass of M to the center of mass m and G is a constant which quantifies the strength of gravity. The minus sign means gravity is always attractive in the direction $\hat{r}$ which points along the line from M to m. Consider a particular case, M is the mass of the earth and m is a small mass a distance r from the center of the earth. It is convenient to write $r = R + h$ where R is the radius of the earth and h is the **altitude** *of m. Here we make the simplifying assumptions that m is a point mass and M is a spherical mass with a homogeneous mass distribution. It turns out that means we can idealize M as a point mass at the center of the earth. All of this said, you may recall that $F = mg$ is the force of gravity in highschool physics where the force points down. But, this is very different then the inverse square law? How are these formulas connected? Focus on a particular ray eminating from the center of the earth so the force depends only on the altitude h. In particular:*

$$F(h) = -\frac{GmM}{(R+h)^2}$$

We calculate,

$$F'(h) = \frac{2GmM}{(R+h)^3}$$

Note that clearly $F''(h) < 0$ hence F' is a decreasing function of h therefore if $0 \leq h \leq h_{max}$ then $F'(0) \geq F'(h) \geq F'(h_{max})$ so $F'(0)$ provides a bound on $F'(h)$. Calculate that

$$F(0) = -\frac{GmM}{R^2} \text{ and } F'(0) = \frac{2GmM}{R^3}$$

Taylor's theorem says that $F(h) = F(0) + E$ and $|E| \leq F'(0)h_{max}$ therefore,

$$F(h) \approx -\frac{GmM}{R^2} \pm \frac{2GmM}{R^3}h$$

Note $G = 6.673 \times 10^{-11}\frac{Nm^2}{kg^2}$ and $R = 6.3675 \times 10^6 m$ and $M = 5.972 \times 10^{24} kg$. You can calculate that $\frac{GmM}{R^2} = 9.83 m/s^2$ which is hopefully familar to some who read this. In contrast, the error term

$$|E| = \frac{2GmM}{R^3}h = (3.1 \times 10^{-6})mh$$

If the altitude doesn't exceed $h = 1,000m$ then the formula $F/m = g$ approximates the true inverse square law to within $0.0031 m/s^2$. At $h = 10,000m$ the error is $0.031 m/s^2$. At $h = 100,000m$ the error is around $0.31 m/s^2$. (100,000 meters is about 60 miles, well above most planes flight ceiling). Taylor's theorem gives us the mathematical tools we need to quantify such nebulous phrases as $F = mg$ "near" the surface of the earth. Mathematically, this is probably the most boring Taylor polynomial you'll ever study, it was just the constant term.

Remark 5.6.15. *transcendental numbers and a look ahead to calculus II.*

Another application of Taylor's theorem is in calculation of transcendental numbers such as π or e. See Apostol pg. 285 problem 10 for a method to approximate π to seven decimals. Or page 281 for the calculation of e to 8 decimal places. On page 282 in Example 2 a proof is offered for the irrationality of e. To be frank, you don't really understand what a real number is until you understand the construction and convergence/divergence of power series. The idea of an unending decimal expansion really has no justification in the mathematics we have thus far discussed. Fortunately most of you will take calculus II so at least then you'll actually learn how to carefully formulate what is required for an unending sum to be reasonable. The idea of a series provides a careful meaning for a sum of infinitely many things. We'll explain why $0.1111... = \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \cdots$ is a real number whereas $\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots$ is not. Taylor's theorem plays an important role in the study of power series. But, as you hopefully see by now it is also useful for gaining deeper insight into the geometry and local behavior of functions.

5.6.2 higher derivative tests

We saw in the previous section that the second derivative test is concretely justified by Taylor's theorem with Lagrange's remainder. The next logical step is the following theorem which is justified by similar analysis. Basically the point is that if you have all the derivatives zero up to some particular order, say $k-1$, then the function $f(x) \approx T_k(x)$ provided x is close to the critical point. Therefore, if k is an even integer then the function is *locally*-shaped like a parabola whereas if k is odd then is *locally*-shaped like a cubic. Hence the following theorem:

Theorem 5.6.16. *higher derivative tests.*

Suppose f has k continuous derivatives such that $f'(c) = f''(c) = \cdots = f^{(k-1)}(c) = 0$ and $f^{(k)}(c) \neq 0$ then

1. if $k \in 2\mathbb{N}$ and $f^{(k)}(c) > 0$ then $f(c)$ is a local minimum.
2. if $k \in 2\mathbb{N}$ and $f^{(k)}(c) < 0$ then $f(c)$ is a local maximum.
3. if $k \in 2\mathbb{N}+1$ then $f(c)$ is not an extrema.

The notation $k \in 2\mathbb{N}$ means that there exists $n \in \mathbb{N}$ such that $k = 2n$. Likewise, the notation $k \in 2\mathbb{N}+1$ means that there exists $n \in \mathbb{N}$ such that $k = 2n+1$. In other words, $2\mathbb{N} = \{2, 4, 6, \dots\}$ whereas $2\mathbb{N}+1 = \{3, 5, 7, \dots\}$. The proof of this theorem is suggested by the examples and general comments about Taylor polynomials and their remainders. However, if you would like to see an explicit proof you can consult C.H. Edwards, Jr. *Advanced Calculus of Several Variables* pages 125-127.

Example 5.6.17. *Consider $f(x) = x^4$. We can calculate $f'(x) = 4x^3$ therefore the only critical number is $c = 0$. Note that $f''(x) = 12x^2$, $f'''(x) = 24x$, $f^{(4)}(x) = 24$. It follows that*

$$f'(0) = f''(0) = f'''(0) = 0$$

but $f^{(4)}(x) = 24 > 0$ therefore, by the higher derivative test, $f(0) = 0$ is a local minimum of $f(x) = x^4$. Notice that this example would not have been covered by the second derivative test (but, the first derivative test would have covered it).

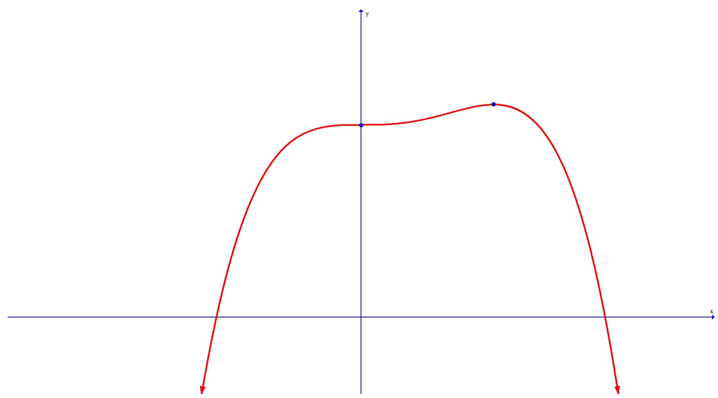
In-Class Example 5.6.18. *Consider $f(x) = x^5$. Analyze $x = 0$ with the higher derivative test.*

Example 5.6.19. *Consider $f(x) = x^3 - x^4 + 1$. We can calculate $f'(x) = 3x^2 - 4x^3 = x^2(3 - 4x)$ thus critical numbers are $c = 0$ and $c = 3/4$. Note that $f''(x) = 6x - 12x^2$, $f'''(x) = 6 - 24x$, $f^{(4)}(x) = -24$. It follows that*

$$f'(0) = f''(0) = 0$$

but $f^{(3)}(x) = 6 \neq 0$ therefore, by the higher derivative test, $f(0) = 3$ is a not a local extrema of $f(x) = x^3 - x^4 + 3$. Continuing to the other critical point notice $f'(3/4) = 0$, $f''(3/4) = 18/4 - 12(3/4)^2 = -9/4$ thus by the second derivative test $f(3/4)$ is a local maximum.

What is the difference between these critical points geometrically? Notice that $y = f''(x) = 6x - 12x^2 = 6x(1 - 2x)$ is a downward opening parabola with zeros at $x = 0$ and $x = 1/2$ therefore we deduce $f''(x) < 0$ for $x < 0$ and $f''(x) > 0$ for $0 < x < 1/2$. This means that $(0, 1)$ is an inflection point of $y = f(x)$. For that reason this example could not be covered by the second derivative test. In contrast, the concavity is downward on a nbhd around $c = 3/4$.



Chapter 6

antiderivatives and the area problem

Let me begin by defining the terms in the title:

1. an antiderivative of f is another function F such that $F' = f$.
2. the area problem is: "find the area of a shape in the plane"

This chapter is concerned with understanding the area problem and then solving it through the fundamental theorem of calculus(FTC).

We begin by discussing antiderivatives. At first glance it is not at all obvious this has to do with the area problem. However, antiderivatives do solve a number of interesting physical problems so we ought to consider them if only for that reason. The beginning of the chapter is devoted to understanding the type of question which an antiderivative solves as well as how to perform a number of basic indefinite integrals. Once all of this is accomplished we then turn to the area problem.

To understand the area problem carefully we'll need to think some about the concepts of finite sums, sequences and limits of sequences. These concepts are quite natural and we will see that the theory for these is easily transferred from some of our earlier work. Once the limit of a sequence and a number of its basic properties are established we then *define area* and the definite integral. Finally, the remainder of the chapter is devoted to understanding the fundamental theorem of calculus and how it is applied to solve definite integrals.

I have attempted to be rigorous in this chapter, however, you should understand that there are superior treatments of integration(Riemann-Stieltjes, Lesbesgue etc..) which cover a greater variety of functions in a more logically complete fashion. The treatment here is more or less typical of elementary calculus texts.

6.1 indefinite integration

Don't worry, the title of this section will make sense later.

6.1.1 why antidifferentiate?

The antiderivative is the opposite of the derivative in the following sense:

Definition 6.1.1. *antiderivative.*

> If f and F are functions such that $F' = f$ then we say that F is an **antiderivative** of f.

Example 6.1.2. *Suppose $f(x) = x$ then an antiderivative of f is a function F such that $\frac{dF}{dx} = x$. We could try x^2 but then $\frac{d}{dx}(x^2) = 2x$ has an unwanted factor of 2. What to do? Just adjust our guess a little: try $F(x) = \frac{1}{2}x^2$. Note that $\frac{d}{dx}(\frac{1}{2}x^2) = \frac{1}{2}\frac{d}{dx}(x^2) = \frac{1}{2}(2x) = x$.*

Example 6.1.3. *Let k be a constant. Suppose $g(t) = e^{kt}$ then we guess $G(t) = \frac{1}{k}e^{kt}$ and note it works; $\frac{d}{dt}(\frac{1}{k}e^{kt}) = e^{kt}$ therefore $g(t) = e^{kt}$ has antiderivative $G(t) = \frac{1}{k}e^{kt}$.*

In-Class Example 6.1.4. *Suppose $h(\theta) = \cos(\theta)$. What is $H(\theta)$?*

In-Class Example 6.1.5. *Suppose $g(\theta) = \sin(\theta)$. What is $G(\theta)$?*

Obviously these guesses are not random. In fact, these are educated guesses. We simply have to think about how we differentiated before and just try to think backwards. Simple enough for now. However, we should stop to notice that the antiderivative is far from unique. You can easily check that $F(x) = \frac{1}{2}x^2 + c_1$, $G(t) = \frac{1}{k}e^{kt} + c_2$ and $H(\theta) = \sin(\theta) + c_3$ are also antiderivatives for any constants $c_1, c_2, c_3 \in \mathbb{R}$.

Proposition 6.1.6. *antiderivatives differ by at most a constant.*

> If f has antiderivatives F_1 and F_2 then there exists $c \in \mathbb{R}$ such that $F_1(x) = F_2(x) + c$.

Proof: We are given that $\frac{dF_1}{dx} = f(x)$ and $\frac{dF_2}{dx} = f(x)$ therefore $\frac{dF_1}{dx} = \frac{dF_2}{dx}$. Hence, by Proposition 5.1.17 we find $F_1(x) = F_2(x) + c$. $\square$

To understand the significance of this constant we should consider a physical question.

Example 6.1.7. *Suppose that the velocity of a particle at position x is measured to be constant. In particular, suppose that $v(t) = \frac{dx}{dt}$ and $v(t) = 1$. The condition $v(t) = \frac{dx}{dt}$ means that x should be an antiderivative of v. For $v(t) = 1$ the form of all antiderivatives is easy enough to guess: $x(t) = t + c$. The value for c cannot be determined unless we are given additional information about this particle. For example, if we also knew that at time zero the particle was at $x = 3$ then we could fit this* **initial data** *to pick a value for c:*

$$x(0) = 0 + c = 3 \quad \Rightarrow \quad c = 3 \quad \Rightarrow \quad \boxed{x(t) = t + 3}.$$

In-Class Example 6.1.8. *Suppose velocity $v(t) = \frac{dx}{dt} = t^2 + e^t$. If $x(0) = 1$ then find $x(t)$.*

For a given velocity function each antiderivative gives a possible position function. To determine the precise position function we need to know both the velocity and some initial position. Often we are presented with a problem for which we do not know the initial condition so we'd like to have a mathematical device to leave open all possible initial conditions.

Definition 6.1.9. *indefinite integral.*

If f has an antiderivative F then the **indefinite integral** of f is given by:

$$\int f(x)dx = \{G(x) \mid G'(x) = f(x)\} = \{F(x) + c | c \in \mathbb{R}\}.$$

However, we will customarily drop the set-notation and simply write

$$\int f(x)dx = F(x) + c \;\; \text{where } F'(x) = f(x).$$

The indefinite integral includes all possible antiderivatives for the given function. Technically the indefinite integral is not a function. Instead, it is a family of functions each of which is an antiderivative of f.

Example 6.1.10. *Consider the constant acceleration problem[1]; we are given that $a = -g$ where $g = 9.8m/s^2$ and $a = \frac{dv}{dt}$. We can take the indefinite integral of the equation:*

$$\frac{dv}{dt} = -g \quad \Rightarrow \quad v(t) = \int -g\, dt = -gt + c_1.$$

[1] here $F = ma$ is $-mg = ma$ so $a = -g$ but that's physics, I supply the equation of motion in calculus. You just have to do the math.

Furthermore, if $v = \frac{dy}{dt}$ *then*

$$\frac{dy}{dt} = -gt + c_1 \quad \Rightarrow \quad y(t) = \int -gt + c_1 \, dt = -\frac{1}{2}gt^2 + c_1 t + c_2.$$

Therefore, we find the velocity and position are given by formulas

$$v(t) = c_1 - gt \qquad y(t) = c_2 + c_1 t - \frac{1}{2}gt^2.$$

If we know the initial velocity is v_o *and the initial position is* y_o *then*

$$v(0) = v_o = c_1 - 0 \quad \Rightarrow \quad \boxed{v(t) = v_o - gt}$$

$$y(0) = y_o = c_2 - 0 - 0 \quad \Rightarrow \quad \boxed{y(t) = y_o + v_o t - \frac{1}{2}gt^2}$$

These formulas were derived by Galileo without the benefit of calculus. Instead, he used experiment and a healthy skepticism of the philosophical nonsense of Aristotle. The ancient Greek's theory of motion said that if something was twice as heavy then it falls twice as fast. This is only true when the objects compared have air friction clouding the dynamics. The equations above say the objects' motion is independent of the mass.

Remark 6.1.11. *redundant comment (again).*

The indefinite integral is a family of antiderivatives: $\int f(x) = F(x) + c$ where $F'(x) = f(x)$. The following equation shows how indefinite integration is undone by differentiation:

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

the function f is called the **integrand** and the variable of indefinite integration is x. Notice the constant is obliterated by the derivative in the equation above. Leibniz' notation intentionally makes you think of cancelling the dx's as if they were tiny quantities. Newton called them *fluxions*. In fact calculus was sometimes called the *theory of fluxions* in the early 19-th century. Newton had in mind that dx was the change in x over a tiny time, it was a fluctuation with respect to a time implicit. We no longer think of calculus in this way because there are easier ways to think about foundations of calculus. That said, it is still an intuitive notation and if you are careful not to overextend intuition it is a powerful mnemonic. For example, the chain rule $\frac{df}{dx} = \frac{df}{du}\frac{du}{dx}$. Is the chain rule just from multiplying by one? No. But, it is a nice way to remember the rule.

A differential equation is an equation which involves derivatives. We have solved a number of differential equations in this section via the process of indefinite integration. The example that follows doesn't quite fit the same pattern. However, I will again solve it by educated guessing[2] .

[2] Actually, the method I use here is rather unusual but the advanced reader will recognize the idea from differential equations. The easier way of solving this is called separation of a variables, but we discuss that method much later

Example 6.1.12. *A simple model of population growth is that the rate of population growth should be directly proportional to the size of the population P. This means there exists $k \in \mathbb{R}$ such that*

$$\frac{dP}{dt} = kP.$$

Fortunately, we just did Example 6.1.3 where we observed that

$$\int e^{kt}\, dt = \frac{1}{k}e^{kt} + c$$

So we know that one solution is given by $P(t) = \frac{1}{k}e^{kt}$. Change variables by substituting $u = \ln(P)$ so $\frac{du}{dt} = \frac{1}{P}\frac{dP}{dt}$ thus $\frac{dP}{dt} = P\frac{du}{dt}$. Hence we can solve $P\frac{du}{dt} = kP$ or $\frac{du}{dt} = k$ instead. This we can antidifferentiate to find $u(t) = kt + c_1$. Thus, $\ln(P) = kt + c_1$ hence $P(t) = e^{kt+c_1} = e^{c_1}e^{kt}$. If the initial population is given to be P_o then we find $P(0) = P_o = e^{c_1}$ thus $\boxed{P(t) = P_o e^{kt}.}$

The same mathematics govern simple radioactive decay, continuously compounded interest, current or voltage in an LR or RC circuit and a host of other simplistic models in the natural sciences. Real human population growth involves many factors beyond just raw population, however for isolated systems this type of model does well. For example, growth of bacteria in a petri dish.

Remark 6.1.13. *why antidifferentiate?*

We antidifferentiate to solve simple differential equations. When one variable (say v) is the instantaneous rate of change of another (say s so $v = \frac{ds}{dt}$) then we can reverse the process of differentiation to discover the formula of s if we are given the formula for v. However, because constants are lost in differentiation we also need an initial condition if we wish to uniquely determine the formula for s. I have emphasized the utility of the concept of antidifferentiation as it applies to physics, but that was just my choice.

Notice, I have yet to even discuss the area problem. We already see that indefinite integration is an important skill to master. The methods I have employed in this section are ad-hoc. We would like a more systematic method. I offer organization for guessing in the next section.

6.1.2 properties of indefinite integration

In this section we list all the basic building blocks for indefinite integration. Some of these we already guessed in specific examples. If you need to see examples you can skip ahead to the section that follows this one.

Proposition 6.1.14. *basic properties of indefinite integration.*

Suppose f, g are functions with antiderivatives and $c \in \mathbb{R}$ then

$$\int [f(x) + g(x)]dx = \int f(x)\, dx + \int g(x)\, dx$$

$$\int cf(x)\, dx = c\int f(x)\, dx$$

Proof: Suppose $\int f(x)\,dx = F(x) + c_1$ and $\int g(x)\,dx = G(x) + c_2$ note that

$$\frac{d}{dx}[F(x) + G(x)] = \frac{d}{dx}[F(x)] + \frac{d}{dx}[G(x)] = f(x) + g(x)$$

hence $\int [f(x) + g(x)]dx = F(x) + G(x) + c_3 = \int f(x)\,dx + \int g(x)\,dx$ where the constant c_3 is understood to be included in either the $\int f(x)\,dx$ or the $\int g(x)\,dx$ integral as a matter of custom. $\square$

Proposition 6.1.15. *power rule for integration. suppose $n \in \mathbb{R}$ and $n \neq -1$ then*

$$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + c.$$

Proof: $\frac{d}{dx}[\frac{1}{n+1}x^{n+1}] = \frac{n+1}{n+1}x^{n+1-1} = x^n$. Note that $n+1 \neq 0$ since $n \neq -1$. $\square$

Note that the special case of $n = -1$ stands alone. You should recall that $\frac{d}{dx}\ln(x) = \frac{1}{x}$ provided $x > 0$. In the case $x < 0$ then by the chain rule applied to the positive case: $\frac{d}{dx}\ln(-x) = \frac{1}{-x}(-1) = \frac{1}{x}$. Observe then that for all $x \neq 0$ we have $\frac{d}{dx}\ln|x| = \frac{1}{x}$. Therefore the proposition below follows:

Proposition 6.1.16. *reciprocal function is special case.*

$$\int \frac{1}{x}\,dx = \ln|x| + c.$$

Note that it is common to move the differential into the numerator of such expressions. We could just as well have written that $\int \frac{dx}{x} = \ln|x| + c$. I leave the proof of the propositions in the remainder of this section to the reader. They are not difficult.

Proposition 6.1.17. *exponential functions. suppose $a > 0$ and $a \neq 1$,*

$$\int a^x\,dx = \frac{1}{\ln(a)}a^x + c \qquad \text{in particular:} \qquad \int e^x\,dx = e^x + c$$

The exponential function has base $a = e$ and $\ln(e) = 1$ so the formulas are consistent.

Proposition 6.1.18. *trigonometric functions.*

$$\begin{aligned}
&\int \sin(x)\,dx = -\cos(x) + c && \int \cos(x)\,dx = \sin(x) + c \\
&\int \sec^2(x)\,dx = \tan(x) + c && \int \sec(x)\tan(x)\,dx = \sec(x) + c \\
&\int \csc^2(x)\,dx = -\cot(x) + c && \int \csc(x)\cot(x)\,dx = -\csc(x) + c.
\end{aligned}$$

You might notice that many trigonometric functions are missing. For example, how would you calculate[3] $\int \tan(x)\,dx$? We do not have the tools for that integration at this time. For now we are simply cataloguing the basic antiderivatives that stem from reading basic derivative rules backwards.

Proposition 6.1.19. *hyperbolic functions.*

$$\int \sinh(x)\,dx = \cosh(x) + c \qquad \int \cosh(x)\,dx = \sinh(x) + c$$

Naturally there are also basic antiderivatives for $sech^2(x), sech(x)tanh(x), csch^2(x)$ and $csch(x)coth(x)$ however I omit them for brevity and also as to not antagonize the struggling student at this juncture.

Proposition 6.1.20. *special algebraic and rational functions*

$$\int \frac{dx}{1+x^2} = \tan^{-1}(x) + c \qquad \int \frac{dx}{\sqrt{1-x^2}} = \sin^{-1}(x) + c.$$

$$\int \frac{dx}{\sqrt{x^2-1}} = \cosh^{-1}(x) + c \qquad \int \frac{dx}{\sqrt{1+x^2}} = \sinh^{-1}(x) + c.$$

$$\int \frac{dx}{1-x^2} = \tanh^{-1}(x) + c.$$

Recall Example 1.5.6 explored alternate formulas for inverse hyperbolics.

6.1.3 examples of indefinite integration

Example 6.1.21.

$$\int dx = \int x^0 dx = \boxed{x + c}$$

Example 6.1.22.

$$\int \left(\sqrt{x} + \frac{1}{\sqrt[3]{x}}\right) dx = \int x^{\frac{1}{2}} dx + \int x^{\frac{-1}{3}}\,dx = \boxed{\frac{2}{3}x^{\frac{3}{2}} + \frac{3}{2}x^{\frac{2}{3}} + c}$$

In-Class Example 6.1.23. *Calculate* $\int \sqrt{13x^7}dx$.

[3] the answer is $\ln|\sec(x)| + c$ if you're curious and impatient.

Example 6.1.24.

$$\int \frac{dx}{3x^2} = \frac{1}{3}\int x^{-2}dx = \frac{-1}{3}x^{-1} = \boxed{\frac{-1}{3x} + c}$$

Example 6.1.25.

$$\int \frac{2x dx}{x^2} = 2\int \frac{dx}{x} = 2\ln|x| + c = \boxed{\ln(x^2) + c}$$

Note that $|x| = \pm x$ *thus* $|x|^2 = (\pm x)^2 = x^2$ *so it was logical to drop the absolute value bars after bringing in the factor of two by the property* $\ln(A^c) = c\ln(A)$.

Example 6.1.26.

$$\int 3e^{x+2}dx = 3\int e^2 e^x dx = 3e^2\int e^x dx = 3e^2(e^x + c_1) = \boxed{3e^{x+2} + c}$$

In-Class Example 6.1.27. *Calculate* $\int(2x^3 + 3)dx$.

In-Class Example 6.1.28. *Calculate* $\displaystyle\int \frac{2x^3+3}{x}dx$.

In-Class Example 6.1.29. *Calculate* $\displaystyle\int (x+2)^2 dx$.

Example 6.1.30.

$$\int(2^x + 3\cosh(x))dx = \int 2^x dx + 3\int \cosh(x)dx = \boxed{\frac{1}{\ln(2)}2^x + 3\sinh(x) + c}$$

Example 6.1.31.

$$\int \frac{x^2}{1+x^2}dx = \int \frac{1+x^2-1}{1+x^2}dx = \int [1 - \frac{1}{1+x^2}]dx = \boxed{x - \tan^{-1}(x) + c}$$

Example 6.1.32.

$$\begin{aligned}
\int \sin(x+3)dx &= \int \big[\sin(x)\cos(3) + \sin(3)\cos(x)\big]dx \\
&= \cos(3)\int \sin(x)dx + \sin(3)\int \cos(x)dx \\
&= -\cos(3)[\cos(x) + c_1] + \sin(3)[\sin(x) + c_2] \\
&= \sin(3)\sin(x) - \cos(3)\cos(x) + c \\
&= \boxed{-\cos(x+3) + c}
\end{aligned}$$

Incidentally, we find a better way to do this later with the technique of u-substitution.

In-Class Example 6.1.33. *Calculate* $\displaystyle\int \frac{1}{\cos^2(x)}dx$.

Example 6.1.34.

$$\int \frac{dx}{x^2 + \cos^2(x) + \sin^2(x)} = \int \frac{dx}{x^2+1} = \boxed{\tan^{-1}(x) + c}$$

In-Class Example 6.1.35. *Calculate* $\displaystyle\int \frac{\sqrt{x^2-1}}{(x+1)(x-1)}dx$.

6.2 area problem

The area of a general shape in the plane can be approximately calculated by dividing the shape into a bunch of rectangles or triangles. Since we know how to calculate the area of a rectangle $[A = lw]$ or a triangle $[A = \frac{1}{2}bh]$ we simply add together all the areas to get an approximation of the total area. In the special case that the shape has flat sides then we can find the exact area since any shape with flat sides can be subdivided into a finite number of triangles. Generally shapes have curved edges so no finite number of approximating rectangles or triangles will capture the exact area. Archimedes realized this some two milennia ago in ancient Syracuse. He argued that if you could find two approximations of the area one larger than the true area and one smaller than the true area then you can be sure that the exact area is somewhere between those approximations. By such squeeze-theorem type argumentation he was able to demonstrate that the value of π must be between $\frac{223}{71}$ and $\frac{22}{7}$ (in decimals $3.1408 < \pi \approxeq 3.1416 < 3.1429$). In Apostol's calculus text he discusses axioms for area and he uses Archimedes' squeezing idea to define both area and definite integrals. Our approach will be less formal and less rigorous.

Our goal in this section is to careful construct a method to calculate the area bounded by a function on some interval $[a, b]$. Since the function could take on negative values in the interval we actually are working on a method to calculate signed area under a graph. Area found beneath the x-axis is counted negative whereas area above the x-axis is counted positive. Shapes more general than those described by the graph of a simple function are treated in the next chapter.

6.2.1 sums and sequences in a nutshell

A sequence is function which corresponds uniquely to an ordered list of values. We consider real-valued sequences but the concept extends to many other objects[4].

Definition 6.2.1. *sequence of real numbers.*

If $U \subseteq \mathbb{Z}$ has a smallest member and the property that $n \in U$ implies $n + 1 \in U$ then a function $f : U \to \mathbb{R}$ is a **sequence**. Moreover, we may denote the sequence by listing its values

$$f = \{f(u_1), f(u_2), f(u_3), \dots\} = \{f_{u_1}, f_{u_2}, f_{u_3}, \dots\} = \{f_{u_j}\}_{j=1}^{\infty}$$

Typically $U = \mathbb{N}$ or $U = \mathbb{N} \cup \{0\}$ and we study sequences of the form

$$\{a_j\}_{j=0}^{\infty} = \{a_0, a_1, a_2, \dots\} \qquad \{b_n\}_{n=1}^{\infty} = \{b_1, b_2, b_3 \dots\}$$

Example 6.2.2. *Sequences may defined by a formula: $a_n = n$ for all $n \in \mathbb{N}$ gives*

$$\{a_n\}_{n=1}^{\infty} = \{1, 2, 3, \dots\}.$$

[4]sequences of functions, matrices or even spaces are studied in modern mathematics

Or by an iterative rule: $f_1 = 1, f_2 = 1$ *then* $f_n = f_{n-1} + f_{n-2}$ *for all* $n \geq 3$ *defines the Fibonacci sequence:*

$$\{f_n\}_{n=1}^{\infty} = \{1, 1, 2, 3, 5, 8, 13, 21, \dots\}.$$

Beyond this we can add, subtract and sometimes divide sequences because a sequence is just a function with a discrete domain.

Definition 6.2.3. *finite sum notation.*

Suppose $a_j \in \mathbb{R}$ for $j \in \mathbb{N}$. Then define:

$$\sum_{j=1}^{1} a_j = a_1 \qquad \sum_{j=1}^{n} a_j = \sum_{j=1}^{n-1} a_j + a_n$$

for $n \geq 2$. This iterative definition gives us the result that

$$\sum_{j=1}^{n} a_j = a_1 + a_2 + \cdots + a_n.$$

The variable j is called the **dummy index of summation**. Moreover, sums such as

$$\sum_{j=j_1}^{j_N} a_j = \underbrace{a_{j_1} + a_{j_2} + \cdots + a_{j_N}}_{N \text{ summands}}$$

can be carefully defined by a similar iterative formula.

Example 6.2.4. *Sums can give particularly interesting sequences. Consider* $a_n = \sum_{j=1}^{n} j$ *for* $n = 1, 2 \dots$.

$$\{a_n\}_{n=1}^{\infty} = \{1, 1+2, 1+2+3, 1+2+3+4, \dots\} = \{1, 3, 6, 10, \dots\}.$$

The greatest mathematician of the 19-th century is generally thought to be Gauss. As a child Gauss was tasked with computing a_{100}. *The story goes that just as soon as the teacher asked for the children to calculate the sum Gauss wrote the answer* 5050 *on his slate. How did he know how to calculate the sum* $1 + 2 + 3 + \cdots + 50$ *with such ease? Gauss understood that generally*

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$$

For example,

$$a_1 = \frac{1(1+1)}{2} = 1, \quad a_2 = \frac{2(2+1)}{2} = 3, \quad a_3 = \frac{3(3+1)}{2} = 6,$$
$$a_4 = \frac{4(4+1)}{2} = 10, \quad \dots, a_{100} = \frac{(100)(101)}{2} = 50(101) = 5050.$$

What method of proof is needed to prove results such as this? The method is called "proof by mathematical induction". In short, the idea is this: you prove the result you interested in is true for $n = 1$ then you prove that if n is true then $n + 1$ is also true for an arbitrary $n \in \mathbb{N}$. Let's see how this plays out for the preceding example:

Proof of Gauss' Formula by induction: note that $n = 1$ is clearly true since $a_1 = 1$. Assume that $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$ $(\star)$ is valid and consider that, by the recursive definition of the finite sum,

$$\sum_{j=1}^{n+1} j = \sum_{j=1}^{n} j + n + 1 = \underbrace{\frac{n(n+1)}{2}}_{\text{using } \star} + n + 1 = \frac{1}{2}(n^2 + 3n + 2) = \frac{([n+1])([n+1]+1)}{2}$$

which is precisely the claim for $n + 1$. Therefore, by proof by mathematical induction, Gauss' formula is true for all $n \in \mathbb{N}$. $\square$

Formulas for simple sums such as $\sum 1, \sum n, \sum n^2, \sum n^3$ are also known and can be proven via induction. Let's collect these results for future reference:

Proposition 6.2.5. *special formulas for finite sums.*

$$\sum_{k=1}^{n} 1 = n, \qquad \sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \qquad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}, \qquad \sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}.$$

The following results also follow from inductive arguments. These are very useful:

Proposition 6.2.6. *finite sum properties. suppose $a_k, b_k, c \in \mathbb{R}$ for all k and let $n, m \in \mathbb{N}$ such that $m < n$,*

$$\begin{aligned}
&(i.)\ \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k = \sum_{k=1}^{n} (a_k + b_k), \\
&(ii.)\ \sum_{k=1}^{n} c a_k = c \sum_{k=1}^{n} a_k, \\
&(iii.)\ \sum_{k=1}^{n} a_k = \sum_{k=1}^{m} a_k + \sum_{k=m+1}^{n} a_k.
\end{aligned}$$

We would like to have sums with $n \to \infty$ in the sections that follow. The definition that follows is essentially the same we gave previously for functions of a continuous variable. The main difference is that only integers are considered in the limiting process.

Definition 6.2.7. *limit of a sequence.*

We say the sequence $\{a_n\}$ converges to $L \in \mathbb{R}$ and denote

$$\lim_{n\to\infty} a_n = L$$

iff for each $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n > N$ we find $|a_n - L| < \epsilon$

The skills you developed in studying functions of a continuous variable transfer to the study of sequential limits because of the following fundamental lemma:

Lemma 6.2.8. *correspondence of limits of functions on $\mathbb{R}$ and sequences.*

Suppose $\{a_n\}$ is a sequence and f is a function such that $f(n) = a_n$ for all $n \in dom(\{a_n\})$. If $\lim_{x\to\infty} f(x) = L \in \mathbb{R}$ then $\lim_{n\to\infty} a_n = L$.

Proof: assume $\lim_{x\to\infty} f(x) = L \in \mathbb{R}$ and $f(n) = a_n$ for all $n \in \mathbb{N}$. Let $\epsilon > 0$ and note that by the given limit there exists $M \in \mathbb{R}$ such that $|f(x) - L| < \epsilon$ for all $x > M$. Choose N to be the next integer beyond M so $N \in \mathbb{N}$ and $N > M$. Suppose that $n \in \mathbb{N}$ and $n > N$ then $|f(n) - L| = |a_n - L| < \epsilon$. Therefore, $\lim_{n\to\infty} a_n = L$. $\square$

In-Class Example 6.2.9. *Calculate* $\lim\limits_{n\to\infty}(e^{-n} + \tan^{-1}(n))$.

Definition 6.2.10. *infinite sum.*

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

Given a particular formula for a_k it is generally not an easy matter to determine if the limit above exists. These sums without end are called **series**. In particular, we define $\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \cdots$ to converge iff the limit $\lim_{n\to\infty} \sum_{k=1}^{n} a_k$ converges to a real number. We discuss a number of various criteria to analyze this question in calculus II. I believe this amount of detail is sufficient for our purposes in solving the area problem. Our focus will soon shift away from explicit calculation of these sums.

6.2.2 left, right and midpoint rules

We aim to calculate the signed-area bounded by $y = f(x)$ for $a \leq x \leq b$. In this section we discuss three methods to approximate the signed-area. To begin we should settle some standard notation which we will continue to use for several upcoming sections. Let's begin with a picture:

Definition 6.2.11. *partition of $[a, b]$.*

Suppose $a < b$ then $[a, b] \subset \mathbb{R}$. Define $\Delta x = \frac{b-a}{n}$ for $n \in \mathbb{N}$ and let $x_j = a + j\Delta x$ for $j = 0, 1, \dots, n$. In particular, $x_o = a$ and $x_n = b$.

The closed interval $[a, b]$ is a union of n-subintervals of length Δx. Note that the closed interval $[a, b] = [x_o, x_1] \cup [x_1, x_2] \cup \cdots \cup [x_{n-1}, x_n]$.

Definition 6.2.12. *left endpoint rule (L_n).*

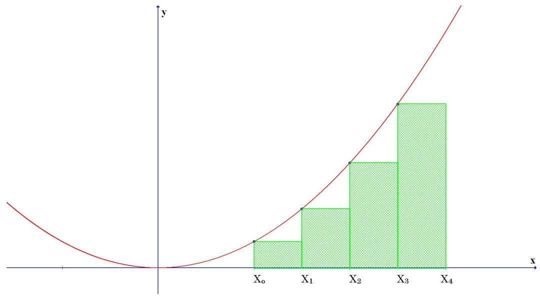
Suppose that $[a, b] \subseteq dom(f)$ then we define

$$L_n = \sum_{j=0}^{n-1} f(x_j)\Delta x = [f(x_0) + f(x_1) + \cdots + f(x_{n-1})]\Delta x.$$

Example 6.2.13. *Let $f(x) = x^2$ and estimate the signed-area bounded by f on $[1, 3]$ by the left-endpoint rule. To keep things simple I'll just illustrate the calculation with $n = 4$. Note $\Delta x = \frac{3-1}{4} = 0.5$ thus $x_o = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ and $x_4 = 3$.*

$$L_4 = [f(1) + f(1.5) + f(2) + f(2.5)]\Delta x = [1 + 2.25 + 4 + 6.25](0.5) = 6.75$$

It's clear from the picture below that L_4 underestimates the true area under the curve.



Definition 6.2.14. *right endpoint rule (R_n).*

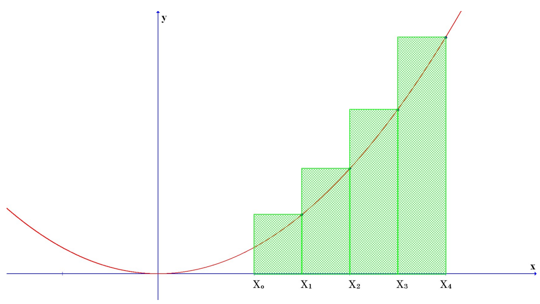
Suppose that $[a, b] \subseteq dom(f)$ then we define

$$R_n = \sum_{j=1}^{n} f(x_j)\Delta x = [f(x_1) + f(x_2) + \cdots + f(x_n)]\Delta x.$$

Example 6.2.15. *Let $f(x) = x^2$ and estimate the signed-area bounded by f on $[1, 3]$ by the right end-point rule. To keep things simple I'll just illustrate the calculation with $n = 4$. Note $\Delta x = \frac{3-1}{4} = 0.5$ thus $x_o = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ and $x_4 = 3$.*

$$R_4 = [f(1.5) + f(2) + f(2.5) + f(3)]\Delta x = [2.25 + 4 + 6.25 + 9](0.5) = 10.75$$

It's clear from the picture below that R_4 overestimates the true area under the curve.



Definition 6.2.16. *midpoint rule (M_n).*

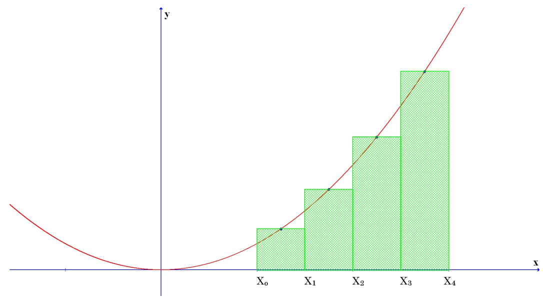
Suppose that $[a, b] \subseteq dom(f)$ and denote the midpoints by $\bar{x}_k = \frac{1}{2}(x_k + x_{k-1})$ and define

$$M_n = \sum_{j=1}^{n} f(\bar{x}_j)\Delta x = [f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)]\Delta x.$$

Example 6.2.17. *Let $f(x) = x^2$ and estimate the signed-area bounded by f on $[1, 3]$ by the midpoint rule. To keep things simple I'll just illustrate the calculation with $n = 4$. Note $\Delta x = \frac{3-1}{4} = 0.5$ thus $\bar{x}_1 = 1.25, \bar{x}_2 = 1.75, \bar{x}_3 = 2.25$ and $\bar{x}_4 = 2.75$.*

$$M_4 = [f(1.25) + f(1.75) + f(2.25) + f(2.75)]\Delta x = [1.5625 + 3.0625 + 5.0625 + 7.5625](0.5) = 8.625$$

Clearly $L_4 < M_4 < R_4$ and if you study the errors you can see $L_4 < M_4 < A < R_4$.



Notice that the size of the errors will shrink if we increase n. In particular, it is intuitively obvious that as $n \to \infty$ we will obtain the precise area bounded by the curve. Moreover, we expect that the distinction between L_n, R_n and M_n should vanish as $n \to \infty$. Careful proof of this seemingly obvious claim is beyond the scope of this course.

Example 6.2.18. *Let $f(x) = x^2$ and calculate the signed-area bounded by f on $[1, 3]$ by the right end-point rule. To perform this calculation we need to set up R_n for arbitrary n and then take the limit as $n \to \infty$. Note $x_k = 1 + k\Delta x$ and $\Delta x = 2/n$ thus $x_k = 1 + 2k/n$. Calculate,*

$$f(x_k) = \left(1 + \frac{2k}{n}\right)^2 = 1 + \frac{4k}{n} + \frac{4k^2}{n^2}$$

thus,

$$\begin{aligned} R_n &= \sum_{k=1}^{n} f(x_k)\Delta x \\ &= \sum_{k=1}^{n} \left[1 + \frac{4k}{n} + \frac{4k^2}{n^2}\right]\frac{2}{n} \\ &= \frac{2}{n}\sum_{k=1}^{n} 1 + \frac{8}{n^2}\sum_{k=1}^{n} k + \frac{8}{n^3}\sum_{k=1}^{n} k^2 \\ &= \frac{2}{n}n + \frac{8}{n^2}\frac{n(n+1)}{2} + \frac{8}{n^3}\frac{n(n+1)(2n+1)}{6} \\ &= 2 + 4\left(1 + \frac{1}{n}\right) + \frac{8}{6}\left(2 + \frac{3}{n} + \frac{1}{n^2}\right) \end{aligned}$$

Note that $\frac{1}{n}$ and $\frac{1}{n^2}$ clearly tend to zero as $n \to \infty$ thus

$$\lim_{n\to\infty} R_n = 2 + 4 + \frac{16}{6} = \frac{26}{3} \approxeq 8.6667.$$

Challenge: show L_n and M_n also have limit $\frac{26}{3}$ as $n \to \infty$.

Notice that the error in M_4 is simply $E = 8.6667 - 8.625 = 0.0417$ which is within 0.5% of the true area. I will not attempt to give an quantitative analysis of the error in L_n, R_n or M_n at this time. Stewart discusses the issue in §8.7. Qualitatively, if the function is monotonic then we should expect that the area is bounded between L_n and R_n.

The example below illustrates that using rectangles is just a convenience of our exposition.

In-Class Example 6.2.19. *Inscribe a regular polygon formed by n-triangles inside a circle of radius R. Let A_n be the area of this polygon and calculate* $\lim_{n\to\infty} A_n$.

6.2.3 Riemann sums and the definite integral

In the last section we claimed that it was intuitively clear that as $n \to \infty$ all the different approximations of the signed-area converge to the same value. You could construct other rules to select the height of the rectangles. Riemann's definition of the definite integral is made to exploit this freedom in the limit. Again, it should be mentioned that this begs an analytical question we are unprepared to answer. For now I have to ask you to trust that the following definition is meaningful. In other words, you have to trust me that it doesn't matter the details of how the point in each subinterval is chosen. Intuitively this is reasonable as $\Delta x \to 0$ as $n \to \infty$. Therefore, the subinterval $[x_j, x_j + \Delta x] \to \{x_j\}$ so the choice between the left, right and midpoints is lost in the limit. Actually, special functions which are very discontinuous could cause problems to the intuitive claim I just made. For that reason we insist that the function below is continuous on $[a, b]$ in order that we avoid certain pathologies.

Definition 6.2.20. *Riemann sum and the definite integral of continuous function on* $[a, b]$.

Suppose that f is continuous on $[a, b]$ suppose $x_k^* \in [x_{k-1}, x_k]$ for all $k \in \mathbb{N}$ such that $1 \leq k \leq n$ then an n-th Riemann sum is defined to be

$$\mathcal{R}_n = \sum_{j=1}^{n} f(x_k^*)\Delta x = [f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)]\Delta x.$$

Notice that no particular restriction is placed on the sample points x_k^*. This means a Riemann sum could be a left, right or midpoint rule. This freedom will be important in the proof of the Fundamental Theorem of Calculus I offer in a later section.

Definition 6.2.21. *definite integrals.*

Suppose that f is continuous on $[a, b]$, the **definite integral** of f from a to b is defined to be $\lim_{n\to} \mathcal{R}_n$ in particular we denote:

$$\int_a^b f(x)\,dx = \lim_{n\to} \mathcal{R}_n = \lim_{n\to\infty}\left[\sum_{j=1}^{n} f(x_k^*)\Delta x\right].$$

The function f is called the **integrand**. The variable x is called the **dummy variable of integration**. We say a is the **lower bound** and b is the **upper bound**. The symbol dx is the **measure**. We also define for $a < b$

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx \qquad \text{and} \qquad \int_a^a f(x)\,dx = 0.$$

The **signed-area** bounded by $y = f(x)$ for $a \leq x \leq b$ is defined to be $\int_a^b f(x)\,dx$.

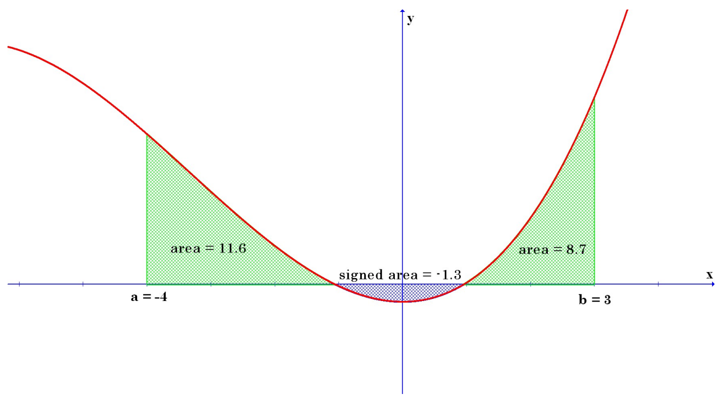
The integral above is known as the Riemann-integral. Other definitions are possible[5].

If f is continuous on the intervals $(a_1, a_2), (a_2, a_3), \dots (a_k, a_{k+1})$ and each discontinuity is a finite-jump discontinuity then the definite integral of f on $[a_1, a_{k+1}]$ is defined to be the sum of the integrals:

$$\int_{a_1}^{a_{k+1}} f(x)\,dx = \sum_{j=1}^{k} \int_{a_j}^{a_{j+1}} f(x)\,dx.$$

Technically this leaves something out since we have only carefully defined integration over a closed interval and here we need the concept of integration over a half-open or open interval. To be careful one has the limit of the end points tending to the points of discontinuity. We discuss this further in Calculus II when we study *improper integrals*

In the graph of $y = f(x)$ below I have shaded the positive signed-area green and the negative signed-area blue for the region $-4 \le x \le 3$. The total signed-area is calculated by the definite integral and can also be found from the sum of the three regions: $11.6 - 1.3 + 8.7 = 19.0 = \int_{-4}^{3} f(x)\,dx$.



Example 6.2.22. *Suppose $f(x) = \sin(x)$. Set-up the definite integral from $[0, \pi]$. We choose $\mathcal{R} = R_n$ for convenience. Note $\Delta x = \pi/n$ and the typical sample point is $x_j^* = j\pi/n$. Thus*

$$R_n = \sum_{j=1}^{n} \sin(x_j^*)\Delta x = \sum_{j=1}^{n} \sin\left(\frac{j\pi}{n}\right)\frac{\pi}{n} \quad \Rightarrow \quad \int_0^{\pi} \sin(x)\,dx = \lim_{n\to\infty} \sum_{j=1}^{n} \sin\left(\frac{j\pi}{n}\right)\frac{\pi}{n}.$$

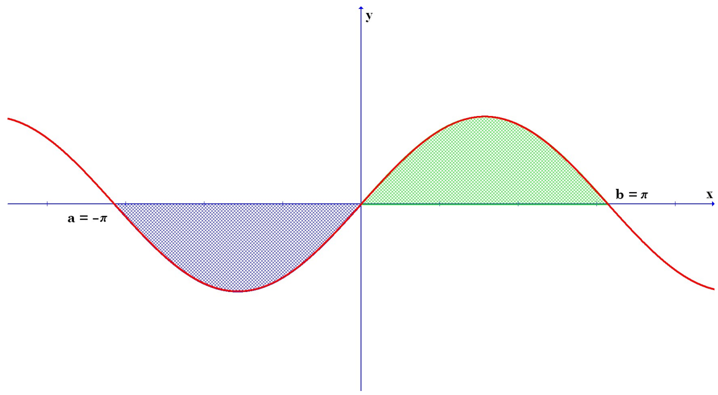
At this point, most of us would get stuck. In order to calculate the limit above we need to find some identity to simplify sums such as

$$\sin\left(\frac{\pi}{n}\right) + \sin\left(\frac{2\pi}{n}\right) + \cdots + \sin\left(\frac{(n-1)\pi}{n}\right) = ?.$$

If you figure it out please show me.

[5]the Riemann-Stieltjes integral or Lesbesque are generalizations of this the basic Riemann integral. Riemann-Stieltjes integral might be covered in some undergraduate analysis courses whereas Lesbesque's measure theory is typically a graduate analysis topic.

Symmetry can help integrate. Note that by the symmetry of the sine function it is clear that $\int_0^{\pi} \sin(x)\,dx = \int_{-\pi}^{0} \sin(x)\,dx$ and consequently the signed area bounded by $y = \sin(x)$ on $[-\pi, \pi]$ is simply zero.



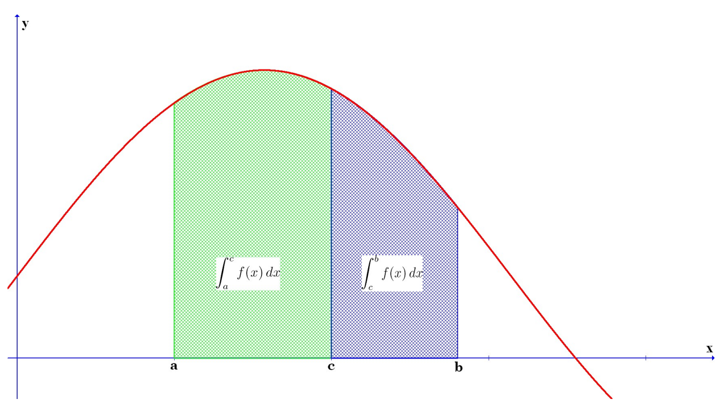
6.2.4 properties of the definite integral

As we just observed a particular Riemann integral can be very difficult to calculate *directly* even if the integrand is a relatively simple function. That said, there are a number of intuitive properties for the definite integral whose proof is easier in general than the preceding specific case.

Proposition 6.2.23. *algebraic properties of definite integration.*

Suppose f, g are continuous on $[a, b]$ and $a < c < b$, $\alpha \in \mathbb{R}$

$$(i.)\ \int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx,$$
$$(ii.)\ \int_a^b \alpha f(x)dx = \alpha \int_a^b f(x)dx,$$
$$(iii.)\ \int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$



Proof: since f, g are continuous it follows $f+g$ is likewise continuous hence $f, g, f+g$ are all bounded on $[a,b]$ and consequently their definite integrals exist (the limit of the Riemann sums must converge to a real value). Consider then,

$$\begin{aligned}
\int_a^b \big[f(x)+\alpha g(x)\big]\,dx &= \lim_{n\to\infty}\left[\sum_{j=1}^{n}\big[f(x_k^*)+\alpha g(x_k^*)\big]\Delta x\right] \\
&= \lim_{n\to\infty}\left[\sum_{j=1}^{n} f(x_k^*)\Delta x + \alpha\sum_{j=1}^{n} g(x_k^*)\Delta x\right] \\
&= \lim_{n\to\infty}\left[\sum_{j=1}^{n} f(x_k^*)\Delta x\right] + \alpha\lim_{n\to\infty}\left[\sum_{j=1}^{n} g(x_k^*)\Delta x\right] \\
&= \int_a^b f(x)\,dx + \alpha\int_a^b g(x)\,dx
\end{aligned}$$

We used the linearity properties of finite sums and the linearity properties of sequential limits in the calculation above. In the case $\alpha = 1$ we obtain a proof for $(i.)$. In the case $g = 0$ we obtain a proof for $(ii.)$.

Proof 1 of (iii.): this is geometrically obvious. Draw a picture, done. $\square$

Proof 2 of (iii.): The proof of $(iii.)$ will require additional thinking. We need to think about a partition of $[a,b]$ and split it into two partitions, one for $[a,c]$ and the other for $[c,b]$. Since $a<c<b$ the value of c must appear somewhere in the partition:

$$x_o = a \;<\; x_1 \;<\; x_2 \;<\; \cdots \;<\; x_j \;\le\; c \;\le\; x_{j+1} \;<\; \cdots \;<\; x_n = a + n\Delta x = b.$$

for some $j<n$. Note $x_k = a + k\Delta x$ and $\Delta x = \frac{b-a}{n}$ for $k = 1,2,\dots,n$. Note that as $n\to\infty$ the following ratios hold (if $x_j = c$ then these are exact, however clearly $x_j \to c$ as $n\to\infty$):

$$\Delta x = \frac{b-a}{n} = \frac{c-a}{j} = \frac{b-c}{n-j}$$

these simply express the fact that the partition of $[a,b]$ has equal length in each region. In what follows the x_j is the particular point in each partition of $[a,b]$ close to the midpoint c:

$$\begin{aligned}
\int_a^b f(x)\,dx &= \lim_{n\to\infty}\left[\sum_{k=1}^{n} f(x_k^*)\Delta x\right] \\
&= \lim_{n\to\infty}\left[\sum_{k=1}^{j} f(x_k^*)\Delta x + \sum_{k=j+1}^{n} f(x_k^*)\Delta x\right] \\
&= \lim_{j\to\infty}\left[\sum_{k=1}^{j} f(z_k^*)\frac{c-a}{j}\right] + \lim_{p\to\infty}\left[\sum_{l=1}^{p} f(y_l^*)\frac{b-c}{p}\right]
\end{aligned}$$

where $z_k^* = x_k^*$ and $y_l^* = x_{l+j}^*$ for $j \approx n\frac{c-a}{b-a}$ and we have replaced the limit of $n \to \infty$ with that of $p = n - j \to \infty$ which is reasonable since $j \approx n\frac{c-a}{b-a}$ gives $n - j \approx n - n\frac{c-a}{b-a} = n\frac{b-a-c+a}{b-a} = n\frac{b-c}{b-a}$ hence $n \to \infty$ implies $n - j \to \infty$ as $b > c$ and $b > a$ by assumption. Likewise, we replaced $n \to \infty$ with $j \to \infty$ for the first sum. This substitution is again justified since $c > a$ and $b > a$ thus $j \approx n\frac{c-a}{b-a}$ suggests $n \to \infty$ implies $j \to \infty$. Finally, denote $\Delta y = \frac{c-a}{j}$ and $\Delta z = \frac{b-c}{p}$ to obtain

$$\begin{aligned}
\int_a^b f(x)\,dx &= \lim_{j\to\infty}\left[\sum_{k=1}^{j} f(z_k^*)\Delta z\right] + \lim_{p\to\infty}\left[\sum_{l=1}^{p} f(y_l^*)\Delta y\right] \\
&= \int_a^c f(z)\,dz + \int_c^b f(y)\,dy \\
&= \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.
\end{aligned}$$

This concludes the proof of $(iii.)$. $\square$

It's interesting that what is intuitively obvious is not necessarily so intuitive to prove. Another example of this pattern is the Jordan curve lemma from complex variables. Basically the lemma simply states that you can divide the plane into two regions, one inside the curve and one outside the curve. The proof isn't typically offered until the graduate course on topology. It's actually a technically challenging thing to prove precisely. This is one of the reasons that rigor is so important to mathematics: what is intuitive maybe be wrong. Historically, appeal to intuition has trapped us for centuries with wrong ideas. However, without intuition we'd probably not advance much either. My personal belief is that for good mathematics to progress we need many different types of mathematicians working in concert. We need visionaries to forge ahead sometimes without proof and we also need careful analytical types to make sure the visionaries are not just going in circles. In this modern age it is no longer feasible to expect all major progress be made by people like Gauss who both propose the idea and provide the proof at levels of rigor sufficient to convince the whole mathematical community. In any event, whether you are a math major or not, I hope this course helps you understand what mathematics is about. By now you should be convinced it's not just about secret formulas and operations on equations.

Proposition 6.2.24. *inequalities of definite integration.*

Suppose f, g are continuous on $[a, b]$ and $m, M \in \mathbb{R}$,

$(i.)$ if $f(x) \geq 0$ for all $x \in [a, b]$ then $\displaystyle\int_a^b f(x)dx \geq 0$,

$(ii.)$ if $f(x) \geq g(x)$ for all $x \in [a, b]$ then $\displaystyle\int_a^b f(x)dx \geq \int_a^b g(x)dx$,

$(iii.)$ if $m \leq f(x) \leq M$ for all $x \in [a, b]$ then $\displaystyle m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$.

Proof: since f, g are continuous we can be sure that the limits defining the definite integrals exist. We need the existence of the limits in order to apply the limit laws in the arguments that follow. Begin with $(i.)$, assume $f(x) \geq 0$ and partition $[a,b]$ as usual $a = x_o, b = x_n$ and $x_k = a + \Delta x$. Sample points x_k^* are chosen from each subinterval $[x_{k-1}, x_k]$. Consider, for any particular $n \in \mathbb{N}$ it is clear that:

$$f(x_k^*) \geq 0 \text{ and } \Delta x = \frac{b-a}{n} > 0 \quad \Rightarrow \quad \sum_{k=1}^{n} f(x_k^*)\Delta x \geq 0$$

Consequently, $\mathcal{R}_n = \sum_{k=1}^{n} f(x_k^*)\Delta x \geq 0$ for all $n \in \mathbb{N}$ hence by comparison property for sequential limits, $\lim_{n\to\infty} \mathcal{R}_n \geq \lim_{n\to\infty}(0) = 0$ and $(i.)$ follows immediately.

To prove $(ii.)$ construct $h(x) = f(x) - g(x)$ and note $f(x) \geq g(x)$ for all $x \in [a,b]$ implies $h(x) = f(x) - g(x) \geq 0$ for all $x \in [a,b]$. We apply $(i.)$ to the clearly continuous function h and obtain:

$$\int_a^b h(x)dx \geq 0 \quad \Rightarrow \quad \int_a^b \big[f(x) - g(x)\big]dx \geq 0 \quad \Rightarrow \quad \int_a^b f(x)dx - \int_a^b g(x)dx \geq 0$$

and $(ii.)$ clearly follows.

Proof of $(iii.)$ follows from observing that if f is bounded by $m \leq f(x) \leq M$ for all $x \in [a,b]$ then $m \leq f(x_k^*) \leq M$ for each $x_k^* \in [a,b]$. Hence,

$$\sum_{k=1}^{n} m \leq \sum_{k=1}^{n} f(x_k^*) \leq \sum_{k=1}^{n} M.$$

But $m, M \in \mathbb{R}$ so the summations on the edges are easy:

$$mn \leq \sum_{k=1}^{n} f(x_k^*) \leq Mn.$$

Finally, we can multiply by $\Delta x = \frac{b-a}{n}$ to obtain

$$mn\frac{b-a}{n} \leq \sum_{k=1}^{n} f(x_k^*)\Delta x \leq Mn\frac{b-a}{n} \quad \Rightarrow \quad m(b-a) \leq \sum_{k=1}^{n} f(x_k^*)\Delta x \leq M(b-a).$$

Apply the sequential limit squeeze theorem and take the limit as $n \to \infty$ to find

$$m(b-a) \leq \lim_{n\to\infty}\left[\sum_{k=1}^{n} f(x_k^*)\Delta x\right] \leq M(b-a)$$

This proves $(iii.)$. $\square$

One easy fact to glean from the proof of $(iii.)$ is the following:

Corollary 6.2.25. *integral of a constant. Let $m \in \mathbb{R}$,*

$$\int_a^b m\, dx = m(b-a).$$

Given that the definite integral was constructed to calculate area this result should not be surprising. Note $0 \leq y \leq m$ for $a \leq x \leq b$ describes a rectangle of width $b-a$ and height m.

6.3 fundamental theorem of calculus

In the preceding section we detailed a careful procedure for calculating the signed area between $y = f(x)$ and $y = 0$ for $a \leq x \leq b$. Unless the function happened to be very simple or enjoyed some obvious symmetry it was difficult to actually calculate the area. We can write the limits but we typically have no way of simplifying the sum to evaluate the limit. In this section we will prove the Fundamental Theorem of Calculus (FTC) which amazingly shows us how to calculate signed-areas without explicit simplification of the Riemann sum or evaluation of the limit. I begin by studying area functions. I show how the FTC part I is seen naturally for both the rectangular and triangular area functions. These two simple cases are discussed to help motivate why we would even expect to find such a thing as the FTC. Then we regurgitate the standard arguments found in almost every elementary calculus text these days to prove "FTC part I" and "FTC part II". Finally, I offer a constructive proof of FTC part II and I argue why FTC part I follows intuitively.

6.3.1 area functions and FTC part I

In that discussion the endpoints a and b were given and fixed in place. We now shift gears a bit. We study **area functions** in this section. The idea of an area function is simply this: if we are given a function f then we can define an area function for f once we pick some base point a. Then $A(x)$ will be defined to be the signed-area bounded by $y = f(t)$ for $a \leq t \leq x$. I use t in the place of x since we wish to use x in a less general sense in the pictures that follow here.

Definition 6.3.1. *area function.*

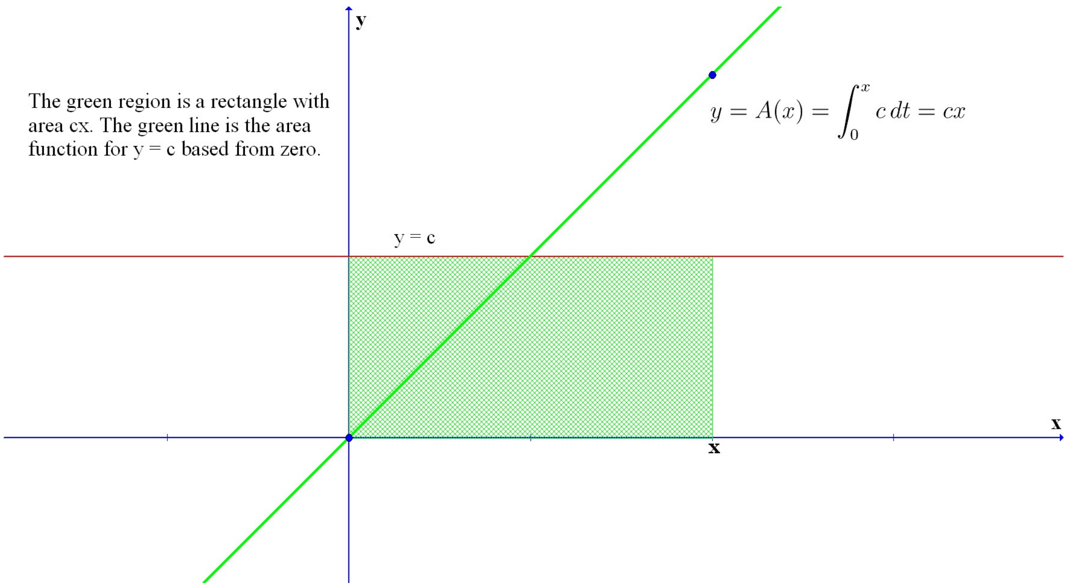
Given f and a point a we define the **area function** of f relative to a as follows:

$$A(x) = \int_a^x f(t)\, dt.$$

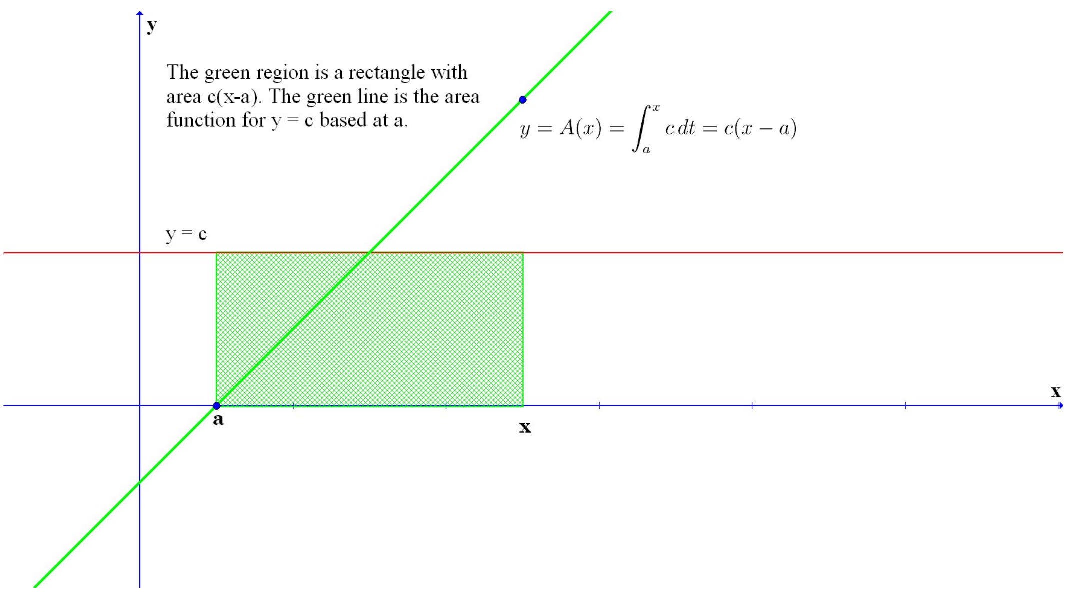
We say that $A(x)$ is the signed-area bounded by f on $[a, x]$.

We would like to look for patterns about area functions. We've seen already that direct calculation is difficult. However, we know two examples from geometry where the area is easily calculated without need of calculus.

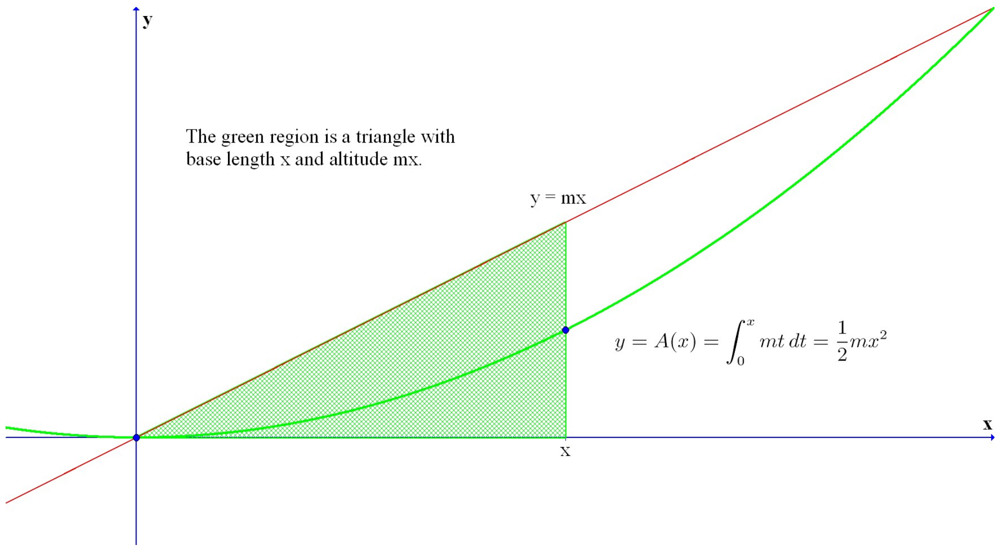
Area function of rectangle: let $f(t) = c$ then the area bounded between $t = 0$ and $t = x$ is simply length (x) times height (c). By geometry we have that $A(x) = \int_0^x c\,dt = cx$, see the picture below:



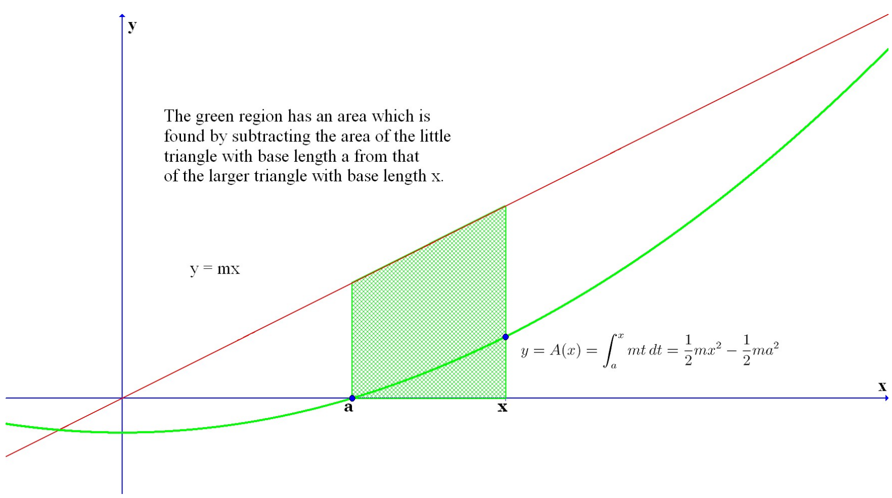
If we positioned the rectangle at $a \le t \le x$ then length becomes $(x - a)$ and the height is still (c). Therefore, by geometry, $A(x) = \int_a^x c\,dt = c(x - a) = cx - ca$. Again, see the picture below where I have pictured a particular x but I have graphed $y = A(t)$ for many t besides x. You can imagine other choices of x and you should find the area function agrees with the area under the curve.



Area function of triangle: I begin with a triangle formed at the origin with the t-axis and the line $y = mt$ and $t = x$. For a particular x, we have base length x and height $y = mx$ thus the area of the triangle is given by geometry: $A(x) = \int_0^x mt\, dt = \frac{1}{2}mx^2$. I picture the function ($y = mt$ in red) as well as the area function ($y = \frac{1}{2}mt^2$ in green) in the picture below:



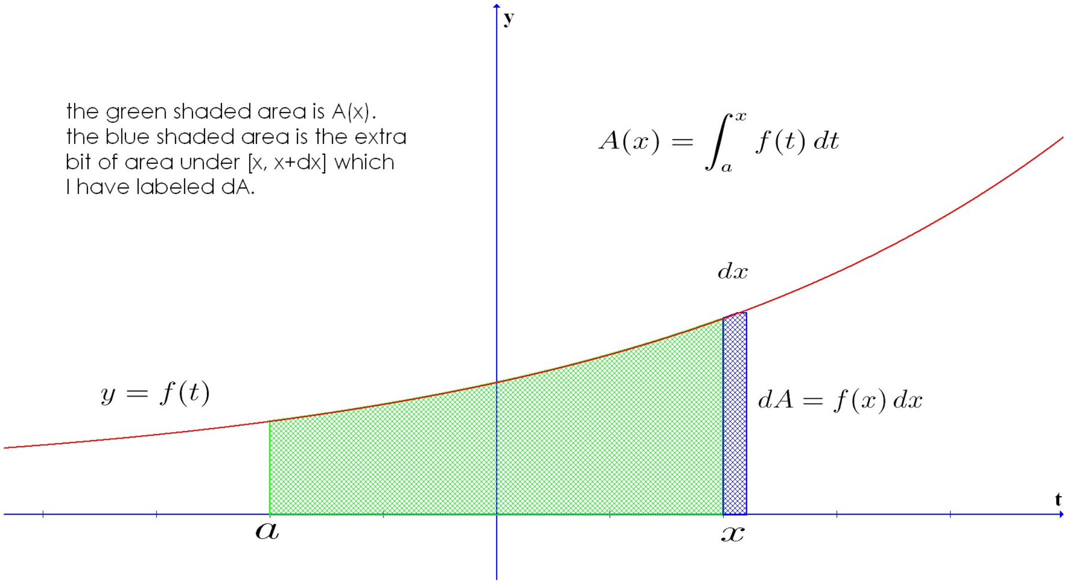
We calculate the area bounded by $y = mt$ for $a \leq t \leq x$ by subtracting the area of the small triangle from $0 \leq t \leq a$ from the area of the larger triangle $0 \leq t \leq x$ as pictured below. Thus from geometry we find $A(x) = \int_a^x mt\, dt = \frac{1}{2}mx^2 - \frac{1}{2}ma^2$.



The area under a parabola could also be calculate without use of further theory. We could work out from the special summation formulas that the area function for $y = t^2$ for $a \leq t \leq x$ is given by $A(x) = \int_a^x t^2\, dt = \frac{1}{3}x^3 - \frac{1}{3}a^3$. I suspect this is beyond the scope of constructive geometry (compass/straight-edge and paper). We should notice a pattern:

1. $A(x) = \int_a^x c\,dt = cx$ has $\frac{dA}{dx} = c$.
2. $A(x) = \int_a^x mt\,dt = \frac{1}{2}mx^2$ has $\frac{dA}{dx} = mx$.
3. $A(x) = \int_a^x t^2\,dt = \frac{1}{3}x^3$ has $\frac{dA}{dx} = x^2$.

We suspect that if $A(x) = \int_a^x f(t)\,dt$ then $\frac{dA}{dx} = f(x)$. Let's examine an intuitive graphical argument for why this is true for an arbitrary function:



Formally, $dA = A(x+dx) - A(x) = f(x)dx$ hence $dA/dx = f(x)$. This proof made sense to you (if it did) because you believe in Leibniz' notation. We should offer a rigorous proof since this is one of the most important theorems in all of calculus.

Theorem 6.3.2. *Fundamental Theorems of Calculus part I (FTC I).*

Suppose f is continuous on $[a,b]$ and $x \in [a,b]$ then,

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x).$$

Proof: let $A(x) = \int_a^x f(t)\,dt$ and note that

$$A(x+h) = \int_a^{x+h} f(t)\,dt = \int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt = A(x) + \int_x^{x+h} f(t)\,dt$$

Therefore, the difference quotient for the area function is simply as follows:

$$\frac{A(x+h) - A(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,dt$$

However, note that by continuity of f we can find bounds for f on $J = [x, x+h]$ (if $h > 0$) or $J = [x+h, x]$ (if $h < 0$). By the extreme value theorem, there exist $u, v \in J$ such that $f(u) \leq f(x) \leq f(v)$ for all $x \in J$. Therefore, if $h > 0$, we can apply the inequality properties of definite integrals and find

$$(x+h-x)f(u) \leq \int_x^{x+h} f(t)\,dt \leq (x+h-x)f(v) \quad \Rightarrow \quad f(u) \leq \frac{1}{h}\int_x^{x+h} f(t)\,dt \leq f(v)$$

If $h < 0$ then dividing by h reverses the inequalities hence $f(v) \leq \frac{1}{h}\int_x^{x+h} f(t)\,dt \leq f(u)$. Finally, observe that $\lim_{h\to 0} u = x$ and $\lim_{h\to 0} v = x$. Therefore, by continuity of f, $\lim_{h\to 0} f(u) = f(x)$ and $\lim_{h\to 0} f(v) = f(x)$. Remember, $f(u) \leq \frac{1}{h}\int_x^{x+h} f(t)\,dt \leq f(v)$ and apply the squeeze theorem to deduce:

$$\lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,dt = f(x)$$

Consequently,

$$\lim_{h\to 0} \frac{A(x+h)-A(x)}{h} = f(x)$$

Which, by definition of the derivative for A, gives $\frac{dA}{dx} = f(x)$. $\square$

The FTC part I is hardly a solution to the area problem. It's just a curious formula. The FTC part II takes this curious formula and makes it useful. It is true there are a few functions defined as area functions hence the differentiation in the FTC I is physically interesting. For example, the Fresnel function can be defined in terms of an integral with a variable bound.

Remark 6.3.3. *a method to derive antiderivatives without guessing.*

Notice that the FTC I also gives us a method to calculate antiderivatives without guessing. But, I can only derive a few very simple antiderivatives. For example, here is a **derivation** of the antiderivative of $f(x) = 3$. I calculate that $\int 3\,dx = 3x + c$ without guessing:

$$\begin{aligned}
\int_0^x f(u)du &= \lim_{n\to\infty}\sum_{i=1}^{n} f(u_i^*)\triangle u \qquad \triangle u = (x-0)/n = x/n \\
&= \lim_{n\to\infty}\sum_{i=1}^{n} 3\frac{x}{n} \\
&= \lim_{n\to\infty}\frac{3x}{n}\sum_{i=1}^{n} 1 \\
&= \lim_{n\to\infty}\frac{3x}{n}n \\
&= \lim_{n\to\infty} 3x \\
&= 3x.
\end{aligned}$$

6.3.2 FTC part II, the standard arguments

The fact that $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$ is just half of what we observed in our examination of the rectangular and triangular area functions. If the area was measured away from the origin on some region $a \leq t \leq x$ then we can observe another pattern: **the area was given by the difference of the antiderivative of the integrand at the end points**

1. $\int_a^x c\,dt = cx - ca$
2. $\int_a^x mt\,dt = \frac{1}{2}mx^2 - \frac{1}{2}ma^2$

This suggests the following theorem may be true:

Theorem 6.3.4. *Fundamental Theorems of Calculus part II (FTC II).*

Suppose f is continuous on $[a, b]$ and has antiderivative F then

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

Proof: consider the area function based at a: $A(x) = \int_a^x f(t)\,dt$. The FTC I says that A is an antiderivative of f. Since F is given to be another antiderivative we know that $F'(x) = A'(x) = f(x)$ which means F and A differ by at most a constant $c \in \mathbb{R}$: $F(x) = A(x) + c$. Since F and A are differentiable on $[a, b]$ it follows they are also continuous on $[a, b]$ hence,

$$F(a) = \lim_{x\to a^+} F(x) = \lim_{x\to a^+}[A(x) + c] = A(a) + c = \int_a^a f(t)\,dt + c = c$$

and

$$F(b) = \lim_{x\to b^-} F(x) = \lim_{x\to b^-}[A(x) + c] = A(b) + c = \int_a^b f(t)\,dt + c.$$

Hence, $F(b) - F(a) = \int_a^b f(t)\,dt + c - c = \int_a^b f(t)\,dt$. Of course, t is just the dummy variable of integration so we can change it to x at this point to complete the proof of the FTC part II. □

In-Class Example 6.3.5. *Calculate the integral* $\int_0^\pi \sin(x)\,dx$ *(recall Example 6.2.22 humbled us in our foolish quest for direct calculation)*

Definition 6.3.6. *evaluation notation.*

We define the symbols below to denote evaluation of an expression:

$$F(x)\Big|_a^b = F(b) - F(a)$$

In this notation the FTC part II is written as follows: $\displaystyle\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a).$

6.3.3 FTC part II an intuitive constructive proof

Let me restate the theorem to begin:

FTC II: Suppose f is continuous on $[a, b]$ and has antiderivative F then

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

Proof: We seek to calculate $\int_a^b f(x)\,dx$. Use the usual partition for the n-th Riemann sum of f on $[a, b]$; $x_o = a, x_1 = a + \Delta x, \dots, x_n = b$ where $\Delta x = \frac{b-a}{n}$. Suppose that f has an antiderivative F on $[a, b]$. Recall the Mean Value Theorem for $y = F(x)$ on the interval $[x_o, x_1]$ tells us that there exists $x_1^* \in [x_o, x_1]$ such that

$$F'(x_1^*) = \frac{F(x_1) - F(x_o)}{x_1 - x_o} = \frac{F(x_1) - F(x_o)}{\Delta x}$$

Notice that this tells us that $F'(x_1^*)\Delta x = F(x_1) - F(x_o)$. But, $F'(x) = f(x)$ so we have found that $f(x_1^*)\Delta x = F(x_1) - F(x_o)$. In other words, the area under $y = f(x)$ for $x_o \leq x \leq x_1$ is well approximated by the difference in the antiderivative at the endpoints. Thus we choose the sample points for the n-th Riemann sum by applying the MVT on each subinterval to select x_j^* such that $f(x_j^*)\Delta x = F(x_j) - F(x_{j-1})$. With this construction in mind calculate:

$$\begin{aligned}
\int_a^b f(x)\,dx &= \lim_{n\to\infty}\left(\sum_{j=1}^n f(x_j^*)\Delta x\right) \\
&= \lim_{n\to\infty}\left(\sum_{j=1}^n \big[F(x_j) - F(x_{j-1})\big]\right) \\
&= \lim_{n\to\infty}\big(F(x_1) - F(x_o) + F(x_2) - F(x_1) + \cdots + F(x_n) - F(x_{n-1})\big) \\
&= \lim_{n\to\infty}\big(F(x_n) - F(x_o)\big) \\
&= \lim_{n\to\infty}\big(F(b) - F(a)\big) \\
&= F(b) - F(a). \square
\end{aligned}$$

This result clearly extends to piecewise continuous functions which have only finite jump discontinuities. We can apply the FTC to each piece and take the sum of those results. This Theorem is amazing. We can calculate the area under a curve based on the values of the antiderivative at the endpoints. Think about that, if $a = 1$ and $b = 3$ then $\int_1^3 f(x)\,dx$ depends only on $F(3)$ and $F(1)$. Doesn't it seem intuitively likely that what value $f(2)$ takes should matter as well? Why don't we have to care about $F(2)$? The values of the function at $x = 2$ certainly went into the calculation of the area, if we calculate a left sum we would need to take values of the function between the endpoints. The cancellation that occurs in the proof is the root of why my naive intuition is bogus.

Next, let me show you how to derive FTC I from FTC II [6]. We have just proved that

$$\int_a^b f(t)\,dt = F(b) - F(a).$$

Suppose $b = x$ and consider differentiating with respect to x,

$$\frac{d}{dx}\int_a^x f(t)\,dt = \frac{d}{dx}[F(x) - F(a)] = \frac{dF}{dx} = f(x).$$

thus we obtain FTC I simply by differentiating FTC II. Moreover, we can obtain a more general result without doing much extra work:

Theorem 6.3.7. *differentiation of integral with variable bounds. (FTC III for fun)*

Suppose u, v are differentiable functions of x and f is continuous where it is integrated,

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(t)\,dt = f(v(x))\frac{dv}{dx} - f(u(x))\frac{du}{dx}.$$

Proof: let f have antiderivative F and apply FTC II at each x to obtain:

$$\int_{u(x)}^{v(x)} f(t)\,dt = F(v(x)) - F(u(x))$$

now differentiate with respect to x and apply the chain-rule,

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(t)\,dt = \frac{dF}{dx}(u(x))\frac{dv}{dx} - \frac{dF}{du}(u(x))\frac{du}{dx}$$

But, $\frac{dF}{dx} = f(x)$ hence $\frac{d}{dx}\int_{u(x)}^{v(x)} f(t)\,dt = f(u(x))\frac{dv}{dx} - f(u(x))\frac{du}{dx}$. $\square$

The examples based on FTC III are embarrassingly simple once you understand what's happening.

[6] note I didn't need to use FTC I in the argument for the FTC II in this section, instead I needed only assume that there existed an antiderivative for the given integrand

Example 6.3.8.

$$\frac{d}{dx}\int_3^x \cos(\sqrt{t})\,dt = \cos(\sqrt{x})\frac{d(\sqrt{x})}{dx} - \cos(\sqrt{3})\frac{d(3)}{dx} = \boxed{\cos(\sqrt{x})\frac{1}{2\sqrt{x}}.}$$

Example 6.3.9.

$$\frac{d}{dx}\int_{e^x}^{x^3} \tanh(t^2)\,dt = \tanh((x^3)^2)\frac{d(x^3)}{dx} - \tanh((e^x)^2)\frac{d(e^x)}{dx} = \boxed{3x^2\tanh(x^6) - e^x\tanh(e^{2x}).}$$

Example 6.3.10. *The function Si is defined by* $Si(x) = \int_0^x \frac{\sin(t)}{t}dt$ *for* $x \neq 0$ *and* $Si(0) = 0$. *This function arises in Electrical Engineering in the study of optics.*

$$\frac{d}{dx}(Si(x)) = \frac{d}{dx}\int_0^x \frac{\sin(t)}{t}dt = \boxed{\frac{\sin(x)}{x}.}$$

In-Class Example 6.3.11. *Calculate* $\dfrac{d}{dx}\displaystyle\int_{\sin(x)}^{x^2+3} \sqrt{t}\,dt$.

In-Class Example 6.3.12. *Suppose* f *is continuous on* $\mathbb{R}$. *It follows that* f *has an antiderivative hence the FTC III applies. Calculate* $\dfrac{d}{dx}\displaystyle\int_{x^2}^{-x} f(u)\,du$.

6.4 definite integration

Example 6.4.1.

$$\int_1^9 \frac{dx}{\sqrt{5x}} = \frac{1}{\sqrt{5}}\int_1^9 \frac{dx}{\sqrt{x}} = \frac{2\sqrt{x}}{\sqrt{5}}\bigg|_1^9 = \frac{2\sqrt{9}}{\sqrt{5}} - \frac{2\sqrt{1}}{\sqrt{5}} = \boxed{\frac{4}{\sqrt{5}}.}$$

In-Class Example 6.4.2.

$$\int_0^1 2^x\,dx =$$

In-Class Example 6.4.3. *Let a, b be constants,*

$$\int_a^b \sinh(t)\,dt =$$

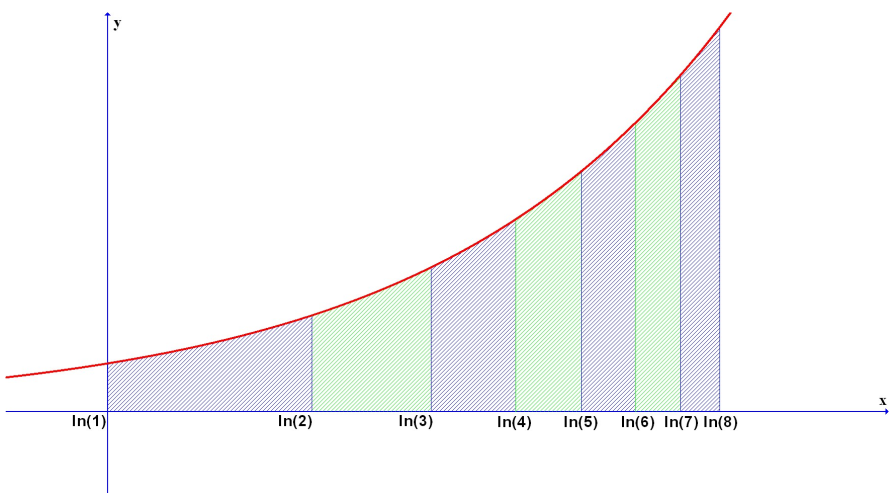
In-Class Example 6.4.4.

$$\int_{-4}^{-2} \frac{dx}{x} =$$

Example 6.4.5. *Let $n > 0$ and consider,*

$$\int_{\ln(n)}^{\ln(n+1)} e^x\,dx = e^{\ln(n+1)} - e^{\ln(n)} = n + 1 - n = \boxed{1.}$$

This is an interesting result. I've graphed a few examples of it below. Notice how as n increases the distance between $\ln(n)$ and $\ln(n+1)$ decreases, yet the exponential increases such that the bounded area still works out to one-unit.

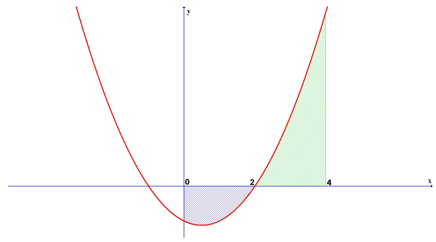


6.4.1 area vs. signed-area

Example 6.4.6. *Calculate the signed-area bounded by $y = 3x^2 - 3x - 6$ for $0 \leq x \leq 2$.*

$$\int_0^2 (3x^2 - 3x - 6)dx = (x^3 - \frac{3}{2}x^2 - 6x)\Big|_0^2 = 8 - \frac{3}{2}(4) - 12 = 8 - 18 = \boxed{-10.}$$

Here's an illustration of the calculation (the blue part):



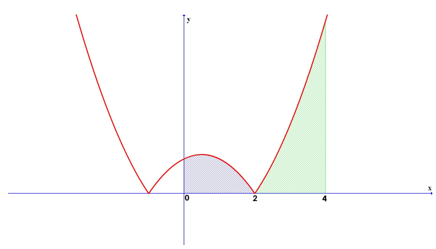
The green area is calculated by

$$\int_2^4 (3x^2 - 3x - 6)dx = (x^3 - \frac{3}{2}x^2 - 6x)\Big|_2^4 = (64 - \frac{3}{2}(16) - 24) + 10 = 64 - 48 + 10 = \boxed{26.}$$

Example 6.4.7. *If we wanted to calculate the area bounded by $y = f(x) = 3x^2 - 3x - 6$ and $y = 0$ for $0 \leq x \leq 4$ then we need to also count negative-signed-area as positive. This is nicely summarized by stating we should integrate the absolute value of the function to obtain the area bounded between the function and the x-axis. Generally analyzing an absolute value of a function takes some work, but given the previous example it is clear how to break up the positive and negative cases:*

$$\begin{aligned}
\int_0^4 |3x^2 - 3x - 6|dx &= \int_0^2 |3x^2 - 3x - 6|dx + \int_2^4 |3x^2 - 3x - 6|dx \\
&= \int_0^2 [-(3x^2 - 3x - 6)]dx + \int_2^4 (3x^2 - 3x - 6)dx \\
&= 10 + 26 \\
&= \boxed{36.}
\end{aligned}$$

Here's a picture of the function we just integrated. You can see how the absolute value flips the negative part of the original function up above the x-axis.



Remark 6.4.8. *absolute values and areas.*

To calculate the area bounded by $y = f(x)$ for $a \leq x \leq b$ we may calculate

$$Area = \int_a^b |f(x)|\, dx.$$

Example 6.4.9. *Calculate the area bounded by $y = \cos(x)$ on $0 \leq x \leq \frac{5\pi}{2}$.*

$$\begin{aligned}
\int_0^{3\pi} |\cos(x)| dx &= \int_0^{\frac{\pi}{2}} \cos(x) dx - \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos(x) dx + \int_{\frac{3\pi}{2}}^{\frac{5\pi}{2}} \cos(x) dx \\
&= \sin(x)\Big|_0^{\frac{\pi}{2}} - \sin(x)\Big|_{\frac{\pi}{2}}^{\frac{3\pi}{2}} + \sin(x)\Big|_{\frac{3\pi}{2}}^{\frac{5\pi}{2}} \\
&= \sin\left(\frac{\pi}{2}\right) - \sin(0) - \sin\left(\frac{3\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right) + \sin\left(\frac{5\pi}{2}\right) - \sin\left(\frac{3\pi}{2}\right) \\
&= \boxed{5.}
\end{aligned}$$

In-Class Example 6.4.10. *Calculate the area bounded by* $f(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ -e^x & x > 1 \end{cases}$ *over the interval* $[0, \ln 4]$.

6.4.2 average of a function

To calculate the average of finitely many things we can just add all the items together then divide by the number of items. If you draw a bar chart and find the area of all the bars and then divide by the number of bars then that gives the average. A function $f(x)$ takes on infinitely many values on a closed interval so we cannot just add the values, however, we can calculate the area and divide by the length. This is the continuous extension of the averaging concept:

Definition 6.4.11. *average of a function over a closed interval.*

The average value of f on $[a, b]$ is defined by

$$f_{avg} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

In-Class Example 6.4.12. *Suppose $f(x) = 4x^3$. Find the average of f on $[0, 2]$.*

Example 6.4.13. *Suppose $f(x) = \sin(x)$. Find the average of f on $[0, 2\pi]$.*

$$f_{avg} = \frac{1}{2\pi}\int_0^{2\pi} \sin(x)\,dx = \frac{-1}{2\pi}\cos(x)\Big|_0^{2\pi} = \boxed{0.}$$

Example 6.4.14. *In the case of constant acceleration $a = -g$ we calculated that $v(t) = v_o - gt$ where v_o, g were constants. Let's calculate the average velocity over some time interval $[t_1, t_2]$,*

$$\begin{aligned}
v_{avg} &= \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} (v_o - gt)\,dt \\
&= \frac{1}{t_2 - t_1}[v_o t - \frac{g}{2}t^2]\Big|_{t_1}^{t_2} \\
&= \frac{1}{t_2 - t_1}\left[v_o(t_2 - t_1) - \frac{g}{2}\left(t_2^2 - t_1^2\right)\right] \\
&= \frac{1}{t_2 - t_1}\left[v_o t_2 - \frac{g}{2}t_2^2 - v_o t_1 + \frac{g}{2}t_1^2\right] \\
&= \frac{y(t_2) - y(t_1)}{t_2 - t_1}
\end{aligned}$$

where I have used a little imagination and a recollection that $y(t) = y_o + v_o t - \frac{g}{2}t^2$. The result is comforting, we find the average velocity is the average of the average velocity function.

There is a better way to calculate the last example. It will provide the first example of the next subsection.

6.4.3 net-change theorem

Combining FTC I and FTC II we find a very useful result: the net-change theorem.

Theorem 6.4.15. *net change theorem.*

$$\int_a^b \frac{df}{dt}\,dt = f(b) - f(a).$$

Example 6.4.16. *Let $v(t)$ be the instantaneous velocity where $v(t) = \frac{dy}{dt}$ then we can calculate the average velocity over some time interval $[t_1, t_2]$,*

$$\begin{aligned} v_{avg} &= \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} v(t)\,dt \\ &= \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} \frac{dy}{dt}\,dt \\ &= \frac{1}{t_2 - t_1}\big(y(t_2) - y(t_1)\big) \\ &= \frac{y(t_2) - y(t_1)}{t_2 - t_1}. \end{aligned}$$

Notice we didn't even need to know the details of the velocity function.

In-Class Example 6.4.17. *Suppose water flows into a tank at a rate $\frac{dm}{dt} = 3 + t^2$ for time $0 \leq t \leq 1$. Find the net mass of water transferred into the tank during the interval $[0, 1]$.*

6.5 u-substitution

The integrations we have done up to this point have been elementary. Basically all we have used is linearity of integration and our basic knowledge of differentiation. We made educated guesses as to what the antiderivative was for a certain class of rather special functions. Integration requires that you look ahead to the answer before you get there. For example, $\int \sin(x)\,dx$. To reason this out we think about our basic derivatives, we note that the derivative of $\cos(x)$ gives $-\sin(x)$ so we need to multiply our guess by -1 to fix it. We conclude that $\int \sin(x)\,dx = -\cos(x) + c$. The logic of this is essentially educated guessing. You might be a little concerned at this point. Is that all we can do? Just guess? Well, no. There is more. But, those basic guesses remain, They form the basis for all elementary integration theory.

The new idea we look at in this section is called "u-substitution". It amounts to the reverse chain rule. The goal of a properly posed u-substitution is to change the given integral to a new integral which is elementary. Typically we go from an integration in x which seems incalculable to a new integration in u which is elementary. For the most part we will make direct substitutions, these have the form $u = g(x)$ for some function g however, this is not strictly speaking the only sort of substitution that can be made. Implicitly defined substitutions such as $x = f(\theta)$ play a critical role in many interesting integrals, we will deal with those more subtle integrations in a later chapter when we discuss trigonometric substitution.

Finally, I should emphasize that when we do a u-substitution we must be careful to convert each and every part of the integral to the new variable. This includes both the integrand($f(x)$) and the measure(dx) in an indefinite integral $\int f(x)\,dx$. Or the integrand($f(x)$), measure(dx) and upper and lower bounds a, b in a definite integral $\int_a^b f(x)\,dx$. I will provide a proof of the method at the conclusion of the section for a change of pace. Examples first this time.

6.5.1 u-substitution in indefinite integrals

Example 6.5.1.

$$\int xe^{x^2}\,dx = \int xe^u \frac{du}{2x} \qquad \boxed{\textit{let } u = x^2,\ \frac{du}{dx} = 2x \textit{ and } dx = \frac{du}{2x}}$$

$$= \frac{1}{2}\int e^u du \qquad \textit{see how all the x's cancelled, this has to happen.}$$

$$= \frac{1}{2}e^u + c \qquad \textit{not done yet.}$$

$$= \boxed{\frac{1}{2}e^{x^2} + c} \qquad \textit{differentiate to check if in doubt.}$$

In-Class Example 6.5.2. *Let a, b be constants. If $a \neq 0$ then,*

$$\int (ax+b)^{13} dx$$

Example 6.5.3.

$$\begin{aligned}\int 5^{\frac{x}{3}}\,dx &= \int 5^u(3du) \\ &= \frac{3}{\ln(5)}5^u + c \\ &= \boxed{\frac{3}{\ln(5)}5^{\frac{x}{3}} + c.}\end{aligned}$$

$$\boxed{\textit{let } u = \frac{x}{3},\ \frac{du}{dx} = \frac{1}{3} \textit{ and } dx = 3du}$$

In-Class Example 6.5.4.

$$\int \tan(x)dx$$

Example 6.5.5.

$$\begin{aligned}\int \frac{2x}{1+x^2}dx &= \int \frac{du}{u} \\ &= \ln(|u|) + c \\ &= \boxed{\ln(1+x^2) + c.}\end{aligned}$$

$$\boxed{\textit{let } u = 1 + x^2,\ \frac{du}{dx} = 2x \textit{ and } 2xdx = du}$$

Notice that $x^2+1>0$ *for all* $x \in \mathbb{R}$ *thus* $|x^2+1| = x^2+1$. *We should only drop the absolute value bars if we have good reason.*

In-Class Example 6.5.6.

$$\int \sqrt[3]{1-3x}dx$$

Example 6.5.7.

$$\begin{aligned} \int \frac{dx}{x+b} &= \int \frac{du}{u} \qquad \boxed{\textit{let } u = x+b \textit{ thus } du = dx} \\ &= \ln|u| + c \\ &= \boxed{\ln|x+b| + c.} \end{aligned}$$

Example 6.5.8. *suppose* $x > 0$.

$$\begin{aligned} \int \frac{x^2dx}{\sqrt{x^2-x^4}} &= \int \frac{x^2dx}{x\sqrt{1-x^2}} \\ &= \int \frac{xdx}{\sqrt{1-x^2}} \\ &= \int \frac{-du}{2\sqrt{u}} \qquad \boxed{\textit{let } u = 1-x^2 \textit{ thus } -du/2 = xdx} \\ &= \frac{-1}{2}2\sqrt{u} + c \\ &= \boxed{-\sqrt{1-x^2} + c.} \end{aligned}$$

In-Class Example 6.5.9. *Suppose* $x > 0$*, calculate:*

$$\int \frac{\ln(x)dx}{x}$$

Example 6.5.10.

$$\begin{aligned} \int \sin(3\theta)d\theta &= \int \sin(u)\frac{du}{3} \qquad \boxed{\textit{let } u = 3\theta \textit{ thus } d\theta = \frac{du}{3}} \\ &= \frac{-1}{3}\cos(u) + c \\ &= \boxed{\frac{-1}{3}\cos(3\theta) + c.} \end{aligned}$$

In-Class Example 6.5.11.

$$\int \frac{\sin^{-1}(z)}{\sqrt{1-z^2}}dz$$

Example 6.5.12.

$$\begin{aligned}\int t\cos(t^2+\pi)dt &= \frac{1}{2}\int \cos(u)du \\ &= \frac{1}{2}\sin(u)+c \\ &= \boxed{\frac{1}{2}\sin(t^2+\pi)+c.}\end{aligned}$$

$$\boxed{\textit{let } u = t^2+\pi \textit{ thus } tdt = \frac{du}{2}}$$

The example below generalizes to allow integration of any odd power of sine or cosine.

In-Class Example 6.5.13.

$$\int \sin^3(x)dx$$

Example 6.5.14. *suppose* $a \neq 0$

$$\begin{aligned}\int \frac{dx}{x^2+a^2} &= \frac{1}{a^2}\int \frac{dx}{\frac{x^2}{a^2}+1} \\ &= \frac{1}{a^2}\int \frac{adu}{u^2+1} \\ &= \frac{1}{a}\tan^{-1}(u)+c \\ &= \boxed{\frac{1}{a}\tan^{-1}\left[\frac{x}{a}\right]+c.}\end{aligned}$$

$\boxed{\textit{let } u = \frac{x}{a} \textit{ thus } adu = dx}$

In-Class Example 6.5.15.

$$\int \frac{xdx}{x^2+6x+13}$$

Example 6.5.16. *suppose* $a \neq 0$

$$\begin{aligned}\int \cos(ae^x+3)e^x dx &= \frac{1}{a}\int \cos(u)du \\ &= \frac{1}{a}\sin(u)+c \\ &= \boxed{\frac{1}{a}\sin(ae^x+3)+c.}\end{aligned}$$

$\boxed{\textit{let } u = ae^x+3 \textit{ thus } du/a = e^x dx}$

Example 6.5.17.

$$\begin{aligned}
\int \sin^2(\theta)d\theta &= \int \frac{1}{2}\Big[1 - \sin^2(\theta)\Big]d\theta && \textit{by trigonmetry.} \\
&= \frac{1}{2}\int d\theta - \frac{1}{2}\int \cos(2\theta)d\theta \\
&= \boxed{\frac{\theta}{2} - \frac{1}{4}\sin(2\theta) + c.}
\end{aligned}$$

In the preceding example I omitted a u-substitution because it was fairly obvious.

In-Class Example 6.5.18.

$$\int \cos^2(\theta)d\theta =$$

Example 6.5.19.

$$\begin{aligned}
\int 4\sinh^2(x)dx &= 4\int \Big[\frac{1}{2}\big(e^x - e^{-x}\big)\Big]^2 dx && \textit{by definition of } \sinh(x). \\
&= \int \Big[(e^x)^2 - 2e^x e^{-x} + (e^{-x})^2\Big]dx \\
&= \int \Big[e^{2x} - 2 + e^{-2x}\Big]dx \\
&= \int e^{2x}dx - 2\int dx + \int e^{-2x}dx \\
&= \frac{1}{2}\int e^{2x}d(2x) - 2x - \frac{1}{2}\int e^{-2x}d(-2x) \\
&= \frac{1}{2}e^{2x} - 2x - \frac{1}{2}e^{-2x} + c. \\
&= \boxed{\sinh(2x) - 2x + c.}
\end{aligned}$$

Interesting, if you trust my calculation then we may deduce

$$4\sinh^2(x) = \frac{d}{dx}[\sinh(2x) - 2x] = 2\cosh(2x) - 2$$

thus $\sinh^2(x) = \frac{1}{2}[\cosh(2x) - 1]$.

6.5.2 u-substitution in definite integrals

There are two ways to do these. I expect you understand both methods.

1. Find the antiderivative via u-substitution and then use the FTC to evaluate in terms of the given upper and lower bounds in x. (see Example 6.5.20 below)

2. Do the u-substitution and change the bounds all at once, this means you will use the FTC and evaluate the upper and lower bounds in u. (see Example 6.5.21 below)

I will deduct points if you write things like a definite integral is equal to an indefinite integral (just leave off the bounds during the u-substitution). The notation is not decorative, it is necessary and important to use correct notation.

Example 6.5.20. *We previously calculated that* $\int t\cos(t^2+\pi)dt = \frac{1}{2}\sin(t^2+\pi)+c$. *We can use this together with the FTC to calculate the following definite integral:*

$$\begin{aligned}\int_0^{\sqrt{\frac{\pi}{2}}} t\cos(t^2+\pi)dt &= \frac{1}{2}\sin(t^2+\pi)\Big|_0^{\sqrt{\frac{\pi}{2}}} \\ &= \frac{1}{2}\sin(\tfrac{\pi}{2}+\pi) - \frac{1}{2}\sin(\pi) \\ &= \boxed{\frac{-1}{2}.}\end{aligned}$$

This illustrates method (1.) we find the antiderivative off to the side then calculate the integral using the FTC in the x-variable. Well, the t-variable here. This is a two-step process. In the next example I'll work the same integral using method (2.). In contrast, that is a one-step process but the extra step is that you need to change the bounds in that scheme. Generally, some problems are easier with both methods. Also, sometimes you may be faced with an abstract question which demands you understand method 2.).

Example 6.5.21.

$$\begin{aligned}\int_0^{\sqrt{\frac{\pi}{2}}} t\cos(t^2+\pi)dt &= \frac{1}{2}\int_{\pi}^{\frac{3\pi}{2}} \cos(u)du \\ &= \frac{1}{2}\sin(u)\Big|_{\pi}^{\frac{3\pi}{2}} \\ &= \frac{1}{2}\sin(\tfrac{3\pi}{2}) - \frac{1}{2}\sin(\pi) \\ &= \boxed{\frac{-1}{2}.}\end{aligned}$$

$\boxed{\textit{let } u = t^2+\pi \textit{ thus } tdt = \frac{du}{2}}$

$\boxed{\textit{also } u\left(\frac{\pi}{2}\right) = \frac{3\pi}{2} \textit{ and } u(0) = \pi}$

Example 6.5.22.

$$\begin{aligned} \int_{4\pi^2}^{9\pi^2} \frac{\sin(\sqrt{x})dx}{\sqrt{x}} &= \int_{2\pi}^{3\pi} \sin(u)(2du) \\ &= -2\cos(u)\Big|_{2\pi}^{3\pi} \\ &= -2\cos(3\pi) + 2\cos(2\pi) \\ &= \boxed{4.} \end{aligned}$$

$\boxed{\text{let } u = \sqrt{x} \text{ thus } 2du = \frac{dx}{\sqrt{x}}}$

$\boxed{\text{also } u(9\pi^2) = \sqrt{9\pi^2} = 3\pi \text{ and } u(4\pi^2) = \sqrt{4\pi^2} = 2\pi}$

6.5.3 theory of u-substitution

In the past 20 examples we've seen how the technique of u-substitution works. To summarize, you take an integrand and measure in terms of x (say $g(f(x))dx$) and propose a new variable $u = f(x)$ for some function f. Then we differentiate $\frac{du}{dx} = f'(x)$ and solve for $dx = \frac{du}{f'(x)}$ which gives us

$$\int g(f(x))\,dx = \int g(u)\,\frac{du}{f'(x)}$$

and if our choice of u is well thought out then the expression $\frac{g(u)}{f'(x)}$ can be simplified into a nice elementary integrable function $h(u)$ (meaning $\int h(u)\,du$ was on our list of elementary integrals). In a nutshell, that is what we did in each example. Let's me raise a couple questions to criticize the method:

1. what in the world do I mean by $dx = \frac{du}{f'(x)}$? This sort of division is not rigorous.
2. what if $f'(x) = 0$? Especially if we were doing an integration with bounds, is it permissible to have a point in the domain of integration where the substitution seems to indicate division by zero?

Question (1.) is not too hard to answer. Let me propose the formal result as a theorem.

Theorem 6.5.23. *change of variables in integration.*

Suppose g is continuous on the connected interval J with endpoints $f(a)$ and $f(b)$ and f is differentiable on a, b then

1. $$\left[\int g(u)\,du\right]\Big|_{u=f(x)} = \int g(f(x))\frac{df}{dx}\,dx$$

2. $$\int_{f(a)}^{f(b)} g(u)\,du = \int_a^b g(f(x))\frac{df}{dx}\,dx.$$

Proof: Note that g continuous indicates the existence of an antiderivative G on J. Let $u = f(x)$ and apply the chain-rule to differentiate $G(u)$,

$$\frac{d}{dx}[G(u)] = G'(u)\frac{du}{dx} = g(u)\frac{df}{dx} = g(f(x))\frac{df}{dx}$$

At this stage we have already proved the indefinite integral substitution rule:

$$G(f(x)) = \left[\int g(u)\,du\right]\bigg|_{u=f(x)} = \int g(f(x))\frac{df}{dx}\,dx = H(x) + c.$$

Use the result above and FTC II to see why (2.) is true:

$$\int_a^b g(f(x))\frac{df}{dx}\,dx = H(b) - H(a) = G(f(b)) - G(f(a)) = \int_{f(a)}^{f(b)} g(u)\,du. \square$$

I assumed continuity for simplicity of argument. One could prove a more general result for piecewise continuous functions. Furthermore, note we never really divided by $f'(x)$ thus $f'(x) = 0$ does not rule out the applicability of this theorem.

Example 6.5.24. *Consider the following problem: calculate*

$$\int_0^{2\pi} e^{\sin(x)} \cos(x)\,dx.$$

In this case we should identify $u = f(x) = \sin(x)$ and $g(u) = e^u$. Clearly the hypotheses of the theorem above are met. Moreover, $f(0) = \sin(0) = 0$ and $f(2\pi) = \sin(2\pi) = 0$ hence

$$\int_0^{2\pi} e^{\sin(x)} \cos(x)\,dx = \int_0^{2\pi} e^{\sin(x)} \frac{d(\sin(x))}{dx}\,dx = \int_0^0 e^u\,du = 0.$$

For whatever reason, using the notation above seems unnatural to most people so we instead think about substituting formulas with u into the integrand. Same calculation, but this time with our usual approach:

$$\int_0^{2\pi} e^{\sin(x)} \cos(x)\,dx = \int_0^0 e^u \cos(x)\,\frac{du}{\cos(x)} \qquad \boxed{\textit{let } u = \sin(x) \textit{ thus } dx = \frac{du}{\cos(x)}}$$

$$= \int_0^0 e^u\,du \qquad \boxed{\textit{also } u(0) = \sin(0) \textit{ and } u(2\pi) = \sin(2\pi).}$$

$$= \boxed{0.}$$

The apparent division by zero was just a sloppy way of communicating application of the theorem for variable change.

This phenomenon of the bounds collapsing to a point will only occur if $\frac{du}{dx} = 0$ somewhere along $a \leq x \leq b$. Otherwise, $\frac{du}{dx} \neq 0$ hence u is strictly monotonic on $[a, b]$ hence either $u(a) < u(b)$ or $u(b) > u(a)$.

6.6 integrals of trigonometric functions

In this section we return to the problem of integrating trigonometric functions. The tools used here are a combination of basic u-substitution, judiciously chosen trigonometric identities[7].

Example 6.6.1.

$$\begin{aligned}\int \sin^4(x)dx &= \int \left[\sin^2(x)\right]^2 dx \\ &= \int \left[\frac{1}{2}(1-\cos(2x))\right]^2 dx \\ &= \frac{1}{4}\int \left[1 - 2\cos(2x) + \cos^2(2x)\right] dx \\ &= \frac{x}{4} - \frac{1}{4}\sin(2x) + \frac{1}{8}\int (1+\cos(4x))dx \\ &= \frac{x}{4} - \frac{1}{4}\sin(2x) + \frac{x}{8} + \frac{1}{32}\sin(4x) + c \\ &= \boxed{\frac{3x}{8} - \frac{1}{4}\sin(2x) + \frac{1}{32}\sin(4x) + c.}\end{aligned}$$

It is important to remember $\tan^2(x) + 1 = \sec^2(x)$ and $\int \sec^2(x)dx = \tan(x) + c$ in the examples that follow.

Example 6.6.2.

$$\begin{aligned}\int \tan^2(x)dx &= \int (\sec^2(x) - 1)dx \\ &= \boxed{\tan(x) - x + c.}\end{aligned}$$

Example 6.6.3. *We let $u = \tan(x)$ so $du = \sec^2(x)dx$,*

$$\begin{aligned}\int \sec^2(x)\tan^2(x)dx &= \int u^2 du \\ &= \frac{1}{3}u^3 + c \\ &= \boxed{\frac{1}{3}\tan^3(x) + c.}\end{aligned}$$

[7] next semester you will learn to extend this section a bit by the method of Integration By Parts (IBP)

In-Class Example 6.6.4. *Calculate* $\displaystyle\int \tan^4(x)dx$

The notation used in the third line of the calculation above is a slick implicit notation for indicating a $u = \tan(x)$ substitution. Every so often I make use of this notation. In any event, you should be able to integrals of expressions like $\int \sec^6(x)dx$ or $\int \cot^2(x)dx$ or $\int \cot^2(x)\csc^2(x)dx$ using arguments paralelling the previous triple of examples. I'll leave integrals of odd powers of secant to Calculus II since those integrals are a pain.

Remark 6.6.5.

See Section 7.2 for an account of how to use and derive trigonometric identities. If you invest a little time to understand how the complex exponential function $e^{ix} = \cos x + i\sin x$ encodes both sine and cosine together in a unified object subject to the expected laws of exponents $e^{ix}e^{iy} = e^{ix+iy}$ then you can derive trig. identities. The trouble of remembering dozens of identities is replaced with the trouble of remembering:

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix}) \qquad \text{and} \qquad \cos x = \frac{1}{2}(e^{ix} + e^{-ix})$$

Alternatively, you can memorize the adding angle formulas and derive most everything from that pair of identities. In some sense these approaches are just alternate notations for the same underlying structure. Naturally, using these formulas without justification is no more logical than utilizing the adding angles formulas without deriving them. Options aside, these formulas are correct, meaningful and have been worthwhile to science and mathematics for a couple centuries.

We keep in mind that the adding angles formula for cosine is $\boxed{\cos(\theta+\beta) = \cos\theta\cos\beta - \sin\theta\sin\beta}$ whereas the adding angles formula for sine is $\boxed{\sin(\theta+\beta) = \sin\theta\cos\beta + \cos\theta\sin\beta}$. Together these adding angles formulas for sine and cosine yield another for tangent; $\boxed{\tan(\theta+\beta) = \dfrac{\tan\theta + \tan\beta}{1 - \tan\theta\tan\beta}}$. Finally the product identities for sine and cosine are also very useful and for most of us far from obvious;

$$\boxed{\cos(ax)\cos(bx) = \frac{1}{2}\cos[(a+b)x] + \frac{1}{2}\cos[(a-b)x]}$$

and

$$\boxed{\sin(ax)\sin(bx) = \frac{1}{2}\cos[(a-b)x] - \frac{1}{2}\cos[(a+b)x]}$$

and

$$\boxed{\cos(ax)\sin(bx) = \frac{1}{2}\sin[(a+b)x] + \frac{1}{2}\sin[(a-b)x]}.$$

The product formulas are very important to the study of constructive and destructive inteference in waves. They explain where beats come from among other things. Also, it is worth mentioning that if you remember one of these carefully then you can get others from differentiating. Try differentiating $\sin(a+x)$ to derive the adding angles formula for $\cos(a+x)$.

Example 6.6.6.

$$\begin{aligned}
\int \cos(3x)\sin(5x)dx &= \int \left[\frac{1}{2}\sin(8x) + \frac{1}{2}\sin(-2x)\right]dx \\
&= \frac{1}{2}\int \sin(8x)dx - \frac{1}{2}\int \sin(2x)dx \\
&= \boxed{\frac{-1}{16}\cos(8x) - \frac{1}{4}\cos(2x) + c.}
\end{aligned}$$

In-Class Example 6.6.7. *Calculate* $\displaystyle\int \cos(3x)\cos(5x)dx$.

Example 6.6.8.

$$\begin{aligned}
\int \sin(3x)\sin(5x)dx &= \int \left[\frac{1}{2}\cos(8x) - \frac{1}{2}\cos(-2x)\right]dx \\
&= \frac{1}{2}\int \cos(8x)dx - \frac{1}{2}\int \cos(2x)dx \\
&= \boxed{\frac{1}{16}\sin(8x) - \frac{1}{4}\sin(2x) + c.}
\end{aligned}$$

What about $\int \sin(x)\cos(3x)\cos(6x)dx$? How would you attack such a problem?

Example 6.6.9. *Here we use the adding angles identity for tangent followed by a* $u = \cos(4x)$ *substitution.*

$$\begin{aligned}
\int \frac{\tan(x)+\tan(3x)}{1-\tan(x)\tan(3x)}dx &= \int \tan(4x)dx \\
&= \int \frac{\sin(4x)}{\cos(4x)}dx \\
&= \int \frac{-du}{4u} \\
&= \boxed{\frac{-1}{4}\ln|\cos(4x)| + c.}
\end{aligned}$$

Finally, I would just comment that there are many integrations of the hyperbolic trigonometric functions which follow arguments paralell to those given in this section.

6.7 area bounded by curves

Examples to be given in lecture.

Chapter 7

Appendix

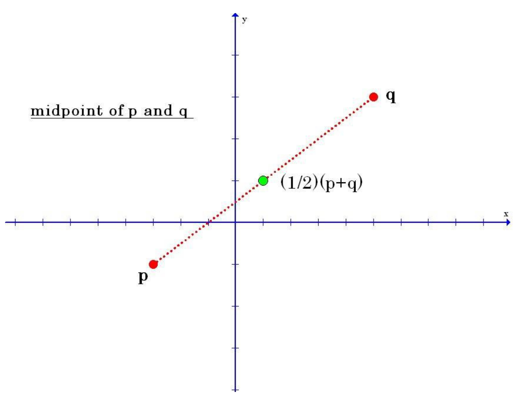
7.1 analytical geometry

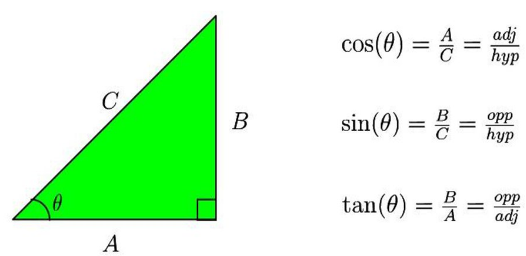
The difference between analytic geometry and the formal geometry of Euclid is that analytic geometry is based primarily on numbers and algebra whereas the method of Euclid involves mainly straight-edge and compass constructions. Analytic geometry is far more useful. As a concrete example, it is impossible to trisect an angle in general using constructive methods however, in analytic geometry trisecting an angle is as easy as dividing by three and using your handy-dandy protractor. Of course the history and beauty of Euclidean geometry ought not be neglected, you'll see the beauty in a course on modern geometry. Also, abstract algebra has much to say about the non-existence of certain constructions in Euclidean geometry.

The geometry of the plane is easily described by various operations on the set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$. If $p = (a, b) \in \mathbb{R}^2$ then we say that the **x-coordinate** of p is a and the **y-coordinate** of p is b. Typically we call the y direction the vertical and the x direction the horizontal. If we are given two points, say $p = (a_1, a_2)$ and $q = (b_1, b_2)$ then the distance between them is given by the **distance formula**

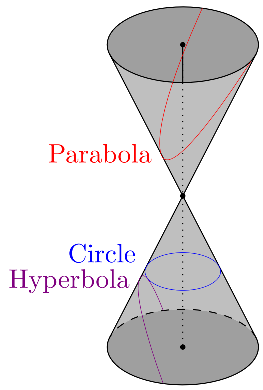
$$\boxed{d(p, q) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}}$$

Notice the distance to the origin to a point $p = (x, y)$ is given by $d(p, 0) = \sqrt{x^2 + y^2}$, you can appreciate the similarity to the distance in the one-dimensional case where $d(x, 0) = |x| = \sqrt{x^2}$. We can also calculate the midpoint of $p, q \in \mathbb{R}^2$ by simply calculating their average; $m = \frac{1}{2}(p + q)$. We define addition of points in the natural manner: if $p = (p_1, p_2)$ and $q = (q_1, q_2)$ then $p + q = (p_1 + q_1, p_2 + q_2)$ and multiplying by $\frac{1}{2}$ is likewise defined to mean $\frac{1}{2}(p+q) = \left(\ \frac{1}{2}(p_1+q_1),\ \frac{1}{2}(p_2+q_2)\ \right)$





The angle between two rays or line segments is often of interest. The sine, cosine and tangent functions are key tools in such analysis. A **right triangle** is one which has a **right-angle** at one corner (a right angle is measured to be 90 degrees or $\pi/2$ radians.) I have provided a quick reminder of how sine, cosine and tangent are defined in for a right triangle in the diagram above. I should also mention a little about the theory of **conic sections**. If you study the curves formed by cutting a double-cone by a plane then you find the possibilities include points, lines, parabolas, ellipse and hyperbolas. I created this picture with the Tikz package in LaTeX. Thanks to Mark Wibrow for posting code which I found at this website.



In fact, there are **geometric definitions** which are at times important to know:

(1.) a **line** is the collection of all points in the plane equidistant from a pair of distinct focal points.

(2.) a **parabola** is the collection of all points in the plane equidistant from a given line and focal point.

(3.) a **ellipse** is the collection of all points in the plane for which the sum of the distance from one focal point and the distance from a second focal point is held constant.

(4.) a **hyperbola** is the collection of all points in the plane for which the difference of the distance from one focal point and the distance from a second focal point is held constant.

Even more important, there are **algebraic definitions** in the special cases that the focal points or lines fall on a line parallel to the coordinate axes.

(1.) a **line** through (x_1, y_1) and (x_2, y_2) is the set of all (x, y) for which:

$$(x - x_2)(y - y_1) = (x - x_1)(y - y_2)$$

If $x_1 = x_2$ then the line is **vertical** and has equation $x = x_1$. Otherwise, the slope of the line is given by $m = \frac{y_2 - y_1}{x_2 - x_1}$ and the equation of the line in **point-slope** form is $y = y_1 + m(x - x_1)$. If $y = mx + b$ then the equation of the line is in **slope-intercept** form.

(2.) If $A \neq 0$, then a vertical parabola with vertex (h, k) is the solution of $y = A(x-h)^2 + k$. Note $A > 0$ gives an **upward opening parabola** and $A < 0$ gives a downward opening parabola. Likewise, for $B \neq 0$ we obtain a **horizontal parabola** with vertex (h, k) as the solution set of $x = h + B(y - k)^2$. If $B > 0$ the horizontal parabola is **right-opening** whereas if $B < 0$ the parabola is **left-opening**.

(3.) Let $a, b > 0$ then the solution set of $\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{a^2} = 1$ is an **ellipse** centered at (h, k).

(4.) Let $a, b > 0$ then the solution set of $\frac{(x-h)^2}{a^2} - \frac{(y-k)^2}{a^2} = \pm 1$ is an **hyperbola** centered at (h, k) with **slant asymptotes** $y = k \pm \frac{b}{a}(x - h)$

Only the definition of line given above is general. Parabolas, hyperbolas and ellipse can have more complicated equations than the ones given above. The complication comes from rotating the curves. Still, it is possible to give one overarching equation to rule them all and in the algebra bind them:

Definition 7.1.1. *Given constants A, B, C, D, E, F the solution set of*

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

is a **conic section***. We say the conic section is degenerate if it is a line or a point.*

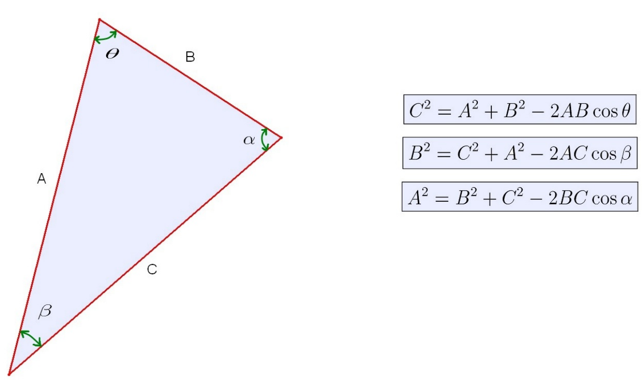
Please see this Desmos demonstration to see how varying the constants A, B, C, D, E, F produces all the conic sections. To properly unwind this precious equation we need the linear algebra of eigenvectors and the real spectral theorem. However, I believe detailed analysis of the rotated case can also be found in the excellent and far from lazy calculus text by Howard Anton[1]. This concludes our short tour of analytic geometry. We will find occasion to use these tools throughout this course. If you want to read more from a calculus II level perspective, you might take a look at this little article I wrote on how to parametrize such curves.

7.2 trigonometry

In this section I try to present most if not all the useful trigonometric identities for calculus. It is not too hard to prove that the law of cosines follows from the Pythagorean Theorem: if A, B, C are the lengths of the sides of a triangle with angle θ opposite C then

$$\boxed{C^2 = A^2 + B^2 - 2AB\cos\theta}$$

Note that when $\theta = \frac{\pi}{2}$ we recover the usual identity $C^2 = A^2 + B^2$. The law of cosines applies to arbitrary triangles whereas the Pythagorean theorem only applies to right-triangles.



[1]Anton puts most books these days to shame

With a little trouble and ingenuity you can use the Law of cosines applied to certain pictures to deduce the fundamental identities which I refer to as the **adding angles identities**

$$\boxed{\cos(\theta+\beta) = \cos\theta\cos\beta - \sin\theta\sin\beta}$$

$$\boxed{\sin(\theta+\beta) = \sin\theta\cos\beta + \cos\theta\sin\beta}$$

With these two identities we can derive most anything we want. The examples that follow are in no particular order. I only use the adding angle identities and the definitions of tangent plus a little algebra.

Example 7.2.1.

$$\begin{aligned} \tan(\theta+\beta) &= \frac{\sin(\theta+\beta)}{\cos(\theta+\beta)} \\ &= \frac{\sin\theta\cos\beta + \cos\theta\sin\beta}{\cos\theta\cos\beta - \sin\theta\sin\beta} \\ &= \frac{\frac{\sin\theta\cos\beta}{\cos\theta\cos\beta} + \frac{\cos\theta\sin\beta}{\cos\theta\cos\beta}}{\frac{\cos\theta\cos\beta}{\cos\theta\cos\beta} - \frac{\sin\theta\sin\beta}{\cos\theta\cos\beta}} \quad \Rightarrow \quad \boxed{\tan(\theta+\beta) = \frac{\tan\theta+\tan\beta}{1-\tan\theta\tan\beta}} \end{aligned}$$

While we are on this example, note if $\theta = \beta$ then we find

$$\boxed{\tan(2\theta) = \frac{2\tan\theta}{1-\tan^2\theta}}$$

Example 7.2.2. *The case $\theta = \beta$ gives interesting formulas for sine and cosine,*

$$\cos(\theta+\theta) = \cos\theta\cos\theta - \sin\theta\sin\theta \quad \Rightarrow \quad \boxed{\cos(2\theta) = \cos^2\theta - \sin^2\theta.}$$

Likewise,

$$\sin(\theta+\theta) = \sin\theta\cos\theta + \cos\theta\sin\theta \quad \Rightarrow \quad \boxed{\sin(2\theta) = 2\sin\theta\cos\theta.}$$

Since $\cos^2\theta + \sin^2\theta = 1$ thus $\sin^2\theta = 1 - \cos^2\theta$ it follows that $\cos(2\theta) = 2\cos^2\theta - 1$ hence

$$\boxed{\cos^2\theta = \frac{1}{2}\Big(1+\cos(2\theta)\Big)}$$

Similarly we can solve for $\sin^2\theta$ to obtain,

$$\boxed{\sin^2\theta = \frac{1}{2}\Big(1-\cos(2\theta)\Big)}$$

7.3 complex numbers and trigonometry

Naturally, we can continue in the fashion of the previous section to derive a great variety of trigonometric identities. However, there is something somewhat unsatisfying about this method. The calculation is indirect. Suppose we wanted to simplify the expression $\sin(\theta)\cos(4\theta)$. How would we do it? To be fair, there are identities for $\sin(\theta)\sin(\beta), \cos(\theta)\cos(\beta)$ and $\sin(\theta)\cos(\beta)$ so we could just look those up and go from there. But, is there a better way to remember all these facts? Is there some elegant formula which encapsulates all these trigonometric identities and reduces these problems to little more than algebra? In fact, yes. However, it comes at the price of understanding a bit of basic complex variables. I would argue that this is a worthy price since most students need to learn more about complex numbers anyway.

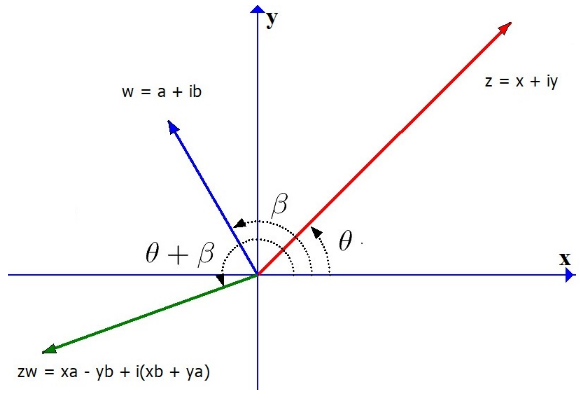
We usually denote a complex numbers $a + ib$ for $a, b \in \mathbb{R}$. Alternatively, perhaps you've see the notation $a + b\sqrt{-1}$. But, what is a complex number[2]? In terms of the axioms of real numbers we can prove $\sqrt{-1} \notin \mathbb{R}$. What then is this odd quantity of $\sqrt{-1}$? Gauss gave an answer to this question in terms of explicitly real mathematics. Gauss showed how to build complex numbers from real numbers. In particular, he said complex numbers could be identified with pairs of real numbers that enjoy a certain rather beautiful multiplication; $\mathbb{C} = \mathbb{R}^2$ where $(a, b) * (c, d) = (ac - bd, ad + bc)$. This is usually denoted

$$\boxed{(a + ib)(c + id) = ac + iad + ibc + i^2bd = ac - bd + i(ad + bc)}$$

Where we denoted $i = (0, 1)$ hence $i^2 = (0, 1) * (0, 1) = -(1, 0)$ and since $(1, 0) * (a, b) = (a, b)$ we denote $(1, 0) = 1$ hence the relation $i^2 = -1$. In fact, that was the whole reason to define this funny multiplication $*$, Gauss wanted a formal system to construct a number with the property $i^2 = -1$. This number i was termed "imaginary" since it didn't fall into the category of the real numbers, it has different properties. In truth, imaginary numbers are just as real in the philosophical sense as real numbers. In any event, we should remember $(a, b) = a + ib$ guides our visualization of $\mathbb{C}$ as the xy-plane $\mathbb{R}^2$.

Complex numbers can be added, subtracted, multiplied and divided just the same as real numbers. Geometrically the multiplication of complex numbers is very interesting. When we multiply complex numbers z and w the length of zw is found to be the product of the lengths of z and w. In addition, the standard angle of zw is the sum of the standard angles of z and w respective. This tight correspondence between geometry and algebra is part of what makes complex numbers so incredible useful. Here is a picture of my claim:

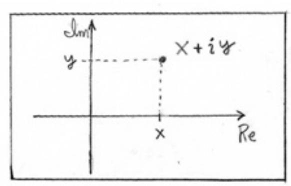
[2] the answer I give here is just one of several popular constructions. We could also build complex numbers from 2×2 matrices or a rather abstract construction called a "field extension". This is the construction most accessible at this point in your education



Continuing, each complex number have a real and imaginary part,

$$\boxed{Re(a,b) = Re(a+ib) = a \qquad Im(a,b) = Im(a+ib) = b}$$

In general if $z \in \mathbb{C}$ then $z = Re(z) + iIm(z)$. It should be emphasized that $Re(z), Im(z) \in \mathbb{R}$ so there is a natural correspondence between complex numbers and the Cartesian Plane $\mathbb{R}^2$; I use this correspondence when I write $(x, y) = x + iy$. This plane is called the complex plane. The x-axis is called the real-axis, the y-axis is called the imaginary-axis. Sometimes also called an Argand diagram,



Suppose $x, y \in \mathbb{R}$ in what follows. Every complex number $z = x + iy$ has a complex-conjugate $\bar{z} = x - iy$. In the complex plane the mapping $z \to \bar{z}$ is a reflection across the x-axis. We define the **modulus** or **length** of a complex number as follows:

$$\boxed{|x+iy| = \sqrt{x^2+y^2}}.$$

Notice $z = x + iy$ and $\bar{z} = x - iy$ have $z\bar{z} = (x + iy)(x - iy) = x^2 + y^2$ thus $|z| = \sqrt{z\bar{z}}$. However, you can also verify $\overline{zw} = \overline{zw}$ and it follows easily that $|zw| = |z||w|$. This algebra verifies my earlier assertion that the length $|zw|$ of zw is the product of the lengths $|z|$ and $|w|$.

7.3.1 the complex exponential

It is likely you will motivate this formula in the complex variables course. Ultimately there are many ways to understand the definition given below is the only definition which is natural, however most of those explanations involve calculus. That said, we can understand the necessity of the definition from a purely algebraic/geometric viewpoint: if the exponential function is to be defined on the complex plane then

1. any complex exponential function should restrict to the real exponential function on the real axis in $\mathbb{C}$.

2. rotations in the plane transform a point (x, y) to a new point $(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$ and in complex notation that factors to $(x+iy)(\cos\theta + i\sin\theta)$. If we rotated again by angle β then the point would be transformed to $(x+iy)(\cos(\theta+\beta) + i\sin(\theta+\beta))$. This means the transformation is like the real exponential function which also has $e^a e^b = e^{a+b}$.

These two ingredients go together to suggest the following definition (of course, definitions don't have to be motivated, I'm just trying to give you some idea of how you could derive such a rule).

Definition 7.3.1. *complex exponential function.*

We define $exp : \mathbb{C} \to \mathbb{C}$ by the following formula:

$$exp(z) = exp(Re(z) + iIm(z)) = e^{Re(z)}\big[\cos(Im(z)) + i\sin(Im(z))\big].$$

We can show that this definition yields the following desirable properties:

1. $e^{Re(z)} = Re(exp(z))$

2. $exp(iIm(z)) = \cos(Im(z)) + i\sin(Im(z))$

3. $exp(0) = 1$

4. $exp(z+w) = exp(z)exp(w)$

5. $exp(-z) = \frac{1}{exp(z)}$

6. $exp(z) \neq 0$ for all $z \in \mathbb{C}$

Here $e^{Re(z)}$ denotes the plain-old real exponential function which we will investigate in depth as this course progresses. Essentially, the second condition says that the complex exponential function must reproduce the real exponential function when the input is a complex number with zero imaginary part. The proof of (1.) is simple, just note $\cos(0) = 1$ and $\sin(0) = 0$ hence (2.) follows. Condition 2.) is called Euler's identity. The proof of (2.) is simple as well, just notice $e^0 = 1$ then observe that the definition reduces to Euler's identity. Again, the proof of (3.) is simple, $e^0 = e^{0+i0} = e^0(\cos(0) + i\sin(0)) = 1$.

Let's examine the proof of 4.). Suppose that $z = x+iy$ and $w = a+ib$ where $x, y, a, b \in \mathbb{R}$. Observe:

$$\begin{aligned}
exp(z+w) &= exp(x+iy+a+ib) \\
&= exp(x+a+i(y+b)) \\
&= e^{x+a}\big(\cos(y+b)+i\sin(y+b)\big) && \text{defn. of complex exp.} \\
&= e^{x+a}\big(\cos y\cos b-\sin y\sin b+i[\sin y\cos b+\sin b\cos y]\big) && \text{adding angles formulas} \\
&= e^{x+a}\big(\cos y+i\sin y\big)\big(\cos b+i\sin b\big) && \text{algebra} \\
&= e^{x}e^{a}\big(\cos y+i\sin y\big)\big(\cos b+i\sin b\big) && \text{law of exponents} \\
&= e^{x+iy}e^{a+ib} && \text{defn. of complex exp.} \\
&= exp(z)exp(w).
\end{aligned}$$

To prove (5.) we can make use of (3.) and (4.),

$$exp(z)exp(-z) = exp(z-z) = exp(0) = 1 \quad \Rightarrow \quad exp(-z) = \frac{1}{exp(z)}.$$

Note that the equation above implies that $exp(z) \neq 0$ for all $z \in \mathbb{C}$ so we have proof for (6.). I will use the notation $e^z = exp(z)$ from this point onward[3]

7.3.2 polar form of a complex number

We argued that sine and cosine are defined in Quadrants II,III and IV in order to extend right triangle geometry from Quadrant I in the natural way. In other words, sine and cosine are defined to force the polar coordinate formulas to be valid[4]

$$x = r\cos\theta \qquad y = r\sin\theta$$

To make connection with complex numbers unambiguously let's suppose we have $r = \sqrt{x^2+y^2}$ and $0 \leq \theta \leq 2\pi$. Consider a complex number $z = x+iy$, convert it to polar coordinates by substituting the polar coordinate transformations above:

$$z = x+iy = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Definition 7.3.2. *polar form of complex number.*

[3] I have emphasized the ways in which the complex exponential is similar to the real exponential, but be warned there is much more to say. For example, $exp(z+2n\pi i) = exp(z)$ because the sine and cosine functions are 2π-periodic. But, this means that the exponential is not 1-1 and consequently one cannot solve the equation $e^z = e^w$ uniquely. This introduces all sorts of ambiguities into the study of complex equations. Given $e^z = e^w$, you cannot conclude that $z = w$, however you can conclude that there exists $n \in \mathbb{Z}$ and $z = w + 2n\pi i$. In the complex variables course you'll discuss local inverses of the complex exponential function, instead of just one natural logarithm there are infinitely many to use.

[4] my viewpoint, it doesn't have to be yours, there are lots of ways to think about sine and cosine

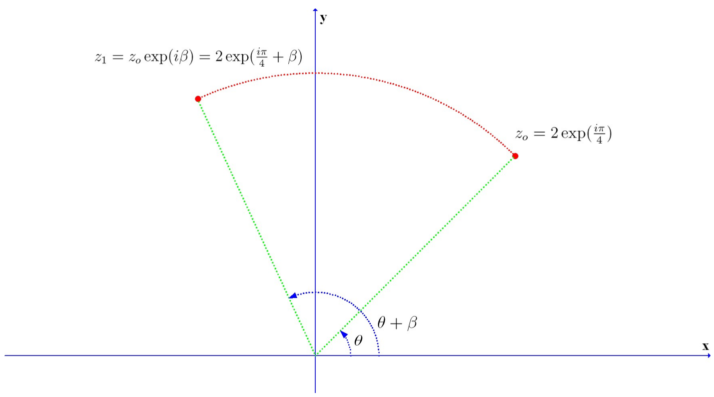
Suppose the Cartesian form of a complex number z is given $z = x + iy$ then the polar form of the complex number is $z = rexp(i\theta)$ where $r = \sqrt{x^2 + y^2}$ and θ is the standard angle of (x, y) measured counterclockwise from the positive real axis.

Example 7.3.3. *Let* $z = 2 + 2i$ *then* $r = \sqrt{4+4} = \sqrt{8}$ *whereas* $\tan\theta = \frac{y}{x} = \frac{2}{2} = 1$ *hence* $\theta = \frac{\pi}{4}$. *The polar form is* $z = \sqrt{8}\ exp(\ i\frac{\pi}{4}\ )$.

Example 7.3.4. *Let* $z = 2 + 2i$ *and multiply by* $exp(i\beta)$. *We found the polar form of* z *in the last example is* $z = \sqrt{8}\ exp(\ i\frac{\pi}{4}\ )$.

$$zw = \sqrt{8}\ exp(\ i\tfrac{\pi}{4}\ )exp(i\beta) = \sqrt{8}\ exp\big[\ i(\tfrac{\pi}{4} + \beta)\ \big]$$

Multiplication of a complex number z *by* $exp(i\beta)$ *rotates* z *by an angle of* β *in the counterclockwise direction.*



In electrical engineering complex numbers are used to represent the **impedance** of some circuit. Inductance and capacitance are give a complex resistance which depends on the frequency of the current present in the circuit. This **phasor** method allows you to solve alternating current problems as if they were direct current. Beware, $j = \sqrt{-1}$ in their formalism because i is used for current. If I was a electrical engineering major then I would make it a point to take linear algebra and complex variables and differential equations as soon as possible. It would help you to see past the math and focus on the engineering[5].

7.3.3 the algebra of sine and cosine

Euler's identity is beautiful on its own, but the following formulas are the most of the reason I'm bothering to type up these notes. Simply add and subtract $e^{i\theta} = \cos\theta + i\sin\theta$ and $e^{-i\theta} = \cos\theta - i\sin\theta$ to obtain,

$$\boxed{\cos\theta = \frac{1}{2}\Big(e^{i\theta} + e^{-i\theta}\Big) \qquad \sin\theta = \frac{1}{2i}\Big(e^{i\theta} - e^{-i\theta}\Big).}$$

[5] of course, given all that you might as well add a math major so you have more options later in life.

Example 7.3.5. *Suppose you want to derive a nice formula for the square of cosine. Just plug in the boxed formula and use the laws of exponents we proved for the complex exponential:*

$$\begin{aligned}
\cos^2\theta &= \left[\frac{1}{2}\left(e^{i\theta}+e^{-i\theta}\right)\right]^2 \\
&= \frac{1}{4}\left(e^{i\theta}e^{i\theta}+2e^{i\theta}e^{-i\theta}+e^{-i\theta}e^{-i\theta}\right) \\
&= \frac{1}{4}\left(e^{2i\theta}+2+e^{-2i\theta}\right) \\
&= \frac{1}{2}+\frac{1}{2}\frac{1}{2}\left(e^{2i\theta}+e^{-2i\theta}\right) \\
&= \frac{1}{2}+\frac{1}{2}\cos 2\theta \\
&= \frac{1}{2}\left(1+\cos 2\theta\right).
\end{aligned}$$

Example 7.3.6. *Suppose you want to derive a nice formula for the square of sine. Just plug in the boxed formula and use the laws of exponents we proved for the complex exponential:*

$$\begin{aligned}
\sin^2\theta &= \left[\frac{1}{2i}\left(e^{i\theta}-e^{-i\theta}\right)\right]^2 \\
&= \frac{-1}{4}\left(e^{i\theta}e^{i\theta}-2e^{i\theta}e^{-i\theta}+e^{-i\theta}e^{-i\theta}\right) \\
&= \frac{-1}{4}\left(e^{2i\theta}-2+e^{-2i\theta}\right) \\
&= \frac{1}{2}-\frac{1}{2}\frac{1}{2}\left(e^{2i\theta}+e^{-2i\theta}\right) \\
&= \frac{1}{2}-\frac{1}{2}\cos 2\theta \\
&= \frac{1}{2}\left(1-\cos 2\theta\right).
\end{aligned}$$

The identities above you should have memorized anyway, but I don't have to memorize them since I can **derive** them in a pinch. In contrast, the next example is not one for which I could typically quote the answer off the top of my head:

Example 7.3.7. *Same method again. Covert given functions to imaginary exponentials and do* *algebra until you see sines and cosines again. Simple as that.*

$$\begin{aligned}
\cos(x)\sin(4x) &= \frac{1}{2}\left(e^{ix}+e^{-ix}\right)\frac{1}{2i}\left(e^{4ix}-e^{-4ix}\right) \\
&= \frac{1}{4i}\left(e^{5ix}-e^{-3ix}+e^{3ix}-e^{-5ix}\right) \\
&= \frac{1}{2}\left[\frac{1}{2i}\left(e^{5ix}-e^{-5ix}\right)+\frac{1}{2i}\left(e^{3ix}-e^{-3ix}\right)\right] \\
&= \frac{1}{2}\sin(5x)+\frac{1}{2}\sin(3x)
\end{aligned}$$

You could calculate identities for $\cos(ax)\cos(bx), \sin(ax)\sin(bx)$ by much the same calculation and you'd find a sum of cosines for each:

$$\boxed{\cos(ax)\cos(bx) = \frac{1}{2}\cos[(a+b)x]+\frac{1}{2}\cos[(a-b)x]}$$

$$\boxed{\sin(ax)\sin(bx) = \frac{1}{2}\cos[(a-b)x]-\frac{1}{2}\cos[(a+b)x]}$$

On the other hand, generally $\cos(ax)\sin(bx)$ yields a sum of sines,

$$\boxed{\cos(ax)\sin(bx) = \frac{1}{2}\sin[(a+b)x]+\frac{1}{2}\sin[(a-b)x]}$$

Naturally, we could also apply the method to calculate formulas for higher powers or products of sine and cosine. Just for a flavor:

Example 7.3.8.

$$\begin{aligned}
\cos^3\theta &= \left[\frac{1}{2}\left(e^{i\theta}+e^{-i\theta}\right)\right]^3 \\
&= \frac{1}{8}\left(e^{3i\theta}+3e^{i\theta}+3e^{-i\theta}+e^{-3i\theta}\right) \\
&= \frac{3}{4}\sin(\theta)+\frac{1}{4}\sin(3\theta).
\end{aligned}$$

DeMoivres' theorem in complex notation is simply $(e^{i\theta})^n = e^{in\theta}$. When you unfold this into sines and cosines the result is amazing:

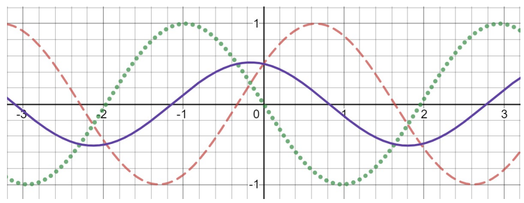
$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$$

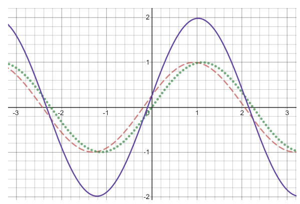
You can try plugging in $n = 2$ or $n = 3$ and you'll find yet more identities which are less than obvious from other approaches.

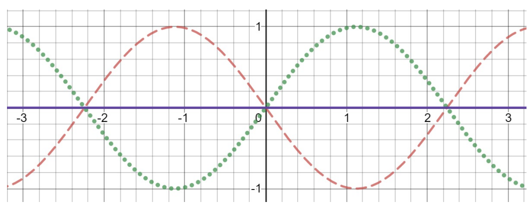
7.3.4 superposition of waves and the method of phasors

Sinusoidal waves on a string have the form $y = A\sin(kx - \omega t) + \phi)$. This wave has amplitude A, wave number k, angular frequency ω and phase ϕ. If we have two such waves on a string or some other medium then they combine to create a new wave.

I made a few examples in Desmos to explore interference. click here to explore what is pictured below. The graph above has $y = \sin(\omega t + \phi_1)$ as the red-dotted curve and $y = \sin(\omega t + \phi_2)$ as the green-dots curve. The sum $y = \sin(\omega t + \phi_1) + \sin(\omega t + \phi_2)$ is the purple curve. I illustrate superposition in several cases below:





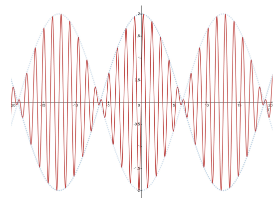


The picture above illustrates **destructive interference**. In all the cases above we are adding waves with the same frequency, only the phase differs.

The mathematics of a simple case is encapsulated in the following trigonometric identity:

$$\boxed{\sin(a) + \sin(b) = 2\sin\left(\frac{a+b}{2}\right)\cos\left(\frac{a-b}{2}\right)}$$

This formula explains what we saw above and it also leads to the phenomenon of beats. When we add two waves with different frequencies the result is the sine and cosine terms oscillate at different rates. The sine is high frequency whereas the cosine is low frequency if we set $a = \omega_1 t$ and $b = \omega_2 t$ where $\omega_1 > \omega_2$. Note $a+b = (\omega_1+\omega_2)t$ whereas $a-b = (\omega_1-\omega_2)t$. We can think of the cosine term as a slowly varying amplitude. I plot $y = \pm\cos(\omega_1 - \omega_2)t$ in the blue dots and $y = \sin\omega_1 t + \sin\omega_2 t$ as the red curve. The plot below has $\omega_1 = 5$ and $\omega_2 = 4.5$.



See this Desmos demonstration to adjust or animate difference cases for ω_1, ω_2 in the plot above.

7.3.5 standing waves

Suppose we have two waves of equal amplitude ($A_1 = A_2 = A$), frequency (ω) and wavenumber (k) traveling in opposite directions, y_1 travels right and y_2 travels left,

$$y_1 = A_1 \sin\big[kx - \omega t + \phi_1\big] \qquad \text{and} \qquad y_2 = A_2 \sin\big[kx + \omega t + \phi_2\big]$$

Consider the superposition of these waves,

$$\begin{aligned} y_1 + y_2 &= A\sin\big[kx - \omega t + \phi_1\big] + A\sin\big[kx + \omega t + \phi_2\big] \\ &= 2A\sin\left[\frac{(kx - \omega t + \phi_1) + (kx + \omega t + \phi_2)}{2}\right]\cos\left[\frac{(kx - \omega t + \phi_1) - (kx + \omega t + \phi_2)}{2}\right] \\ &= 2A\sin\left[kx + \frac{\phi_1 + \phi_2}{2}\right]\cos\left[\frac{\phi_1 - \phi_2}{2} - \omega t\right] \\ &= 2A\sin\left[kx + \frac{\phi_1 + \phi_2}{2}\right]\cos\left[\omega t - \frac{\phi_1 - \phi_2}{2}\right] \end{aligned}$$

This is a standing wave with amplitude $2A$. The shape of the wave is given by the sine factor then as time evolves the second factor oscillates between -1 and 1. Perhaps you've see such a pattern, if you fix a rope to a wall then swing the free end you can set up two waves, one created by your waving, the other created by the reflection of your wave off the wall. The net result is the appearance of a wave that stands still. A standing wave. Similar mathematics applies to the patterns of pressure variation in pipe organs. Again a addition of sines or cosines will describe how the notes combine within the instrument.

Problem: find the amplitude and phase of $y_1 + y_2$ given that $y_1 = A_1 \sin(kx + \phi_1)$ and $y_2 = A_2 \sin(kx + \phi_2)$. Also, derive a similar result for cosine.

Solution: Define $\widetilde{y_1} = A_1 e^{i(kx+\phi_1)}$ and $\widetilde{y_2} = A_2 e^{i(kx+\phi_2)}$. Notice that these complex functions contain both the sine and cosine functions we wish to add: in particular

$$y_1 = Im(\widetilde{y_1}) = A_1 \sin(kx + \phi_1), \qquad y_2 = Im(\widetilde{y_2}) = A_2 \sin(kx + \phi_2)$$

The real parts will give us cosines instead. We can calculate the sum of the sine functions by instead adding the corresponding complex functions,

$$\begin{aligned}
\widetilde{y_1} + \widetilde{y_2} &= A_1 e^{i(kx+\phi_1)} + A_2 e^{i(kx+\phi_2)} \\
&= A_1 e^{ikx} e^{i\phi_1} + A_2 e^{ikx} e^{i\phi_2} \\
&= [A_1 e^{i\phi_1} + A_2 e^{i\phi_2}] e^{ikx} \\
&= A e^{i\gamma} e^{ikx} && (A \text{ and } \gamma \text{ from picture}) \\
&= A e^{i[kx+\gamma]} \\
&= A\cos[kx+\gamma] + iA\sin[kx+\gamma]
\end{aligned}$$

As is often the case with complex variables we solved two real problems at once. Equating the real and imaginary parts of the equation above yields

$$\boxed{A_1 \sin(kx+\phi_1) + A_2 \sin(kx+\phi_2) = A\sin[kx+\gamma]}$$

$$\boxed{A_1 \cos(kx+\phi_1) + A_2 \cos(kx+\phi_2) = A\cos[kx+\gamma]}$$

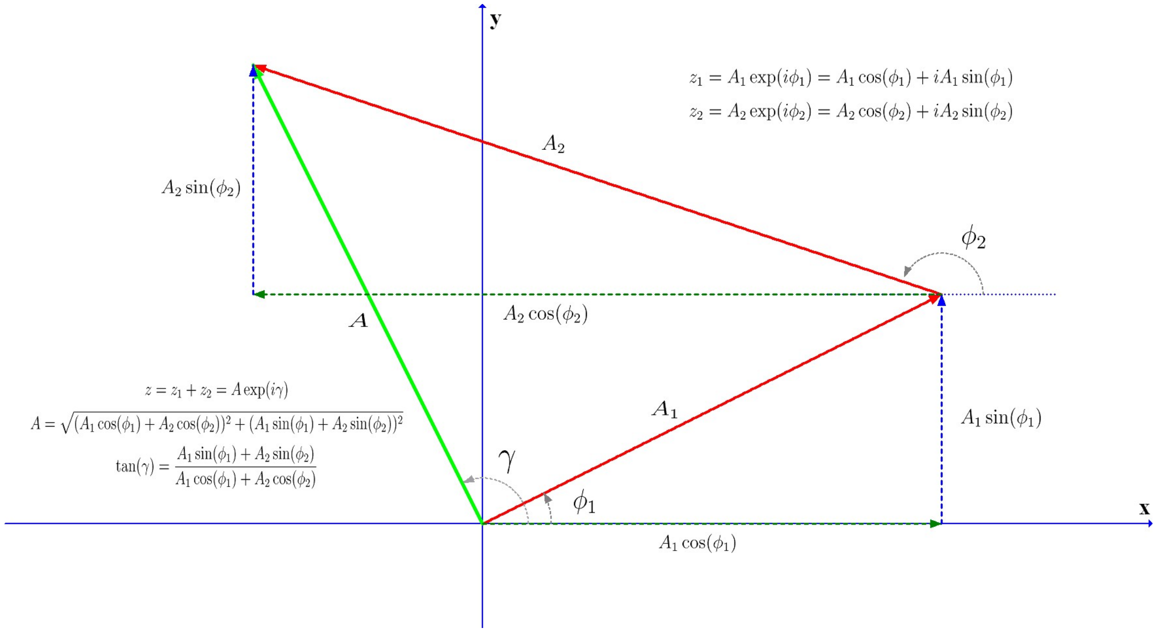
where γ is implicitly defined by

$$\tan\gamma = \frac{A_1 \sin(\phi_1) + A_2 \sin(\phi_2)}{A_1 \cos(\phi_1) + A_2 \cos(\phi_2)}.$$

And A is defined by

$$A = \sqrt{[A_1 \cos(\phi_1) + A_2 \cos(\phi_2)]^2 + [A_1 \sin(\phi_1) + A_2 \sin(\phi_2)]^2}$$

I've drawn the picture in quadrants I and II but the argument is general.



If you had highschool physics you should recognize the construction above as the so-called "tip-2-tail" method of vector addition. Have no fear, it's just trigonometry. Moreover, it is no hard to see the calculation above easily generalizes to three or more vectors.

Finally, I would just mention that sines and cosines are important even though most waves are not sinusoidal. Typically waves come in finite packets and their precise mathematical account requires much more sophisticated terminology. The *Fourier decomposition* breaks down a waveform into a sum of sines or cosines. Most digital formats of music are based on transforming the music into its Fourier equivalent then devising clever methods to compress this data. In contrast, compression of visual data is better accomplished with something called *wavelets*. The popular jpg-format is based on wavelets.

7.4 functions

The term **function** is about a third of a milennia old. It was first used by Leibniz in about 1700. More recently the term function has gained a rigorous and precise meaning. To say f is a **function from** A **to** B means that for each $a \in A$ the function f assigns a particular element $b \in B$. We denote this by saying that $f(a) = b$ or we can equivalently denote $a \mapsto f(a)$.

Definition 7.4.1. *function*

We say f is a **function from A to B** if $f(a) \in B$ for each $a \in A$ and the value $f(a)$ is a single value. We denote $f : A \to B$ in this case and we say that $\mathbf{A = domain(f)}$ and $\mathbf{B = codomain(f)}$. Furthermore, we say that f is an **B-valued function of A**. If $A = B$ then we may say that f is a function **on** A. If $A \subseteq \mathbb{R}$ then f is said to be a **function of a real variable**. If $B \subseteq \mathbb{R}$ then f is said to be a **real-valued function**. If $B \subset \mathbb{C}$ then f is said to be a **complex-valued function**.

7.4.1 local inverses

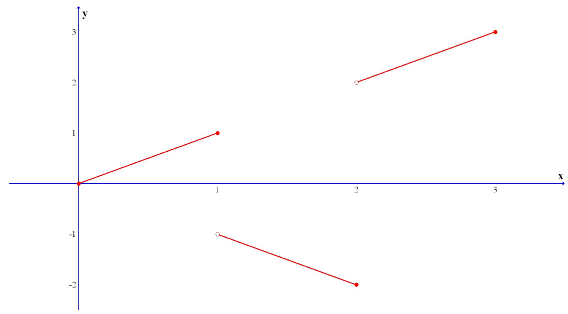
It does seem geometrically obvious that if the restriction of a function passes the horizontal line test with respect to a connected set then the same function ought to be either strictly decreasing or strictly increasing on that set.

Proposition 7.4.2. *strictly monotonic functions are injective.*

If f is either strictly increasing or strictly decreasing on $U \subseteq \mathbb{R}$ then f is injective on U.

Proof: assume that f is strictly increasing on U then for all $x, y \in U$ such that $x < y$ we have that $f(x) < f(y)$. Let $a, b \in U$ and suppose $f(a) = f(b)$ (*we seek to show $a = b$ since that proves that f is injective on U*). If $a = b$ then we're done. Suppose that $a < b$ then $f(a) < f(b)$ which contradicts $f(a) = f(b)$. Likewise, if $b < a$ then $f(b) < f(a)$ which contradicts $f(a) = f(b)$. Therefore, since otherwise we find a contradiction, the only possibility is that $a = b$. Thus f is 1-1 on U. If f is decreasing then the proof is similar. $\square$

I would like to offer a converse to this proposition. If a function is $1-1$ then it is either increasing or decreasing, however, there are counter-examples. For example, $f(x) = \begin{cases} x & 0 \leq x \leq 1, \\ -x & 1 < x \leq 2 \\ x & 2 < x \leq 3 \end{cases}$ is injective but is neither increasing nor decreasing on $[0, 3]$. Here is a graph of this funny function:



If we wish to obtain a converse to the proposition then we will need to add additional hypothesis to avoid the counter-examples like the one offered above.

Proposition 7.4.3. *inverse functions also increase or decrease.*

> Suppose $f : U \to V$ is either strictly increasing or strictly decreasing on $U \subseteq \mathbb{R}$ then $f^{-1} : V \to U$ is likewise either strictly increasing or decreasing on V.

Proof: suppose $f : U \to V$ is strictly increasing with inverse $f^{-1} : V \to U$. Suppose $a, b \in V$ such that $a < b$ and suppose $f^{-1}(a) = x$ and $f^{-1}(b) = y$. There exist three possibilities:

1. $f^{-1}(a) = f^{-1}(b)$ which implies $f(f^{-1}(a)) = f(f^{-1}(b))$ thus $a = b$ which contradicts our assumption $a < b$.

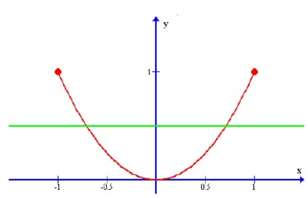
2. $f^{-1}(a) > f^{-1}(b)$ which implies $f(f^{-1}(a)) > f(f^{-1}(b))$ thus $a > b$ which contradicts our assumption $a < b$.

3. $f^{-1}(a) < f^{-1}(b)$ which implies $f(f^{-1}(a)) < f(f^{-1}(b))$ thus $a < b$ which is without contradiction of our assumption $a < b$.

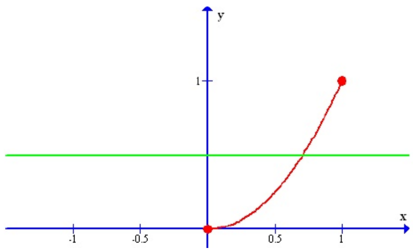
Therefore, we find for all $a, b \in V$, if $a < b$ then $f^{-1}(a) < f^{-1}(b)$ which proves f^{-1} is strictly increasing. The proof for the strictly decreasing case is similar. $\square$

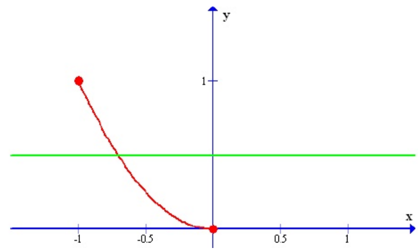
We now examine a number of examples to elaborate on the concept of a local inverse. We should see the propositions above made manifest in each case.

Example 7.4.4. *Consider $f(x) = x^2$ with dom$(f) = [-1, 1]$. We can argue algebraically that this function is not one-one since $f(a) = f(b)$ gives $a^2 = b^2$ which implies $a = \pm b$ (we needed $a = b$ to obtain injectivity). Or observe that it fails the horizontal line test:*



In contrast, the same formula with reduced domain $[-1,0]$ *or* $[0,1]$ *will pass the horizontal line test,*





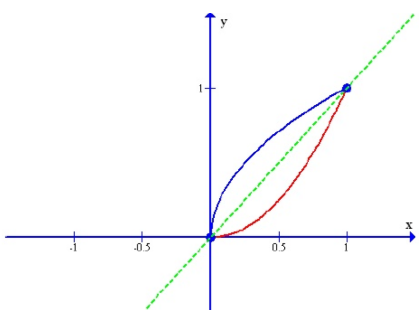
So then what is the formula for the inverse functions? We need,

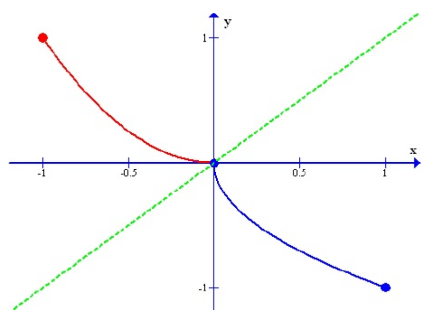
$$(i.)\ f^{-1}(f(x)) = f^{-1}(x^2) = x \qquad (ii.)\ f(f^{-1}(x)) = (f^{-1}(x))^2 = x$$

Notice that $(ii.)$ *gives* $f^{-1}(x) = \pm\sqrt{x}$*. Then substituting into* $(i.)$ *yields:* $\pm\sqrt{x^2} = x$*. But, recall that* $\sqrt{x^2} = |x|$ *so we can see that the two solutions are,*

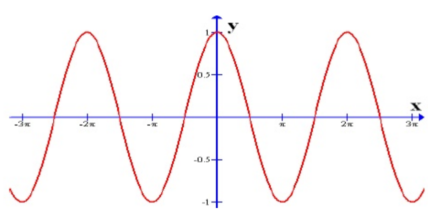
1. *If* $x \geq 0$ *then* $\sqrt{x^2} = x$ *so we choose the + solution;* $f^{-1}(x) = \sqrt{x}$
2. *If* $x \leq 0$ *then* $\sqrt{x^2} = -x$ *so we choose the - solution;* $f^{-1}(x) = -\sqrt{x}$

We find that the inverse of $f(x) = x^2$ *on* $[0,1]$ *is* $f^{-1}(x) = \sqrt{x}$ *and the inverse of* $f(x) = x^2$ *on* $[-1,0]$ *is* $f^{-1}(x) = -\sqrt{x}$*. Notice that the graphs of inverses (blue) are symmetric about the line (green).*

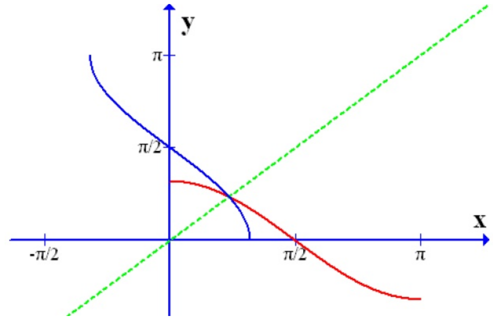




Example 7.4.5. *Let* $f(x) = \cos(x)$*. Recall the graph of the cosine function is:*

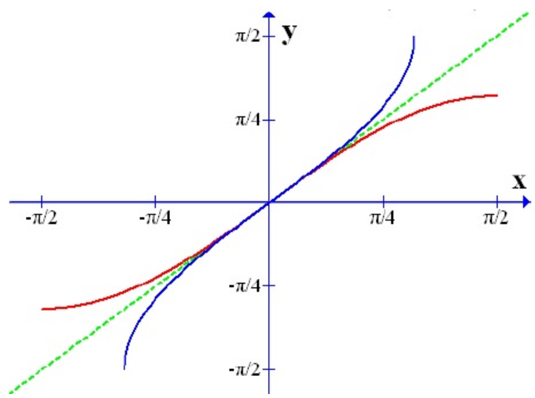


note that f *cannot have a global inverse since it is not 1-1. However, if we reduce the domain to* $[0,\pi]$ *we obtain a 1-1 function on that interval. I have graphed the local inverse in blue , and you can see that the inverse is the reflection of the graph of cosine about the line* $y = x$ *(green).*



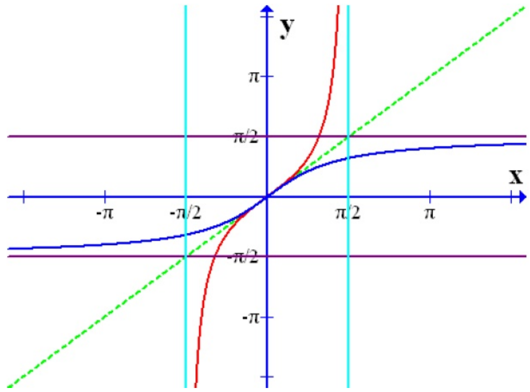
It should be understood that when we speak of inverse cosine we actually refer the local inverse for cosine on the interval $[0, \pi]$. The domain of inverse cosine is $[-1, 1]$ and the range is $[0, \pi]$. In principle one could construct other inverses for cosine based on other intervals, the choice of is simply one of convention.

Example 7.4.6. *Let $f(x) = \sin(x)$ with dom$(f) = \mathbb{R}$. This is not 1-1 because sine oscillates just like cosine. However, if we reduce the domain to $[-\pi/2, \pi/2]$ we obtain a 1-1 function on that interval (red), so we can find an inverse function(blue),*



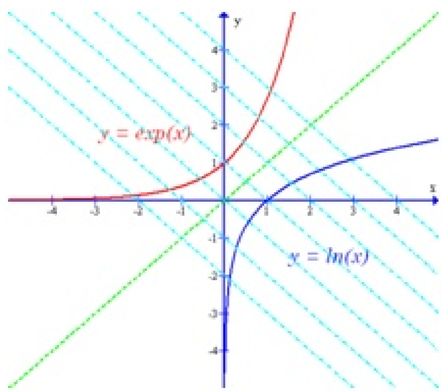
and you can see that the inverse is the reflection of the graph of sine about the line (green). The domain of inverse sine is $[-1, 1]$ and the range is $[-\pi/2, \pi/2]$. In principle one could construct other inverses for sine based on other intervals, the choice of $[-\pi/2, \pi/2]$ is simply one of convention.

Example 7.4.7. *Let $f(x) = tan(x)$ with dom$(f) = \mathbb{R} - \{n\pi + \pi/2 \mid n \in \mathbb{Z}\}$. This is not 1-1 because tangent function oscillates just like sine and cosine. However, if we reduce the domain to $(-\pi/2, \pi/2)$ we obtain a 1-1 function on that interval (red), so we can find an inverse function(blue),*



and you can see that the inverse is the reflection of the graph of tangent about the line $y = x$ (green). The domain of inverse tangent is $(-\infty, \infty)$ and the range is $(-\pi/2, \pi/2)$. I have added the vertical asymptotes of tangent in cyan at $x = \pm\frac{\pi}{2}$ you can see that the inverse tangent has horizontal asymptotes at $y = \pm\frac{\pi}{2}$. This illustrates a general pattern, vertical asymptotes for a function will morph into horizontal asymptotes for the inverse function. This helps us understand the limit of $\tan^{-1}(x)$ is as $x \to \infty$ (it's $\pi/2$).

By now you should have noticed that we can construct the inverse function's graph by reflection about the line $y = x$ (assuming that the function is 1-1 on the interval of interest). I actually use this fact to construct certain graphs.



You can draw the graph $y = e^x$ (red) then draw the line $y = x$ (green) and a bunch of perpendicular bisectors (cyan) then the graph of the inverse function $y = ln(x)$ follows. If we travel one unit from the red graph to the green line along the cyan line then the corresponding point on the blue graph is one unit further past the green line. That is the green line should intersect the cyan line at the midpoint between the intersection points of the red and blue graphs. Now, I should warn you that this advice is given for graphs with horizontal and vertical directions given the same scale. The cyan lines and the green line would take a different slant if x-axis and y-axis used a different scale.

7.5 limit proofs

Theorem 7.5.1. *two-sided limit holds if and only if both left and right limits hold.*

Let f be a function with interior limit point a. Let $L \in \mathbb{R}$,

$$\lim_{x \to a} f(x) = L \qquad \Leftrightarrow \qquad \left\{ \lim_{x \to a^+} f(x) = L \quad \text{and} \quad \lim_{x \to a^-} f(x) = L \right\}$$

Proof: to prove $\Leftrightarrow$ we must show both $\Rightarrow$ and $\Leftarrow$.

($\Rightarrow$) Begin by assuming $\lim_{x \to a} f(x) = L$ then for each $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - L| < \varepsilon$. Note for each $\varepsilon > 0$ that if $0 < x - a < \delta$ it follows $0 < |x - a| < \delta$ so $|f(x) - L| < \varepsilon$ hence $\lim_{x \to a^+} f(x) = L$. Likewise, note for each $\varepsilon > 0$ that if $-\delta < x - a < 0$ it follows $0 < |x - a| < \delta$ so $|f(x) - L| < \varepsilon$ hence $\lim_{x \to a^-} f(x) = L$.

($\Leftarrow$) We assume that both $\lim_{x \to a^+} f(x) = L$ and $\lim_{x \to a^-} f(x) = L$. Let $\varepsilon > 0$ and choose $\delta = min(\delta_+, \delta_-)$ where we use the givens to choose $\delta_+, \delta_- > 0$ such that

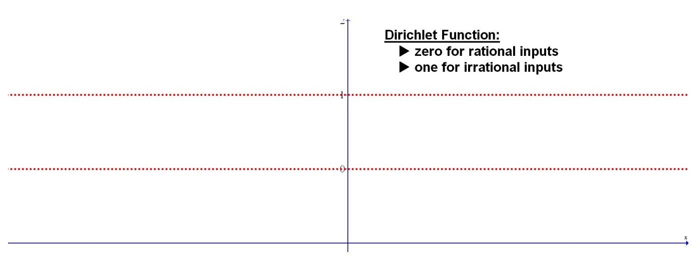
1. $0 < x - a < \delta_+$ implies $|f(x) - L| < \varepsilon$,
2. $-\delta_- < x - a < 0$ implies $|f(x) - L| < \varepsilon$

Therefore, if $x \in \mathbb{R}$ such that $0 < |x - a| < \delta \leq \delta_+, \delta_-$ then either $0 < x - a < \delta < \delta_+$ or $-\delta_- < -\delta < x - a < 0$ so by (1.) or (2.) it follows $|f(x) - L| < \varepsilon$. Therefore, the two-sided limit exists and $f(x) \to L$ as $x \to a$. $\square$.

Half the reason I include this proof is to get the math majors thinking about how to unfold the logic of the symbol $\Leftrightarrow$.

Limits work out well until mathematicians do weird stuff. For example, the function below has domain $\mathbb{R}$ hence every point is a limit point for the function, yet, there is not even one point where the limit exists.

Example 7.5.2. *The* **Dirichlet function***;* $f(x) = \begin{cases} 1 & if\ x \in \mathbb{Q} \\ 0 & if\ x \in \mathbb{J} = \mathbb{R} - \mathbb{Q} \end{cases}$. *The best I can do to the limit of the resolution is as follows:*



This function misbehaves! Think about $f(x) \to L$ as $x \to 0$. Should we expect that $L = 1$ or $L = -1$? No matter how close you get to $x = 0$ there are both rational and irrational numbers closer to $x = 0$.

7.5.1 proofs from the definition

Example 7.5.3. Problem: prove $\lim_{x\to 2}(3x+2) = 8$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $|x-2| < \delta$ implies $|f(x) - L| < \varepsilon$ for $f(x) = 3x+2$ and $L = 8$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |3x+2-8| = |3x-6| = |3(x-2)| = 3|x-2| < 3\delta = \varepsilon.$$

So, we should choose $\delta = \varepsilon/3$ since $\varepsilon > 0$ it is clear that $\delta = \varepsilon/3 > 0$. In view of these calculations we are ready to state the proof.

Proof: *Let $\varepsilon > 0$ and choose $\delta = \varepsilon/3$. Suppose $x \in \mathbb{R}$ such that $0 < |x-2| < \delta$. Observe that*

$$|3x+2-8| = |3(x-2)| = 3|x-2| < 3\delta = \varepsilon.$$

Thus $0 < |x-2| < \delta$ implies $|3x+2-8| < \varepsilon$ and it follows by the definition of the limit that $\lim_{x\to 2}(3x+2) = 8$. $\square$

Students sometimes ask me which part is the answer. My answer is that the whole proof is the answer. It is important that it contains all the proper logical statements put in the logical order. Basically, a "proof" is simply a complete explanation of why some statement is true. I will admit there is ambiguity as to what constitutes a "complete" proof in general.

Example 7.5.4. Problem: prove $\lim_{x\to 3}(2-x) = -1$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $|x-3| < \delta$ implies $|f(x)-L| < \varepsilon$ for $f(x) = 2-x$ and $L = -1$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |2-x-(-1)| = |-x+3| = |-1(x-3)| = |x-3| < \delta = \varepsilon.$$

So, we should choose $\delta = \varepsilon$.

Proof: *Let $\varepsilon > 0$ and choose $\delta = \varepsilon$. Suppose $x \in \mathbb{R}$ such that $0 < |x-3| < \delta$. Observe that*

$$|2-x-(-1)| = |-x+3)| = |x-3| < \delta = \varepsilon.$$

Thus $0 < |x-3| < \delta$ implies $|2-x-(-1)| < \varepsilon$ and it follows by the definition of the limit that $\lim_{x\to 3}(2-x) = -1$. $\square$

Example 7.5.5. Problem: prove $\lim_{x \to 0}(x^2) = 0$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $|x - 0| < \delta$ implies $|f(x) - L| < \varepsilon$ for $f(x) = x^2$ and $L = 0$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |x^2 - 0| = |x|^2 < \delta^2 = \varepsilon.$$

So, we should choose $\delta = \sqrt{\varepsilon}$. Since $\varepsilon > 0$ we can be assured that the squareroot gives $\delta > 0$.

Proof: *Let $\varepsilon > 0$ and choose $\delta = \sqrt{\varepsilon}$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. Observe that*

$$|x^2 - 0| = |x|^2 < (\sqrt{\varepsilon})^2 = \varepsilon.$$

Thus $0 < |x - 0| < \delta$ implies $|x^2 - 0| < \varepsilon$ and it follows by the definition of the limit that $\lim_{x\to 0}(x^2) = 0$. $\square$

Example 7.5.6. Problem: prove $\lim_{x \to 3}(x^2) = 9$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $|x - 3| < \delta$ implies $|f(x) - L| < \varepsilon$ for $f(x) = x^2$ and $L = 9$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |x^2 - 9| = |(x - 3)(x + 3)| < \delta|x + 3|$$

Ok, so $|x + 3|$ is annoying. But, have no fear, we control the δ. Note that $0 < |x - 3| < \delta$ gives $3 - \delta < x < 3 + \delta$ so $6 - \delta < x + 3 < 6 + \delta$. Suppose $\delta < 1$ then we certainly have that $5 < x + 3 < 7$ which gives $-7 < 5 < x + 3 < 7$ so $|x + 3| < 7$ which is very nice because, given our assumption $\delta < 1$ we find:

$$|f(x) - L| =< \delta|x + 3| < 7\delta.$$

now the choice should be clear, we use $\delta = \varepsilon/7$. However, we do need that $\varepsilon/7 < 1$, remember we don't control ε, all we know is that $\varepsilon > 0$. The solution is simple, to be careful about the possibility of large ε we choose $\delta = min(\varepsilon/7, 1)$. If $\delta = 1$ then we still find $|x+3| \leq 7$ and so $|f(x)-L| \leq 7\delta < \varepsilon$ provide that $\delta = min(\varepsilon/7, 1)$ so we knew $\delta < \varepsilon/7$ hence $7\delta < \varepsilon$.

Proof: *Let $\varepsilon > 0$ and choose $\delta = min(\varepsilon/7, 1)$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 3| < \delta$. Observe that $\delta \leq 1$ thus $0 < |x - 3| < \delta \leq 1$ yields $-1 \leq x - 3 \leq 1$ from which it follows $5 < x + 3 \leq 7$ hence $-7 < x + 3 \leq 7$ so $|x + 3| \leq 7$. Therefore,*

$$|x^2 - 9| = |(x - 3)(x + 3)| = |x - 3||x + 3| < \delta|x + 3| < 7\delta$$

Moreover, as $\delta \leq \varepsilon/7$ we have $7\delta \leq \varepsilon$. Thus, $0 < |x - 3| < \delta$ implies that $|x^2 - 9| < \varepsilon$ and it follows by the definition of the limit that $\lim_{x\to 3}(x^2) = 9$. $\square$

Sometimes we are called upon to calculate a limit which has an arbitrary limit point. In the example below the limit point is denoted by "a". We must make arguments which hold for all possible values of a since no particular restriction on a is offered.

Example 7.5.7. Problem: prove $\lim_{x \to a}(3x+2) = 3a+2$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $|x-a| < \delta$ implies $|f(x) - L| < \varepsilon$ for $f(x) = 3x+2$ and $L = 3a+2$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |3x + 2 - (3a + 2)| = |3(x - a)| = 3|x - a| < 3\delta = \varepsilon.$$

So, we should choose $\delta = \varepsilon/3$ since $\varepsilon > 0$ it is clear that $\delta = \varepsilon/3 > 0$. In view of these calculations we are ready to state the proof.

Proof: *Let $\varepsilon > 0$ and choose $\delta = \varepsilon/3$. Suppose $x \in \mathbb{R}$ such that $0 < |x - a| < \delta$. Observe that*

$$|3x + 2 - (3a + 2)| = |3(x - a)| = 3|x - a| < 3\delta = \varepsilon.$$

Thus $0 < |x - a| < \delta$ implies $|3x + 2 - (3a + 2)| < \varepsilon$ and it follows by the definition of the limit that $\lim_{x \to a}(3x + 2) = 3a + 2$. $\square$

Example 7.5.8. Problem: prove $\lim_{x \to 1^+}(\sqrt{x-1}) = 0$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *We need to show that $1 < x < 1 + \delta$ implies $|f(x) - L| < \varepsilon$ for $f(x) = \sqrt{x-1}$ and $L = 0$ and a particular choice of δ. Consider then*

$$|f(x) - L| = |\sqrt{x-1} - 0| = |\sqrt{x-1}| = \sqrt{|x-1|}.$$

where we used $1 < x < 1 + \delta$ to deduce $0 < x - 1$ hence $|x - 1| = x - 1$. We should choose $\delta = \varepsilon^2$.

Proof: *Let $\varepsilon > 0$ and choose $\delta = \varepsilon^2$. Suppose $x \in \mathbb{R}$ such that $0 < x - 1 < \delta$. Observe that*

$$|\sqrt{x-1}| = \sqrt{|x-1|} < \sqrt{\delta} = \sqrt{\varepsilon^2} = \varepsilon.$$

Thus $0 < x - 1 < \delta$ implies $|\sqrt{x-1}| < \varepsilon$ and it follows by the definition of the right-sided limit that $\lim_{x \to 1^+} \sqrt{x-1} = 0$. $\square$

Notice $f(x) = \sqrt{x-1}$ has $dom(f) = [1, \infty)$ and $x = 1$ is the boundary point of the domain. The two-sided limit is not defined at one because the function is not real-valued for $x < 1$.

Example 7.5.9. Problem: prove $\lim_{x \to 0} \frac{1}{x} \notin \mathbb{R}$ via the $\varepsilon\delta$-definition of the limit.

Preparatory calculations: *think about it. What do we need to show to show it is impossible for any real number to be the limit of $\frac{1}{x}$ as $x \to 0$?. By the proposition we just proved it would suffice to show that the right-limit failed to exist no matter what our choice of L is. Let's proceed from that angle. We want to show that $\lim_{x \to 0^+} \frac{1}{x}$ cannot be a real number. The natural thing to try here is contradiction, we suppose that there does exist $L \in \mathbb{R}$ such that $\lim_{x \to 0^+} \frac{1}{x} = L$ and then we hunt for something insane. Once we find the insanity we see that believing in the existence of $L \in \mathbb{R}$*

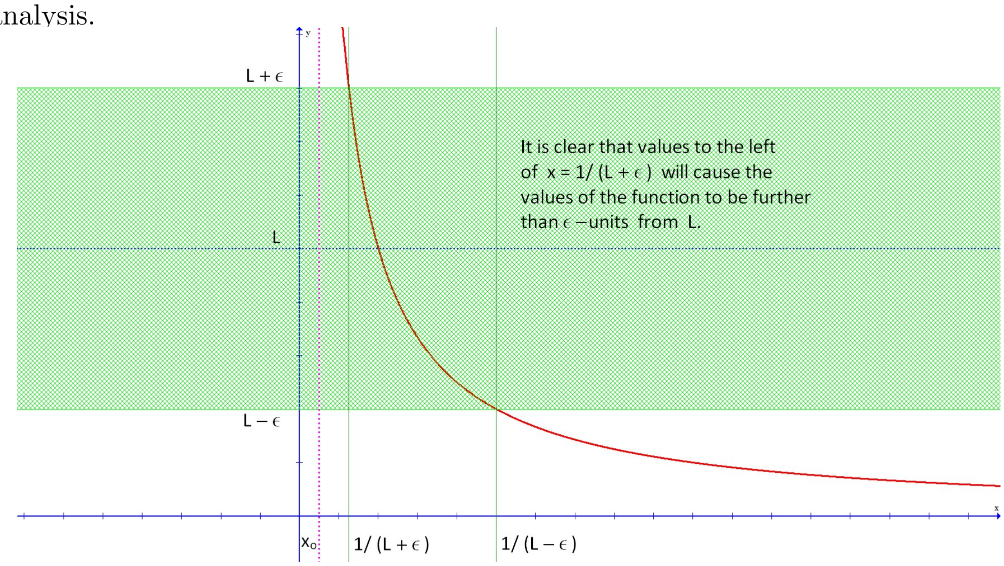
is madness so we can safely assume $L \notin \mathbb{R}$. This is the outline of the logic. Let's get into the details:

Proof: *assume that $L \in \mathbb{R}$ such that $\frac{1}{x} \to L$ as $x \to 0^+$. This means that for each $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < x < \delta$ implies $\left|\frac{1}{x} - L\right| < \varepsilon$. We seek a contradiction, suppose $\varepsilon = L$ and let $\delta > 0$ be some number such that all $x \in \mathbb{R}$ satisfying $0 < x < \delta$ force $\left|\frac{1}{x} - L\right| < \varepsilon$. Define $x_o = min\left(\frac{1}{2(L+\varepsilon)}, \frac{\delta}{2}\right)$ thus $x_o \leq \frac{1}{2(L+\varepsilon)}$ and $x_o < \delta$. Clearly $0 < x_o < \delta$ so it follows that*

$$-\varepsilon < \frac{1}{x_o} - L < \varepsilon$$

and as $\varepsilon = L$ we add ε to find $0 < \frac{1}{x_o} < 2\varepsilon$. On the other hand, we have constructed x_o to satisfy the inequality $x_o \leq \frac{1}{2(L+\varepsilon)} = \frac{1}{4\varepsilon}$ thus $\frac{1}{x_o} \geq 4\varepsilon$. But, this is a contradiction since we cannot have both $\frac{1}{x_o} < 2\varepsilon$ and $\frac{1}{x_o} \geq 4\varepsilon$. Therefore, be proof by contradiction, there does not exist such an $L \in \mathbb{R}$ and we conclude that the $\lim_{x \to 0^+} \frac{1}{x}$ does not exist, hence $\lim_{x \to 0} \frac{1}{x}$ does not exist. These limits **diverge**. $\square$

If you're wondering how I thought of the argument in the last example then perhaps the following picture will help you understand why I chose x_o as I did. In fact, the picture is what I used to think of the proof. Pictures are often helpful, you ought not forget that graphing can be a powerful tool for analysis.



Example 7.5.10. Problem: prove $\lim_{x \to 0^+} \frac{1}{x} = \infty$.**.**

Preparatory calculations: *we need to find δ such that $M > \frac{1}{x}$ for all $x \in \mathbb{R}$ such that $0 < x < \delta$. Note $M > \frac{1}{x}$ implies $\frac{1}{M} < x$. Looks like $\delta = \frac{1}{2M}$ will do nicely.*

Proof: *suppose $M > 0$ and let $\delta = \frac{1}{2M}$. If $0 < x < \delta = \frac{1}{2M}$ then $\frac{1}{x} > 2M > M$. Therefore, for each $M > 0$ there exists $\delta > 0$ such that $\frac{1}{x} > M$ whenever $0 < x < \delta$. It follows by definition that $\lim_{x \to 0^+} \frac{1}{x} = \infty$.* $\square$

We learned in Example 7.5.9 this limit does not exist in $\mathbb{R}$. Now we have shown that it actually diverges to ∞. Notice that $\infty \notin \mathbb{R}$, rather, ∞ is simply a notation to indicate a function has a particular behavior at a point.

Remark 7.5.11. *Another concept of infinity is discussed in the study of* **cardnality**. *Intuitively speaking the cardnality of a set describes the size of the set. For example, $S = \{1, 2, 3\}$ has cardnality 3. The natural numbers have cardnality $\aleph_o$ which is infinite. Then the real numbers are even larger, the cardnality of $\mathbb{R}$ is called the continuum c. Some authors denote the continuum by $c = \aleph_1$ and it does make sense to say that $\aleph_o < c$. However, the idea that the continuum is the next infinity past $\aleph_o$ is called the* **continuum hypothesis**.

7.6 limit laws

We assume $a \in \mathbb{R}$ and f, g are functions with limit point a throughout this section unless otherwise explicitly stated. Let us begin by proving a limit has a single value.

Proposition 7.6.1. *limit is unique.*

If $\lim\limits_{x \to a} f(x) = L_1$ and $\lim\limits_{x \to a} f(x) = L_1$ then $L_1 = L_2$.

Proof: let $\varepsilon > 0$. Suppose $\lim\limits_{x \to a} f(x) = L_1$ and $\lim\limits_{x \to a} f(x) = L_1$. Choose $\delta_1 > 0$ for which $0 < |x - a| < \delta_1$ implies $|f(x) - L_1| < \varepsilon/2$. Likewise, choose $\delta_2 > 0$ for which $0 < |x - a| < \delta_2$ implies $|f(x) - L_2| < \varepsilon/2$. Let $\delta = min(\delta_1, \delta_2)$ and suppose $0 < |x - a| < \delta \leq \delta_1, \delta_2$ hence

$$\begin{aligned} |L_1 - L_2| &= |L_1 - f(x) + f(x) - L_2| \\ &\leq |L_1 - f(x)| + |f(x) - L_2| \\ &= |f(x) - L_1| + |f(x) - L_2| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \tag{7.1}$$

Thus $|L_1 - L_2| < \varepsilon$ for arbitary $\varepsilon > 0$ and that implies $|L_1 - L_2| = 0$ hence $L_1 = L_2|$. $\square$

Very similar arguments can be used to show right and left limits which exist in $\mathbb{R}$ are unique. It is very amusing that the proof of Proposition 7.6.4 rests on nearly the same calculation as the uniqueness result above.

Proposition 7.6.2. *limit of identity function.*

$$\lim_{x \to a} x = a.$$

Proof: Fix $a \in \mathbb{R}$. Let $f(x) = x$ for all $x \in \mathbb{R}$. Let $\varepsilon > 0$ and choose $\delta = \varepsilon$. If $0 < |x - a| < \delta$ then $|f(x) - a| = |x - a| < \varepsilon$ thus $\lim_{x \to a} f(x) = a$ which is to say $\lim_{x \to a} x = a$. $\square$

Proposition 7.6.3. *limit of constant function.*

$$\lim_{x \to a} c = c.$$

Proof: Fix $a \in \mathbb{R}$ and define $f(x) = c$ for all $x \in \mathbb{R}$. Suppose $\varepsilon > 0$ and choose $\delta = 42$. If $x \in \mathbb{R}$ with $0 < |x - a| < 42$ then $|f(x) - c| = |c - c| = 0 < \varepsilon$ thus $\lim_{x \to a} f(x) = c$ by the definition of the limit. Thus $\lim_{x \to a} c = c$. $\square$

Proposition 7.6.4. *additivity of the limit.*

Suppose $\lim_{x \to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \to a} g(x) = L_g \in \mathbb{R}$ then

$$\lim_{x \to a}[f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x).$$

Proof: we are given that $\lim\limits_{x \to a} f(x) = L_f$ and $\lim\limits_{x \to a} g(x) = L_g$. Let $\varepsilon > 0$ and choose $\delta_f > 0$ such that $0 < |x - a| < \delta_f$ implies $|f(x) - L_f| < \frac{\varepsilon}{2}$. Likewise, choose $\delta_g > 0$ for which $0 < |x - a| < \delta_g$ implies $|g(x) - L_g| < \frac{\varepsilon}{2}$. Let $\delta = min(\delta_f, \delta_g)$ then $\delta \leq \delta_f$ and $\delta \leq \delta_g$. Suppose $x \in \mathbb{R}$ and $0 < |x - a| < \delta$ then $|f(x) - L_f| < \frac{\varepsilon}{2}$ and $|g(x) - L_g| < \frac{\varepsilon}{2}$. Consider that

$$\begin{aligned} |f(x) + g(x) - (L_f + L_g)| &= |f(x) - L_f + g(x) - L_g| \\ &\leq |f(x) - L_f| + |g(x) - L_g| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \tag{7.2}$$

Therefore, by the definition of the limit, $\lim\limits_{x \to a}[f(x) + g(x)] = \lim\limits_{x \to a} f(x) + \lim\limits_{x \to a} g(x)$. $\square$.

Proposition 7.6.5. *homogeneity of the limit.*

Suppose $c \in \mathbb{R}$ and $\lim_{x \to a} f(x) = L \in \mathbb{R}$ then $\lim\limits_{x \to a} cf(x) = c \lim\limits_{x \to a} f(x)$.

Proof: Suppose $c \in \mathbb{R}$ and $\lim_{x \to a} f(x) = L \in \mathbb{R}$. Let $\varepsilon > 0$. If $c \neq 0$ then choose $\delta > 0$ for which $0 < |x - a| < \delta$ implies $|f(x) - L| < \frac{\varepsilon}{|c|}$. Observe

$$|cf(x) - cL| = |c||f(x) - L| < |c|\frac{\varepsilon}{|c|} = \varepsilon \tag{7.3}$$

If $c = 0$ then $|cf(x) - cL| = 0 < \varepsilon$ for all $x \in dom(f)$. Thus, by the definition of the limit, $\lim\limits_{x \to a} cf(x) = c \lim\limits_{x \to a} f(x)$. $\square$

I often collectively refer to the previous two theorems as the *linearity* of the limit. In calculus we will learn that most major constructions obey the linearity rules. We can also extend the rules to give the limit law for a finite linear combination of convergent functions.

Proposition 7.6.6. *limit of linear combination of convergent functions.*

Suppose $a \in \mathbb{R}$ and $f_i(x) \to L_i \in \mathbb{R}$ as $x \to a$ for $i = 1, 2, \dots, n$. Then,

$$\lim_{x\to a} (c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x)) = c_1 \lim_{x\to a} f_1(x) + c_2 \lim_{x\to a} f_2(x) + \cdots + c_n \lim_{x\to a} f_n(x).$$

Proof: Suppose $f_i(x) \to L_i \in \mathbb{R}$ as $x \to a$ for $i = 1, 2, \dots, n$. We claim

$$\lim_{x\to a} (c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x)) = c_1 \lim_{x\to a} f_1(x) + c_2 \lim_{x\to a} f_2(x) + \cdots + c_n \lim_{x\to a} f_n(x)$$

for all $n \in \mathbb{N}$. We will prove this claim by induction on n. Notice the claim is true for $n = 1$ since Proposition 7.6.5 provides that $\lim_{x\to a}(c_1 f_1(x)) = c_1 \lim_{x\to a} f_1(x)$. Inductively suppose the claim is true for some $n \in \mathbb{N}$. Consider the linear combination of $n + 1$ functions,

$$\begin{aligned}
\lim_{x\to a}\big(c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) + c_{n+1} f_{n+1}(x)\big) = \\
&= \lim_{x\to a}\big(c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x)\big) + \lim_{x\to a}\big(c_{n+1} f_{n+1}(x)\big) && (7.4)\\
&= c_1 \lim_{x\to a} f_1(x) + c_2 \lim_{x\to a} f_2(x) + \cdots + c_n \lim_{x\to a} f_n(x) + c_{n+1} \lim_{x\to a} f_{n+1}(x) && (7.5)
\end{aligned}$$

We used Proposition 7.6.4 for Equation 7.4 and we applied the induction hypothesis and Proposition 7.6.5 for Equation 7.5. Thus we have shown the claim holds for $n + 1$ and it follows the result is true for all $n \in \mathbb{N}$ by induction on n. $\square$

Proposition 7.6.7. *limit of product is product of limits.*

If $\lim_{x\to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x\to a} g(x) = L_g \in \mathbb{R}$ then

$$\lim_{x\to a}[f(x)g(x)] = \left(\lim_{x\to a} f(x)\right)\left(\lim_{x\to a} g(x)\right).$$

Preparing for Proof: Consider that we wish to find $\delta > 0$ that forces $x \in B_\delta(a)_o$ to satisfy

$$|f(x)g(x) - L_f L_g| < \varepsilon \tag{7.6}$$

we have control over $|f(x) - L_f|$ and $|g(x) - L_g|$. If we can somehow factor these out then we have something to work with. Add and subtract $L_f g(x)$ towards that goal:

$$\begin{aligned}
|f(x)g(x) - L_f L_g| &= |f(x)g(x) - L_f g(x) + L_f g(x) - L_f L_g| && (7.7)\\
&\leq |f(x) - L_f||g(x)| + |L_f||g(x) - L_g|
\end{aligned}$$

Proof: let $\varepsilon > 0$ and suppose $f(x) \to L_f$ and $g(x) \to L_g$ as $x \to a$. Observe we may select positive constants δ_1, δ_2 and δ_3 for which:

(i.) $0 < |x - a| < \delta_1$ implies $|f(x) - L_f| < \dfrac{\varepsilon}{2(1 + |L_g|)}$,

(ii.) $0 < |x-a| < \delta_2$ implies $|g(x) - L_g| < \dfrac{\varepsilon}{2(1+|L_f|)}$,

(iii.) $0 < |x-a| < \delta_3$ implies $|g(x) - L_g| < 1$.

Observe, from (iii.) we also have the bound below:

$$|g(x)| = |g(x) - L_g + L_g| \leq |g(x) - L_g| + |L_g| < 1 + |L_g| \tag{7.8}$$

Let $\delta = \min(\delta_1, \delta_2, \delta_3)$ and suppose $0 < |x-a| < \delta$ thus (i.), (ii.) and (iii.) hold true and $|g(x)| < 1 + |L_g|$. Thus calculate:

$$\begin{aligned} |f(x)g(x) - L_f L_g| &= |f(x)g(x) - L_f g(x) + L_f g(x) - L_f L_g| \\ &\leq |f(x) - L_f||g(x)| + |L_f||g(x) - L_g| \\ &\leq \frac{\varepsilon}{2(1+|L_g|)}(1+|L_g|) + |L_f|\frac{\varepsilon}{2(1+|L_f|)} \\ &< \varepsilon \end{aligned} \tag{7.9}$$

where the last inequality stems from the observation that $|L_f|/(1+|L_f|) < 1$. Therefore, we have shown $f(x)g(x) \to L_f L_g$ as $x \to a$ and this completes the proof. $\square$

The proof given above is fairly standard. I found the argument in this Wikibook.

Proposition 7.6.8. *power function limit (for powers $n \in \mathbb{N}$).*

> Let $a \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$, $\lim_{x \to a} x^n = a^n$.

Proof: is by induction on n. Observe $n = 0$ is true by Proposition 7.6.3. Inductively suppose $\lim\limits_{x \to a} x^n = a^n$ for some $n \in \mathbb{N}$. Consider the $(n+1)$ case,

$$\lim_{x \to a} x^{n+1} = \lim_{x \to a} x^n x = \left(\lim_{x \to a} x^n\right)\left(\lim_{x \to a} x\right) = a^n a = a^{n+1}$$

where I used the Proposition 7.6.7 based on the induction hypothesis and Proposition 7.6.2. We find the statement true for n implies it is likewise true for $n+1$ hence the theorem is true for all $n \in \mathbb{N}$ by proof by mathematical induction. $\square$

Proposition 7.6.9. *polynomial function limit.*

> Suppose $c_n, \dots, c_1, c_0 \in \mathbb{R}$ and $p(x) = c_n x^n + \cdots + c_1 x + c_0$ then $\lim\limits_{x \to a} p(x) = p(a)$.

Proof: by Proposition 7.6.8 we note $f_i(x) = x^i$ has $\lim\limits_{x \to a} f_i(x) = a^i$ for $i = 0, 1, 2, \dots, n$. Changing numbering slightly on Proposition 7.6.6 with $f_i(x) = x^i$ for $i = 0, 1, \dots, n$ we obtain:

$$\lim_{x \to a} (c_n x^n + \cdots + c_1 x + c_0) = c_n a^n + \cdots + c_1 a + c_0 = p(a). \quad \square$$

Proposition 7.6.10. *limit of composite. Suppose f has limit point a and g has limit point L_1,*

If $\lim\limits_{x\to a} f(x) = L_1$ and $\lim\limits_{y\to L_1} g(y) = L_2$ then $\lim_{x\to a} g(f(x)) = L_2$.

Proof: let $\varepsilon > 0$. Since $\lim_{y\to L_1} g(y) = L_2$ we may choose $\delta_g > 0$ such that $0 < |y-L_1| < \delta_g$ implies $|g(y)-L_2| < \varepsilon$. Likewise, since $\lim\limits_{x\to a} f(x) = L_1$ we may select $\delta_f > 0$ for which $0 < |x-a| < \delta_f$ implies $|f(x) - L_1| < \delta_g$. Suppose $0 < |x-a| < \delta_f$ and let $y = f(x)$ then $|y - L_1| = |f(x) - L_1| < \delta_g$ hence $|g(y) - L_2| < \varepsilon$. Thus $|g(f(x)) - L_2| < \varepsilon$. Therefore, by definition of limit, $\lim_{x\to a} g(f(x)) = L_2$. $\square$

This proposition can be written without use of L_1 and L_2 but the statement is a bit clunky:

$$\lim_{x\to a} [g(f(x))] \;=\; \lim_{y\to \lim_{x\to a} f(x)} [g(y)]\,. \tag{7.10}$$

Notice the proof and application of the composite limit rule both rest on the substitution $y = f(x)$. When we make the subsitution of $y = f(x)$ we have to swap $f(x)$ for y as we trade $g(f(x))$ for $g(y)$. Likewise, we exchange $x \to a$ for the corresponding limit in y of $y \to \lim_{x\to a} f(x)$.

Proposition 7.6.11. *reciprocal function limit.*

If $a \neq 0$ then $\lim_{x\to a} \frac{1}{x} = \frac{1}{a}$.

Proof: Suppose $a > 0$. Let $\varepsilon > 0$ and choose $\delta = min\left(\frac{a}{2}, \frac{a^2\varepsilon}{2}\right) > 0$. Observe $\delta \leq a/2$ then $|x-a| < \delta \leq \frac{a}{2}$ implies $-\frac{a}{2} < x - a$ hence $\frac{a}{2} < x = |x|$. Therefore, $\frac{1}{|x|} < \frac{2}{a}$. Consequently, if $x \in \mathbb{R}$ with $0 < |x-a| < \delta$ we find:

$$\left|\frac{1}{x} - \frac{1}{a}\right| = \left|\frac{a-x}{ax}\right| = \frac{|x-a|}{a|x|} < \frac{2|x-a|}{a^2} < \frac{2}{a^2}\delta \leq \frac{2}{a^2}\frac{a^2\varepsilon}{2} = \varepsilon. \tag{7.11}$$

Thus, by the definition of the limit, $\lim\limits_{x\to a} \dfrac{1}{x} = \dfrac{1}{a}$. The proof in the case $a < 0$ is similar and we leave it as an exercise for the reader. $\square$

Proposition 7.6.12. *limit of quotient is quotient of limits.*

Suppose $\lim_{x\to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x\to a} g(x) = L_g \in \mathbb{R}$ with $L_g \neq 0$ then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}.$$

Proof: Let $h(y) = \frac{1}{y}$ and note Proposition 7.6.11 provides $\lim_{y\to L_g} h(y) = \frac{1}{L_g}$ since $L_g \neq 0$. Furthermore, by Proposition 7.6.10 we find the limit of the composite function $h(g(x)) = \frac{1}{g(x)}$ is given by $\lim_{y\to L_g} h(y) = \frac{1}{L_g}$. Proposition 7.6.7 completes the proof since:

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a}\left[f(x)\cdot\frac{1}{g(x)}\right] = \left(\lim_{x\to a} f(x)\right)\left(\lim_{x\to a}\frac{1}{g(x)}\right) = L_f \cdot \frac{1}{L_g} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}. \;\square$$

Beyond these rules you will find a number of other "limit laws" in various texts. In one way or another they boil down to proving a particular function has a natural limit then you combine that data together with the composite limit law. So, to complete our catalog of basic limit math we ought to calculate limits of the elementary functions.

Proposition 7.6.13. *limit of square root function.*

If $a > 0$ then $\lim_{x\to a} \sqrt{x} = \sqrt{a}$. In addition, $\lim_{x\to 0^+} \sqrt{x} = 0$.

Proof: Notice the following algebraic identity for $a > 0$,

$$\left|\sqrt{x} - \sqrt{a}\right| = \frac{|\sqrt{x} - \sqrt{a}||\sqrt{x} + \sqrt{a}|}{|\sqrt{x} + \sqrt{a}|} = \frac{|(\sqrt{x} - \sqrt{a})(\sqrt{x} + \sqrt{a})|}{|\sqrt{x} + \sqrt{a}|} = \frac{|x-a|}{|\sqrt{x} + \sqrt{a}|} \tag{7.12}$$

Let $\varepsilon > 0$ and choose $\delta = \varepsilon\sqrt{a} > 0$. If $0 < |x-a| < \delta = \varepsilon\sqrt{a}$ then following Equation 7.12 we find

$$\left|\sqrt{x} - \sqrt{a}\right| = \frac{|x-a|}{|\sqrt{x} + \sqrt{a}|} < \frac{|x-a|}{\sqrt{a}} < \frac{\varepsilon\sqrt{a}}{\sqrt{a}} = \varepsilon.$$

Therefore, by the definition of the limit we find $\lim_{x\to a} \sqrt{x} = \sqrt{a}$. We leave the proof of $\lim_{x\to 0^+} \sqrt{x} = 0$ to the reader $\square$

Notice the proof for $\lim_{x\to 0^+} \sqrt{x} = 0$ will be similar to that given in Exercise 7.5.8. We could continue on to show $\lim_{x\to a} \sqrt[n]{x} = \sqrt[n]{x}$ for $n = 3, 4, 5, \dots$, but the algebraic difficulty is nontrivial and we will soon find a more efficient way to calculate such limits. As such we content ourselves to merely prove the limit law for the square root for the time being. That said, if we had a proof that $\lim_{x\to a} \sqrt[n]{x} = \sqrt[n]{x}$ for $n = 3, 4, 5, \dots$ then we could use Proposition 7.6.8 and Proposition 7.6.10 for $x > 0$ since $x^{m/n} = (\sqrt[n]{x})^m$.

Proposition 7.6.14. *rational power limit (power function with power m/n for $m, n \in \mathbb{N}$).*

Let $a \in \mathbb{R}$ with $a > 0$ and $n \in \mathbb{N}$, $\lim_{x\to a} x^{\frac{m}{n}} = a^{\frac{m}{n}}$.

Notice that fractional powers are problematic for negative numbers. If you agree that $\frac{1}{3} = \frac{1}{2}\frac{2}{3}$ then you should ask yourself what domain would you assign $f(x) = x^{\frac{1}{3}}$? What about $g(x) = x^{\frac{1}{2}\frac{2}{3}}$? What about $h(x) = (\sqrt{x})^{\frac{2}{3}}$? I would argue that $dom(f) = \mathbb{R}$ whereas $dom(h) = [0, \infty)$. But, the only difference between these formulas is that I applied the exponent law $a^{st} = (a^s)^t$. My point? Laws of exponents presuppose a positive base a. In fact, h and f are different functions because the "law" I used was incorrect for the base considered. Another good example of laws of exponents breaking down is the following:

$$-1 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{1} = 1$$

oops. Exponential functions for negative bases are meaningful from the viewpoint of complex variables, however it comes at the cost of losing the function property. For example, $(-1)^{\frac{1}{2}} = \{i, -i\}$

where i is the imaginary unit classically denoted $\sqrt{-1} = i$. Enough about that, I'm just trying to make you aware of some boundaries in our thinking about exponents. Hopefully you'll get a chance to take a good course in complex analysis to unwrap the mysteries of complex arithmetic.

I suppose I should mention that we can calculate the $(2k+1)$-th root of any real number. In other words, $f(x) = \sqrt[2k+1]{x}$ has $dom(f) = \mathbb{R}$ for any $k \in \mathbb{N}$. Cube roots of negative numbers are defined to be real in this course without trouble. For example, $\sqrt[3]{-27} = -3$ since $(-3)^3 = -27$. It turns out the limit works just fine for such functions in the negative case. For $a < 0$ we could prove:

$$\lim_{x \to a} \sqrt[2k+1]{x} = \sqrt[2k+1]{a}. \tag{7.13}$$

From this point we could go on and prove dozens of propositions about limits of your favorite algebraic functions. Let me summarize:

Theorem 7.6.15. *algebraic functions have nice limits.*

Let $f(x)$ be defined by a finite number of algebraic operations (possibly including addition, multiplication, division, taking integer or fractional roots) then $\lim_{x \to a} f(x) = f(a)$.

I think we've seen enough detail in this direction so we now turn to limits of sine and cosine.

Proposition 7.6.16. *limits of sine and cosine.*

Let $a \in \mathbb{R}$, $\lim_{x \to a} \sin(x) = \sin(a)$ and $\lim_{x \to a} \cos(x) = \cos(a)$.

Proof: We begin by proving limits for sine and cosine at $a = 0$. We note it can be shown geometrically that for $0 < x \leq \pi/2$ we have $|\sin x| \leq |x|$ and $|\cos x - 1| \leq |x|$.

Let $\varepsilon > 0$ and choose $\delta = min(\varepsilon, \pi/4)$. Suppose that $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. Since $\delta \leq \pi/4 < \pi/2$ we have $|\sin x - 0| = |\sin x| < |x| < \delta \leq \varepsilon$. Thus $\lim_{x \to 0} \sin x = 0 = \sin 0$.

Let $\varepsilon > 0$ and choose $\delta = min(\varepsilon, \pi/4)$. Suppose that $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. Since $\delta \leq \pi/4 < \pi/2$ we have $|\cos x - 1| < |x| < \delta \leq \varepsilon$. Thus $\lim_{x \to 0} \cos x = 1 = \cos 0$.

Lemma 7.6.17. *substitution of limiting variable.*

If either one of the following limits exist then so does the other and

$$\lim_{x \to a} f(x) = \lim_{h \to 0} f(a + h).$$

Proof of lemma: Suppose $\lim_{x \to a} f(x) = L \in \mathbb{R}$. Let $\varepsilon > 0$ choose $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - L| < \varepsilon$. Let $h = x - a$ then $x = a + h$ and $0 < |h| < \delta$ implies $|f(a + h) - L| < \epsilon$. Thus $\lim_{h \to 0} f(a + h) = L$ by the definition of the limit.

Next, suppose $\lim_{h\to 0} f(a+h) = L \in \mathbb{R}$. Let $\varepsilon > 0$ and choose $\delta > 0$ such that $0 < |h| < \delta$ implies $|f(a+h) - L| < \varepsilon$. Let $x = a + h$ then $h = x - a$. If $0 < |x-a| = |h| < \delta$ then $|f(a+h) - L| = |f(x) - L| < \varepsilon$. Therefore, $\lim\limits_{x\to a} f(x) = L$. [6] $\triangledown$

Apply the Lemma and recall the trigonometric formula for adding angles in cosine,

$$\begin{aligned}
\lim_{x\to a} \cos(x) &= \lim_{h\to 0} \cos(a+h) \\
&= \lim_{h\to 0} \cos(a)\cos(h) - \sin(a)\sin(h) \\
&= \cos(a) \lim_{h\to 0} \cos(h) - \sin(a) \lim_{h\to 0} \sin(h) \\
&= \cos(a).
\end{aligned}$$

I leave the proof that $\lim_{x\to a} \sin(x) = \sin(a)$ as an exercise for the reader. This completes the proof of proposition 7.6.16. $\square$

It should be fairly clear that we can calculate limits of tangent, cotangent, cosecant and secant since we have a quotient rule for limits and we know the limits for sine and cosine. I leave working out such limits as an elementary exercise for the reader.

Proposition 7.6.18. *limit of exponential function.*

Let $b > 0$, $\lim_{x\to a} b^x = b^a$.

Proof: We begin by proving $\lim_{x\to 0} 2^x = 1$. Let $\varepsilon > 0$ and choose $\delta = \log_2(1+\varepsilon)$. Note that $1 + \varepsilon > 1$ hence $\log_2(1+\varepsilon) > 0$. Suppose that $x \in \mathbb{R}$ such that $0 < |x| < \delta$, it follows that

$$-\log_2(1+\varepsilon) = \log_2\left(\frac{1}{1+\varepsilon}\right) < x < \log_2(1+\varepsilon)$$

but surely[7] $x < y$ implies $2^x < 2^y$ thus

$$\frac{1}{1+\varepsilon} < 2^x < 1 + \varepsilon$$

subtracting one from each inequality yields,

$$\frac{1}{1+\varepsilon} - 1 < 2^x - 1 < \varepsilon$$

Note that $\frac{1}{1+\varepsilon} - 1 = -\frac{\varepsilon}{1+\varepsilon} > -\varepsilon$ thus $-\varepsilon < 2^x - 1 < \varepsilon$ which is equivalent to $|2^x - 1| < \varepsilon$. Hence, $0 < |x| < \delta$ implies $|2^x - 1| < \varepsilon$. Therefore, $\lim_{x\to 0} 2^x = 1$.

[6] the upside-down triangle indicates the proof of the lemma is complete however the proposition's proof is still unfinished. I should mention, this Lemma equally well applies for other limiting behavior, if either one has limit $\pm\infty$ then do does the other and if either limit fails to exist then so goes the other. Please forgive me for omitting the proof of those assertions.

[7] admitably there is a gap here, I invite the reader to supply a proof

To cover other bases than 2 we can use the identity $b^x = 2^{\log_2(b^x)} = 2^{\log_2(b)x}$ for any $b > 0$. Since $\log_b(2)$ is a constant we can deduce that $\log_2(b)x \to 0$ as $x \to 0$. Moreover, using the composition of limits proposition we find that $b^x = 2^{\log_2(b)x} \to 1$ as $x \to 0$. Thus, $\lim_{x\to 0} b^x = 1$.

The laws of exponents complete the proof for limit points other than zero:

$$\lim_{h\to 0}(b^{a+h}) = \lim_{h\to 0}(b^a b^h) = b^a \lim_{h\to 0} b^h = b^a.$$

Then by Lemma 7.6.17, $\lim_{x\to a}(b^x) = b^a$. □

The use of 2^x was simply a choice on my part. We could just as well have used the identity $x^p = 3^{x\log_3(p)}$ to drive the proof. We defer limits of inverse functions such as logarithms and inverse trigonometric functions as they require some sophisticated ideas we have not yet explored.

7.7 squeeze theorem

There are limits not easily solved through algebraic trickery. Sometimes the "Squeeze" or "Sandwich" Theorem allows us to calculate the limit.

Proposition 7.7.1. *squeeze theorem*[8].

Let $f(x) \leq g(x) \leq h(x)$ for all x near a then we find that the limits at a follow the same ordering,

$$\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x) \leq \lim_{x\to a} h(x).$$

Moreover, if $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L \in \mathbb{R}$ then $\lim_{x\to a} f(x) = L$.

Proof: Suppose $f(x) \leq g(x)$ for all $x \in B_{\delta_1}(a)_o$ for some $\delta_1 > 0$ and also suppose $\lim_{x\to a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x\to a} g(x) = L_g \in \mathbb{R}$. We wish to prove that $L_f \leq L_g$. Suppose otherwise towards a contradiction. That is, suppose $L_f > L_g$. Note that $\lim_{x\to a}[g(x) - f(x)] = L_g - L_f$ by the linearity of the limit. It follows that for $\varepsilon = \frac{1}{2}(L_f - L_g) > 0$ there exists $\delta_2 > 0$ such that $x \in B_{\delta_2}(a)_o$ implies $|g(x) - f(x) - (L_g - L_f)| < \varepsilon = \frac{1}{2}(L_f - L_g)$. Expanding this inequality we have

$$-\frac{1}{2}(L_f - L_g) < g(x) - f(x) - (L_g - L_f) < \frac{1}{2}(L_f - L_g)$$

adding $L_g - L_f$ yields,

$$-\frac{3}{2}(L_f - L_g) < g(x) - f(x) < -\frac{1}{2}(L_f - L_g) < 0.$$

Thus, $f(x) > g(x)$ for all $x \in B_{\delta_2}(a)_o$. But, $f(x) \leq g(x)$ for all $x \in B_{\delta_1}(a)_o$ so we find a contradiction for each $x \in B_\delta(a)$ where $\delta = min(\delta_1, \delta_2)$. Hence $L_f \leq L_g$. The same proof can be applied to

[8] this theorem can also be stated for $x \to \pm\infty$ or $x \to a^{\pm}$ provided the inequality holds for appropriate values where the limit is taken

g and h thus the first part of the theorem follows.

Next, we suppose that $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L \in \mathbb{R}$ and $f(x) \leq g(x) \leq h(x)$ for all $x \in B_{\delta_1}(a)$ for some $\delta_1 > 0$. We seek to show that $\lim_{x\to a} f(x) = L$. Let $\varepsilon > 0$ and choose $\delta_2 > 0$ such that $|f(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$ for all $x \in B_\delta(a)_o$. We are free to choose such a $\delta_2 > 0$ because the limits of f and h are given at $x = a$. Choose $\delta = min(\delta_1, \delta_2)$ and note that if $x \in B_\delta(a)_o$ then $f(x) \leq g(x) \leq h(x)$ hence,

$$f(x) - L \leq g(x) - L \leq h(x) - L$$

but $|f(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$ imply $-\varepsilon < f(x) - L$ and $h(x) - L < \varepsilon$ thus

$$-\varepsilon < f(x) - L \leq g(x) - L \leq h(x) - L < \varepsilon.$$

Therefore, for each $\varepsilon > 0$ there exists $\delta > 0$ such that $x \in B_\delta(a)_o$ implies $|g(x) - L| < \varepsilon$ so $\lim_{x\to a} g(x) = L$. $\square$

7.8 continuity theorems

Theorem 7.8.1. *Each function below is continuous on its domain:*

1.) *Polynomial functions,*

2.) *Rational functions,*

3.) *Algebraic functions,*

4.) *Trigonmetric functions and their reciprocal functions.*

5.) *Exponential functions,*

6.) *Hyperbolic trigonmetric functions and their reciprocal functions.*

Proof: We have already done most of the work in Section 7.6. Proposition 7.6.9gives continuity of polynomials; if $p(x)$ is a polynomial then $\lim_{x\to a} p(x) = p(a)$. Likewise, if $f(x) = \frac{p(x)}{q(x)}$ for polynomials $p(x), q(x)$ and if $q(a) \neq 0$ then we note by the limit law for quotients:

$$\lim_{x\to a} \frac{p(x)}{q(x)} = \frac{\lim_{x\to a} p(x)}{\lim_{x\to a} q(x)} = \frac{p(a)}{q(a)} \tag{7.14}$$

thus rational functions are continuous. Arguments supporting the continuity of algebraic functions are given in the discussion summarized by Theorem 7.6.15. Continuity of sine and cosine is given by Proposition 7.6.16 and continuity of tangent was shown in Example **??**. Continuity of $\cot x = \frac{1}{\tan x}$, $\csc x = \frac{1}{\sin x}$ and $\sec x = \frac{1}{\cos x}$ follow from the quotient rule for limits and the known limits for tangent, sine and cosine. Continuity of exponential functions is given by Proposition 7.6.18. Finally, recall hyperbolic functions are defined in terms of exponential functions;

$$\cosh x = \frac{1}{2}\left(e^x + e^{-x}\right) \qquad \& \qquad \sinh x = \frac{1}{2}\left(e^x - e^{-x}\right). \tag{7.15}$$

Furthermore, $\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. Proposition 7.6.18 gives $\lim_{x\to a} e^x = e^a$. Note $e^{-x} = \frac{1}{e^x}$ thus by the quotient rule for limits $\lim_{x\to a} e^{-x} = \frac{1}{\lim_{x\to a} e^x} = \frac{1}{e^a} = e^{-a}$. Use Proposition 7.6.6 to calculate

$$\lim_{x\to a} \cosh x = \lim_{x\to a} \frac{1}{2}\left(e^x + e^{-x}\right) = \frac{1}{2}\left(e^a + e^{-a}\right) = \cosh a, \tag{7.16}$$

$$\lim_{x\to a} \sinh x = \lim_{x\to a} \frac{1}{2}\left(e^x - e^{-x}\right) = \frac{1}{2}\left(e^a - e^{-a}\right) = \sinh a. \tag{7.17}$$

Continuity of $\tanh x = \frac{\sinh x}{\cosh x}$ and $\text{sech}x = \frac{1}{\cosh x}$ on $\mathbb{R}$ follow from the result above and the observation $\cosh x \neq 0$ for all $x \in \mathbb{R}$. Note $\sinh x = 0$ if and only if $x = 0$ thus the limits above for cosh and sinh serve to establish continuity of $\coth x = \frac{\cosh x}{\sinh x}$ and $\text{csch}x = \frac{1}{\sinh x}$ on $\mathbb{R} - \{0\}$. $\square$

7.9 proof of intermediate value theorem

Proposition 7.9.1.

Let f be continuous at c such that $f(c) \neq 0$ then there exists $\delta > 0$ such that either $f(x) > 0$ or $f(x) < 0$ for all $x \in (c-\delta, c+\delta)$.

Proof: we are given that $\lim_{x\to c} f(x) = f(a) \neq 0$.

1.) Assume that $f(a) > 0$. Choose $\varepsilon = \frac{f(a)}{2}$ and use existence of the limit $\lim_{x\to c} f(x) = f(a)$ to select $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|f(x) - f(a)| < \frac{f(a)}{2}$ hence $-\frac{f(a)}{2} < f(x) - f(a) < \frac{f(a)}{2}$. Adding $f(a)$ across the inequality yields $0 < \frac{f(a)}{2} < f(x) < \frac{3f(a)}{2}$.

2.) If $f(a) < 0$ then we can choose $\varepsilon = -\frac{f(a)}{2} > 0$ and select $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|f(x) - f(a)| < -\frac{f(a)}{2}$ hence $\frac{f(a)}{2} < f(x) - f(a) < -\frac{f(a)}{2}$. It follows that $\frac{3f(a)}{2} < f(x) < \frac{f(a)}{2} < 0$.

The proposition follows. $\square$

Bolzano understood there was a gap in the arguments of the founders of calculus. Often, theorems like those stated in this section would merely be claimed without proof. The work of Bolzano and others like him ultimately gave rise to the careful rigorous study of the real numbers and more generally the study of *real analysis* [9]

Proposition 7.9.1 is clearly extended to sets which have boundary points. If we know a function is continuous on $[a, b)$ and $f(a) \neq 0$ then we can find $\delta > 0$ such that $f([a, a+\delta)) > 0$. (*This is needed in the proof below in the special case that $c = a$ and a similar comment applies to $c = b$.*)

Theorem 7.9.2. *Bolzano's theorem*

[9]the Bolzano-Weierstrauss theorem is one of the central theorems of real analysis, in 1817 Bolzano used it to prove the IVT. It states every bounded sequence contains a convergent subsequence. Sequences can also be used to formulate limits and continuity. Sequential convergence is dealt with properly in undergraduate real analysis.

Let f be continuous on $[a.b]$ such that $f(a)f(b) < 0$ then there exists $c \in (a, b)$ such that $f(c) = 0$.

Proof: suppose $f(a) < f(b)$ then $f(a)f(b) < 0$ implies $f(a) < 0$ and $f(b) > 0$. We can use axiom A11 for the heart of this proof. Our goal is to find a nonempty subset $S \subseteq \mathbb{R}$ which has an upper bound. Axiom A11 will then provides the existence of the least upper bound. We should like to construct a set which has the property desired in this theorem. Define $S = \{x \in [a, b] \mid f(x) < 0\}$. Notice that $a \in S$ since $f(a) < 0$ thus $S \neq \emptyset$. Moreover, it is clear that $x \leq b$ for all $x \in S$ thus S is bounded above. Axiom A11 states that there exists a least upper bound $c \in S$. To say c is the least upper bound means that any other upperbound of S is larger than c.

We now seek to show that $f(c) = 0$. Consider that there exist three possibilities:

1. if $f(c) < 0$ then the continuous function f has $f(c) \neq 0$ so by prop. 7.9.1 there exists $\delta > 0$ such that $x \in (c - \delta, c + \delta) \cap [a, b]$ implies $f(x) < 0$. However, this implies there is a value $x \in [c, c + \delta)$ such that $f(x) < 0$ and $x > c$ which means x is in S and is larger than the upper bound c. Therefore, c is not an upper bound of S. Obviously this is a contradiction therefore $f(c) \not< 0$.

2. if $f(c) > 0$ then the continuous function f has $f(c) \neq 0$ so by prop. 7.9.1 there exists $\delta > 0$ such that $x \in (c - \delta, c + \delta) \cap [a, b]$ implies $f(x) > 0$. However, this implies that all values $x \in (c - \delta, c]$ have $f(x) > 0$ and thus $x \notin S$ which means $x = c - \delta/2 < c$ is an upper bound of S which is smaller than the least upper bound c. Therefore, c is not the least upper bound of S. Obviously this is a contradiction therefore $f(c) \not> 0$.

3. if $f(c) = 0$ then no contradiction is found. The theorem follows. $\square$

My proof here essentially follows the argument in CALCULUS, Volume 1 by Apostol[10] However I suspect this argument in one form or another can be found in many serious calculus texts. With Bolzano's theorem settled we can prove the IVT without much difficulty.

(IVT): Suppose that f is continuous on an interval $[a, b]$ with $f(a) \neq f(b)$ and let N be a number such that N is between $f(a)$ and $f(b)$ then there exists $c \in (a, b)$ such that $f(c) = N$.

Proof: let N be as described above and define $g(x) = f(x) - N$. Note that g is clearly continuous. Suppose that $f(a) < f(b)$ then we must have $f(a) < N < f(b)$ which gives $f(a) - N \leq 0 \leq f(b) - N$ hence $g(a) < 0 < g(b)$. Applying Bolzano's theorem to g gives $c \in (a, b)$ such that $g(c) = 0$. But, $g(c) = f(c) - N = 0$ therefore $f(c) = N$. If $f(a) > f(b)$ then a similar argument applies. $\square$.

[10] CALCULUS Volumes 1 and 2 are a worthy resource for any math major, I highly recomend reading them as a follow-up to calculus. Those volumes capture a time when we were much more serious about math at the undergraduate level. Much of the rest of the world still uses Apostol for the text in the university calculus course. International editions of the text are inexpensive and a pdf is freely available online.

7.10 inverse function theorem

It would be wise to read Section 7.4.1 to review what we can say about inverse functions with algebraic techniques. In this section we bring additional arguments anchored in the theory of continuous functions. In particular, the Intermediate Value Theorem is central to the proof of the inverse function theorem for continuous functions.

Proposition 7.10.1. *continuous injections are strictly monotonic.*

> If $S \subseteq \mathbb{R}$ is connected and $f : S \to T$ is continuous then f is one-to-one if and only if f is either strictly increasing or strictly decreasing on S.

Proof: Assume S is a connected subset of $\mathbb{R}$ and $f : S \to T$ is a continuous function. I assume S is an interval as introduced in Definition 1.3.2. Connected subsets of $\mathbb{R}$ are either intervals or sets containing a single point $\{p\}$. I will assume $S \neq \{p\}$ in the interest of being interesting.

($\Rightarrow$) If f is either strictly increasing or decreasing then Proposition 7.4.2 proves f is one-to-one.

($\Leftarrow$) Conversely, suppose $f : S \to T$ is one-to-one. We seek to show that f is either increasing or decreasing. Suppose f is strictly increasing on the connected subsets $U_j \subseteq S$ for $j = 1, 2, \dots$. Likewise, suppose f is strictly decreasing on connected subsets $V_k \subseteq S$ for $k = 1, 2, \dots$. The union of sets U_j and V_k for all j, k should yield S. Of particular interest are the points which are on the edge between U_j and V_k. Suppose that U, V are two subsets such that $U \cap V = \{z_o\}$ and U is to the left of V on the number line. I continue to use the notation U indicates strictly increasing and V strict decrease of f. We can show that f is not one-to-one if there exists such a point. We choose sets small enough such that $[w_o, z_o] \subset U$ whereas $[z_o, q_o] \subset V$. By construction $w_o < z_o$ and as f increases on U it follows that $f(w_o) < f(z_o)$. By the continuity of f the intermediate value theorem yields $[f(w_o), f(z_o)] \subseteq f[w_o, z_o]$. Likewise, by construction $z_o < q_o$ and as f decreases on V it follows $f(z_o) > f(q_o)$. Again, by the continuity of f the intermediate value theorem yields $[f(q_o), f(z_o)] \subseteq f[z_o, q_o]$. Suppose that $p \in [f(w_o), f(z_o)] \cap [f(q_o), f(z_o)]$ such that $p \neq f(z_o)$ then we have both $p < f(z_o)$ and $p > f(z_o)$ which is a contradiction. It follows that we either have disjoint intervals of increase and decrease or we have just one interval of strict increase or decrease. Our assumption that U is connected rules out the possibility of disjoint subsets whose union cover the whole set. Therefore, f is either strictly increasing or strictly decreasing. $\square$

Theorem 7.10.2. *invertible continuous function have continuous inverses.*

> Suppose $S, T \subseteq \mathbb{R}$ and S is connected. If $f : S \to T$ is continuous with inverse $f^{-1} : T \to S$ then f^{-1} is continuous.

Proof: we seek to show f^{-1} is continuous at $y_o \in T$. Let $\varepsilon > 0$ and suppose $x_o = f^{-1}(y_o)$, choose $\delta = min\big[f(x_o) - f(x_o - \varepsilon), f(x_o + \varepsilon) - f(x_o)\big]$ and suppose $0 < |y - y_o| < \delta$. Note that

$$y < y_o + \delta \leq f(x_o) + [f(x_o + \varepsilon) - f(x_o)] = f(x_o + \varepsilon)$$

Then on the other side,

$$y > y_o - \delta \geq f(x_o) - [f(x_o - \varepsilon) - f(x_o)] = f(x_o - \varepsilon)$$

Putting together the inequalities above yields $f(x_o - \varepsilon) < y < f(x_o + \varepsilon)$. Since f is continuous and invertible it follows from the previous proposition (and ultimately the IVT) that f^{-1} and f are either strictly increasing or strictly decreasing on S. Suppose f^{-1} is strictly increasing then:

$$x_o - \varepsilon < f^{-1}(y) < x_o + \varepsilon \quad \Rightarrow \quad |f^{-1}(y) - x_o| < \varepsilon \quad \Rightarrow \quad |f^{-1}(y) - f^{-1}(y_o)| < \varepsilon.$$

If f^{-1} is strictly decreasing then we again find that $0 < |y - y_o| < \delta$ implies $|f^{-1}(y) - y_o| < \varepsilon$. Therefore, $\lim_{y \to y_o} f^{-1}(y) = f^{-1}(y_o)$ for each $y_o \in T$ hence f^{-1} is continuous. $\square$