

CALCULUS

James S. Cook
Liberty University
Department of Mathematics

Fall 2010

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 do not follow the linear order of the text, if this is a problem for you then you should take calculus I with another professor. Why do I not follow text directly? Simply this, precalculus is a prerequisite for this course and you ought to already know what the sine, cosine, exponential and natural log are from the outset. As such I integrate their discussion throughout the course. 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. read the assigned reading (will have short quizzes to check you've read).
3. come to class, take notes, **think**.
4. attempt Problem Sets, you will likely find forming a study group is essential for success here. I ask some very hard questions. The majority of the hwk grade comes from Problem Sets.

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, noncommutative 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, differential forms, vector bundles, gauge theory, poisson algebras, homotopy, Atiyah-Singer index theorem, nonstandard analysis, construction of the real numbers... 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 of 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 the math is applied in the real world, working out examples of a general definition in specific contexts, develop an algebraic or geometric intuition,... Generally, is it better to give the definition then show examples which illustrate a concept, or is it better to give examples then provide the general definition at the end of the story? To summarize, basically all these suggestions boil down to spending time with 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

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
\S	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
$A \Rightarrow B$	A implies B
$A \Leftrightarrow B$	A and B are equivalent statements
\therefore	therefore
\forall	for all
\equiv	definition
\approx	approximately
eq^{n}	equation
sol^{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

Finally, 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. This is my third set of notes for calculus I. I have given up on the Microsoft Word edition because it has a file size of over 100Mb and \LaTeX takes only about a tenth the space. Also, \LaTeX automatically numbers sections, chapters and creates hyperlinks in the table of contents. If you read the pdf you ought to be able to navigate by clicking on the links in the table of contents. I plan to add an index as well once these notes mature a bit more.

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. I have been sloppier in previous offerings, I am refining the course this semester. In particular I intend to treat the so-called $\epsilon\delta$ -type arguments with much additional rigor. I also plan to include a more systematic discussion of limits and the major existence theorems in this course. What this means for you is that as you look over old tests you ought not make too much of them. To whom much is given much is required, I will be giving much more detail this semester on topics which I have previously neglected.

James Cook, August 12, 2010.

version 3.0

Calculus II notes are now added. Although, only about 100pages. This means the remaining half of the Math 132 notes remain in the 2009/2010-format. Also, be warned the section on asymptotics needs work in the calculus I notes. The condition necessary for proper understanding is more subtle than my current notes indicate. I haven't had time to fix that yet.

James Cook, January 13, 2011.

No major changes since Calculus II edit, however, removed asymptotic section for time being. Hope to add improved section later in Fall 2011 semester.

James Cook, August 18, 2011.

Contents

1	history and applications of calculus	13
1.1	geometry	13
1.2	numbers	14
1.3	algebra and physics	15
1.4	discovery of calculus	15
1.5	a selection of mathematical stories	17
1.6	a short overview of calculus in basic physics	21
2	foundations of mathematics	25
2.1	set theory	25
2.2	numbers	26
2.3	algebra	32
2.4	analytical geometry	33
2.5	functions	34
2.6	graphing and inequalities	38
2.7	local inverses	40
2.8	elementary functions	46
2.8.1	polynomial functions	46
2.8.2	power functions	46
2.8.3	rational functions	47
2.8.4	algebraic functions	47
2.8.5	trigonometric functions	48
2.8.6	reciprocal trigonometric functions	50
2.8.7	inverse trigonometric functions	51
2.8.8	exponential functions	51
2.8.9	logarithmic functions	52
2.8.10	hyperbolic functions	54
2.9	trigonometry	55
2.9.1	complex numbers	56
2.9.2	the complex exponential*	58
2.9.3	polar form of a complex number*	59
2.9.4	the algebra of sine and cosine*	61

2.9.5	superposition of waves and the method of phasors*	63
3	limits	67
3.1	graphical motivation of limit	67
3.2	definition of the limit	72
3.2.1	two-sided limit	72
3.2.2	one-sided limits	77
3.2.3	divergent limits	80
3.3	continuity and limit laws	82
3.4	limit calculation	92
3.5	squeeze theorem	98
3.6	intermediate value theorem	99
3.6.1	a deeper look at the intermediate value theorem	101
4	differential calculus	103
4.1	tangent lines	104
4.2	definition of the derivative function	108
4.3	linearity of the derivative and the power rule	110
4.3.1	derivative of a constant	111
4.3.2	derivative of identity function	111
4.3.3	derivative of quadratic function	112
4.3.4	derivative of cubic function	112
4.3.5	power rule	113
4.4	the exponential function	116
4.5	derivatives of sine and cosine	118
4.6	product rule	122
4.7	quotient rule	123
4.8	chain rule	125
4.9	higher derivatives	130
4.10	implicit differentiation and derivatives of inverse functions	134
4.11	logarithmic differentiation	137
4.11.1	proof of power rule	139
4.12	summary of basic derivatives	140
4.13	related rates	142
5	derivatives and linear approximations	145
5.1	linearizations	145
5.2	differentials and error	146
5.3	Newton's method	148

6	geometry and differential calculus	153
6.1	graphing with derivatives	153
6.1.1	first derivative test	159
6.1.2	concavity and the second derivative test	164
6.1.3	closed interval method	171
6.2	optimization	173
6.3	to $\pm\infty$ and beyond	178
6.3.1	algebraic techniques for calculating limits at $\pm\infty$	182
6.3.2	asymptotes in general	188
6.4	l'Hopital's rule	190
6.4.1	concerning why l'Hopital's rule works	193
6.4.2	indeterminant powers	194
6.5	Taylor's Theorem about polynomial approximation	197
6.5.1	constant functions	197
6.5.2	linearizations again	197
6.5.3	quadratic approximation of function	197
6.5.4	cubic approximation of function	197
6.5.5	general case	198
6.5.6	error in Taylor approximations	201
6.5.7	higher derivative tests	208
7	antiderivatives and the area problem	211
7.1	indefinite integration	212
7.1.1	why antidifferentiate?	212
7.1.2	properties of indefinite integration	215
7.1.3	examples of indefinite integration	217
7.2	area problem	218
7.2.1	sums and sequences in a nutshell	219
7.2.2	left, right and midpoint rules	222
7.2.3	Riemann sums and the definite integral	225
7.2.4	properties of the definite integral	227
7.3	fundamental theorem of calculus	230
7.3.1	area functions and FTC part I	231
7.3.2	FTC part II, the standard arguments	235
7.3.3	FTC part II an intuitive constructive proof	236
7.4	definite integration	239
7.4.1	area vs. signed-area	240
7.4.2	average of a function	241
7.4.3	net-change theorem	242
7.5	u-substitution	243
7.5.1	u-substitution in indefinite integrals	243
7.5.2	u-substitution in definite integrals	247

7.5.3	theory of u -substitution	248
8	applications of integral calculus	251
8.1	a brief tour of infinitesimal methods	252
8.2	area	258
8.3	volume	268
9	Integration Techniques	281
9.1	integration by parts (IBP)	281
9.2	integrals of trigonometric functions	287
9.3	implicit substitutions	291
9.4	partial fractions	299
9.4.1	existence of partial fractions decomposition	309
10	Numerical Integration	311
10.1	left, right and midpoint rules	311
10.2	trapezoid rule	314
10.3	Simpson's Rule	315
11	Improper Integrals	319
11.1	divergent integrands	320
11.1.1	on the integration of finite jump discontinuous functions	324
11.2	divergent bounds	326
12	Curves	335
12.1	graphs	335
12.2	level curves	336
12.3	parametrized curves	339
12.4	converting to and from the parametric viewpoint	343
12.4.1	how can we find the Cartesian form for a given parametric curve?	344
12.4.2	how can we find a parametrization of a given Cartesian curve?	345
13	Calculus of Parametrized Curves	347
13.1	vectors in a nutshell	348
13.2	calculus for vector-valued functions	350
13.3	position, velocity and acceleration in two dimensions	353
13.3.1	position vs. displacement vs. distance traveled	355
13.4	geometric analysis in parametric setting	359
13.4.1	motivating the arclength formula	364
13.5	arclength examples	365

14 Further Applications of Integral Calculus	369
14.1 surface area	370
14.2 physics	373
14.2.1 work and force with calculus	373
14.2.2 center of mass	379
15 Geometry and Coordinates	383
15.1 polar coordinates	383
15.2 parametric curves in polar coordinates	383
15.3 conic sections	383
15.4 rotated coordinates	383
15.5 other coordinate systems	383
16 Introduction to Differential Equations	385
16.1 what is a solution?	385
16.2 separation of variables	385
16.3 exponential growth and decay	385
16.4 integrating factor method	385
16.5 constant coefficient second order DEqns	385
17 Sequences and Series	387
17.1 sequence examples	387
17.2 convergence of sequences	387
17.3 series and convergence tests	387
17.4 error tests	387
17.5 conditional and absolute convergence	387
18 Basics of Power Series	389
18.1 what is a power series?	389
18.2 power series via geometric series	389
18.3 power series representations	389
19 Taylor Series and Techniques	391
19.1 Taylor Polynomials	391
19.2 Taylor's Theorem	391
19.3 binomial series	391
19.4 numerical applications of Taylor series	391
19.5 calculus applications of Taylor series	391
19.6 physical applications of Taylor series	391

Chapter 1

history and applications of calculus

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 BC¹. 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 millennia. 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 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.

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

1.2 numbers

What about numbers? The ancients certainly knew about whole numbers and fractions. The pythagoreans 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 money or lose it.
3. fractions also come from commerce or manufacture; you 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. Later, Gibbs and others supplanted this formalism with the notation of vectors which we still use to this day.
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. Quaternions and supernumbers are less likely to arise this semester.

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.

1.4 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 advances³ 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 ∞ .

²page 418 of Katz' text

³a point made by Joseph L. McCauley in his historical section of *Classical Mechanics*. If you want to read a really bitter history of what the church did wrong to science then this is not a bad source. The bias is glaring, I can respect that.

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.

If you'd like to see more about fluxions or early history of calculus there are many good books. Or, you could just download the original works from the internet. Much is available for download at this time. Beware historians, all too often they have some ulterior motive in their story telling. Of course the history I give here is purely objective⁴.

1.5 a selection of mathematical stories

Mathematics has enjoyed an incredible expansion of thought since the time of Newton. I'll just say a word or three about some of the more notable names. There are two sort of developments in this list, First, there are mathematicians who see past the ad-hoc methods of earlier generations to put in place a better explanation which has more logical consistency. Second, there are those who push forward to solve or ask new problems which generalize the method.

1. The Bernoulli family solved or were involved in the solution of many of the most difficult problems of Newton's time. The main three you hear about are the brothers Jakob(1654-1705) and Johann(1667-1748) and Daniel(1700-1782). The problem you often hear about is the following: what shape will a chain hang if suspended between two points? DiVinci and Galileo tried unsuccessfully to solve it. Galileo though hanging chain should take shape of parabola, but was proven wrong by a Jesuit priest Ignatius Pardies(1636-1673). Apparently, Jakob raised the problem again around 1690, but his brother Johann solved it. He had good company, Leibniz and Christiaan Huygens(1629-1695) also solved the problem, apparently Leibniz solved it first but gave notice in a popular journal that he would withhold his solution to give other mathematicians a chance to solve it as well. Daniel and the brothers Jakob and Johann made significant contributions to differential equations and solve the problem of motion through a medium with friction.

⁴just joking, this history is a work in progress, I welcome corrections if there are mistakes concerning content (grammatical mistakes are for your amusement).

2. Pierre-Simon de Laplace(1749-1827) in his *Celestial Mechanics* (english translation, the original title in French since Laplace was very French) put forth nearly complete solutions for the motion of planets. He founded a method called *perturbation theory* which would prove necessary to correctly apply Newtonian mechanics to the solar system as a whole. Laplace also was very proud to abandon God as an explanation for physics. It is sad that so many people still accept the flawed logic of Laplace in this sense: God was not invented to explain things. More than that, why should there be one explanation for everything? Is it not possible that God did something and physics explains how He did it? But, I believe in God, so I guess that is why I differ with Laplace. Laplace also popularized one of the earlier versions of the naturalist's creation myth: he put forth the "nebular hypothesis" which basically says the solar system formed from a giant cloud of gas shrinking to form planets etc...
3. Jean D'Alembert(1717-1783) was one of the first people to think of time as a fourth dimension in 1754. He also found interesting solutions to a variety of physical problems.
4. Joseph-Louis Lagrange(1736-1813) was a contemporary of Laplace, but in contrast did believe that God could and should have a place in explaining events in nature. He did agree with D'Alembert. Lagrange's work in physics combined the genius of Newton with the variational calculus pioneered by Euler to produce what is now known as Lagrangian mechanics. Classical field theories are largely based on generalizations of Lagrange's formalism. Lagrange attempted to give a treatment of calculus free of fluxions in 1797. His fundamental assumption was that all functions could be written as a power series. It's interesting that this assumption is still made in many contexts by modern authors as a method to get past difficulties. It turns out not all functions are analytic, but honestly, not too many in my experience.
5. William Rowan Hamilton(1805-1865) quaternions and physics.
6. Leonhard Euler(1707-1783) was a prolific mathematician who spent the later years of his life completely blind. However, his capacity to calculate without writing meant that blindness was not so much of a hindrance to his work. His memory was apparently amazing, he committed to memory all the powers up to order 6 for the numbers from 1 to 100 just for the sake of quizzing his grandchildren. His blindness is thought to stem from a period of over exertion on some astronomical calculation. You can read more in the 1801 translation of Euler's text on algebra. It's free to download as a pdf. In it you will find solutions to an impressive array of algebra problems. For example, he shows how to solve quadric equations. Euler did many of the things we find in modern texts. For example, Euler introduced notation for the functions sine and cosine and presented the calculus of these trigonometric functions in 1739. He introduced the word "function" in a sloppy way, more or less he said the function was the formula of the function, an abuse of terminology we still advocate for convenience. He studied double integrals and coordinate change. He found the fundamental equations of

variational calculus through some, in retrospect, questionable methods. Early in life he was tutored by Johann Bernoulli who recognized his genius and later the Bernoulli's helped him secure a position in the newly opened Russian school in St. Petersburg. Catherin Gsell was married Euler in 1733 and in all they had 13 children.

7. Hermann Gunther Grassmann(1809-1877) in 1844 discovered exterior algebra. He put forth an operation which satisfied the weird relation $AB = -BA$. Ok, admittedly this has little to do with calculus. On the other hand, it was probably the earliest example of super math.
8. Jean Baptiste Joseph Fourier (1768-1830) introduces sums of sines and cosines to solve heat diffusion problems in physics. These Fourier series helped bring new questions to analysis that were only answered in the late 19-th century.
9. Carl Friedrich Gauss(1777-1855) did everything. Complex variable foundations, algebra, number theory, physics, noneuclidean geometry, ...
10. Bernhard Bolzano(1781-1848) independently discovered foundational ideas about continuity similar to Cauchy's.
11. Karl Weierstrauss(1815-1897) very popular teacher. Pushed forward theory of convergence of functions by introducing concept of *uniform convergence*.
12. Niels Henrik Abel(1802-1829) proved that quintic equations could not be solved by radicals. In other words, there is no quintic formula in the same sense that we know a quadratic formula. His proof involved the permutation groups studied by Cauchy and also Lagrange.
13. Evariste Galois(1811-1832) found methods which said which quintic equations allowed solutions by radicals and which did not. Galois' work was so deep that we still only typically cover it in the second semester of abstract algebra.
14. Camille Jordan(1838-1922) wrote the first book on Galois theory. Did foundational work in linear algebra and helped launch the imaginations of Klein and Lie. Developed a classification of matrices which we now call the Jordan-form-decomposition.
15. Augustin-Louis Cauchy(1789-1857) permutation groups, introduced term "determinant" and wrote the foundational work on the topic in 1815. Also, found early example of eigenvalues

and vectors for 2×2 system. In about 1820 showed Lagrange's approach was based on an assumption which was not reasonable for some functions. In 1823 introduced limits, continuity and much of what we will study early in this course. In complex variables he proved useful theorems about complex integration.

16. Arthur Cayley(1821-1895) foundations of group theory, theory of determinants. Beginning work in linear algebra. In 1858, promoted notation for matrix in terms of single letter.
17. Adrien Marie Legendre(1752-1833) wrote understandable treatment for the method of least squares in 1805. In 1809 Gauss claimed to have invented it himself and have used it since 1795. Not surprisingly, Legendre was not happy about this. From what I read in Katz' text it's not so easy to really pin down who first thought of least squares.
18. Kronecker(1823-1891) fundamental theorem of finite Abelian groups. Not a fan of Cantor's ideas about infinite sets.
19. George Green(1793-1841) found and published in 1828 what Russians classically called Ostrogradsky's theorem.
20. George Stokes(1819-1903) generalization of Green/Ostrogradsky theorem.
21. Mikhail Ostrogradsky(1801-1861) found and proved in 1826 what we usually call Green's theorem.
22. James Clerk Maxwell(1831-1901) used quaternionic notation, unified electricity and magnetism by adding a correction term, one of the greatest physicists of all time.
23. Georg Bernhard Riemann(1826-1866) complex variables, analysis, integration and noneuclidean geometry.
24. Nikolai Ivanovich Lobachevsky(1792-1856) noneuclidean geometry.
25. Janos Bolyai(1802-1860) noneuclidean geometry.
26. James Joseph Sylvester(1814-1897) coined the term "matrix".

- 27. Richard Dedekind(1831-1916) ideals for abstract algebra and construction of real numbers through what we now call Dedekind cuts.
- 28. Georg Frobenius(1849-1917) gave general proof of the Cayley Hamilton theorem. In 1878 wrote the book on matrix math.
- 29. Heaviside(1850-1925) popularized the vector notation.

I should mention my use of the word "discovered" begs a philosophical question: **is mathematics discovered or created?** Plato and his followers would emphatically claim it is discovered. They believed in the existence of mathematical theorems and objects in some sort of metaphysical reality. When we prove a theorem we merely find aspects of that reality. On the other hand, other philosophers of mathematics emphatically state that math is merely a human invention. The critics of the platonic school believe there is no world of platonic forms, instead, math is what we make of it. Noneuclidean geometry is often held up as an apt example of the invention of math. Bolyai, Gauss, Lobachetsky and Riemann turned the tide against the most simplistic form of platonic thought. Many thinkers found the uniqueness of Euclidean geometry as evidence for the existence of math outside the physical realm. From another perspective, mathematicians discovered Euclidean geometry *because* there was no other geometry to discover. However, Euclidean geometry was not unique or even necessary to model the world. Euclidean geometry had such an elevated status in the minds of mathematicians that Gauss actually held back his work in the discovery for fear of being critized by his small-minded peers. It didn't help that famous philosopher Kant offered "proofs" that only Euclidean geometry could be conceived in the mind of man.

Well, Kant was dead wrong. Riemann gave us the framework to create literally thousands of different geometries. I guess the nail in the coffin of the Euclidean-only crowd must have been when Hilbert proved through careful arguments that consistency of Euclidean geometry was equivalent to consistency of a certain noneuclidean geometry. That means both Euclidean geometry and noneuclidean geometry are reasonable or both are not. But, does this disprove Platonic philosophy of math? Not really. The world of platonic ideals just got a little bigger. Personally, I think the question of whether math is created or discovered is wrong. But, I'll leave it at that for here.

Remark 1.5.1.

I'm not a math historian. But, I do like math history and I'm trying to learn more. This chapter is in part just an exercise towards that end. I don't expect you study this, or even read it necessarily. Read it if you wish, ignore it if you want. I will likely talk about part of this the first day of class. What follows from here is much more serious.

1.6 a short overview of calculus in basic physics

I'll speak to what I know a little about.

1. **Newtonian Mechanics** is based on Newton's Second Law which is stated in terms of a time derivative of three functions. We use vector notation to say it succinctly as

$$\frac{d\vec{P}}{dt} = \vec{F}_{net}$$

where \vec{P} is the momentum and \vec{F}_{net} is the force applied.

2. **Lagrangian Mechanics** is the proper way of stating Newtonian mechanics. It centers its focus on energy and conserved quantities. It is mathematically equivalent to Newtonian Mechanics for some systems. The fundamental equations are called the Euler Lagrange equations they follow from Hamilton's principle of least action $\delta S = \delta \int L dt = 0$,

$$\frac{d}{dt} \left[\frac{\partial L}{\partial \dot{y}} \right] = \frac{\partial L}{\partial y}.$$

Lagrangian mechanics allows you to derive equations of physics in all sorts of curvy geometries. Geometric constraints are easily implemented by Lagrange multipliers. In any event, the mathematics here is integration, differentiation and to see the big picture variational calculus (I sometimes cover variational calculus in the Advanced Calculus course Math 332)

3. **Electricity and Magnetism** boils down to solving Maxwell's equations subject to various boundary conditions:

$$\nabla \cdot \vec{B} = 0, \quad \nabla \cdot \vec{E} = \frac{\rho}{\epsilon_o}, \quad \nabla \times \vec{B} = \mu_o \vec{J} - \mu_o \epsilon_o \frac{\partial \vec{E}}{\partial t} \quad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}.$$

Again, the mathematics here is calculus of several variables and vector notations. In other words, the mathematics of electromagnetism is vector calculus.

4. **Special Relativity** also uses vector calculus. However, linear algebra is really needed to properly understand the general structure of Lorentz transformations. Mathematically this is actually not so far removed from electromagnetism. In fact, electromagnetism as discovered by Maxwell around 1860 naturally included Einstein's special relativity. In relativistic coordinate free differential form language Maxwell's equations are simply stated as

$$dF = 0, \quad d * F = *J.$$

Newtonian mechanics is inconsistent with these equations thus Einstein's theory was inevitable.

5. **General Relativity** uses calculus on manifolds. A manifold is a curved surface which allows for calculus in local coordinates. The geometry of the manifold encodes the influence of gravity and conversely the presence of mass curves space and time.
6. **Quantum Mechanics** based on Schrodinger's equation which is a partial differential equation (much like Maxwell's equations) governing a complex wave function. Alternatively, quantum mechanics can be formulated through the path integral formalism as championed by Richard Feynman.
7. **Quantum Field Theory** is used to frame modern physics. The mathematics is not entirely understood. However, Lie groups, Lie algebras, supermanifolds, jet-bundles, algebraic geometry are likely to be part of the correct mathematical context. Physicists will say this is done, but mathematicians do not in general agree. To understand quantum field theory one needs to master calculus, differential equations and more generally develop an ability to conquer very long calculations.

I speak of basic physics simply to illustrate how correct Galileo was when he said that mathematics was the language of nature. In fact, all modern technical fields in one way or another have calculus-based models at their core. This is why you are expected to take calculus in a proper university education.

Chapter 2

foundations of mathematics

2.1 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 ". 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 in a set. In such a case we may be able to use **roster notation** to define a set.

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

For example, the open interval $(1, \infty) = \{x \mid x \in \mathbb{R}, x > 1\}$. We may also use the equivalent notation $(1, \infty) = \{x \in \mathbb{R} \mid x > 1\}$. The set with no elements is called the **empty set** and it is denoted by $\{\}$ or \emptyset . We say that two sets S and T are **equal** when they have identical elements, this is what $S = T$ is meant to denote when S and T are sets.

Definition 2.1.1. *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 2.1.2. *union, intersection and difference of sets*

Let S and T be sets,

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

Set theory is a deep and interesting subject, however we will conclude our discussion here.

2.2 numbers

Real numbers can be constructed from set theory and about a semester of mathematics. We will accept the following as **axioms**¹

Definition 2.2.1. *real numbers*

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 $M \geq s$ for all $s \in S$. Similarly a lower bound on S is a number $m \in \mathbb{R}$ such that $m \leq s$ for all $s \in S$. If a set S is bounded above and below then the set is said to be **bounded**. For example, the open set (a, b) is bounded above by b and it is bounded below by a . In contrast, rays such as $(0, \infty)$ are not bounded above. Closed intervals contain their least upper bound and greatest lower bound. The bounds for an open interval are outside the set.

¹an 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.

Definition 2.2.2. *standard subsets of real numbers*

- natural numbers (positive integers); $\mathbb{N} = \{1, 2, 3, \dots\}$.
- integers; $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. Note, $\mathbb{Z}_{>0} = \mathbb{N}$.
- non-negative integers; $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\}$.
- negative integers; $\mathbb{Z}_{<0} = \{-1, -2, -3, \dots\} = -\mathbb{N}$.
- rational numbers; $\mathbb{Q} = \{\frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0\}$.
- irrational numbers; $\mathbb{J} = \{x \in \mathbb{R} \mid x \notin \mathbb{Q}\}$.
- open interval from a to b ; $(a, b) = \{x \mid a < x < b\}$.
- half-open interval; $(a, b] = \{x \mid a < x \leq b\}$. (oposed interval)
- half-open interval; $[a, b) = \{x \mid a \leq x < b\}$. (clopen interval)
- closed interval; $[a, b] = \{x \mid a \leq x \leq b\}$.

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 an 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 $(a, b) \subseteq \mathbb{R}$. Question: is $(4, 3)$ a point or an open interval? Why is there no danger of ambiguity in this case?

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. To complete the rational numbers you have to throw in the irrational numbers which gives the whole real number system. Beyond the real and rational fields we can consider the complex number field.

Definition 2.2.3. *complex numbers*

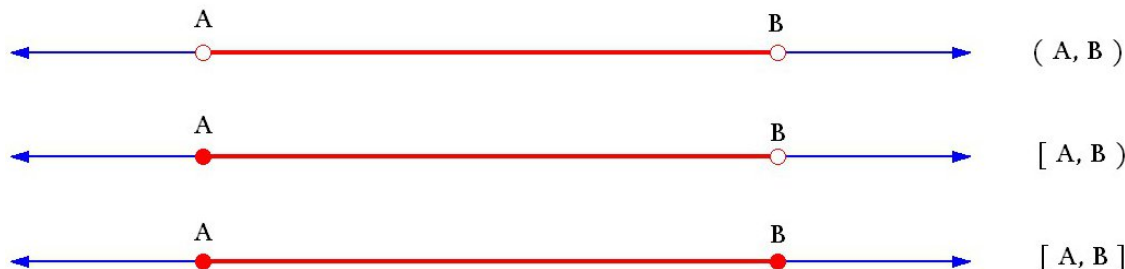
$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R} \text{ and } i^2 = -1\}$$

The complex numbers are not ordered, however they are **algebraically complete** this means we can factor any polynomial into linear factors with complex numbers. In contrast, the real numbers only allow us to factor a polynomial into linear factors together with irreducible quadratic factors. For example, $x^2 + 1$ cannot be factored over \mathbb{R} but $x^2 + 1 = (x + i)(x - i)$. The proof that the complex numbers are algebraically complete was provided by Gauss in the nineteenth century. We often prove it in the complex variables (Math 331) course here at LU. We will find many occasions to use complex numbers in calculus and differential equations. Like it or not real problems often have complex solutions. Take the quadratic formula as a prime example of this phenomenon. The concept of a variable is so fundamental it bears mention at this juncture (I already used this concept in the definition of \mathbb{R} if you think about it):

Definition 2.2.4. *variable*

A **real variable** x is a symbol which is allowed to assume any value in \mathbb{R} . A **complex variable** z is a symbol which is allowed to assume any value in \mathbb{C} .

Other types of variables are interesting but these two are the primary types we'll be interested in for this semester. In addition, usually these variables are not entirely **free**. Typically a variable is restricted by some equality or inequality. Another ubiquitous concept is the **number line**². 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.



It is useful to catalogue the following properties of inequalities:

Theorem 2.2.5. *properties of inequalities:*

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

1. square of real number is non-negative; $a^2 \geq 0$,
2. square zero only if number is zero; $a^2 = 0$ iff $a = 0$,
3. add or subtract from both sides at once; if $a < b$ then $a + c < b + c$ and $a - c < b - c$,
4. add inequalities; $a < b$ and $c < d$ implies $a + c < b + d$,
5. transitivity; $a < b$ and $b < c$ implies $a < c$,
6. if $ab > 0$ then $a < b$ implies $1/a > 1/b$.

The last statement $ab > 0$ is just a tricky way of saying that a and b are either both positive or both negative. This theorem can be proven from the axioms of the real numbers, but I will not offer those details here. You should not be surprised to hear that a similar theorem also holds if we replace $<$ with $>$ or \leq with \geq .

Definition 2.2.6. *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:

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

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

²Rene' Descartes popularized this concept in the early 17th 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.

We should recall that the square root function is by definition the positive root; $\sqrt{x} \geq 0$. Therefore, we can characterize a nonzero **positive** number by the equation $x = |x|$ whereas a nonzero **negative** number x has $|x| = -x$. It is useful to catalogue the following properties absolute values:

Theorem 2.2.7. *properties of absolute value:*

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

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 same as double inequality; $|a| < \epsilon \Leftrightarrow -\epsilon < a < \epsilon$,
5. triangle inequalities ;

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

Proof:

Item (1.) is immediately obvious from the definition $|x| = \sqrt{x^2}$.

To prove (2.) consider that if $a = 0$ then clearly $|0| = \sqrt{0^2} = 0$. Conversely, $\sqrt{a^2} = 0$ implies $a = 0$.

To prove (3.) note that $|ab| = \sqrt{(ab)^2} = \sqrt{a^2b^2} = \sqrt{a^2}\sqrt{b^2} = |a||b|$.

I leave (4.) for the reader to prove.

The proof of (5.) requires a bit more thought, I'll prove part (i.) and leave (ii.) and (iii.) for the reader.

Notice that $|a + b|^2 = (a + b)^2$. Consider then

$$|a + b|^2 = (a + b)^2 = a^2 + 2ab + b^2 = |a|^2 + 2ab + |b|^2$$

To complete the proof of (4.) part (i.) we need to break into cases:

1. If $a, b > 0$ then $|a| = a$ and $|b| = b$ thus the equation above yields $|a + b|^2 = |a|^2 + 2|a||b| + |b|^2 = (|a| + |b|)^2$ hence $|a + b| = |a| + |b|$.
2. If both $a, b < 0$ then we have $|a| = -a$ and $|b| = -b$ thus $2ab = 2(-a)(-b) = 2|a||b|$ which gives us that $|a + b|^2 = (|a| + |b|)^2$ which again yields $|a + b| = |a| + |b|$.
3. If $a > 0$ and $b < 0$ then $2ab = -2a(-b) = -2|a||b|$ therefore $|a + b|^2 = |a|^2 - 2|a||b| + |b|^2$. Since $|a|, |b| > 0$ it is certainly true that adding $4|a||b|$ to the r.h.s. of the equality gives the following inequality, $|a + b|^2 = |a|^2 - 2|a||b| + |b|^2 < |a|^2 + 2|a||b| + |b|^2 = (|a| + |b|)^2$. Therefore, $|a + b| < |a| + |b|$.
4. If $a < 0$ and $b > 0$ then by the argument above with $a \leftrightarrow b$ shows $|a + b| < |a| + |b|$.
5. If either $a = 0$ or $b = 0$ then the (4.) is clearly true.

Thus $|a + b| \leq |a| + |b|$ for all possible cases hence the proposition is true³. \square

³Note that in mathematics our standard for true and false is much stricter than other disciplines. When we ask if something is true it is usually the case that we implicitly mean to ask "is this true for all possible cases". If we wish to ask a question relative to a restricted set of cases then we are obligated to reduce the set of answers for the question which is asked. This is often a source of confusion between professors and students. Typically I'll answer the question which was literally asked whether or not that was the intended question.

It is probably useful to study the geometric significance of the theorem on absolute values. Note that $|x| = |x - 0| = d(x, 0)$ so **the absolute value gives the distance to the origin**. This makes (4.) easy to understand: it simply says that if the distance of a to the origin is less than ϵ then the point a must reside between $-\epsilon$ and ϵ .

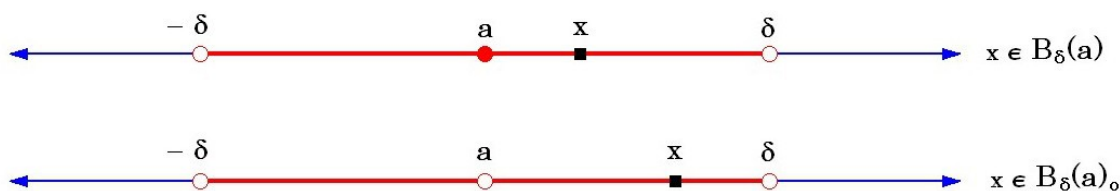
Definition 2.2.8. *neighborhoods*

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. ⁴

⁴I'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. Other texts merely insist that a neighborhood of a is a set N such that there exists a $\delta > 0$ with $B_\delta(a) \subseteq N$. I don't believe we'll need such sophistication in this course.

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 2.2.9. *connected subsets of real numbers.*

We say $U \subseteq \mathbb{R}$ is connected iff

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

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 students⁵. 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.

Definition 2.2.10. *boundary points.*

We say $p \in U \subseteq \mathbb{R}$ is a **boundary point** of U iff every open nbhd. 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 ∂U .

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

1. boundary points
2. interior points

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

Definition 2.2.11. *interior points.*

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

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

Definition 2.2.12. *open sets, closed sets.*

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

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

⁵Equivalently, 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 .

2.3 algebra

The fundamental rules of algebra were revealed in our definition of real numbers. Factoring follows from the fact that if $a, b \in \mathbb{R}$ and $ab = 0$ then $a = 0$ or $b = 0$. The factor theorem provides a simple test to check if a factor appears in a particular polynomial.

Theorem 2.3.1. *factor theorem*

If $p(x)$ is a polynomial and $p(r) = 0$ then there is another polynomial $q(x)$ such that $p(x) = (x - r)q(x)$. The polynomials could have real or complex coefficients and the root r could either be real or complex.

Actually, the theorem holds in a context which is more general than just \mathbb{R} or \mathbb{C} but we only need the result claimed above. I will not provide a proof of this theorem here.

Example 2.3.2. Suppose $f(x) = x^{33} + x^2 - 1$, is $(x - 1)$ a factor of f ? Well, observe that $f(1) = 1 + 1 - 1 = 1 \neq 0$ hence $(x - 1)$ is not a factor of $f(x)$. In contrast, if $g(x) = x^{33} + x^2 - 2$ then we could calculate that $g(1) = 1 + 1 - 2 = 0$ so the factor theorem tells us that we can write $g(x) = (x - 1)h(x)$ where $h(x) = x^{32} + \dots + 2$. If you really want to you could use long division to calculate that:

$$\begin{aligned} g(x) = (x - 1) & \left(x^{32} + x^{31} + x^{30} + x^{29} + x^{28} + x^{27} + x^{26} + x^{25} + x^{24} + x^{23} \right. \\ & + x^{22} + x^{21} + x^{20} + x^{19} + x^{18} + x^{17} + x^{16} + x^{15} + x^{14} + x^{13} \\ & \left. + x^{12} + x^{11} + x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + 2x + 2 \right) \end{aligned}$$

Theorem 2.3.3. *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 2.3.4.

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

Example 2.3.5.

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

Example 2.3.6.

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

There is more to say, but this will suffice for the moment.

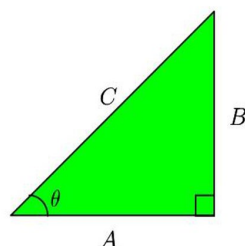
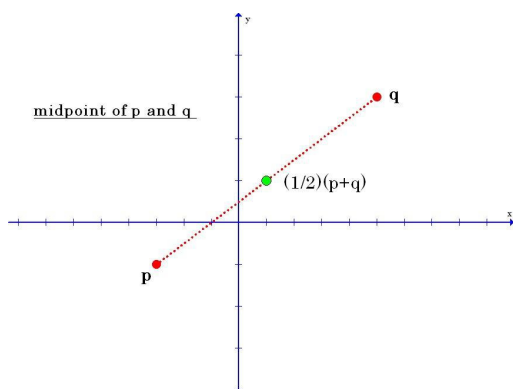
2.4 analytical geometry

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 our course on modern geometry here at LU. Also, abstract algebra has much to say about the non-existence of certain constructions in Euclidean geometry, take Math 422 to see about that.

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**

$$d(p, q) = \sqrt{(b_1 - a_1)^2 + (a_2 - a_2)^2}$$

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 $\frac{1}{2}(p + q) = (\frac{1}{2}(p_1 + q_1), \frac{1}{2}(p_2 + q_2))$



$$\cos(\theta) = \frac{A}{C} = \frac{\text{adj}}{\text{hyp}}$$

$$\sin(\theta) = \frac{B}{C} = \frac{\text{opp}}{\text{hyp}}$$

$$\tan(\theta) = \frac{B}{A} = \frac{\text{opp}}{\text{adj}}$$

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. This concludes our short tour of analytic geometry. We will find occasion to use these tools throughout this course.

2.5 functions

The term **function** is about a third of a millennia 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 2.5.1. *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 $f : A \rightarrow B$ in this case and we say that $\mathbf{A} = \mathbf{domain}(f)$ and $\mathbf{B} = \mathbf{codomain}(f)$. 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 $B \subseteq \mathbb{C}$ then f is said to be a **complex-valued function**.

Often it is convenient to put the subset notation together with the function notation: $f : U \subseteq \mathbb{R} \rightarrow V \subseteq \mathbb{R}$ means that $f : U \rightarrow V$ and $U \subseteq \mathbb{R}$ and $V \subseteq \mathbb{R}$. Additionally, it is a common abuse of terminology to refer to the formula of the function as the function: for instance consider $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(x) = x^2$:

1. "the function is f "
2. "the function is x^2 "

usually there is no danger of confusion provided we all understand that the function actually is a rule which maps points from one set to another.

Definition 2.5.2. *implicit domain for function defined by formula*

If only a formula $f(x)$ is given then it is customary to choose $domain(f)$ to be the largest set of values for which the formula $f(x)$ is well-defined.

Let $f(x) = \sqrt{x}$, this function has implicit domain $dom(f) = [0, \infty)$ since otherwise the formula would not yield a real number. Naturally we can always choose a smaller domain by an additional statement; $g(x) = \sqrt{x}$ with $dom(g) = [0, 1]$ has same formula as f but obviously the domain is smaller. One application that might make you choose the domain $[0, 1]$ would be if x was a probability.

Definition 2.5.3. *function equality*

We say two functions f and g are equal iff $dom(f) = dom(g)$ and for all $x \in dom(f)$ we have $f(x) = g(x)$.

If two functions disagree at even just one point then we say they are not equal. For example, $f(x) = x$ and $g(x) = x^2/x$ are not the same function since they do not share the same domain, $dom(f) = \mathbb{R}$ whereas $dom(g) = \mathbb{R} - \{0\} = (-\infty, 0) \cup (0, \infty)$. If we define $h : (-\infty, 0) \cup (0, \infty) \rightarrow \mathbb{R}$ by $h(x) = x$ then it is true that $g = h$ since $g(x) = x^2/x = x$ for each $x \neq 0$.

Definition 2.5.4. *graph and range of a function*

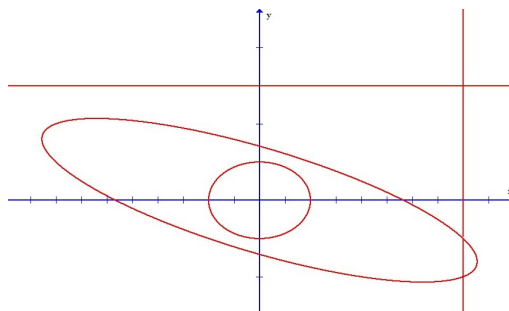
Suppose $f : U \subseteq \mathbb{R} \rightarrow V \subseteq \mathbb{R}$ then

1. $\text{graph}(f) = \{(x, y) \in \mathbb{R}^2 \mid x \in \text{dom}(f) \text{ and } y = f(x)\}$
2. $\text{range}(f) = f(U) = \{y \in \mathbb{R} \mid \exists x \in \text{dom}(f) \text{ s.t. } y = f(x)\}.$

Notice that the codomain and the range are not necessarily the same set. What we can say is that the range is the smallest possible codomain for a given domain and formula. Furthermore, you should recall that the graph of a function is severely limited in the possible shapes it can assume. The **vertical line test** states that if you draw a vertical line through the graph then it may hit at most one point on the graph of a function. We can also discuss graphs of equations. The graph of $x^2 + y^2 = 1$ is the unit-circle, clearly it fails the vertical line test and so we can conclude it is not the graph of a single function⁶. Generally the graph of an equation can be most anything you can think of. Consider,

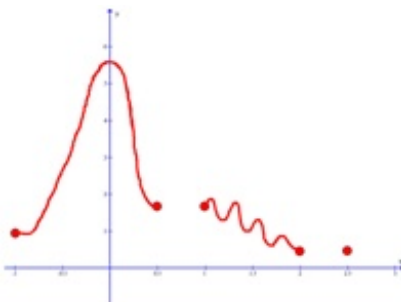
$$(x - 4)(y - 3)(x^2 + y^2 - 1)(x^2 + 3xy + 4y^2 - 8) = 0$$

has graph as follows:



Clearly this fails the vertical line test⁷.

Sometimes a function may be defined by a graph. A graph is simply a visual representation of a table of values. If the point (a, b) is on the graph then that tells us that $f(a) = b$. This is good news since we may not want to write the formula for certain functions, for example:



⁶ in fact it can be described as $x = \cos(t)$ and $y = \sin(t)$ for $0 \leq t \leq 2\pi$ this is a *parametrization* of the circle. Parametric curves provide a better geometric framework, we will use them at times.

⁷can you see how I made the equation work? If you understand my silly trick you can build a lot of pictures with equations.

Remark 2.5.5.

Finding the domain and range of a given function requires thinking. Tools you should already have at your disposal are sign-charts, graphs of polynomials and rational functions without the help of a graphing calculator. You should be able to solve the inequality $x^2 + 5x + 6 > 0$. You should also be able to graph those basic transcendental functions such as sine, cosine and the exponential function. Please understand the purpose of this chapter is to remind and reinforce. If you need more examples then you should come to office hours or perhaps review your precalculus text. That said, I will give a few examples which illustrate the main algebraic techniques for algebra and graphing in the next section. If you need a review of long-division, sign-charts and so forth you could read my college algebra notes. They're easily downloaded from my website at your convenience.

Given two functions f and g we can create new functions by adding, subtracting, dividing or multiplying by a constant.

Definition 2.5.6. *new functions from old.*

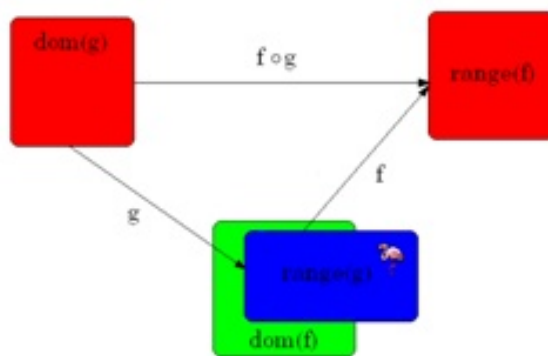
function	defining formula	domain of new function
$f + g$	$(f + g)(x) = f(x) + g(x)$	$\text{dom}(f) \cap \text{dom}(g)$
$f - g$	$(f - g)(x) = f(x) - g(x)$	$\text{dom}(f) \cap \text{dom}(g)$
fg	$(fg)(x) = f(x)g(x)$	$\text{dom}(f) \cap \text{dom}(g)$
f/g	$(f/g)(x) = f(x)/g(x)$	$\text{dom}(f) \cap \text{dom}(g) \cap \{x \mid g(x) \neq 0\}$
cf	$(cf)(x) = cf(x)$	$\text{dom}(f)$

These formulas go to show that functions are a lot like numbers, we can add, subtract, multiply and even divide functions and the result will be a function. Functions are different than numbers of course, for example, I'm not sure what the analogue for the following would be in terms of numbers

Definition 2.5.7. *composite function*

Suppose $f : U \rightarrow V$ and $g : W \rightarrow U$ then we define $f \circ g : W \rightarrow V$ by $(f \circ g)(x) = f(g(x))$ for each $x \in W$. If no domains are explicitly given for f and g then it is customary to take $\text{dom}(f \circ g) = \text{dom}(g) \cap \{x \mid g(x) \in \text{dom}(f)\}$

Here's a silly picture to explain why we may need to exclude part of the domain of the inside function. The flamingo might not be covered by the domain of the outside function.



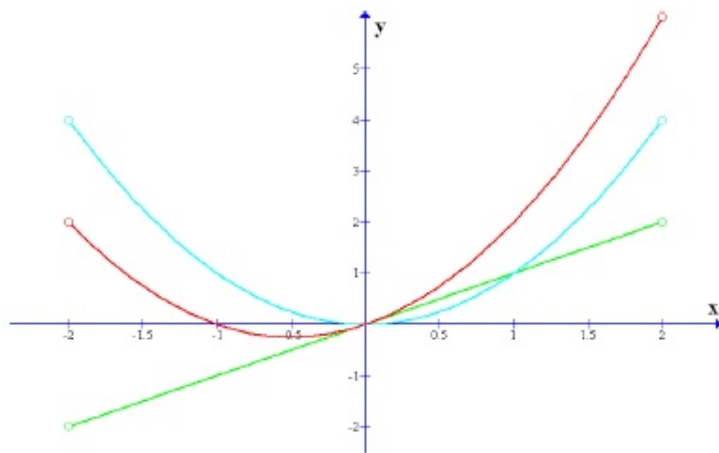
Consider $g(x) = x^2$ and $f(x) = \sqrt{x-1}$. The domain of g is \mathbb{R} however the domain of $f \circ g$ is necessarily smaller. We need $g(x) - 1 \geq 0$ for $g(x) \in \text{dom}(f)$. That is, $x^2 - 1 = (x-1)(x+1) \geq 0$. This inequality has solution $x \leq -1$ or $x \geq 1$ therefore, $\text{dom}(f \circ g) = (-\infty, -1] \cup [1, \infty)$. In this case, the natural domain of the formula $(f \circ g)(x) = \sqrt{x^2 - 1}$ is also $(-\infty, -1] \cup [1, \infty)$.

Suppose $g(x) = \frac{1}{x}$ and $f(x) = \frac{1}{x}$. Calculate that $(f \circ g)(x) = f(1/x) = \frac{1}{1/x} = x$. The domain suggested by $(f \circ g)(x) = x$ is \mathbb{R} , however that is incorrect. The true domain is $\mathbb{R} - \{0\}$, the simplification $\frac{1}{1/x} = x$ fails to be correct when $x = 0$. Moral of story? Be careful when you simplify to keep track of cases that are excluded from the algebraic steps you perform.

Definition 2.5.8. *even or odd functions*

1. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **even** iff $f(-x) = f(x)$ for all $-x, x \in U$,
2. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **odd** iff $f(-x) = -f(x)$ for all $-x, x \in U$,

Some functions are even, for example $f(x) = x^2$ (see how it mirrors the y-axis). Other functions are odd, for example $g(x) = x$ (see how it reflects through the origin). We can also have functions which are neither even nor odd, for example $h(x) = f(x) + g(x) = x + x^2$ (no simple symmetry evident).



Perhaps you'll be surprised to hear that any function defined on a domain which is symmetric about zero (which means that if $x \in \text{dom}(f)$ then $-x \in \text{dom}(f)$ for each $x \in \text{dom}(f)$) can be written as a sum of an even function and an odd function. Don't believe it? Consider:

$$f(x) = \underbrace{\frac{1}{2}[f(x) + f(-x)]}_{\text{even}} + \underbrace{\frac{1}{2}[f(x) - f(-x)]}_{\text{odd}}$$

you can verify that the terms are correctly labeled, the even part is an even function and the odd part is an odd function. All I did was to add zero and use $1 = \frac{1}{2} + \frac{1}{2}$.

2.6 graphing and 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 2.6.1. *algebraic critical numbers.*

Suppose $f : \text{dom}(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is a function then we say $c \in \text{dom}(f) \cup \partial(\text{dom}(f))$ is an **algebraic critical number** iff either $c \notin \text{dom}(f)$ 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 point⁸.

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 2.6.2. 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

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

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 $x \in \text{dom}(g)$ iff $6 - x - x^2 > 0$. But, this is the same as stating $x \in \text{dom}(g)$ iff $x^2 + x - 6 < 0$ hence, by the sign chart, $\text{dom}(g) = (-3, 2)$.

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.

⁸critical numbers are algebraic critical numbers of the derivative function, but you're not allowed to know that just yet... oops.

Example 2.6.3. Solve the following inequality:

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

We define $f(x) = \frac{(x^2+3x)(x^2+4x+5)}{x^2-2x}$ and factor the formula as much as possible,

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

for $x \neq 0$. The quadratic factor is irreducible. I completed the square to make it explicitly clear that the quadratic could not be factored⁹. We have three algebraic critical numbers: $c = -3, 0, 2$. Again, pick test points to the left and right of each algebraic critical number,

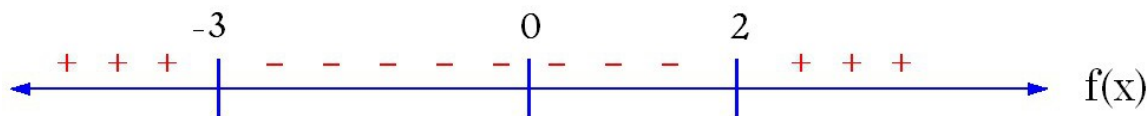
$$f(-4) = \frac{(-4+3)((-4+2)^2+1)}{-4-2} > 0$$

$$f(-2) = \frac{(-2+3)((-2+2)^2+1)}{-2-2} < 0$$

$$f(1) = \frac{(1+3)((1+2)^2+1)}{1-2} < 0$$

$$f(3) = \frac{(3+3)((3+2)^2+1)}{3-2} > 0$$

Hence,



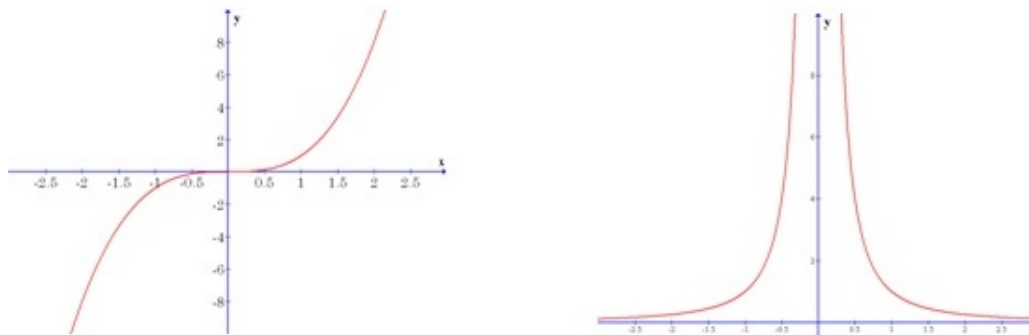
From which we deduce $\frac{(x^2+3x)(x^2+4x+5)}{x^2-2x} \leq 0$ for $x \in [3, 0) \cup (0, 2]$. Notice we had to exclude $x = 0$ since $f(0)$ is undefined.



⁹completing the square on a quadratic for numerically reasonable quadratics is almost always the fastest way to find the roots in the complex case. Given $(x+2)^2 + 1 = 0$ I can read the solution is $x = -2 \pm i$ without further calculation. This is much faster than the quadratic formula. If the quadratic was factorable over \mathbb{R} then completing the square gives a minus instead of a plus. For example, $x^2 + 6x + 5 = (x+3)^2 - 4$. Then I can read the solution is $x = -3 \pm 2 = -1, -5$. Obviously factoring is also a good way to solve in the case of real zeros; $x^2 + 6x + 5 = (x+1)(x+5) = 0$ implies $x = -1, -5$.

2.7 local inverses

Notice that the concepts of even and odd are global concepts for a given function, they apply to the whole domain of the function. Increasing and decreasing are local concepts. A function might increase on one interval and decrease on another. For example, the function $f(x) = \frac{1}{x^2}$ decreases on $(0, \infty)$ and it increases on $(-\infty, 0)$. Of course, a function could increase on all of \mathbb{R} , for example: the cubic function $g(x) = x^3$ is increasing everywhere.



Definition 2.7.1. *increasing or decreasing*

1. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **increasing** on $I \subseteq U$ iff for all $a, b \in I$ with $a < b$ we can show $f(a) \leq f(b)$,
2. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **strictly increasing** on $I \subseteq U$ iff for all $a, b \in I$ with $a < b$ we can show $f(a) < f(b)$,
3. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **decreasing** on $I \subseteq U$ iff for all $a, b \in I$ with $a < b$ we can show $f(a) \geq f(b)$,
4. $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is **strictly decreasing** on $I \subseteq U$ iff for all $a, b \in I$ with $a < b$ we can show $f(a) > f(b)$,

If f is either increasing, strictly increasing, decreasing or strictly decreasing on $I = \text{dom}(f)$ 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.

Remark 2.7.2.

CAUTION: I disagree with the text on this definition. Please use my definition.

Notice that increasing and decreasing are most meaningful with respect to some connected interval I . Connected means that the interval can be written as either an open, closed, clopen or oposed possibly infinite interval. A connected interval on the number line is one for which we can sketch the whole interval without ever lifting our pencil. A **singleton** is a set containing just one element, if you take a singleton $I = \{a\} \subseteq \text{dom}(f)$ then we find that f is both increasing and decreasing on I (the conditions are trivially satisfied).

We will later learn how to use calculus to characterize if a function increases or decreases, for now all we can do is use graphing or argument from explicit inequalities. One thing worth noticing is that if a function is strictly increasing or decreasing on a connected interval then it will pass a horizontal line test. If you draw a horizontal line then it will hit at most one point in the graph. The cubic function passes the horizontal line test. The volcano graph does not have a connected domain, but if you just look at the right or left half of $f(x) = 1/x^2$ then those parts separately pass the test. If a piece of a function can pass a horizontal line test then we'd like to have some term to quantify that behavior. For our future reference I define precisely what I mean by "part" of a function.

Definition 2.7.3. *restriction and extension*

Let $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be a function then we say that g is the **restriction of f to $V \subseteq U$** iff $g : V \rightarrow \mathbb{R}$ and $g(x) = f(x)$ for all $x \in V$. We use the notation $g = f|_V$ to indicate that g is the restriction of f to the set V . Generally, if g is a restriction of f then we say that f is an extension of g .

A restriction is just a part of a function. An extension makes a given function bigger. If the restriction of a function to V passes the horizontal line test then that function is said to be injective on V .

Definition 2.7.4. *injective or 1-1*

Let f be a function, if $U \subseteq \text{dom}(f)$ and for all $a, b \in U$ we find $f(a) = f(b)$ implies $a = b$ then we say f is **injective** on U . If f is injective on its domain then we say f is injective. The terms one-to-one or 1-1 are synonymous with injective.

When a function is injective on U we can prove that there is an inverse function for the restriction of the function to U . If we **define** $f^{-1}(y) = x$ **such that** $f(x) = y$ then this provides a single-valued inverse function. Why? Suppose that $f^{-1}(y) = x_1$ and $f^{-1}(y) = x_2$ then by our definition $f(x_1) = f(x_2) = y$ hence by the 1-1 property $x_1 = x_2$. This little argument goes to show that the inverse function below is well-defined. We say a function is *well-defined* if it is single-valued and it has a rule which makes the output unambiguous.

Definition 2.7.5. *local inverse*

We say that f has a local inverse f^{-1} on $U \subseteq \text{dom}(f)$ if it satisfies the following two equations,

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

for each $x \in U$, and

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

for each $y \in f(U)$ where $f(U) = \{y \mid \exists x \in U \text{ s.t. } f(x) = y\}$. If the set $U = \text{dom}(f)$ then we say that f^{-1} is a global inverse of f and the function f is said to be invertible.

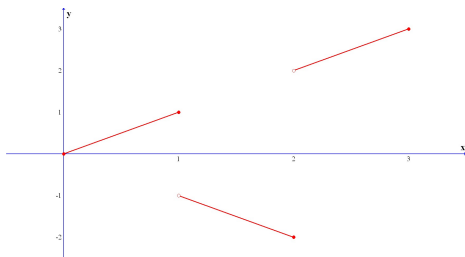
It does seem geometrically obvious that if the restriction of a function passes the horizontal line test with respect to a connected set then the same function ought to be either strictly decreasing or strictly increasing on that set.

Proposition 2.7.6. *if f is strictly increasing or decreasing then f is 1-1.*

Suppose 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 2.7.7. *inverse functions also increase or decrease.*

Suppose $f : U \rightarrow V$ is either strictly increasing or strictly decreasing on $U \subseteq \mathbb{R}$ then $f^{-1} : V \rightarrow U$ is likewise either strictly increasing or decreasing on V .

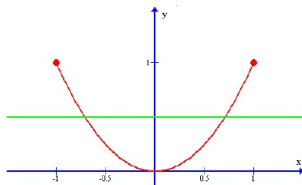
Proof: suppose $f : U \rightarrow V$ is strictly increasing with inverse $f^{-1} : V \rightarrow U$. Suppose $a, b \in V$ such that $a < b$ and suppose $f^{-1}(a) = x$ and $f^{-1}(b) = y$. There exist three possibilities:

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 $a < b$.

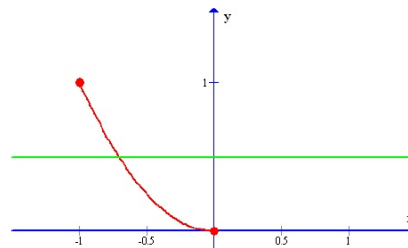
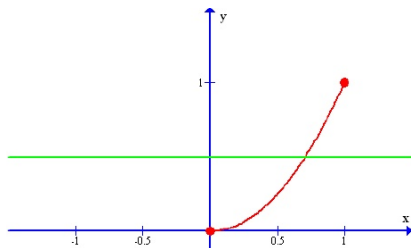
Therefore, we find for all $a, b \in V$, if $a < b$ then $f^{-1}(a) < f^{-1}(b)$ which proves f^{-1} is strictly increasing. The proof for the strictly decreasing case is similar. \square

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 2.7.8. Consider $f(x) = x^2$ with $\text{dom}(f) = [-1, 1]$. 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,



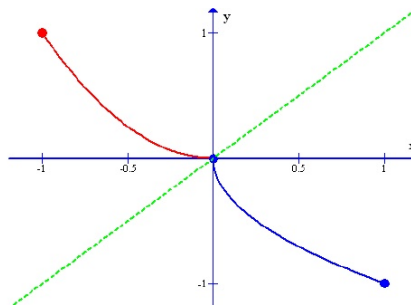
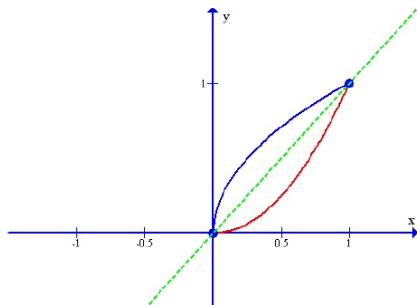
So then what is the formula for the inverse functions? We need,

$$(i.) f^{-1}(f(x)) = f^{-1}(x^2) = x \quad (ii.) f(f^{-1}(x)) = (f^{-1}(x))^2 = x$$

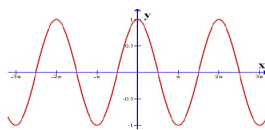
Notice that (ii.) gives $f^{-1}(x) = \pm\sqrt{x}$. Then substituting into (i.) yields: $\pm\sqrt{x^2} = x$. But, recall that $\sqrt{x^2} = |x|$ so we can see that the two solutions are,

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

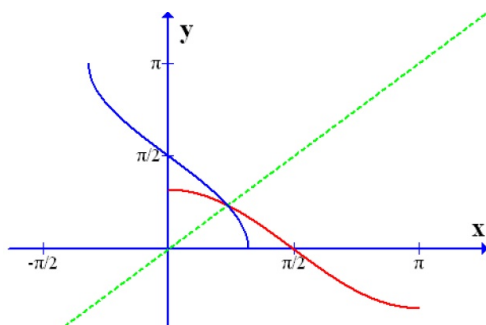
We find that the inverse of $f(x) = x^2$ on $[0, 1]$ is $f^{-1}(x) = \sqrt{x}$ and the inverse of $f(x) = x^2$ on $[-1, 0]$ is $f^{-1}(x) = -\sqrt{x}$. Notice that the graphs of inverses (blue) are symmetric about the line (green).



Example 2.7.9. Let $f(x) = \cos(x)$. Recall the graph of the cosine function is:

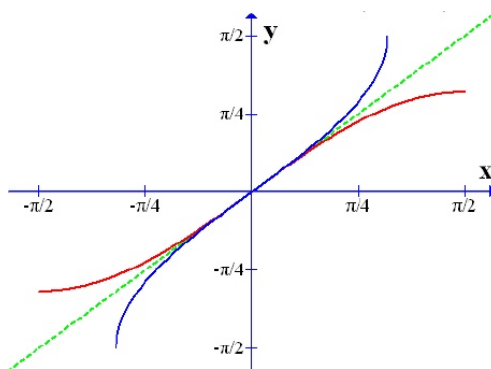


note that f cannot have a global inverse since it is not 1-1. However, if we reduce the domain to $[0, \pi]$ we obtain a 1-1 function on that interval. I have graphed the local inverse in **blue**, and you can see that the inverse is the reflection of the graph of cosine about the line $y = x$ (**green**).



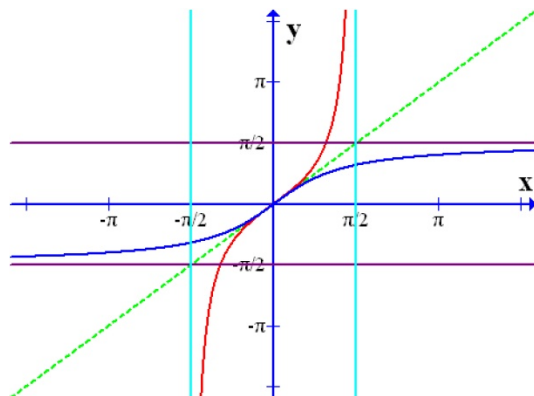
It should be understood that when we speak of inverse cosine we actually refer the local inverse for cosine on the interval $[0, \pi]$. The domain of inverse cosine is $[-1, 1]$ and the range is $[0, \pi]$. In principle one could construct other inverses for cosine based on other intervals, the choice of is simply one of convention.

Example 2.7.10. Let $f(x) = \sin(x)$ with $\text{dom}(f) = \mathbb{R}$. 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**),



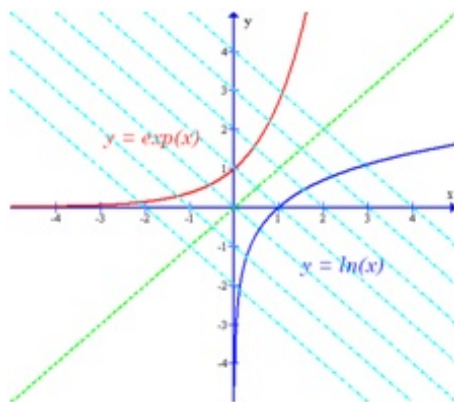
and you can see that the inverse is the reflection of the graph of sine about the line (**green**). The domain of inverse sine is $[-1, 1]$ and the range is $[-\pi/2, \pi/2]$. In principle one could construct other inverses for sine based on other intervals, the choice of $[-\pi/2, \pi/2]$ is simply one of convention.

Example 2.7.11. Let $f(x) = \tan(x)$ with $\text{dom}(f) = \mathbb{R} - \{n\pi + \pi/2 \mid n \in \mathbb{Z}\}$. 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 $x = \pm \pi/2$ you can see that the inverse tangent has horizontal asymptotes at $y = \pm \pi/2$. This illustrates a general pattern, vertical asymptotes for a function will morph into horizontal asymptotes for the inverse function. We will make use of this example in later chapters. It helps us understand what the limit of $\tan^{-1}(x)$ is as $x \rightarrow \infty$ (its $\pi/2$).

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.






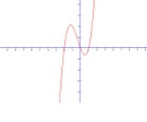
You can draw the graph $y = e^x$ (red) then draw the line $y = x$ (green) and a bunch of perpendicular bisectors (cyan) then the graph of the inverse function $y = \ln(x)$ follows. If we travel one unit from the red graph to the green line along the cyan line then the corresponding point on the blue 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.

2.8 elementary functions

The functions we discuss in this section are the most common functions used in calculus. We can model a great variety of phenomena with these functions.

2.8.1 polynomial functions

We say p is a **polynomial function** of degree n if it has the form $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ where $a_n \neq 0$ and we call $a_n, a_{n-1}, \dots, a_0 \in \mathbb{R}$ the **coefficients** of the polynomial. where 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.	

2.8.2 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**.

2.8.3 rational functions

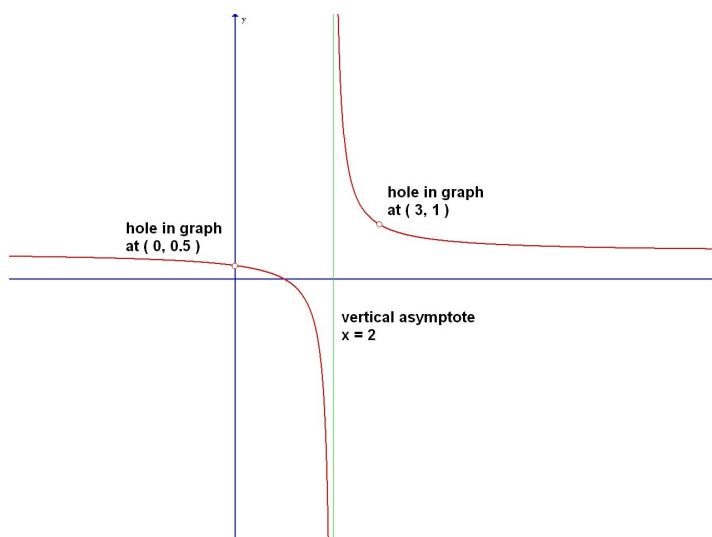
We say that f is a **rational function** if it has the form $f(x) = p(x)/q(x)$ for a pair of polynomial functions p and q . The zeros of f occur at the zeros of p if anywhere. However, it is possible that a zero of p is also a zero of q in which case the point could be a zero, a hole in the graph or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

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

The reciprocal function is a rational function. 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 $(1, 0)$.



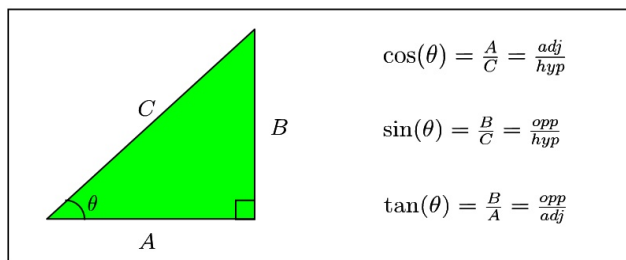
Note, $\text{dom}(f) = (-\infty, 0) \cup (0, 2) \cup (2, 3) \cup (3, \infty) = \mathbb{R} - \{0, 2, 3\}$. Can you tell me the formula for a function that agrees with f on $\mathbb{R} - \{2\}$ but has no holes? Its not a hard question (that function is often called the **reduced function** in precalculus)

2.8.4 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 all the preceding examples in 1,2 and 3. The domain for an algebraic function is simply all the inputs which result in a real number output. That means we must avoid taking the square root of a negative number and also division by zero. A silly example of an algebraic function is $f(x) = \sqrt{\sqrt{x}} - \sqrt{x}$. What is the difference between this function and $g(x) = 0$? Ill give you a clue, its just the domain that is different.

2.8.5 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**.

Theorem 2.8.1. *Pythagorean Theorem*

Let A, B, C be the sides of a right triangle with hypotenuse C then $A^2 + B^2 = C^2$.

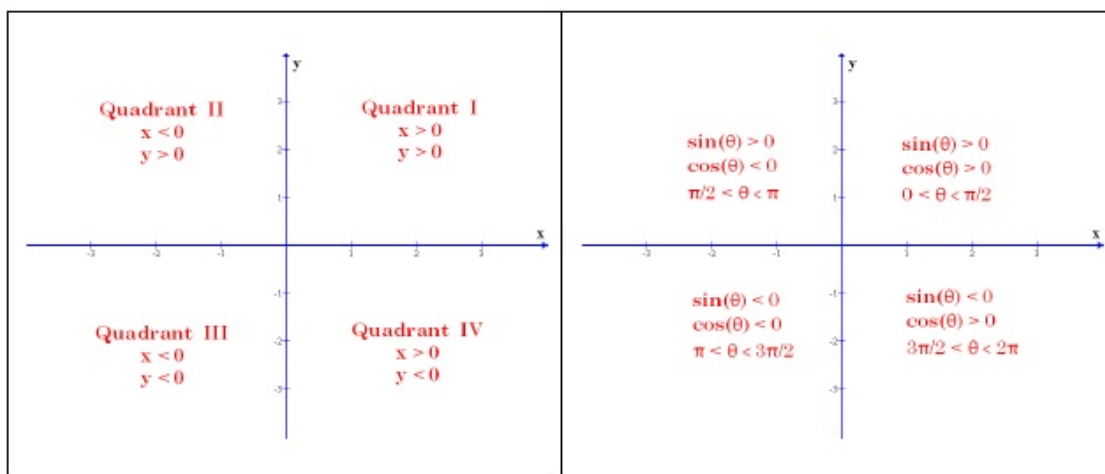
We could go on and list many more facts that are known about triangles and the geometric ratios of sine, cosine and tangent. Instead, I now introduce you to the functions which are defined for any value of θ .

notation	name	Zeros	graph
$\sin(x)$	sine	$x = 0, \pm\pi, \pm2\pi, \dots$ Equivalently, $x = n\pi, n \in \mathbb{Z}$	
$\cos(x)$	cosine	$x = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \dots$ 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.	

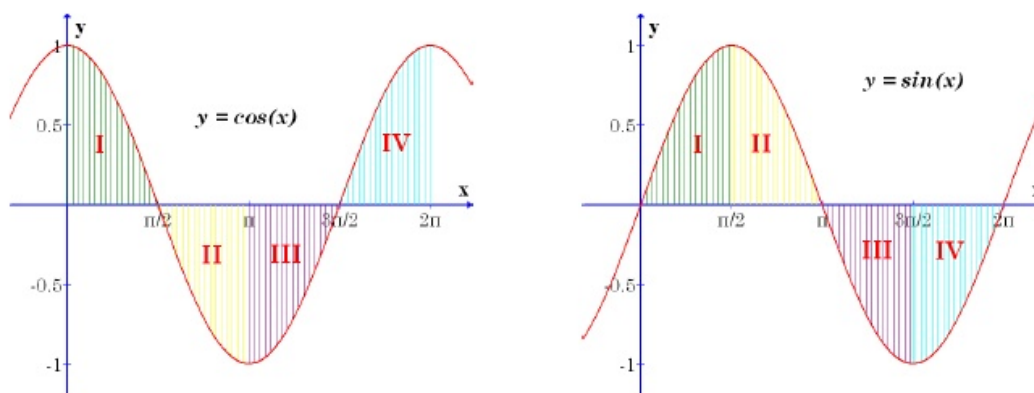
These functions extend the quadrant I geometric quantities to the other three quadrants. The definitions also make **polar coordinates** work. The polar coordinates of $P = (x, y)$ are r, θ where

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

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 usually¹⁰ 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 $\cos(\theta)$.



Or perhaps the following diagrams make more sense to you,



¹⁰admittedly, 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.

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.

Remark 2.8.2.

Calculator Warning: Given the Cartesian coordinates of a point it is a common task to find the standard angle θ , we can solve $\tan(\theta) = \frac{y}{x}$ for θ by taking the inverse tangent to obtain $\theta = \tan^{-1}(\frac{y}{x})$. Let me explain some of the dangers of this formula. Notice that $\tan(\theta)$ is positive in quadrants I and III and is negative in quadrants II and IV. If you try to solve for with a calculator it cannot detect the difference between I and III or II and IV. Lets see how the formula is ambiguous if you are not careful,

i.) Suppose $x = 1, y = 1$ then $\tan(\theta) = 1/1 = 1$. We can solve for θ by taking the inverse tangent of both sides, $\tan^{-1}(\tan(\theta)) = \theta = \tan^{-1}(1)$ now most scientific calculators will calculate the inverse tangent to be $\tan^{-1}(1) = \pi/4$. In this case the calculator has not misled, the standard angle is $\theta = \pi/4$.

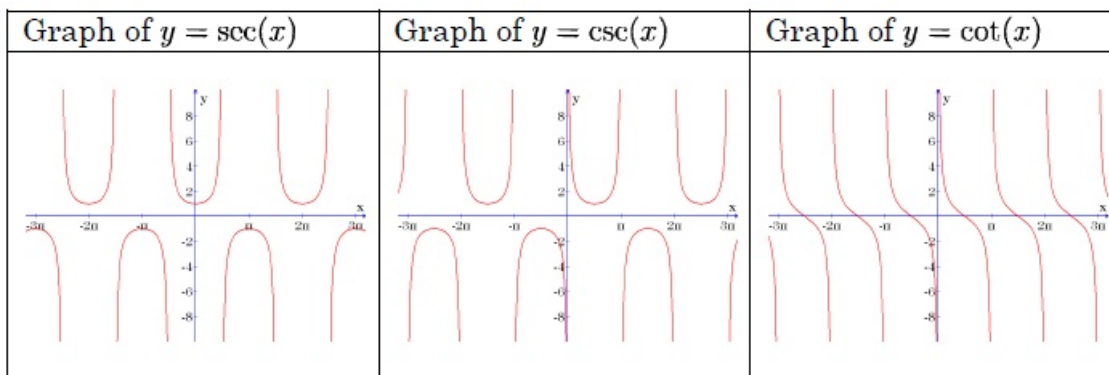
ii.) Suppose $x = -1, y = -1$ then $\tan(\theta) = -1/-1 = 1$. We can solve for θ by taking the inverse tangent of both sides, $\tan^{-1}(\tan(\theta)) = \theta = \tan^{-1}(1)$ and again the calculator will calculate the inverse tangent to be $\tan^{-1}(1) = \pi/4$. In this case the calculator might mislead us, the standard angle is not $\theta = \pi/4$. In fact the standard angle here lies in quadrant III and so we have to add π to the angle the calculator found to get the correct angle of $\theta = 5\pi/4$.

2.8.6 reciprocal trigonometric functions

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

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

The graphs of these functions are given below:



2.8.7 inverse trigonometric functions

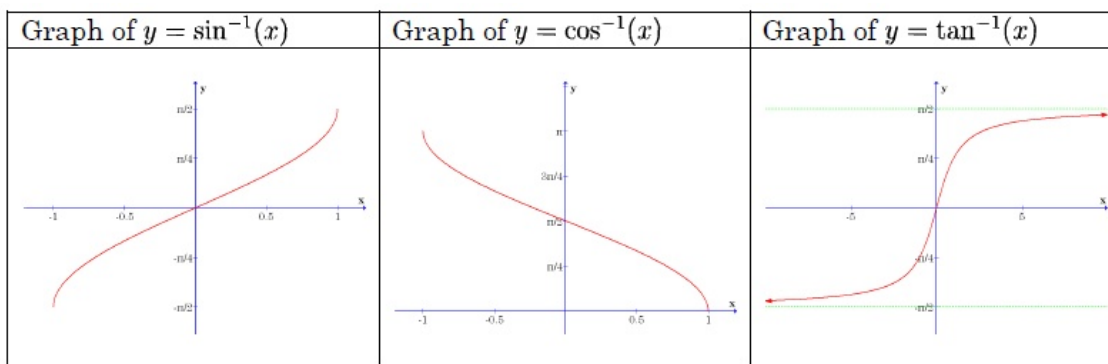
Inverse trigonometric functions: we should be careful to distinguish the inverse trigonometric functions from the reciprocal trig functions. The inverse trig functions are denoted by \sin^{-1} , \cos^{-1} and \tan^{-1} which I refer to as inverse sine, inverse cosine and inverse tangent respectively.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

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

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

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

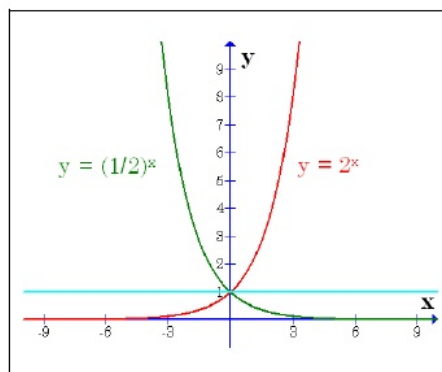
For $x \in [-1, 1]$, $y \in [-1, 1]$ and $z \in \mathbb{R}$. Let us collect the graphs of the inverse trig functions for future reference.



The green lines illustrate horizontal asymptotes of inverse tangent. They 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 $-\pi/2 < \theta < \pi/2$.

2.8.8 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 $a \neq 1$,

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

If $f(x) = e^x$ then this is **the exponential function**, more often than not we will work with this particular base, the number $e \approx 2.71 \dots$ is called **Eulers 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 in some depth in later chapters.

2.8.9 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 were talking out the math) satisfies the following equations,

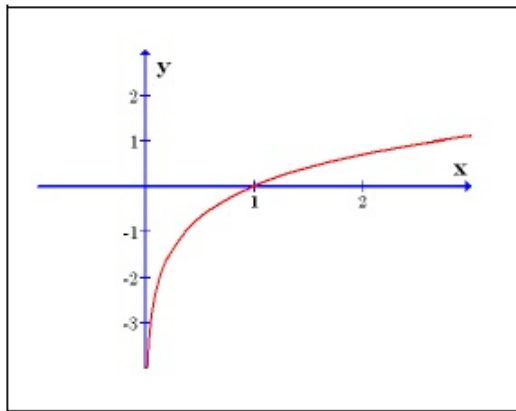
$$\boxed{\log_a(a^x) = x \quad a^{\log_a(x)} = x}$$

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; $\text{dom}(\log_a(x)) = (0, \infty)$.

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

The **natural log function** is denoted $\ln(x)$, this the logarithmic function with base $e = 2.71 \dots$ 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 $\text{dom}(\ln(x)) = (0, \infty)$ and the $\text{range}(\ln(x)) = (-\infty, \infty)$.

The following table is full of things you should already know.

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.

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

2.8.10 hyperbolic functions

Hyperbolic functions: these are little less common than 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:** $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$,

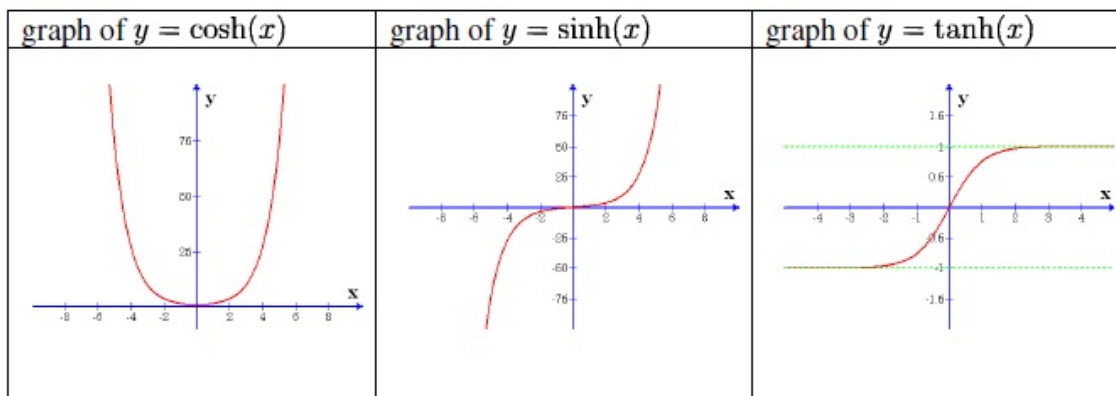
2. **hyperbolic sine:** $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$,

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

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

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.



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

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

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

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

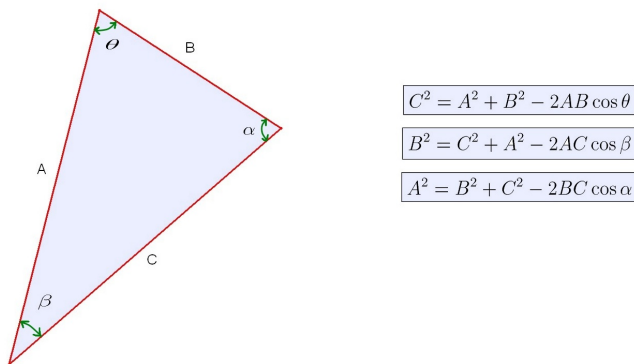
for $x \in [0, \infty)$, $y \in \mathbb{R}$ and $z \in \mathbb{R}$. The hyperbolic sine and tangent functions are injective so they have a global inverse. In contrast, the hyperbolic cosine is not injective and it is customary to let $\cosh^{-1}(x)$ denote the local inverse for hyperbolic cosine restricted to $[0, \infty)$.

2.9 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

$$C^2 = A^2 + B^2 - 2AB \cos \theta$$

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.



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**

$$\cos(\theta + \beta) = \cos \theta \cos \beta - \sin \theta \sin \beta$$

$$\sin(\theta + \beta) = \sin \theta \cos \beta + \cos \theta \sin \beta$$

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 2.9.1.

$$\begin{aligned} \tan(\theta + \beta) &= \frac{\sin(\theta + \beta)}{\cos(\theta + \beta)} \\ &= \frac{\sin \theta \cos \beta + \cos \theta \sin \beta}{\cos \theta \cos \beta - \sin \theta \sin \beta} \\ &= \frac{\frac{\sin \theta \cos \beta}{\cos \theta \cos \beta} + \frac{\cos \theta \sin \beta}{\cos \theta \cos \beta}}{\frac{\cos \theta \cos \beta}{\cos \theta \cos \beta} - \frac{\sin \theta \sin \beta}{\cos \theta \cos \beta}} \Rightarrow \tan(\theta + \beta) = \frac{\tan \theta + \tan \beta}{1 - \tan \theta \tan \beta} \end{aligned}$$

While we are on this example, note if $\theta = \beta$ then we find

$$\tan(2\theta) = \frac{2 \tan \theta}{1 - \tan^2 \theta}$$

Example 2.9.2. The case $\theta = \beta$ gives interesting formulas for sine and cosine,

$$\cos(\theta + \theta) = \cos \theta \cos \theta - \sin \theta \sin \theta \Rightarrow \boxed{\cos(2\theta) = \cos^2 \theta - \sin^2 \theta.}$$

Likewise,

$$\sin(\theta + \theta) = \sin \theta \cos \theta + \cos \theta \sin \theta \Rightarrow \boxed{\sin(2\theta) = 2 \sin \theta \cos \theta.}$$

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

$$\boxed{\cos^2 \theta = \frac{1}{2} \left(1 + \cos(2\theta) \right)}$$

Similarly we can solve for $\sin^2 \theta$ to obtain,

$$\boxed{\sin^2 \theta = \frac{1}{2} \left(1 - \cos(2\theta) \right)}$$

Naturally, we can continue in this fashion to derive a great variety of trigonometric identities. 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.

2.9.1 complex numbers

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 number¹¹? 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

$$\boxed{(a + ib)(c + id) = ac + iad + ibc + i^2 bd = ac - bd + i(ad + bc)}$$

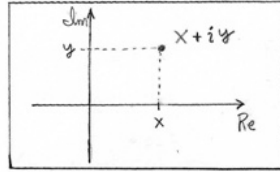
Where we denoted $i = (0, 1)$ hence $i^2 = (0, 1) * (0, 1) = -(1, 0)$ and since $(1, 0) * (a, b) = (a, b)$ we denote $(1, 0) = 1$ hence the relation $i^2 = -1$. In fact, that was the whole reason to define this funny multiplication $*$, Gauss wanted a formal system to construct a number with the property $i^2 = -1$. This number i was termed "imaginary" since it didn't fall into the category of the real numbers, it has different properties. You don't need to remember $a + ib = (a, b)$ for most questions and it is doubtful you'll see it again until you take the complex variables course. I include this somewhat nonstandard material here to drive home the point that there is nothing at all imaginary about complex numbers.

¹¹the 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

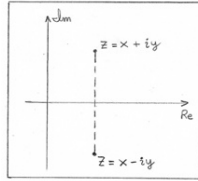
Complex numbers can be added, subtracted, multiplied and divided just the same as real numbers. Complex numbers have a real and imaginary part,

$$\boxed{Re(a, b) = Re(a + ib) = a \quad Im(a, b) = Im(a + ib) = b}$$

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 $z = x - iy$. In the complex plane the mapping $z \rightarrow z^*$ is a reflection across the x-axis.



Gauss proved that any polynomial with real-coefficients can be completely factored over the complex numbers (his thesis work in about 1800). For example, we usually say that $x^2 + 1$ is an irreducible quadratic. This is true with respect to real numbers, however if we use complex numbers to assist with the factorization then we can factor $x^2 + 1 = (x + i)(x - i)$. Generally, a quadratic polynomial $ax^2 + bx + c$ with $b^2 - 4ac < 0$ is called **irreducible** because we cannot factor it over the real numbers. Notice that the quadratic formula still makes sense in this case it just gives complex solutions. We can pull¹² an $i = \sqrt{-1}$ out of the square root;

$$\sqrt{b^2 - 4ac} = \sqrt{(-1)(4ac - b^2)} = \sqrt{-1}\sqrt{4ac - b^2} = i\sqrt{4ac - b^2}$$

where the quantity $\sqrt{4ac - b^2} \in \mathbb{R}$ since $4ac - b^2 > 0$. If $ax^2 + bx + c = 0$ then it can be shown,

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

where I have defined $Re(x) = \alpha = -\frac{b}{2a}$ and $\beta = \frac{\sqrt{4ac - b^2}}{2a}$. The quadratic polynomial factors as follows:

$$ax^2 + bx + c = a[x - (\alpha + i\beta)][x - (\alpha - i\beta)]$$

The roots $\alpha + i\beta$ and $\alpha - i\beta$ form a conjugate pair. Any polynomial with real coefficients can be completely factored with the help of complex numbers. When an irreducible quadratic appears in the factorization it gives rise to a pair of linear factors whose roots form a conjugate pair.

¹²generally $\sqrt{zw} \neq \sqrt{z}\sqrt{w}$ for $z, w \in \mathbb{C}$. For example, $1 = (-1)(-1)$ but $-1 = i^2 = \sqrt{-1}\sqrt{-1} \neq \sqrt{(-1)(-1)} = \sqrt{1} = 1$. Laws of exponents are subtle in complex variables, rest assured the quadratic calculation is true for reasons I'm not going to expose here.

2.9.2 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 2.9.3. *complex exponential function.*

We define $\exp : \mathbb{C} \rightarrow \mathbb{C}$ by the following formula:

$$\exp(z) = \exp(\operatorname{Re}(z) + i\operatorname{Im}(z)) = e^{\operatorname{Re}(z)} [\cos(\operatorname{Im}(z)) + i \sin(\operatorname{Im}(z))].$$

We can show that this definition yields the following desirable properties:

1. $e^{\operatorname{Re}(z)} = \operatorname{Re}(\exp(z))$
2. $\exp(i\operatorname{Im}(z)) = \cos(\operatorname{Im}(z)) + i \sin(\operatorname{Im}(z))$
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 $e^{\operatorname{Re}(z)}$ 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:

$$\begin{aligned}
 \exp(z + w) &= \exp(x + iy + a + ib) \\
 &= \exp(x + a + i(y + b)) \\
 &= e^{x+a} (\cos(y + b) + i \sin(y + b)) && \text{defn. of complex exp.} \\
 &= e^{x+a} (\cos y \cos b - \sin y \sin b + i[\sin y \cos b + \sin b \cos y]) && \text{adding angles formulas} \\
 &= e^{x+a} (\cos y + i \sin y) (\cos b + i \sin b) && \text{algebra} \\
 &= e^x e^a (\cos y + i \sin y) (\cos b + i \sin b) && \text{law of exponents} \\
 &= e^{x+iy} e^{a+ib} && \text{defn. of complex exp.} \\
 &= \exp(z) \exp(w).
 \end{aligned}$$

To prove (5.) we can make use of (3.) and (4.),

$$\exp(z) \exp(-z) = \exp(z - z) = \exp(0) = 1 \Rightarrow \exp(-z) = \frac{1}{\exp(z)}.$$

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.

Remark 2.9.4.

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.

2.9.3 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 valid¹³

$$x = r \cos \theta \quad y = r \sin \theta$$

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:

$$z = x + iy = r \cos \theta + ir \sin \theta = r(\cos \theta + i \sin \theta) = r e^{i\theta}.$$

¹³my viewpoint, it doesn't have to be yours, there are lots of ways to think about sine and cosine

Definition 2.9.5. *polar form of complex number.*

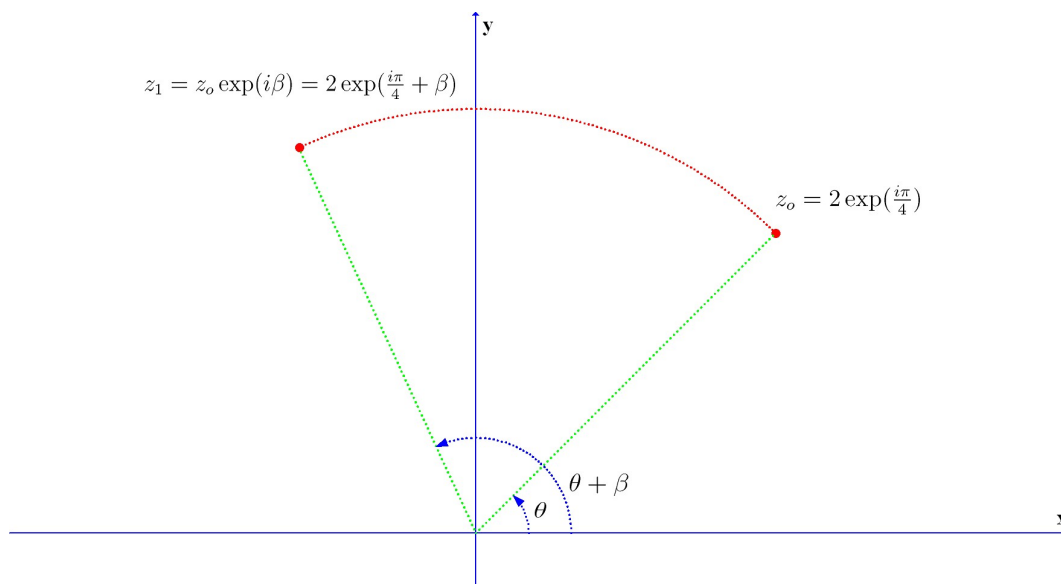
Suppose the Cartesian form of a complex number z is given $z = x + iy$ then the polar form of the complex number is $z = r \exp(i\theta)$ where $r = \sqrt{x^2 + y^2}$ and θ is the standard angle of (x, y) measured counterclockwise from the positive real axis.

Example 2.9.6. 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 2.9.7. Let $z = 2 + 2i$ and multiply by $\exp(i\beta)$. We found the polar form of z in the last example is $z = \sqrt{8} \exp(i \frac{\pi}{4})$.

$$zw = \sqrt{8} \exp(i \frac{\pi}{4}) \exp(i\beta) = \sqrt{8} \exp[i(\frac{\pi}{4} + \beta)]$$

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 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 ECE 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 engineering¹⁴.

¹⁴of course, given all that you might as well add a math major so you have more options later in life.

2.9.4 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,

$$\cos \theta = \frac{1}{2} \left(e^{i\theta} + e^{-i\theta} \right) \quad \sin \theta = \frac{1}{2i} \left(e^{i\theta} - e^{-i\theta} \right).$$

Example 2.9.8. 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 2.9.9. 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 2.9.10. *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:

$$\cos(ax) \cos(bx) = \frac{1}{2} \cos[(a+b)x] + \frac{1}{2} \cos[(a-b)x]$$

$$\sin(ax) \sin(bx) = \frac{1}{2} \cos[(a+b)x] - \frac{1}{2} \cos[(a-b)x]$$

On the other hand, generally $\cos(ax)\sin(bx)$ yields a sum of sines,

$$\cos(ax) \sin(bx) = \frac{1}{2} \sin[(a+b)x] + \frac{1}{2} \sin[(a-b)x]$$

Naturally, we could also apply the method to calculate formulas for higher powers or products of sine and cosine. Just for a flavor:

Example 2.9.11.

$$\begin{aligned}
 \cos^3 \theta &= \left[\frac{1}{2} \left(e^{i\theta} + e^{-i\theta} \right) \right]^3 \\
 &= \frac{1}{8} \left(e^{3i\theta} + 3e^{i\theta} + 3e^{-i\theta} + e^{-3i\theta} \right) \\
 &= \frac{3}{4} \cos(\theta) + \frac{1}{4} \cos(3\theta).
 \end{aligned}$$

DeMoivre's 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.

2.9.5 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. The mathematics of a simple case is encapsulated in the following trigonometric identity:

$$\sin(a) + \sin(b) = 2 \sin\left(\frac{a+b}{2}\right) \cos\left(\frac{a-b}{2}\right)$$

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,

$$y_1 = A_1 \sin[kx - \omega t + \phi_1] \quad \text{and} \quad y_2 = A_2 \sin[kx + \omega t + \phi_2]$$

Consider the superposition of these waves,

$$\begin{aligned} y_1 + y_2 &= A \sin[kx - \omega t + \phi_1] + A \sin[kx + \omega t + \phi_2] \\ &= 2A \sin\left[\frac{(kx - \omega t + \phi_1) + (kx + \omega t + \phi_2)}{2}\right] \cos\left[\frac{(kx - \omega t + \phi_1) - (kx + \omega t + \phi_2)}{2}\right] \\ &= 2A \sin\left[kx + \frac{\phi_1 + \phi_2}{2}\right] \cos\left[\frac{\phi_1 - \phi_2}{2} - \omega t\right] \\ &= 2A \sin\left[kx + \frac{\phi_1 + \phi_2}{2}\right] \cos\left[\omega t - \frac{\phi_1 - \phi_2}{2}\right] \end{aligned}$$

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 seen 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.

What about two arbitrary sine waves or cosine waves of the same wave number but possibly different phases. I'll eliminate the time term to reduce clutter. I think once we solve this problem it's easy to add time to our result.

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 $\tilde{y}_1 = A_1 e^{i(kx + \phi_1)}$ and $\tilde{y}_2 = A_2 e^{i(kx + \phi_2)}$. Notice that these complex functions contain both the sine and cosine functions we wish to add: in particular

$$y_1 = \text{Im}(\tilde{y}_1) = A_1 \sin(kx + \phi_1), \quad y_2 = \text{Im}(\tilde{y}_2) = A_2 \sin(kx + \phi_2)$$

The real parts will give us cosines instead. We can calculate the sum of the sine functions by instead adding

the corresponding complex functions,

$$\begin{aligned}
 \tilde{y}_1 + \tilde{y}_2 &= A_1 e^{i(kx + \phi_1)} + A_2 e^{i(kx + \phi_2)} \\
 &= A_1 e^{ikx} e^{i\phi_1} + A_2 e^{ikx} e^{i\phi_2} \\
 &= [A_1 e^{i\phi_1} + A_2 e^{i\phi_2}] e^{ikx} \\
 &= A e^{i\gamma} e^{ikx} && (A \text{ and } \gamma \text{ from picture}) \\
 &= A e^{i[kx + \gamma]} \\
 &= A \cos[kx + \gamma] + iA \sin[kx + \gamma]
 \end{aligned}$$

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

$$A_1 \sin(kx + \phi_1) + A_2 \sin(kx + \phi_2) = A \sin[kx + \gamma]$$

$$A_1 \cos(kx + \phi_1) + A_2 \cos(kx + \phi_2) = A \cos[kx + \gamma]$$

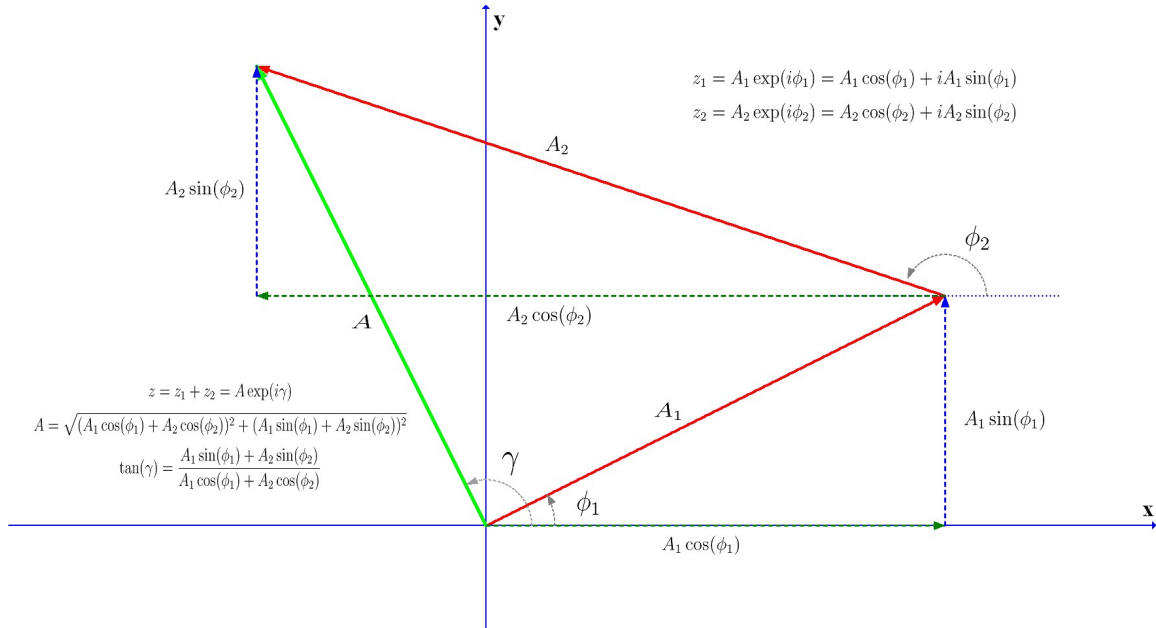
where γ is implicitly defined by

$$\tan \gamma = \frac{A_1 \sin(\phi_1) + A_2 \sin(\phi_2)}{A_1 \cos(\phi_1) + A_2 \cos(\phi_2)}.$$

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.

Theorem 2.9.12. *superposition of waves.*

Two waves with identical wave number will combine when added to give another wave with the same wave number and an amplitude which is between zero and the sum of the individual wave amplitudes.

It's tempting to say more here, but I'll leave it for physics. My intention here was merely to explore a difficult problem in trigonometry to help you push some boundaries of your knowledge. This treatment of waves is neither clever nor comprehensive. If you think about it you could just as well apply the calculation to waves traveling in time in the same direction or opposite directions. You'd recover the boxed formula earlier in this section as a special case. Also, you could consider waves of different frequencies $\omega_1 \neq \omega_2$ interfering, the result is a wave of frequency $\frac{\omega_1 + \omega_2}{2}$ modulated by a *beat frequency* of $\frac{\omega_1 - \omega_2}{2}$. Pictures of these phenomena may be found at:

<http://paws.kettering.edu/~drussell/Demos/superposition/superposition.html>

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. Fourier analysis is heavily calculus-based. In contrast, from what I can gather from a talk I heard last year, the Wavelet method for visual data is primarily linear algebra.

Remark 2.9.13.

The sections marked with a * are optional in the following sense:

1. you could do the analysis using sines and cosines with no mention of complex exponentials,
2. complex exponentials are not on tests except possibly as a bonus question,

I happen to think this material should be integrated in precalculus mathematics and consequently we should weave a certain amount of complex variables throughout the calculus sequence. It's not that hard if you work on it a little. However, most instructors disagree with me on this point so I behave and just write these sections for the most elusive creature: the interested reader.¹⁵

Chapter 3

limits

There are two major questions we must answer in this chapter

1. what is a limit of a function? In other words, how should the limiting process be carefully defined?
2. how do we calculate limits in the most efficient and reliable fashion? If rigor is not required how should we allow intuition to guide our thoughts?

I begin with a few pictures and words to set-up the idea of a limit. Then we give a careful definition. After that we'll build the theory from scratch theorem by theorem. Once the theory is settled then we'll turn our focus to item (2.). Limits are important because all of calculus is traditionally¹ formulated at its foundation by limiting processes. The tangent line is the limit of secant lines, the Riemann integral can be viewed as the limit of an approximating sum. Therefore, to understand calculus properly we will need a good understanding of the limit.

3.1 graphical motivation of limit

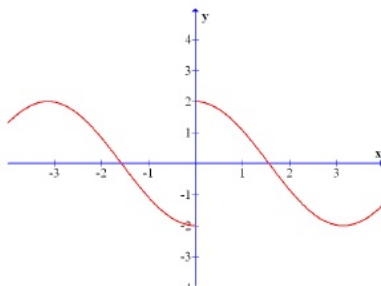
At this point I suspect words are unnecessary, they can only do harm. Instead some pictures should illustrate what is meant by the terms:

1. "taking the limit from the left at $x = a$ "
2. "taking the limit from the right at $x = a$ "
3. "taking the limit at $x = a$ "

¹it may be worth mentioning that there is a formulation of calculus which does not use limits and yet reproduces the same theorems and results. It's called *non-standard analysis*. The text is free online if you're curious.

(I.) The function $f(x)$ graphed below has

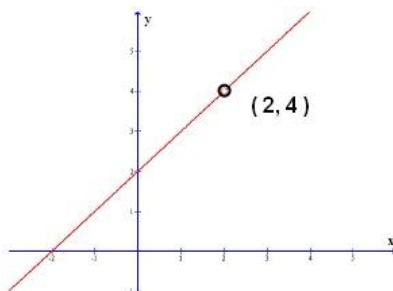
1. " $f(x) \rightarrow 2$ as x approaches 0 from the right"
2. " $f(x) \rightarrow -2$ as x approaches 0 from the left"
3. "the limit of $f(x)$ as x approaches 0 does not exist because the left and right limits at zero do not exist"



It may not be entirely clear to you if the function is defined at $x = 0$ in the picture. It actually would not change the result if both the left and right graphs had open circles at $(0, -2)$ and $(0, 2)$ respectively. The idea of the limit is to look at values of the function near the limit point, but not at the limit point. In this discussion the limit point was $x = 0$.

(II.) The function $f(x)$ graphed below has

1. " $f(x) \rightarrow 4$ as x approaches 2 from the right"
2. " $f(x) \rightarrow 4$ as x approaches 2 from the left"
3. "the limit of $f(x)$ as x approaches 2 does exist and is equal to 4 because the left and right limits both exist and are also equal to 4."

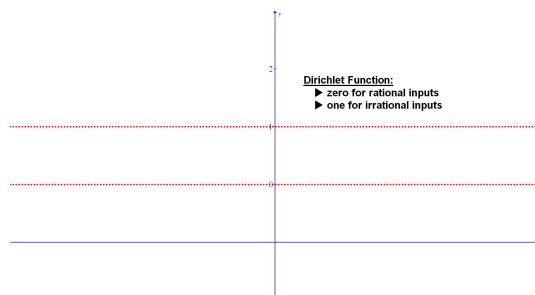


The little black circle is meant to denote the fact that $2 \notin \text{dom}(f)$. We can take the limit at 2 even though $f(2)$ is not defined. To say that $f(x) \rightarrow 4$ as $x \rightarrow 2$ means that as we take values of $f(x)$ close to 4 we can find x near 2 which give those values close to 4.

Warning: you cannot reverse the sentence above. The idea is that as we take values close to the alleged limiting value L we can find inputs near the limit point which returns values close to the limit

However, even my warning fails to capture part of the idea. Let's look at a picture, this is called the **Dirichlet function**;

$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{J} = \mathbb{R} - \mathbb{Q} \end{cases}$. The best I can do to the limit of the resolution is as follows:



This function misbehaves. Don't misunderstand my graph. I don't mean to say that f is not single-valued. It truly is since either $x \in \mathbb{Q}$ or $x \notin \mathbb{Q}$ hence the value of the function is either 1 or 0. Think about $f(x) \rightarrow L$ as $x \rightarrow 0$. Should we expect that $L = 1$ or $L = -1$? Certainly if we look at values close to $L = 1$, for example $L = 1$ we can find x_o close to $x = 0$ such that $f(x_o) = 1$. However, no matter how close x_o is to zero we can always find $x_1 \notin \mathbb{Q}$ which is closer to zero and has $f(x_1) = 0$. Therefore, it's not enough for values of the function near the limit point to take values close to L . We need the values "close" to the limit point to be reached by **all** the points near to the limit point. Perhaps you can start to see we need to use the concept of a set to properly understand the limit.

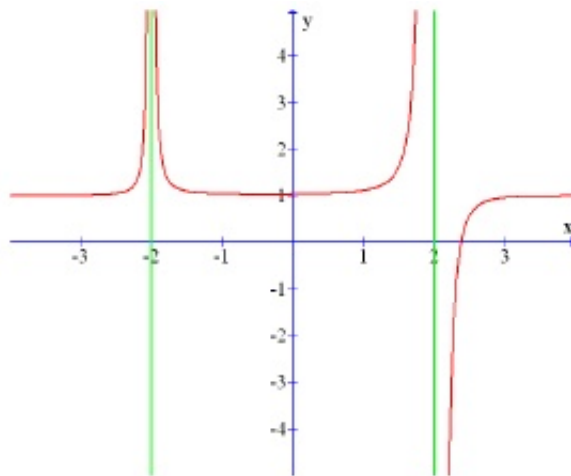
The terms ∞ or $-\infty$ are simply nice notation for expressing a certain type of behavior in a graph. I introduce them here because they help us discuss limits. The examples up to now all involved *jump-discontinuities* or *removable-discontinuities*. The example below has vertical asymptotes which spoil the limits in question.

(III.) The function $f(x)$ graphed below has

1. " $f(x) \rightarrow \infty$ as x approaches 2 from the right"
2. " $f(x) \rightarrow -\infty$ as x approaches 2 from the left"
3. "the limit of $f(x)$ as x approaches 2 does not exist because the left and right limits do not agree."

But, in contrast,

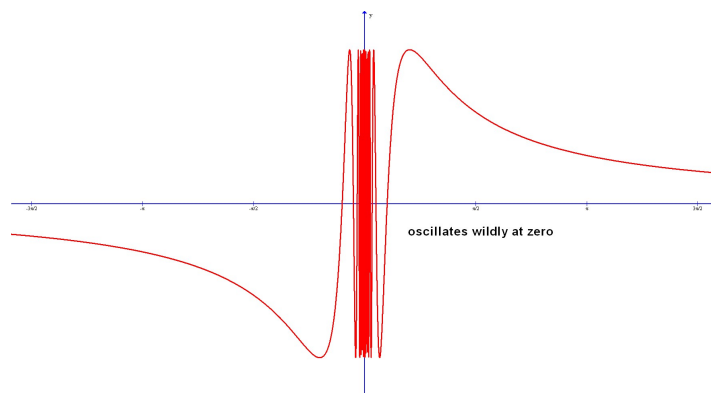
1. " $f(x) \rightarrow \infty$ as x approaches -2 from the right"
2. " $f(x) \rightarrow \infty$ as x approaches -2 from the left"
3. "the limit of $f(x)$ as x approaches -2 does not exist in \mathbb{R} , and it is ∞ because the left and right limits both are ∞ ."



When we say "does not exist in \mathbb{R} " we mean just that; ∞ is not a number. When a limit is "equal to infinity" that is simply a way of communicating that the limit does not exist as a real number and it does so in a particular manner. We saw before there are other ways the limit may fail to exist. Let us examine one more misbehaving function.

(IV.) The function $f(x)$ graphed below has

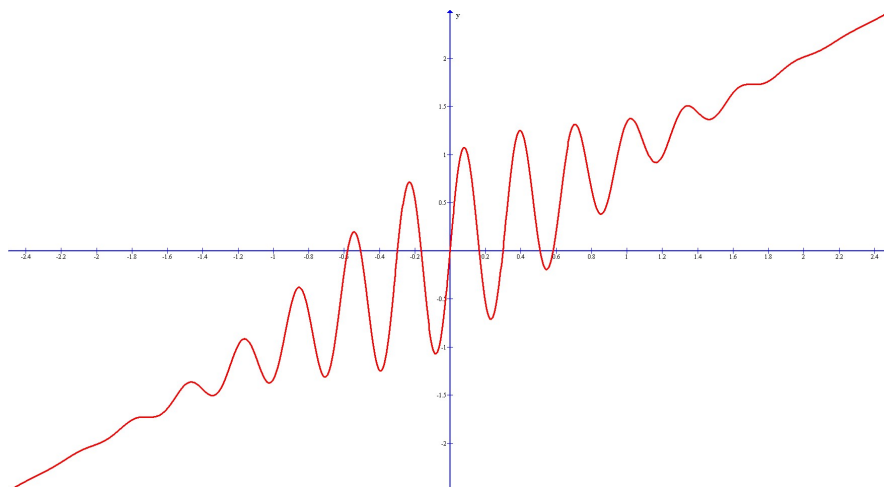
1. "the limit of $f(x)$ as x approaches 0 from the right does not exist due to oscillation at the limit point"
2. "the limit of $f(x)$ as x approaches 0 from the left does not exist due to oscillation at the limit point"
3. "the limit of $f(x)$ as x approaches 0 does not exist due to oscillation at the limit point"



You could also imagine an example where we have oscillation just on the left or just on the right. The examples given thus far should serve to illustrate the typical ways which limits either exist or d.n.e. as real numbers. Let me conclude this section with a less exciting, but far more common situation,

(V.) The function $f(x)$ graphed below has

1. " $f(x) \rightarrow f(a)$ as x approaches a from the right"
2. " $f(x) \rightarrow f(a)$ as x approaches a from the left"
3. "the limit of $f(x)$ as x approaches a exists and is equal to $f(a)$ because the left and right limits exist and are equal to $f(a)$."



In the graph above any point $a \in \text{dom}(f)$ has values near $f(a)$ which are close to $f(a)$, there are no jumps, asymptotes or oscillations which get bunched up on some point. This function is called *continuous* because it has a connected graph which could be drawn with one uninterrupted stroke of a pen. This graphical definition of continuous is not the one I wish to see on the exam. We will soon offer a better description in terms of limits. But, first we need to settle just what a limit is. Remember, this section was simply to motivate the one that follows. We usually have more than a graph to reason with so we can give better arguments than the ones offered in this section. In fact, to be honest, I have yet to give an argument about a limit. This section was basically just a bunch of name-calling (but in a good way).

3.2 definition of the limit

I am certainly indebted to the excellent text *Calculus* by Apostol. A worthy purchase if you're a math major.

3.2.1 two-sided limit

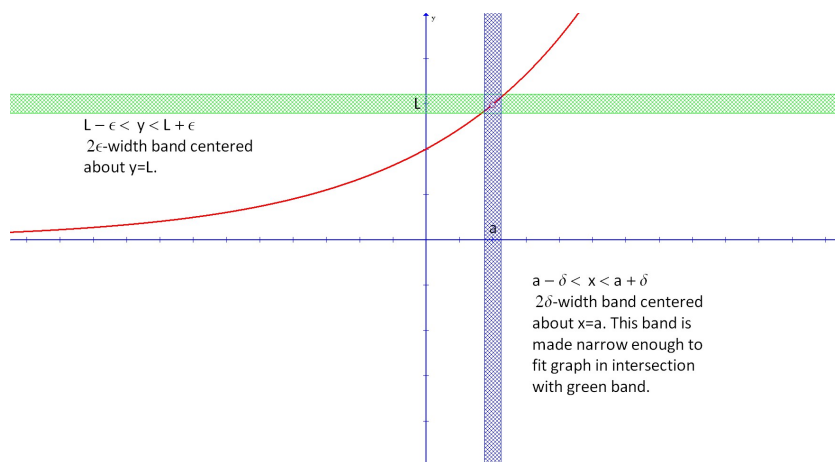
Recall that $B_\delta(a) = \{x \in \mathbb{R} \mid d(a, x) < \delta\} = (a - \delta, a + \delta)$ is a neighborhood² centered at a with radius $\delta > 0$. Likewise, $B_\delta(a)_o = \{x \in \mathbb{R} \mid 0 < d(a, x) < \delta\} = (a - \delta, a) \cup (a, a + \delta)$ is a deleted nbhd. centered at a .

Definition 3.2.1. *limit*

Let f be a function and $a, L \in \mathbb{R}$. We say that $f(x) \rightarrow L$ as $x \rightarrow a$ iff for each nbhd. $B_\epsilon(L)$ centered at L there exists a deleted nbhd. $B_\delta(a)_o$ centered at a such that $f(B_\delta(a)_o) \subseteq B_\epsilon(L)$. In the case that the condition above is met we say that the limit exists and denote this by

$$\lim_{x \rightarrow a} f(x) = L.$$

This definition emphasizes the geometry of the limit. To understand this, meditate on the following diagram:



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 $(a - \delta, a) \cup (a, a + \delta)$. Pragmatically, we would like an easier formulation of the limit to prove theorems and solve problems. For that end we restate the definition in terms of inequalities and absolute values. I invite the reader to verify this is nothing more than a change of notation from Definition 3.2.1.

²clearly an abbreviation is warranted here, we'll use nbhd. from here on out

Definition 3.2.2. *limit*

Let f be a function and $a, L \in \mathbb{R}$. We say that $f(x) \rightarrow L$ as $x \rightarrow a$ iff for each $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$ it follows $|f(x) - L| < \epsilon$. In the case that the condition above is met we say that the limit exists and denote this by

$$\lim_{x \rightarrow a} f(x) = L.$$

Sometimes the condition that must be met by the function is instead stated:

$$|f(x) - L| < \epsilon \text{ whenever } 0 < |x - a| < \delta$$

This is equivalent to the language in my definition provided the reader realizes that we must have this condition hold for **all** $x \in \mathbb{R}$ that meet the condition $0 < |x - a| < \delta$. If even one value fails then you have to find a better δ or perhaps give up and prove the limit does not exist³

We will eventually have theorems which do the calculations that follow in this section with ease. However, do not ignore the calculations that follow. This material is challenging and required. Most students have to really do some thinking to become proficient in the arguments which are offered in the upcoming examples.

Example 3.2.3. Problem: prove $\lim_{x \rightarrow 2} (3x + 2) = 8$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 2| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = 3x + 2$ and $L = 8$ and a particular choice of δ . Consider then

$$|f(x) - L| = |3x + 2 - 8| = |3x - 6| = |3(x - 2)| = 3|x - 2| < 3\delta = \epsilon.$$

So, we should choose $\delta = \epsilon/3$ since $\epsilon > 0$ it is clear that $\delta = \epsilon/3 > 0$. In view of these calculations we are ready to state the proof.

Proof: Let $\epsilon > 0$ and choose $\delta = \epsilon/3$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 2| < \delta$. Observe that

$$|3x + 2 - 8| = |3(x - 2)| = 3|x - 2| < 3\delta = \epsilon.$$

Thus $0 < |x - 2| < \delta$ implies $|3x + 2 - 8| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow 2} (3x + 2) = 8$. \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. However, in the context of this course there is less ambiguity since I am giving you examples which show you how much detail is required.

³remember, "exist" in this context means "exist as a real number", it certainly could be argued that all limits exist in the sense of being a concept held by many rational beings.

Example 3.2.4. Problem: prove $\lim_{x \rightarrow 3}(2 - x) = -1$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 3| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = 2 - x$ and $L = -1$ and a particular choice of δ . Consider then

$$|f(x) - L| = |2 - x - (-1)| = |-x + 3| = |-1(x - 3)| = |x - 3| < \delta = \epsilon.$$

So, we should choose $\delta = \epsilon$.

Proof: Let $\epsilon > 0$ and choose $\delta = \epsilon$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 3| < \delta$. Observe that

$$|2 - x - (-1)| = |-x + 3| = |x - 3| < \delta = \epsilon.$$

Thus $0 < |x - 3| < \delta$ implies $|2 - x - (-1)| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow 3}(2 - x) = -1$. \square

Example 3.2.5. Problem: prove $\lim_{x \rightarrow 0}(x^2) = 0$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 0| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = x^2$ and $L = 0$ and a particular choice of δ . Consider then

$$|f(x) - L| = |x^2 - 0| = |x|^2 < \delta^2 = \epsilon.$$

So, we should choose $\delta = \sqrt{\epsilon}$. Since $\epsilon > 0$ we can be assured that the squareroot gives $\delta > 0$.

Proof: Let $\epsilon > 0$ and choose $\delta = \sqrt{\epsilon}$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. Observe that

$$|x^2 - 0| = |x|^2 < (\sqrt{\epsilon})^2 = \epsilon.$$

Thus $0 < |x - 0| < \delta$ implies $|x^2 - 0| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow 0}(x^2) = 0$. \square

Example 3.2.6. Problem: prove $\lim_{x \rightarrow 3}(x^2) = 9$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 3| < \delta$ implies $|f(x) - L| < \epsilon$ 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 $\delta = \epsilon/7$. However, we do need that $\epsilon/7 < 1$, remember we don't control ϵ , all we know is that $\epsilon > 0$. The solution is simple, to be careful about the possibility of large ϵ we choose $\delta = \min(\epsilon/7, 1)$. If $\delta = 1$ then we still find $|x + 3| \leq 7$ and so $|f(x) - L| \leq 7\delta < \epsilon$ provide that $\delta = \min(\epsilon/7, 1)$ so we knew $\delta < \epsilon/7$ hence $7\delta < \epsilon$.

Proof: Let $\epsilon > 0$ and choose $\delta = \min(\epsilon/7, 1)$. 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 $\delta \leq \epsilon/7$ we have $7\delta \leq \epsilon$. Thus, $0 < |x - 3| < \delta$ implies that $|x^2 - 9| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow 3}(x^2) = 9$. \square

Example 3.2.7. Problem: prove $\lim_{x \rightarrow 5} |x| = 5$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 5| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = |x|$ and $L = 5$ and a particular choice of δ . I find that $||x| - 5|$ is best covered by forcing $\delta < 1$ for convenience: notice that $\delta < 1$ and the assumption $0 < |x - 5| < \delta$ yields $-1 < -\delta < x - 5 < \delta < 1$ hence $4 < x < 6$ which is nice because that means that x is positive so $|x| = x$. Suppose $\delta < 1$ and $0 < |x - 5| < \delta$ it follows that

$$||x| - 5| = |x - 5| < \delta.$$

we can choose $\epsilon = \delta$. However, you might worry, what if $\epsilon > 1$? In that case we can just choose $\delta = 1$ then $||x| - 5| = |x - 5| < \delta = 1 < \epsilon$. So, to take care of both cases we should simply choose $\delta = \min(1, \epsilon)$. The notation "min" means to take the minimum of the values.

Proof: Let $\epsilon > 0$ and choose $\delta = \min(1, \epsilon)$. Suppose $x \in \mathbb{R}$ such that $0 < |x - 5| < \delta$. Observe that we have $\delta < 1$ hence $0 < |x - 5| < \delta < 1$ yields $-1 < x - 5 < 1$ hence $4 < x < 6$ so $|x| = x$. Consider then,

$$||x| - 5| = |x - 5| < \delta \leq \epsilon.$$

Thus $0 < |x - 5| < \delta$ implies $||x| - 5| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow 5} |x| = 5$. \square

Example 3.2.8. Problem: Let $f(x) = x|x - 2|$, prove $\lim_{x \rightarrow 3} f(x) = 3$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - 3| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = x|x - 2|$ and $L = 3$ and a particular choice of δ . Consider then

$$\begin{aligned} |f(x) - L| &= |x|x - 2| - 3| \\ &\leq |x|x - 2| + |-3| && \text{triangle inequality} \\ &= |x||x - 2| + 3 && \text{used } |ab| = |a||b| \text{ and } ||a| = |a| \text{ since } |a| \geq 0 \\ &= |x - 3 + 3||x - 3 + 1| + 3 && \text{used } x - 2 = x - 3 + 1 \\ &\leq (|x - 3| + 3)(|x - 3| + 1) + 3 && \text{triangle inequality twice} \\ &= |x - 3|^2 + 4|x - 3| + 6 && \text{just algebra} \\ &= \delta^2 + 4\delta + 6 && \text{supposing } |x - 3| < \delta \end{aligned}$$

So, we should insist that $\delta^2 + 4\delta + 6 \leq \epsilon$. This is a quadratic equation in δ . If I tinker with the 4 I can make it factor. If $\delta > 0$ it is clear that

$$\delta^2 + 4\delta + 6 < \delta^2 + 5\delta + 6 = (\delta + 2)(\delta + 3) < (\delta + 3)^2.$$

We can solve $(\delta + 3)^2 = \epsilon$ for $\delta = \sqrt{\epsilon} - 3$.

Question: why will my choice $\delta = \sqrt{\epsilon} - 3$ fail? The point of this example is to show you that we can make correct steps and not find our way to a correct choice of δ . Maybe you can repair my argument and find a better choice for δ .

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 3.2.9. Problem: prove $\lim_{x \rightarrow a} (3x + 2) = 3a + 2$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - a| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = 3x + 2$ and $L = 3a + 2$ and a particular choice of δ . Consider then

$$|f(x) - L| = |3x + 2 - (3a + 2)| = |3(x - a)| = 3|x - a| < 3\delta = \epsilon.$$

So, we should choose $\delta = \epsilon/3$ since $\epsilon > 0$ it is clear that $\delta = \epsilon/3 > 0$. In view of these calculations we are ready to state the proof.

Proof: Let $\epsilon > 0$ and choose $\delta = \epsilon/3$. Suppose $x \in \mathbb{R}$ such that $0 < |x - a| < \delta$. Observe that

$$|3x + 2 - (3a + 2)| = |3(x - a)| = 3|x - a| < 3\delta = \epsilon.$$

Thus $0 < |x - a| < \delta$ implies $|3x + 2 - (3a + 2)| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow a} (3x + 2) = 3a + 2$. \square

The preceding example was no harder with arbitrary a than it was with $a = 2$ in Example 3.2.3. In contrast, we'll have to think a bit more in the next example. The arguments given in Example 3.2.7 will need some tweaking.

Example 3.2.10. Problem: prove $\lim_{x \rightarrow x_o} |x| = |x_o|$ directly by the definition of the limit.

Preparatory calculations: We need to show that $|x - x_o| < \delta$ implies $|f(x) - L| < \epsilon$ for $f(x) = |x|$ and $L = |x_o|$ and a particular choice of δ . I find that $||x| - |x_o||$ is best covered by forcing $\delta < 1$ for convenience: notice that $\delta < 1$ and the assumption $0 < |x - x_o| < \delta$ yields $-1 < -\delta < x - x_o < \delta < 1$ hence $x_o - 1 < x < x_o + 1$. Before we were able to conclude that x is positive so $|x| = x$, but $x_o = -2$ is possible now and in that case $x < 0$. We need to break-up into cases:

1. if $x_o \geq 1$ then $x_o - 1 \geq 0$ and $0 \leq x_o - 1 < x < x_o + 1$ yields $x \geq 0$ hence $|x| = x$
2. if $x_o \leq -1$ then $x_o + 1 \leq 0$ then $x_o - 1 < x \leq x_o + 1 < 0$ yields $x \leq 0$ hence $|x| = -x$
3. if $-1 < x_o < 1$ then $x_o - 1 < 0$ whereas $0 < x_o + 1$ thus $x_o - 1 < x < x_o + 1$ has no solutions since we cannot have both $x > 0$ and $x < 0$. In this case we cannot assume $\delta < 1$. Apparently some other argument is needed here.

Case (1.) (assume $x_o \geq 1$) follows the same pattern as in Example 3.2.7, we can simply use $\delta = \min(1, \epsilon)$ and as $x_o - 1 \geq 0$ and $0 \leq x_o - 1 < x < x_o + 1$ yields $x \geq 0$ hence $|x| = x$ therefore we'll find $||x| - |x_o|| = |x - x_o| < \delta < \epsilon$.

Case (2.) (assume $x_o \leq -1$) also follows a similar argument. Use $\delta = \min(1, \epsilon)$ again and note $x_o + 1 \leq 0$ then $x_o - 1 < x \leq x_o + 1 \leq 0$ yields $x \leq 0$ hence $|x| = -x$ therefore we'll find $||x| - |x_o|| = |-x - (-x_o)| =$

$$|x - x_o| < \delta < \epsilon.$$

Case (3.)(assume $-1 < x_o < 1$) left to reader.

Proof: partly left to reader, I'll cover cases 1 and 2. Suppose $|x_o| \geq 1$. Let $\epsilon > 0$ and choose $\delta = \min(1, \epsilon)$. Suppose $x \in \mathbb{R}$ and $0 < |x - x_o| < \delta$. Since $|x_o| \geq 1$ it follows that either $x_o \geq 1$ or $x_o \leq -1$. We treat each case separately:

1. if $x_o \geq 1$ then $|x_o| = x_o$. Notice that $\delta \leq 1$ and the assumption $0 < |x - x_o| < \delta$ yields $-1 < -\delta < x - x_o < \delta < 1$ hence $x_o - 1 < x < x_o + 1$ and as $0 \leq x_o - 1$ we find $0 \leq x$ so $|x| = x$. Thus,

$$||x| - |x_o|| = |x - x_o| < \delta \leq \epsilon.$$

2. if $x_o \leq -1$ then $|x_o| = -x_o$. Notice that $\delta \leq 1$ and the assumption $0 < |x - x_o| < \delta$ yields $-1 < -\delta < x - x_o < \delta < 1$ hence $x_o - 1 < x < x_o + 1$ and as $x_o + 1 \leq 0$ we find $x \leq 0$ so $|x| = -x$. Thus,

$$||x| - |x_o|| = |-x - (-x_o)| = |x - x_o| < \delta \leq \epsilon.$$

Therefore we have shown in the case $|x_o| < 1$ that for each $\epsilon > 0$ we can choose $\delta > 0$ such that $0 < |x - x_o| < \delta$ implies $||x| - |x_o|| < \epsilon$ and it follows by the definition of the limit that $\lim_{x \rightarrow a} |x| = x_o$. \square

3.2.2 one-sided limits

If you examine the definition of limit in the preceding section you'll notice it doesn't make much sense for boundary points of the $\text{dom}(f)$. We say $p \in \mathbb{R}$ is a **boundary point** of $\text{dom}(f)$ iff every deleted open interval centered at p intersects points in $\mathbb{R} - \text{dom}(f)$ and $\text{dom}(f)$. In other words, boundary points are positioned so that they are close to points both inside and outside $\text{dom}(f)$. We can define one-sided limits at boundary points.

Definition 3.2.11. *limit*

Let f be a function and $a, L \in \mathbb{R}$. We say that $f(x) \rightarrow L$ as $x \rightarrow a^+$ iff for each $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ with $a < x < a + \delta$ it follows $|f(x) - L| < \epsilon$. In the case that the condition above is met we say that the **right limit** exists and denote this by

$$\lim_{x \rightarrow a^+} f(x) = L.$$

Likewise, we say that $f(x) \rightarrow L$ as $x \rightarrow a^-$ iff for each $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{R}$ with $a - \delta < x < a$ it follows $|f(x) - L| < \epsilon$. In the case that the condition above is met we say that the **left limit** exists and denote this by

$$\lim_{x \rightarrow a^-} f(x) = L.$$

The logic is very similar to the two-sided examples. I'll just do this example.

Example 3.2.12. Problem: prove $\lim_{x \rightarrow 1^+} (\sqrt{x-1}) = 0$ directly by the definition of the limit.

Preparatory calculations: We need to show that $1 < x < 1 + \delta$ implies $|f(x) - L| < \epsilon$ 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 $\delta = \epsilon^2$.

Proof: Let $\epsilon > 0$ and choose $\delta = \epsilon^2$. Suppose $x \in \mathbb{R}$ such that $0 < x - 1 < \delta$. Observe that

$$|\sqrt{x-1}| = \sqrt{|x-1|} < \sqrt{\delta} = \sqrt{\epsilon^2} = \epsilon.$$

Thus $0 < x - 1 < \delta$ implies $|\sqrt{x-1}| < \epsilon$ and it follows by the definition of the right-sided limit that $\lim_{x \rightarrow 1^+} \sqrt{x-1} = 0$. \square

Notice that $f(x) = \sqrt{x-1}$ has implicit domain $\text{dom}(f) = [1, \infty)$ and $x = 1$ is the boundary point of the domain. We could not consider a two-sided limit at one because the function is not real-valued for $x < 1$.

Proposition 3.2.13. *two-sided limit holds iff both left and right limits hold.*

Let $f : \text{dom}(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be a function and suppose $a \in \text{dom}(f)$ is not a boundary point so both the left and right limits of f can be defined at a .

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

Proof: to prove \Leftrightarrow we must show both \Rightarrow and \Leftarrow .

(\Rightarrow) Begin by assuming $\lim_{x \rightarrow a} f(x) = L$ then for each $\epsilon > 0$ there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - L| < \epsilon$. Note for each $\epsilon > 0$ that if $0 < x - a < \delta$ it follows $0 < |x - a| < \delta$ so $|f(x) - L| < \epsilon$ hence $\lim_{x \rightarrow a^+} f(x) = L$. Likewise, note for each $\epsilon > 0$ that if $-\delta < x - a < 0$ it follows $0 < |x - a| < \delta$ so $|f(x) - L| < \epsilon$ hence $\lim_{x \rightarrow a^-} f(x) = L$.

(\Leftarrow) We assume that both $\lim_{x \rightarrow a^+} f(x) = L$ and $\lim_{x \rightarrow a^-} f(x) = L$. Let $\epsilon > 0$ and choose $\delta = \min(\delta_+, \delta_-)$ where we use the givens to choose $\delta_+, \delta_- > 0$ such that

1. $0 < x - a < \delta_+$ implies $|f(x) - L| < \epsilon$,
2. $-\delta_- < x - a < 0$ implies $|f(x) - L| < \epsilon$

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 $|f(x) - L| < \epsilon$. Therefore, the two-sided limit exists and $f(x) \rightarrow L$ as $x \rightarrow a$. \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 .

Example 3.2.14. Problem: prove $\lim_{x \rightarrow 0} \frac{1}{x} \notin \mathbb{R}$ directly by the definition of the limit.

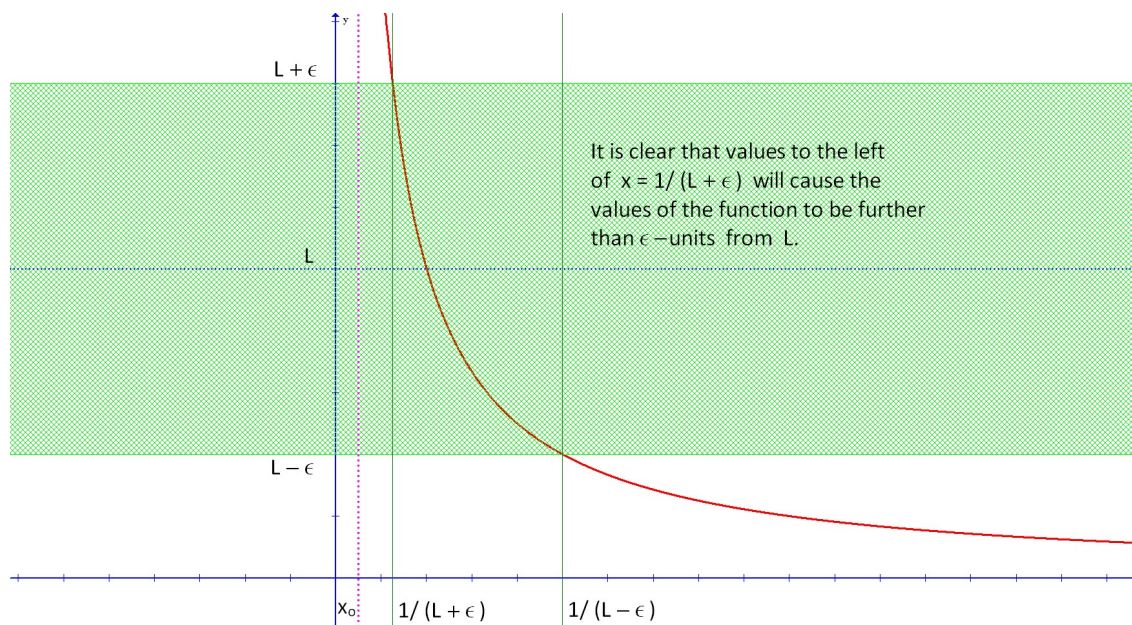
Preparatory calculations: think about it. What do we need to show to show it is impossible for any real number to be the limit of $\frac{1}{x}$ as $x \rightarrow 0$? By the proposition we just proved it would suffice to show that the right-limit failed to exist no matter what our choice of L is. Let's proceed from that angle. We want to show that $\lim_{x \rightarrow 0^+} \frac{1}{x}$ 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 $\lim_{x \rightarrow 0^+} \frac{1}{x} = L$ 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 $\frac{1}{x} \rightarrow L$ as $x \rightarrow 0^+$. This means that for each $\epsilon > 0$ there exists $\delta > 0$ such that $0 < x < \delta$ implies $|\frac{1}{x} - L| < \epsilon$. We seek a contradiction, suppose $\epsilon = L$ and let $\delta > 0$ be some number such that all $x \in \mathbb{R}$ satisfying $0 < x < \delta$ force $|\frac{1}{x} - L| < \epsilon$. Define $x_o = \min(\frac{1}{2(L+\epsilon)}, \frac{\delta}{2})$ thus $x_o \leq \frac{1}{2(L+\epsilon)}$ and $x_o < \delta$. Clearly $0 < x_o < \delta$ so it follows that

$$-\epsilon < \frac{1}{x_o} - L < \epsilon$$

and as $\epsilon = L$ we add ϵ to find $0 < \frac{1}{x_o} < 2\epsilon$. On the other hand, we have constructed x_o to satisfy the inequality $x_o \leq \frac{1}{2(L+\epsilon)} = \frac{1}{4\epsilon}$ thus $\frac{1}{x_o} \geq 4\epsilon$. But, this is a contradiction since we cannot have both $\frac{1}{x_o} < 2\epsilon$ and $\frac{1}{x_o} \geq 4\epsilon$. Therefore, by proof by contradiction, there does not exist such an $L \in \mathbb{R}$ and we conclude that the $\lim_{x \rightarrow 0^+} \frac{1}{x}$ does not exist, hence $\lim_{x \rightarrow 0} \frac{1}{x}$ 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.



3.2.3 divergent limits

Definition 3.2.15. *limit*

Let f be a function and $a \in \mathbb{R}$. We say that $f(x) \rightarrow \infty$ as $x \rightarrow a$ iff 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

$$\lim_{x \rightarrow a} f(x) = \infty.$$

If for each $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ whenever $a < x < a + \delta$ then we say $f(x) \rightarrow \infty$ as $x \rightarrow a^+$.

Likewise, if for each $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ whenever $a - \delta < x < a$ then we say $f(x) \rightarrow \infty$ as $x \rightarrow a^-$.

The definitions of $f(x) \rightarrow -\infty$ as $x \rightarrow a$ or $x \rightarrow a^\pm$ 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 3.2.16. *two-sided limit diverges to $\pm\infty$ iff both left and right limits diverge to $\pm\infty$.*

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

The proof is similar to that given in the last section, the details are left to the reader. Notice that the notation \pm is meant to denote case-wise logic. I mean to state that either both limits are $+\infty$ or both limits are $-\infty$ if they are to match the two-sided limit.

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

Example 3.2.18. Problem: prove $\lim_{x \rightarrow 0^+} \frac{1}{x} = \infty$.

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.

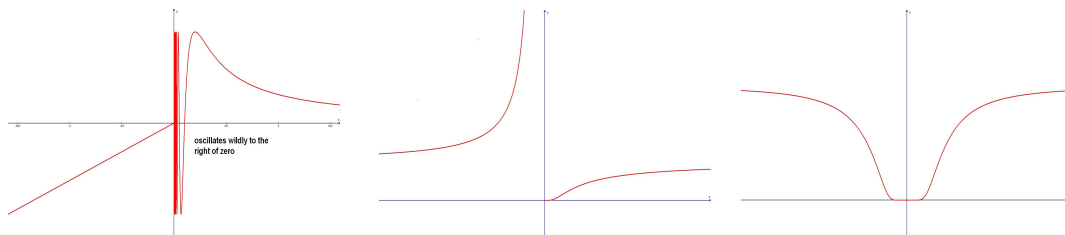
Proof: suppose $M > 0$ and let $\delta = \frac{1}{2M}$. If $0 < x < \delta = \frac{1}{2M}$ then $\frac{1}{x} > 2M > M$. Therefore, for each $M > 0$ there exists $\delta > 0$ such that $\frac{1}{x} > M$ whenever $0 < x < \delta$. It follows by definition that $\lim_{x \rightarrow 0^+} \frac{1}{x} = \infty$. \square

We learned in Example 3.2.14 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 3.2.19.

Another concept of infinity is discussed in the study of **cardinality**. Intuitively speaking the cardinality of a set describes the size of the set. For example, $S = \{1, 2, 3\}$ has cardinality 3. The natural numbers have cardinality \aleph_0 which is infinite. Then the real numbers are even larger, the cardinality 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 $\aleph_0 < c$. However, the idea that the continuum is the *next* infinity past \aleph_0 is called the **continuum hypothesis**.

Here are three other cases we might encounter:



1. the left graph has $f(x) \rightarrow 0$ as $x \rightarrow 0^-$ whereas the right limit fails to exist due to oscillation to the right of zero.
2. the middle has $f(x) \rightarrow \infty$ as $x \rightarrow 0^-$ whereas the $f(x) \rightarrow 0$ as $x \rightarrow 0^+$.
3. the right graph has $f(x) \rightarrow 0$ as $x \rightarrow 0^-$ and also $f(x) \rightarrow 0$ as $x \rightarrow 0^+$. In fact, $f(x) \rightarrow 0$ as $x \rightarrow 0$ since both the left and right limits agree.

Don't mistake this short list for a complete list of possible types of limits. There possibilities are too numerous to list. Generally, we just have to THINK.

3.3 continuity and limit laws

The definition of the limit is important, however, if the only way to calculate limits was with direct ϵ, δ arguments then I doubt the concept of a limit would enjoy much interest from the mathematical community. The properties we discover in this section go to show that the definition given in the last section was the proper, *natural* definition.

You were probably told that a *continuous function* was one for which the graph could be drawn without lifting the pencil. There are no vertical asymptotes, jump-discontinuities or holes in the graph of a continuous function. We are now able to give a precise definition:

Definition 3.3.1. *continuity at a point, on a set, and for a function.*

Suppose $f : \text{dom}(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ and $U \subseteq \text{dom}(f)$. We say that f is **continuous at** $a \in \text{int}(U)$ iff $\lim_{x \rightarrow a} f(x) = f(a)$. If $a \in U$ is a boundary point of U such that $[a, a + \epsilon) \subset U$ for some $\epsilon > 0$ then we say f is **continuous at** a iff $\lim_{x \rightarrow a^+} f(x) = f(a)$. Likewise, if $a \in U$ is a boundary point of U such that $(a - \epsilon, a] \subset U$ for some $\epsilon > 0$ then we say f is **continuous at** a iff $\lim_{x \rightarrow a^-} f(x) = f(a)$. If f is continuous for each $a \in U$ and $U \subseteq \text{dom}(f)$ then we say that f is continuous on U . Moreover, if f is continuous on $\text{dom}(f)$ then we say that f is a **continuous function**.

I should caution you that it is often the case that you want to affix the qualifier *connected* to the domain of the function to avoid jumps in the graph. Technically, this definition does allow for jumps in the graph if the domain is not connected⁴.

Proposition 3.3.2. *additivity of the limit.*

Let $a \in \mathbb{R}$. Suppose f, g are functions and $\lim_{x \rightarrow a} f(x) = L_f$ and $\lim_{x \rightarrow a} g(x) = L_g$ then

$$\lim_{x \rightarrow a} [f(x) + g(x)] = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x).$$

Proof: we are given that $\lim_{x \rightarrow a} f(x) = L_f$ and $\lim_{x \rightarrow a} g(x) = L_g$. Let $\epsilon > 0$, because of the given limits for f and g , we can find $\delta_f, \delta_g > 0$ such that $|f(x) - L_f| < \epsilon/2$ whenever $x \in B_{\delta_f}(a)_o$ and $|g(x) - L_g| < \epsilon/2$ whenever $x \in B_{\delta_g}(a)_o$. We would like for both conditions to hold at once so we choose $\delta = \min(\delta_f, \delta_g)$. Suppose then $x \in \mathbb{R}$ and $0 < |x - a| < \delta$ it follows that $|f(x) - L_f| < \epsilon/2$ and $|g(x) - L_g| < \epsilon/2$. Consider that

$$|f(x) + g(x) - (L_f + L_g)| = |f(x) - L_f + g(x) - L_g| \leq |f(x) - L_f| + |g(x) - L_g| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Therefore, by the definition of the limit, $\lim_{x \rightarrow a} [f(x) + g(x)] = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x)$. \square .

Corollary 3.3.3. *sum of continuous functions is continuous.*

If $\text{dom}(f) = \text{dom}(g)$ and f, g are continuous functions then $f + g$ is a continuous function.

A "corollary" to a proposition is simply a fact which is so connected to the proposition that it follows with almost no proof. The proof of this corollary follows immediately once you see continuity gives that $L_f = f(a)$ and $L_g = g(a)$.

⁴in particular, if the domain is just a set of disconnected points then this definition says the function is continuous by default.

Proposition 3.3.4. *homogeneity of the limit.*

Let $a, c \in \mathbb{R}$. Suppose f is a function and $\lim_{x \rightarrow a} f(x) = L$ then

$$\lim_{x \rightarrow a} cf(x) = c \lim_{x \rightarrow a} f(x).$$

Proof: we are given that $\lim_{x \rightarrow a} f(x) = L$.

First, suppose $c \neq 0$. Let $\epsilon > 0$ and choose $\delta > 0$ such that $|f(x) - L| < \epsilon/|c|$ whenever $0 < |x - a| < \delta$. We can choose such a δ because $\epsilon/|c| > 0$ and $\lim_{x \rightarrow a} f(x) = L$ means we can choose a small enough nbhd of a to obtain values for $f(x)$ as close as we wish to L (in this case we wish the values to be within $\epsilon/|c|$ -units. Suppose that $x \in \mathbb{R}$ such that $0 < |x - a| < \delta$,

$$|cf(x) - cL| = |c(f(x) - L)| = |c||f(x) - L| < |c|(\epsilon/|c|) = \epsilon.$$

Therefore, by the definition of the limit, if $c \neq 0$ then $\lim_{x \rightarrow a} cf(x) = c \lim_{x \rightarrow a} f(x)$.

Second, consider the case $c = 0$. Let $\epsilon > 0$ and note that $cf(x) = 0$ for all $x \in \text{dom}(f)$. Choose⁵ $\delta = 1$ and suppose $0 < |x - a| < \delta$. Note

$$|cf(x) - cL| = |0 - 0| = 0 < \epsilon.$$

Hence, if $c = 0$ then $\lim_{x \rightarrow a} cf(x) = c \lim_{x \rightarrow a} f(x)$. Therefore the theorem holds true for all possible values of $c \in \mathbb{R}$. \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.

Corollary 3.3.5. *constant multiple of continuous function is continuous.*

If f is a continuous function then cf is also continuous.

Proof: since f is continuous $\lim_{x \rightarrow a} f(x) = f(a)$ for each $a \in U$ then by the homogeneity of the limit, $\lim_{x \rightarrow a} cf(x) = cf(a)$. Therefore, cf is continuous at each point in its domain. \square

Proposition 3.3.6. *limit of composites.*

Suppose f and g are functions such that $\lim_{x \rightarrow a} f(x) = L_1$ and $\lim_{y \rightarrow L_1} g(y) = L_2$ then $\lim_{x \rightarrow a} g(f(x)) = L_2$.

Proof: let $\epsilon > 0$ and choose $\delta > 0$ such that if $0 < |x - a| < \delta$ then $|f(x) - L_1| < \delta_2$ where $\delta_2 > 0$ is small enough to force $|g(y) - L_2| < \epsilon$ for all $y \in \mathbb{R}$ such that $0 < |y - L_1| < \delta_2$. We can choose $\delta_2 > 0$ as above because we were given that $\lim_{y \rightarrow L_1} g(y) = L_2$ and we can choose $\delta > 0$ to force $|f(x) - L_1| < \delta_2$ because we were also given that $\lim_{x \rightarrow a} f(x) = L_1$. Suppose that $x \in \mathbb{R}$ such that $0 < |x - a| < \delta$ and observe that $|g(f(x)) - L_2| < \epsilon$. Therefore, by the definition of the limit, $\lim_{x \rightarrow a} g(f(x)) = L_2$. \square

I was tempted to relabel the proposition above "limit of composites(sometimes)". The term "sometimes" might be included to encourage the reader to think about cases other than the one covered by this proposition. For example, if one of the factors in the product has a divergent limit is this proposition true? More generally,

⁵Note that any choice of $\delta > 0$ will suffice in this trivial case, I use 1 for no particular reason. You could just as well make $\delta = 42$.

can this proposition be extended to the case that the one or more of the factors has a limit which does not exist? In any event, we should always be mindful of what presuppositions are made in the statement of some theorem.

Corollary 3.3.7. *composite of continuous functions is continuous.*

If $f \circ g$ is well-defined and f and g are continuous functions then $f \circ g$ is a continuous function.

Proof: Note that continuity of g yields $\lim_{x \rightarrow a} g(x) = g(a)$. Since $f \circ g$ is well-defined I know $g(a) \in \text{dom}(f)$ hence $\lim_{y \rightarrow g(a)} f(y) = f(g(a))$. By the preceding proposition $\lim_{x \rightarrow a} (f \circ g)(x) = f(g(a))$ thus $f \circ g$ is a continuous function since this holds for each point in the domain of $f \circ g$. \square

Proposition 3.3.8. *limit of product is product of limits.*

Let $a \in \mathbb{R}$. Suppose f, g are functions and $\lim_{x \rightarrow a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \rightarrow a} g(x) = L_g \in \mathbb{R}$ then

$$\lim_{x \rightarrow a} [f(x)g(x)] = \left(\lim_{x \rightarrow a} f(x) \right) \left(\lim_{x \rightarrow a} g(x) \right).$$

Proof: Consider that we wish to find $\delta > 0$ that forces $x \in B_\delta(a)_o$ to satisfy

$$|f(x)g(x) - L_f L_g| < \epsilon$$

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:

$$\begin{aligned} |f(x)g(x) - L_f L_g| &= |f(x)g(x) - L_f g(x) + L_f g(x) - L_f L_g| \\ &\leq |f(x) - L_f| |g(x)| + |L_f| |g(x) - L_g| \end{aligned}$$

Very well, most things above are easy to control, however the $|g(x)|$ requires a bit of thought. Not too much though, since $\lim_{x \rightarrow a} g(x) = L_g$ it follows that $|g(x)|$ can be made close to $|L_g|$ for a particularly small nbhd of a . In fact, combining Proposition 3.3.6 and Example 3.2.10 we have a proof that $\lim_{x \rightarrow a} |g(x)| = |L_g|$ (take the outside function to be the absolute value function in the proposition). After a little scratch work I found the following argument:

Let $\epsilon > 0$, observe that since $\lim_{x \rightarrow a} |g(x)| = |L_g|$, $\lim_{x \rightarrow a} g(x) = L_g$ and $\lim_{x \rightarrow a} |f(x)| = |L_f|$ we can choose $\delta_{|g|}, \delta_g, \delta_f > 0$ such that $x \in B_{\delta_{|g|}}(a)_o$ implies $||g(x)| - |L_g|| < \beta$, $x \in B_{\delta_g}(a)_o$ implies $|g(x) - L_g| < \beta$ and $x \in B_{\delta_f}(a)_o$ implies $|f(x) - L_f| < \beta$. Simply choose $\delta = \min(\delta_{|g|}, \delta_g, \delta_f)$ to obtain that $x \in B_\delta(a)_o$ implies $|f(x) - L_f|, |g(x) - L_g|, |g(x) - |L_g|| < \beta$ for any $\beta > 0$. I propose we choose $\delta > 0$ such that

$$\beta = \frac{-(|L_f| + |L_g|) + \sqrt{(|L_f| + |L_g|)^2 + 4\epsilon}}{2}.$$

I leave it to the reader to convince themselves that $\beta > 0$. Suppose that $x \in B_\delta(a)_o$ then,

$$\begin{aligned} |f(x)g(x) - L_f L_g| &= |f(x)g(x) - L_f g(x) + L_f g(x) - L_f L_g| && \text{added zero} \\ &\leq |f(x) - L_f| |g(x)| + |L_f| |g(x) - L_g| && \text{triangle inequality} \\ &\leq \beta |g(x)| + |L_f| \beta && \text{construction of } \delta \\ &\leq \beta (|L_g| + \beta) + |L_f| \beta && \text{note } |g(x)| < |L_g| + \beta \\ &= \beta^2 + (|L_f| + |L_g|) \beta \\ &= \epsilon \end{aligned}$$

note $\beta^2 + (|L_f| + |L_g|)\beta - \epsilon = 0$ has solution $\beta = \frac{-(|L_f| + |L_g|) + \sqrt{(|L_f| + |L_g|)^2 + 4\epsilon}}{2}$ then solve for ϵ to understand the last step above. To summarize, we have shown for each $\epsilon > 0$ that there exists a $\delta > 0$ such that $x \in B_\delta(a)_o$ implies $|f(x)g(x) - L_f L_g| < \epsilon$. The proposition follows. \square

The proof given above can be shortened by a few clever moves. See Apostle for another way to attack the proof. The proof given in Appendix F of Stewart's *Calculus* is similar to the proof I gave here.

Corollary 3.3.9. *product of continuous functions is continuous.*

If $\text{dom}(f) = \text{dom}(g)$ and f, g are continuous functions then fg is a continuous function.

Proof: Note that continuity of f and g yields $\lim_{x \rightarrow a} f(x) = f(a)$ and $\lim_{x \rightarrow a} g(x) = g(a)$ respective. By the preceding proposition $\lim_{x \rightarrow a} f(x)g(x) = \lim_{x \rightarrow a} f(x) \lim_{x \rightarrow a} g(x) = f(a)g(a) = (fg)(a)$ thus fg is a continuous function since this holds for each point in the domain of fg . \square

Proposition 3.3.10. *limit of quotient is quotient of limits.*

Let $a \in \mathbb{R}$. Suppose f, g are functions and $\lim_{x \rightarrow a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \rightarrow a} g(x) = L_g \neq 0 \in \mathbb{R}$ with $L_g \neq 0$ then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)}.$$

Proof: note that if we show that $\lim_{y \rightarrow b} \frac{1}{y} = \frac{1}{b}$ then the proposition follows since $h(x) = \frac{1}{g(x)}$ is the composite of g and the reciprocal function and $f(x)h(x) = \frac{f(x)}{g(x)}$ and we already proved the product of existent limits it the limit of the product. I leave the proof that $\lim_{y \rightarrow b} \frac{1}{y} = \frac{1}{b}$ as a problem in your homework⁶. \square

Corollary 3.3.11. *quotient of continuous functions is continuous.*

If $\text{dom}(f) = \text{dom}(g)$ and f, g are continuous functions then $\frac{f}{g}$ is a continuous function on connected subsets of $\text{dom}(g) - \{x \in \mathbb{R} \mid g(x) = 0\}$.

Proof: Note that continuity of f and g yields $\lim_{x \rightarrow a} f(x) = f(a)$ and $\lim_{x \rightarrow a} g(x) = g(a)$ respective. If $g(a) \neq 0$ then By the preceding proposition

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)} = \frac{f(a)}{g(a)} = \left(\frac{f}{g}\right)(a)$$

thus $\frac{f}{g}$ is a continuous function on each connected subset of $\text{dom}(\frac{f}{g})$ since this holds for each point in the domain of $\frac{f}{g}$. \square

Remark 3.3.12.

The concept of continuity can be extended to non-connected sets, however the interesting theorems in this course all assume continuity of a function with respect to some connected subset of the real numbers. For this reason I refuse to discuss continuity beyond that context. This is why I did not say the quotient was continuous function. If $U \subset \mathbb{R}$ is disconnected then I have yet to provide a definition for continuity with respect to U . Most rational functions have dis-connected domains; for example $f(x) = \frac{1}{x}$ has $\text{dom}(f) = (-\infty, 0) \cup (0, \infty)$.

⁶Stewart has a self-contained proof in Appendix F for the proof of Law 5. You might be able to twist his argument to do this homework problem.

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 3.3.13. *limit of constant is the constant.*

Let $a, c \in \mathbb{R}$, $\lim_{x \rightarrow a} c = c$.

Proof: let $\epsilon > 0$ choose $\delta = 2$ and note $0 < |x - a| < \delta$ implies $|c - c| = |0| = 0 < \epsilon$. Therefore, by the definition of the limit, $\lim_{x \rightarrow a} c = c$. \square .

Hopefully at this point you know how to prove the following corollary.

Corollary 3.3.14. *constant functions are continuous.*

If $\text{dom}(f)$ is connected and $f(x) = c$ for all $x \in \text{dom}(f)$ then f is a continuous function.

Proposition 3.3.15. *limit identity function returns the limit point.*

Let $a \in \mathbb{R}$, $\lim_{x \rightarrow a} x = a$.

Proof: let $\epsilon > 0$ choose $\delta = \epsilon$ and note $0 < |x - a| < \delta$ implies $|x - a| < \delta = \epsilon$. Therefore, by the definition of the limit, $\lim_{x \rightarrow a} x = a$. \square .

The function $\text{id}(x) = x$ is called the identity function because it returns an output of x which is identical to its input x . Also, $g \circ \text{id} = g$, so it behaves like the number 1 which is the multiplicative identity for \mathbb{R} . Again, given the proposition above the following corollary is obvious:

Corollary 3.3.16. *constant functions are continuous.*

The identity function $f(x) = x$ is a continuous function.

Naturally, the restriction of the identity function to any connected subset of \mathbb{R} is also continuous. Moving on,

Proposition 3.3.17. *power function limit (for powers $n \in \mathbb{N}$).*

Let $a \in \mathbb{R}$ and $n \in \mathbb{N}$, $\lim_{x \rightarrow a} x^n = a^n$.

Proof: whenever we want to show something is true for arbitrary $n \in \mathbb{N}$ we use a proof method called induction. The way it works is that we have to show the statement is true for $n = 1$, which we already proved in our last theorem. Then we must show that if the statement is true for $n \in \mathbb{N}$ then it is also true for $n + 1$. If we can make that "induction step" then proof by mathematical induction applies and the theorem is valid for all $n \in \mathbb{N}$. You can look at the appendix on induction if you want to see more about induction for inductions sake. Let's assume this theorem holds for $n \in \mathbb{N}$ then $\lim_{x \rightarrow a} x^n = a^n$. Consider the $(n + 1)$ case,

$$\lim_{x \rightarrow a} x^{n+1} = \lim_{x \rightarrow a} x^n x = \left(\lim_{x \rightarrow a} x^n \right) \left(\lim_{x \rightarrow a} x \right) = a^n a = a^{n+1}$$

where I used the product of limits theorem and the identity function limit theorem once more. 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

Corollary 3.3.18. *continuity of integer power functions.*

Let $n \in \mathbb{Z}$, a power function $f(x) = x^n$ is continuous at each point in the interior of its domain.

Proof: notice that $\text{dom}(x^n) = \mathbb{R}$ if $n \geq 0$ whereas $\text{dom}(x^n) = (-\infty, 0) \cup (0, \infty)$ if $n < 0$. In the case $n \in \mathbb{N}$ we have $n > 0$ and the proposition preceding this coro. gives $\lim_{x \rightarrow a} x^n = a^n$ hence $f(x) = x^n$ is continuous on \mathbb{R} . In contrast, if $n < 0$ then $-n > 0$ thus $f(x) = x^n = \frac{1}{x^{-n}}$ is the quotient of continuous functions thus f is continuous on $(-\infty, 0)$ and $(0, \infty)$. \square

Proposition 3.3.19. *root function limit (power function with power $1/n$ for $n \in \mathbb{N}$).*

Let $a \in \mathbb{R}$ with $a > 0$ and $n \in \mathbb{N}$, $\lim_{x \rightarrow a} \sqrt[n]{x} = \sqrt[n]{a}$.

Proof: Let $\epsilon > 0$ and choose $\delta = \min(a/2, \epsilon \sqrt[n]{a})$. Suppose $x \in B_\delta(a)_o$ then we have $0 < |x - a| < \delta$ hence $0 < a - \delta < x < a + \delta$. Note that our construction of δ insures that $\delta \leq a/2 < a$ hence $0 < a - \delta$. Continuing,

$$|\sqrt[n]{x} - \sqrt[n]{a}| = \left| \frac{x - a}{\sqrt[n]{x} + \sqrt[n]{a}} \right| = \frac{|x - a|}{\sqrt[n]{x} + \sqrt[n]{a}} < \frac{|x - a|}{\sqrt[n]{a}} < \frac{\delta}{\sqrt[n]{a}} \leq \frac{\epsilon \sqrt[n]{a}}{\sqrt[n]{a}} = \epsilon.$$

To summarize, for each $\epsilon > 0$ we have shown there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|\sqrt[n]{x} - \sqrt[n]{a}| < \epsilon$ hence $\lim_{x \rightarrow a} \sqrt[n]{x} = \sqrt[n]{a}$. \square

For $n = 2k \in 2\mathbb{N}$ an even power we can consider right limits at $a = 0$ and a argument similar to the one offered above will prove $\sqrt[2k]{x} \rightarrow 0$ as $x \rightarrow 0^+$. However, for an even power the root function is not real-valued to the left of the origin so the double sided limit at zero does not exist. In contrast, the proposition above could be extended for all $a \in \mathbb{R}$ if it is the case that $n = 2k + 1 \in 2\mathbb{N} + 1$ is an odd power. Moreover, given the limit law about composite limits and our previous work on the reciprocal function we can prove the following:

Proposition 3.3.20. *rational power limit (power function with power m/n for $m, n \in \mathbb{N}$).*

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

Notice that fractional powers are problematic for negative numbers. If you agree that $\frac{1}{3} = \frac{1}{2} \cdot \frac{2}{3}$ then you should ask yourself what domain would you assign $f(x) = x^{\frac{1}{3}}$? What about $g(x) = x^{\frac{1}{2} \cdot \frac{2}{3}}$? What about $h(x) = (\sqrt{x})^{\frac{2}{3}}$? I would argue that $\text{dom}(f) = \mathbb{R}$ whereas $\text{dom}(h) = [0, \infty)$. 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 Math 331 sometime soon and all the mysteries of complex arithmetic will become clear.

From this point we could go on and prove dozens of propositions about limits of your favorite algebraic functions. Let me summarize: if you can plug in the limit point and avoid division by zero or an even root of a negative number then the formula of the function gives the output of the limit by simple function evaluation. In other words:

Theorem 3.3.21. *continuity of algebraic functions.*

Let $f(x)$ be defined by a finite number of algebraic operations (possibly including addition, multiplication, division, taking integer or fractional roots) then f is continuous at each point in the interior of its domain.

I think we've seen enough detail in this direction so we now turn to limits of sine and cosine.

Proposition 3.3.22. *limits of sine and cosine.*

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

Proof: in your homework you will prove that if $0 \leq |x| \leq \pi/2$ then $|\sin(x)| \leq |x|$. To begin this proof we show that $\lim_{x \rightarrow 0} \sin(x) = \sin(0) = 0$.

Let $\epsilon > 0$ and choose $\delta = \min(\epsilon, \pi/4)$. Suppose that $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. Since $\delta \leq \pi/4 < \pi/2$ your homework gives us the result $|\sin(x) - 0| = |\sin(x)| < |x| < \delta \leq \epsilon$. Therefore, for each $\epsilon > 0$ we have shown there exists $\delta > 0$ such that $0 < |x - 0| < \delta$ implies $|\sin(x) - 0| < \epsilon$ hence $\lim_{x \rightarrow 0} \sin(x) = 0$.

$$\begin{aligned}
 \lim_{x \rightarrow 0} \cos(x) &= \lim_{x \rightarrow 0} [1 - 2\sin^2(x/2)] && \text{trig. identity} \\
 &= \lim_{x \rightarrow 0} 1 - 2 \lim_{x \rightarrow 0} \sin(x/2) \lim_{x \rightarrow 0} \sin(x/2) && \text{limit laws} \\
 &= 1 - 2 \sin\left(\lim_{x \rightarrow 0} x/2\right) \sin\left(\lim_{x \rightarrow 0} x/2\right) && \text{composition limit law} \\
 &= 1 - 2 \sin\left(\frac{1}{2} \lim_{x \rightarrow 0} x\right) \sin\left(\frac{1}{2} \lim_{x \rightarrow 0} x\right) && \text{homogeneity limit law} \\
 &= 1 - 2 \sin(0) \sin(0) && \text{limit of identity function} \\
 &= 1. && \text{definition of sine function}
 \end{aligned}$$

Lemma 3.3.23. *substitution of limiting variable.*

$$\lim_{x \rightarrow a} f(x) \in \mathbb{R} \Leftrightarrow \lim_{h \rightarrow 0} f(a+h) \in \mathbb{R}$$

Proof of lemma: suppose $\lim_{x \rightarrow a} f(x) = L_2 \in \mathbb{R}$. Let $g(h) = a+h$ and note $g(h) \rightarrow a$ as $h \rightarrow 0$. Therefore, $\lim_{h \rightarrow 0} f \circ g(h) = \lim_{h \rightarrow 0} f(a+h) = L_2$ by the limit of composites law.

Conversely, suppose $\lim_{h \rightarrow 0} f(a+h) = L_2 \in \mathbb{R}$. Let $p(h) = f(a+h)$ hence $\lim_{h \rightarrow 0} p(h) = L_2$. Furthermore, define $q(x) = x - a$. Observe that $q(x) \rightarrow 0$ as $x \rightarrow a$. Consider then

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} f(a + x - a) = \lim_{x \rightarrow a} f(a + q(x)) = \lim_{x \rightarrow a} p(q(x)) = L_2$$

where the last step again uses the limit composition law⁷ ∇

Finally, using the facts already proven together with the adding angle formulas suffices to complete the proof:

$$\begin{aligned}\lim_{x \rightarrow a} \cos(x) &= \lim_{h \rightarrow 0} \cos(a + h) \\ &= \lim_{h \rightarrow 0} \cos(a) \cos(h) - \sin(a) \sin(h) \\ &= \cos(a) \lim_{h \rightarrow 0} \cos(h) - \sin(a) \lim_{h \rightarrow 0} \sin(h) \\ &= \cos(a).\end{aligned}$$

I leave the proof that $\lim_{x \rightarrow a} \sin(x) = \sin(a)$ as a homework problem⁸. This completes the proof of proposition 3.3.22. \square

Remark 3.3.24.

It is sometimes quite annoying how difficult it is to prove something as graphically obvious as the preceding proposition. Of course the sine and cosine function have limits which are nothing more than the sine or cosine function evaluated at the limit point. Make no mistake, the preceding proof was not superfluous, if we skip it then something is missing. We're laying a foundation currently. Once it's built then we we'll just use it.

In our lexicon of basic functions there are two things we have yet to cover in this section:

1. exponential functions
2. inverse functions

Once we have the exponential function and the inverse function then we can pretty much complete our list of basic limits. For example, $f(x) = x^\pi = e^{\pi \ln(x)}$ can be taken as the definition of x^π for $x > 0$. Currently we have only covered fractional power functions.

Proposition 3.3.25. *limit of exponential function.*

Let $b > 0$, $\lim_{x \rightarrow a} b^x = b^a$.

Proof: We begin by proving $\lim_{x \rightarrow 0} 2^x = 1$. Let $\epsilon > 0$ and choose $\delta = \log_2(1 + \epsilon)$. Note that $1 + \epsilon > 1$ hence $\log_2(1 + \epsilon) > 0$. Suppose that $x \in \mathbb{R}$ such that $0 < |x| < \delta$, it follows that

$$-\log_2(1 + \epsilon) = \log_2\left(\frac{1}{1 + \epsilon}\right) < x < \log_2(1 + \epsilon)$$

but surely⁹ $x < y$ implies $2^x < 2^y$ thus

$$\frac{1}{1 + \epsilon} < 2^x < 1 + \epsilon$$

subtracting one from each inequality yields,

$$\frac{1}{1 + \epsilon} - 1 < 2^x - 1 < \epsilon$$

⁷the upside-down triangle indicates the proof of the lemma is complete however the proposition's proof is still unfinished. The method of proof we just used here is called a "biconditional proof". To prove \Leftrightarrow we proved \Rightarrow and \Leftarrow .

⁸it should be easy if you understand the argument we just made for cosine here!

⁹admitably there is a gap here, I invite the reader to supply a proof

Note that $\frac{1}{1+\epsilon} - 1 = -\frac{\epsilon}{1+\epsilon} > -\epsilon$ thus $-\epsilon < 2^x - 1 < \epsilon$ which is equivalent to $|2^x - 1| < \epsilon$. Hence, $0 < |x| < \delta$ implies $|2^x - 1| < \epsilon$. Therefore, $\lim_{x \rightarrow 0} 2^x = 1$.

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 $\log_2(b)x \rightarrow 0$ as $x \rightarrow 0$. Moreover, using the composition of limits proposition we find that $b^x = 2^{\log_2(b)x} \rightarrow 1$ as $x \rightarrow 0$. Thus, $\lim_{x \rightarrow 0} b^x = 1$.

The laws of exponents complete the proof for limit points other than zero:

$$\lim_{h \rightarrow 0} (b^{a+h}) = \lim_{h \rightarrow 0} (b^a b^h) = b^a \lim_{h \rightarrow 0} b^h = b^a.$$

Then by Lemma 3.3.23, $\lim_{x \rightarrow a} (b^x) = b^a$. \square

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. The interesting thing about this proof is that in retrospect we can replace the proof for the root function, reciprocal function, identity function and natural number power function with this one proof.

Proposition 3.3.26. *continuous injections are strictly monotonic.*

If U is connected and $f : U \rightarrow V$ is continuous then f is 1-1 iff f is either strictly increasing or strictly decreasing on U .

Remark 3.3.27.

You can skip the proof of this proposition until later. It would be better to read this after you read the section on the intermediate value theorem. I'm leaving it here because I aim to state an important theorem at the end of this section which is not possible without the last couple propositions in this section.

Proof: if we drop the condition of continuity then we could jump from increasing to decreasing on different components of U . However, continuity should keep f from jumping so if f is increasing on part of U it should continue to increase over the whole set. This is not a proof, rather just an argument for plausibility. The proof is somewhat technical but the key is the intermediate value theorem (IVT). Forgive me if we don't prove the IVT just yet.

Suppose $f : U \rightarrow V$ is continuous and strictly increasing or decreasing. We have shown before that strict monotonicity implies injectivity. See prop. 2.7.6.

Conversely, suppose $f : (a, b) \rightarrow (c, d)$ is continuous and 1-1. We seek to show that f is either increasing or decreasing. Suppose f is strictly increasing on the connected subsets $U_j \subseteq U$ for $j = 1, 2, \dots$. Likewise, suppose f is strictly decreasing on connected subsets $V_k \subseteq U$ for $k = 1, 2, \dots$. The union of sets U_j and V_k for all j, k should yield U . Of particular interest are the points which are on the edge between U_j and V_k . Suppose in particular 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 1-1 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 $p > f(x_o)$ 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, we have shown that f is either strictly increasing or strictly decreasing. \square

Proposition 3.3.28. *invertible continuous function have continuous inverses.*

Let $U, V \subseteq \mathbb{R}$, if $f : U \rightarrow V$ is continuous with inverse $f^{-1} : V \rightarrow U$ then f^{-1} is continuous.

Proof: we seek to show f^{-1} is continuous at $y_o \in V$. Let $\epsilon > 0$ and suppose $x_o = f^{-1}(y_o)$, choose $\delta = \min[f(x_o) - f(x_o - \epsilon), f(x_o + \epsilon) - f(x_o)]$ and suppose $0 < |y - y_o| < \delta$. Note that

$$y < y_o + \delta \leq f(x_o) + [f(x_o + \epsilon) - f(x_o)] = f(x_o + \epsilon)$$

Then on the other side,

$$y > y_o - \delta \geq f(x_o) - [f(x_o - \epsilon) - f(x_o)] = f(x_o - \epsilon)$$

Putting together the inequalities above yields $f(x_o - \epsilon) < y < f(x_o + \epsilon)$. 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 U . Suppose f^{-1} is strictly increasing then it follows:

$$x_o - \epsilon < f^{-1}(y) < x_o + \epsilon \Rightarrow |f^{-1}(y) - x_o| < \epsilon \Rightarrow |f^{-1}(y) - f^{-1}(y_o)| < \epsilon.$$

If f^{-1} is strictly decreasing then we again find that $0 < |y - y_o| < \delta$ implies $|f^{-1}(y) - y_o| < \epsilon$. Therefore, $\lim_{y \rightarrow y_o} f^{-1}(y) = f^{-1}(y_o)$ for each $y_o \in V$ hence f^{-1} is continuous. \square

Proposition 3.3.29. *continuity of power function for arbitrary power.*

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

Proof: the following identity could be used as the definition if $p > 0$, $x^p = 2^{p \log_2(x)}$. By prop. 3.3.28 we know logarithms are continuous since each logarithm is inverse function of an exponential function which we already proved was continuous. Note that,

$$\lim_{x \rightarrow a} (x^p) = \lim_{x \rightarrow a} (2^{p \log_2(x)}) = 2^{\lim_{x \rightarrow a} (p \log_2(x))} = 2^{p \log_2(a)} = a^p. \quad \square$$

Remark 3.3.30.

The beauty of the logarithm is that it changes products to sums: $\log(fg) = \log(f) + \log(g)$. Note that $fg = 2^{\log_2(fg)} = 2^{\log_2(f) + \log_2(g)}$. If we knew that $f(x) = \log_2(x)$ was a continuous function then we would have a proof that the product of two continuous function is continuous since it is built using the composite and the sum of continuous functions. In other words, continuity of the logarithm yields an easy proof of that the product of continuous functions is continuous. A similar proof is possible for the quotient. However, these proofs are not as general as the one already offered since we can only make these arguments for functions with positive outputs.

The functions listed in chapter 2 are called *elementary functions*. Perhaps the best summary of this section is as follows:

Proposition 3.3.31. *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.

I'm sometimes tempted to say that all functions whose formula is constructed from finitely many elementary functions are continuous at each point in the interior of their domains. However, I think a clever student could find a counter-example. This much is almost always true:

Remark 3.3.32.

If a function is defined at the limit point then the value of the function at the limit point is simply given by function evaluation. In other words, functions are usually continuous.

3.4 limit calculation

In the preceding section we made painstaking arguments to prove most of the basic theorems about limits. Fortunately we will not usually find that level of detail is necessary to calculate a given limit. In fact, the point of this section is that we don't have to use the definition to calculate most limits. Rigorous arguments can be built via combining the propositions of the preceding section together with some crafty algebraic techniques. Non-rigorous, intuitive arguments are also possible through either numerical calculation (table of values) or by leading term analysis. Which argument is best depends on your audience.

Example 3.4.1. *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.*

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

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

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

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

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.

Example 3.4.2.

$$\begin{aligned} \lim_{x \rightarrow 0} (\sin(x) + \cos(e^x)) &= \lim_{x \rightarrow 0} (\sin(x)) + \lim_{x \rightarrow 0} (\cos(e^x)), \quad (\text{additivity.}) \\ &= \sin(0) + \cos(\lim_{x \rightarrow 0} e^x), \quad (\text{continuity of sine and cosine.}) \\ &= \sin(0) + \cos(e^0) \quad (\text{continuity of exponential.}) \\ &= \cos(1). \end{aligned}$$

I may ask you to calculate a particular limit a particular way. However, if I don't say one way or the other you are free to think for yourself. Sometimes a graph is a good solution, sometimes a table of values is convenient, sometimes we can use propositions from section 3.3. The example below illustrates the table of values idea.

Example 3.4.3. The following table of values indicates that $\lim_{x \rightarrow 0^+} \frac{\sin(x)}{x} = 1$

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 consider in Example 3.4.3 is not nearly as obvious as the limits in Example 3.4.1. 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 chapter 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 3.4.4. *indeterminant forms.*

1. we say $\lim \frac{f}{g}$ is of "type $\frac{0}{0}$ " iff $\lim f = 0$ and $\lim g = 0$
2. we say $\lim \frac{f}{g}$ is of "type $\frac{\infty}{\infty}$ " iff $\lim f = \pm\infty$ and $\lim g = \pm\infty$
3. we say $\lim fg$ is of "type 0∞ " iff $\lim f = 0$ and $\lim g = \pm\infty$
4. 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.

Example 3.4.5. Calculate $\lim_{x \rightarrow -2} \left(\frac{x+2}{x^2+3x+2} \right)$. Notice that this limit is of type $0/0$ since the numerator and denominator are both zero when take the limit at -2 .

$$\begin{aligned}
 \lim_{x \rightarrow -2} \left(\frac{x+2}{x^2+3x+2} \right) &= \lim_{x \rightarrow -2} \left(\frac{x+2}{(x+2)(x+1)} \right) \\
 &= \lim_{x \rightarrow -2} \left(\frac{1}{x+1} \right) \\
 &= \frac{1}{-2+1} \\
 &= -1.
 \end{aligned}$$

The second step where we cancelled $(x+2)$ with $(x+2)$ is valid inside the limit because we do not have $x = -2$ in the limit. We get very close, but that is the difference, this is not division by zero.

Example 3.4.6. *The limit below is also type 0/0 to begin with*

$$\begin{aligned}\lim_{x \rightarrow 0} \left(\frac{3x + x^2}{x^3 + 2x^2 + x} \right) &= \lim_{x \rightarrow 0} \left(\frac{x(x + 3)}{x(x^2 + 2x + 1)} \right) \\ &= \lim_{x \rightarrow 0} \left(\frac{(x + 3)}{(x^2 + 2x + 1)} \right) \\ &= \frac{0 + 3}{0 + 0 + 1} \\ &= 3.\end{aligned}$$

I reiterate, we can cancel the x/x inside the limit because $x \neq 0$ within the limit. Again we see that factoring and cancellation has allowed us to modify the limit so that we could reasonably plug in the limit point in the simplified limit. This is often the goal.

We observe that algebraic manipulations may change an undetermined form to a determined form which does not violate the laws of real arithmetic when you plug in the limit point.

Example 3.4.7. *This limit also has form 0/0 to begin.*

$$\begin{aligned}\lim_{\theta \rightarrow 0} \left(\frac{\tan(\theta)}{\sin(\theta)} \right) &= \lim_{\theta \rightarrow 0} \left(\frac{\sin(\theta)}{\cos(\theta)} \frac{1}{\sin(\theta)} \right) \\ &= \lim_{\theta \rightarrow 0} \left(\frac{1}{\cos(\theta)} \right) \\ &= \frac{1}{\cos(0)} \\ &= 1.\end{aligned}$$

Example 3.4.8. *The first step is a time-honored trick, it is nothing more than multiplication by 1. So if you encounter a similar problem try a similar trick.*

$$\begin{aligned}\lim_{h \rightarrow 0} \left(\frac{\sqrt{4+h} - 2}{h} \right) &= \lim_{h \rightarrow 0} \left(\frac{(\sqrt{4+h} - 2)(\sqrt{4+h} + 2)}{h(\sqrt{4+h} + 2)} \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{4 + h + 2\sqrt{4+h} - 2\sqrt{4+h} - 4}{h(\sqrt{4+h} + 2)} \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{h}{h(\sqrt{4+h} + 2)} \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{1}{\sqrt{4+h} + 2} \right) \\ &= \frac{1}{\sqrt{4} + 2} \\ &= \frac{1}{4}.\end{aligned}$$

Example 3.4.9. Here the trick is to combine the fractions in the numerator by finding the common denominator of $4x$

$$\begin{aligned}\lim_{x \rightarrow -4} \left(\frac{\frac{1}{4} + \frac{1}{x}}{4 + x} \right) &= \lim_{x \rightarrow -4} \left(\frac{\frac{x+4}{4x}}{4+x} \right) \\ &= \lim_{x \rightarrow -4} \left(\frac{x+4}{4x} \cdot \frac{1}{4+x} \right) \\ &= \lim_{x \rightarrow -4} \left(\frac{1}{4x} \right) \\ &= \frac{-1}{16}.\end{aligned}$$

Example 3.4.10.

$$\begin{aligned}\lim_{x \rightarrow 3} \left(\frac{(x-3)\cos(x-3)}{x(x^2-5x+6)} \right) &= \lim_{x \rightarrow 3} \left(\frac{(x-3)\cos(x-3)}{x(x-3)(x-2)} \right) \\ &= \lim_{x \rightarrow 3} \left(\frac{\cos(x-3)}{x(x-2)} \right) \\ &= \frac{\cos(3-3)}{3(3-2)} \\ &= \frac{1}{3}.\end{aligned}$$

Example 3.4.11. Piecewise defined functions can require a bit more care. Sometimes we need to look at one-sided limits.

$$\lim_{x \rightarrow 0} \left[\frac{|x|}{x} \right] = ?$$

recall that the notation $|x|$ is the absolute value of x , it is the distance from zero to x on the number line.

$$|x| = \begin{cases} -x & : x < 0. \\ x & : x \geq 0. \end{cases}$$

In the left limit $x \rightarrow 0^-$ we have $x < 0$ so $|x| = -x$ thus,

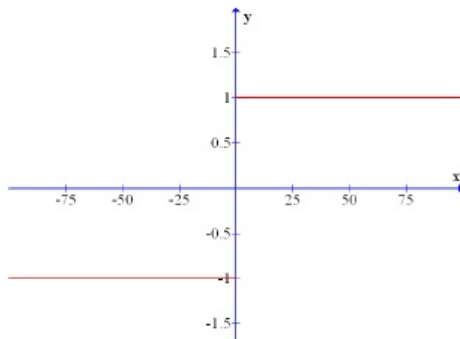
$$\lim_{x \rightarrow 0^-} \left[\frac{|x|}{x} \right] = \lim_{x \rightarrow 0^-} \left[\frac{-x}{x} \right] = \lim_{x \rightarrow 0^-} \left[\frac{-1}{1} \right] = -1.$$

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

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

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

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 3.4.12. *Good to know your trig. identities:*

$$\begin{aligned}\lim_{x \rightarrow 0} \left(\frac{\sin(x)}{\sin(2x) + \sin(x)} \right) &= \lim_{x \rightarrow 0} \left(\frac{\sin(x)}{2 \sin(x) \cos(x) + \sin(x)} \right) \\ &= \lim_{x \rightarrow 0} \left(\frac{1}{2 \cos(x) + 1} \right) \\ &= \frac{1}{3}.\end{aligned}$$

Example 3.4.13. *This limit below is not indeterminate, the type $\infty/0$ will diverge. The question is merely how does it diverge? It becomes clear this limit is positive after we simplify,*

$$\begin{aligned}\lim_{x \rightarrow 0} \left(\frac{\cot(x)}{\tan(x)} \right) &= \lim_{x \rightarrow 0} \left(\frac{1}{\cot^2(x)} \right) \\ &= \infty.\end{aligned}$$

Example 3.4.14. *This limit below is not indeterminate, the type $\infty/0$ will diverge. The question is merely how does it diverge?*

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

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

Example 3.4.15. *This limit exhibits yet another case that can arise, notice we start with the indeterminate form $0/0$ but in the next to last line we have reduced to the problem to $0/\infty$ which intuitively must yield zero.*

$$\begin{aligned}\lim_{x \rightarrow 0^+} \left(\frac{x + \frac{1}{\sqrt{x}}}{\frac{1}{\sqrt{x^2 + x}}} \right) &= \lim_{x \rightarrow 0^+} \left(\frac{x + \frac{1}{\sqrt{x}}}{\frac{1}{\sqrt{x^2}}} \right) \\ &= \lim_{x \rightarrow 0^+} \left(\frac{x\sqrt{x} + \sqrt{x}}{\frac{\sqrt{x}}{\sqrt{x^2}}} \right) \\ &= \lim_{x \rightarrow 0^+} \left(\frac{x\sqrt{x} + \sqrt{x}}{\sqrt{\frac{1}{x}}} \right) \\ &= 0.\end{aligned}$$

We may make the result obvious without appeal to intuition. Just make one more step, multiply the by \sqrt{x}/\sqrt{x} :

$$\begin{aligned}\lim_{x \rightarrow 0^+} \left(\frac{x + \frac{1}{\sqrt{x}}}{\frac{1}{\sqrt{x^2 + x}}} \right) &= \lim_{x \rightarrow 0^+} \left(\frac{x\sqrt{x} + \sqrt{x}}{\sqrt{\frac{1}{x}}} \right) \\ &= \lim_{x \rightarrow 0^+} \left(\frac{x^2 + x}{1} \right) \\ &= 0.\end{aligned}$$

I would not fuss as much at you skipping the last step as I would if you skipped the earlier steps. Again, the point of the earlier steps was to remove the indeterminacy whereas the point of the later steps was to clarify that the limit did indeed converge to zero.

Remark 3.4.16.

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 indeterminacy is removed for a given problem. I did that in every example in this section.

3.5 squeeze theorem

There are limits not easily solved through algebraic trickery. Sometimes the "Squeeze" or "Sandwich" Theorem allows us to calculate the limit.

Proposition 3.5.1. *squeeze theorem.*

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,

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

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

Proof: Suppose $f(x) \leq g(x)$ for all $x \in B_{\delta_1}(a)_o$ for some $\delta_1 > 0$ and also suppose $\lim_{x \rightarrow a} f(x) = L_f \in \mathbb{R}$ and $\lim_{x \rightarrow a} g(x) = L_g \in \mathbb{R}$. We wish to prove that $L_f \leq L_g$. Suppose otherwise towards a contradiction. That is, suppose $L_f > L_g$. Note that $\lim_{x \rightarrow a} [g(x) - f(x)] = L_g - L_f$ by the linearity of the limit. It follows that for $\epsilon = \frac{1}{2}(L_f - L_g) > 0$ there exists $\delta_2 > 0$ such that $x \in B_{\delta_2}(a)_o$ implies $|g(x) - f(x) - (L_g - L_f)| < \epsilon = \frac{1}{2}(L_f - L_g)$. 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 g and h thus the first part of the theorem follows.

Next, we suppose that $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L \in \mathbb{R}$ 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 $\lim_{x \rightarrow a} g(x) = L$. Let $\epsilon > 0$ and choose $\delta_2 > 0$ such that $|f(x) - L| < \epsilon$ and $|h(x) - L| < \epsilon$ for all $x \in B_{\delta_2}(a)_o$. 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 $|f(x) - L| < \epsilon$ and $|h(x) - L| < \epsilon$ imply $-\epsilon < f(x) - L$ and $h(x) - L < \epsilon$ thus

$$-\epsilon < f(x) - L \leq g(x) - L \leq h(x) - L < \epsilon.$$

Therefore, for each $\epsilon > 0$ there exists $\delta > 0$ such that $x \in B_{\delta}(a)_o$ implies $|g(x) - L| < \epsilon$ so $\lim_{x \rightarrow a} g(x) = L$. \square

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

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

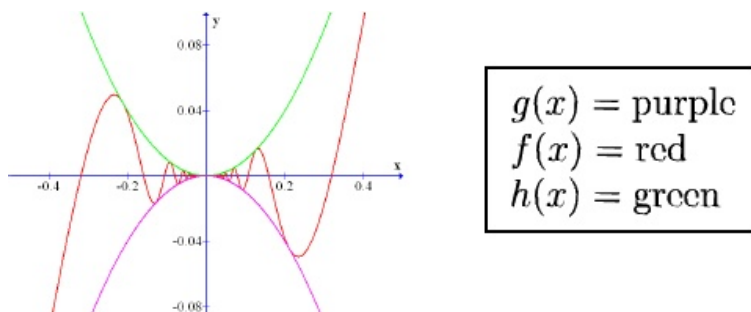
Substitute $\theta = 1/x$ and multiply by x^2 which is positive if $x \neq 0$ so the inequality is maintained,

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

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

$$\lim_{x \rightarrow 0} (x^2) = 0 \quad \lim_{x \rightarrow 0} (-x^2) = 0.$$

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



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 $\sin(\frac{1}{x})$ at zero does not exist due to wild oscillation at zero. Therefore, we have no right to apply the limit proposition.

Example 3.5.3. 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 $x \rightarrow -1$? Well, notice that

$$\lim_{x \rightarrow -1} (1) = 1 \quad \lim_{x \rightarrow -1} (x^2 + 2x + 2) = 1.$$

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

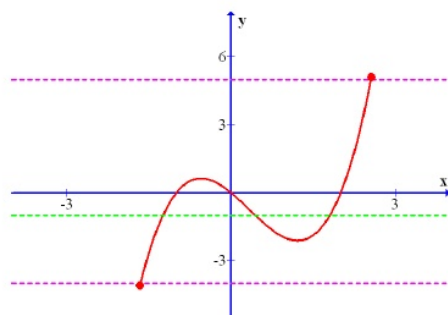
3.6 intermediate value theorem

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

Theorem 3.6.1. *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 can't 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 3.6.2. 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.

Example 3.6.3. 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,

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

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

(3.1)

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).

Remark 3.6.4. *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 $a_2 = a_1$.
- \vdots \vdots \vdots \vdots

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 .

3.6.1 a deeper look at the intermediate value theorem

Proposition 3.6.5.

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 $\lim_{x \rightarrow c} f(x) = f(a) \neq 0$.

- 1.) Assume that $f(a) > 0$. Choose $\epsilon = \frac{f(a)}{2}$ and use existence of the limit $\lim_{x \rightarrow c} f(x) = f(a)$ 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 $\epsilon = -\frac{f(a)}{2} > 0$ 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*¹⁰

¹⁰the Bolzano-Weierstrauss theorem is one of the central theorems of real analysis, in 1817 Bolzano used it to

Proposition 3.6.5 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 3.6.6. *Bolzano's theorem*

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. 3.6.5 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. 3.6.5 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 Apostol's argument, 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 .

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 Math 431 at LU.

Chapter 4

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 use Leibniz' notation for the most part. Anyway, I digress as usual.

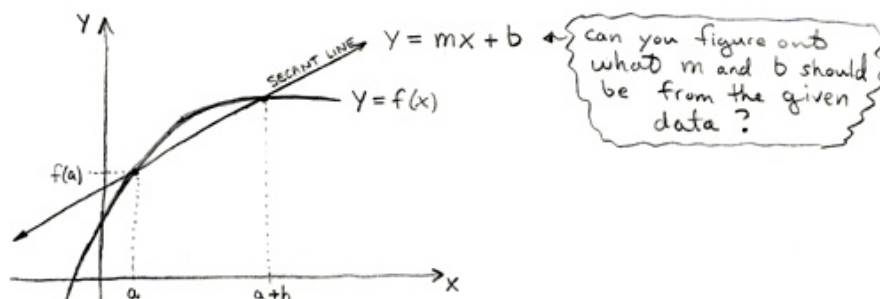
Let me briefly describe the content of this chapter. We begin by defining tangent lines and infinitesimal rates of change. Then the derivative as a function is defined and several examples exhibiting the tangent line construction are given. Next, linearity and the power rule are developed. Breaking from 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.

4.1 tangent lines

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. I have pictured a particular secant line below,



You can imagine that as h increases or decreases we will get a different secant line. In fact, there are infinitely many secant lines. Notice that the slope of the pictured secant line is just the rise over the run, that is

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

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** .

Now imagine that $h \rightarrow 0$. As we take that limit we will get the **tangent line** which just kisses the function at the point $(a, f(a))$ ¹. Moreover, we should notice that in the limit as $h \rightarrow 0$ the average rate of change is replaced with the **instantaneous rate of change of y with respect to x** , this is precisely what the slope of the tangent line means from an analytical standpoint.

The interpretations of y and x are too numerous to list. However, the most important case is arguably $y = s$ (position) and $x = t$ (time). In that context the slope of a secant line between (t_1, s_1) and (t_2, s_2) is called the **average velocity**; $v_{avg} = \frac{\Delta s}{\Delta t} = \frac{s_2 - s_1}{t_2 - t_1}$ whereas the slope of the tangent line is called the **instantaneous velocity**. Let me be precise,

¹it may however intersect the graph elsewhere depending on how the graph curves away from the point of tangency

Definition 4.1.1. *tangent line.*

The tangent line to $y = f(x)$ is the line that passes through $(a, f(a))$ and has the slope

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

if the limit exists, otherwise we say there is no tangent line at that point. If there is a tangent line through $(a, f(a))$ to the curve $y = f(x)$ then the equation for the tangent line is

$$y = f(a) + m(x - a).$$

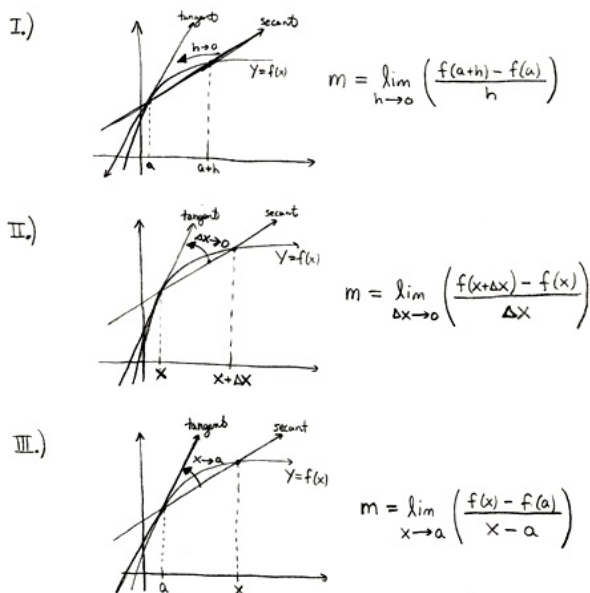
the notation $f'(a)$ draws our attention to the fact that the quantity m is also **”the derivative at a point”**. We define the **slope of a function** at a point to be the derivative of the function at that point (when it exists).

Continuing our discussion about velocity, the instantaneous velocity at time t for position s is thus defined,

$$v(t) = \lim_{h \rightarrow 0} \left(\frac{s(t+h) - s(t)}{h} \right).$$

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. The interplay between mathematics and physics continues to this day. Anyway, let’s get back to the math...

The tangent line is unique when it exists because limits are unique when they exist. There are other equivalent ways of looking at the limit which gives the slope of the tangent line. For examples:



I reiterate, the slope of the tangent line characterizes how quickly y is changing with respect to x . The slope of the tangent line gives us the *instantaneous* rate of change of y with respect to x .

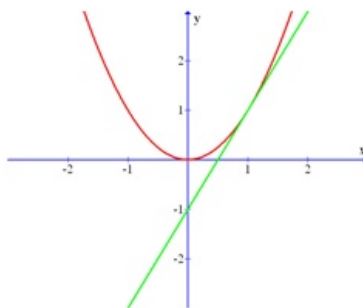
Example 4.1.2. Find the slope of the parabolic function $f(x) = x^2$ at $x = 1$. In other words, find the slope of the graph $y = f(x) = x^2$ when $x = 1$. We defined the slope of the graph at a point to be the slope of the tangent line at that point. So we calculate,

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

I have listed more steps than I typically do for such limits. Notice the critical thing here is that once the 1 cancels with -1 then all terms have a factor of h so it cancels with the h in the denominator. We see that the slope of the parabola at the point $(1, 1)$ is $f'(1) = 2$. Moreover, we can even find the equation of the tangent line follows,

$$y = f(1) + f'(1)(x - 1) = 1 + 2(x - 1) = 2x - 1.$$

It is possible to find the tangent line approximately through drawing a careful graph and using a ruler and graph paper. But, our results are not approximate. We found the exact result using calculus. Here is what it looks like,



You may be wondering, when does the derivative at a point fail to exist? What sort of function would make that happen? The example that follows illustrates one culprit, a "kink" or "corner" in the graph. This means that a function does not have a well-defined slope at a kink or corner in the graph because the left and right tangents have different slopes (see picture below).

Example 4.1.3. The absolute value function is $f(x) = |x|$. As we have discussed it is really a piece-wise defined function. We have

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

It turns out that this function has a kink at zero where it changes from a negative slope to a positive slope. This means that the difference quotient has different left and right limits at zero. In particular,

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

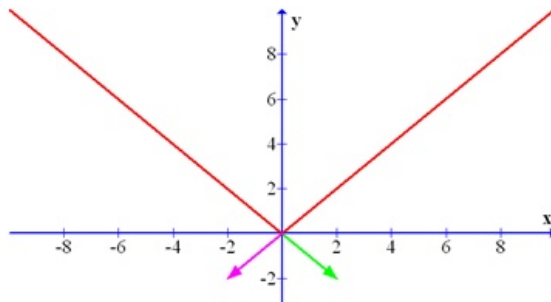
Notice that we replace $|h|$ with h because in this left limit we allow values to the left of zero on the number line, those are negative numbers. Similarly,

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

Therefore we can conclude,

$$f'(0) = \lim_{h \rightarrow 0} \frac{f(0+h) - f(0)}{h} \lim_{h \rightarrow 0} \frac{|h|}{h} = d.n.e.$$

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. A popular example of that is $f(x) = \sqrt{x}$. If you look at the graph the tangent line is vertical. Vertical lines do not have a well-defined slope².

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.

Proposition 4.1.4.

If $f'(a)$ exists for a function f then $\lim_{x \rightarrow a} f(x) = f(a)$. In other words, differentiability of f at a implies continuity of f at a .

Proof: Now, our goal is to show that $\lim_{h \rightarrow 0} f(a+h) = f(a)$ since the substitution lemma stated $\lim_{h \rightarrow 0} f(a+h) = \lim_{x \rightarrow a} f(x)$. We are given that $f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ and clearly $\lim_{h \rightarrow 0} h = 0$ so we have two limits which exist. Consider then

$$0 = \lim_{h \rightarrow 0} (h) \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \rightarrow 0} h \frac{f(a+h) - f(a)}{h} = \lim_{h \rightarrow 0} (f(a+h) - f(a))$$

Thus, $\lim_{h \rightarrow 0} (f(a+h)) = \lim_{h \rightarrow 0} (f(a))$ hence $\lim_{h \rightarrow 0} f(a+h) = f(a)$. \square

Notice that the equality above only holds true because we know the limit of the difference quotient exists. In the case of a function like $f(x) = |x|$ the limit of that product is not necessarily the product of the limits; remember $\lim f g = \lim f \lim g$ only if both $\lim f$ and $\lim g$ exist.

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.

4.2 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 4.2.1. *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,

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

We also may use the notation $f' = df/dx = \frac{df}{dx}$. Let $U \subseteq \mathbb{R}$. When a function 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.

Example 4.2.2. Suppose $f(x) = \sqrt{x}$. Calculate $f'(x)$ directly from the definition. By definition,

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

In other notation,

$$\frac{df}{dx} = \frac{1}{2\sqrt{x}} \quad \text{or} \quad \frac{d}{dx} [\sqrt{x}] = \frac{1}{2\sqrt{x}}.$$

The notation $\frac{d}{dx}$ is read "derivative with respect to x " or often in conversation I'll just say "d-d-x of square root of x minus three is one over two times the square root of x minus three". I'm sure glad we have algebraic

language for calculation, good notation is the starting point of much good mathematics. Note that $f'(x)$ is continuous on $(0, \infty)$ and we may rightly conclude that $f \in C^1(0, \infty)$.

Let me make a few more comments about our calculation: notice that the h is not present in the final answer, the h is a variable which should go away after we complete the limiting process. Moreover, note that the difference quotient was an indeterminate form of type $0/0$ for most of the calculation. If you think about it that will almost always be the case for any derivative calculation. For this reason the limit must be defined carefully. The indeterminate form case is not the exception to the rule, rather it is the primary case of interest to differential calculus. The purpose of this example is not that you should always calculate from the definition. I merely include it to illustrate the definition explicitly, the purpose of similar homework problems is the same. Obviously if you are instructed to calculate from the definition then you must do such, but if allowed to use properties and power rules then you would be foolish to use the definition. That said, we must use the definition for now since that's all we know³.

Example 4.2.3. Suppose $f(x) = \frac{1}{x^2}$. Calculate $f'(x)$ directly from the definition, assume $x \neq 0$. By definition,

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

In other notation,

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

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,

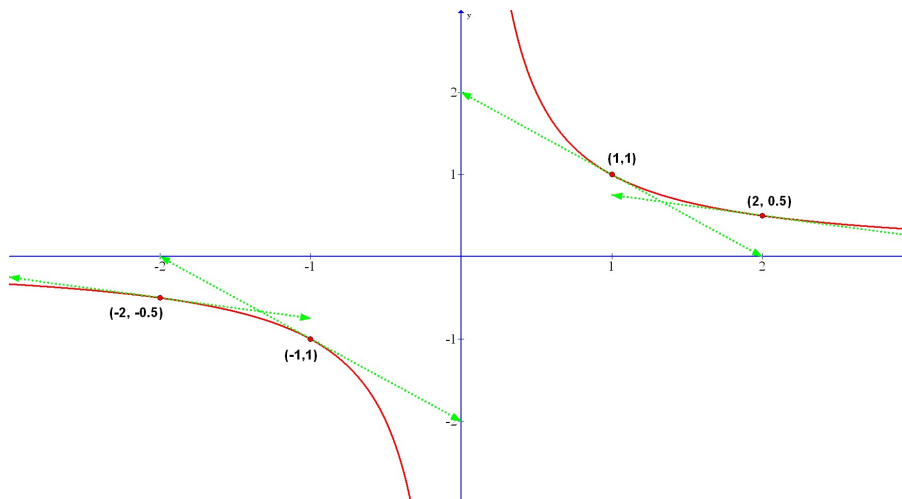
$$f'(-2) = -1/9 \quad f'(-1) = -1 \quad f'(1) = -1 \quad f'(2) = -1/4$$

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:

³sorry if you remember the "easy" way from highschool, we have not earned the right to use such short-cut formulae yet.



4.3 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 4.3.1.

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

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

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

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

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

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

Likewise, homogeneity:

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

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. \square

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.

4.3.1 derivative of a constant

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

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

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.

4.3.2 derivative of identity function

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

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

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.

4.3.3 derivative of quadratic function

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

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

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

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

4.3.4 derivative of cubic function

Let $f(x) = x^3$ for all $x \in \mathbb{R}$,

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

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

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

4.3.5 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 4.3.2. *power rule*

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

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

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*,

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

The symbol $\binom{n}{k} \equiv \frac{n(n-1)(n-2)\dots(n-k+1)}{k(k-1)\dots 3 \cdot 2 \cdot 1}$ 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. If you look in my college algebra notes you'll find some examples of how to use Pascal's triangle to multiply things like $(x+y)^7$ quickly.

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

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

This proof is no good if $n = 1/2$ since we have no binomial theorem in that case⁴. However, we proved in Example 4.2.2 that $\frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}}$. In other words, $\frac{d}{dx}(x^{\frac{1}{2}}) = \frac{1}{2}x^{\frac{1}{2}-1}$ (power rule works). You should also note we also proved the case $n = -1$ in Example 4.2.3. In fact, the power rule is still true in the case that $n \in \mathbb{R} - \mathbb{N}$ ⁵, we just need another method of proof. I will give the general proof towards the end of this chapter.

⁴We will learn in a later calculus course that the binomial expansion has infinitely many terms when $n \notin \mathbb{N}$.

⁵for example, $\frac{d}{dy}(y^{\pi+2}) = (\pi+2)y^{\pi+1} \approx 5.142y^{4.142}$

Example 4.3.3. Using the power rule correctly mostly boils down to you having a good grasp of laws of exponents.

$$\frac{d}{dx}(xx^4) = \frac{d}{dx}(x^5) = 5x^4$$

Example 4.3.4. We can use linearity in conjunction with the power rule for added fun,

$$\frac{d}{dx} \left[\frac{3x^3}{x} + \sqrt{4x} \right] = 3 \frac{d}{dx}(x^2) + 2 \frac{d}{dx}(\sqrt{x}) = 3(2x) + \frac{2}{2\sqrt{x}} = 6x + \frac{1}{\sqrt{x}}.$$

Example 4.3.5. Sometimes the independent variable is not "x", rather t, or c or even μ

$$\frac{d}{dt}(ttt) = \frac{d}{dt}(t^3) = 3t^2.$$

$$\frac{d}{dc}(c) = 1.$$

$$\frac{d}{d\mu}(\mu^k) = k\mu^{k-1}.$$

Proposition 4.3.6. *extended linearity.*

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

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

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 reader in homework exercise.

Example 4.3.7.

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

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

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

Why didn't I include a $3c^2$ in the answer above?

Example 4.3.8. We will find other ways to do this one later, but now algebra is our only hope.

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

Example 4.3.9. What is the slope of the line $y = mx + b$ at the point $(x_o, mx_o + b)$? Consider that,

$$\frac{d}{dx}(mx + b) = m \frac{dx}{dx} + 0 = m.$$

We find that the slope of the function $f(x) = mx + b$ is the same at all points along the line, it is simply m . This is good news, it verifies that there is no disagreement between our new calculus-based definition of the slope and the old standard definition we used in algebra and precalculus. Guess what the tangent line to the line is?

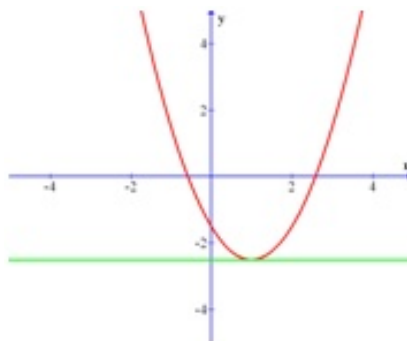
$$y = f(x_o) + f'(x_o)(x - x_o) = mx_o + b + m(x - x_o) = mx + b.$$

Of course graphically this is obvious, but it is nice to see the algebra works out.

Example 4.3.10. What is the slope of $y = f(x) = ax^2 + bx + c$ at the point $(x, f(x))$? Lets calculate the derivative at x ,

$$f'(x) = \frac{d}{dx}(ax^2 + bx + c) = 2ax + b.$$

We see that a parabola will have different slopes at different points. Where is the slope zero? Well we can just set $2ax + b = 0$ and solve to find $x = -b/2a$. If you are familiar with the formulas from algebra for the vertex of a parabola you'll recall that $h = -b/2a$ which makes a lot of sense. The vertex will have a horizontal tangent line.



What is the equation of the tangent line at x_o ? The derivative at x_o is $f'(x_o) = 2ax_o + b$. Therefore, the equation of the tangent line is

$$\begin{aligned} y &= f(x_o) + f'(x_o)(x - x_o) \\ &= ax_o^2 + bx_o + c + (2ax_o + b)(x - x_o). \end{aligned} \tag{4.1}$$

Why did I avoid asking you what the tangent line was at $(x, f(x))$? (subtle)

4.4 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.

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

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 4.4.1. *Euler's number; e .*

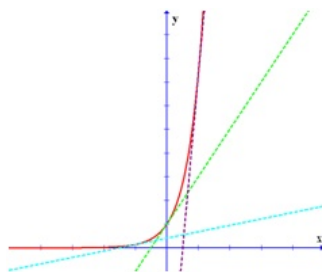
The number e is the real number such that

$$\lim_{h \rightarrow 0} \left(\frac{e^h - 1}{h} \right) = 1.$$

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

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

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 4.4.2.

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. We could 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 essentially what we did in this section.
2. The following limit is a more direct description of what the value of e is,

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

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. (there is another argument in Stewart's calculus on page 444 which is closely related). This definition arises naturally in the study of repeated multiplication, or 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 $e = 2.71\dots$. 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 $\lim_{h \rightarrow 0} \frac{a^h - 1}{h}$ 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.

4.5 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:

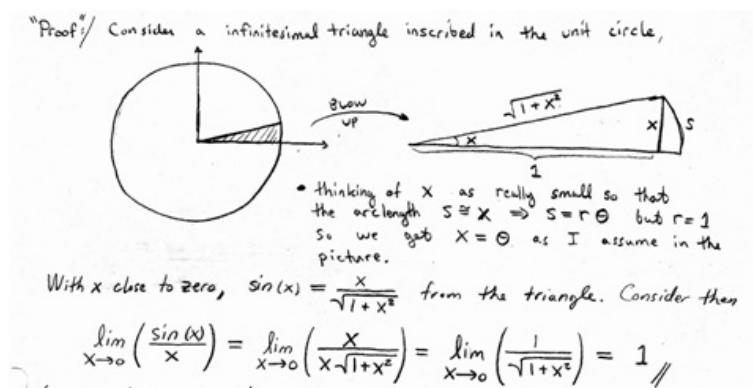
$$\lim_{x \rightarrow 0} \left(\frac{\sin(x)}{x} \right) = 1.$$

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

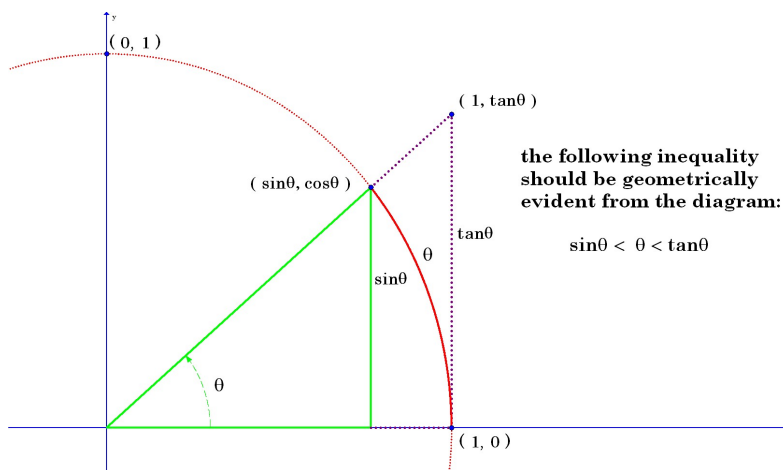
$$\lim_{x \rightarrow 0^-} \frac{\sin(x)}{x} = \lim_{x \rightarrow 0^-} \frac{\sin(-x)}{-x} = \lim_{y \rightarrow 0^+} \frac{\sin(y)}{y}$$

where in the last step we made the substitution $y = -x$ which naturally changes the left-limit of $x \rightarrow 0^-$ to the right limit $y \rightarrow 0^+$. If you require a more formal proof of this substitution rule then you should think about the composition of limits rule. Composition of limits justifies substitutions like this and others. I'll give two arguments which show $\lim_{x \rightarrow 0^+} \frac{\sin(x)}{x} = 1$. One intuitive, the other with mathematical rigor.

Intuitive Proof: imagine a tiny triangle in the unit circle. In such a case the arclength subtended and the vertical leg of the triangle are \approx equal and the limit follows:



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 $\lim_{\theta \rightarrow 0^+} \cos(\theta) = 1$ and $\lim_{\theta \rightarrow 0^+} 1 = 1$ hence by the squeeze theorem it follows that $\lim_{\theta \rightarrow 0^+} \frac{\sin(\theta)}{\theta} = 1$. \square

Some mathematicians would perceive a gap in the argument above. However, other mathematicians would accept this sort of minor geometric leap as a reasonable proof step. Rest assured that you can prove the red arclength is indeed sandwiched between $\sin(\theta)$ and $\tan(\theta)$. If you would like to see a more detailed argument feel free to consult the Appendix in Stewart. I learned the argument above from Dr. Honore Mavinga and it is found in many calculus texts, it's better than my initial physicsy argument since it only makes obvious statements about a finite triangle. The limiting is all handled by the squeeze theorem. In contrast my physicsy "proof" selectively ignored certain infinitesimals while emphasizing others.

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

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

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: $\sin(a \pm b) = \sin(a)\cos(b) \pm \sin(b)\cos(a)$

and $\cos(a \pm b) = \cos(a) \cos(b) \mp \sin(a) \sin(b)$,

$$\begin{aligned}
 \frac{d}{dx}(\sin(x)) &= \lim_{h \rightarrow 0} \left(\frac{\sin(x+h) - \sin(x)}{h} \right) \\
 &= \lim_{h \rightarrow 0} \left(\frac{\sin(x) \cos(h) + \sin(h) \cos(x) - \sin(x)}{h} \right) \\
 &= \lim_{h \rightarrow 0} \left(\sin(x) \cdot \frac{\cos(h) - 1}{h} + \cos(x) \cdot \frac{\sin(h)}{h} \right) \\
 &= \sin(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\cos(h) - 1}{h} \right) + \cos(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right) \\
 &= \sin(x) \cdot 0 + \cos(x) \cdot 1 \\
 &= \cos(x).
 \end{aligned}$$

I think it is interesting that we had to use both of the limits we just found.

$$\begin{aligned}
 \frac{d}{dx}(\cos(x)) &= \lim_{h \rightarrow 0} \left(\frac{\cos(x+h) - \cos(x)}{h} \right) \\
 &= \lim_{h \rightarrow 0} \left(\frac{\cos(x) \cos(h) - \sin(h) \sin(x) - \cos(x)}{h} \right) \\
 &= \lim_{h \rightarrow 0} \left(\cos(x) \cdot \frac{\cos(h) - 1}{h} - \sin(x) \cdot \frac{\sin(h)}{h} \right) \\
 &= \cos(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\cos(h) - 1}{h} \right) - \sin(x) \cdot \lim_{h \rightarrow 0} \left(\frac{\sin(h)}{h} \right) \\
 &= \cos(x) \cdot 0 - \sin(x) \cdot 1 \\
 &= -\sin(x).
 \end{aligned}$$

I think you will agree with me that these were harder to derive than the power rule. The neat thing is that armed with the few basic derivatives we have derived so far we will be able to differentiate just about anything once we learn a few more tools such as the product, quotient and chain rules. Barring the derivation of those rules this will be one of the last times we use the definition of the derivative to calculate a derivative. You see ultimately our goal is to calculate things without doing these tiresome limits. What I find really interesting is that after we get further into the subject we can make the limits disappear. Now, don't misunderstand me here. The limiting concept is important. There are even certain applications where you don't even have a formula for the function, all you have is raw data from some experiment. In those sort of cases you might need to apply the definition directly through some numerical methods. In this course we are mostly interested with those less interesting problems which allow pen and paper solutions. So-called analytic problems. Ok, enough philosophy of calculus, let's get back to work.

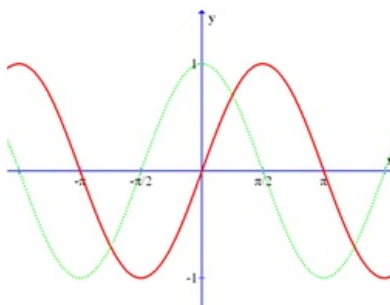
To summarize this section so far it's pretty simple,

Proposition 4.5.1. *derivatives of (radian-based) sine and cosine.*

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

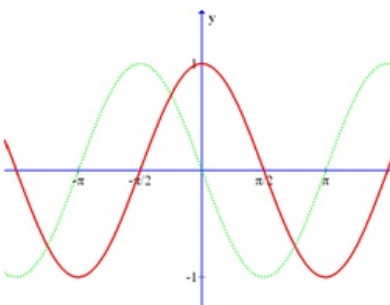
The function called "sine" for degree measure of angles is not the same function as the "sine" for radian-measured angle. We can relate them by a simple conversion: $\sin(\theta) = \sin_{degrees}(\frac{180\theta}{\pi})$. For example, $\sin(\pi/2) = \sin_{degrees}(90)$. Even your calculator knows these 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.

4.6 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 4.6.1. *product rule.*

Let f and g be differentiable functions then

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

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

Proof: start with the definition of the derivative and then after a sneaky step or two we'll have it.

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

I added zero in the third line, a very sneaky move. Then in the next to last step I pulled out $f(x)$ which is sensible since it does not depend on h . Then in the very last step I used that $\lim_{h \rightarrow 0} g(x+h) = g(x)$ which is true since g is a continuous function. I know g is continuous at x as we were given $g'(x)$ exists and differentiability at x implies continuity at x for g . \square

Example 4.6.2. *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 a mathematical truth, it helps when doubt ensues.

Example 4.6.3. *Identify that in the problem that follows $f(x) = x$ and $g(x) = e^x$ thus by the product rule,*

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

Example 4.6.4. *Observe that $f(x) = \sin(x)$ and $g(x) = \cos(x)$ so by the product rule,*

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

You might wonder what happens if we have a product of three things, suppose that are differentiable then,

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

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

Example 4.6.5.

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

Example 4.6.6. You can combine the product rule with linearity,

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

4.7 quotient rule

Proposition 4.7.1. *product 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: This rule actually follows from the product rule. Let $Q(x) = f(x)/g(x)$ then since $g(x) \neq 0$ it follows that $f(x) = Q(x)g(x)$. That's a product so we can use the product rule; $f' = (Qg)' = Q'g + Qg'$. Solve this for Q' ,

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

But, $Q' = (f/g)'$ so the proof is complete. \square

Example 4.7.2. We already know the derivatives of sine and cosine, with the help of the quotient rule we can differentiate the tangent function.

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

This is the secant function squared. I expect you to remember this derivative. You are of course free to derive it if you have time.

Example 4.7.3.

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

Example 4.7.4.

$$\frac{d}{dx} \left(\frac{1}{3x + 5} \right) = \frac{0(3x + 5) - 1(3)}{(3x + 5)^2} = \frac{-3}{(3x + 5)^2}.$$

Example 4.7.5. The reciprocal trigonometric functions' derivatives all follow from the quotient rule,

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

Likewise the derivative of the cosecant follows from the quotient rule

$$\begin{aligned}\frac{d}{dx}(\csc(x)) &= \frac{d}{dx} \left(\frac{1}{\sin(x)} \right) \\ &= \frac{\frac{d}{dx}(1) \sin(x) - 1 \frac{d}{dx}(\sin(x))}{\sin^2(x)} \\ &= \frac{-\cos(x)}{\sin^2(x)} \\ &= -\frac{1}{\sin(x)} \frac{\cos(x)}{\sin(x)} \\ &= -\csc(x) \cot(x).\end{aligned}$$

Example 4.7.6. *the quotient rule is used in conjunction with other rules sometimes, here I use linearity to start,*

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

The last couple lines were just algebraic simplification, the most important thing here was that you understood how the quotient rule was applied.

4.8 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.

$$f'(x) = \frac{df}{dx} = \frac{df}{dx}(x) = \frac{df}{dx} \Big|_x$$

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

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

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

$$\frac{df}{dx}(u) = \frac{df}{dx} \Big|_u = 2u$$

Proposition 4.8.1. *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

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

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 4.8.2. 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 $f(x) = x^5$.

$$\frac{d}{dx}(3x + 7)^5 = \frac{df}{dx} \Big|_{3x+7} \frac{d}{dx}(3x + 7) \quad (4.2)$$

$$= 5x^4 \Big|_{3x+7} \cdot 3 \quad (4.3)$$

$$= 15(3x + 7)^4 \quad (4.4)$$

$$(4.5)$$

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.

Example 4.8.3.

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

Example 4.8.4.

$$\begin{aligned}
 \frac{d}{dx}(\exp(3x^2 + x)) &= \frac{d}{dx}(\exp(u)) \\
 &= \exp(u) \frac{du}{dx} \\
 &= \exp(3x^2 + x) \frac{d}{dx}(3x^2 + x) \\
 &= (6x + 1) \exp(3x^2 + x).
 \end{aligned}$$

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 $u'(x) = \lim_{h \rightarrow 0} \left(\frac{u(x+h) - u(x)}{h} \right)$ and we can rewrite the l.h.s. in terms of a matching limit $u'(x) = \lim_{h \rightarrow 0} \left(\frac{hu'(x)}{h} \right)$. Thus

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

This shows that if $h \rightarrow 0$ 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 $h \rightarrow 0$). Consider then,

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

So the proof of the chain rule relies on approximating both the inside and outside function by their tangent line. Let's get back to the examples.

Example 4.8.5.

$$\frac{d}{dx}(e^{\sqrt{x}}) = \frac{d}{dx}(e^u) = e^u \frac{du}{dx} = e^{\sqrt{x}} \frac{d}{dx}(\sqrt{x}) = e^{\sqrt{x}} \frac{1}{2\sqrt{x}}$$

Example 4.8.6. Let a be a constant,

$$\frac{d}{dx}(\sin(ax)) = \frac{d}{dx}(\sin(u)) = \cos(u) \frac{du}{dx} = \cos(ax) \frac{d}{dx}(ax) = a \cos(ax).$$

Example 4.8.7. 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 4.8.8. 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.

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

Example 4.8.9. Let a, b, c be constants,

$$\begin{aligned} \frac{d}{dx} \left(\sqrt{ax^2 + bx + c} \right) &= \frac{1}{2\sqrt{ax^2 + bx + c}} \cdot \frac{d}{dx}(ax^2 + bx + c) \\ &= \frac{2ax + b}{2\sqrt{ax^2 + bx + c}}. \end{aligned}$$

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.

Example 4.8.10.

$$\begin{aligned} \frac{d}{dx} \left(\sqrt{x^2 + \sqrt{x^2 + 3}} \right) &= \frac{1}{2\sqrt{x^2 + \sqrt{x^2 + 3}}} \cdot \frac{d}{dx} \left(x^2 + \sqrt{x^2 + 3} \right) \\ &= \frac{1}{2\sqrt{x^2 + \sqrt{x^2 + 3}}} \left(2x + \frac{1}{2\sqrt{x^2 + 3}} \frac{d}{dx}(x^2 + 3) \right) \\ &= \frac{1}{2\sqrt{x^2 + \sqrt{x^2 + 3}}} \left(2x + \frac{x}{\sqrt{x^2 + 3}} \right). \end{aligned}$$

Example 4.8.11. Let a, b, c be constants,

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

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.

Example 4.8.12.

$$\begin{aligned} \frac{d}{dx}(x^3 e^{2x} \cos(x^2)) &= \frac{d}{dx}(x^3) e^{2x} \cos(x^2) + x^3 \frac{d}{dx}(e^{2x}) \cos(x^2) + x^3 e^{2x} \frac{d}{dx}(\cos(x^2)) \\ &= 3x^2 e^{2x} \cos(x^2) + x^3 e^{2x} \frac{d(2x)}{dx} \cos(x^2) + x^3 e^{2x} (-\sin(x^2) \frac{d(x^2)}{dx}) \\ &= 3x^2 e^{2x} \cos(x^2) + 2x^3 e^{2x} \cos(x^2) - 2x^4 e^{2x} \sin(x^2). \end{aligned}$$

We can rearrange this expression using $\sin^2(x^2) = 1 - \cos^2(x^2)$

$$\frac{d}{dx}(x^3 e^{2x} \cos(x^2)) = x^2 e^{2x} \left(\cos(x^2)[3 + 2x + 2x^2] - 2x^2 \right).$$

Example 4.8.13.

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

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

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

Example 4.8.14.

$$\begin{aligned}
\frac{d}{d\theta} \left(\frac{\sin(3\theta)}{\sqrt{\theta+4}} \right) &= \frac{3 \cos(3\theta) \sqrt{\theta+4} - \sin(3\theta) \frac{1}{2\sqrt{\theta+4}}}{(\sqrt{\theta+4})^2} \\
&= \frac{3 \cos(3\theta) \sqrt{\theta+4} \sqrt{\theta+4} - \sin(3\theta) \frac{\sqrt{\theta+4}}{2\sqrt{\theta+4}}}{(\sqrt{\theta+4})^3} \\
&= \frac{6(\theta+4) \cos(3\theta) - \sin(3\theta)}{2(\theta+4)^{\frac{3}{2}}}.
\end{aligned}$$

Example 4.8.15. Observe we can derive the power rule from the product rule.

$$\begin{aligned}
\frac{d}{dx}(x^n) &= \frac{d}{dx}(xx \cdots x) = \frac{dx}{dx} x^{n-1} + x \frac{dx}{dx} x^{n-2} + \cdots + x^{n-1} \frac{dx}{dx} \\
&= x^{n-1} + x^{n-1} + \cdots + x^{n-1} \\
&= nx^{n-1}.
\end{aligned}$$

Example 4.8.16.

$$\begin{aligned}
\frac{d}{dt}(\sin(\sqrt{2t-1})) &= \cos(\sqrt{2t-1}) \frac{d(\sqrt{2t-1})}{dt} \\
&= \cos(\sqrt{2t-1}) \frac{1}{2\sqrt{2t-1}} \frac{d(2t-1)}{dt} \\
&= \frac{\cos(\sqrt{2t-1})}{\sqrt{2t-1}}.
\end{aligned}$$

Example 4.8.17.

$$\begin{aligned}
\frac{d}{dt}(t^2 \cos(\sin(t))) &= 2t \cos(\sin(t)) + t^2 (-\sin(\sin(t))) \frac{d}{dt}(\sin(t)) \\
&= 2t \cos(\sin(t)) - t^2 \sin(\sin(t)) \cos(t).
\end{aligned}$$

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 4.8.18. 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 4.8.19. 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 $x \neq 0$). The position vector is $\vec{r} = (x, y)$ and velocity vector is $\vec{v} = (\frac{dx}{dt}, \frac{dy}{dt})$. The **dot-product** is

$$\vec{r} \cdot \vec{v} = (x, y) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right) = x\frac{dx}{dt} + y\frac{dy}{dt}$$

We will learn that when $\vec{r} \cdot \vec{v} = 0$ the vectors \vec{r} and \vec{v} are perpendicular. So the equation we found involving dx/dt and dy/dt expresses that particles traveling in a circle have velocity vectors which are tangent to the circle. (Tangents to a circle meet radial vectors at right angles)

4.9 higher derivatives

Higher derivatives are defined iteratively.

Definition 4.9.1. the n -th derivative of a function.

Suppose $f : \text{dom}(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ and $U \subseteq \text{dom}(f)$. We define $f^{(0)}(x) = f(x)$ and $f^{(1)}(x) = \frac{df}{dx}$ for all such $x \in \text{dom}(f)$ 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 $x \in \text{dom}(f)$ that $f^{(n+1)}(x) \in \mathbb{R}$. If f has continuous derivatives $f', f'', \dots, f^{(k)}$ on $U \subseteq \text{dom}(f)$ then $f \in C^k(U)$. If we can take arbitrarily many derivatives of f and those derivatives are continuous on $U \subseteq \text{dom}(f)$ 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} .

Example 4.9.2. Suppose $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$

$$\frac{d}{dx}[f(x)] = 5x^4 + 4x^3 + 3x^2 + 2x + 1$$

$$\frac{d}{dx}\left[\frac{d}{dx}[f(x)]\right] = \frac{d}{dx}[5x^4 + 4x^3 + 3x^2 + 2x + 1] = 20x^3 + 12x^2 + 6x + 2$$

$$\frac{d^3}{dx^3}[f(x)] = \frac{d}{dx}[20x^3 + 12x^2 + 6x + 2] = 60x^2 + 24x + 6$$

$$\frac{d^4}{dx^4}[f(x)] = \frac{d}{dx}[60x^2 + 24x + 6] = 120x + 24$$

$$\frac{d^5}{dx^5}[f(x)] = \frac{d}{dx}[120x + 24] = 120$$

Note that $f^{(k)}(x) = 0$ for all $k \geq 6$. It follows that $f \in C^\infty(\mathbb{R})$.

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 great importance. For mechanics only a few derivatives are typically required.

Example 4.9.3. Suppose $s : \mathbb{R} \rightarrow \mathbb{R}$ 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)

$$v(t) = \frac{ds}{dt} = \dot{s}.$$

The second derivative with respect to time is called the **acceleration** at time t and it is defined by

$$a(t) = \frac{d^2s}{dt^2} = \ddot{s}.$$

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 $F_{\text{net}} = ma$ 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 $j(t) = \frac{da}{dt} = \frac{d^3s}{dt^3}$. If you've ever rode a train and transitioned from a straight track to a half-circle track then you've felt the jerk. The jerk will be big if there is a discontinuity in the acceleration. We cannot complete these thoughts this semester. To correctly discuss mechanics we need three dimensional mathematics. We need vectors⁶.

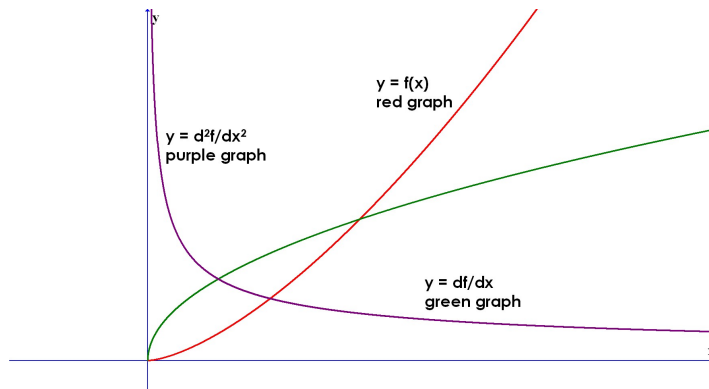
Example 4.9.4. How many times is $f(x) = x^{\frac{3}{2}}$ differentiable at zero⁷? Calculate,

$$f'(x) = \frac{3}{2}x^{\frac{1}{2}}, \quad f''(x) = \frac{3}{4}x^{-\frac{1}{2}}$$

⁶there is no reason you can't learn vectors now. It is merely the custom of mathematics departments to make you wait until calculus III to see the neat part. One could envision a calculus education which simultaneously implemented the first third of calculus III throughout the earlier topics. True multivariate calculus really only begins when you discuss multivariate functions in my mind. The calculus of space curves is just calculus I together with vectors.

⁷In such a case we ask only that the right limit of the difference quotient exists. We define that $f'(0) = \lim_{h \rightarrow 0^+} \frac{f(0+h) - f(0)}{h}$ in the case $0 \in \partial(\text{dom}(f))$

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.



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 $x > 0$.

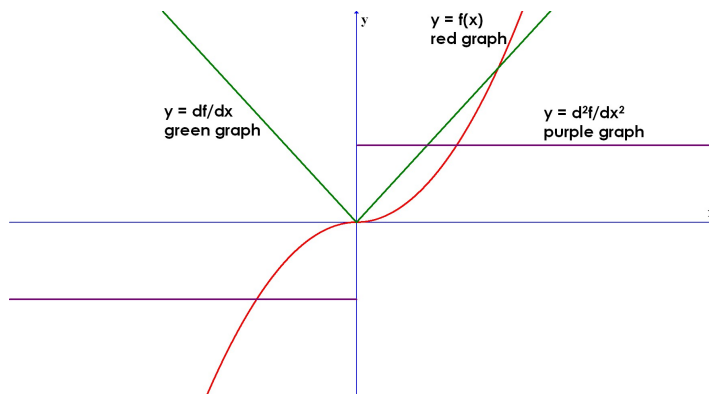
Example 4.9.5. Another interesting function which fails to be smooth is $f(x) = x|x|$. The graph resembles a cubic function but it is actually a pair of half-parabolas glued at the origin. For $x > 0$ we have $f(x) = x^2$ and for $x < 0$ we have $f(x) = -x^2$. It follows that

$$f'(x) = \begin{cases} 2x & x \geq 0 \\ -2x & x \leq 0 \end{cases}$$

In this case $f'(0) = 0$ since $\lim_{h \rightarrow 0^-} \frac{f(0+h)-f(0)}{h} = \lim_{h \rightarrow 0^+} \frac{f(0+h)-f(0)}{h} = 0$. 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 $\lim_{h \rightarrow 0^-} \frac{f(0+h)-f(0)}{h} = -2$ whereas $\lim_{h \rightarrow 0^+} \frac{f(0+h)-f(0)}{h} = 2$. 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.

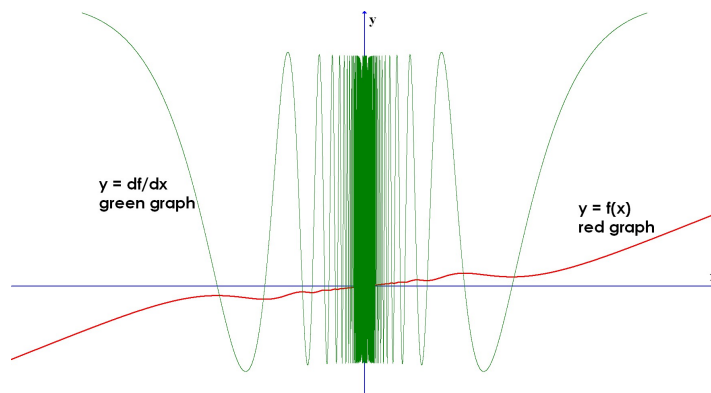
Example 4.9.6. *I found this example in Hubbard's advanced calculus text (see Ex. 1.9.4, pg. 123). It is a source of endless odd examples, notation and bizarre quotes. Let $f(x) = 0$ and*

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

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 $\text{dom}(f') = \mathbb{R}$. 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 ⁸

⁸Continuity 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 lack of continuous differentiability usually signifies a lack of a well-defined tangent space. I discuss this in some depth in the advanced calculus course

4.10 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 4.10.1. 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 .

Example 4.10.2. Observe that the equation $xy + \sin(x) = e^{xy}$ implicitly defines y as a function of x . Let's find $\frac{dy}{dx}$.

$$\begin{aligned} \frac{d}{dx}(xy + \sin(x)) &= \frac{d}{dx}(e^{xy}) \\ \Rightarrow \frac{dx}{dx}y + x \frac{dy}{dx} + \cos(x) &= e^{xy} \frac{d}{dx}(xy) \\ \Rightarrow y + x \frac{dy}{dx} + \cos(x) &= e^{xy} \left(y + x \frac{dy}{dx} \right) \end{aligned}$$

Now solve for $\frac{dy}{dx}$,

$$y + \cos(x) - ye^{xy} = (xe^{xy} - x) \frac{dy}{dx} \Rightarrow \boxed{\frac{dy}{dx} = \frac{y + \cos(x) - ye^{xy}}{xe^{xy} - x}}.$$

You might question why such differentiation is interesting. One good reason is that it is what we use to solve related rates problems.

Example 4.10.3. 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,

$$\frac{dV}{dt} = \frac{d}{dt} \left(\frac{4\pi}{3} R^3 \right) = 4\pi R^2 \frac{dR}{dt} = 4\pi (10m)^2 \frac{m}{min} \approx 1200 \frac{m^3}{min}.$$

We'll do more of these in a later section.

I hope you get the idea about these sort of problems. I'm going to shift back to the other type of problem that implicit differentiation is great for. That is the problem of calculating the inverse function's derivative. We know the derivatives of e^x , $\cos(x)$, $\sin(x)$, $\tan(x)$, $\sec(x)$. I will now systematically derive the derivatives of $\ln(x)$, $\cos^{-1}(x)$, $\sin^{-1}(x)$, $\tan^{-1}(x)$, $\sec^{-1}(x)$ using essentially the same technique every time.

Example 4.10.4. Let $y = \ln(x)$ we wish to derive $\frac{d}{dx}(\ln(x))$. To begin we take the exponential of both sides of $y = \ln(x)$ to obtain

$$e^y = e^{\ln(x)} = x$$

Now differentiate with respect to x and solve for $\frac{dy}{dx}$,

$$e^y \frac{dy}{dx} = 1 \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{e^y}$$

Now remember that we found $e^y = x$ so we have shown that

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

Example 4.10.5. 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 $\frac{dy}{dx}$

$$-\sin(y) \frac{dy}{dx} = 1 \quad \Rightarrow \quad \frac{dy}{dx} = \frac{-1}{\sin(y)}$$

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

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

Example 4.10.6. Let $y = \sin^{-1}(x)$ we wish to derive $\frac{d}{dx}(\sin^{-1}(x))$. To begin we take the sine of both sides of $y = \sin^{-1}(x)$ to obtain

$$\sin(y) = \sin(\sin^{-1}(x)) = x$$

Now differentiate with respect to x and solve for $\frac{dy}{dx}$

$$\cos(y) \frac{dy}{dx} = 1 \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{\cos(y)}$$

Now $\sin^2(y) + \cos^2(y) = 1$ thus $\cos(y) = \sqrt{1 - \sin^2(y)}$ but remember that we found $\sin(y) = x$ so $\cos(y) = \sqrt{1 - x^2}$ thus we find

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

Example 4.10.7. Let $y = \tan^{-1}(x)$ we wish to derive $\frac{d}{dx}(\tan^{-1}(x))$. To begin we take the tangent of both sides of $y = \tan^{-1}(x)$ to obtain

$$\tan(y) = \tan(\tan^{-1}(x)) = x$$

Now differentiate with respect to x and solve for $\frac{dy}{dx}$

$$\sec^2(y) \frac{dy}{dx} = 1 \Rightarrow \frac{dy}{dx} = \frac{1}{\sec^2(y)}$$

Now $\sin^2(y) + \cos^2(y) = 1$ thus if we divide this equation by $\cos^2(y)$ we'll obtain the less familiar identity $\tan^2(y) + 1 = \sec^2(y)$. But we know that in this example $\tan(x) = y$ hence $\sec^2(y) = 1 + x^2$. To conclude,

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

Example 4.10.8. 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 $\frac{dy}{dx}$

$$\sec(y) \tan(y) \frac{dy}{dx} = 1 \Rightarrow \frac{dy}{dx} = \frac{1}{\sec(y) \tan(y)}$$

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,

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

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 4.10.9. 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 4.10.4,

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

Now solve for $\frac{dy}{dx}$,

$$\frac{dy}{dx} = \ln(a)y = \ln(a)a^x \Rightarrow \boxed{\frac{d}{dx}(a^x) = \ln(a)a^x.}$$

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 4.10.10. 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 ,

$$\frac{1}{y} \frac{dy}{dx} = \ln(x) + x \frac{1}{x} \implies \frac{dy}{dx} = y(\ln(x) + 1)$$

Therefore we find,

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

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 4.10.11. 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,

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

Notice in the case $a = e$ we have $\log_e(x) = \ln(x)$ and $\ln(e) = 1$. Therefore, this result agrees with Example 4.10.4.

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.

4.11 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.

Example 4.11.1. Find the derivative of $y = xe^{x^2+9}\sqrt{3x+7}$ using logarithmic differentiation. Take the natural log to begin,

$$\begin{aligned} \ln(y) &= \ln(xe^{x^2+9}\sqrt{3x+7}) = \ln(x) + \ln(e^{x^2+9}) + \ln(\sqrt{3x+7}) \\ &= \ln(x) + x^2 + 9 + \frac{1}{2} \ln(3x+7). \end{aligned}$$

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.

$$\frac{1}{y} \frac{dy}{dx} = \frac{1}{x} + 2x + \frac{3}{2(3x+7)}$$

$$\Rightarrow \boxed{\frac{dy}{dx} = xe^{x^2+9}\sqrt{3x+7}\left(\frac{1}{x} + 2x + \frac{3}{2(3x+7)}\right)}.$$

This is much easier than the 3-term product rule for this problem.

Example 4.11.2. 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

$$\begin{aligned}\frac{1}{y} \frac{dy}{dx} &= \frac{1}{2-x} + \frac{1}{4(x+32)} + \frac{4(2x)}{x^2-3} \\ \Rightarrow \boxed{\frac{dy}{dx} &= \left(\frac{1}{2-x}\right)(x+32)^{\frac{1}{4}}(x^2-3)^4 \left(\frac{1}{2-x} + \frac{1}{4(x+32)} + \frac{8x}{x^2-3}\right)}.\end{aligned}$$

Again, this is much easier than the 3-term product rule for this problem

Example 4.11.3. Let a, b, c be constants. Differentiate y .

$$y = \left(\frac{1}{x-a}\right)\left(\frac{1}{x-b}\right)^2\left(\frac{1}{x-c}\right)^3$$

Take the natural log to begin,

$$\ln(y) = -\ln(x-a) - 2\ln(x-b) - 3\ln(x-c)$$

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

$$\begin{aligned}\frac{1}{y} \frac{dy}{dx} &= \frac{-1}{x-a} - \frac{2}{x-b} - \frac{3}{x-c} \\ \Rightarrow \boxed{\frac{dy}{dx} &= \left(\frac{1}{x-a}\right)\left(\frac{1}{x-b}\right)^2\left(\frac{1}{x-c}\right)^3 \left(\frac{-1}{x-a} - \frac{2}{x-b} - \frac{3}{x-c}\right)}.\end{aligned}$$

Example 4.11.4. Differentiate y .

$$y = (x^2+1)(x-3)^2(x^3+x)^3(x-1)^4$$

Take the natural log to begin,

$$\begin{aligned}\ln(y) &= \ln(x^2+1) + 2\ln(x-3) + 3\ln(x^3+x) + 4\ln(x-1) \\ \Rightarrow \frac{1}{y} \frac{dy}{dx} &= \frac{2x}{x^2+1} + \frac{2}{x-3} + \frac{3(3x^2+1)}{x^3+x} + \frac{4}{x-1} \\ \Rightarrow \boxed{\frac{dy}{dx} &= y \left(\frac{2x}{x^2+1} + \frac{2}{x-3} + \frac{3(3x^2+1)}{x^3+x} + \frac{4}{x-1} \right)}.\end{aligned}$$

Example 4.11.5. 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 $y = \ln\left(\frac{\sin(x)\sqrt{x}}{x^2+3x-2}\right)$.

$$\begin{aligned} y &= \ln\left(\frac{\sin(x)\sqrt{x}}{x^2+3x-2}\right) \\ &= \ln(\sin(x)) + \frac{1}{2}\ln(x) - \ln(x^2+3x-2). \end{aligned}$$

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 4.11.6. 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

$$\frac{dy}{dx} = \frac{1}{(x+1)^{30} + 2} \frac{d}{dx} \left((x+1)^{30} + 2 \right) = \frac{30(x+1)^{29}}{(x+1)^{30} + 2}.$$

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 $\ln(A^c) = c \ln(A)$.

4.11.1 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. \square

Another method to derive rules such as $\frac{d}{dx}\left(\frac{1}{x^n}\right) = \frac{-n}{x^{n+1}}$ is apply the product rule n -times for the reciprocal function for which we have already shown $\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{-1}{x^2}$.

Example 4.11.7.

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

Example 4.11.8.

$$\frac{d}{dy}(y^{\pi+2}) = (\pi + 2)y^{\pi+1} \approx 5.142y^{4.142}$$

4.12 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	$\sinh(x) = \frac{1}{2}(e^x - e^{-x})$
$\tanh(x)$	$\text{sech}^2(x)$	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 need⁹.

⁹you can ask Dr. Wang about his numerical method course

4.13 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 4.13.1. Problem: *Imagine a circular oil slick which continuous to grow 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 $r = 10\text{m}$?*

Solution: *since the oil slick is circular we know that the area of the slick (call it A) is related to the radius of the slick (call it r) by $A = \pi r^2$. In this context both A and r are functions of time so we may differentiate implicitly to find*

$$\frac{dA}{dt} = \frac{d}{dt}(\pi r^2) = 2\pi r \frac{dr}{dt}$$

We were given that $\frac{dA}{dt} = 3\text{m}^2/2$ thus,

$$\left. \frac{dr}{dt} \right|_{r=10\text{m}} = \frac{1}{2\pi(10\text{m})} 3 \frac{\text{m}^2}{\text{s}} = \frac{3}{20\pi} \frac{\text{m}}{\text{s}} \approx \boxed{0.0159\text{m/s.}}$$

Example 4.13.2. Problem: *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 $A = 15\text{ft}^2$. We can relate the time-rate of change of V and h :*

$$\frac{dV}{dt} = A \frac{dh}{dt} \quad \Rightarrow \quad \frac{dh}{dt} = \frac{1}{A} \frac{dV}{dt}.$$

We were given $\frac{dV}{dt} = 5\text{ft}^3/\text{min}$ thus $\boxed{\frac{dh}{dt} = \frac{1}{3} \frac{\text{ft}}{\text{min}}.}$

Example 4.13.3. Problem: *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: *We should define the variables and draw a picture; call the volume of water in the trough V and the height of water in the trough h . Since the trough is triangular we have a typical vertical cross section of area $A = \frac{1}{2} \text{BaseHeight} = \frac{1}{2} \frac{h}{2} h$. The volume $V = Al$ where $l = 4$, thus $V = h^2$. We can relate the time-rate of change of V and h :*

$$\frac{dV}{dt} = \frac{d}{dt}(h^2) = 2h \frac{dh}{dt} \quad \Rightarrow \quad \frac{dh}{dt} = \frac{1}{2h} \frac{dV}{dt}.$$

We were given $\frac{dV}{dt} = 5\text{ft}^3/\text{min}$ thus $\boxed{\frac{dh}{dt} = 2.5 \frac{\text{ft}}{\text{min}}.}$

Example 4.13.4. 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 arclength 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 $\frac{ds}{dt} = (93,000,000\text{miles})\frac{\pi}{12hrs} \approx \boxed{24,300,000\text{mph.}}$

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 $v = \frac{2\pi(4000\text{miles})}{24hrs} \approx 1050\text{mph}$ (at the equator). The circumference at the equator is about 25,000 miles. In contrast, the land at the North or South poles rotates at a much slower tangential speed. For this reason the Earth is actually an *oblate spheroid*¹⁰ 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.

Example 4.13.5. Problem: 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: We label the distance from the ground to the top of the ladder to be y and the distance from the base of the wall to the base of the ladder to be x . We are given that $dy/dt = -3\text{ft/s}$. On the other hand, by the pythagorean theorem $x^2 + y^2 = 100$. Differentiating with respect to time we find

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0$$

Solve for dx/dt and note that if $x = 4\text{ft}$ then $y = \sqrt{84}$ thus

$$\frac{dx}{dt} = -\frac{y}{x}\frac{dy}{dt} = -\frac{\sqrt{84}}{4}(-3\frac{ft}{s}) \approx \boxed{6.9\frac{ft}{s}}.$$

Notice that common sense also suggests that if $dy/dt < 0$ then $dx/dt > 0$.

Example 4.13.6. 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 $x = 2\text{ft}$?

¹⁰which is basically just a squished sphere

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, $\boxed{\frac{dx}{dt} = 0.25 \text{ ft/s.}}$

Remark 4.13.7.

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 $y = 6 \text{ ft} - (\frac{1}{\text{ft}})x^2$ 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 $[x] = [y] = \text{ft}$ and $[t] = \text{s}$ in the last example (I use $[f]$ to denote the customary units of f).

Example 4.13.8. 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 $t = 1\text{hr}$?

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 $\boxed{s'(1) = 69.41 \text{ mph}}$

Remark 4.13.9.

I will likely add pictures as I lecture on this material. You can also consult Stewart §3.8 or the previous edition of my notes for additional examples.

Chapter 5

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.

5.1 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 usually¹ provides a good approximation if we are near to the point. The linearization of a function f at a point $a \in \text{dom}(f)$ is denoted by L_f^a or simply L_f in this course,

$$L_f^a(x) \equiv f(a) + f'(a)(x - a)$$

The graph of L_f^a is the tangent line to $y = f(x)$ at $(a, f(a))$.

Example 5.1.1. *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, $4.01 = 4 + 0.01$ 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,*

$$f(x) \approx L_f^4(x) = f(4) + f'(4)(x - 4) = 2 + \frac{1}{4}(x - 4).$$

¹we discuss limitations of the tangent line approximation at the conclusion of this section

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

$$\sqrt{4.01} = 2.00249843945007857276972121483226054214864513129159...$$

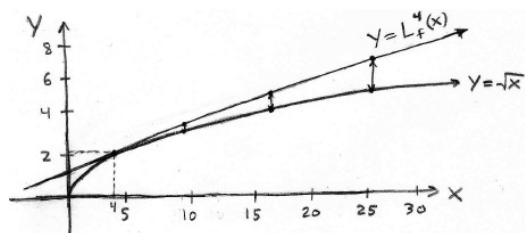
As you can easily see we did very well considering the crudeness of our method (in fact the error is only about 0.0001%). Is a line a parabola? Certainly not. But that is the heart of what I just did. I said you can replace a curve with a line locally and get good approximations. But, what is "local" 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 $a = 4$.

5.2 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 5.2.1.

Suppose $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ then if f has a derivative at a then it also has a **differential** $df_a : \mathbb{R} \rightarrow \mathbb{R}$ 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 5.2.2. Estimate the uncertainty in the volume of a cubical box if you measure the length of the side to be $20\text{in} \pm 0.2\text{in}$. 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 $a = 20\text{in}$ and $\Delta x = 0.2\text{in}$ thus,

$$\Delta V \approx 3(20)^2(0.2)\text{in}^3 = 240\text{in}^3.$$

Thus the uncertainty in the volume of the cubical box is approximately $\boxed{\pm 240\text{in}^3}$

Example 5.2.3. Suppose that we have 5 resistors all given to have a resistance of $R = 10\Omega$. Furthermore, we know that the given resistance values are known to within 5% by the resistors color coding. If the resistors are arranged so that 2 of the resistors are in series and the other 3 are in parallel with the series combination then what is the error in the equivalent resistance R_{eq} ? In second semester physics you can learn to calculate that

$$R_{eq} = \left[\frac{1}{R} + \frac{1}{R} + \frac{1}{R} + \frac{1}{R+R} \right]^{-1} = \left[\frac{7}{2R} \right]^{-1} = \frac{2R}{7}.$$

We calculate $\Delta R_{eq} = \frac{2}{7}\Delta R = 0.286\Delta R$. Since 5% of 10Ω is 0.5Ω we are given that $\Delta R = 0.5\Omega$ thus $\Delta R_{eq} = (0.286)(0.5)\Omega = 0.143\Omega$. Notice that $R = 10\Omega$ gives $R_{eq} = 2.86\Omega$ so we have $R_{eq} = 2.86 \pm 0.143\Omega$. The uncertainty in the value of the equivalent resistance is $[0.143/2.86][100\%] = 5\%$.

It should be mentioned that the total and correct analysis of error propagation is more involved than 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.

Remark 5.2.4.

It is also true that $df_a = \frac{df}{dx}(a)dx_a$, but beware this equation is not just multiplication and division of tiny variables. Let $g(x) = x$ for all $x \in \mathbb{R}$. Note that $g'(x) = 1$ and it follows that $dg_a(x) = 1 \cdot x = x$ for all $x \in \mathbb{R}$. Therefore, $dg = g$. If we denote $g = x$ so that $dx = x$ in this notation. Note then we can write the differential in terms of the derivative function:

$$df(a)(h) = df_a(h) = f'(a)h = f'(a)dx_a(h) \quad \text{for all } h \in \mathbb{R}$$

Hence $df(a) = f'(a)dx_a$ for all $a \in \mathbb{R}$ hence $df = f'dx$ or we could denote this by the deceptively simple formula $df = \frac{df}{dx}dx$.

Most elementary calculus texts give you the idea that $f'(a)$ is primary and df_a is just some application tacked on at the end, however it turns out that in the big scheme of things df_a is the primary object and it has a definition which generalizes even to infinite dimensional calculus. In general the differential is a linear operator and the derivative is the matrix of that operator. See my advanced calculus notes or Edward's excellent text for a deeper perspective on the concept of a differential.

5.3 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

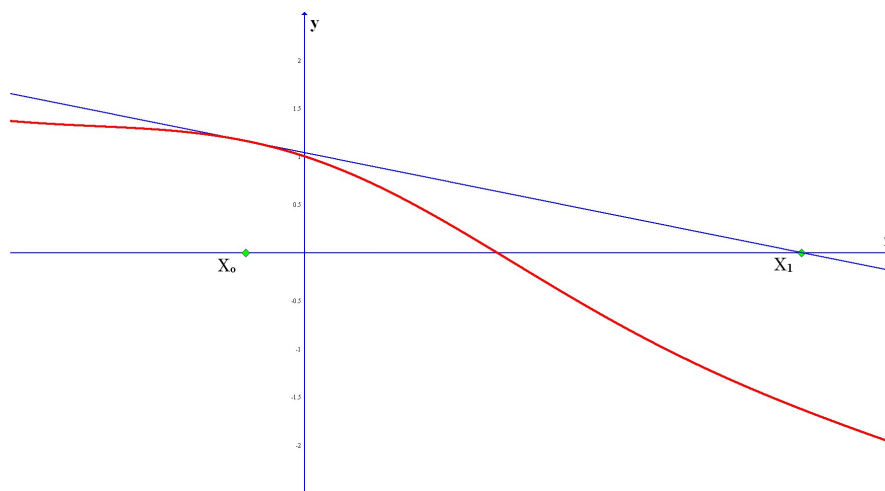
1. guess a solution x_o and calculate $f(x_o)$ and if it is close enough to zero then stop.
2. construct $L_o(x) = f(x_o) + f'(x_o)(x - x_o)$ and solve $L_o(x_1) = 0$ to find solution $x_1 = x_o - f(x_o)/f'(x_o)$.
3. calculate $f(x_1)$ and if it is close enough to zero then stop.
4. construct $L_1(x) = f(x_1) + f'(x_1)(x - x_1)$ 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 $L_2(x_3) = 0$ 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

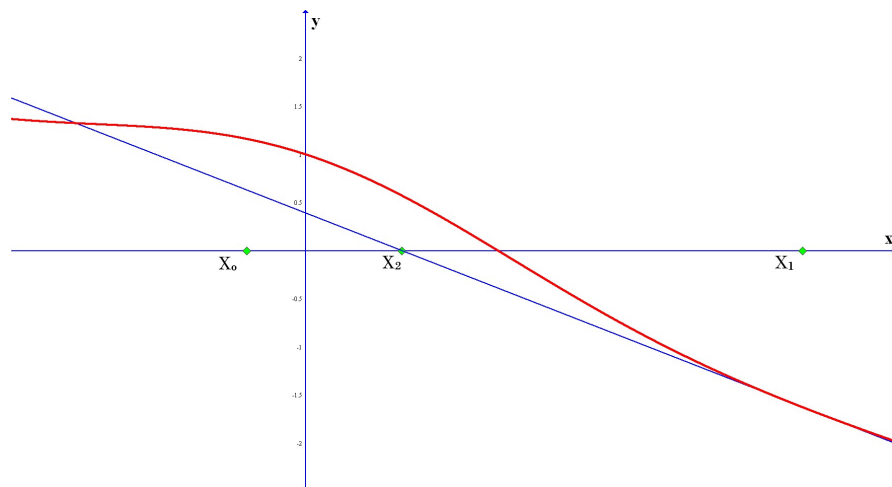
$$x_{n+1} = x_n - f(x_n)/f'(x_n)$$

until $|f(x_n)| < \epsilon$ where ϵ is an upper bound on the error you allow for the approximate solution.

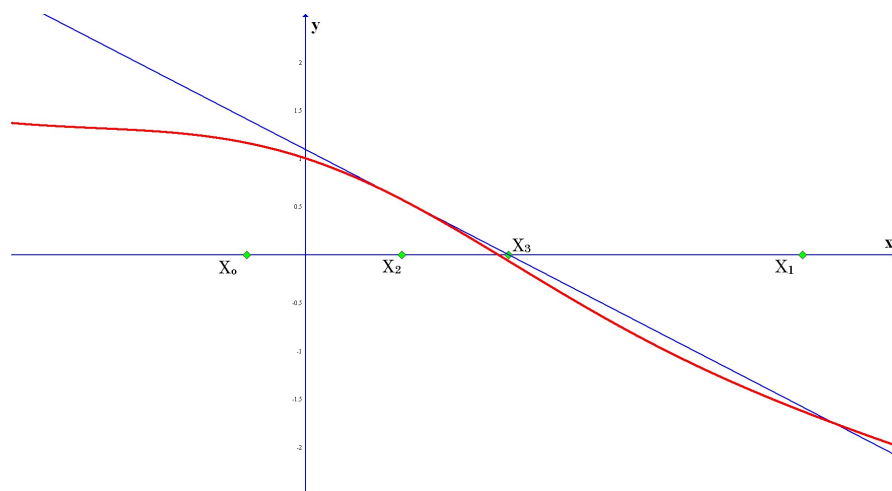
Example 5.3.1. Let's see how to solve the equation $e^{-x^2} = x$ to within ± 0.01 . First construct $f(x) = e^{-x^2} - x$ and note that the problem becomes solving $f(x) = 0$. Calculate $f'(x) = -2xe^{-x^2} - 1$. To begin we guess $x_o = -0.2$. Note $f(-0.2) \approx 1.16$. Calculate $x_1 = x_o - f(x_o)/f'(x_o) = 1.69$. I have pictured the initial guess x_o as well as the first iterate x_1 with green diamonds on the x -axis:



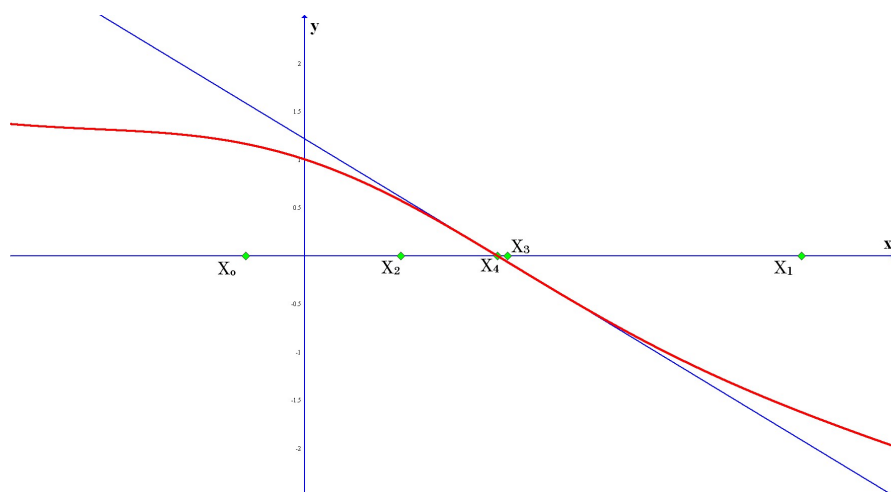
You can see that x_1 is the x -intercept of the tangent line from x_o . Next, we can calculate $x_2 = x_1 - f(x_1)/f'(x_1) = 0.324$.



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_0 .

Example 5.3.2. 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 $x_0 = 2.5$ since we know that our answer must be somewhere between 2 and 3 since $2^3 = 8$ and $3^3 = 27$. Note $f'(x) = 3x^2$.

$$x_1 = x_0 - f(x_0)/f'(x_0) = 2.5 + 4.375/18.75 = 2.733$$

$$x_2 = x_1 - f(x_1)/f'(x_1) = 2.733 - 0.4136/22.41 = 2.715$$

$$x_3 = x_2 - f(x_2)/f'(x_2) = 2.715 - 0.013/22.11 = 2.714$$

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 5.3.3. *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 be inverted. \square

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. I have a Mathematica notebook which demonstrated the algorithm from Edwards' text. 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 6

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 Theorem¹ 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.

6.1 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 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 model² of the bridge would reveal motion which is *unbounded*. The idea of bounded motion is closely connected with the following:

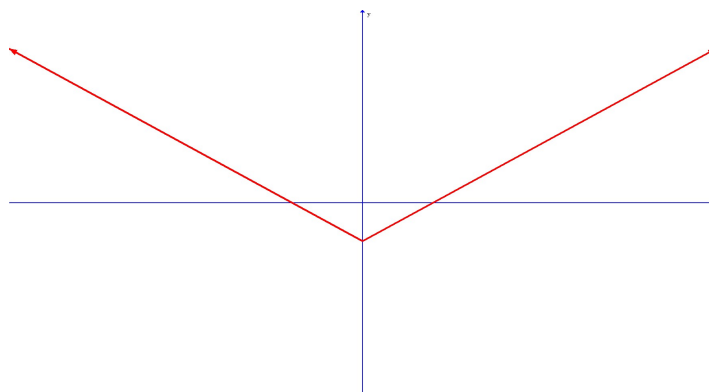
¹note we do not discuss series at this juncture, the totality of that topic waits until calculus II, I simply include some discussion here in the interest of deeper geometric insight. As a side consequence I also hope this inclusion strengthens the student for calculus II's travails

²a good mathematical model is the sort which anticipates these sort of problems before they occur

Definition 6.1.1. *absolute extrema.*

1. We say a function f has an absolute maximum at c if $f(c) \geq f(x)$ for all $x \in \text{dom}(f)$. The absolute maximum is $f(c)$ in this case.
2. We say a function f has an absolute minimum at d if $f(d) \leq f(x)$ for all $x \in \text{dom}(f)$. The absolute minimum is $f(d)$ in this case.
3. If there exist absolute maximum and minimum values then we call them the **global extrema** of f .

One subtle point is that a function could be bounded and yet no global extrema are realized. Think about it. What shape of graph would be bounded and yet have no global max. or min. ? A less subtle comment is that if a function has a VA then it is not bounded above and below and it also does not possess global a maximum and minimum (it might have just a global max. or just a min.). It can happen that function has a global minimum but not global maximum. For example the function below has $f(0) = -2$ and that is the global minimum for $f(x) = |x| - 2$.

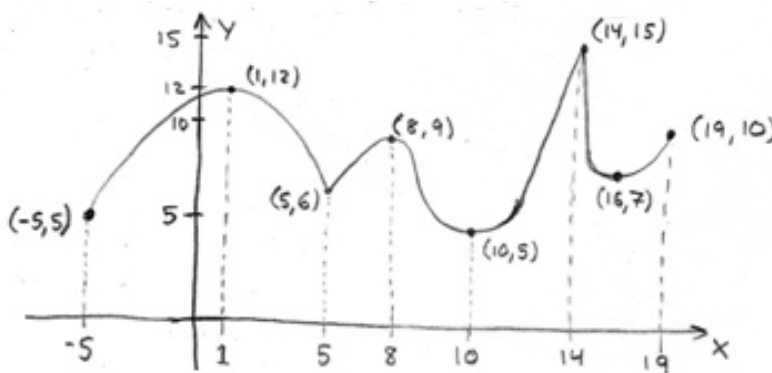


Suppose we are given a model plus some additional condition so that we know the model must have variables whose values are near some given data point. In a case such as that it is interesting to know what the largest positive or negative values the function takes near the given data point. Mathematically this is encapsulated by the idea of a local extreme value:

Definition 6.1.2. *local extrema.*

1. A function f has a **local maximum** at c if there exists a connected set J with $c \in J$ and $J \subseteq \text{dom}(f)$ such that $f(c) \geq f(x)$ for all $x \in J$. The local maximum is $f(c)$ in this case.
2. A function f has a **local minimum** at c if there exists a connected set J with $c \in J$ and $J \subseteq \text{dom}(f)$ such that $f(c) \leq f(x)$ for all $x \in J$. The local minimum is $f(c)$ in this case.
3. If $f(c)$ is either a local maximum or a local minimum then we say $f(c)$ is a **local extrema** at c .

In the graph below there are local minima at $x = -5, 5, 10, 16$ and the local maxima are at $1, 8, 14, 19$. The global maximum is 15 it is reached at $x = 14$. The global minimum is 5 and it is reached at $x = -5$ and $x = 10$.



The example above is just intended to illustrate the definitions. Usually we are not given a picture of the graph. More often we are given the formula for the function then we endeavor to discover the graph through a careful, calculus-assisted analysis.

The following theorem is at the heart of most everything that follows in this chapter.

Proposition 6.1.3. *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 6.1.4. *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 $c \in \text{dom}(f)$ is a critical number then $(c, f(c))$ is a **critical point** of f .

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 number as $f'(0)$ does not exist in \mathbb{R} . Clearly $0 \notin \text{dom}(f)$ either. It is usually the case that a vertical asymptote of the function will likewise be a vertical asymptote of the derivative function.

Proposition 6.1.5. *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 $\lim_{h \rightarrow 0} \frac{f(c+h)-f(c)}{h} = f'(c) \in \mathbb{R}$. 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 $f'(c) = \lim_{h \rightarrow 0^+} \frac{f(c+h)-f(c)}{h} \leq \lim_{h \rightarrow 0}(0) = 0$. Likewise, if $h < 0$ and $h \in B_{\delta_1}(0)$ then $f(c+h) - f(c) \leq 0$ hence,

$$\frac{f(c+h) - f(c)}{h} \geq 0$$

Using the squeeze theorem we find $f'(c) = \lim_{h \rightarrow 0^-} \frac{f(c+h)-f(c)}{h} \geq \lim_{h \rightarrow 0}(0) = 0$. 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".

The converse of this Theorem is not true. We can have a critical number c such that $f(c)$ is **not** a local maximum or minimum. For example, $f(x) = x^3$ has critical number $c = 0$ yet $f(0) = 0$ which is neither a local max. nor min. value of $f(x) = x^3$. It turns out that $(0, 0)$ is actually an *inflection point* as we'll discuss soon. Another example of a critical point which yields something funny is a constant function; if $g(x) = k$ then $g'(x) = 0$ for each and every $x \in \text{dom}(g)$. Technically, $y = k$ is both the minimum and maximum value of g . Constant functions are a sort of exceptional case in this game we are playing.

Proposition 6.1.6. *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 $f(x) = k$ for all $x \in [a, b]$ then every point is a critical point and the theorem is trivially satisfied. Suppose f is not constant, apply the Extreme Value Theorem to show there exists $c, d \in [a, b]$ such that $f(c) \geq f(x)$ for all $x \in [a, b]$ and $f(d) \leq f(x)$ for all $x \in [a, b]$. Since $f(x_o) \neq f(a)$ for at least one $x_o \in (a, b)$ it follows that $f(x_o) > f(a)$ or $f(x_o) < f(a)$. If $x_o \in (a, b)$ and $f(x_o) > f(a)$ then $f(a)$ is not the absolute maximum therefore we deduce $c \in (a, b)$ is the absolute maximum. Likewise, if $x_o \in (a, b)$ and $f(x_o) < f(a)$ then $f(a)$ is not the absolute minimum therefore we deduce $d \in (a, b)$ is the absolute maximum. In all cases there is an absolute extremum in the open set (a, b) hence there exists a critical point in the interior of the set. Moreover, since f is differentiable on (a, b) it follows that either $f'(c) = 0$ or $f'(d) = 0$ and Rolle's theorem follows. \square

Let's think about this 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.

Proposition 6.1.7. *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:(essentially borrowed from Stewart pg. 216-217, this proof is common to a host of calculus texts). The equation of the secant line to $y = f(x)$ on the interval $[a, b]$ is $y = s(x)$ where $s(x)$ is defined via the standard formula for a line going from $(a, f(a))$ to $(b, f(b))$

$$s(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

The Mean Value Theorem proposes that there is some point on the interval $[a, b]$ such that the slope of the tangent line is equal to the slope of the secant line $y = s(x)$. Consider a new function defined to be the difference of the secant line and the given function, call it h :

$$h(x) = f(x) - s(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

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

$$h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \implies \boxed{f'(c) = \frac{f(b) - f(a)}{b - a}}. \quad \square$$

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.

Proposition 6.1.8. *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 $\text{dom}(f)$

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 6.1.9. *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,

$$0 = \frac{f(b_o) - f(a_o)}{b_o - a_o} \implies f(b_o) = f(a_o)$$

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.

Theorem 6.1.10. *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 6.1.9 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 6.1.10 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 $(2, 3)$).

Physical Significance: If we have equal velocities over some time interval then the displacement between our positions at any time will be constant.

Proposition 6.1.11. *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 6.1.8 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. \square

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 4.9.6 for this bad behavior).

6.1.1 first derivative test

The following theorem naturally follows from the sign-test theorem.

Theorem 6.1.12. *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 6.1.13. 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,

$$f'(-4) = (-4 - 2)(-2 + 3) = 8 > 0 \quad (6.1)$$

$$f'(0) = (-2)(3) = -6 < 0 \quad (6.2)$$

$$f'(3) = (3 - 2)(3 + 3) = 6 > 0 \quad (6.3)$$

$$(6.4)$$

Hence the completed sign-chart,



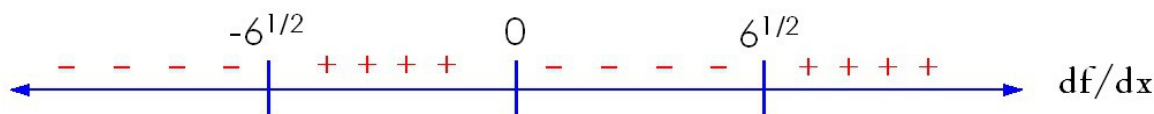
By the First Derivative Test we conclude,

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.

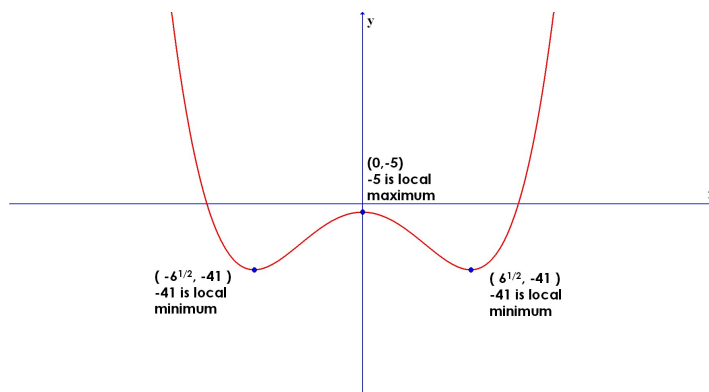
Example 6.1.14. Let $f(x) = e^x + x$. Note that $f'(x) = e^x + 1$. This function has no critical points since the equation $e^x + 1 = 0$ has no solutions. It follows that $y = e^x + 1$ has no local extrema. However, we can deduce that $f(x)$ is increasing on \mathbb{R} since $f'(x) = e^x + 1 \geq 2$ for all $x \in \mathbb{R}$.

Example 6.1.15. Observe that $f(x) = x \frac{\sqrt{(x-1)^2}}{x-1}$ for $x \neq 1$ and $f(1) = 1$ has critical number $c = 1$. Moreover, the derivative changes signs at $c = 1$ since $f'(x) = -1$ for $x < 1$ whereas $f'(x) = 1$ for $x > 1$. Is it the case that $f(1) = 1$ is a local maximum? Does this contradict the First Derivative Test? Explain.

Example 6.1.16. Let $f(x) = x^4 - 12x^2 - 5$. Calculate $f'(x) = 4x^3 - 24x = 4x(x^2 - 6) = 3(x + \sqrt{6})(x - \sqrt{6})$ 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 maxima whereas $f(0) = -5$ is a local minimum. 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.

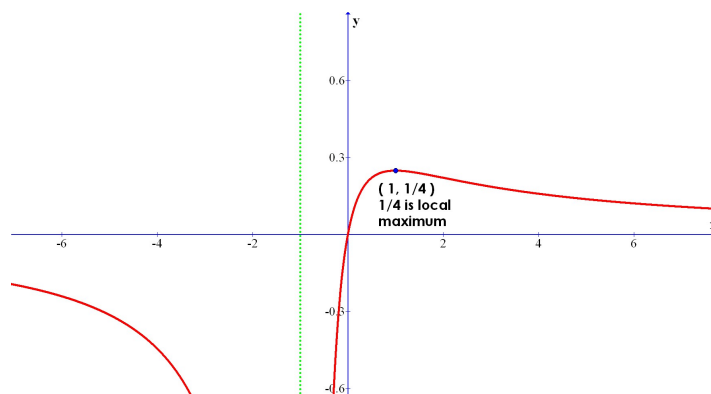
Example 6.1.17. Let $f(x) = \frac{x}{(1+x)^2}$. By quotient rule

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

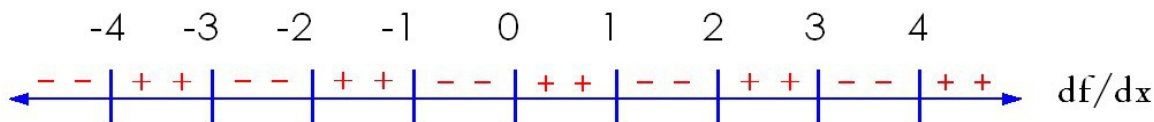
Thus the critical points are $c = 1$ and $c = -1$. The sign-chart is



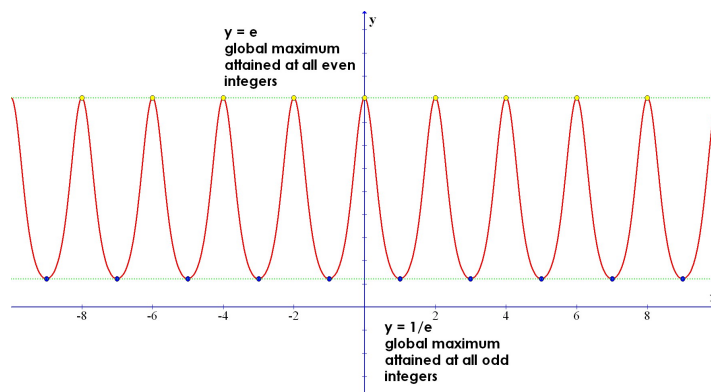
Observe that $x = -1$ is a VA and by the first derivative test $f(1) = 1/4$ is a local maximum. The function is increasing on $(-1, 1)$ and it is decreasing on $(-\infty, -1) \cup (1, \infty)$



Example 6.1.18. 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

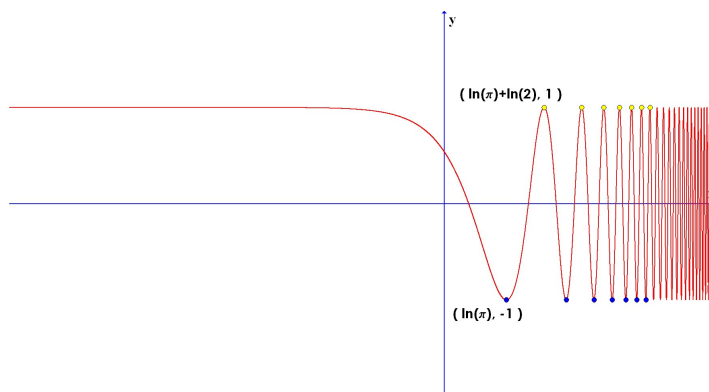


For each even integer $2n$ we apply first derivative test to find $f(2n) = e$ is the global maximum of f and for each odd integer $2n + 1$ we apply first derivative test to find $f(2n + 1) = 1/e$ is the global minimum of f . The graph is sort of like an cosine graph, although it is bounded by $1/e \leq e^{\cos(\pi x)} \leq e$ 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 6.1.19. 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 $x \rightarrow \infty$. 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 $n \in \mathbb{N}$.

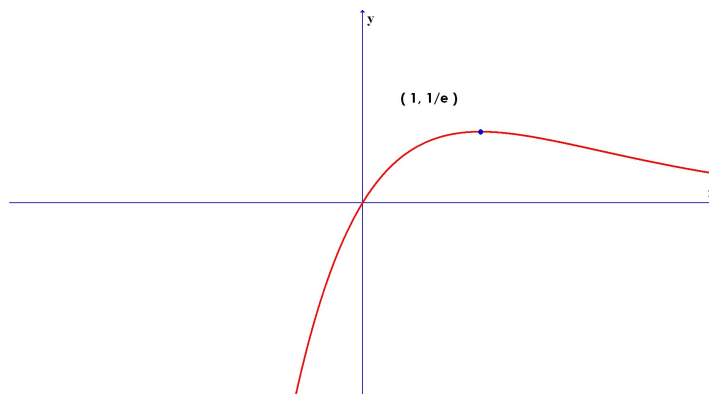


The yellow dots are at $(c_{2n}, 1)$ and the blue dots are at $(c_{2n-1}, -1)$ for $n = 1, 2, 3, 4, 5$.

Example 6.1.20. Let $f(x) = xe^{-x}$. By the product rule, $f'(x) = e^{-x} - xe^{-x} = (1-x)e^{-x}$ thus the critical number is $c = 1$. The sign-chart is very simple here:



Therefore, by the first derivative test $f(1) = 1/e$ is a local maximum. Moreover, the function is increasing on $(-\infty, 1)$ and it is decreasing on $(1, \infty)$. We can see that $f(1) = 1/e$ is also a global maximum for f since $f(x) \leq 1/e$ for all $x \in \mathbb{R}$.



Example 6.1.21. Let $f(x) = \sqrt{(x-1)^2} - \sqrt{(x-2)^2}$. You should recognize³ 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:

$$\sqrt{(x-1)^2} = |x-1| = \begin{cases} x-1 & \text{if } x > 1 \\ 1-x & \text{if } x \leq 1 \end{cases} \quad \sqrt{(x-2)^2} = |x-2| = \begin{cases} x-2 & \text{if } x > 2 \\ 2-x & \text{if } x \leq 2 \end{cases}$$

Therefore,

$$f(x) = |x-1| - |x-2| = \begin{cases} -1 & \text{if } x \leq 1 \\ 2x-1 & \text{if } 1 < x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}.$$

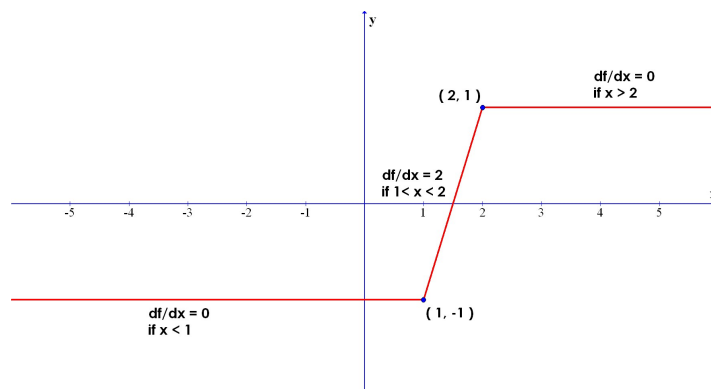
It follows that

$$f'(x) = \begin{cases} 0 & \text{if } x < 1 \\ 2 & \text{if } 1 < x < 2 \\ 0 & \text{if } x > 2 \end{cases}.$$

we can show that f is continuous on \mathbb{R} however the derivative f' is discontinuous at $x = 1$ and $x = 2$. In fact, $\text{dom}(f') = \mathbb{R} - \{1, 2\}$. 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

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

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.)



6.1.2 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.

Remark 6.1.22. (pre-calculus definition of geometric concavity, do not use later).

Suppose $J = (a, b)$ then we define f to be concave down on (a, b) if

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

for all $x \in (a, b)$. Likewise, we define f is concave-up on (a, b) if

$$f(x) \geq f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

If $p \in \text{dom}(f)$ is a point for which the concavity of $(p - \delta_1)$ is different than the concavity of $(p, p + \delta_2)$ for some $\delta_1, \delta_2 > 0$ then we call $(p, f(p))$ an **inflection point**

We should note these definitions are in some sense more general than the definition I offer below (however, I argue this is not a good thing). For example, $f(x) = 3 - |x|$ would be judged concave down on $(-a, a)$ for any $a > 0$ by the definition above. On the other hand, $f(x) = 3 - |x|$ is both concave up and down on any connected set not containing zero. In other words, functions made from lines piecewise joined yield some seemingly contradictory statements. We should like our description to be locally consistent. In other words, if a function is concave up on J then f ought also to be concave up on a connected subset of J . In my view that makes the following a better definition:

Definition 6.1.23. (*calculus definition of geometric concavity, use later*).

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 $p \in \text{dom}(f)$ 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.

Equivalence of definitions proof sketch: Consider this, if a function has the shape of a bowl right side up then the slopes of the tangent lines will increase as we increase x . On the other hand, if a function has the shape of a bowl upside down then the slopes of the tangent lines will decrease as x increases. In other words, the derivative is an increasing function where the function is concave-up and the derivative is a decreasing function where the function is concave-down. This is the heart of the proof that the definition given above is equivalent to the geometric definition for concavity.

Notice that the example $f(x) = 3 - |x|$ escapes trouble in view of the definition above because we cannot say that f is concave down on $(-a, a)$ based on the calculus definition because the calculus-based criteria below does not even apply since $f'(0)$ and $f''(0)$ do not exist. Because increase and decrease allow equality in these notes it is still the case that a line will be found both concave up and down since the derivative of a line is a constant function. Intuitively, a line is between the cases of CU and CD. If we bend a line upwards then it morphs into a CU graph whereas if we bend the line downward it morphs into a CD graph.

Remark 6.1.24.

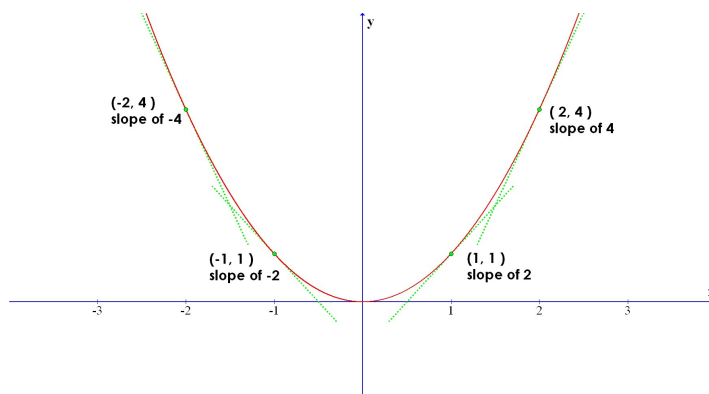
One easy way to remember which is up and which is down is the following slogan:

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.

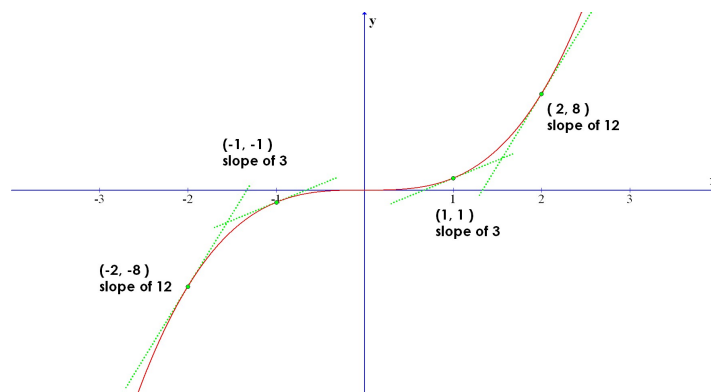
Incidentally, the term "convex" was historically used for concave down and this term is still used in physics particularly in the study of optics.

Example 6.1.25. 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 6.1.26. 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 6.1.27. sign-charts for derivatives reveal increase and decrease of function.

Suppose $f \in C^2(a, b)$ (has continuous 2nd derivative) and (a, b) contains a critical number c then

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, but it might be something else, there are countless other options⁴. The following theorem is geometrically obvious.

Theorem 6.1.28. Second Derivative Test.

Suppose f has a critical number c such that $f'(c) = 0$ and $f''(x)$ 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

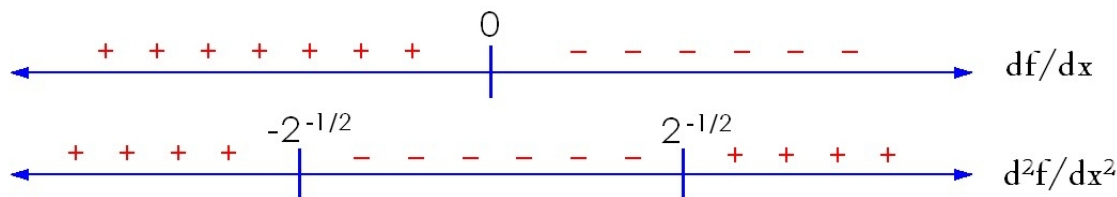
⁴challenge: find me an example of a continuous function which has a nonzero derivative and a critical point which is neither at a local maximum, minimum or inflection point.

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

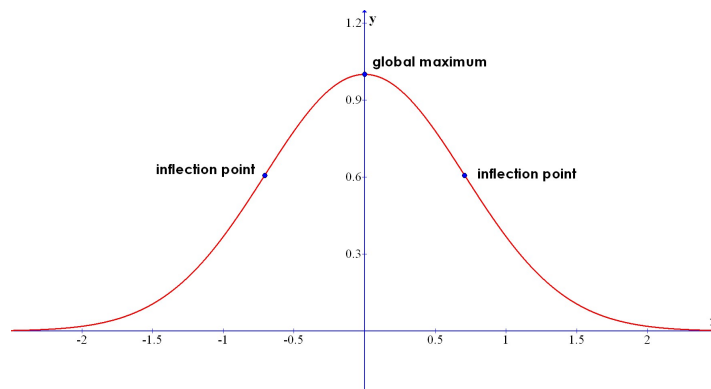
We will discover another proof for the second derivative test when we discuss Taylor's Theorem later in this chapter.

In the examples below we aim to analyze the graph of the given function with the aid of concavity and the second derivative test.

Example 6.1.29. Suppose $f(x) = e^{-x^2}$ then $f'(x) = -2xe^{-x^2}$ and $f''(x) = -2e^{-x^2} + 4x^2e^{-x^2} = 2(2x^2 - 1)e^{-x^2}$. Therefore, we find $c = 0$ is the critical point and $x = \pm 1/\sqrt{2}$ are possible points of inflection. We assemble the sign charts for f' and f'' to guide our thoughts



We identify that $f(0) = 1$ is a local maximum since $f''(0) < 0$. We can read from the sign-chart for f'' that f is CU on $(-\infty, -1/\sqrt{2})$ and $(1/\sqrt{2}, \infty)$ and f is CD on $(-1/\sqrt{2}, 1/\sqrt{2})$. Therefore, $(\pm 1/\sqrt{2}, e^{-1/2})$ are inflection points.

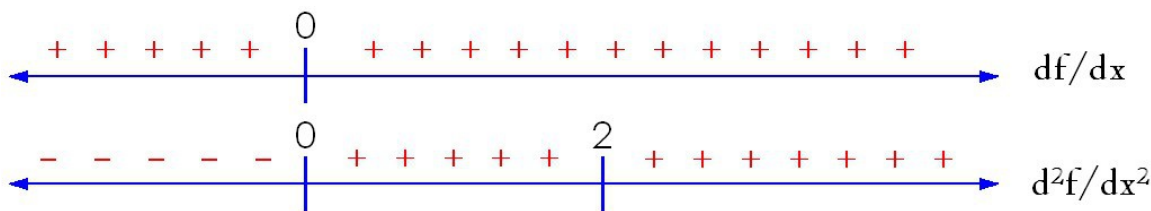


Perhaps you recognize this function, it's the Gaussian function. The graph above is called the Gaussian distribution for its application to probability.

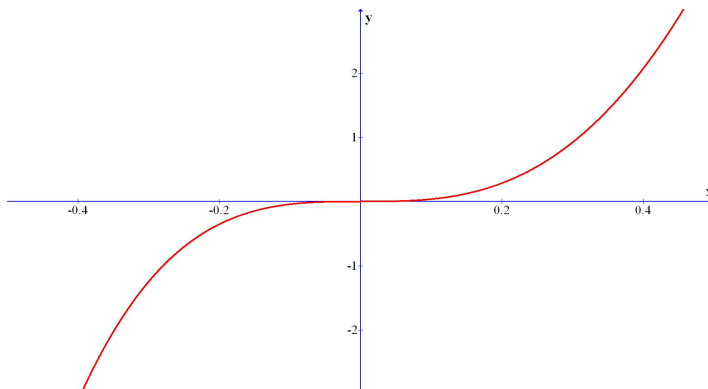
Example 6.1.30. Suppose $f(x) = 3x^5 - 20x^4 + 40x^3$. Differentiating once we find $f'(x) = 15x^4 - 80x^3 + 120x^2$. Differentiate once more to find $f''(x) = 60x^3 - 240x^2 + 240x$. We can factor $f'(x)$ and $f''(x)$ to make clear the critical points and possible inflection points:

$$f'(x) = 5x^2(3x^2 - 16x + 24) \quad f''(x) = 60x(x^2 - 4x + 4) = 60x(x - 2)^2$$

Thus the sign-charts for f' and f'' are: (need to use a few test points to pin down the signs, I leave such details to the reader)



We find the function is increasing on \mathbb{R} and of the two possible inflection points only $(0,0)$ is a point of inflection. This quintic polynomial has a graph that much resembles the standard cubic function. We find $y = f(x)$ is CU on $(0, \infty)$ and CD on $(-\infty, 0)$.



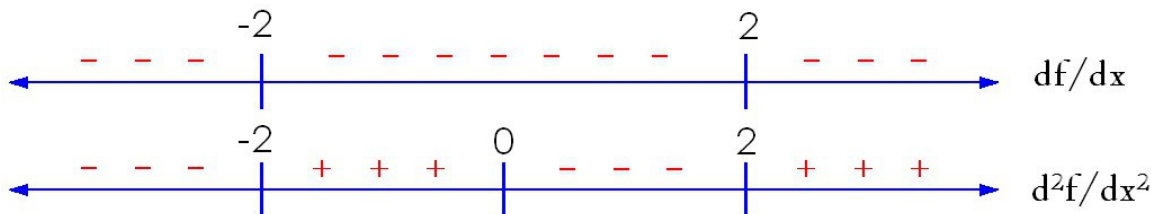
Example 6.1.31. Suppose $f(x) = \frac{1}{x+2} + \frac{1}{x-2}$. If we make a common denominator we find $f(x) = \frac{2x}{x^2-4}$. We differentiate (the original given formula),

$$f'(x) = \frac{-1}{(x+2)^2} - \frac{1}{(x-2)^2} = -\frac{2x^2+8}{(x+2)^2(x-2)^2}$$

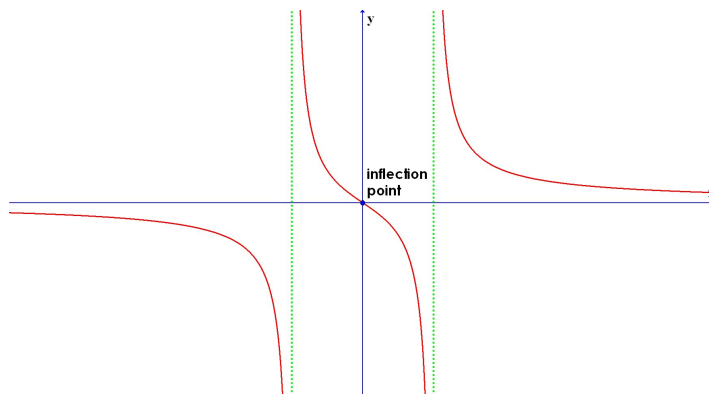
then differentiate again (using the unsimplified $f'(x)$ as starting point),

$$f''(x) = \frac{2}{(x+2)^3} + \frac{2}{(x-2)^3} = 2 \frac{2x^3+48x}{(x+2)^3(x-2)^3} = \frac{4x(x^2+24)}{(x+2)^3(x-2)^3}.$$

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 $(0, 2)$.

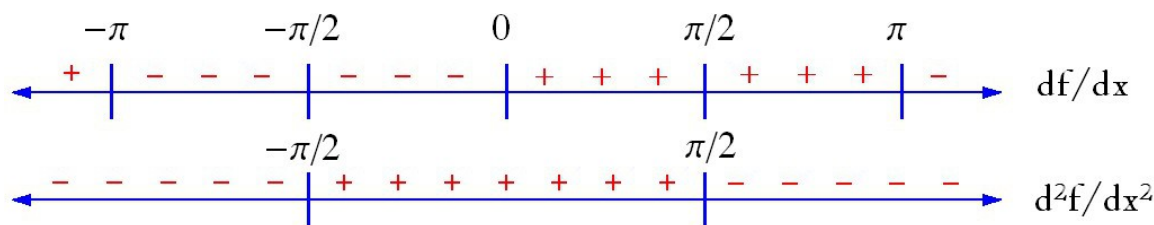


One physical interpretation mathematics found in the previous example is that $y = f(x)$ could be a graph of the electric potential along the x -axis for two positive point charges placed at $x = -2$ and $x = 2$. A divergence in the potential signals the presence of localized charge.

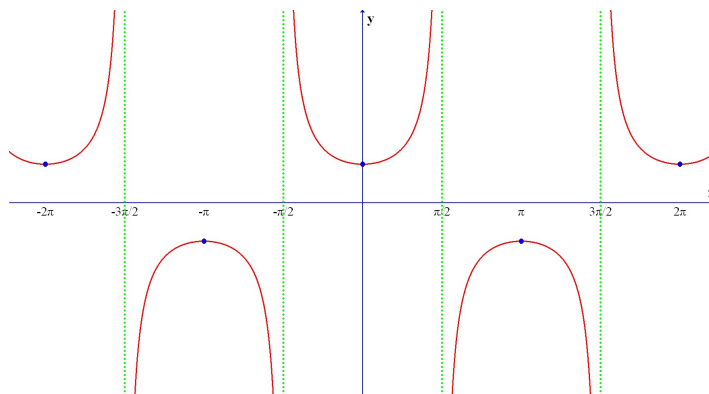
Example 6.1.32. Suppose $f(x) = \sec(x)$ then $f'(x) = \sec(x)\tan(x)$ and $f''(x) = \sec^2(x)\tan(x) + \sec^3(x)$ by the product rule. Let us write these in terms of sine and cosine since we have a complete and working knowledge of all the zeros for sine and cosine.

$$f(x) = \frac{1}{\cos(x)} \quad \frac{df}{dx} = \frac{\sin(x)}{\cos^2(x)} \quad \frac{d^2f}{dx^2} = \frac{\sin(x) + 1}{\cos^3(x)}$$

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 6.1.33. *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 .

6.1.3 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 6.1.34. *closed interval method.*

If we are given function f which is continuous on a closed interval $[a, b]$ then 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 6.1.35. 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 $x = \frac{\pi}{2}, \frac{3\pi}{2} \in [0, 2\pi]$.

$$f(0) = \sin(0) = 0, \quad f\left(\frac{\pi}{2}\right) = \sin\left(\frac{\pi}{2}\right) = 1, \quad f\left(\frac{3\pi}{2}\right) = \sin\left(\frac{3\pi}{2}\right) = -1, \quad \sin(2\pi) = 0$$

Therefore, by closed interval method $f\left(\frac{\pi}{2}\right) = 1$ is the maximum and $f\left(\frac{3\pi}{2}\right) = -1$ is the minimum of $f(x) = \sin(x)$ on the interval $[0, 2\pi]$.

Example 6.1.36. Let $f(x) = (x - 3)(x - 4)$ find absolute extrema of f on $[0, 1]$. Calculate $f'(x) = (x - 4) + (x - 3) = 2x - 7$ thus $c = -\frac{7}{2}$ is a critical point. Compute the values of $f(x)$ at the critical points inside $[0, 1]$ and the endpoints (there are no critical points in $[0, 1]$):

$$f(0) = 12, \quad f(1) = 6.$$

Therefore, $f(0) = 12$ is the absolute maximum and $f(1) = 6$ is the absolute minimum of $f(x) = (x - 3)(x - 4)$ on $[0, 1]$.

Example 6.1.37. Let $f(x) = x^4 - 2x^2 + 3$ find absolute extrema of f on $[0, 2]$. Note that $f'(x) = 4x^3 - 4x = 4x(x^2 - 1) = 4x(x + 1)(x - 1)$ thus $c = 0, -1, 1$ are critical points for f . Only $0, 1 \in [0, 2]$. Calculate the values of the potential extrema:

$$f(0) = 3, \quad f(1) = 2$$

Thus, $f(1) = 2$ is the minimum and $f(0) = 3$ is the maximum of f on $[0, 2]$.

Example 6.1.38. 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,

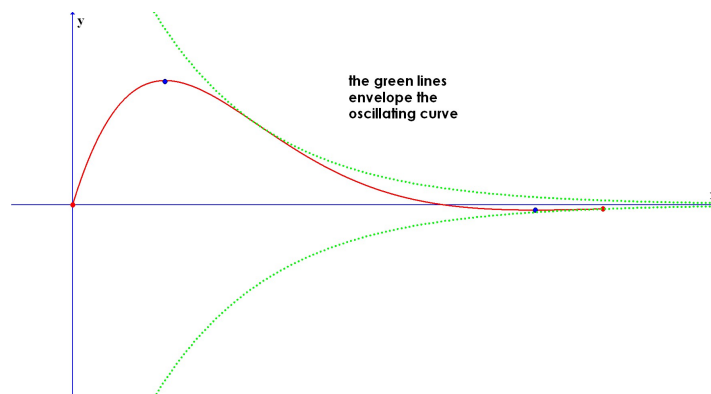
$$f\left(\frac{\pi}{4}\right) = e^{-\frac{\pi}{4}} \sin\left(\frac{\pi}{4}\right) \approx 0.32$$

$$f\left(\frac{5\pi}{4}\right) = e^{-\frac{5\pi}{4}} \sin\left(\frac{5\pi}{4}\right) \approx -0.0139$$

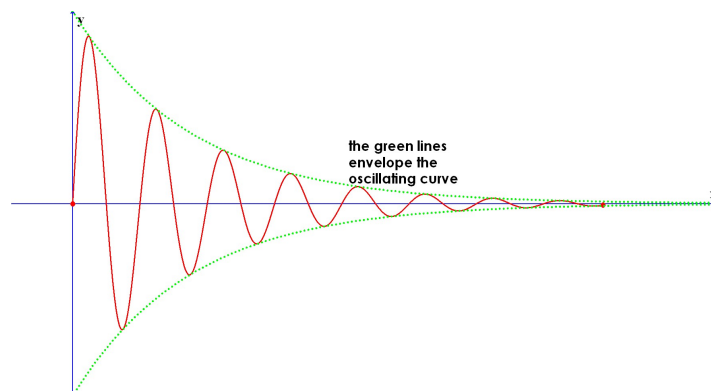
$$f(0) = e^0 \sin(0) = 0$$

$$f(4) = e^{-4} \sin(4) = -0.138$$

We find $f\left(\frac{5\pi}{4}\right) = -0.0139$ is the minimum and $f\left(\frac{\pi}{4}\right) = 0.32$ 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.



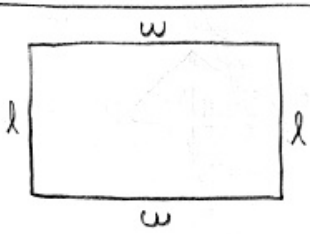
Physically these are very interesting functions. You should see it again when you study springs with friction or RLC circuits.

6.2 optimization

I have preserved the format of these examples from an earlier edition of my notes. If I have any advice about optimization problems or word problems more generally it is simply to write down your thoughts. Draw a picture. Label unknown quantities. Once you find a solution, check it against common sense. Anyway, people have written whole books on the proper way to teach problem solving. I assume the reader is mature enough that no large amount of coddling is required. IN A WORD: THINK.

Example 6.2.1.

E1 Given 400ft of fencing what dimensions should you give a rectangular pen as to maximize the area?



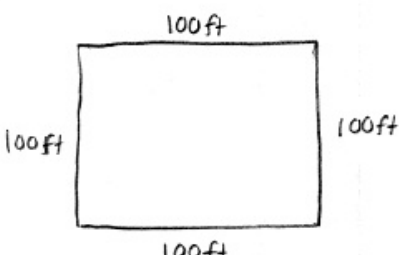
$2l + 2w = 400$
 $A = lw$
 Notice that $l = 200 - w$
 $\Rightarrow A = (200 - w)w = 200w - w^2$

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 2nd derivative test $w = 100$ gives max. area. Thus

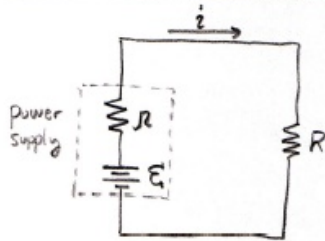


$l = 200 - 100 = 100$

These dimensions maximize the area of the pen.

Example 6.2.2.

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?



$$\mathcal{E} = ir + iR \quad : \text{Kirchoff's Voltage Law}$$

$$P = i^2 R \quad : \text{Power delivered to } R \text{ by current } i$$

$$\text{Solve to find } i = \frac{\mathcal{E}}{r+R} \text{ then substitute into } P = i^2 R = \frac{\mathcal{E}^2 R}{(r+R)^2}$$

Now do calculus on P as a function of R ,

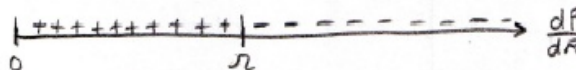
$$\frac{dP}{dR} = \mathcal{E}^2 \frac{d}{dR} \left(\frac{R}{(r+R)^2} \right)$$

$$= \mathcal{E}^2 \left[\frac{(r+R)^2 - 2(r+R)R}{(r+R)^4} \right]$$

$$= \frac{\mathcal{E}^2}{(r+R)^4} [r^2 + 2rR + R^2 - 2rR - 2R^2]$$

$$= \frac{\mathcal{E}^2}{(r+R)^4} [r^2 - R^2]$$

$$= \frac{\mathcal{E}^2}{(r+R)^2} [(r-R)(r+R)] \quad : r, R > 0 \text{ for physical reasons so } R=r \text{ is only interesting critical point.}$$



Thus by 1st derivative test the max power is delivered to R when $R = r$. This is a simple case of the

"Max Power Transfer Th"

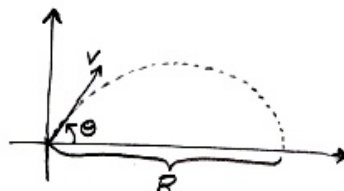
Notice it implies that the Max Efficiency of a simple power supply is 50%. Electrical engineers can build more efficient power supplies called "switching" power supplies, in those supplies the output is automatically tailored as to be best for the load.

Example 6.2.3.

E3 The range of a projectile fired at angle Θ with velocity V can be shown to be $R = \frac{V \sin(2\Theta)}{g}$ assuming no friction and $g = 9.8 \text{ m/s}^2$. What angle Θ maximizes the range. What is R_{\max} ?

We notice that V and g are constants, R is the dependent (like y) variable and Θ is the independent variable (like x).

$$\begin{aligned} \frac{dR}{d\Theta} &= \frac{d}{d\Theta} \left(\frac{V}{g} \sin(2\Theta) \right) \\ &= \frac{V}{g} \frac{d}{d\Theta} (\sin(2\Theta)) \\ &= \frac{2V}{g} \cos(2\Theta) \end{aligned}$$



For physical reasons we restrict Θ to $0 \leq \Theta \leq \pi/2$.

$$\cos(2\Theta) = 0 \quad \text{for} \quad 2\Theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots$$

$$\Rightarrow \Theta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \dots$$

only $\Theta = \pi/4$ is a physically interesting critical value.

$$\text{Calculate } \frac{d^2R}{d\Theta^2} = -\frac{4V}{g} \sin(2\Theta) \Rightarrow \frac{d^2R}{d\Theta^2}(\Theta = \pi/4) = -\frac{4V}{g} \sin(\pi/2)$$

Thus we note $\frac{d^2R}{d\Theta^2}(\Theta = \pi/4) = -\frac{4V}{g} < 0$ hence by 2nd derivative test we have that $R(\pi/4)$ is the max. range. $\boxed{\Theta = \pi/4}$

$$R_{\max} = \frac{V \sin(2 \cdot \frac{\pi}{4})}{g} = \boxed{\frac{V}{g} = R_{\max}}$$

Example 6.2.4.

E4 The height of a projectile launched straight up with velocity V_0 is given by:

$$Y = V_0 t - \frac{1}{2} g t^2$$

What is the maximum height and at what time is it reached?

Notice that V_0 and g are just numbers so,

$$\frac{dy}{dt} = \frac{d}{dt} \left(V_0 t - \frac{1}{2} g t^2 \right) = V_0 - g t \quad (\text{the velocity})$$

$$\frac{d^2 y}{dt^2} = \frac{d}{dt} (V_0 - g t) = -g \quad (\text{the acceleration})$$

Now $\frac{dy}{dt} = 0 = V_0 - g t \Rightarrow t_{\max} = V_0/g$ is critical time.

We know that $Y(t_{\max}) = Y_{\max} = V_0 \left(\frac{V_0}{g} \right) - \frac{1}{2} g \left(\frac{V_0}{g} \right)^2 = \frac{V_0^2}{2g} = Y_{\max}$
 is indeed the max height because $Y'' = -g < 0$. (2nd der. test.)

Example 6.2.5.

ES Find the point on the line $Y = 4x + 7$ that is closest to the origin $(0,0)$.

We want to minimize the distance $S = \sqrt{x^2 + y^2}$ where $Y = 4x + 7$ so consider

$$\frac{d}{dx} \left(\sqrt{x^2 + (4x+7)^2} \right) = \frac{2x + 8(4x+7)}{2\sqrt{x^2 + (4x+7)^2}} = \frac{ds}{dx}$$

So then $\frac{ds}{dx} = 0 \Leftrightarrow 2x + 8(4x+7) = 0 \Leftrightarrow 34x = -56$
 Thus $x = -28/17$ is the only critical point.

Thus $s(-28/17 = x)$ is the min. distance to $Y = 4x + 7$ from the origin by 1st derivative test. The Y -coordinate is

$$Y(-28/17) = 4(-28/17) + 7 = 7/17$$

So the closest point is in fact $\left(-\frac{28}{17}, \frac{7}{17} \right)$

Example 6.2.6.

Find two positive numbers whose product is 100 and whose sum is a minimum

Let m and n be the numbers. We have that

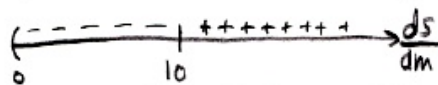
$$mn = 100$$

$$S = m + n = m + \frac{100}{m}$$

Now minimize S as a function of m ,

$$\frac{dS}{dm} = 1 - \frac{100}{m^2} = 0 \Rightarrow m = \pm 10 \text{ are critical values of } m.$$

Notice that $m=0$ is also a critical value. However only $m=10$ matters because we're looking for positive #'s.



Thus S is minimized for $m=10$. Thus the numbers are $m=10$ and $n=10$.

6.3 to $\pm\infty$ and beyond

The behavior a function for $x \gg 0$ or for $x \ll 0$ is captured by the limit⁵ of the function at $\pm\infty$,

Definition 6.3.1. *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

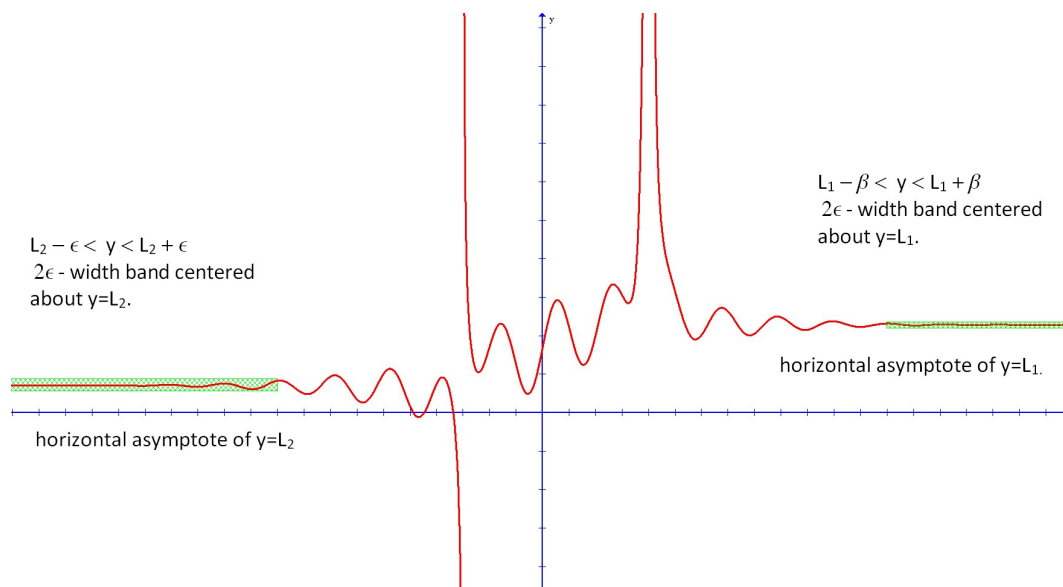
$$\lim_{x \rightarrow \infty} f(x) = L$$

in this case. To be more precise we should say that $\lim_{x \rightarrow \infty} f(x) = L$ iff for each $\epsilon > 0$ there exists $N \in \mathbb{R}$ with $N > 0$ such that if $x > N$ then $|f(x) - L| < \epsilon$. Likewise,

$$\lim_{x \rightarrow -\infty} f(x) = L$$

iff for each $\epsilon > 0$ there exists $M \in \mathbb{R}$ with $M < 0$ such that if $x < M$ then $|f(x) - L| < \epsilon$.

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 $\epsilon > 0$ 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 $\lim_{x \rightarrow \infty} f(x) = L_1$ and $\lim_{x \rightarrow -\infty} f(x) = L_2$.

⁵we 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; $\lim_{x \rightarrow \infty} e^x = \infty$.

Example 6.3.2. Let $f(x) = \frac{1}{x}$. Calculate the limit of $f(x)$ at ∞ . Observe that,

$$f(10) = 0.1, \quad f(100) = 0.01, \quad f(1000) = 0.001.$$

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,

$$\lim_{x \rightarrow \infty} \frac{1}{x} = 0.$$

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 $\epsilon > 0$ choose $N = 1/\epsilon$ and observe that for $x > N = 1/\epsilon$ it follows that $1/x < \epsilon$. Consequently, $x > N$ implies $|f(x) - 0| = |\frac{1}{x}| = \frac{1}{x} < \epsilon$. Hence by the precise definition $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$. \square

The limits at $-\infty$ are much the same.

Example 6.3.3. Let $f(x) = \frac{1}{x}$. Calculate the limit of $f(x)$ at $-\infty$. Observe that,

$$f(-10) = -0.1, \quad f(-100) = -0.01, \quad f(-1000) = -0.001.$$

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,

$$\lim_{x \rightarrow -\infty} \frac{1}{x} = 0.$$

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 $\epsilon > 0$ choose $M = -1/\epsilon$ and observe that for $x < M = -1/\epsilon$ it follows that $-1/x < \epsilon$. Consequently, $x < N$ implies $|f(x) - 0| = |\frac{1}{x}| = -\frac{1}{x} < \epsilon$. Hence by the precise definition $\lim_{x \rightarrow -\infty} \frac{1}{x} = 0$. \square

Clearly we'd prefer to avoid the picky *eps*-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 6.3.4. Let $f(x) = 1/x^n$ where $n > 0$. Calculate the limit of $f(x)$ at ∞ . Observe that,

$$f(10) = 1/10^n, \quad f(100) = 1/100^n, \quad f(1000) = 1/1000^n.$$

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,

$$\lim_{x \rightarrow \infty} \frac{1}{x^n} = 0.$$

Let $\epsilon > 0$ choose $N = 1/\epsilon^{\frac{1}{n}}$. 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, $\lim_{x \rightarrow \infty} \frac{1}{x^n} = 0$. \square

The graphical significance of all three examples thus far considered is that the function has a *horizontal asymptote* of $y = 0$ as $x \rightarrow \pm\infty$.

Definition 6.3.5. *horizontal asymptotes.*

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

Example 6.3.6. 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 $x \gg 0$ and $y = -\frac{\pi}{2}$ for $x \ll 0$. Therefore,

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

Vertical asymptotes of the function correspond to horizontal asymptotes for the inverse function⁶ 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 6.3.7. *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

$$\lim_{x \rightarrow \infty} f(x) = \infty.$$

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

$$\lim_{x \rightarrow -\infty} f(x) = \infty.$$

Similarly, if for each $M < 0$ there exists $N > 0$ such that $x > N$ implies $f(x) < M$ we say $\lim_{x \rightarrow \infty} f(x) = -\infty$. Finally, if for each $M < 0$ there exists $N < 0$ such that $x < N$ implies $f(x) < M$ we say $\lim_{x \rightarrow -\infty} f(x) = -\infty$.

Example 6.3.8. 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.

$\lim_{x \rightarrow \infty} \frac{1}{x} = 0$	$\lim_{x \rightarrow -\infty} \frac{1}{x} = 0$	$\lim_{x \rightarrow 0^+} \frac{1}{x} = \infty$	$\lim_{x \rightarrow 0^-} \frac{1}{x} = -\infty$	$\lim_{x \rightarrow 0} \frac{1}{x} = dne$
$\lim_{x \rightarrow \infty} \frac{1}{x^2} = 0$	$\lim_{x \rightarrow -\infty} \frac{1}{x^2} = 0$	$\lim_{x \rightarrow 0^+} \frac{1}{x^2} = \infty$	$\lim_{x \rightarrow 0^-} \frac{1}{x^2} = \infty$	$\lim_{x \rightarrow 0} \frac{1}{x^2} = \infty$
$\lim_{x \rightarrow \infty} \frac{1}{\sqrt{x}} = 0$	$\lim_{x \rightarrow -\infty} \frac{1}{\sqrt{x}} = ?$	$\lim_{x \rightarrow 0^+} \frac{1}{\sqrt{x}} = \infty$	$\lim_{x \rightarrow 0^-} \frac{1}{\sqrt{x}} = ?$	$\lim_{x \rightarrow 0} \frac{1}{\sqrt{x}} = dne$
$\lim_{x \rightarrow \infty} \sqrt{x} = \infty$	$\lim_{x \rightarrow -\infty} \sqrt{x} = ?$	$\lim_{x \rightarrow 0^+} \sqrt{x} = 0$	$\lim_{x \rightarrow 0^-} \sqrt{x} = ?$	$\lim_{x \rightarrow 0} \sqrt{x} = dne$
$\lim_{x \rightarrow \infty} x^2 = \infty$	$\lim_{x \rightarrow -\infty} x^2 = \infty$	$\lim_{x \rightarrow 0^+} x^2 = 0$	$\lim_{x \rightarrow 0^-} x^2 = 0$	$\lim_{x \rightarrow 0} x^2 = 0$
$\lim_{x \rightarrow \infty} x^3 = \infty$	$\lim_{x \rightarrow -\infty} x^3 = -\infty$	$\lim_{x \rightarrow 0^+} x^3 = 0$	$\lim_{x \rightarrow 0^-} x^3 = 0$	$\lim_{x \rightarrow 0} x^3 = 0.$

⁶Challenge: what are the horizontal asymptotes of $y = \tan^{-1}(3x)$?

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 6.3.9. *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.*

$$\begin{array}{lll} \lim_{x \rightarrow \infty} \sin(x) = d.n.e. & \lim_{x \rightarrow \infty} \cos(x) = d.n.e. & \lim_{x \rightarrow \infty} \tan(x) = d.n.e. \\ \lim_{x \rightarrow -\infty} \sin(x) = d.n.e. & \lim_{x \rightarrow -\infty} \cos(x) = d.n.e. & \lim_{x \rightarrow -\infty} \tan(x) = d.n.e. \\ \lim_{x \rightarrow \infty} \csc(x) = d.n.e. & \lim_{x \rightarrow \infty} \sec(x) = d.n.e. & \lim_{x \rightarrow \infty} \sec(x) = d.n.e. \\ \lim_{x \rightarrow -\infty} \csc(x) = d.n.e. & \lim_{x \rightarrow -\infty} \sec(x) = d.n.e. & \lim_{x \rightarrow -\infty} \sec(x) = d.n.e. \end{array}$$

Example 6.3.10. *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.*

$$\begin{array}{lll} \lim_{x \rightarrow \infty} \sin^{-1}(x) = d.n.e. & \lim_{x \rightarrow \infty} \cos^{-1}(x) = d.n.e. & \lim_{x \rightarrow \infty} \tan^{-1}(x) = \pi/2 \\ \lim_{x \rightarrow -\infty} \sin^{-1}(x) = d.n.e. & \lim_{x \rightarrow -\infty} \cos^{-1}(x) = d.n.e. & \lim_{x \rightarrow -\infty} \tan^{-1}(x) = -\pi/2 \\ \lim_{x \rightarrow \infty} e^x = \infty & \lim_{x \rightarrow \infty} e^{-x} = 0 & \lim_{x \rightarrow \infty} (1/2)^x = 0 \\ \lim_{x \rightarrow -\infty} e^x = 0 & \lim_{x \rightarrow -\infty} e^{-x} = \infty & \lim_{x \rightarrow -\infty} (1/2)^x = \infty \end{array}$$

The domain of $\sin^{-1}(x)$ and $\cos^{-1}(x)$ will be the range of sine and cosine respectively; that is $\text{dom}(\sin^{-1}(x)) = [-1, 1]$ and $\text{dom}(\cos^{-1}(x)) = [-1, 1]$ 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

$$\lim_{x \rightarrow -\infty} \ln(x) = d.n.e. \quad \lim_{x \rightarrow 0^+} \ln(x) = -\infty \quad \lim_{x \rightarrow \infty} \ln(x) = \infty$$

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. I've assigned the proof as homework.

The following lemma connects limits at $\pm\infty$ with one-sided limits at zero.

Lemma 6.3.11.

$$\lim_{x \rightarrow \infty} f(x) = \lim_{t \rightarrow 0^+} f(1/t) \quad \text{and} \quad \lim_{x \rightarrow -\infty} f(x) = \lim_{t \rightarrow 0^-} f(1/t).$$

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 $\lim_{x \rightarrow \infty} f(x) = L \in \mathbb{R}$. Let $\epsilon > 0$ and note the following inequalities are equivalent:

$$0 < M < x \quad \Leftrightarrow \quad 0 < \frac{1}{x} < \frac{1}{M}$$

Therefore, $0 < \frac{1}{x} < \frac{1}{m}$ implies $|f(x) - L| < \epsilon$ which indicates that

$$\lim_{\frac{1}{x} \rightarrow 0^+} f(x) = L \quad \text{hence using } t = 1/x \text{ we find} \quad \lim_{t \rightarrow 0^+} f(1/t) = L.$$

The proof $\lim_{x \rightarrow -\infty} f(x) = \lim_{t \rightarrow 0^-} f(1/t) \in \mathbb{R}$ is similar.

Suppose $\lim_{x \rightarrow -\infty} f(x) = \infty$. 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

$$\lim_{\frac{1}{x} \rightarrow 0^-} f(x) = \infty \quad \text{hence using } t = 1/x \text{ we find} \quad \lim_{t \rightarrow 0^-} f(1/t) = \infty.$$

Proof for other cases are similar and left to the reader. The basic point is that if $x \rightarrow \pm\infty$ then $t = \frac{1}{x} \rightarrow 0^\pm$. \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.

6.3.1 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 indeterminacy. All the same indeterminant forms (see defn. 3.4.4) arise again and most of the algebraic techniques we used back in section 3.4 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,

$$\lim_{x \rightarrow \infty} (x - 2x) = \lim_{x \rightarrow \infty} (x) + \lim_{x \rightarrow \infty} (-2x) = \infty - \infty$$

is not valid because you might be tempted to cancel and find $\lim_{x \rightarrow \infty} (x - 2x) = 0$ yet $\lim_{x \rightarrow \infty} (x - 2x) = \lim_{x \rightarrow \infty} (-x) = -\infty$ 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 calculus⁷

⁷in complex variables one can actually add the point at infinity and use the extended complex numbers. In fact, 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 6.3.12. *this one is type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{x+3}{x-2} \right) &= \lim_{x \rightarrow \infty} \left(\frac{\frac{x}{x} + \frac{3}{x}}{\frac{x}{x} - \frac{2}{x}} \right) && \text{divided top and bottom by } x \\ &= \lim_{x \rightarrow \infty} \left(\frac{1+0}{1-0} \right) && c/x \rightarrow 0 \text{ as } x \rightarrow \infty \\ &= 1.\end{aligned}$$

Example 6.3.13. *this one is also of type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{x^3 + 3x - 2}{x^4 - 2x + 1} \right) &= \lim_{x \rightarrow \infty} \left(\frac{\frac{1}{x} + \frac{3}{x^3} - \frac{2}{x^4}}{1 - \frac{2}{x^3} + \frac{1}{x^4}} \right) && \text{divided top and bottom by } x^4 \\ &= \lim_{x \rightarrow \infty} \left(\frac{0+0-0}{1-0+0} \right) && \text{for } n=1,2,4, c/x^n \rightarrow 0 \text{ as } x \rightarrow \infty \\ &= 0.\end{aligned}$$

Example 6.3.14. *again, type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}\lim_{x \rightarrow -\infty} \left(\frac{x^3 + 3x - 2}{x^2 - x + 7} \right) &= \lim_{x \rightarrow -\infty} \left(\frac{\frac{x}{x} + \frac{3}{x} - \frac{2}{x^2}}{1 - \frac{2}{x} + \frac{7}{x^2}} \right) && \text{divided top and bottom by } x^2 \\ &= \lim_{x \rightarrow -\infty} \left(\frac{\frac{x}{1}}{1} \right) && \text{for } n=1,2, c/x^n \rightarrow 0 \text{ as } x \rightarrow -\infty \\ &= -\infty.\end{aligned}$$

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 $f(x) \rightarrow 0$ as $x \rightarrow -\infty$. To be safer you should experiment with more than one number, or better yet THINK.

Example 6.3.15. *you guessed it, type $\frac{\infty}{\infty}$ to begin.*

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{\sqrt{2x^4 + 3x - 2}}{x^2 - x + 7} \right) &= \lim_{x \rightarrow \infty} \left(\frac{\frac{1}{x^2} \sqrt{2x^4 + 3x - 2}}{\frac{1}{x^2} (x^2 - x + 7)} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{\sqrt{\frac{2x^4 + 3x - 2}{x^4}}}{1 - \frac{1}{x} + \frac{7}{x^2}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{\sqrt{2 + \frac{3}{x^3} - \frac{2}{x^4}}}{1 - \frac{1}{x} + \frac{7}{x^2}} \right) \\ &= \sqrt{2}.\end{aligned}$$

Example 6.3.16. *this has type $0 \cdot \infty$ to begin.*

$$\begin{aligned} \lim_{x \rightarrow \infty} (e^{-x} 2^x) &= \lim_{x \rightarrow \infty} (e^{\ln(2^x)} e^{-x}) && \text{sneaky step} \\ &= \lim_{x \rightarrow \infty} (e^{x \ln(2)} e^{-x}) \\ &= \lim_{x \rightarrow \infty} (e^{x(\ln(2)-1)}) \\ &= 0. \end{aligned}$$

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 indeterminacy 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.

Example 6.3.17. *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*

$$P(x) = a_p x^p + \dots + a_1 x + a_0 \quad Q(x) = b_q x^q + \dots + b_1 x + b_0$$

Consider $f(x) = P(x)/Q(x)$. There are three cases.

1. *If $p > q$ then $p - q > 0$ hence*

$$\begin{aligned} \lim_{x \rightarrow \infty} \left(\frac{P(x)}{Q(x)} \right) &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^p + \dots + a_1 x + a_0}{b_q x^q + \dots + b_1 x + b_0} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^{p-q} + \dots + \frac{a_1}{x^{q-1}} + \frac{a_0}{x^q}}{b_q + \dots + \frac{b_1}{x^{q-1}} + \frac{b_0}{x^q}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p}{b_q} x^{p-q} + \dots + \frac{a_1}{b_q x^{q-1}} + \frac{a_0}{b_q x^q} \right) \\ &= \pm \infty. \end{aligned}$$

In particular if $a_p/b_q > 0$ then ∞ is obtained whereas if $a_p/b_q < 0$ then $-\infty$ is the answer.

2. *If $p < q$ then $q - p > 0$ hence*

$$\begin{aligned} \lim_{x \rightarrow \infty} \left(\frac{P(x)}{Q(x)} \right) &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^p + \dots + a_1 x + a_0}{b_q x^q + \dots + b_1 x + b_0} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^{p-q} + \dots + \frac{a_1}{x^{q-1}} + \frac{a_0}{x^q}}{b_q + \dots + \frac{b_1}{x^{q-1}} + \frac{b_0}{x^q}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p}{b_q} \frac{1}{x^{q-p}} + \dots + \frac{a_1}{b_q x^{q-1}} + \frac{a_0}{b_q x^q} \right) \\ &= 0. \end{aligned}$$

3. If $p = q$ then

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{P(x)}{Q(x)} \right) &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^p + \cdots + a_1 x + a_0}{b_q x^q + \cdots + b_1 x + b_0} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p x^{p-q} + \cdots + \frac{a_1}{x^{q-1}} + \frac{a_0}{x^q}}{b_q + \cdots + \frac{b_1}{x^{q-1}} + \frac{b_0}{x^q}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{a_p}{b_q} + \cdots + \frac{a_1}{b_q x^{q-1}} + \frac{a_0}{b_q x^q} \right) \\ &= \frac{a_p}{b_q}.\end{aligned}$$

In each case my goal was to simplify the denominator so I could focus on the behavior of the numerator. Very similar arguments will work for $x \rightarrow -\infty$.

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 $x \in \text{dom}(f)$.

Example 6.3.18. 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:

$$\begin{aligned}\lim_{x \rightarrow \infty} (\sin(x) + e^x) &= \lim_{x \rightarrow \infty} (e^x) = \infty \\ \lim_{x \rightarrow -\infty} (x + 2) &= \lim_{x \rightarrow -\infty} (x) = -\infty.\end{aligned}$$

Example 6.3.19. if we take a function $f(x)$ with a known limit of $L \in \mathbb{R}$ or $\pm\infty$ as $x \rightarrow \pm\infty$ then the limit of $f(x+a)$ for $a \in \mathbb{R}$ is the same for $x \rightarrow \pm\infty$. For example,

$$\begin{aligned}\lim_{x \rightarrow \infty} (e^x) = \infty &\implies \lim_{x \rightarrow \infty} (e^{x+3}) = \infty \\ \lim_{x \rightarrow -\infty} \left(\frac{1}{x^2} \right) = 0 &\implies \lim_{x \rightarrow -\infty} \left(\frac{1}{(x-7)^2} \right) = 0 \\ \lim_{x \rightarrow \infty} (\tan^{-1}(x)) = \frac{\pi}{2} &\implies \lim_{x \rightarrow \infty} (\tan^{-1}(x+2)) = \frac{\pi}{2}.\end{aligned}$$

Example 6.3.20. in a contest between power functions the largest degree wins.

$$\begin{aligned}\lim_{x \rightarrow \infty} (x^3 - x^2) &= \lim_{x \rightarrow \infty} (x^3) = \infty. \\ \lim_{x \rightarrow \infty} (x^3 - x^4) &= \lim_{x \rightarrow \infty} (-x^4) = -\infty.\end{aligned}$$

On the other hand the exponential function will win against a polynomial because eventually the exponential function's values will totally dwarf the power function's values.

$$\begin{aligned}\lim_{x \rightarrow \infty} (x^3 - e^x) &= \lim_{x \rightarrow \infty} (-e^x) = -\infty. \\ \lim_{x \rightarrow -\infty} (2^{-x} + x) &= \lim_{x \rightarrow -\infty} (2^{-x}) = \infty.\end{aligned}$$

How would we prove such a claim?

Example 6.3.21. *limits of type $\infty - \infty$ can sometimes be dealt with via the rationalization technique:*

$$\begin{aligned}
 \lim_{x \rightarrow \infty} (x - \sqrt{x}) &= \lim_{x \rightarrow \infty} \left(\frac{x + \sqrt{x}}{x + \sqrt{x}} (x - \sqrt{x}) \right) \\
 &= \lim_{x \rightarrow \infty} \left(\frac{x^2 - x}{x + \sqrt{x}} \right) \\
 &= \lim_{x \rightarrow \infty} \left(\frac{x - 1}{1 + \frac{\sqrt{x}}{x}} \right) \\
 &= \lim_{x \rightarrow \infty} \left(\frac{x - 1}{1 + \sqrt{1/x}} \right) \\
 &= \lim_{x \rightarrow \infty} (x) \\
 &= \infty.
 \end{aligned}$$

Example 6.3.22. *this limit is also of type $\infty - \infty$ but in this case the $-\infty$ wins.*

$$\begin{aligned}
 \lim_{x \rightarrow -\infty} (x + \sqrt{x^2 + 4x}) &= \lim_{x \rightarrow -\infty} \left(\frac{x - \sqrt{x^2 + 4x}}{x - \sqrt{x^2 + 4x}} (x + \sqrt{x^2 + 4x}) \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{x^2 - x^2 - 4x}{x - \sqrt{x^2 + 4x}} \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{-4x}{2x - \sqrt{x^2 + 4x}} \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{-4}{1 - \sqrt{1 + 4/x}} \right) \\
 &= -\infty.
 \end{aligned}$$

Of course, similar looking problems might have ∞ as the answer:

$$\begin{aligned}
 \lim_{x \rightarrow -\infty} (x - \sqrt{2x^2 + 4x}) &= \lim_{x \rightarrow -\infty} \left(\frac{x + \sqrt{2x^2 + 4x}}{x + \sqrt{2x^2 + 4x}} (x - \sqrt{2x^2 + 4x}) \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{x^2 - 2x^2 - 4x}{x + \sqrt{2x^2 + 4x}} \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{-x^2 - 4x}{x + \sqrt{2x^2 + 4x}} \right) \\
 &= \lim_{x \rightarrow -\infty} \left(\frac{-x - 4/x}{1 + \sqrt{2 - 4/x}} \right) \\
 &= \infty.
 \end{aligned}$$

It is also possible for the type $\infty - \infty$ to resolve to a finite limit.

Example 6.3.23. *when dealing with square roots it is important that you remember that the laws of exponents indicate $\frac{1}{x}\sqrt{a+b} = \sqrt{\frac{1}{x^2}(a+b)}$. We assume that $a, c > 0$ in this problem. Consider,*

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{\sqrt{ax^2 + bx + c}}{\sqrt{cx^2 + dx + e}} \right) &= \lim_{x \rightarrow \infty} \left(\frac{\frac{1}{x}\sqrt{ax^2 + bx + c}}{\frac{1}{x}\sqrt{cx^2 + dx + e}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{\sqrt{\frac{ax^2 + bx + c}{x^2}}}{\sqrt{\frac{cx^2 + dx + e}{x^2}}} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{\sqrt{a + b/x + c/x^2}}{\sqrt{c + d/x + e/x^2}} \right) \\ &= \sqrt{\frac{a}{c}}.\end{aligned}$$

Example 6.3.24. *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

$$\begin{aligned}\lim_{x \rightarrow \infty} \left(\frac{2}{\pi} \tan^{-1}(x) \right) &= \frac{2}{\pi} \cdot \frac{\pi}{2} = 1 \\ \lim_{x \rightarrow \infty} \left(\frac{\sqrt{4x^2 + 1}}{2x - 3} \right) &= \lim_{x \rightarrow \infty} \left(\frac{\sqrt{4 + 1/x^2}}{2 - 3/x} \right) = \sqrt{4}/2 = 1.\end{aligned}$$

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

Example 6.3.25. *(an example of what we can't do easily in this section) the infinite limit view of e . Consider the following limit:*

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x} \right)^x = e$$

If you can show that this definition is compatible with our previous implicit definition of e :

$$\lim_{h \rightarrow 0} \left(\frac{e^h - 1}{h} \right) = 1.$$

then I'd be impressed⁸. The $\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x} \right)^x$ is type 1^∞ and we have yet to develop the best tools to deal with such limits.

⁸page 444 of Stewart together with some thinking will do it

6.3.2 asymptotes in general

This section currently being remodelled. Should post new version later this semester.

.

6.4 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 $\pm\infty$.

I will give a good proof of the l'Hopital's rule in a later section, but my proof in this section is only for a relatively special case. l'Hopital's Rule holds in a context more general than the assumptions for my proof. It turns out that the *extended law of the mean* is needed to prove both l'Hopital's rule and Taylor's polynomial approximation theorem with Lagrange's form of the remainder. We postpone those arguments for later.

Theorem 6.4.1. 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 6.4.2. Notice $\lim_{x \rightarrow 0} \frac{\sin(x)}{x}$ is type $\frac{0}{0}$. Observe that

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\sin(x)}{x} &= \lim_{x \rightarrow 0} \frac{\cos(x)}{1} && \text{L'Hopital with } \frac{0}{0} \\ &= 1. \end{aligned}$$

We gave a geometric argument to prove this limit in the discussion leading up to the derivatives of sine and cosine. Given that the derivatives of sine and cosine require knowledge of this limit it is not surprising that this limit is trivially reproduced by l'Hopital's rule with the help of the derivative of sine and cosine. I used to think this proved this limit, but it is circular logic since we cannot know the derivative of sine is cosine unless we have already derived this limit.

Remark 6.4.3. *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 L^AT_EX. You should notice my notation in lecture is less cumbersome.

Example 6.4.4. *In this example we'll apply l'Hopital's rule twice to remove the indeterminacy.*

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{x^2 + x - 2}{2 + 3x} &= \lim_{x \rightarrow \infty} \frac{2x + 1}{2x} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\ &= \lim_{x \rightarrow \infty} \frac{2}{2} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\ &= 1.\end{aligned}$$

Please notice that the rule says to differentiate the numerator and denominator separately. There is no such rule as $\lim(f(x)) = \lim(f'(x))$.

Example 6.4.5.

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{\ln(x^2)}{\sqrt[3]{x}} &= \lim_{x \rightarrow \infty} \frac{\frac{2}{x}}{\frac{1}{3}x^{-\frac{2}{3}}} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\ &= \lim_{x \rightarrow \infty} \frac{6x^{\frac{2}{3}}}{x} \\ &= \lim_{x \rightarrow \infty} \frac{6}{\sqrt[3]{x}} \\ &= 0.\end{aligned}$$

Example 6.4.6.

$$\begin{aligned}\lim_{x \rightarrow 0^+} x e^{\frac{1}{x}} &= \lim_{x \rightarrow 0^+} \frac{e^{\frac{1}{x}}}{\frac{1}{x}} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\ &= \lim_{x \rightarrow 0^+} \frac{e^{\frac{1}{x}} \left(\frac{-1}{x^2} \right)}{\frac{-1}{x^2}} \\ &= \lim_{x \rightarrow 0^+} e^{\frac{1}{x}} \\ &= \infty.\end{aligned}$$

Remark 6.4.7. *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.

Example 6.4.8.

$$\begin{aligned}
\lim_{x \rightarrow \infty} e^{-x} x^2 &= \lim_{x \rightarrow \infty} \frac{x^2}{e^x} \\
&= \lim_{x \rightarrow \infty} \frac{2x}{e^x} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\
&= \lim_{x \rightarrow \infty} \frac{2}{e^x} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\
&= 0.
\end{aligned}$$

Apparently exponentials do grow faster than quadratic functions. It's not hard to see that if we replaced x with x^n for $n > 1$ then we could again find the same result from n -applications of $l'Hopital$'s rule.

Example 6.4.9.

$$\begin{aligned}
\lim_{x \rightarrow \infty} \frac{x^2}{\ln(x)} &= \lim_{x \rightarrow \infty} \frac{2x}{\frac{1}{x}} && L'Hopital \text{ on type } \frac{\infty}{\infty} \\
&= \lim_{x \rightarrow \infty} (2x^2) \\
&= \infty.
\end{aligned}$$

Apparently the natural logarithm grows slower than a quadratic function.

Example 6.4.10. I posed the question of how to prove that $\lim_{x \rightarrow \infty} (2^x - x) = \infty$. I simply claimed that exponentials grow faster than polynomials but I offered no justification in the preceding section. $l'Hopital$'s rule helps us argue that

$$\lim_{x \rightarrow \infty} (2^x/x) = \lim_{x \rightarrow \infty} (\ln(2)2^x/1) = \infty.$$

The limit above is useful to help analyze the following

$$\lim_{x \rightarrow \infty} (2^x - x) = \lim_{x \rightarrow \infty} \frac{\frac{2^x}{x} - 1}{\frac{1}{x}}.$$

Clearly the denominator tends to zero and the numerator tends to ∞ since the $2^x/x$ will dominate the -1 as $x \rightarrow \infty$. Therefore, $\lim_{x \rightarrow \infty} (2^x - x) = \infty$.

The type $\infty/0$ is divergent, the question is just if it goes to ∞ , $-\infty$ or oscillates. In this example it was clear the function was positive for $x \gg 0$ once I rewrote the expression.

Example 6.4.11. Here's what **not** to do.

$$\lim_{x \rightarrow 1} \frac{x}{x-1} = \lim_{x \rightarrow 1} \frac{1}{1} = 1.$$

Notice here I was wrong to apply $l'Hopital$'s rule because this limit was not indeterminate of type $0/0$ or ∞/∞ . Rather, this limit has the form $1/0$ which is divergent. Analysis from graph of $f(x) = \frac{x}{x-1} = 1 + \frac{1}{x-1}$ quickly shows that the limit point is at a VA and in fact the limit d.n.e.

Example 6.4.12.

$$\begin{aligned}
 \lim_{\theta \rightarrow 0^+} (\csc(\theta) - \cot(\theta)) &= \lim_{\theta \rightarrow 0^+} \left(\frac{1}{\sin(\theta)} - \frac{\cos(\theta)}{\sin(\theta)} \right) \\
 &= \lim_{\theta \rightarrow 0^+} \left(\frac{1 - \cos(\theta)}{\sin(\theta)} \right) \\
 &= \lim_{\theta \rightarrow 0^+} \left(\frac{\sin(\theta)}{\cos(\theta)} \right) \quad L'Hopital \text{ on type } \frac{0}{0} \\
 &= 0.
 \end{aligned}$$

Example 6.4.13.

$$\begin{aligned}
 \lim_{x \rightarrow 1} \left(\frac{1}{\ln(x)} - \frac{1}{x-1} \right) &= \lim_{x \rightarrow 1} \left(\frac{x-1-\ln(x)}{(x-1)\ln(x)} \right) \\
 &= \lim_{x \rightarrow 1} \left(\frac{1-\frac{1}{x}}{\frac{x-1}{x} + \ln(x)} \right) \quad L'Hopital \text{ on type } \frac{0}{0} \\
 &= \lim_{x \rightarrow 1} \left(\frac{1-\frac{1}{x}}{1-\frac{1}{x} + \ln(x)} \right) \quad L'Hopital \text{ on type } \frac{0}{0} \\
 &= \lim_{x \rightarrow 1} \left(\frac{\frac{1}{x^2}}{\frac{1}{x^2} + \frac{1}{x}} \right) \\
 &= \frac{1}{2}.
 \end{aligned}$$

6.4.1 concerning why l'Hopital's rule works

I'll begin with an intuitive argument. Suppose $\lim_{x \rightarrow a} f(x)/g(x)$ is type $(0/0)$. Since $x \rightarrow a$ in the limit it is reasonable to replace f and g with their linearizations based at $x = a$ hence

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \left[\frac{f(a) + f'(a)(x-a)}{g(a) + g'(a)(x-a)} \right] = \lim_{x \rightarrow a} \left[\frac{f'(a)(x-a)}{g'(a)(x-a)} \right] = \lim_{x \rightarrow a} \frac{f'(a)}{g'(a)} = \frac{f'(a)}{g'(a)}.$$

Note, I knew $f(a) = g(a) = 0$ since differentiability of f, g at $x = a$ implies continuity at $x = a$ and since the limit is form $0/0$ we know $0 = \lim_{x \rightarrow a} f(x) = f(a)$ and $0 = \lim_{x \rightarrow a} g(x) = g(a)$.

The proof for type ∞/∞ is technical and beyond the scope of this course. The proof of l'Hopital's rule in for type $(0/0)$ follows from an interesting generalization of the MVT which is due to Cauchy.

Theorem 6.4.14. *Cauchy's law of the mean.*

If f, g are continuous on $[a, b]$ and are differentiable on (a, b) then there exists $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Proof: the proof is similar to that for the MVT. Begin by defining

$$h(x) = (f(b) - f(a))(g(x) - g(a)) - (g(b) - g(a))(f(x) - f(a)).$$

Observe that h is continuous on $[a, b]$ and differentiable on (a, b) by virtue of what is given. Furthermore, $h(a) = 0$ and $h(b) = 0$ hence we may apply Rolle's theorem to conclude there exists $c \in (a, b)$ such that $h'(c) = 0$ hence

$$(f(b) - f(a))g'(c) - (g(b) - g(a))f'(c) \Rightarrow \boxed{\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}} \quad \square$$

There is a beautiful interpretation of this theorem in terms of parametric curves and their tangent vectors. See Taylor's *Advanced Calculus* page 109 (I happen to have the first edition, if you have another then look for "Cauchy's generalized law of the mean").

Proof of l'Hopital's Rule for type $\frac{0}{0}$ at finite limit point: Suppose we are given that $\lim_{x \rightarrow a} f'(x)/g'(x)$ exists and $f(a) = g(a) = 0$ and $g'(x) \neq 0$ for all $x \in B_\delta(a)$ for some $\delta > 0$. Suppose $x \in B_\delta(a)$ and $x > a$ then $g'(x) \neq 0$. Apply Cauchy's law of the mean to intervals of the form $[a, x]$, for each such interval we find c such that $a < c < x$ and

$$\frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)}$$

since $f(a) = g(a) = 0$. Therefore,

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{c \rightarrow a^+} \frac{f'(c)}{g'(c)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$$

and since the limit on the right exists it follows that the limit on the left exists and this proves half of l'Hopital's rule. If $x \in B_\delta(a)$ and $x < a$ then by almost the same argument we can show that $\lim_{x \rightarrow a^-} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^-} \frac{f'(x)}{g'(x)}$. \square

Proof of l'Hopital's Rule for type $\frac{0}{0}$ at infinite limit point: Apply the lemma to switch limits $x \rightarrow \pm\infty$ to corresponding limits of form $t \rightarrow 0^\pm$:

$$\begin{aligned} \lim_{x \rightarrow \pm\infty} \left[\frac{f(x)}{g(x)} \right] &= \lim_{t \rightarrow 0^\pm} \left[\frac{f(\frac{1}{t})}{g(\frac{1}{t})} \right] && \text{apply l'Hopital's rule} \\ &= \lim_{t \rightarrow 0^\pm} \left[\frac{f'(\frac{1}{t}) \cdot \frac{-1}{t^2}}{g'(\frac{1}{t}) \cdot \frac{-1}{t^2}} \right] \\ &= \lim_{t \rightarrow 0^\pm} \left[\frac{f'(\frac{1}{t})}{g'(\frac{1}{t})} \right] \\ &= \lim_{t \rightarrow \pm\infty} \left[\frac{f'(x)}{g'(x)} \right] && \square \end{aligned}$$

6.4.2 indeterminate powers

We have discussed indeterminate forms of type $0/0$, ∞/infy , $0 \cdot \infty$ and $\infty - \infty$ in some depth. There are three more forms to consider.

Definition 6.4.15. *forms of indeterminate 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 6.4.16. *the power lemma.*

Suppose that $f(x) > 0$ for points considered in limit,

$$\lim [f(x)]^{g(x)} = \exp(\lim g(x) \ln(f(x)))$$

where equality includes all cases including those divergent cases. In particular,

1. if $\lim[g(x) \ln(f(x))] = c \in \mathbb{R}$ then $\lim[f(x)]^{g(x)} = e^c$.
2. if $\lim[g(x) \ln(f(x))] = \infty$ then $\lim[f(x)]^{g(x)} = \infty$.
3. if $\lim[g(x) \ln(f(x))] = -\infty$ then $\lim[f(x)]^{g(x)} = 0$.

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} :

$$\lim [f(x)]^{g(x)} = \lim \left[\exp(\ln [f(x)]^{g(x)}) \right] = \exp(\lim \ln [f(x)]^{g(x)}) = \exp(\lim g(x) \ln(f(x))).$$

I leave the proof of the divergent cases for the reader. \square

Example 6.4.17. Calculate $\lim_{x \rightarrow 0^+} x^x$. We use the power lemma, consider

$$\lim_{x \rightarrow 0^+} x^x = \exp(\underbrace{\lim_{x \rightarrow 0^+} (x \ln(x))}_{\star})$$

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

$$\star = \lim_{x \rightarrow 0^+} (x \ln(x)) = \lim_{x \rightarrow 0^+} \frac{\ln(x)}{\frac{1}{x}} = \lim_{x \rightarrow 0^+} \frac{\frac{1}{x}}{\frac{-1}{x^2}} = \lim_{x \rightarrow 0^+} (-x) = 0$$

Hence, we find $\star = 0$ and returning to our original limit,

$$\lim_{x \rightarrow 0^+} x^x = \exp(0) = 1.$$

Example 6.4.18. Calculate $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{nx}$. We use the power lemma, consider we can pull out x since it is independent of n ,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^{nx} = \exp \left[x \underbrace{\lim_{n \rightarrow \infty} n \ln \left(1 + \frac{1}{n}\right)}_{\star} \right]$$

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

$$\star = \lim_{n \rightarrow \infty} n \ln \left(1 + \frac{1}{n}\right) = \lim_{n \rightarrow \infty} \frac{\ln(1 + \frac{1}{n})}{\frac{1}{n}} = \lim_{n \rightarrow \infty} \frac{\frac{1}{1+\frac{1}{n}} \cdot \frac{-1}{n^2}}{\frac{-1}{n^2}} = \lim_{n \rightarrow \infty} \frac{1}{1 + \frac{1}{n}} = 1.$$

Hence, we find $\star = 1$ and returning to our original limit,

$$\lim_{x \rightarrow 0^+} x^x = \exp(x\star) = \exp(x).$$

Example 6.4.19. Calculate $\lim_{x \rightarrow \infty} x^{\frac{1}{x}}$. We use the power lemma,

$$\lim_{x \rightarrow \infty} x^{\frac{1}{x}} = \exp \left[\underbrace{\lim_{x \rightarrow \infty} \frac{1}{x} \ln(x)}_{\star} \right]$$

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

$$\star = \lim_{x \rightarrow \infty} \frac{1}{x} \ln(x) = \lim_{x \rightarrow \infty} \frac{\ln(x)}{x} = \lim_{x \rightarrow \infty} \frac{\frac{1}{x}}{1} = 0.$$

Hence, we find $\star = 0$ and returning to our original limit,

$$\lim_{x \rightarrow 0^+} x^{\frac{1}{x}} = \exp(0) = 1.$$

Remark 6.4.20. notation for l'Hopital's rule

Some students prefer the method presented in Stewart and other texts. The idea is to call the desired limit y then take natural logarithm of limit and y and apply l'Hopital's rule etc... to determine the modified limit. Then you exponentiate both sides to find the answer. This technique is entirely equivalent to the \star -method I propose in this section and as such you are free to use it if for some reason you find my approach distasteful.

6.5 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. Let's see how it works order by order starting with the most silly case.

6.5.1 constant functions

Suppose we are given $f(a) = y_o$ then $T_o(x) = y_o$ is the zeroth Taylor polynomial for f centered at $x = a$. Usually you have to be very close to the center of the approximation for this to match the function.

6.5.2 linearizations again

Suppose we are given values for $f(a), f'(a)$ we seek to find $T_1(x) = c_o + c_1(x - a)$ which fits the given data. Note that

$$\begin{aligned} T_1(a) &= c_o + c_1(a - a) = f(a) & c_o &= f(a). \\ T'_1(a) &= c_1 = f'(a) & c_1 &= f'(a). \end{aligned}$$

Which gives us the first Taylor polynomial for f centered at a : $T_1(x) = f(a) + f'(a)(x - a)$. This function, I hope, is familiar from our earlier study of linearizations. The linearization at a is the best linear approximation to f near a .

6.5.3 quadratic approximation of function

Suppose we are given values for $f(a), f'(a)$ and $f''(a)$ we seek to find $T_2(x) = c_o + c_1(x - a) + c_2(x - a)^2$ which fits the given data. Note that

$$\begin{aligned} T_2(a) &= c_o + c_1(a - a) + c_2(a - a)^2 = f(a) & c_o &= f(a). \\ T'_2(a) &= c_1 + 2c_2(a - a) = f'(a) & c_1 &= f'(a). \\ T''_2(a) &= 2c_2 = f''(a) & c_2 &= \frac{1}{2}f''(a). \end{aligned}$$

Which gives us the first Taylor polynomial for f centered at a : $T_1(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$. We would hope this is the best quadratic approximation for f near the point $(a, f(a))$.

6.5.4 cubic approximation of function

Suppose we are given values for $f(a), f'(a), f''(a)$ and $f'''(a)$ we seek to find $T_3(x) = c_o + c_1(x - a) + c_2(x - a)^2 + c_3(x - a)^3$ which fits the given data. Note that

$$\begin{aligned} T_3(a) &= c_o + c_1(a - a) + c_2(a - a)^2 + c_3(a - a)^3 = f(a) & c_o &= f(a). \\ T'_3(a) &= c_1 + 2c_2(a - a) + 3c_3(a - a)^2 = f'(a) & c_1 &= f'(a). \\ T''_3(a) &= 2c_2 + 3 \cdot 2c_3(a - a) = f''(a) & c_2 &= \frac{1}{2}f''(a). \\ T'''_3(a) &= 3 \cdot 2c_3 = f'''(a) & c_3 &= \frac{1}{3 \cdot 2}f'''(a). \end{aligned}$$

Which gives us the first Taylor polynomial for f centered at a : $T_1(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + \frac{1}{6}f'''(a)(x - a)^3$. We would hope this is the best cubic approximation for f near the point $(a, f(a))$.

6.5.5 general case

Hopefully by now a pattern is starting to emerge. We see that $T_k(x) = T_{k-1}(x) + \frac{1}{k!}f^{(k)}(a)(x-a)^k$ where $k! = k(k-1)(k-2)\cdots 3 \cdot 2 \cdot 1$.

Definition 6.5.1. *Taylor polynomials.*

Suppose f is a function which has k -derivatives defined at a then the k -th Taylor polynomial for f is defined to be $T_k(x)$ where

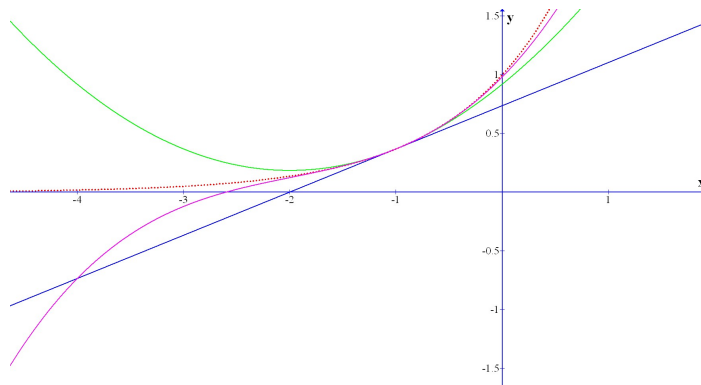
$$T_k(x) = \sum_{j=0}^k \frac{f^{(j)}(a)}{j!}(x-a)^j = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots + \frac{1}{k!}f^{(k)}(a)(x-a)^k$$

Let's examine a few examples before continuing with the theory.

Example 6.5.2. Suppose $f(x) = e^x$. Calculate the first few Taylor polynomials centered at $a = -1$. Derivatives of the exponential are easy enough to calculate; $f'(x) = f''(x) = f'''(x) = e^x$ therefore we find

$$\begin{aligned} T_0(x) &= \frac{1}{e} \\ T_1(x) &= \frac{1}{e} + \frac{1}{e}(x+1) \\ T_2(x) &= \frac{1}{e} + \frac{1}{e}(x+1) + \frac{1}{2e}(x+1)^2 \\ T_3(x) &= \frac{1}{e} + \frac{1}{e}(x+1) + \frac{1}{2e}(x+1)^2 + \frac{1}{6e}(x+1)^3. \end{aligned}$$

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. You can see that the cubic is the best approximation.



Example 6.5.3. Suppose $f(x) = \frac{1}{x-2} + 1$. Calculate the first few Taylor polynomials centered at $a = 1$. Observe

$$f(x) = \frac{1}{x-2} + 1, \quad f'(x) = \frac{-1}{(x-2)^2}, \quad f''(x) = \frac{2}{(x-2)^3}, \quad f'''(x) = \frac{-6}{(x-2)^4}$$

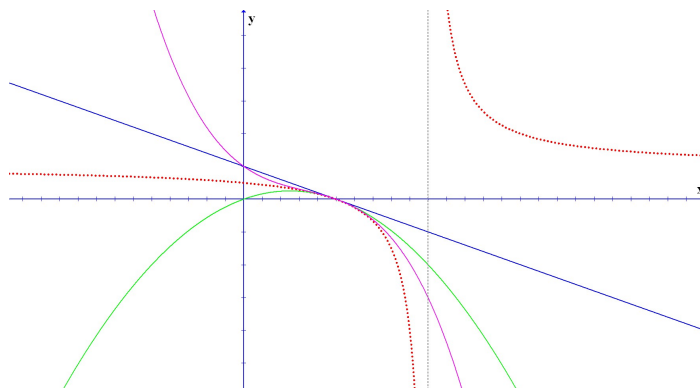
thus $f(1) = 0$, $f'(1) = -1$, $f''(1) = -2$ and $f'''(1) = -6$. Hence,

$$T_1(x) = -(x - 1)$$

$$T_2(x) = -(x - 1) + (x - 1)^2$$

$$T_3(x) = -(x - 1) + (x - 1)^2 - (x - 1)^3$$

The graph below shows $y = \frac{1}{x-2} + 1$ 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. You can see that the cubic is the best approximation. Also, you can see that the Taylor polynomials will not give a good approximation to $f(x)$ to the right of the VA at $x = 2$.



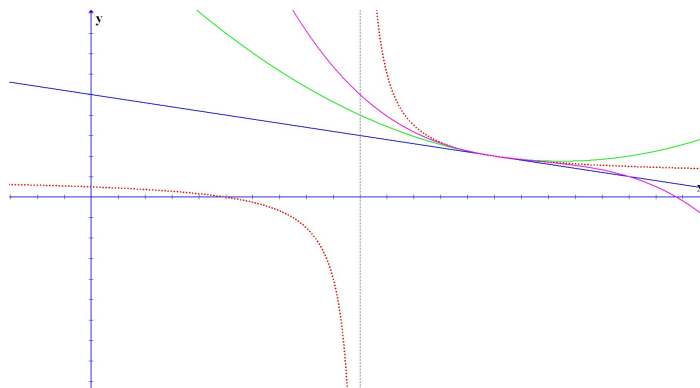
Now, for a given function we can find a Taylor polynomial relative to any point in the domain. They certainly are not unique. For example, we could expand about the center $a = 3$ to find

$$T_1(x) = 2 + (3 - x)$$

$$T_2(x) = 2 + (3 - x) + (3 - x)^2$$

$$T_3(x) = 2 + (3 - x) + (3 - x)^2 + (3 - x)^3.$$

The graph below uses the same color scheme. Notice this time the Taylor polynomials only work well to the right of the VA.



Example 6.5.4. Let $f(x) = \sin(x)$. It follows that

$$f'(x) = \cos(x), f''(x) = -\sin(x), f'''(x) = -\cos(x), f^{(4)}(x) = \sin(x), f^{(5)}(x) = \cos(x)$$

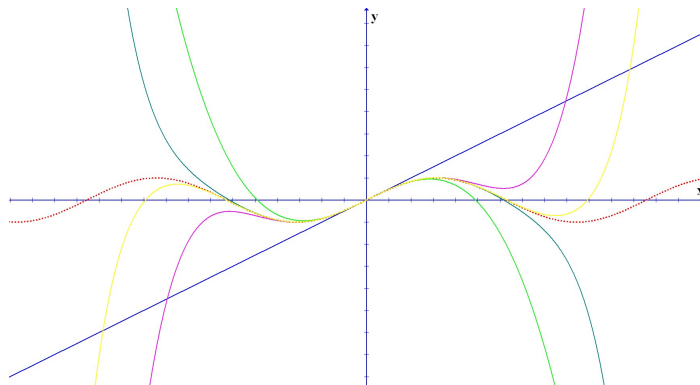
Hence, $f(0) = 0, f'(0) = 1, f''(0) = 0, f'''(0) = -1, f^{(4)}(0) = 0, f^{(5)}(0) = 1$. Therefore the Taylor polynomials of orders 1, 3, 5 are

$$T_1(x) = x \quad \text{blue graph}$$

$$T_3(x) = x - \frac{1}{6}x^3 \quad \text{green graph}$$

$$T_5(x) = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 \quad \text{purple graph}$$

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*⁹ function through a certain linear combination of its derivatives. The idea is that we can approximate a function by a polynomial¹⁰, 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.

Example 6.5.5. 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,

$$\frac{dE}{dv} = \frac{m_o v}{(1 - v^2/c^2)^{\frac{3}{2}}}$$

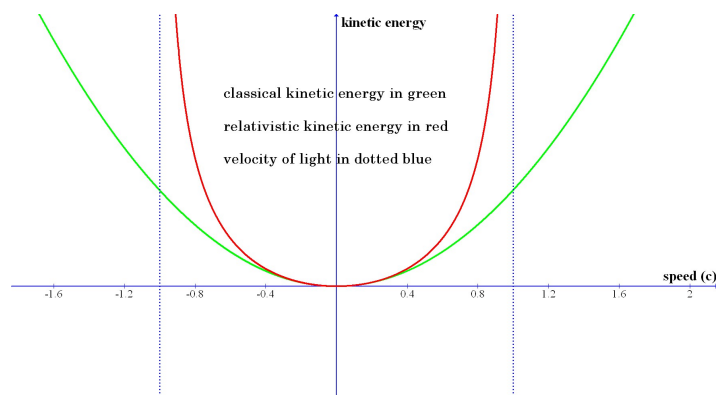
⁹for $p \in \mathbb{R}$ the notation $f \in C^\infty(p)$ means there exists a nbhd. of $p \in \mathbb{R}$ on which f has infinitely many continuous derivatives.

¹⁰there 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 $f(0) = 0$

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

$$E(v) \approx m_0 c^2 + \frac{1}{2} m_0 v^2 + \frac{3m_0}{8c^2} v^4$$

The constant term is the source of the famous equation $E = m_0 c^2$ and the quadratic term is precisely the classical kinetic energy. The last term is very small if $v \approx 0$. As $|v| \rightarrow c$ the values of the last term become more significant and they signal a departure from classical physics. I have graphed the relativistic kinetic energy $K = E - m_0 c^2$ (red) as well as the classical kinetic energy $K_{\text{Newtonian}} = \frac{m_0}{2} v^2$ (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? Recently some materialists have turned to something called the *anthropomorphic principle* as a tool to describe how this fortunate accident occurred. To the hardcore materialist the allowance of supernatural intervention is abhorrent. They prefer a universe without purpose. Personally, I prefer purpose. Moreover, it is my understanding of my place in this universe and our purpose to glorify God that make me expect to find laws of physics. Laws, or more correctly, approximations of physics reveal the glory of a God we cannot fully comprehend. I guess I digress... back to the math.

6.5.6 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 theorem¹¹, the one below is partially due to Lagrange.

Theorem 6.5.6. *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 $c \in \text{int}(I)$ 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 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 guaranteed 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:

$$f(b) = f(a) + f'(a)(b-a) + \dots + \frac{f^{(k-1)}(a)}{(k-1)!} (b-a)^{(k-1)} + \frac{M(b-a)^k}{k!}.$$

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

$$g(t) = -f(b) + f(t) + f'(t)(b-t) + \dots + \frac{f^{(k-1)}(t)}{(k-1)!} (b-t)^{(k-1)} + \frac{M(b-t)^k}{k!}$$

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

$$g(b) = -f(b) + f(b) + f'(b)(b-b) + \dots + \frac{f^{(k-1)}(b)}{(k-1)!} (b-b)^{(k-1)} + \frac{M(b-b)^k}{k!} = 0.$$

¹¹Chapter 7 of Apostol or Chapter II.6 of Edwards would be good additional readings if you wish to understand this material in added depth.

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,

$$\begin{aligned}
 g'(t) &= \frac{d}{dt}[-f(b) + f(t)] + \frac{d}{dt}[f'(t)(b-t)] + \cdots + \\
 &\quad + \frac{d}{dt}\left[\frac{f^{(k-1)}(t)}{(k-1)!}(b-t)^{(k-1)}\right] + \frac{d}{dt}\left[\frac{M(b-t)^k}{k!}\right] \\
 &= f'(t) - f'(t) + f''(t)(b-t) - \frac{1}{2}f''(t)2(b-t) + \cdots + \\
 &\quad + \frac{f^{(k)}(t)}{(k-1)!}(b-t)^{(k-1)} - \frac{f^{(k-1)}(t)}{(k-1)!}k(b-t)^{(k-2)} - \frac{Mk(b-t)^{k-1}}{k!} \\
 &= \frac{f^{(k)}(t)}{(k-1)!}(b-t)^{(k-1)} - \frac{Mk(b-t)^{k-1}}{k!} \\
 &= \frac{(b-t)^{(k-1)}}{(k-1)!} \left[f^{(k)}(t) - M \right]
 \end{aligned}$$

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 6.5.7. *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

$$|f(x) - f(c) - \frac{1}{2}f''(c)(x-c)^2| = \left| \frac{1}{6}f'''(\zeta_x)(x-c)^3 \right| \leq \frac{4M\delta^3}{3}$$

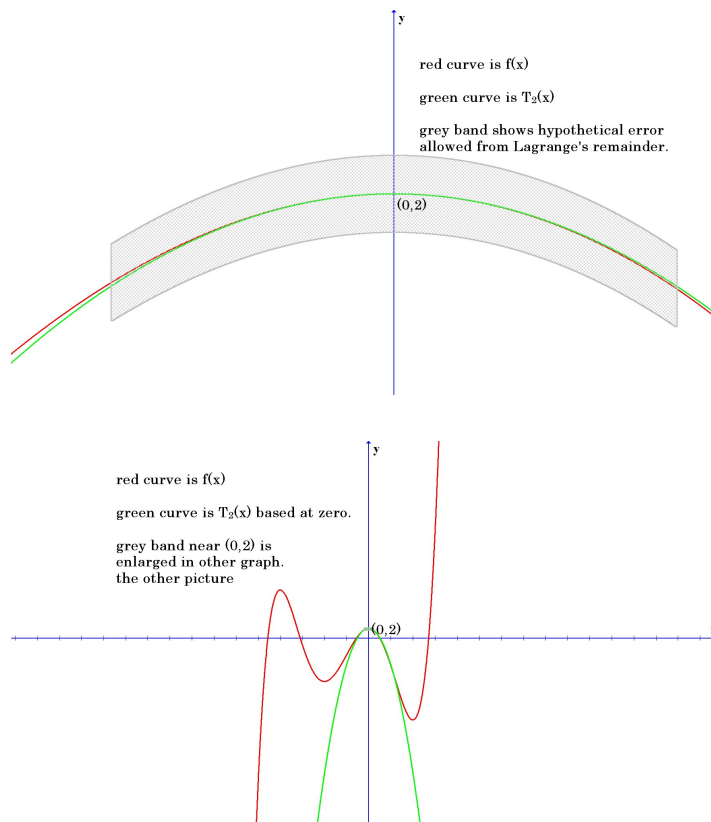
for all $x \in [c - \delta, c + \delta]$. This inequality reveals that we have $f(x) \approx f(c) + \frac{1}{2}f''(c)(x-c)^2$ as $\delta \rightarrow 0$. Therefore, locally the graph of the function resembles a parabola which either opens up or down at the critical point.

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 6.5.8. 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

$$f'''(x) = 360x^2 + 360x - 60 \quad \text{and} \quad f^{(4)}(x) = 720x + 360$$

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 6.5.9. 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 $x \in \mathbb{R}$.

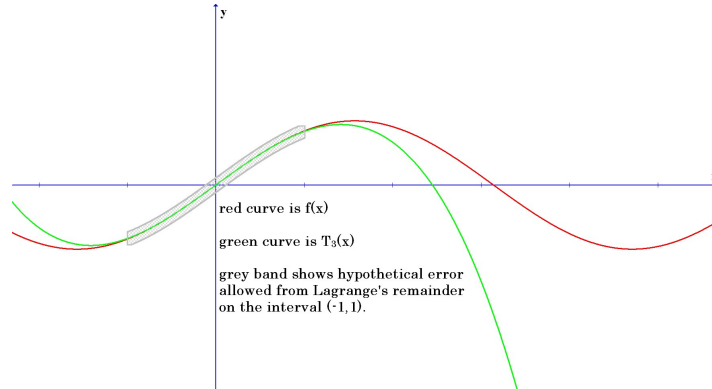
If we wish to bound the error to 0.1 on $-1 < x < 1$ then we to bound the remainder term as follows: (note $-1 < x < 1$ implies $l = 2$ and we just argued $M = 1$ is a good bound for any k)

$$|f(x) - T_k(x)| \leq \frac{Ml^{k+1}}{(k+1)!} = \frac{2^{k+1}}{(k+1)!} = E_k \leq 0.1$$

At this point I just start plugging various values of k until I find a value smaller than the desired bound. For this case,

$$E_1 = \frac{2^2}{2!} = 2, \quad E_2 = \frac{2^3}{3!} = \frac{4}{3}, \quad E_3 = \frac{2^4}{4!} = \frac{2}{3}, \quad E_4 = \frac{2^5}{5!} = \frac{32}{120} \approx 0.25, \quad E_5 = \frac{2^6}{6!} = \frac{64}{720} \approx 0.1$$

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:



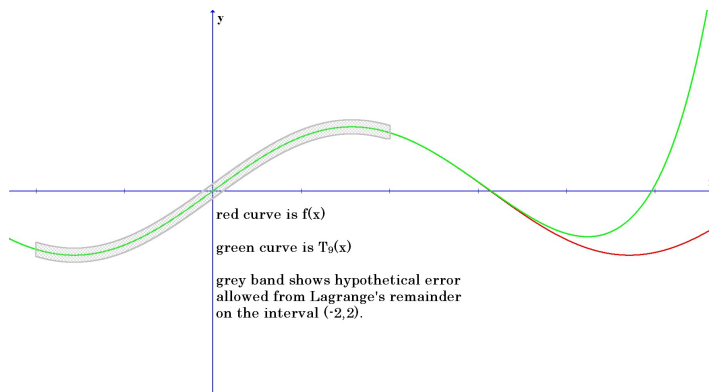
If we wish to bound the error to 0.1 on $-2 < x < 2$ then we to bound the remainder term as follows: (note $-2 < x < 2$ implies $l = 4$)

$$|f(x) - T_k(x)| \leq \frac{Ml^{k+1}}{(k+1)!} = \frac{4^{k+1}}{(k+1)!} = E_k \leq 0.1$$

At this point I just start plugging various values of k until I find a value smaller than the desired bound. For this case,

$$E_7 = \frac{4^8}{8!} \approx 1.6, \quad E_9 = \frac{4^{10}}{10!} \approx 0.3, \quad E_{11} = \frac{2^{12}}{12!} \approx 0.035$$

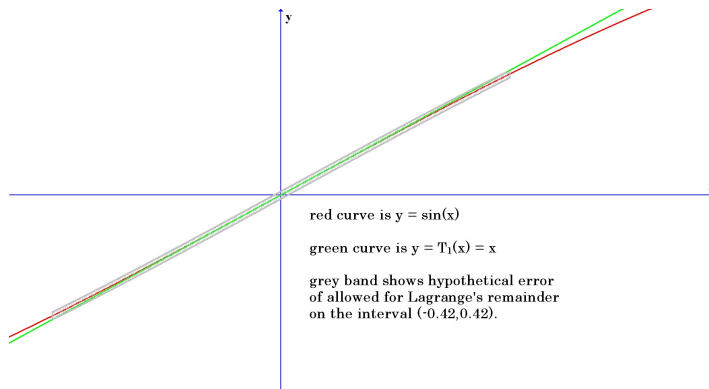
This shows that $T_{10}(x)$ will provide the desired accuracy. But, it just so happens that $T_9 = T_{10}$ in this case so we find $T_9(x) = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \frac{1}{5040}x^7 + \frac{1}{362880}x^9$ will suffice. In fact, as you can see below it fits the ± 0.1 tolerance band quite nicely well beyond the target interval of $-2 < x < 2$:



Example 6.5.10. 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 6.5.11. 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,

$$\begin{aligned} \sqrt{4.03} &= f(4) + f'(4)(4.03 - 4) + \frac{1}{2}f''(4)(4.03 - 4)^2 + R \\ &= 2 + \frac{1}{4} \frac{3}{100} - \frac{1}{64} \frac{9}{10000} + R \\ &= 2 + 0.0075 - 0.000014062 + R \\ &= 2.007485938 + R \end{aligned}$$

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 $\sqrt{4.03} = 2.00748598999$.

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 6.5.12. 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 emanating 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 $\frac{GmM}{R^2} = 9.83m/s^2$ which is hopefully familiar 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 $h = 1,000\text{m}$ then the formula $F/m = g$ approximates the true inverse square law to within 0.0031m/s^2 . At $h = 10,000\text{m}$ the error is 0.031m/s^2 . At $h = 100,000\text{m}$ the error is around 0.31m/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 6.5.13. *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 $0.1111\dots = \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \dots$ is a real number whereas $\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots$ 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.

6.5.7 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 6.5.14. *higher derivative tests.*

Suppose f has k continuous derivatives such that $f'(c) = f''(c) = \dots = f^{(k-1)}(c) = 0$ 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 6.5.15. 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).

Example 6.5.16. Consider $f(x) = x^5$. We can calculate $f'(x) = 5x^4$ therefore the only critical number is $c = 0$. Note that $f''(x) = 20x^3$, $f'''(x) = 60x^2$, $f^{(4)}(x) = 120x$, $f^{(5)}(x) = 120$. It follows that

$$f'(0) = f''(0) = f'''(0) = f^{(4)}(0) = 0$$

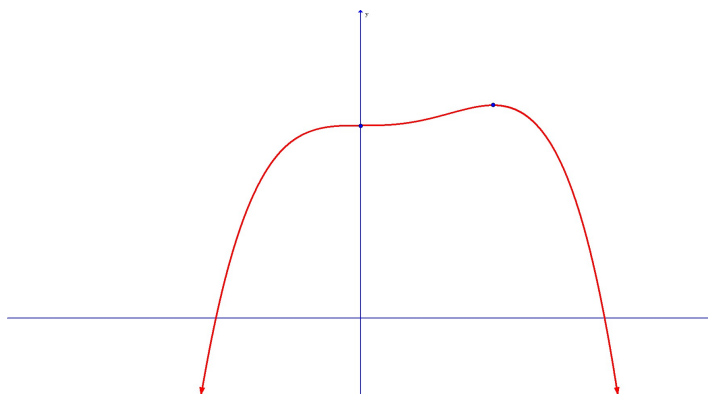
but $f^{(5)}(x) = 120 \neq 0$ therefore, by the higher derivative test, $f(0) = 0$ is not a local extrema of $f(x) = x^5$. Notice that this example would not have been covered by the second derivative test (but, the first derivative test would have covered it).

Example 6.5.17. 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 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 7

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, Lebesgue 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.

7.1 indefinite integration

Don't worry, the title of this section will make sense later.

7.1.1 why antidifferentiate?

The antiderivative is the opposite of the derivative in the following sense:

Definition 7.1.1. *antiderivative.*

If f and F are functions such that $F' = f$ then we say that F is an **antiderivative** of f .

Example 7.1.2. 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 7.1.3. 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 $G(t) = \frac{1}{k}e^{kt}$.

Example 7.1.4. Suppose $h(\theta) = \cos(\theta)$. Guess $H(\theta) = \sin(\theta)$ and note it works; $\frac{d}{d\theta}(\sin(\theta)) = \cos(\theta)$.

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 7.1.5. *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 6.1.10 we find $F_1(x) = F_2(x) + c$. \square

To understand the significance of this constant we should consider a physical question.

Example 7.1.6. 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 :

$$x(0) = 0 + c = 3 \quad \Rightarrow \quad c = 3 \quad \Rightarrow \quad \boxed{x(t) = t + 3}.$$

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 7.1.7. *indefinite integral.*

If f has an antiderivative F then the **indefinite integral** of f is given by:

$$\int f(x)dx = \{G(x) \mid G'(x) = f(x)\} = \{F(x) + c \mid c \in \mathbb{R}\}.$$

However, we will customarily drop the set-notation and simply write

$$\int f(x)dx = F(x) + c \quad \text{where } F'(x) = f(x).$$

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 7.1.8. *Consider the constant acceleration problem¹; we are given that $a = -g$ where $g = 9.8\text{m/s}^2$ and $a = \frac{dv}{dt}$. We can take the indefinite integral of the equation:*

$$\frac{dv}{dt} = -g \Rightarrow v(t) = \int -g dt = -gt + c_1.$$

Furthermore, if $v = \frac{dy}{dt}$ then

$$\frac{dy}{dt} = -gt + c_1 \Rightarrow y(t) = \int -gt + c_1 dt = -\frac{1}{2}gt^2 + c_1t + c_2.$$

Therefore, we find the velocity and position are given by formulas

$$v(t) = c_1 - gt \quad y(t) = c_2 + c_1t - \frac{1}{2}gt^2.$$

If we know the initial velocity is v_o and the initial position is y_o then

$$v(0) = v_o = c_1 - 0 \Rightarrow \boxed{v(t) = v_o - gt}$$

$$y(0) = y_o = c_2 - 0 - 0 \Rightarrow \boxed{y(t) = y_o + v_ot - \frac{1}{2}gt^2}$$

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.

¹here $F = ma$ 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.

Remark 7.1.9. *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 guessing².

Example 7.1.10. *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 7.1.3 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 $P(t) = P_o e^{kt}$.

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 7.1.11. *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.

²Actually, 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

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.

7.1.2 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 7.1.12. *basic properties of indefinite integration.*

Suppose f, g are functions with antiderivatives and $c \in \mathbb{R}$ then

$$\begin{aligned}\int [f(x) + g(x)] dx &= \int f(x) dx + \int g(x) dx \\ \int cf(x) dx &= c \int f(x) dx\end{aligned}$$

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 7.1.13. *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 7.1.14. *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 7.1.15. *exponential functions.* suppose $a > 0$ and $a \neq 1$,

$$\int a^x dx = \frac{1}{\ln(a)} a^x + c \quad \text{in particular:} \quad \int e^x dx = e^x + c$$

The exponential function has base $a = e$ and $\ln(e) = 1$ so the formulas are consistent.

Proposition 7.1.16. *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 calculate³ $\int \tan(x) dx$? 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 7.1.17. *hyperbolic functions.*

$$\int \sinh(x) dx = \cosh(x) + c \quad \int \cosh(x) dx = \sinh(x) + c$$

Naturally there are also basic antiderivatives for $\operatorname{sech}^2(x)$, $\operatorname{sech}(x)\tanh(x)$, $\operatorname{csch}^2(x)$ and $\operatorname{csch}(x)\coth(x)$ however I omit them for brevity and also as to not antagonize the struggling student at this juncture.

Proposition 7.1.18. *special algebraic and rational functions*

$$\begin{aligned} \int \frac{dx}{1+x^2} &= \tan^{-1}(x) + c & \int \frac{dx}{\sqrt{1-x^2}} &= \sin^{-1}(x) + c. \\ \int \frac{dx}{\sqrt{x^2-1}} &= \cosh^{-1}(x) + c & \int \frac{dx}{\sqrt{1+x^2}} &= \sinh^{-1}(x) + c. \\ \int \frac{dx}{1-x^2} &= \tanh^{-1}(x) + c. \end{aligned}$$

One can replace the expressions above with natural logs of certain algebraic functions. These identities are explored on page 466 of Stewart's 6-th edition. Page 488 has a nice summary of these basic integrals that we ought to memorize (although we have not covered $\tan(x)$ and $\cot(x)$ at this point)

³the answer is $\ln|\sec(x)| + c$ if you're curious and impatient.

7.1.3 examples of indefinite integration

Example 7.1.19.

$$\int dx = \int x^0 dx = \boxed{x + c}$$

Example 7.1.20.

$$\int \left[\sqrt{x} + \frac{1}{\sqrt[3]{x}} \right] dx = \int x^{\frac{1}{2}} dx + \int x^{\frac{-1}{3}} dx = \boxed{\frac{2}{3}x^{\frac{3}{2}} + \frac{3}{2}x^{\frac{2}{3}} + c}$$

Example 7.1.21.

$$\int \sqrt{13x^7} dx = \sqrt{13} \int x^{\frac{7}{2}} dx = \frac{\sqrt{13}}{\frac{9}{2}} x^{\frac{9}{2}} + c = \boxed{\frac{2\sqrt{13}}{9} x^4 \sqrt{x} + c}$$

Example 7.1.22.

$$\int \frac{dx}{3x^2} = \frac{1}{3} \int x^{-2} dx = \frac{-1}{3} x^{-1} = \boxed{\frac{-1}{3x} + c}$$

Example 7.1.23.

$$\int \frac{2x dx}{x^2} = 2 \int \frac{dx}{x} = 2 \ln |x| + c = \boxed{\ln(x^2) + c}$$

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 7.1.24.

$$\int 3e^{x+2} dx = 3 \int e^2 e^x dx = 3e^2 \int e^x dx = 3e^2(e^x + c_1) = \boxed{3e^{x+2} + c}$$

Example 7.1.25.

$$\begin{aligned} \int (x+2)^2 dx &= \int (x^2 + 4x + 4) dx \\ &= \int x^2 dx + 4 \int x dx + 4 \int dx \\ &= \boxed{\frac{1}{3}x^3 + 2x^2 + 4x + c} \end{aligned}$$

Example 7.1.26.

$$\int (2x^3 + 3) dx = \frac{2}{4}x^4 + 3x + c = \boxed{\frac{1}{2}x^4 + 3x + c}$$

Example 7.1.27.

$$\int (2^x + 3 \cosh(x)) dx = \int 2^x dx + 3 \int \cosh(x) dx = \boxed{\frac{1}{\ln(2)} 2^x + 3 \sinh(x) + c}$$

Example 7.1.28.

$$\int \frac{2x^3 + 3}{x} dx = \int \left[\frac{2x^3}{x} + \frac{3}{x} \right] dx = \int 2x^2 dx + 3 \int \frac{dx}{x} = \boxed{\frac{2}{3}x^3 + 3 \ln |x| + c}$$

Example 7.1.29.

$$\int \frac{x^2}{1+x^2} dx = \int \frac{1+x^2-1}{1+x^2} dx = \int \left[1 - \frac{1}{1+x^2}\right] dx = \boxed{x - \tan^{-1}(x) + c}$$

Example 7.1.30.

$$\begin{aligned} \int \sin(x+3) dx &= \int [\sin(x) \cos(3) + \sin(3) \cos(x)] dx \\ &= \cos(3) \int \sin(x) dx + \sin(3) \int \cos(x) dx \\ &= -\cos(3) [\cos(x) + c_1] + \sin(3) [\sin(x) + c_2] \\ &= \sin(3) \sin(x) - \cos(3) \cos(x) + c \\ &= \boxed{-\cos(x+3) + c} \end{aligned}$$

Incidentally, we find a better way to do this later with the technique of u -substitution.

Example 7.1.31.

$$\int \frac{1}{\cos^2(x)} dx = \int \sec^2(x) dx = \boxed{\tan(x) + c}$$

Example 7.1.32.

$$\int \frac{dx}{x^2 + \cos^2(x) + \sin^2(x)} = \int \frac{dx}{x^2 + 1} = \boxed{\tan^{-1}(x) + c}$$

Every example in this section is easily checked by differentiation.

7.2 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 millennia 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 $3.1408 < \pi \approx 3.1416 < 3.1429$). 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 carefully 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.

7.2.1 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 objects⁴.

Definition 7.2.1. *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 $f : U \rightarrow \mathbb{R}$ 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

$$\{a_j\}_{j=0}^{\infty} = \{a_0, a_1, a_2, \dots\} \quad \{b_n\}_{n=1}^{\infty} = \{b_1, b_2, b_3, \dots\}$$

Example 7.2.2. *Sequences may defined by a formula: $a_n = n$ for all $n \in \mathbb{N}$ gives*

$$\{a_n\}_{n=1}^{\infty} = \{1, 2, 3, \dots\}.$$

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 7.2.3. *finite sum notation.*

Suppose $a_j \in \mathbb{R}$ for $j \in \mathbb{N}$. Then define:

$$\sum_{j=1}^1 a_j = a_1 \quad \sum_{j=1}^n a_j = \sum_{j=1}^{n-1} a_j + a_n$$

for $n \geq 2$. This iterative definition gives us the result that

$$\sum_{j=1}^n a_j = a_1 + a_2 + \dots + a_n.$$

The variable j is called the **dummy index of summation**. Moreover, sums such as

$$\sum_{j=j_1}^{j_N} a_j = \underbrace{a_{j_1} + a_{j_2} + \dots + a_{j_N}}_{N \text{ summands}}$$

can be carefully defined by a similar iterative formula.

⁴sequences of functions, matrices or even spaces are studied in modern mathematics

Example 7.2.4. 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 $1+2+3+\dots+50$ with such ease? Gauss understood that generally

$$\sum_{j=1}^n j = \frac{n(n+1)}{2}$$

For example,

$$\begin{aligned} a_1 &= \frac{1(1+1)}{2} = 1, & a_2 &= \frac{2(2+1)}{2} = 3, & a_3 &= \frac{3(3+1)}{2} = 6, \\ a_4 &= \frac{4(4+1)}{2} = 10, & \dots, & a_{100} &= \frac{(100)(101)}{2} = 50(101) = 5050. \end{aligned}$$

What method of proof is needed to prove results such as this? The method is called "proof by mathematical induction". We discuss it in some depth in the Math 200 course. 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 7.2.5. special formulas for finite sums.

$$\sum_{k=1}^n 1 = n, \quad \sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4}.$$

The following results are less surprising but are even more useful:

Proposition 7.2.6. *finite sum properties. suppose $a_k, b_k, c \in \mathbb{R}$ for all k and let $n, m \in \mathbb{N}$ such that $m < n$,*

$$\begin{aligned} (i.) \quad & \sum_{k=1}^n a_k + \sum_{k=1}^n b_k = \sum_{k=1}^n (a_k + b_k), \\ (ii.) \quad & \sum_{k=1}^n c a_k = c \sum_{k=1}^n a_k, \\ (iii.) \quad & \sum_{k=1}^n a_k = \sum_{k=1}^m a_k + \sum_{k=m+1}^n a_k. \end{aligned}$$

Proof: begin with (i.). The proof is by induction on n . Note that (i.) is true for $n = 1$ since $\sum_{k=1}^1 a_k + \sum_{k=1}^1 b_k = a_1 + b_1 = \sum_{k=1}^1 (a_k + b_k)$. Suppose that (i.) is true for n and consider

$$\begin{aligned} \sum_{k=1}^{n+1} a_k + \sum_{k=1}^{n+1} b_k &= \sum_{k=1}^n a_k + a_{n+1} + \sum_{k=1}^n b_k + b_{n+1} && \text{by defn. of } \sum \\ &= \sum_{k=1}^n (a_k + b_k) + a_{n+1} + b_{n+1} && \text{by induction hypothesis for } n \\ &= \sum_{k=1}^{n+1} (a_k + b_k) && \text{by defn. of } \sum \end{aligned}$$

Therefore, (i.) true for n implies (i.) is true for $n + 1$ hence by proof by mathematical induction we conclude (i.) is true for all $n \in \mathbb{N}$. The proof for (ii.) is similar. We leave the proof of (iii.) to the reader. \square

We would like to have sums with many terms in the sections that follow. In fact, we will want to let $n \rightarrow \infty$. 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 7.2.7. *limit of a sequence.*

We say the sequence $\{a_n\}$ converges to $L \in \mathbb{R}$ and denote

$$\lim_{n \rightarrow \infty} a_n = L$$

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 7.2.8. *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 $n \in \text{dom}(\{a_n\})$. If $\lim_{x \rightarrow \infty} f(x) = L \in \mathbb{R}$ then $\lim_{n \rightarrow \infty} a_n = L$.

Proof: assume $\lim_{x \rightarrow \infty} f(x) = L \in \mathbb{R}$ 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, $\lim_{n \rightarrow \infty} a_n = L$. \square

The converse is not true. You could extend a sequence so that it gave a function of a continuous variable which diverged. Just imagine a function which oscillates wildly between the natural numbers.

Definition 7.2.9. *infinite sum.*

$$\sum_{k=1}^{\infty} a_k = \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k.$$

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 $\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k$ 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.

7.2.2 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.

Definition 7.2.10. *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, $x_0 = a$ and $x_n = b$.

The closed interval $[a, b]$ is a union of n -subintervals of length Δx . Note that the closed interval $[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup \cdots \cup [x_{n-1}, x_n]$. The following rule is an intuitively obvious way to calculate the signed-area.

Definition 7.2.11. *left endpoint rule (L_n).*

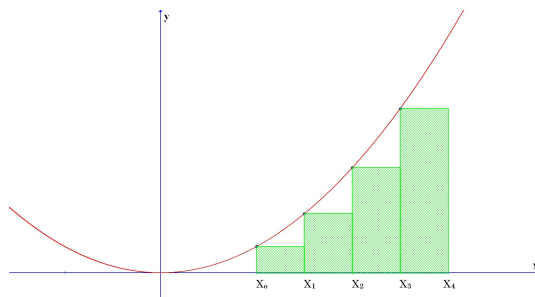
Suppose that $[a, b] \subseteq \text{dom}(f)$ then we define

$$L_n = \sum_{j=0}^{n-1} f(x_j) \Delta x = [f(x_0) + f(x_1) + \cdots + f(x_{n-1})] \Delta x.$$

Example 7.2.12. 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 $x_0 = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ 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 7.2.13. *right endpoint rule (R_n).*

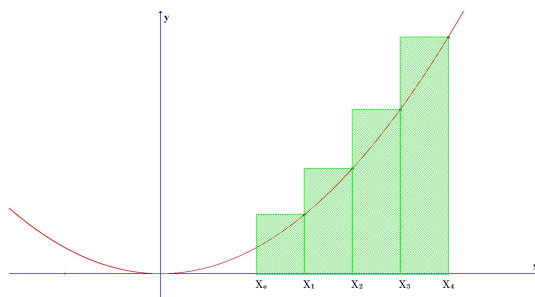
Suppose that $[a, b] \subseteq \text{dom}(f)$ then we define

$$R_n = \sum_{j=1}^n f(x_j) \Delta x = [f(x_1) + f(x_2) + \cdots + f(x_n)] \Delta x.$$

Example 7.2.14. 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 $x_0 = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ 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 7.2.15. *midpoint rule (M_n).*

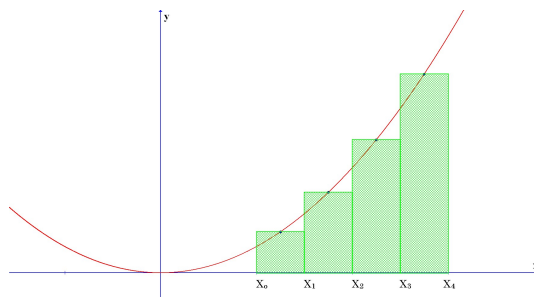
Suppose that $[a, b] \subseteq \text{dom}(f)$ and denote the midpoints by $\bar{x}_k = \frac{1}{2}(x_k + x_{k-1})$ and define

$$M_n = \sum_{j=1}^n f(\bar{x}_j) \Delta x = [f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)] \Delta x.$$

Example 7.2.16. 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 $n \rightarrow \infty$ 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 $n \rightarrow \infty$. Careful proof of this seemingly obvious claim is beyond the scope of this course.

Example 7.2.17. 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 $n \rightarrow \infty$. 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 $n \rightarrow \infty$ thus

$$\lim_{n \rightarrow \infty} R_n = 2 + 4 + \frac{16}{6} = \frac{26}{3} \cong 8.6667.$$

Challenge: show L_n and M_n also have limit $\frac{26}{3}$ as $n \rightarrow \infty$.

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 .

7.2.3 Riemann sums and the definite integral

In the last section we claimed that it was intuitively clear that as $n \rightarrow \infty$ 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 $\Delta x \rightarrow 0$ as $n \rightarrow \infty$. Therefore, the subinterval $[x_j, x_j + \Delta x] \rightarrow \{x_j\}$ 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 7.2.18. *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 7.2.19. *definite integrals.*

Suppose that f is continuous on $[a, b]$, the **definite integral** of f from a to b is defined to be $\lim_{n \rightarrow \infty} \mathcal{R}_n$ in particular we denote:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \mathcal{R}_n = \lim_{n \rightarrow \infty} \left[\sum_{j=1}^n f(x_k^*) \Delta x \right].$$

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 $a < b$

$$\int_b^a f(x) dx = - \int_a^b f(x) dx \quad \text{and} \quad \int_a^a f(x) dx = 0.$$

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 possible⁵.

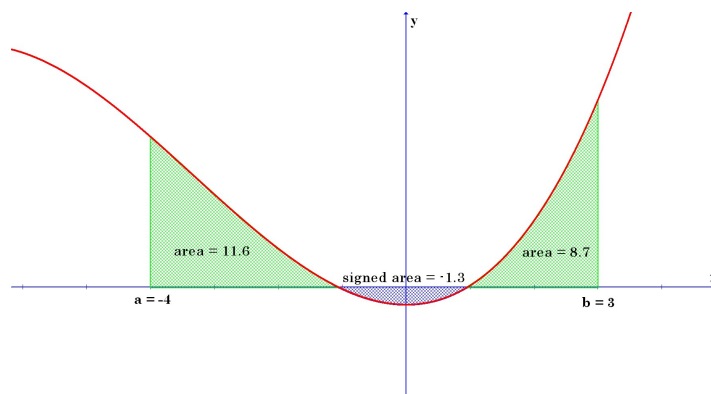
⁵ the Riemann-Stieltjes integral or Lebesgue are generalizations of this the basic Riemann integral. Riemann-Stieltjes integral might be covered in some undergraduate analysis courses whereas Lebesgue's measure theory is typically a graduate analysis topic.

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 $-4 \leq x \leq 3$. 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 7.2.20. 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

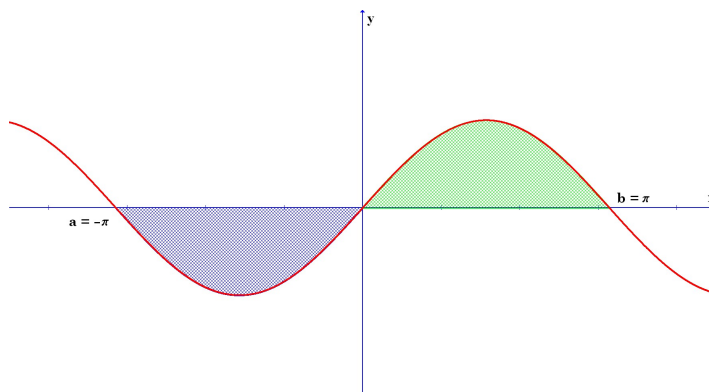
$$R_n = \sum_{j=1}^n \sin(x_j^*) \Delta x = \sum_{j=1}^n \sin\left(\frac{j\pi}{n}\right) \frac{\pi}{n} \Rightarrow \int_0^\pi \sin(x) dx = \lim_{n \rightarrow \infty} \sum_{j=1}^n \sin\left(\frac{j\pi}{n}\right) \frac{\pi}{n}.$$

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.

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.



7.2.4 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.

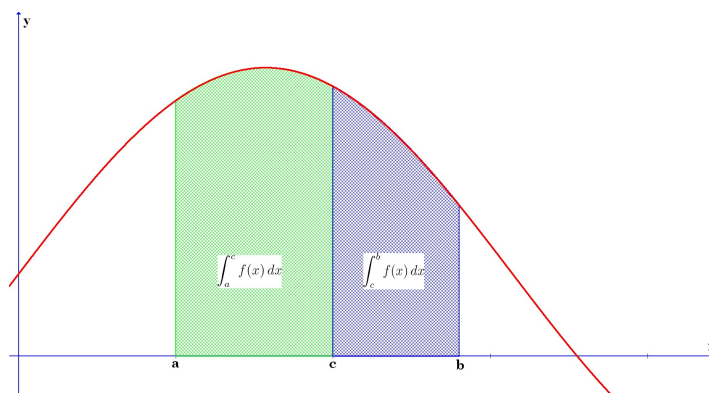
Proposition 7.2.21. *algebraic properties of definite integration.*

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,

$$\begin{aligned}
 \int_a^b [f(x) + \alpha g(x)] dx &= \lim_{n \rightarrow \infty} \left[\sum_{j=1}^n [f(x_k^*) + \alpha g(x_k^*)] \Delta x \right] \\
 &= \lim_{n \rightarrow \infty} \left[\sum_{j=1}^n f(x_k^*) \Delta x + \alpha \sum_{j=1}^n g(x_k^*) \Delta x \right] \\
 &= \lim_{n \rightarrow \infty} \left[\sum_{j=1}^n f(x_k^*) \Delta x \right] + \alpha \lim_{n \rightarrow \infty} \left[\sum_{j=1}^n g(x_k^*) \Delta x \right] \\
 &= \int_a^b f(x) dx + \alpha \int_a^b g(x) dx
 \end{aligned}$$

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.). 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:

$$x_0 = a < x_1 < x_2 < \cdots < x_j \leq c \leq x_{j+1} < \cdots < x_n = a + n\Delta x = b.$$

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 $n \rightarrow \infty$ the following ratios hold (if $x_j = c$ then these are exact, however clearly $x_j \rightarrow c$ as $n \rightarrow \infty$):

$$\Delta x = \frac{b-a}{n} = \frac{c-a}{j} = \frac{b-c}{n-j}$$

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 :

$$\begin{aligned}
 \int_a^b f(x) dx &= \lim_{n \rightarrow \infty} \left[\sum_{k=1}^n f(x_k^*) \Delta x \right] \\
 &= \lim_{n \rightarrow \infty} \left[\sum_{k=1}^j f(x_k^*) \Delta x + \sum_{k=j+1}^n f(x_k^*) \Delta x \right] \\
 &= \lim_{j \rightarrow \infty} \left[\sum_{k=1}^j f(z_k^*) \frac{c-a}{j} \right] + \lim_{p \rightarrow \infty} \left[\sum_{l=1}^p f(y_l^*) \frac{b-c}{p} \right]
 \end{aligned}$$

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 $n \rightarrow \infty$ with that of $p = n - j \rightarrow \infty$ 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 $n \rightarrow \infty$ implies $n - j \rightarrow \infty$ as $b > c$ and $b > a$ by assumption. Likewise, we replaced $n \rightarrow \infty$ with $j \rightarrow \infty$ 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 $n \rightarrow \infty$

implies $j \rightarrow \infty$. Finally, denote $\Delta y = \frac{c-a}{j}$ and $\Delta z = \frac{b-c}{p}$ to obtain

$$\begin{aligned} \int_a^b f(x) dx &= \lim_{j \rightarrow \infty} \left[\sum_{k=1}^j f(z_k^*) \Delta z \right] + \lim_{p \rightarrow \infty} \left[\sum_{l=1}^p f(y_l^*) \Delta y \right] \\ &= \int_a^c f(z) dz + \int_c^b f(y) dy \\ &= \int_a^c f(x) dx + \int_c^b f(x) dx. \end{aligned}$$

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 (Edward Witten is probably the most famous example of this type currently) and then we 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 7.2.22. *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 $\int_a^b f(x) dx \geq 0$,
- (ii.) if $f(x) \geq g(x)$ for all $x \in [a, b]$ then $\int_a^b f(x) dx \geq \int_a^b g(x) dx$,
- (iii.) if $m \leq f(x) \leq M$ for all $x \in [a, b]$ then $m(b-a) \leq \int_a^b f(x) dx \leq M(b-a)$.

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 $a = x_0, b = x_n$ 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:

$$f(x_k^*) \geq 0 \text{ and } \Delta x = \frac{b-a}{n} > 0 \Rightarrow \sum_{k=1}^n f(x_k^*) \Delta x \geq 0$$

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, $\lim_{n \rightarrow \infty} \mathcal{R}_n \geq \lim_{n \rightarrow \infty} (0) = 0$ 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:

$$\int_a^b h(x)dx \geq 0 \Rightarrow \int_a^b [f(x) - g(x)]dx \geq 0 \Rightarrow \int_a^b f(x)dx - \int_a^b g(x)dx \geq 0$$

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

$$mn \frac{b-a}{n} \leq \sum_{k=1}^n f(x_k^*) \Delta x \leq Mn \frac{b-a}{n} \Rightarrow m(b-a) \leq \sum_{k=1}^n f(x_k^*) \Delta x \leq M(b-a).$$

Apply the sequential limit squeeze theorem and take the limit as $n \rightarrow \infty$ to find

$$m(b-a) \leq \lim_{n \rightarrow \infty} \left[\sum_{k=1}^n f(x_k^*) \Delta x \right] \leq M(b-a)$$

This proves (iii.). \square

One easy fact to glean from the proof of (iii.) is the following:

Corollary 7.2.23. *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 .

7.3 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.

7.3.1 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 7.3.1. *area function.*

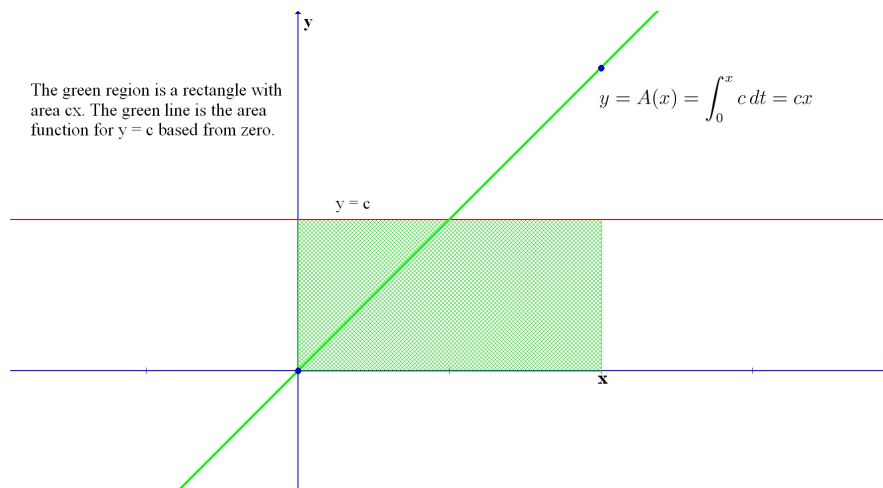
Given f and a point a we define the **area function** of f relative to a as follows:

$$A(x) = \int_a^x f(t) dt.$$

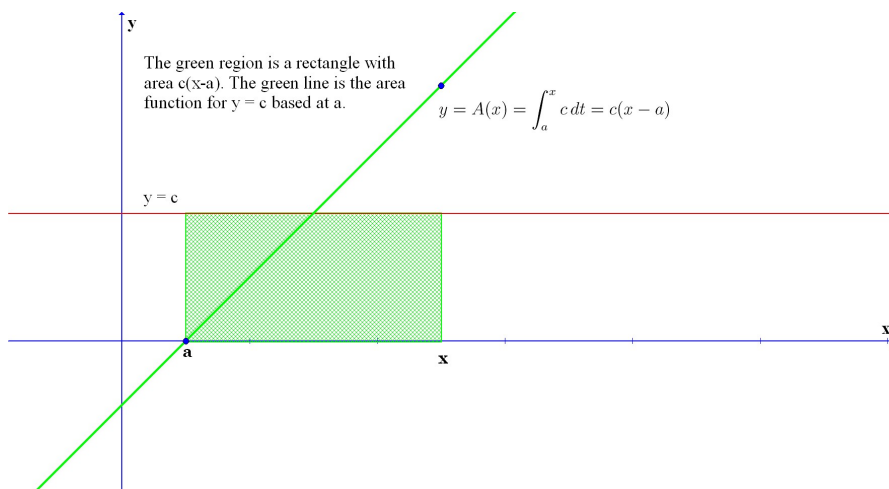
We say that $A(x)$ is the signed-area bounded by f on $[a, x]$.

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.

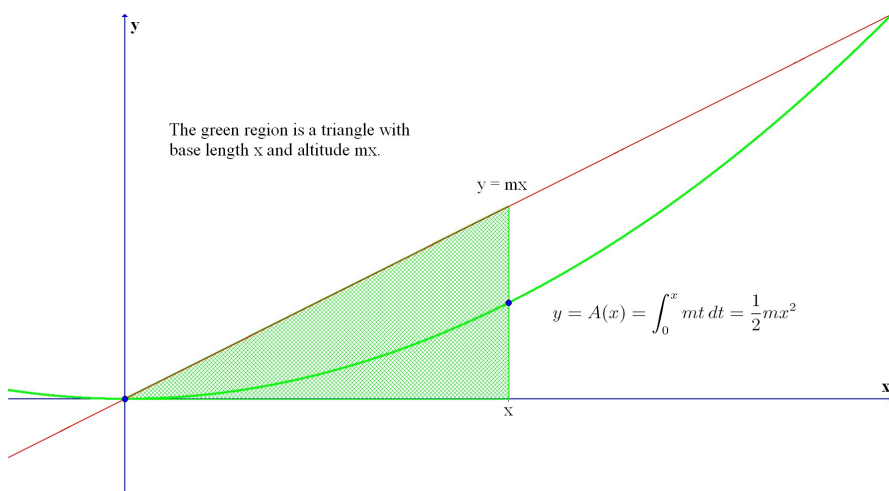
Area function of rectangle: let $f(t) = c$ then the area bounded between $t = 0$ and $t = x$ is simply length (x) times height (c). By geometry we have that $A(x) = \int_0^x c dt = cx$, see the picture below:



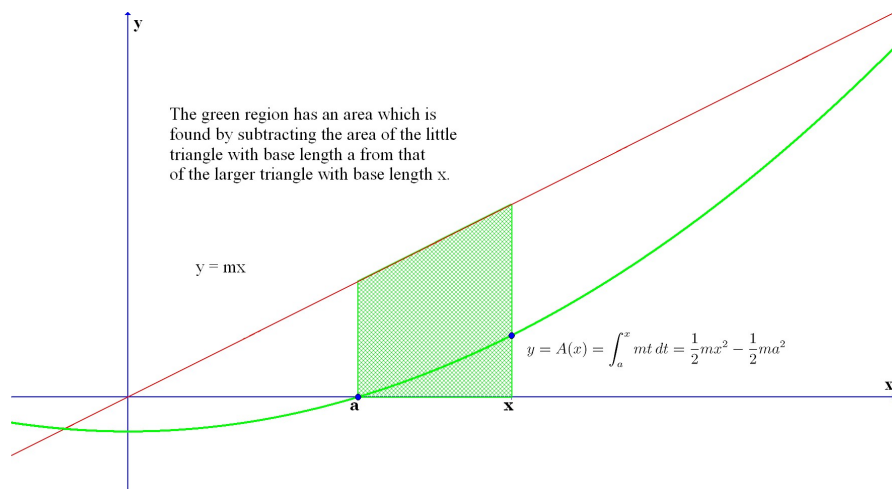
If we positioned the rectangle at $a \leq t \leq x$ then length becomes $(x - a)$ and the height is still (c). Therefore, by geometry, $A(x) = \int_a^x c dt = c(x - a) = cx - ca$. Again, see the picture below where I have pictured a particular x but I have graphed $y = A(t)$ for many t besides x . You 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:



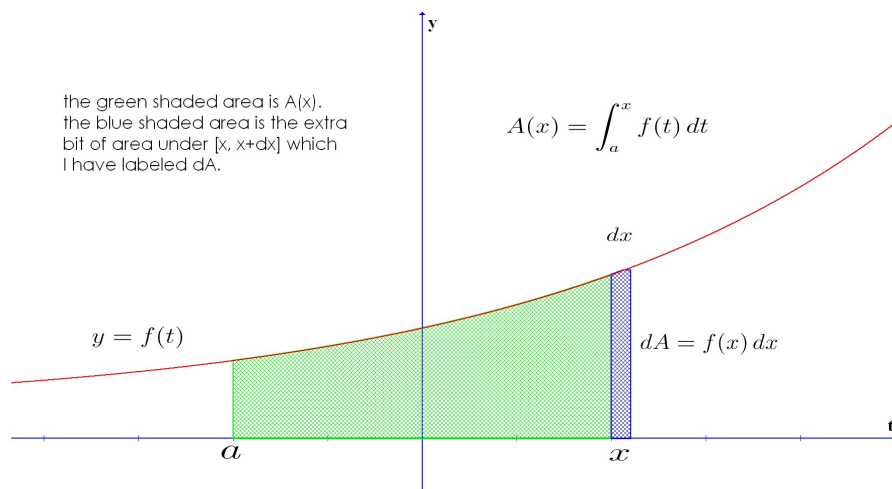
We calculate the area bounded by $y = mt$ for $a \leq t \leq x$ by subtracting the area of the small triangle from $0 \leq t \leq a$ from the area of the larger triangle $0 \leq t \leq x$ as pictured below. Thus from geometry we find $A(x) = \int_a^x mt \, dt = \frac{1}{2}mx^2 - \frac{1}{2}ma^2$.



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 might ask you to show this in a homework). 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 7.3.2. *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

$$(x+h-x)f(u) \leq \int_x^{x+h} f(t) dt \leq (x+h-x)f(v) \Rightarrow f(u) \leq \frac{1}{h} \int_x^{x+h} f(t) dt \leq f(v)$$

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 $\lim_{h \rightarrow 0} u = x$ and $\lim_{h \rightarrow 0} v = x$. Therefore, by continuity of f , $\lim_{h \rightarrow 0} f(u) = f(x)$ and $\lim_{h \rightarrow 0} f(v) = f(x)$. Remember, $f(u) \leq \frac{1}{h} \int_x^{x+h} f(t) dt \leq f(v)$ and apply the squeeze theorem to deduce:

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} f(t) dt = f(x)$$

Consequently,

$$\lim_{h \rightarrow 0} \frac{A(x+h) - A(x)}{h} = f(x)$$

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. However, such problems are fairly rare. You can read about the Fresnel function in the text.

Remark 7.3.3. *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:

$$\begin{aligned}\int_0^x f(u) du &= \lim_{n \rightarrow \infty} \sum_{i=1}^n f(u_i^*) \Delta u & \Delta u &= (x - 0)/n = x/n \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n 3 \frac{x}{n} \\ &= \lim_{n \rightarrow \infty} \frac{3x}{n} \sum_{i=1}^n 1 \\ &= \lim_{n \rightarrow \infty} \frac{3x}{n} n \\ &= \lim_{n \rightarrow \infty} 3x \\ &= 3x.\end{aligned}$$

7.3.2 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 7.3.4. *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,

$$F(a) = \lim_{x \rightarrow a^+} F(x) = \lim_{x \rightarrow a^+} [A(x) + c] = A(a) + c = \int_a^a f(t) dt + c = c$$

and

$$F(b) = \lim_{x \rightarrow b^-} F(x) = \lim_{x \rightarrow b^-} [A(x) + c] = A(b) + c = \int_a^b f(t) dt + c.$$

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. \square

Example 7.3.5. We return to Example 7.2.20 where we were stuck due to an incalculable summation. We wish to calculate $\int_0^\pi \sin(x) dx$. Observe that $F(x) = -\cos(x)$ has $F'(x) = \sin(x)$ hence this is a valid antiderivative for the given integrand $\sin(x)$. Apply the FTC part II to find the area:

$$\int_0^\pi \sin(x) dx = F(\pi) - F(0) = -\cos(\pi) + \cos(0) = 2.$$

Obviously this is much easier than calculation from the definition of the Riemann integral.

Definition 7.3.6. *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:

$$\int_a^b f(x) dx = F(x) \Big|_a^b = F(b) - F(a).$$

7.3.3 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 $[a, b]$; $x_0 = a, x_1 = a + \Delta x, \dots, x_n = b$ 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 $[x_0, x_1]$ tells us that there exists $x_1^* \in [x_0, x_1]$ such that

$$F'(x_1^*) = \frac{F(x_1) - F(x_0)}{x_1 - x_0} = \frac{F(x_1) - F(x_0)}{\Delta x}$$

Notice that this tells us that $F'(x_1^*)\Delta x = F(x_1) - F(x_0)$. But, $F'(x) = f(x)$ so we have found that $f(x_1^*)\Delta x = F(x_1) - F(x_0)$. In other words, the area under $y = f(x)$ for $x_0 \leq x \leq x_1$ 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:

$$\begin{aligned}
 \int_a^b f(x) dx &= \lim_{n \rightarrow \infty} \left(\sum_{j=1}^n f(x_j^*) \Delta x \right) \\
 &= \lim_{n \rightarrow \infty} \left(\sum_{j=1}^n [F(x_j) - F(x_{j-1})] \right) \\
 &= \lim_{n \rightarrow \infty} (F(x_1) - F(x_0) + F(x_2) - F(x_1) + \cdots + F(x_n) - F(x_{n-1})) \\
 &= \lim_{n \rightarrow \infty} (F(x_n) - F(x_0)) \\
 &= \lim_{n \rightarrow \infty} (F(b) - F(a)) \\
 &= F(b) - F(a). \square
 \end{aligned}$$

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 7.3.7. *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:

$$\int_{u(x)}^{v(x)} f(t) dt = F(v(x)) - F(u(x))$$

⁶note 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

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.

Example 7.3.8.

$$\frac{d}{dx} \int_3^x \cos(\sqrt{t}) dt = \cos(\sqrt{x}) \frac{d(\sqrt{x})}{dx} - \cos(\sqrt{3}) \frac{d(3)}{dx} = \boxed{\cos(\sqrt{x}) \frac{1}{2\sqrt{x}}}.$$

Example 7.3.9.

$$\frac{d}{dx} \int_{e^x}^{x^3} \tanh(t^2) dt = \tanh((x^3)^2) \frac{d(x^3)}{dx} - \tanh((e^x)^2) \frac{d(e^x)}{dx} = \boxed{3x^2 \tanh(x^6) - e^x \tanh(e^{2x})}.$$

Example 7.3.10. 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.

$$\frac{d}{dx}(Si(x)) = \frac{d}{dx} \int_0^x \frac{\sin(t)}{t} dt = \boxed{\frac{\sin(x)}{x}}.$$

Example 7.3.11.

$$\frac{d}{dx} \int_{\sin(x)}^{x^2+3} \sqrt{t} dt = \sqrt{x^2+3} \frac{d(x^2+3)}{dx} - \sqrt{\sin(x)} \frac{d(\sin(x))}{dx} = \boxed{2x\sqrt{x^2+3} - \cos(x)\sqrt{\sin(x)}}.$$

Example 7.3.12. Suppose f is continuous on \mathbb{R} . It follows that f has an antiderivative hence the FTC III applies:

$$\frac{d}{dx} \int_{x^2}^{-x} f(u) du = f(-x) \frac{d(-x)}{dx} - f(x^2) \frac{d(x^2)}{dx} = \boxed{-f(-x) - 2xf(x^2)}.$$

7.4 definite integration

Example 7.4.1.

$$\int_0^1 2^x dx = \frac{1}{\ln(2)} 2^x \Big|_0^1 = \frac{1}{\ln(2)} (2^1 - 2^0) = \boxed{\frac{1}{\ln(2)}}.$$

Example 7.4.2. Let a, b be constants,

$$\int_a^b \sinh(t) dt = \cosh(t) \Big|_a^b = \boxed{\cosh(b) - \cosh(a)}.$$

Example 7.4.3.

$$\int_{-4}^{-2} \frac{dx}{x} = \ln|x| \Big|_{-4}^{-2} = \ln|-2| - \ln|-4| = \ln(2) - \ln(4) = \boxed{\ln(1/2)}.$$

If we had neglected the absolute value function in the antiderivative then we would have obtained an incorrect result. The absolute value bars are important for this integral. Note the answer is negative here because $y = 1/x$ is under the x -axis in the region $-4 \leq x \leq -2$.

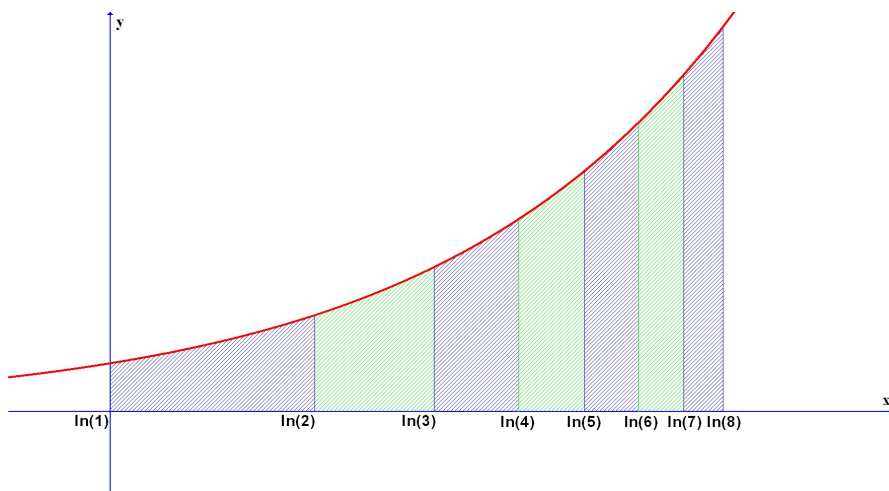
Example 7.4.4.

$$\int_1^9 \frac{dx}{\sqrt{5x}} = \frac{1}{\sqrt{5}} \int_1^9 \frac{dx}{\sqrt{x}} = \frac{2\sqrt{x}}{\sqrt{5}} \Big|_1^9 = \frac{2\sqrt{9}}{\sqrt{5}} - \frac{2\sqrt{1}}{\sqrt{5}} = \boxed{\frac{4}{\sqrt{5}}}.$$

Example 7.4.5. 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.

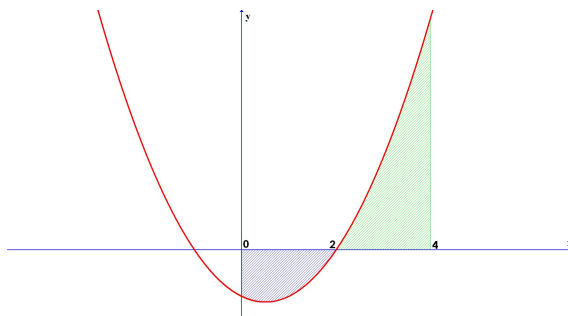


7.4.1 area vs. signed-area

Example 7.4.6. Calculate the signed-area bounded by $y = 3x^2 - 3x - 6$ for $0 \leq x \leq 2$.

$$\int_0^2 (3x^2 - 3x - 6)dx = \left(x^3 - \frac{3}{2}x^2 - 6x\right)\Big|_0^2 = 8 - \frac{3}{2}(4) - 12 = 8 - 18 = \boxed{-10}.$$

Here's an illustration of the calculation (the blue part):



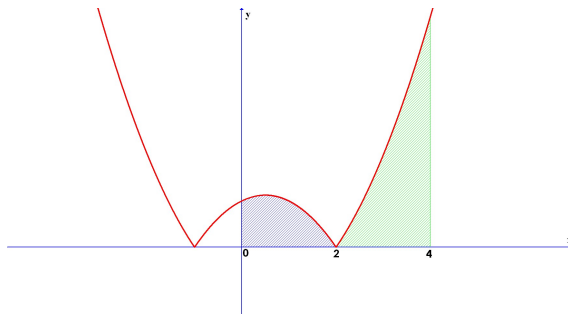
The green area is calculated by

$$\int_2^4 (3x^2 - 3x - 6)dx = \left(x^3 - \frac{3}{2}x^2 - 6x\right)\Big|_2^4 = (64 - \frac{3}{2}(16) - 24) + 10 = 64 - 48 + 10 = \boxed{26}.$$

Example 7.4.7. 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 7.4.8. *absolute values and areas.*

To calculate the area bounded by $y = f(x)$ for $a \leq x \leq b$ we may calculate

$$\text{Area} = \int_a^b |f(x)| dx.$$

Example 7.4.9. Calculate the area bounded by $y = \cos(x)$ on $0 \leq x \leq \frac{5\pi}{2}$.

$$\begin{aligned} \int_0^{3\pi} |\cos(x)| dx &= \int_0^{\frac{\pi}{2}} \cos(x) dx - \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos(x) dx + \int_{\frac{3\pi}{2}}^{\frac{5\pi}{2}} \cos(x) dx \\ &= \sin(x) \Big|_0^{\frac{\pi}{2}} - \sin(x) \Big|_{\frac{\pi}{2}}^{\frac{3\pi}{2}} + \sin(x) \Big|_{\frac{3\pi}{2}}^{\frac{5\pi}{2}} \\ &= (\sin(\frac{\pi}{2}) - \sin(0)) - (\sin(\frac{3\pi}{2}) - \sin(\frac{\pi}{2})) + (\sin(\frac{5\pi}{2}) - \sin(\frac{3\pi}{2})) \\ &= \boxed{5}. \end{aligned}$$

7.4.2 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 7.4.10. *average of a function over a closed interval.*

The average value of f on $[a, b]$ is defined by

$$f_{avg} = \frac{1}{b-a} \int_a^b f(x) dx.$$

Example 7.4.11. Suppose $f(x) = 4x^3$. Find the average of f on $[0, 2]$.

$$f_{avg} = \frac{1}{2} \int_0^2 4x^3 dx = \frac{1}{2} x^4 \Big|_0^2 = \boxed{8}.$$

Example 7.4.12. 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 7.4.13. 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 $[t_1, t_2]$,

$$\begin{aligned} v_{avg} &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (v_o - gt) dt \\ &= \frac{1}{t_2 - t_1} \left[v_o t - \frac{g}{2} t^2 \right]_{t_1}^{t_2} \\ &= \frac{1}{t_2 - t_1} \left[v_o(t_2 - t_1) - \frac{g}{2}(t_2^2 - t_1^2) \right] \\ &= \frac{1}{t_2 - t_1} \left[v_o t_2 - \frac{g}{2} t_2^2 - v_o t_1 + \frac{g}{2} t_1^2 \right] \\ &= \frac{y(t_2) - y(t_1)}{t_2 - t_1} \end{aligned}$$

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 topic.

7.4.3 net-change theorem

Combining FTC I and FTC II we find a very useful result: the net-change theorem.

Theorem 7.4.14. *net change theorem.*

$$\int_a^b \frac{df}{dt} dt = f(b) - f(a).$$

Example 7.4.15. Let $v(t)$ be the instantaneous velocity where $v(t) = \frac{dy}{dt}$ then we can calculate the average velocity over some time interval $[t_1, t_2]$,

$$\begin{aligned} v_{avg} &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} v(t) dt \\ &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{dy}{dt} dt \\ &= \frac{1}{t_2 - t_1} (y(t_2) - y(t_1)) \\ &= \frac{y(t_2) - y(t_1)}{t_2 - t_1}. \end{aligned}$$

Notice we didn't even need to know the details of the velocity function.

We'll see more examples of like this one when we discuss infinitesimal methods in the next Chapter. Your text also has many more examples.

7.5 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.

7.5.1 u-substitution in indefinite integrals

Example 7.5.1.

$$\begin{aligned} \int x e^{x^2} dx &= \int x e^u \frac{du}{2x} \\ &= \frac{1}{2} \int e^u du \\ &= \frac{1}{2} e^u + c \\ &= \boxed{\frac{1}{2} e^{x^2} + c} \end{aligned}$$

$let u = x^2, \frac{du}{dx} = 2x \text{ and } dx = \frac{du}{2x}$

see how all the x's cancelled, this has to happen.

not done yet.

differentiate to check if in doubt.

Example 7.5.2. Let a, b be constants. If $a \neq 0$ then,

$$\begin{aligned} \int (ax + b)^{13} dx &= \int u^{13} \frac{du}{a} \\ &= \frac{1}{14a} u^{14} + c \\ &= \boxed{\frac{1}{14a} (ax + b)^{14} + c.} \end{aligned}$$

$let u = ax + b, \frac{du}{dx} = a \text{ and } dx = \frac{du}{a}$

If $a = 0$ then clearly $\int (ax + b)^{13} dx = \int b^{13} dx = \boxed{b^{13}x + c.}$

Example 7.5.3.

$$\begin{aligned}
 \int 5^{\frac{x}{3}} dx &= \int 5^u (3du) \\
 &= \frac{3}{\ln(5)} 5^u + c \\
 &= \boxed{\frac{3}{\ln(5)} 5^{\frac{x}{3}} + c.}
 \end{aligned}$$

$$\text{let } u = \frac{x}{3}, \frac{du}{dx} = \frac{1}{3} \text{ and } dx = 3du$$

Example 7.5.4.

$$\begin{aligned}
 \int \tan(x) dx &= \int \frac{\sin(x)}{\cos(x)} dx \\
 &= \int \frac{-du}{u} \\
 &= -\ln(|u|) + c \\
 &= \boxed{-\ln(|\cos(x)|) + c.}
 \end{aligned}$$

$$\text{let } u = \cos(x), \frac{du}{dx} = -\sin(x) \text{ and } \sin(x)dx = -du$$

Notice that $-\ln|\cos(x)| = \ln|\cos(x)|^{-1} = \ln|\sec(x)|$ hence $\int \tan(x)dx = \ln|\sec(x)| + c$.

Example 7.5.5.

$$\begin{aligned}
 \int \frac{2x}{1+x^2} dx &= \int \frac{du}{u} \\
 &= \ln(|u|) + c \\
 &= \boxed{\ln(1+x^2) + c.}
 \end{aligned}$$

$$\text{let } u = 1+x^2, \frac{du}{dx} = 2x \text{ and } 2xdx = du$$

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.

Example 7.5.6.

$$\begin{aligned}
 \int \sqrt[3]{1-3x} dx &= \int \sqrt[3]{u} \frac{du}{-3} \\
 &= \frac{3}{4} u^{\frac{4}{3}} + c \\
 &= \boxed{\frac{3}{4} (1-3x)^{\frac{4}{3}}.}
 \end{aligned}$$

$$\text{let } u = 1-3x, \frac{du}{dx} = -3 \text{ and } dx = \frac{du}{-3}$$

Example 7.5.7.

$$\begin{aligned}
 \int \frac{dx}{x+b} &= \int \frac{du}{u} \\
 &= \ln|u| + c \\
 &= \boxed{\ln|x+b| + c.}
 \end{aligned}$$

$$\text{let } u = x+b \text{ thus } du = dx$$

Example 7.5.8. suppose $x > 0$.

$$\begin{aligned}\int \frac{x^2 dx}{\sqrt{x^2 - x^4}} &= \int \frac{x^2 dx}{x\sqrt{1 - x^2}} \\ &= \int \frac{x dx}{\sqrt{1 - x^2}} \\ &= \int \frac{-du}{2\sqrt{u}} \\ &= \frac{-1}{2} 2\sqrt{u} + c \\ &= \boxed{-\sqrt{1 - x^2} + c.}\end{aligned}$$

$$\boxed{\text{let } u = 1 - x^2 \text{ thus } -du/2 = x dx}$$

Notice, we start to see examples where educated guessing alone probably wouldn't have solved it. Of course there are numerous software programs to assist with integration these days but unless you do a bunch of these at some point in your life you'll never really understand what the computer is doing.

Example 7.5.9. suppose $x > 0$.

$$\begin{aligned}\int \frac{\ln(x) dx}{x} &= \int u du \\ &= \frac{1}{2} u^2 + c \\ &= \boxed{\frac{1}{2} \ln(x)^2 + c\sqrt{1 - x^2} + c.}\end{aligned}$$

$$\boxed{\text{let } u = \ln(x) \text{ thus } du = dx/x}$$

Example 7.5.10.

$$\begin{aligned}\int \sin(3\theta) d\theta &= \int \sin(u) \frac{du}{3} \\ &= \frac{-1}{3} \cos(u) + c \\ &= \boxed{\frac{-1}{3} \cos(3\theta) + c.}\end{aligned}$$

$$\boxed{\text{let } u = 3\theta \text{ thus } d\theta = \frac{du}{3}}$$

Example 7.5.11.

$$\begin{aligned}\int \frac{\sin^{-1}(z)}{\sqrt{1 - z^2}} dz &= \int u du \\ &= \frac{1}{2} u^2 + c \\ &= \boxed{\frac{1}{2} [\sin^{-1}(z)]^2 + c.}\end{aligned}$$

$$\boxed{\text{let } u = \sin^{-1}(z) \text{ thus } du = \frac{dz}{\sqrt{1 - z^2}}}$$

Example 7.5.12.

$$\begin{aligned}
 \int t \cos(t^2 + \pi) dt &= \frac{1}{2} \int \cos(u) du \\
 &= \frac{1}{2} \sin(u) + c \\
 &= \boxed{\frac{1}{2} \sin(t^2 + \pi) + c.}
 \end{aligned}$$

$$\text{let } u = t^2 + \pi \text{ thus } t dt = \frac{du}{2}$$

If you understand the example below then you will be able to integrate any odd power of sine or cosine.

Example 7.5.13.

$$\begin{aligned}
 \int \sin^3(x) dx &= \int \sin^2(x) \sin(x) dx \\
 &= \int (1 - \cos^2(x)) \sin(x) dx \\
 &= \int (1 - u^2)(-du) \\
 &= \int (u^2 - 1) du \\
 &= \frac{1}{3} u^3 - u + c \\
 &= \boxed{\frac{1}{3} \cos^3(x) - \cos(x) + c.}
 \end{aligned}$$

$$\text{let } u = \cos(x) \text{ thus } du = -\sin(x) dx$$

Example 7.5.14. suppose $a \neq 0$

$$\begin{aligned}
 \int \frac{dx}{x^2 + a^2} &= \frac{1}{a^2} \int \frac{dx}{\frac{x^2}{a^2} + 1} \\
 &= \frac{1}{a^2} \int \frac{adu}{u^2 + 1} \\
 &= \frac{1}{a} \tan^{-1}(u) + c \\
 &= \boxed{\frac{1}{a} \tan^{-1}\left[\frac{x}{a}\right] + c.}
 \end{aligned}$$

$$\text{let } u = \frac{x}{a} \text{ thus } adu = dx$$

Example 7.5.15. suppose $a \neq 0$

$$\begin{aligned}
 \int \cos(ae^x + 3)e^x dx &= \frac{1}{a} \int \cos(u) du \\
 &= \frac{1}{a} \sin(u) + c \\
 &= \boxed{\frac{1}{a} \sin(ae^x + 3) + c.}
 \end{aligned}$$

$$\text{let } u = ae^x + 3 \text{ thus } du/a = e^x dx$$

Example 7.5.16.

$$\begin{aligned}
\int \sin^2(\theta) d\theta &= \int \frac{1}{2} \left[1 - \sin^2(\theta) \right] d\theta && \text{by trigonometry.} \\
&= \frac{1}{2} \int d\theta - \frac{1}{2} \int \cos(2\theta) d\theta \\
&= \boxed{\frac{\theta}{2} - \frac{1}{4} \sin(2\theta) + c.}
\end{aligned}$$

In the preceding example I omitted a u -substitution because it was fairly obvious. In the next example I demonstrate a notation which allows you to perform u -substitution without even stating explicitly the u . You will not find this notation in many American textbooks.

Example 7.5.17.

$$\begin{aligned}
\int 4 \sinh^2(x) dx &= 4 \int \left[\frac{1}{2} (e^x - e^{-x}) \right]^2 dx && \text{by definition of } \sinh(x). \\
&= \int \left[(e^x)^2 - 2e^x e^{-x} + (e^{-x})^2 \right] dx \\
&= \int \left[e^{2x} - 2 + e^{-2x} \right] dx \\
&= \int e^{2x} dx - 2 \int dx + \int e^{-2x} dx \\
&= \frac{1}{2} \int e^{2x} d(2x) - 2x - \frac{1}{2} \int e^{-2x} d(-2x) \\
&= \frac{1}{2} e^{2x} - 2x - \frac{1}{2} e^{-2x} + c. \\
&= \boxed{\sinh(2x) - 2x + c.}
\end{aligned}$$

Interesting, if you trust my calculation then we may deduce

$$4 \sinh^2(x) = \frac{d}{dx} [\sinh(2x) - 2x] = 2 \cosh(2x) - 2$$

$$\text{thus } \sinh^2(x) = \frac{1}{2} [\cosh(2x) - 1].$$

7.5.2 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 7.5.18 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 7.5.19 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 7.5.18. 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:

$$\begin{aligned} \int_0^{\sqrt{\frac{\pi}{2}}} t \cos(t^2 + \pi) dt &= \left. \frac{1}{2} \sin(t^2 + \pi) \right|_0^{\sqrt{\frac{\pi}{2}}} \\ &= \frac{1}{2} \sin\left(\frac{\pi}{2} + \pi\right) - \frac{1}{2} \sin(\pi) \\ &= \boxed{\frac{-1}{2}}. \end{aligned}$$

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 7.5.19.

$$\begin{aligned} \int_0^{\sqrt{\frac{\pi}{2}}} t \cos(t^2 + \pi) dt &= \frac{1}{2} \int_{\pi}^{\frac{3\pi}{2}} \cos(u) du \\ &= \left. \frac{1}{2} \sin(u) \right|_{\pi}^{\frac{3\pi}{2}} \\ &= \frac{1}{2} \sin\left(\frac{3\pi}{2}\right) - \frac{1}{2} \sin(\pi) \\ &= \boxed{\frac{-1}{2}}. \end{aligned}$$

let $u = t^2 + \pi$ thus $tdt = \frac{du}{2}$

also $u\left(\frac{\pi}{2}\right) = \frac{3\pi}{2}$ and $u(0) = \pi$

Example 7.5.20.

$$\begin{aligned} \int_{4\pi^2}^{9\pi^2} \frac{\sin(\sqrt{x}) dx}{\sqrt{x}} &= \int_{2\pi}^{3\pi} \sin(u) (2du) \\ &= \left. -2 \cos(u) \right|_{2\pi}^{3\pi} \\ &= -2 \cos(3\pi) + 2 \cos(2\pi) \\ &= \boxed{4}. \end{aligned}$$

let $u = \sqrt{x}$ thus $2du = \frac{dx}{\sqrt{x}}$

also $u(9\pi^2) = \sqrt{9\pi^2} = 3\pi$ and $u(4\pi^2) = \sqrt{4\pi^2} = 2\pi$

7.5.3 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 7.5.21. *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:

$$G(f(x)) = \left[\int g(u) du \right] \Big|_{u=f(x)} = \int g(f(x)) \frac{df}{dx} dx = H(x) + c.$$

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 7.5.22. *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:

$$\begin{aligned} \int_0^{2\pi} e^{\sin(x)} \cos(x) dx &= \int_0^0 e^u \cos(x) \frac{du}{\cos(x)} && \boxed{\text{let } u = \sin(x) \text{ thus } dx = \frac{du}{\cos(x)}} \\ &= \int_0^0 e^u du && \boxed{\text{also } u(0) = \sin(0) \text{ and } u(2\pi) = \sin(2\pi).} \\ &= \boxed{0.} \end{aligned}$$

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 $u(b) > u(a)$.

Remark 7.5.23. *geometric meaning of u -substitution.*

The geometric meaning of substitution is an interesting topic that this current version of my notes does not address. You are free to read the text on that topic and it is probable I will spend a few minutes of lecture contemplating the geometry of u -substitution.

Well, that's about it for the mathematics of integration this semester. I do expect you can do many trigonometric integrals which capitalize on your knowledge of trigonometry. For example, $\int \cos^2(x) dx$, $\int \tan^2(x) dx$ and so forth. However, we are necessarily limited this semester. Many integrals I cannot ask simply because we have yet to cover *integration by parts* or *trigonometric substitution* or *partial fractions*. Those techniques are covered in Calculus II. Shortcomings aside, we can solve a great variety of applied problems with tools discovered thus far. This is the focus of the remainder of this course: how can we apply integration to "real world" problems?

Chapter 8

applications of integral calculus

Broadly speaking, the infinitesimal method is the idea of formulating models at an infinitesimal level and then using differentiation and/or integration to decide instantaneous rates of change or total amounts accumulated. For the both the derivative and the integral the same pattern is seen as we transition from the finite model to the continuous model: we replace Δ with d to indicate finite changes becoming infinitesimal changes:

1. we replace the average rate of change $v_{avg} = \frac{\Delta s}{\Delta t}$ with the instantaneous rate of change $v = \frac{ds}{dt}$. As $\Delta t \rightarrow 0$ we have $\Delta t \rightarrow dt$ and correspondingly $\Delta s \rightarrow ds$.

$$v_{avg} = \frac{\Delta s}{\Delta t} \quad \longrightarrow \longrightarrow \longrightarrow \longrightarrow \longrightarrow \longrightarrow \quad v = \frac{ds}{dt}$$

2. as a finite sum of n items has $n \rightarrow \infty$ the pieces of the sum become smaller and smaller, ideally in the limit each element in the sum is "infinitesimal".

$$A_n = \sum_{i=1}^n f(x_i) \Delta x \quad \longrightarrow \longrightarrow \longrightarrow \longrightarrow \longrightarrow \longrightarrow \quad A = \int_a^b f(x) dx$$

the integral is often called a **continuous sum**.

We have established the precise mathematical meaning of $\frac{ds}{dt}$ and $\int_a^b f(x) dx$ in previous chapters. We know that the comments above speak to a conceptual process. A finite sum is not an integral and an average rate of change is not an instantaneous rate of change. Often your first exposure to an application is at the level of averages, for example $v = s/t$ the speed is the distance over the time. But, this is too crude for all but the most boring of physical examples so we insist that the basic quantity to study is $v = ds/dt$. We replace the average quantities with corresponding infinitesimals and this results in a precise model for the system. To make the continuum limit we ask what average laws should hold for infinitesimal quantities? What is the logical way to set up the problem. Much of this chapter is about trying to answer these questions.

To begin we reexamine the net-change theorem in view of the conceptual clarity the infinitesimal method offers. Infinitesimals provide a notation to quickly summarize a limiting process without need of even stating the process. If you need to see the process I invite you to read the text.

In this chapter we return to the problem of calculating the area of some region. In contrast to signed area, this will really be the area which is positive. We will find that infinitesimal arguments provide an efficient shorthand for writing limiting processes. Almost every problem is begun by drawing a graph of the region and a typical approximating rectangle. Then we find the net area by adding up all the infinitesimal areas.

This "adding up" is integration, we should think of integration as a continuous sum.

Once the area problem is settled we turn to the task of calculating the volumes of solids which possess a certain regularity. If a solid is such that the cross-sectional area is of the same type at each value of the axis perpendicular to the cross-section then we can add up the volume of each slice and get the total volume. For example, a sphere has a cross-sections which are disks if we use a diameter as an axis. A tetrahedron has triangular cross-sections relative to the axis which extends perpendicularly from one its faces to an opposing vertex. A solid of revolution has cross-sections which are disks or washers relative to the axis of revolution. Of course, you should look at the pictures in this chapter before you get too worried about the meaning of this paragraph. Integration provides the technology needed to add together all the tiny volumes.

Finally, I issue a word of caution. Infinitesimals are not real numbers. They are merely a notation to indicate a precise mathematical construction which we have spend months to develop with care. Remember this as you use them.

8.1 a brief tour of infinitesimal methods

Concept: average concepts still hold true at the infinitesimal level. Integration then extends these microscopic arguments to macroscopic rules. Often we can find a relation that holds true over an instant dt of time or a little displacement dx etc...

Example 8.1.1. *For example, the displacement for a particle moving with velocity v during a time dt is simply the product of velocity and time; $dy = v dt$. This makes sense because the velocities $v(t)$ and $v(t + dt)$ are equal. To be more precise I should say they are equal in the limit that $dt \rightarrow 0$. During an instant of time the velocity is constant so we can use the constant velocity formula.*

$$\begin{aligned} dy = v dt &\Rightarrow \int_{y_o}^{y(t)} ds = \int_0^t v(\tau) d\tau \\ &\Rightarrow y(t) = y_o + \int_0^t v(\tau) d\tau. \end{aligned}$$

We call $y(t)$ the **position** at time t . The displacement during the time interval $[0, t]$ is the net-change in position; $y(t) - y_o$. If we wanted to know the distance traveled then we need to do a different calculation since neither the displacement nor the position reveal the distance traveled. Imagine a particle oscillating back and forth 20 times. You could have a distance traveled much larger than the displacement. Distance traveled is always positive, let's denote distance by s :

$$\begin{aligned} ds = |v| dt &\Rightarrow \int_{s_o}^{s(t)} ds = \int_0^t |v(\tau)| d\tau \\ &\Rightarrow s(t) = s_o + \int_0^t |v(\tau)| d\tau. \end{aligned}$$

The absolute value signs insure we are calculating distance rather than displacement, the absolute value of the velocity is called the **speed**. Often I set $s_o = 0$. The acceleration $a = \frac{dv}{dt}$ thus $dv = a dt$ and we may

obtain the velocity from integration the acceleration from time zero to time t ,

$$\begin{aligned} dv = a \, dt &\Rightarrow \int_{v_o}^{v(t)} dv = \int_0^t a(\tau) d\tau \\ &\Rightarrow v(t) = v_o + \int_0^t a(\tau) d\tau. \end{aligned}$$

We explore the meaning of the constructions in this example in depth in the physics course. It's much more fun once we have a few more dimensions to explore.

Suppose that $a(t) = \sin(t)$ for some object travelling in the x -direction such that the object is initially moving right at 1m/s and has an initial position $x = 3\text{m}$. We can calculate the position, velocity, speed and distance traveled by integration.

$$\begin{aligned} dv = \sin(t) \, dt &\Rightarrow \int_1^{v(t)} dv = \int_0^t \sin(\tau) d\tau \\ &\Rightarrow v(t) - 1 = -\cos(\tau) \Big|_0^t \\ &\Rightarrow \boxed{v(t) = 2 - \cos(t)}. \end{aligned}$$

To calculate speed we simply take absolute value; speed $= |v(t)| = |2 - \cos(t)|$. Next, integrate the velocity to find position,

$$\begin{aligned} dx = v \, dt &\Rightarrow \int_3^{x(t)} dx = \int_0^t (2 - \cos(\tau)) d\tau \\ &\Rightarrow x(t) - 3 = \left(2\tau - \sin(\tau) \right) \Big|_0^t \\ &\Rightarrow \boxed{x(t) = 3 + 2t - \sin(t)}. \end{aligned}$$

Distance traveled is curiously equal to the displacement in this example since $|v(t)| = v(t)$ (can you see why I know the velocity is positive for this example?). If I had given an initial velocity of $v_o = 0.5\text{m/s}$ then the speed would differ from the velocity. In invite the reader to rework this example for the case $x_o = 3$ and $v_o = 0.5$.

Remark 8.1.2.

Let me comment on the mathematics used in the preceding example. The net-change theorem states:

$$\int_a^b \frac{df}{dx} dx = f(b) - f(a).$$

The technique of u -substitution tell us the same result: just substitute $u = f$ to see,

$$\int_a^b \frac{df}{dx} dx = \int_{f(a)}^{f(b)} df = f(b) - f(a).$$

Generically, if we have $\frac{df}{dx} = g$ then we can solve to obtain $df = g(x)dx$. Then you can integrate this, but to be fair we have to match the bounds to the variables:

$$df = g(x)dx \quad \Rightarrow \quad \int_{f(a)}^{f(b)} df = \int_a^b g(x) dx$$

the bounds on the LHS and RHS should correspond in the way that the change of variables theorem demands. These observations give us a new way to solve some of the problems we previously solved by antidifferentiation followed by algebra to fit the "c". Integration from time zero to time t naturally encodes initial conditions into the solution for time t . The downside is that to be careful we have to take care not to confuse the time t with the dummy variable of integration τ .

Example 8.1.3. Another example is current $I = dQ/dt$. If we wish to calculate the net charge that has flowed from time zero to time t then we simply integrate the current,

$$\begin{aligned} I = \frac{dQ}{dt} &\Rightarrow dQ = I dt \\ &\Rightarrow \int_{Q_o}^{Q(t)} dQ = \int_0^t I(\tau) d\tau \\ &\Rightarrow Q(t) = Q_o + \int_0^t I(\tau) d\tau. \end{aligned}$$

Example 8.1.4. This idea also applies to things which are not from some time-rate of change. For example, the work done by a constant force F over a distance x is given by the formula $W = Fx$. Now this formula is only for constant forces which act in the direction of the displacement. What would we do if the force was a function of position? Then we could not just use the formula since the force is not constant. However, if we look at $F(x)$ and $F(x+dx)$ then those forces are equal in the limit that $dx \rightarrow 0$. So we can conclude that the simple work equation holds at the infinitesimal level; the work dW done by a force $F(x)$ over a displacement dx will be $dW = F(x) dx$. If the force does work from $x = a$ to $x = b$ then the net work done will be the sum of all the infinitesimal works dW , in other words,

$$W = \int dW = \int_a^b F(x) dx.$$

For example, if we have the spring force then $F(x) = kx$ for some constant k called the "stiffness",

$$W = \int dW = \int_a^b kx dx = \frac{1}{2}kb^2 - \frac{1}{2}ka^2.$$

Example 8.1.5. Another application is hydrostatic force. The force on a dam is due to water pressure. The definition of pressure is that it is force per unit area, this gives us

$$P = \frac{F}{A}$$

Now this only makes sense so long as the same force is applied over the whole area. We cannot just apply this equation to the force due to the water pressure on a dam. The pressure at the bottom of the dam is larger than that at the top. In fact it is known that $P = \rho gy$, it is proportional to the depth y . Different depths give different pressures, hopefully this is a familiar fact to everyone. So, if we wish to calculate the net-force due to pressure (this is called the "hydrostatic force") then we should consider horizontal strips of area $dA = l(y)dy$ where $l(y)$ is the width of the dam at depth y . These will have the same pressure all along them so the equation makes sense to apply to the strip, we have

$$P(y) = \rho gy = \frac{dF}{dA}$$

The little force dF is due to the pressure $P(y)$ acting on dA . Then

$$dF = \rho gy dA = \rho gyl(y)dy \quad \Rightarrow \quad F = \int_{y_1}^{y_2} \rho gyl(y)dy$$

Generally the challenging part of these dam examples is finding the actual function for l as function of y . The simplest example is a rectangular dam of constant width L and depth h

$$dF = \rho gy dA = \rho gYLdy \quad \Rightarrow \quad F = \rho gL \int_0^h y dy = \frac{1}{2} \rho gLh^2 = P_{\text{middle}}Lh.$$

Where $P_{\text{middle}} = \rho g \frac{h}{2}$. Effectively, it is as if we had the total area of the dam $A = Lh$ subject to the pressure at the midpoint of the dam. For a triangular or semicircular dam the effective pressure might occur lower or higher on the dam.

Example 8.1.6. What is the **arclength** of a curve $y = f(x)$ for $a \leq x \leq b$? We cannot just calculate the distance between $(a, f(a))$ and $(b, f(b))$ since the curve is probably not a line. However, if we calculate the little distance ds between $(x, f(x))$ and $(x + dx, f(x + dx))$ then it stands to reason we can approximate the function with its tangent line so $dy = f'(x)dx$ and the distance formula yields:

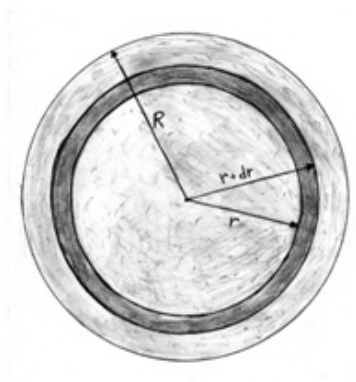
$$ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + \frac{dy^2}{dx^2}} dx$$

Consider then $y = \sqrt{R^2 - x^2}$ for $0 \leq x \leq R$. Calculate $\frac{dy}{dx} = \frac{-x}{\sqrt{R^2 - x^2}}$ hence the arclength of this curve is found by integrating:

$$s = \int_0^R \sqrt{1 + \frac{x^2}{R^2 - x^2}} dx = \int_0^R \frac{R}{\sqrt{R^2 - x^2}} dx = \int_0^R \frac{R}{\sqrt{1 - (\frac{x}{R})^2}} d(\frac{x}{R}) = R \sin^{-1}(\frac{x}{R}) \Big|_0^R = \frac{\pi R}{2}.$$

Notice that the curve we just considered has equation $y^2 = R^2 - x^2$ or $x^2 + y^2 = R^2$. In fact, it was precisely the quarter of the circle of radius R which is in quadrant I. To find the total arclength of the circle we simply multiply by 4; $S = 2\pi R$. With calculus we can **derive** the formula for the circumference of a circle.

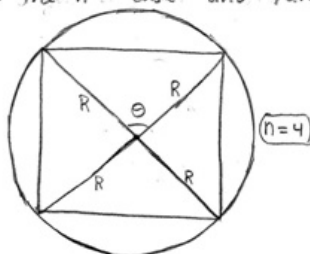
Example 8.1.7. What is the **area** of a circle of radius R ? Notice we can cut a circle into a bunch of circular strips. If the strip from radius r to $r + dr$ has area dA then we can calculate that area from imagining taking the strip and laying it straight to make a rectangular strip of width dr and length $2\pi r$ (using the last example). Hence, $dA = 2\pi r dr$. We have to add up the area of each strip all the way from $r = 0$ up to $r = R$.



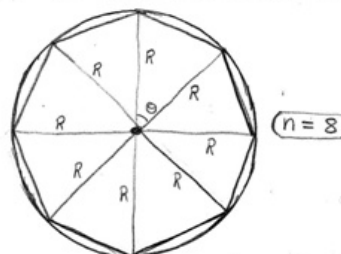
$$A = \int dA = \int_0^R 2\pi r dr = \pi r^2 \Big|_0^R = \boxed{\pi R^2}.$$

We can calculate the area of a circle without integral calculus. We just need trigonometry and a few basic limit theorems:

n-sided polygon inside a circle of radius R . I illustrate the case $n=4$ (square) and $n=8$ (octagon) we then generalize to the n^{th} case and take $n \rightarrow \infty$ to find the area of circle,

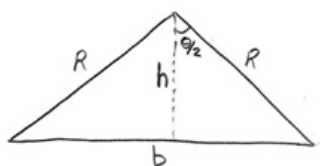


$$\Theta = \frac{2\pi}{n} = \frac{\pi}{2}$$



$$\Theta = \frac{2\pi}{n} = \frac{\pi}{4}$$

Lets use trigonometry to calculate the area of a typical subtriangle. Need to find b and h using R and Θ .



$$b = 2R \sin(\Theta/2)$$

$$h = R \cos(\Theta/2)$$

$$A = \frac{1}{2}bh = R^2 \sin(\Theta/2) \cos(\Theta/2)$$

Now you should know $\sin(2\beta) = 2\sin\beta \cos\beta$ thus

$$A = \frac{1}{2}R^2 \sin(\Theta)$$

Lets write the area of a regular n -gon inscribed in the circle of radius R

$$A_n = n \cdot \frac{1}{2}R^2 \sin(\Theta) = \boxed{\frac{nR^2}{2} \sin\left(\frac{2\pi}{n}\right) = A_n}$$

For finite n we approximate the area of circle, but as $n \rightarrow \infty$ this becomes an exact account of the circle's area,

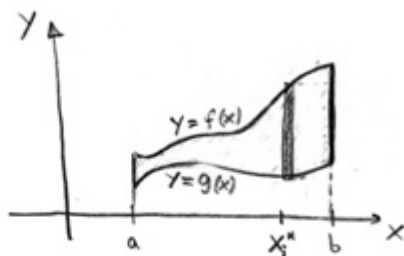
$$\begin{aligned} A_{\infty} &= \lim_{n \rightarrow \infty} \left(\frac{nR^2}{2} \sin\left(\frac{2\pi}{n}\right) \right) \\ &= \pi R^2 \lim_{n \rightarrow \infty} \left(\frac{\sin\left(\frac{2\pi}{n}\right)}{\frac{2\pi}{n}} \right) \\ &= \pi R^2 \lim_{t \rightarrow 0^+} \left(\frac{\sin(t)}{t} \right) = \boxed{\pi R^2} \end{aligned}$$

I hope this selection of examples is enough for you to begin to appreciate the beauty of the infinitesimal formalism. I wanted you to understand that the remainder of the chapter is useful beyond the problems of area and volume. Similar thinking is needed to set-up a host of interesting applied problems.

8.2 area

Example 8.2.1.

Consider a generic example to begin, find the area bounded by $x=a$, $x=b$ and $f(x)$ and $g(x)$ where $f(x) > g(x)$ for all x on $[a, b]$. A picture helps



we can find the area by dividing the area up into as many infinitesimal rectangles. We've drawn a typical box you can see its height is $f(x) - g(x)$, it has area $\Delta A_i = [f(x_i) - g(x_i)] \Delta x$.

$$A = \text{AREA} = \lim_{n \rightarrow \infty} \sum_{i=1}^n (f(x_i^*) - g(x_i^*)) \Delta x$$

But this is precisely the integral of $f(x) - g(x)$,

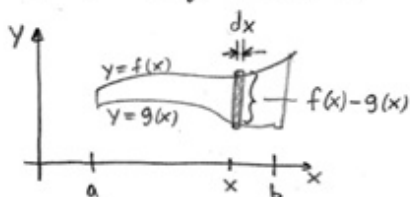
$$A = \int_a^b (f(x) - g(x)) dx$$

- Technically, we should begin all problems with argument about Δx and ΔA_i and so on, these are finite. Then once the object of interest is suitably approximated by n -rectangles we pass to the limit $n \rightarrow \infty$ and find the \sum becomes an \int and Δx becomes a dx . For all calculational purposes the first steps with " Δ "s are unnecessary. We can make these arguments using infinitesimals from the beginning and I will from now on.

The general strategy is to draw a picture to get a handle on the problem, then find the formula for the area dA of a typical infinitesimal rectangle and then add all the little areas together by integrating.

Example 8.2.2. .

Revisited using infinitesimal arguments.



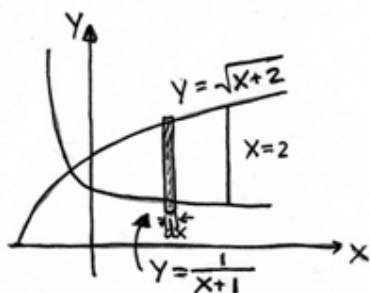
$$dA = (f(x) - g(x)) dx$$

typical area of infinitesimal rectangle.

To obtain area simply add up all the infinitesimal areas,

$$A = \int dA = \int_a^b (f(x) - g(x)) dx$$

Example 8.2.3. .



$$dA = (y_t - y_b) dx$$

$$= \left(\sqrt{x+2} - \frac{1}{x+1} \right) dx$$

There is a vertical slice at each x from 0 to 2. Thus

$$A = \int_0^2 \left(\sqrt{x+2} - \frac{1}{x+1} \right) dx = \left[\frac{2}{3} (x+2)^{3/2} - \ln|x+1| \right]_0^2$$

$$= \left(\frac{2}{3} (4)^{3/2} - \ln(3) \right) - \left(\frac{2}{3} (2)^{3/2} - \ln(1) \right)$$

$$= \boxed{2.349}$$

Example 8.2.4.

find area bounded by $Y = x^2$ and $Y = \sqrt{x}$

Lets find where these curves intersect, set $y = y$ yielding,

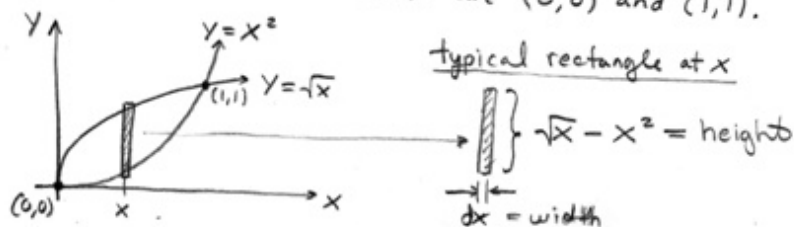
$$x^2 = \sqrt{x}$$

$$x^4 = x$$

$$x^4 - x = 0$$

$$x(x^3 - 1) = 0 \Rightarrow \underline{x = 0 \text{ or } x = 1}$$

Thus the points of intersection are $(0,0)$ and $(1,1)$.



The area of this tiny rectangle is, $dA = (\sqrt{x} - x^2)dx$

$$\text{Thus } A = \int_0^1 (\sqrt{x} - x^2) dx$$

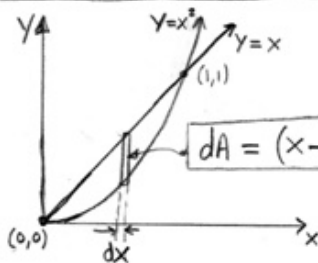
$$= \left(\frac{2}{3} x^{3/2} - \frac{1}{3} x^3 \right) \Big|_0^1$$

$$= \frac{2}{3} - \frac{1}{3}$$

$$= \boxed{\frac{1}{3}}$$

Example 8.2.5.

Area bounded by $Y = x$ and $Y = x^2$ is what?



• Need to find intersection points
here we can simply equate Y ,

$$x = x^2$$

$$x^2 - x = x(x-1) = 0$$

$$x = 0 \text{ or } x = 1$$

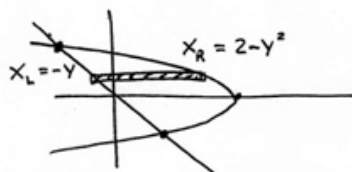
The points of intersection
are $(0,0)$ and $(1,1)$.

$$\begin{aligned} A &= \int dA = \int_0^1 (x - x^2) dx \\ &= \left(\frac{1}{2} x^2 - \frac{1}{3} x^3 \right) \Big|_0^1 \\ &= \frac{1}{2} - \frac{1}{3} \\ &= \boxed{\frac{1}{6}} \end{aligned}$$

Example 8.2.6.

$$\begin{aligned} x &= 2 - y^2 \\ x &= -y \end{aligned}$$

Just switch the
roles of x & y to see
why the graph should be
a line and sideways parabola.



Points of intersection follow from $x_L = x_R$,

$$2 - y^2 = -y$$

$$y^2 - y - 2 = 0$$

$$(y-2)(y+1) = 0 \quad \therefore y = 2 \text{ (or)} y = -1$$

$$dA = (x_R - x_L) dy$$

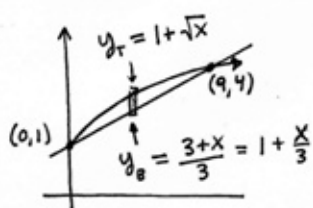
$$= (2 - y^2 - (-y)) dy$$

$$= (2 + y - y^2) dy$$

$$\therefore A = \int_{-1}^2 (2 + y - y^2) dy = \left[2y + \frac{1}{2} y^2 - \frac{1}{3} y^3 \right]_{-1}^2$$

Evaluation yields $\boxed{A = 4.5}$

Example 8.2.7. .



$$\begin{aligned} dA &= (y_T - y_B) dx \\ &= \left(1 + \sqrt{x} - 1 - \frac{x}{3}\right) dx \\ &= \left(\sqrt{x} - \frac{1}{3}x\right) dx \end{aligned}$$

Therefore,
using algebra
below to
explain bounds,

$$\begin{aligned} A &= \int_0^9 \left(\sqrt{x} - \frac{1}{3}x\right) dx \\ &= \left[\frac{2}{3}x^{3/2} - \frac{1}{6}x^2\right]_0^9 \\ &= \left[\frac{2}{3}(9)^{3/2} - \frac{9^2}{6}\right] \\ &= \boxed{4.5} \end{aligned}$$

Points of intersection followed from $y_B = y_T$ because,

$$1 + \sqrt{x} = 1 + \frac{x}{3}$$

$$\sqrt{x} = \frac{x}{3}$$

$$3\sqrt{x} = x$$

$$9x = x^2$$

$$x^2 - 9x = 0$$

$$x(x-9) = 0$$

$$\therefore \underline{x=0 \text{ or } x=9}$$

← cannot just divide by x
you lose information by doing
that. (x might be zero)

Example 8.2.8.

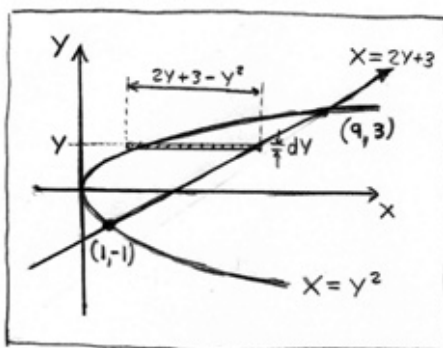
Find area bounded by $Y^2 = X$ and $2Y = X - 3$

Notice that at intersection $X = X \Rightarrow Y^2 = 2Y + 3$
 $\Rightarrow Y^2 - 2Y - 3 = 0$

$$\Rightarrow (Y-3)(Y+1) = 0$$

$$\Rightarrow Y = 3 \text{ or } Y = -1$$

\Rightarrow points of intersection are $(9, 3)$ and $(1, -1)$

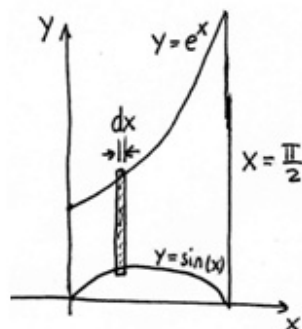


• its better to use horizontal strips because we don't need to break up into cases. If we used vertical strips it would be tricky because $0 \leq X \leq 1$ is different than $1 \leq X \leq 9$ for vertical strips

$$dA = (2Y + 3 - Y^2) dY$$

$$\begin{aligned} \text{Thus } A &= \int_{-1}^3 (2Y + 3 - Y^2) dY \\ &= \left(Y^2 + 3Y - \frac{1}{3} Y^3 \right) \Big|_{-1}^3 \\ &= \left(9 + 9 - \frac{1}{3}(27) \right) - \left(1 - 3 + \frac{1}{3} \right) \\ &= 9 + \frac{5}{3} = \boxed{\frac{32}{3} \approx 10.67} \end{aligned}$$

Example 8.2.9.



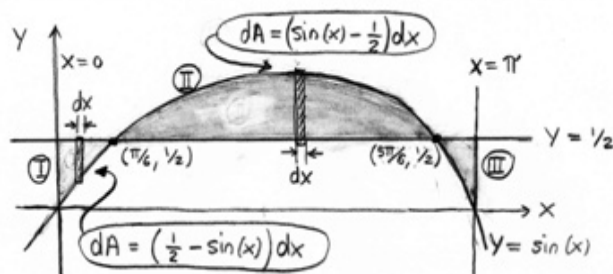
$$\begin{aligned} dA &= (y_t - y_b) dx \\ &= (e^x - \sin(x)) dx \end{aligned}$$

Therefore,

$$\begin{aligned} A &= \int_0^{\pi/2} (e^x - \sin(x)) dx \\ &= \left[e^x + \cos(x) \right]_0^{\pi/2} = \boxed{e^{\pi/2} - 2 = 2.81} \end{aligned}$$

Example 8.2.10. .

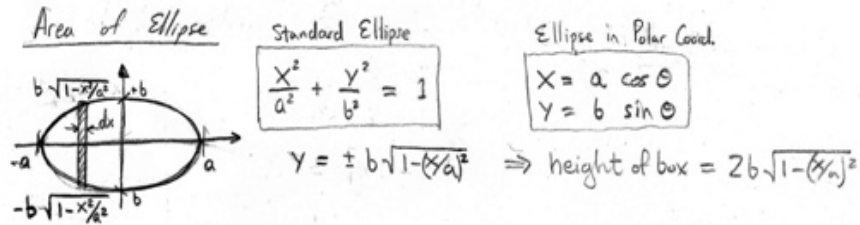
Find area bounded by $Y = \sin(x)$ and $Y = \frac{1}{2}$ and $X = 0$ and $X = \pi$



Notice that $\sin(30^\circ) = \sin(\pi/6) = 1/2$ and by the symmetry of $\sin(x)$ about $x = \pi/2$ it's clear $\sin(5\pi/6) = 1/2$. This reveals the intersection points are $(\pi/6, 1/2)$ and $(5\pi/6, 1/2)$. Clearly we need to divide up into cases I, II and III,

$$\begin{aligned}
 A &= \int_0^{\pi/6} \left(\frac{1}{2} - \sin(x)\right) dx + \int_{\pi/6}^{5\pi/6} \left(\sin(x) - \frac{1}{2}\right) dx + \int_{5\pi/6}^{\pi} \left(\frac{1}{2} - \sin(x)\right) dx \\
 &= \left[\frac{\pi}{12} + \cos(x)\right]_0^{\pi/6} - \left[\frac{4\pi}{12} - \cos(x)\right]_{\pi/6}^{5\pi/6} + \left[\frac{\pi}{12} + \cos(x)\right]_{5\pi/6}^{\pi} \\
 &= \frac{\pi}{12} + \frac{\sqrt{3}}{2} - 1 - \frac{4\pi}{12} + \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2} + \frac{\pi}{12} - 1 + \frac{\sqrt{3}}{2} \\
 &= \frac{-2\pi}{12} + 2\sqrt{3} - 2 \\
 &= 2(\sqrt{3} - 1) - \pi/6 \\
 &\approx \boxed{0.9405 = A}
 \end{aligned}$$

Example 8.2.11. .



$$dA = 2b \sqrt{1 - (x/a)^2} dx$$

$$A = \int_{-a}^a 2b \sqrt{1 - (x/a)^2} dx$$

$$= 2b \int_{-\pi/2}^{\pi/2} \sqrt{1 - \sin^2 \theta} a \cos \theta d\theta$$

$$= 2ab \int_{-\pi/2}^{\pi/2} \cos^2 \theta d\theta$$

$$= 2ab \int_{-\pi/2}^{\pi/2} \frac{1}{2} (1 + \cos 2\theta) d\theta$$

$$= ab \left[\theta + \frac{\sin 2\theta}{2} \right]_{-\pi/2}^{\pi/2}$$

$$= ab \left[\left(\pi/2 + \frac{1}{2} \sin(\pi) \right) - \left(-\pi/2 + \frac{1}{2} \sin(-\pi) \right) \right]$$

$$= \boxed{\pi ab}$$

Notice when $a = b = r$ we get $A = \pi r^2$ which is good.

Discussion: What is the eqⁿ of this ellipse in polar coordinates?

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \Rightarrow \frac{r^2 \cos^2 \theta}{a^2} + \frac{r^2 \sin^2 \theta}{b^2} = 1$$

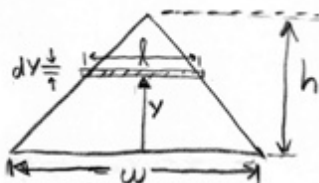
$$r^2 \left(\cos^2 \theta + \frac{a^2}{b^2} \sin^2 \theta \right) = a^2$$

$$\boxed{r = \frac{a}{\sqrt{\cos^2 \theta + \left(\frac{a}{b}\right)^2 \sin^2 \theta}}}$$

We can again notice when $a = b$ we get $r = a$ a very sensible eqⁿ for a circle at the origin.

Example 8.2.12. .

let's find the width of a slice in a triangular region of width w and height h as a function of y .



Again it's clear that for this shape that l varies linearly with y , so $l = my + b$. Additionally we know

$$y = 0 \Rightarrow l = w = m(0) + b \therefore \boxed{b = w}$$

$$y = h \Rightarrow l = 0 = mh + w \therefore \boxed{m = -w/h}$$

Thus we find $\boxed{l = -\frac{w}{h}y + w}$.

We see that $dA = l dy = (-\frac{w}{h}y + w) dy$ thus

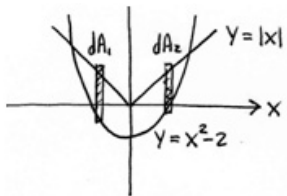
$$A = \int_0^h (-\frac{w}{h}y + w) dy$$

$$= \left(-\frac{w}{2h}y^2 + wy \right) \Big|_0^h$$

$$= -\frac{1}{2}wh + wh$$

$$= \boxed{\frac{1}{2}wh = \frac{1}{2}(\text{base})(\text{height})}$$

Example 8.2.13. .



$$dA_1 = (-x - (x^2 - 2)) dx \quad (x < 0)$$

$$= (2 - x - x^2) dx$$

$$dA_2 = (x - (x^2 - 2)) dx \quad (x > 0)$$

$$= (2 + x - x^2) dx$$

Points of intersection

$x > 0$ $x = x^2 - 2 \Rightarrow x^2 - x - 2 = (x-2)(x+1) = 0$
 $\therefore x = 2$ or $x = -1$ throw that out because $x \neq 0$.

$x < 0$ $-x = x^2 - 2 \Rightarrow x^2 + x - 2 = (x+2)(x-1) = 0$
 $\therefore x = -2$ or $x = 1$ throw out that solⁿ because $x \neq 0$.

$$A = \int_{-2}^0 (2 - x - x^2) dx + \int_0^2 (2 + x - x^2) dx$$

$$= \left(2x - \frac{1}{2}x^2 - \frac{1}{3}x^3 \right) \Big|_{-2}^0 + \left(2x + \frac{1}{2}x^2 - \frac{1}{3}x^3 \right) \Big|_0^2 = \frac{10}{3} + \frac{10}{3} = \boxed{\frac{20}{3}}$$

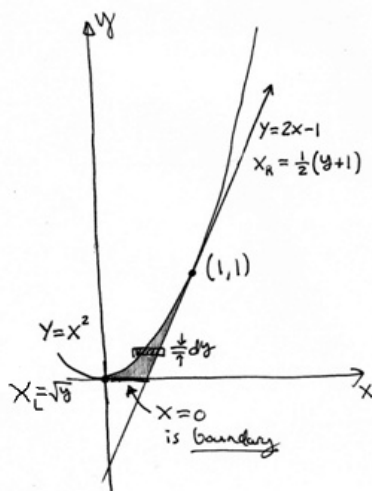
Example 8.2.14. .

Find area bounded by $y = x^2$ and the tangent to $y = f(x) = x^2$ at $(1, 1)$, and the x -axis.
 Recall we can write the tangent to $y = f(x)$ at $(a, f(a))$ in general (provided $f'(a)$ exists)

$$y = f(a) + f'(a)(x - a)$$

Then we have $a = 1$ and $f'(x) = 2x \Rightarrow f'(1) = 2$ then

$$y = 1 + 2(x - 1) = 2x - 1 = y \quad (\text{tangent line})$$



$$dA = (x_R - x_L) dy$$

$$= \left(\frac{1}{2}y + \frac{1}{2} - \sqrt{y} \right) dy$$

Points of intersection:

$$2x - 1 = x^2$$

$$x^2 - 2x + 1 = (x - 1)^2 = 0$$

$$\therefore x = 1 \Rightarrow y = 1 \text{ at intersection}$$

that is the pt. of intersection is $(1, 1)$

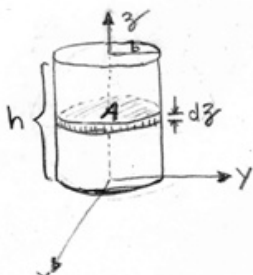
Therefore, noting $0 \leq y \leq 1$ for area,

$$A = \int_0^1 \left(\frac{1}{2}y + \frac{1}{2} - \sqrt{y} \right) dy$$

$$= \left(\frac{1}{4}y^2 + \frac{1}{2}y - \frac{2}{3}y^{3/2} \right) \Big|_0^1 = \frac{1}{4} + \frac{1}{2} - \frac{2}{3} = \boxed{\frac{1}{12}}$$

8.3 volume

Example 8.3.1. .



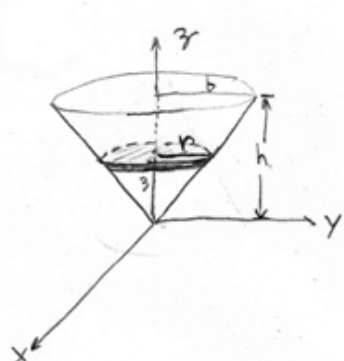
$$A = \pi b^2$$


$$dV = A(z) dz = \pi b^2 dz$$

$$V = \int dV = \int_0^h \pi b^2 dz = \boxed{\pi b^2 h = V}$$

This is the simplest case, the cross-section is constant along the integration axis.

Example 8.3.2. .



$$A = \pi r^2 = \pi \left(\frac{b}{h} z \right)^2 \quad (\text{see arguments below})$$


Notice that r should depend linearly on z thus $r = mz + b$.

$$\left. \begin{array}{l} r(z=0) = 0 \\ r(z=h) = b \end{array} \right\} \begin{array}{l} \text{clear from} \\ \text{picture.} \end{array}$$

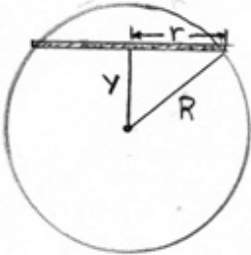
Thus $0 = b$ and $b = mh \therefore m = b/h$.
giving that $\boxed{r = \frac{b}{h} z}$.

The volume of each tiny slice is found to be $dV = \frac{\pi b^2}{h^2} z^2 dz$.
So the total volume is found by adding these up,

$$V = \int_0^h \frac{\pi b^2}{h^2} z^2 dz = \frac{\pi b^2}{3h^2} z^3 \Big|_0^h = \boxed{\frac{\pi b^2 h}{3} = V_{\text{cone}}}$$

Example 8.3.3. .

Find the volume of a sphere of radius R



view of same slice as it intersects (xy)-plane.

$$r^2 + y^2 = R^2$$

$$r^2 = R^2 - y^2$$

$$A = \pi r^2 = \pi(R^2 - y^2)$$

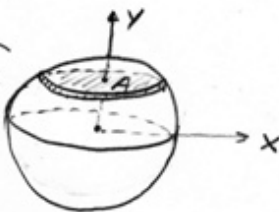
$$dV = A dy = \pi(R^2 - y^2) dy$$

We then find the total volume by summing the volumes of the disks from $y = -R$ up to $y = R$,

$$V = \int_{-R}^R \pi(R^2 - y^2) dy$$

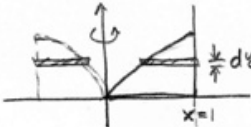
$$= \pi \left(R^2 y - \frac{1}{3} y^3 \right) \Big|_{-R}^R$$

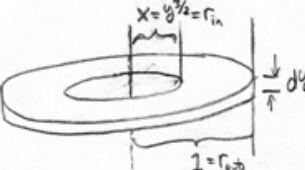
$$= \pi \left[\left(R^3 - \frac{1}{3} R^3 \right) - \left(-R^3 + \frac{1}{3} R^3 \right) \right]$$

$$= \frac{4}{3} \pi R^3 = V_{\text{sphere}}$$


Example 8.3.4. .

Bound region by,
 $y = x^{2/3}$, $x = 1$, $y = 0$
 about the y -axis



$$y = x^{2/3} \Rightarrow x = y^{3/2}$$


Area of annular
 $A = \pi(r_{\text{out}}^2 - r_{\text{in}}^2)$
 $dV = A dy$

$$dV = \pi(1 - (y^{3/2})^2) dy$$

$$= \pi(1 - y^3) dy$$

$$V = \int_0^1 \pi(1 - y^3) dy$$

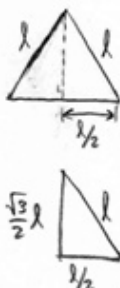
$$= \pi \left(y - \frac{1}{4} y^4 \right) \Big|_0^1$$

$$= \pi \left(1 - \frac{1}{4} \right) = \frac{3\pi}{4} = 2.356$$

Example 8.3.5.

Remark: for many shapes the length of a side (or the radius) varies linearly as we go up the shape. So if we know the length of the side (or radius) for two slices we can find the length of the side (or radius) for any slice. An example to clear up what I mean,

find volume of pyramid of height h with triangular base with side a ,



what's the area of the triangle?
look at the pictures to see

$$A = \frac{1}{2} l \left(\frac{\sqrt{3}}{2} l \right) = \frac{\sqrt{3}}{4} l^2$$

The volume from the slice would be $dV = A dy = \frac{\sqrt{3}}{4} l^2 dy$.

Notice that l is a function of y . More importantly it is a linear function of y .

$$y=0 \Rightarrow l=a \quad (\text{at the base})$$

$$y=h \Rightarrow l=0 \quad (\text{at the top})$$

But we know $l(y) = my + b$. Plug in the data,

$$l(h) = 0 = mh + b \Rightarrow m = -b/h$$

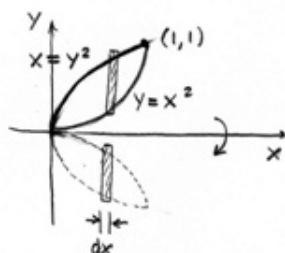
$$l(0) = a = m(0) + b \Rightarrow \underline{a=b} \Rightarrow \underline{m = -a/h}$$

$$\text{Thus } l(y) = \left(-\frac{a}{h}\right)y + a = \frac{a}{h}(h-y)$$

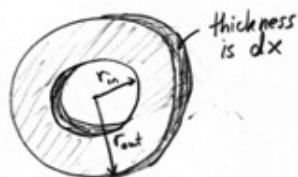
$$\begin{aligned} V &= \int_0^h \frac{\sqrt{3}}{4} l^2 dy = \int_0^h \frac{\sqrt{3}}{4} \frac{a^2}{h^2} (h^2 - 2hy + y^2) dy \\ &= \frac{\sqrt{3} a^2}{4 h^2} \left(h^2 y - h y^2 + \frac{1}{3} y^3 \right) \Big|_0^h = \boxed{\frac{a^2 h}{4\sqrt{3}}} \end{aligned}$$

Example 8.3.6.

Find volume of solid obtained by rotating the region bounded by $y = x^2$ and $y^2 = x$ about the x -axis.



intersection points have $y^2 = y^2$
 meaning $x = (x^2)^2 = x^4$
 $\Rightarrow x^4 - x = x(x^3 - 1) = 0 \Rightarrow \underline{x=0 \text{ or } x=1}$



← this is a typical washer in the volume pictured

$$r_{in} = x^2$$

$$r_{out} = \sqrt{x}$$

the area of the annulus is
 (given by $A = \pi(r_{out}^2 - r_{in}^2) = \pi(x - x^4)$)

thus the volume of a typical washer is,

$$dV = \pi(x - x^4) dx$$

which allows us to find the total volume V ,

$$V = \int_0^1 \pi(x - x^4) dx$$

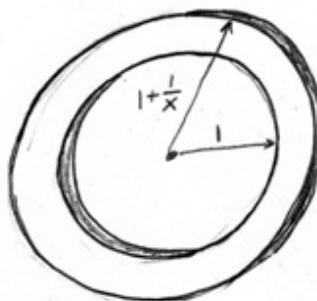
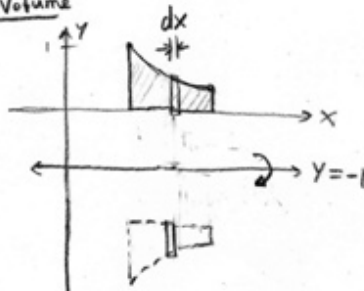
$$= \pi \left(\frac{1}{2} x^2 - \frac{1}{5} x^5 \right) \Big|_0^1$$

$$= \pi \left(\frac{1}{2} - \frac{1}{5} \right)$$

$$= \boxed{\frac{3\pi}{10}}$$

Example 8.3.7. .

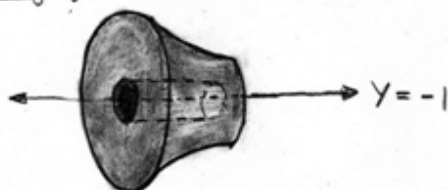
bound region by $Y = 1/x$, $Y = 0$, $X = 1$, $X = 3$ and rotate around $Y = -1$
 • Find its Volume



$$\begin{aligned} dV &= \pi (r_{out}^2 - r_{in}^2) dx \\ &= \pi \left(\left(1 + \frac{1}{x}\right)^2 - 1 \right) dx \\ &= \pi \left(1 + \frac{2}{x} + \frac{1}{x^2} - 1 \right) dx \\ &= \pi \left(\frac{2}{x} + \frac{1}{x^2} \right) dx \end{aligned}$$

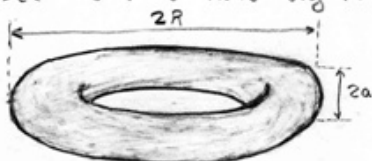
$$\begin{aligned} V &= \int_1^3 \pi \left(\frac{2}{x} + \frac{1}{x^2} \right) dx = \left(2\pi \ln(x) - \frac{\pi}{x} \right) \Big|_1^3 = \\ &= \left(2\pi \ln(x) - \frac{\pi}{x} \right) \Big|_1^3 \\ &= 2\pi \ln(3) - \frac{\pi}{3} + \pi \\ &= \boxed{2\pi \left(\ln(3) + \frac{1}{3} \right)} \end{aligned}$$

Roughly this solid looks like,

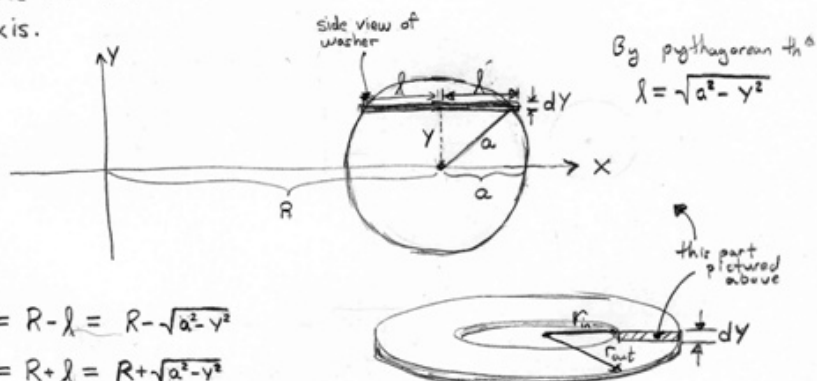


Example 8.3.8.

Find volume of a donut.
 "torus" in math. Let the torus have big radius R and little radius a .



We can see that this shape can be obtained by rotating a circle of radius a centered at $X = R$ around the Y -axis.



$$r_{in} = R - l = R - \sqrt{a^2 - y^2}$$

$$r_{out} = R + l = R + \sqrt{a^2 - y^2}$$

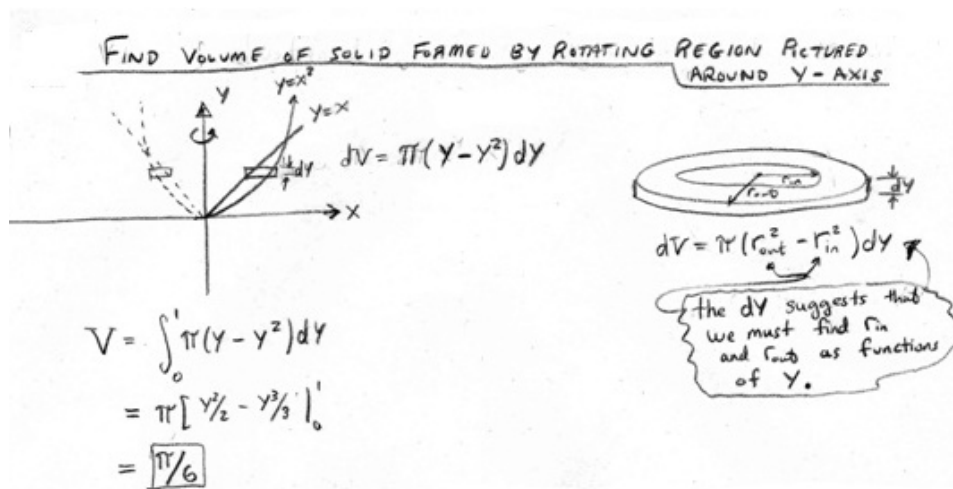
Find the volume of a typical washer,

$$\begin{aligned} dV &= \pi r_{out}^2 dy - \pi r_{in}^2 dy \\ &= \pi [(R+l)^2 - (R-l)^2] dy \\ &= \pi [R^2 + 2Rl + l^2 - (R^2 - 2Rl + l^2)] dy \\ &= 4\pi R l dy \\ &= 4\pi R \sqrt{y^2 - a^2} dy \quad (\text{you can see a washer at each } y \text{ from } y = -a \text{ to } y = a) \end{aligned}$$

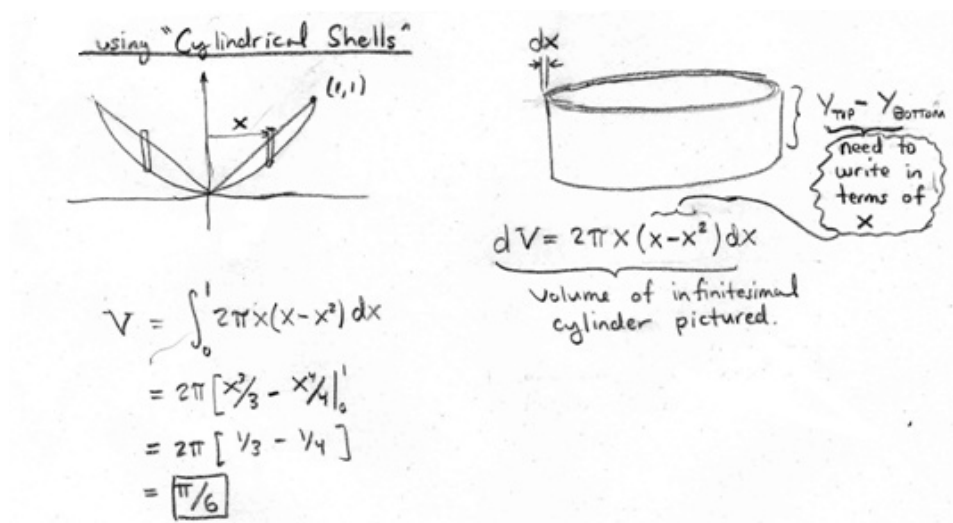
$$\begin{aligned} V &= \int_{-a}^a 4\pi R \sqrt{y^2 - a^2} dy \\ &= \int_{-\pi/2}^{\pi/2} 4\pi R (a \cos \theta) a \cos \theta d\theta \quad \left\{ \begin{array}{l} \text{notice this subst. is geometrically pleasing.} \\ y = a \sin \theta \quad y = a \rightarrow \theta = \pi/2 \\ dy = a \cos \theta d\theta \quad y = -a \rightarrow \theta = -\pi/2 \\ \sqrt{y^2 - a^2} = a \cos \theta \end{array} \right. \\ &= 4\pi R a^2 \int_{-\pi/2}^{\pi/2} \frac{1}{2} (1 + \cos 2\theta) d\theta \\ &= 2\pi R a^2 \left(\theta + \frac{1}{2} \sin(2\theta) \right) \Big|_{-\pi/2}^{\pi/2} = 2\pi^2 R a^2 = V \end{aligned}$$

Remark: intuitively nice
 $V = (2\pi R)(\pi a^2)$

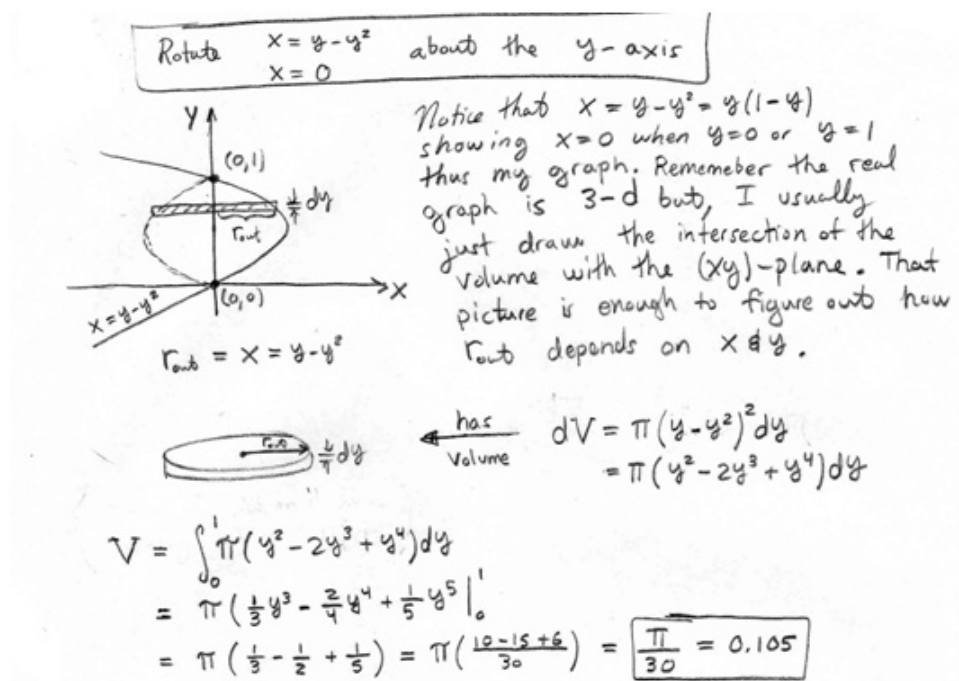
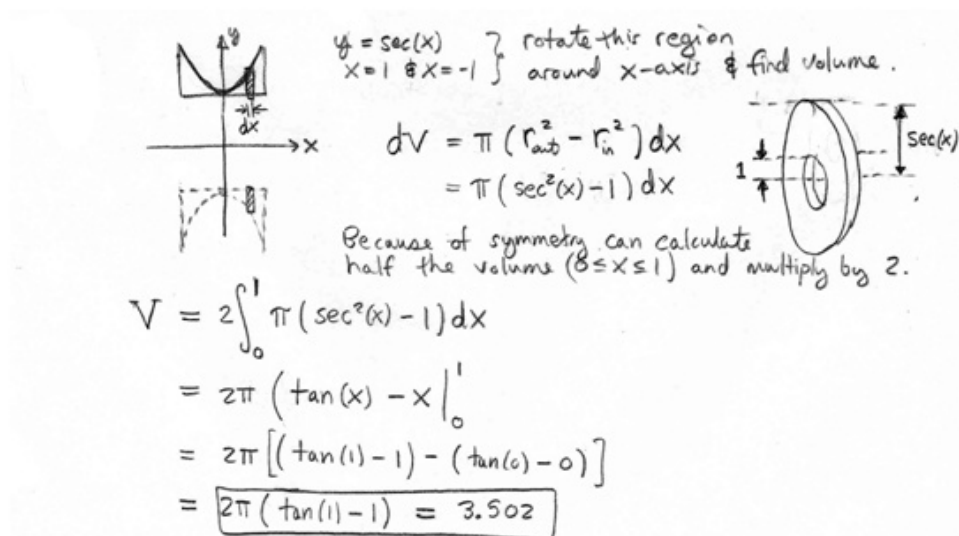
Example 8.3.9. we calculate the volume by the method of washers.



Example 8.3.10. same as the previous example, but this time we calculate by the method of cylindrical shells.

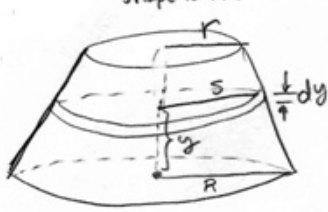


Example 8.3.11. .

Example 8.3.12. Note, we also include the boundary $y = 1$,

Example 8.3.13. .

heights of shape is h .




s depends linearly on y

$$S = my + b$$

$$S(0) = m(0) + b = b = R$$

$$S(h) = mh + R = r$$

$$\Rightarrow mh = r - R \therefore m = \frac{r - R}{h}$$

$$\therefore S = \left(\frac{r - R}{h}\right)y + R$$


$$A = \pi s^2$$

$$dV = \pi s^2 dy$$

$$= \pi \left[\left(\frac{r - R}{h}\right)y + R \right]^2 dy$$

$$= \pi \left[\left(\frac{r - R}{h}\right)^2 y^2 + 2R\left(\frac{r - R}{h}\right)y + R^2 \right] dy$$

$$V = \pi \int_0^h \left[\left(\frac{r - R}{h}\right)^2 y^2 + 2R\left(\frac{r - R}{h}\right)y + R^2 \right] dy$$

$$= \pi \left[\left(\frac{r - R}{h}\right)^2 \frac{1}{3} y^3 + 2R\left(\frac{r - R}{h}\right) \frac{1}{2} y^2 + R^2 y \right]_0^h$$

$$= \pi \left[\frac{1}{3} \left(\frac{r - R}{h}\right)^2 h^3 + R\left(\frac{r - R}{h}\right) h^2 + R^2 h \right]$$

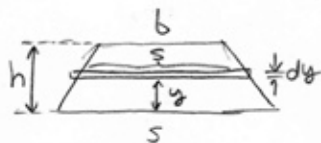
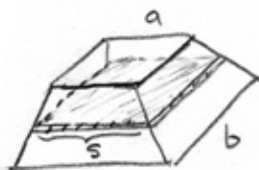
$$= \pi h \left[\frac{1}{3} (r - R)^2 + Rr - R^2 + R^2 \right]$$

$$= \frac{1}{3} \pi h \left[(r - R)^2 + 3Rr \right]$$

$$= \frac{1}{3} \pi h \left[r^2 - 2Rr + R^2 + 3Rr \right]$$

$$= \boxed{\frac{1}{3} \pi h \left[r^2 + Rr + R^2 \right] = V}$$

Example 8.3.14. .



s depends linearly
on y thus

$$s = my + B$$

From picture

$$s(0) = b$$

$$s(h) = a$$

Continuing
$$\left. \begin{aligned} s(0) &= m(0) + B = B = b \\ s(h) &= m(h) + b = a \\ m &= \frac{a-b}{h} \end{aligned} \right\} \Rightarrow s = \left(\frac{a-b}{h} \right) y + b$$

The area of the cross-section is just $A = s^2$ (square) thus

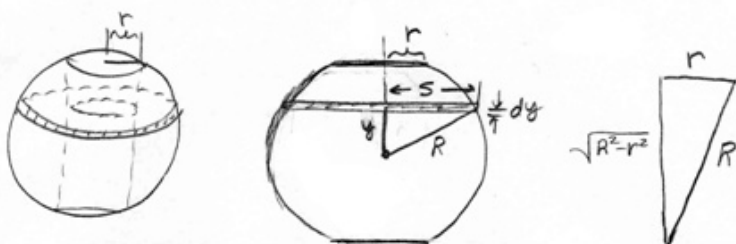
$$dV = s^2 dy = [my + b]^2 dy = [m^2 y^2 + 2mb y + b^2] dy$$

Therefore,

$$\begin{aligned} V &= \int_0^h [m^2 y^2 + 2mb y + b^2] dy \\ &= \left(\frac{1}{3} m^2 y^3 + \frac{2mb}{2} y^2 + b^2 y \right) \Big|_0^h \\ &= \frac{1}{3} m^2 h^3 + mb h^2 + b^2 h \\ &= \frac{1}{3} \left(\frac{a-b}{h} \right)^2 h^3 + \left(\frac{a-b}{h} \right) b h^2 + b^2 h \\ &= h \left(\frac{1}{3} (a-b)^2 + (a-b)b + b^2 \right) \\ &= \frac{1}{3} h \left((a-b)^2 + 3ab \right) \\ &= \frac{1}{3} h \left(a^2 - 2ab + b^2 + 3ab \right) \\ &= \boxed{\frac{1}{3} h \left(a^2 + ab + b^2 \right)} \end{aligned}$$

I figure why
write $\frac{a-b}{h}$ when
I could just write
 m instead.

Example 8.3.15. .



$$\begin{aligned}
 dV &= \pi(r_{\text{out}}^2 - r_{\text{in}}^2)dy \\
 &= \pi(s^2 - r^2)dy \\
 &= \pi(R^2 - y^2 - r^2)dy \\
 &= \pi(R^2 - r^2 - y^2)dy
 \end{aligned}$$

$$\begin{aligned}
 y^2 + s^2 &= R^2 \\
 s^2 &= R^2 - y^2
 \end{aligned}$$

Notice that
 $-\sqrt{R^2 - r^2} \leq y \leq \sqrt{R^2 - r^2}$
 we add volume of washers in this range
 By symmetry can just integrate from zero $\rightarrow \sqrt{R^2 - r^2}$ and double it.

$$\begin{aligned}
 V &= 2 \int_0^{\sqrt{R^2 - r^2}} \pi(R^2 - r^2 - y^2)dy \\
 &= 2\pi \left((R^2 - r^2)y - \frac{1}{3}y^3 \right) \Big|_0^{\sqrt{R^2 - r^2}} \\
 &= 2\pi \left((R^2 - r^2)\sqrt{R^2 - r^2} - \frac{1}{3}(\sqrt{R^2 - r^2})^3 \right) \\
 &= 2\pi \left((R^2 - r^2)^{3/2} - \frac{1}{3}(R^2 - r^2)^{3/2} \right) \\
 &= \boxed{\frac{4\pi}{3}(R^2 - r^2)^{3/2} = V}
 \end{aligned}$$

Notice when $r=0$ we get $V = \frac{4}{3}\pi R^3$ which is a good thing.

Example 8.3.16. .

Rotate $\begin{cases} y=x \\ y=\sqrt{x} \end{cases}$ about $x=2$. Notice $x=\sqrt{x} \Rightarrow x^2=x$
 $\Rightarrow x(x-1)=0$
 $\Rightarrow x=0 \text{ \& } x=1$
 (intersection pts.)

$x_R = y$
 $x_L = y^2$

$$r_{out} = 2 - x_L = 2 - y^2$$

$$r_{in} = 2 - x_R = 2 - y$$

$$dV = \pi(r_{out}^2 - r_{in}^2) dy$$

$$= \pi[(2-y^2)^2 - (2-y)^2] dy$$

$$= \pi[4 - 4y^2 + y^4 - (4 - 4y + y^2)] dy$$

$$= \pi[y^4 - 5y^2 + 4y] dy$$

$$\Rightarrow V = \int_0^1 \pi(y^4 - 5y^2 + 4y) dy = \pi\left(\frac{1}{5}y^5 - \frac{5}{3}y^3 + 2y^2\right)\bigg|_0^1 = \pi\left(\frac{1}{5} - \frac{5}{3} + 2\right)$$

$$= \pi\left(\frac{3-25+30}{15}\right)$$

$$= \boxed{\frac{8\pi}{15} \approx 1.6755}$$

Chapter 9

Integration Techniques

In this chapter it is assumed you have a complete and working knowledge of both basic antidifferentiation theory and the Fundamental Theorems of Calculus parts I and II. You must know the basic integrals as a reflex, if you have to look them up then you are not ready to start this Chapter. In addition, it is assumed you have mastered all the elementary u-substitutions and are ready, willing and able to solve definite and indefinite integrals which require u-substitution. If this is not the case then you need to review.

9.1 integration by parts (IBP)

I'm breaking from tradition a little here. I've decided to use U and V as opposed to u and v for integration by parts. Of course you could use other letters if you don't like these.

Theorem 9.1.1. *integration by parts.*

Suppose that U, V are differentiable functions on some connected subset $J \subseteq \mathbb{R}$. Then,

$$(1.) \int U \frac{dV}{dx} dx = UV - \int V \frac{dU}{dx} dx.$$

Moreover, if $a, b \in J$ then

$$(2.) \int_a^b U \frac{dV}{dx} dx = U(b)V(b) - U(a)V(a) - \int_a^b V(x) \frac{dU}{dx} dx.$$

Notice that is more common to use the following notation to put this theorem into practice:

$$\int U dV = UV - \int V dU.$$

here we mean for the expressions to be evaluated at $x \in J$ once the integrations are complete.

Proof: Consider that by linearity of differentiation we find

$$\frac{d}{dx} \left[UV - \int V \frac{dU}{dx} dx \right] = \frac{d}{dx} [UV] - \frac{d}{dx} \left[\int V \frac{dU}{dx} dx \right]$$

The rightmost term simplifies to $V \frac{dU}{dx}$ by the definition of the indefinite integral and the other term gives $\frac{d}{dx}[UV] = \frac{dU}{dx}V + U \frac{dV}{dx}$ by the product rule. Thus,

$$\frac{d}{dx} \left[UV - \int V \frac{dU}{dx} dx \right] = \frac{dU}{dx}V + U \frac{dV}{dx} - V \frac{dU}{dx} = U \frac{dV}{dx}.$$

On the other hand, by the definition of the indefinite integral,

$$\frac{d}{dx} \left[\int U \frac{dV}{dx} dx \right] = U \frac{dV}{dx}.$$

Therefore, we find the derivatives of the l.h.s and r.h.s. of (1.) agree hence by the definition of indefinite integration the equality is justified (recall indefinite integration is a shorthand for a whole class of antiderivative functions which differ by at most a constant). The proof of (2.) follows immediately from (1.) together with FTC part II. \square

Example 9.1.2.

$$\begin{aligned} \int x \sin(x) dx &= \int \underbrace{x}_U \underbrace{\sin(x) dx}_{dV} = UV - \int V dU \\ &= -x \cos(x) - \int (-\cos(x)) dx \\ &= \boxed{-x \cos(x) + \sin(x) + c.} \end{aligned}$$

I observed that $V = \int dV = \int \sin(x) dx = -\cos(x) + c$ and $dU = dx$. You might wonder why I didn't have to add a constant to V . Let's discuss this in general:

Suppose you replace V with $\tilde{V} = V + c$ for some constant c then the formula for IBP changes to:

$$\begin{aligned} \int U d\tilde{V} &= U\tilde{V} - \int \tilde{V} dU \\ &= U(V + c) - \int (V + c) dU \\ &= UV + cU - c \int dU - \int V dU \\ &= UV - \int V dU. \end{aligned}$$

Therefore, either V or $V + c$ will yield the same answer under IBP so we are free to choose c however we do so desire. For example, in the first example I just took $c = 0$ for convenience. It is very likely I will continue to do so for the foreseeable future. Another way of stating this result is to comment that when we find V from dV we just need an antiderivative, not the indefinite integral. We take account of the indefinite integral at the end of the calculation when we add c in the final step. You may recall from calculus I that the indefinite integral is technically a whole family of functions whereas the antiderivative is a specific function. Since some of you didn't have me for calculus I, I will elaborate a little further: if we were more technical and less traditional we'd have to write something like $\int x dx = \{\frac{1}{2}x^2 + c \mid c \in \mathbb{R}\}$. But, to be less notationally obtuse we have agreed to write $\int x dx = \frac{1}{2}x^2 + c$ to indicate the set of functions. Ok, now that I have you all properly confused, let's do some more examples. Come back to this paragraph later if at first it doesn't make sense.

Example 9.1.3.

$$\begin{aligned}
\int_0^1 x 2^x dx &= \int_0^1 \underbrace{x}_U \underbrace{2^x dx}_{dV} = UV \Big|_0^1 - \int_0^1 V dU \\
&= \frac{1}{\ln(2)} x 2^x \Big|_0^1 - \int_0^1 \frac{1}{\ln(2)} 2^x dx \\
&= \frac{1}{\ln(2)} x 2^x \Big|_0^1 - \frac{1}{[\ln(2)]^2} 2^x \Big|_0^1 \\
&= \boxed{\frac{2}{\ln(2)} - \frac{1}{[\ln(2)]^2} (2 - 1)} \\
&= \boxed{\frac{2}{\ln(2)} - \frac{1}{[\ln(2)]^2}}.
\end{aligned}$$

Example 9.1.4. In the calculation below we use IBP twice.

$$\begin{aligned}
\int x^2 \sin(x) dx &= \int \underbrace{x^2}_{U_1} \underbrace{\sin(x) dx}_{dV_1} = -x^2 \cos(x) - \int (-\cos(x)) 2x dx \\
&= -x^2 \cos(x) + \int \underbrace{2x}_{U_2} \underbrace{\cos(x) dx}_{dV_2} \\
&= -x^2 \cos(x) + 2x \sin(x) - \int \sin(x) 2 dx \\
&= \boxed{-x^2 \cos(x) + 2x \sin(x) + 2 \cos(x) + c}.
\end{aligned}$$

The basic idea is that each application of IBP can reduce x^n to x^{n-1} . If we know how to calculate $\int f(x) dx$ then the integral $\int x^n f(x) dx$ is solvable by repeated application of IBP. **Challenge:** show that

$$\int x^3 \sin(x) dx = -x^3 \cos(x) + 3x^2 \sin(x) + 6x \cos(x) - 6 \sin(x) + c.$$

IBP often allows us to integrate inverse functions: in this type of example there is only one reasonable choice for dV , it must be dx hence $V = x$.

Example 9.1.5.

$$\begin{aligned}
\int \sin^{-1}(x) dx &= \int \underbrace{\sin^{-1}(x)}_U \underbrace{dx}_{dV} = x \sin^{-1}(x) - \int \frac{x dx}{\sqrt{1-x^2}} \\
&= x \sin^{-1}(x) + \frac{1}{2} \int \frac{dw}{\sqrt{w}} \quad (\text{here } w = 1 - x^2) \\
&= x \sin^{-1}(x) + \frac{1}{2} \frac{\sqrt{w}}{\frac{1}{2}} + c \\
&= \boxed{x \sin^{-1}(x) + \sqrt{1-x^2} + c}.
\end{aligned}$$

Challenge: follow the method of the example above to find integrals of $\cos^{-1}(x)$ and $\tan^{-1}(x)$. What about $x \sin^{-1}(x)$? Could you integrate that function? You may not have seen the hyperbolic trig. functions in calculus I. If that is the case then you should probably ask me in office hours.

Example 9.1.6.

$$\begin{aligned} \int \cosh^{-1}(x) dx &= \int \underbrace{\cosh^{-1}(x)}_U \underbrace{dx}_{dV} = x \cosh^{-1}(x) - \int \frac{x dx}{\sqrt{x^2 - 1}} \\ &= x \cosh^{-1}(x) - \frac{1}{2} \int \frac{dw}{\sqrt{w}} \quad (\text{here } w = x^2 - 1) \\ &= x \cosh^{-1}(x) - \frac{1}{2} \frac{\sqrt{w}}{\frac{1}{2}} + c \\ &= \boxed{x \cosh^{-1}(x) - \sqrt{x^2 - 1} + c.} \end{aligned}$$

Challenge: follow the method of the example above to find integrals of $\sinh^{-1}(x)$ and $\tanh^{-1}(x)$.

Example 9.1.7.

$$\begin{aligned} \int \ln(x) dx &= \int \underbrace{\ln(x)}_U \underbrace{dx}_{dV} = x \ln(x) - \int \frac{x dx}{x} \\ &= \boxed{x \ln(x) - x + c.} \end{aligned}$$

The pattern we've observed in the last couple examples is something like this:

$$\int g(x) dx = \int g(x) \frac{dx}{dx} dx = \int \left[\frac{d}{dx}[xg] - x \frac{dg}{dx} \right] dx = xg - \int x \frac{dg}{dx} dx.$$

It's good to remember that all IBP does is to use the product rule. If you forget the formula it should be easy to find it again from the product rule. The examples that follow here are probably more like your homework.

Example 9.1.8. *I encourage the use of brackets to eliminate careless sign errors.*

$$\begin{aligned} \int \sin(\ln(x)) dx &= \int \underbrace{\sin(\ln(x))}_{U_1} \underbrace{dx}_{dV_1} = x \sin(\ln(x)) - \int x \frac{\cos(\ln(x)) dx}{x} \\ &= x \sin(\ln(x)) - \left[\int \underbrace{\cos(\ln(x))}_{U_2} \underbrace{dx}_{dV_2} \right] \\ &= x \sin(\ln(x)) - \left[x \cos(\ln(x)) - \int x \left(-\frac{\sin(\ln(x)) dx}{x} \right) \right] \\ &= x \sin(\ln(x)) - \left[x \cos(\ln(x)) + \int \sin(\ln(x)) dx \right] \\ &= x \sin(\ln(x)) - x \cos(\ln(x)) - \int \sin(\ln(x)) dx. \end{aligned}$$

Ok, what's the answer? You tell me. Hint: let $I = \int \sin(\ln(x)) dx$ and solve for I . Don't forget to add c to the final result since we're calculating an indefinite integral.

This pattern of looping back to where we start is found in a number of common integrals.

Example 9.1.9. *I could switch and use the $\sin(3x + 1)dx$ as the dV , but I think it's easier to choose $dV = e^x dx$ since e^x is trivial to integrate.*

$$\begin{aligned}
 \int e^x \sin(3x + 1) dx &= \int \underbrace{\sin(3x + 1)}_{U_1} \underbrace{e^x dx}_{dV_1} = e^x \sin(3x + 1) - 3 \int \cos(3x + 1) e^x dx \\
 &= e^x \sin(3x + 1) - 3 \left[\int \underbrace{\cos(3x + 1)}_{U_2} \underbrace{e^x dx}_{dV_2} \right] \\
 &= e^x \sin(3x + 1) - 3 \left[e^x \cos(3x + 1) - 3 \int -\sin(3x + 1) e^x dx \right] \\
 &= e^x \sin(3x + 1) - 3e^x \cos(3x + 1) - 9 \int e^x \sin(3x + 1) dx \\
 &= e^x \sin(3x + 1) - 3e^x \cos(3x + 1) - 9 \int e^x \sin(3x + 1) dx.
 \end{aligned}$$

Therefore,

$$\boxed{\int e^x \sin(3x + 1) dx = \frac{1}{10} e^x [\sin(3x + 1) - \cos(3x + 1)] + c.}$$

I'm not sure what will happen in the next example. This is an experiment.

Example 9.1.10.

$$\begin{aligned}
 \int \sin^2(x) dx &= \int \underbrace{\sin(x)}_{U_1} \underbrace{\sin(x) dx}_{dV_1} = -\sin(x) \cos(x) + \int \cos(x) \cos(x) dx \\
 &= -\sin(x) \cos(x) + \int (1 - \sin^2(x)) dx \\
 &= -\sin(x) \cos(x) + x - \int \sin^2(x) dx
 \end{aligned}$$

Therefore, by algebra! ,

$$\boxed{\int \sin^2(x) dx = \frac{1}{2} (x - \sin(x) \cos(x)) + c.}$$

I usually integrate $\sin^2(x)$ via the trigonometric identity $\sin^2(x) = \frac{1}{2}(1 - \cos(2x))$ from which it is obvious that $\int \sin^2(x) dx = \frac{x}{2} - \frac{1}{4} \sin(2x) + c$. Can you see the answer we just found from IBP is the same answer?

A formula which is defined by showing how one step goes to the next step is called a *recursive* formula. There are a number of interesting recursive formulas which allow us to integrate arbitrarily high powers of trigonometric functions through a simple step-by-step procedure. I don't expect you memorize these, but I do expect you could re-derive them if asked. In other words, you should understand the next example.

Example 9.1.11. Let $k \in \mathbb{N}$ and consider:

$$\begin{aligned} \int \sin^k(x) dx &= \int \underbrace{\sin^{k-1}(x)}_{U_1} \underbrace{\sin(x)}_{dV_1} dx = -\sin^{k-1}(x) \cos(x) + \int \cos(x)(k-1) \sin^{k-2}(x) \cos(x) dx \\ &= -\sin^{k-1}(x) \cos(x) + (k-1) \int \sin^{k-2}(x) [1 - \sin^2(x)] dx \\ &= -\sin^{k-1}(x) \cos(x) + (k-1) \int \sin^{k-2}(x) dx - (k-1) \int \sin^k(x) dx \end{aligned}$$

Solve for $\int \sin^k(x) dx$ (note $k \neq 0$ since $k \in \mathbb{N}$)

$$\boxed{\int \sin^k(x) dx = \frac{-1}{k} \sin^{k-1}(x) \cos(x) + \frac{k-1}{k} \int \sin^{k-2}(x) dx + c.}$$

We already have integrals for $\sin^k(x)$ in the cases $k = 1, 2$ so we can use the recursive formula above to calculate the cases $k = 3, 4$.

Example 9.1.12. Use Example 9.1.11 to find for the $k = 4$ case,

$$\begin{aligned} \int \sin^4(x) dx &= \frac{-1}{4} \sin^{4-1}(x) \cos(x) + \frac{4-1}{4} \int \sin^{4-2}(x) dx + c \\ &= \frac{-1}{4} \sin^3(x) \cos(x) + \frac{3}{4} \left(\frac{1}{2} (x - \sin(x) \cos(x)) \right) + c \\ &= \boxed{\frac{3}{8} x - \frac{3}{8} \sin(x) \cos(x) - \frac{1}{4} \sin^3(x) \cos(x) + c.} \end{aligned}$$

Example 9.1.13. Again, use Example 9.1.11 in the $k = 3$ case,

$$\begin{aligned} \int \sin^3(x) dx &= \frac{-1}{3} \sin^{3-1}(x) \cos(x) + \frac{3-1}{3} \int \sin^{3-2}(x) dx + c \\ &= \frac{-1}{3} \sin^2(x) \cos(x) - \frac{2}{3} \cos(x) + c \\ &= \frac{-1}{3} (1 - \cos^2(x)) \cos(x) - \frac{2}{3} \cos(x) + c \\ &= \boxed{\frac{1}{3} \cos^3(x) - \cos(x) + c.} \end{aligned}$$

9.2 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 and as a last resort IBP. I'll begin by attacking the problem of $\sin^3(x)$ which we just solved by IBP in the previous section. This is an easier way:

Example 9.2.1.

$$\begin{aligned}\int \sin^3(x)dx &= \int \sin^2(x) \sin(x)dx \\ &= \int [1 - u^2](-du) \quad (\text{where } u = \cos(x)) \\ &= -u + \frac{1}{3}u^3 + c \\ &= -\cos(x) + \frac{1}{3}\cos^3(x) + c\end{aligned}$$

The integral of $\sin^4(x)$ is not as easy in my view.

Example 9.2.2.

$$\begin{aligned}\int \sin^4(x)dx &= \int [\sin^2(x)]^2 dx \\ &= \int \left[\frac{1}{2}(1 - \cos(2x)) \right]^2 dx \\ &= \frac{1}{4} \int [1 - 2\cos(2x) + \cos^2(2x)] dx \\ &= \frac{x}{4} - \frac{1}{4}\sin(2x) + \frac{1}{8} \int (1 + \cos(4x)) dx \\ &= \frac{x}{4} - \frac{1}{4}\sin(2x) + \frac{x}{8} + \frac{1}{32}\sin(4x) + c \\ &= \boxed{\frac{3x}{8} - \frac{1}{4}\sin(2x) + \frac{1}{32}\sin(4x) + c.}\end{aligned}$$

I invite the reader to verify that Example 9.1.12 and the example above are consistent. Actually, would you know where to start in comparing these answers?

If you ponder the methods we just used to integrate $\sin^k(x)$ you should be able to integrate any sum or product of $\sin(x)$ and $\cos(x)$. For example, see if you can calculate the integrals $\int \sin(x) \cos(x)dx$ or $\int \sin^2(x) \cos^2(x)dx$. Sums of products and reciprocals of sine and cosine require more thought but, many are not too difficult.

Example 9.2.3. Let let $u = \cos(x)$ in the calculation below:

$$\begin{aligned}\int \frac{\sin(x)}{\cos(x)} dx &= \int \frac{-du}{u} \\ &= \boxed{-\ln |\cos(x)| + c.}\end{aligned}$$

Therefore, $\int \tan(x)dx = \ln |\sec(x)| + c$.

I hope you can figure out $\int \cot(x)dx$ with ease. 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 9.2.4.

$$\begin{aligned}\int \tan^2(x)dx &= \int (\sec^2(x) - 1)dx \\ &= \boxed{\tan(x) - x + c.}\end{aligned}$$

Example 9.2.5. We let $u = \tan(x)$ so $du = \sec^2(x)dx$,

$$\begin{aligned}\int \sec^2(x) \tan^2(x)dx &= \int u^2 du \\ &= \frac{1}{3}u^3 + c \\ &= \boxed{\frac{1}{3} \tan^3(x) + c.}\end{aligned}$$

Example 9.2.6.

$$\begin{aligned}\int \tan^4(x)dx &= \int \tan^2(x)(\sec^2(x) - 1)dx \\ &= \int \tan^2(x) \sec^2(x)dx - \int \tan^2(x)dx \\ &= \int \tan^2(x)d(\tan(x)) - \int \tan^2(x)dx \\ &= \boxed{\frac{1}{3} \tan^3(x) + \tan(x) - x + c.}\end{aligned}$$

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 integrate expressions like $\int \sec^6(x)dx$ or $\int \cot^2(x)dx$ or $\int \cot^2(x) \csc^2(x)dx$ using arguments paralleling the previous triple of examples. What lies beneath is scarier.

Example 9.2.7. Observe that if $u = \sec(x) + \tan(x)$ then $\frac{du}{u} = \sec(x)dx$ (work it out for yourself!). With this bit of trivia in mind note:

$$\begin{aligned}\int \sec(x)dx &= \int \frac{du}{u} \\ &= \ln |u| + c \\ &= \boxed{\ln |\sec(x) + \tan(x)| + c.}\end{aligned}$$

Ok, by now you should expect me to ask if you can integrate $\int \csc(x)dx$ given the patterns above. Given our work thus far it ought to be clear that integrating even powers of secant is actually pretty easy. On the other hand, the first odd power above required a stroke of genius. If you try to convert to a sine/cosine integral it does not help much if you were wondering. Maybe we'll return to secant in a future section and I'll show you a less clever way of calculating the integral. For now we move on to the dreaded $\sec^3(x)$.

Example 9.2.8. I use the result of the previous example to go from the 4th to the 5th line of the calculation below.

$$\begin{aligned}
 \int \sec^3(x) dx &= \int \underbrace{\sec(x)}_{U_1} \underbrace{\sec^2(x)}_{dV_1} dx \\
 &= \sec(x) \tan(x) - \int \tan(x) \sec(x) \tan(x) dx \\
 &= \sec(x) \tan(x) - \int (\sec^2(x) - 1) \sec(x) dx \\
 &= \sec(x) \tan(x) - \int \sec^3(x) dx + \int \sec(x) dx \\
 &= \sec(x) \tan(x) + \ln |\sec(x) + \tan(x)| - \int \sec^3(x) dx.
 \end{aligned}$$

Therefore,

$$\int \sec^3(x) dx = \frac{1}{2} \sec(x) \tan(x) + \frac{1}{2} \ln |\sec(x) + \tan(x)| + c.$$

Given the example above you ought to have a shot at completing integrals like $\int \sec^5(x) dx$ or $\int \csc^3(x) dx$. The integrals that follow in this section require a comment:

Remark 9.2.9.

See Section 2.9 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:

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix}) \quad \text{and} \quad \cos x = \frac{1}{2}(e^{ix} + e^{-ix})$$

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 $\cos(\theta + \beta) = \cos \theta \cos \beta - \sin \theta \sin \beta$ whereas the adding angles formula for sine is $\sin(\theta + \beta) = \sin \theta \cos \beta + \cos \theta \sin \beta$. Together these adding angles formulas for sine and cosine yield another for tangent; $\tan(\theta + \beta) = \frac{\tan \theta + \tan \beta}{1 - \tan \theta \tan \beta}$. Finally the product identities for sine and cosine are also very useful and for most of us far from obvious;

$$\cos(ax) \cos(bx) = \frac{1}{2} \cos[(a+b)x] + \frac{1}{2} \cos[(a-b)x]$$

and

$$\sin(ax) \sin(bx) = \frac{1}{2} \cos[(a+b)x] - \frac{1}{2} \cos[(a-b)x]$$

and

$$\cos(ax) \sin(bx) = \frac{1}{2} \sin[(a+b)x] + \frac{1}{2} \sin[(a-b)x].$$

The product formulas are very important to the study of constructive and destructive interference 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 9.2.10.

$$\begin{aligned} \int \cos(3x) \sin(5x) dx &= \int \left[\frac{1}{2} \sin(8x) + \frac{1}{2} \sin(-2x) \right] dx \\ &= \frac{1}{2} \int \sin(8x) dx - \frac{1}{2} \int \sin(2x) dx \\ &= \boxed{\frac{-1}{16} \cos(8x) - \frac{1}{4} \cos(2x) + c.} \end{aligned}$$

Example 9.2.11.

$$\begin{aligned} \int \cos(3x) \cos(5x) dx &= \int \left[\frac{1}{2} \cos(8x) + \frac{1}{2} \cos(-2x) \right] dx \\ &= \frac{1}{2} \int \cos(8x) dx + \frac{1}{2} \int \cos(2x) dx \\ &= \boxed{\frac{1}{16} \sin(8x) + \frac{1}{4} \sin(2x) + c.} \end{aligned}$$

Example 9.2.12.

$$\begin{aligned} \int \sin(3x) \sin(5x) dx &= \int \left[\frac{1}{2} \cos(8x) - \frac{1}{2} \cos(-2x) \right] dx \\ &= \frac{1}{2} \int \cos(8x) dx - \frac{1}{2} \int \cos(2x) dx \\ &= \boxed{\frac{1}{16} \sin(8x) - \frac{1}{4} \sin(2x) + c.} \end{aligned}$$

What about $\int \sin(x) \cos(3x) \cos(6x) dx$? How would you attack such a problem?

Example 9.2.13. 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 parallel to those given in this section. You may be asked to discover some of those for yourself in the homework.

9.3 implicit substitutions

All of the substitutions made thus far have been explicit. This means the new substituted variable is defined explicitly as a function of the given variable of integration. An implicit substitution may give the change of variables in many other forms. Most common among these is the so-called trig-substitution where we usually either use $x = a \cos(\theta)$ or $x = a \sin(\theta)$ or $x = a \sec(\theta)$ or $x = a \tan(\theta)$. In each of these cases the new variable which is analogous to the u of our earlier work is played by the variable θ . We'll also look at substitutions based on the inverse hyperbolic functions. We will need the techniques of the previous section to properly complete many of these problems. It really is a beautiful chapter in the theory of calculation, an art which is slowly but surely being lost to the ever encroaching hob-goblin of mediocrity in the calculus sequence.

Example 9.3.1. We wish to calculate $\int \sqrt{4 - x^2} dx$. The squareroot is trouble. One sneaky way to eliminate it is to let $x = 2 \sin(\theta)$ thus $4 - x^2 = 4 - 4 \sin^2(\theta) = 4 \cos^2(\theta)$ and $dx = 2 \cos(\theta) d\theta$. Hence,

$$\begin{aligned}
 \int \sqrt{4 - x^2} dx &= \int \sqrt{4 \cos^2(\theta)} 2 \cos(\theta) d\theta \\
 &= \int 4 \cos^2(\theta) d\theta \\
 &= \int [2 + \cos(2\theta)] d\theta \\
 &= 2\theta + \frac{1}{2} \sin(2\theta) + c \\
 &= \boxed{2 \sin^{-1}\left(\frac{x}{2}\right) + \frac{1}{2} \sin\left(2 \sin^{-1}\left[\frac{x}{2}\right]\right) + c.}
 \end{aligned}$$

The careful student will have questioned how I knew that $\sqrt{4 \cos^2(\theta)} = 2 \cos(\theta)$. After all, algebraically we ought to have $\sqrt{4 \cos^2(\theta)} = 2 |\cos(\theta)|$. The answer is simply that we insist θ be chosen such that $\cos(2\theta) \geq 0$, otherwise we'd have to introduce a sign in that step. Similar comments apply to future examples, but I will not belabor this point further.

Example 9.3.2. We wish to calculate $\int \sqrt{x^2 - 9} dx$. The squareroot is trouble. One sneaky way to eliminate it is to let $x = 3 \sec(\theta)$ thus $x^2 - 9 = 9 \sec^2(\theta) - 9 = 9 \tan^2(\theta)$ and $dx = 3 \sec(\theta) \tan(\theta) d\theta$. Hence,

$$\begin{aligned} \int \sqrt{x^2 - 9} dx &= \int \sqrt{9 \tan^2(\theta)} 3 \sec(\theta) \tan(\theta) d\theta \\ &= \int 9 \sec(\theta) \tan^2(\theta) d\theta \\ &= \int 9 \sec(\theta) (\sec^2(\theta) - 1) d\theta \\ &= 9 \int \sec^3(\theta) d\theta - 9 \int \sec(\theta) d\theta \\ &= \frac{9}{2} \sec(\theta) \tan(\theta) + \frac{9}{2} \ln |\sec(\theta) + \tan(\theta)| - 9 \ln |\sec(\theta) + \tan(\theta)| + c \\ &= \frac{9}{2} \sec(\theta) \tan(\theta) - \frac{9}{2} \ln |\sec(\theta) + \tan(\theta)| + c \end{aligned}$$

We can simplify this answer nicely if we think about the substitution in terms of a triangle. Note that if $x = 3 \sec(\theta)$ then $\sec(\theta) = \frac{x}{3} = \frac{\text{hyp}}{\text{adj}}$ hence $\text{opp} = \sqrt{x^2 - 9}$ and so $\tan(\theta) = \frac{\sqrt{x^2 - 9}}{3}$. We find,

$$\boxed{\int \sqrt{x^2 - 9} dx = \frac{1}{2} x \sqrt{x^2 - 9} - \frac{9}{2} \ln \left| \frac{x}{3} + \frac{\sqrt{x^2 - 9}}{3} \right| + c.}$$

Example 9.3.3. The fundamental identity for hyperbolic trigonometry is $\cosh^2(\phi) - \sinh^2(\phi) = 1$. We can rearrange this to give $\cosh^2(\phi) - 1 = \sinh^2(\phi)$. This suggests we might be able to work the previous example by making a $x = 3 \cosh(\phi)$ substitution. If $x = 3 \cosh(\phi)$ then $dx = 3 \sinh(\phi) d\phi$ and $x^2 - 9 = 9 \sinh^2(\phi)$. Hence,

$$\int \sqrt{x^2 - 9} dx = \int 9 \sinh^2(\phi) d\phi.$$

I need an identity for $\sinh^2(\phi)$, let's derive it from scratch,

$$\sinh^2(\phi) = \left[\frac{1}{2} (e^\phi - e^{-\phi}) \right]^2 = \frac{1}{4} [e^{2\phi} - 2 + e^{-2\phi}] = \frac{1}{2} \cosh(2\phi) - \frac{1}{2}$$

Returning to our integration with this new-found insight,

$$\int \sqrt{x^2 - 9} dx = \int \left(\frac{9}{2} \cosh(2\phi) - \frac{9}{2} \right) d\phi = \frac{9}{4} \sinh(2\phi) - \frac{9\phi}{2} + c.$$

Note that $2 \sinh(\phi) \cosh(\phi) = \frac{1}{2} (e^\phi - e^{-\phi})(e^\phi + e^{-\phi}) = \frac{1}{2} (e^{2\phi} - e^{-2\phi}) = \sinh(2\phi)$. Furthermore, we began by supposing that $x = 3 \cosh(\phi)$ hence $\cosh(\phi) = \frac{x}{3}$ and $\sinh(\phi) = \sqrt{x^2 - 9}$. We find that $\frac{9}{4} \sinh(2\phi) = \frac{9}{2} \frac{x}{3} \frac{\sqrt{x^2 - 9}}{3} = \frac{1}{2} x \sqrt{x^2 - 9}$. On the other hand, in our current formalism we are led to write $\frac{9\phi}{2} = \frac{9}{2} \cosh^{-1}\left(\frac{x}{3}\right)$. Thus,

$$\boxed{\int \sqrt{x^2 - 9} dx = \frac{1}{2} x \sqrt{x^2 - 9} - \frac{9}{2} \cosh^{-1}\left(\frac{x}{3}\right) + c.}$$

You should consult Stewart pg. 466 equation 4 to help see why the last two examples are in fact consistent. Also, reading section 7.7 of Stewart will almost certainly aid you in your success with the integration technique we currently consider.

Example 9.3.4. We wish to calculate $\int \sqrt{4x^2 + 9} \, dx$. The squareroot is trouble. One sneaky way to eliminate it is to let $2x = 3 \tan(\theta)$ thus $4x^2 + 9 = 9 \tan^2(\theta) + 9 = 9 \sec^2(\theta)$ and $2dx = 3 \sec^2(\theta) d\theta$. Hence,

$$\begin{aligned} \int \sqrt{4x^2 + 9} \, dx &= \int \sqrt{9 \sec^2(\theta)} \frac{3 \sec^2(\theta) d\theta}{2} \\ &= \frac{9}{2} \int \sec^3(\theta) d\theta \\ &= \frac{9}{4} \sec(\theta) \tan(\theta) + \frac{9}{4} \ln |\sec(\theta) + \tan(\theta)| + c \end{aligned}$$

We can simplify this answer nicely if we think about the substitution in terms of a triangle. Note that if $2x = 3 \tan(\theta)$ then $\tan(\theta) = \frac{2x}{3} = \frac{\text{opp}}{\text{adj}}$ hence $\text{hyp} = \sqrt{4x^2 + 9}$ and so $\sec(\theta) = \frac{\sqrt{4x^2 + 9}}{3}$. We find,

$$\int \sqrt{4x^2 + 9} \, dx = \frac{1}{2} x \sqrt{4x^2 + 9} + \frac{9}{4} \ln \left| \frac{\sqrt{4x^2 + 9}}{3} + \frac{2x}{3} \right| + c.$$

Notice, we can simplify this answer by factoring out a $1/3$ in the natural log argument,

$$\int \sqrt{4x^2 + 9} \, dx = \frac{1}{2} x \sqrt{4x^2 + 9} + \frac{9}{4} \ln |\sqrt{4x^2 + 9} + 2x| + c.$$

Example 9.3.5. Another way to calculate $\int \sqrt{4x^2 + 9} \, dx$ is to make a hyperbolic substitution. Observe that $\cosh^2(\phi) - \sinh^2(\phi) = 1$ gives $\cosh^2(\phi) = \sinh^2(\phi) + 1$. This suggests we can make a $2x = 3 \sinh(\phi)$ substitution. If $2x = 3 \sinh(\phi)$ then $4x^2 + 9 = 9 \sinh^2(\phi) + 9 = 9 \cosh^2(\phi)$ and $2dx = 3 \cosh(\phi) d\phi$.

$$\begin{aligned} \int \sqrt{4x^2 + 9} \, dx &= \int \sqrt{9 \cosh^2(\phi)} \frac{3 \cosh(\phi) d\phi}{2} \\ &= \frac{9}{2} \int \cosh^2(\phi) d\phi \end{aligned}$$

I need an identity for $\cosh^2(\phi)$, let's derive it from scratch,

$$\cosh^2(\phi) = \left[\frac{1}{2} (e^\phi + e^{-\phi}) \right]^2 = \frac{1}{4} [e^{2\phi} + 2 + e^{-2\phi}] = \frac{1}{2} \cosh(2\phi) + \frac{1}{2}$$

Returning to our integration with this new-found insight,

$$\begin{aligned} \int \sqrt{4x^2 + 9} \, dx &= \frac{9}{4} \int (\cosh(2\phi) + 1) d\phi \\ &= \frac{9}{8} \sinh(2\phi) + \frac{9}{4} \phi + c. \\ &= \frac{9}{4} \frac{2x}{3} \frac{\sqrt{4x^2 + 9}}{3} + \frac{9}{4} \phi + c. \\ &= \frac{1}{2} x \sqrt{4x^2 + 9} + \frac{9}{4} \sinh^{-1} \left(\frac{2x}{3} \right) + c. \end{aligned}$$

I used the identity $\sinh(2\phi) = 2\sinh(\phi)\cosh(\phi)$ to help simplify the answer.

Again, you can consult Stewart pg. 466 equation 4 to help see why the last two examples are in fact consistent. Also, by now you should start to appreciate that both the hyperbolic and the trig. substitutions have pros and cons. I believe a healthy approach is to be ready to apply either approach to a given problem. Knowledge of this mathematics will also aid you in interpreting Mathematica outputs. Mathematica might give the answer in terms of hyperbolics whereas you know the answer for other reasons in terms of logarithms. Hopefully as you work the problems in this section you'll start to understand that hyperbolic functions and trigonometric functions are intricately linked. The explicit connection and ultimate synthesis for these functions is largely covered in the complex variables course. The following triple of examples all invoke a bit of algebra which is commonly called "completing the square".

Example 9.3.6. *To begin I simplify the quadratic by completing the square:*

$$\int \frac{dx}{x^2 + 4x + 5} = \int \frac{dx}{(x+2)^2 + 1}.$$

Remember that $\tan^2(\theta) + 1 = \sec^2(\theta)$ so this makes me think a $x+2 = \tan(\theta)$ substitution will simplify this problem. If $x+2 = \tan(\theta)$ then $(x+2)^2 + 1 = \tan^2(\theta) + 1 = \sec^2(\theta)$ and $dx = \sec^2(\theta)d\theta$. Thus,

$$\begin{aligned} \int \frac{dx}{x^2 + 4x + 5} &= \int \frac{\sec^2(\theta)d\theta}{\sec^2(\theta)} \\ &= \theta + c \\ &= \boxed{\tan^{-1}(x+2) + c.} \end{aligned}$$

Example 9.3.7. *To begin I simplify the quadratic:*

$$\int \frac{dx}{4x^2 + 8x + 9} = \int \frac{dx}{4(x+1)^2 + 5}.$$

Remember that $\tan^2(\theta) + 1 = \sec^2(\theta)$ so this makes me think a $4(x+1)^2 = 5\tan^2(\theta)$ substitution will simplify this problem. If $2(x+1) = \sqrt{5}\tan(\theta)$ then $4(x+1)^2 + 5 = 5\tan^2(\theta) + 5 = 5\sec^2(\theta)$ and $2dx = \sqrt{5}\sec^2(\theta)d\theta$. Thus,

$$\begin{aligned} \int \frac{dx}{4x^2 + 8x + 9} &= \int \frac{\sqrt{5}\sec^2(\theta)d\theta}{2(5\sec^2(\theta))} \\ &= \frac{1}{2\sqrt{5}}\theta + c \\ &= \boxed{\frac{1}{2\sqrt{5}}\tan^{-1}\left(\frac{2(x+1)}{\sqrt{5}}\right) + c.} \end{aligned}$$

Example 9.3.8. *To begin I simplify the quadratic by completing the square:*

$$\int \frac{dx}{x^2 + 6x + 5} = \int \frac{dx}{(x+3)^2 - 4}.$$

Remember that $\tan^2(\theta) = \sec^2(\theta) - 1$ so this makes me think a $(x+3)^2 = 4\sec^2(\theta)$ substitution will simplify this problem. If $x+3 = 2\sec(\theta)$ then $(x+3)^2 - 4 = 4\sec^2(\theta) - 4 = 4\tan^2(\theta)$ and $dx = 2\sec(\theta)\tan(\theta)d\theta$. Thus,

$$\begin{aligned}\int \frac{dx}{x^2+6x+5} &= \int \frac{2\sec(\theta)\tan(\theta)d\theta}{4\tan^2(\theta)} \\ &= \int \frac{\sec(\theta)d\theta}{2\tan(\theta)} \\ &= \frac{1}{2} \int \csc(\theta)d\theta \\ &= \frac{-1}{2} \int \frac{du}{u} \quad \text{let } u = \csc(\theta) + \cot(\theta) \text{ hence } -du/u = \csc(\theta)d\theta \\ &= \frac{-1}{2} \ln |\csc(\theta) + \cot(\theta)| + c\end{aligned}$$

We let $x+3 = 2\sec(\theta)$ thus $\sec(\theta) = \frac{x+3}{2} = \frac{\text{hyp}}{\text{adj}}$. The other side of the substitution triangle is found by the pythagorean theorem; $\text{opp} = \sqrt{(x+3)^2 - 4}$. Observe that $\cot(\theta) = \frac{\text{adj}}{\text{opp}} = \frac{2}{\sqrt{(x+3)^2 - 4}}$ and $\csc(\theta) = \frac{\text{hyp}}{\text{opp}} = \frac{\frac{x+3}{2}}{\sqrt{(x+3)^2 - 4}}$. Therefore,

$$\int \frac{dx}{x^2+6x+5} = \frac{-1}{2} \ln \left| \frac{x+3}{\sqrt{(x+3)^2 - 4}} + \frac{2}{\sqrt{(x+3)^2 - 4}} \right| + c.$$

Let's attack the preceding example once more, but this time from the angle of a hyperbolic substitution.

Example 9.3.9. If $\cosh^2(\phi) - \sinh^2(\phi) = 1$ then dividing by $\sinh^2(\phi)$ reveals the hyperbolic cotangent/cosecant identity: $\coth^2(\phi) - 1 = \text{csch}^2(\phi)$ where by definition $\coth(\phi) = \frac{\cosh(\phi)}{\sinh(\phi)}$ and $\text{csch}(\phi) = \frac{1}{\sinh(\phi)}$. In view of the identity above, the integral $\int \frac{dx}{x^2+6x+5} = \int \frac{dx}{(x+3)^2 - 4}$ is likely simplified by an $(x+3)^2 = 4\coth^2(\phi)$ substitution. If $x+3 = 2\coth(\phi)$ then $(x+3)^2 - 4 = 4\coth^2(\phi) - 4 = 4\text{csch}^2(\phi)$ and ¹ $dx = -2\text{csch}^2(\phi)d\phi$.

$$\begin{aligned}\int \frac{dx}{x^2+6x+5} &= \int \frac{dx}{(x+3)^2 - 4} \\ &= \int \frac{-2\text{csch}^2(\phi)d\phi}{4\text{csch}^2(\phi)} \\ &= \frac{-1}{2} \phi + c \\ &= \frac{-1}{2} \coth^{-1} \left(\frac{x+3}{2} \right) + c.\end{aligned}$$

I show how $\int \frac{dx}{x^2+6x+5} = \frac{-1}{2} \tanh^{-1} \left(\frac{x+3}{2} \right) + c$ in the next example.

¹just use the quotient rule; $\coth(\phi)' = \frac{\cosh(\phi)'}{\sinh(\phi)} = \frac{\sinh^2(\phi) - \cosh^2(\phi)}{\sinh^2(\phi)} = \frac{-1}{\sinh^2(\phi)} = -\text{csch}^2(\phi)$

Example 9.3.10. If $\cosh^2(\phi) - \sinh^2(\phi) = 1$ then dividing by $\cosh^2(\phi)$ reveals the hyperbolic cotangent/cosecant identity: $1 - \tanh^2(\phi) = \operatorname{sech}^2(\phi)$ where by definition $\tanh(\phi) = \frac{\sinh(\phi)}{\cosh(\phi)}$ and $\operatorname{sech}(\phi) = \frac{1}{\cosh(\phi)}$. In view of the identity above, the integral $\int \frac{dx}{x^2+6x+5} = -\int \frac{dx}{4-(x+3)^2}$ is likely simplified by an $(x+3)^2 = 4 \tanh^2(\phi)$ substitution. If $x+3 = 2 \tanh(\phi)$ then $4 - (x+3)^2 = 4 - 4 \tanh^2(\phi) = 4 \operatorname{sech}^2(\phi)$ and $dx = 2 \operatorname{sech}^2(\phi) d\phi$.

$$\begin{aligned} \int \frac{dx}{x^2+6x+5} &= -\int \frac{dx}{4-(x+3)^2} \\ &= -\int \frac{2 \operatorname{sech}^2(\phi) d\phi}{4 \operatorname{sech}^2(\phi)} \\ &= \frac{-1}{2} \phi + c. \\ &= \boxed{\frac{-1}{2} \tanh^{-1}\left(\frac{x+3}{2}\right) + c.} \end{aligned}$$

Example 9.3.11. We wish to calculate $\int \frac{x^3 dx}{\sqrt{9-x^2}}$. The squareroot is trouble. One sneaky way to eliminate it is to let $x = 3 \sin(\theta)$ thus $9 - x^2 = 9 - 9 \sin^2(\theta) = 9 \cos^2(\theta)$ and $dx = 3 \cos(\theta) d\theta$. Hence,

$$\begin{aligned} \frac{x^3 dx}{\sqrt{9-x^2}} &= \int \frac{27 \sin^3(\theta) (3 \cos(\theta) d\theta)}{3 \cos(\theta)} \\ &= \int 27 \sin^3(\theta) d\theta \\ &= \int 27 (1 - \cos^2(\theta)) \sin(\theta) d\theta \\ &= \int [27u^2 - 27] du \quad \text{let } u = \cos(\theta) \\ &= 9 \cos^3(\theta) - 27 \cos(\theta) + c \end{aligned}$$

Note that $x = 3 \sin(\theta)$ gives $\sin(\theta) = \frac{x}{3} = \frac{\text{opp}}{\text{hyp}}$ where we have in mind a triangle which represents the substitution. Applying the pythagorean theorem we calculate $\text{adj} = \sqrt{9-x^2}$ hence $\cos(\theta) = \frac{\text{adj}}{\text{hyp}} = \frac{\sqrt{9-x^2}}{3}$ and we can simplify our answer nicely: $9 \cos^3(\theta) = \frac{9(\sqrt{9-x^2})^3}{27} = \frac{1}{3}(9-x^2)\sqrt{9-x^2}$ and $9 \cos^3(\theta) - 27 \cos(\theta) = \frac{1}{3}(9-x^2)\sqrt{9-x^2} - \frac{1}{3}27\sqrt{9-x^2}$,

$$\boxed{\int \frac{x^3 dx}{\sqrt{9-x^2}} = \frac{-1}{3} (18+x^2) \sqrt{9-x^2} + c.}$$

²just use the quotient rule; $\tanh(\phi)' = \frac{\sinh(\phi)'}{\cosh(\phi)} = \frac{\cosh^2(\phi) - \sinh^2(\phi)}{\cosh^2(\phi)} = \frac{1}{\cosh^2(\phi)} = \operatorname{sech}^2(\phi)$

Remark 9.3.12.

If we have an integral with bounds then we change the integrand, measure and the bounds under an implicit substitution. We have not faced that difficulty in this section so far because I have focused on indefinite integration. Also, once bounds enter in then one may hope for an explicit geometric interpretation of the substitution. I have emphasized an algebraic understanding in this section because I believe it will lead less students astray. In my experience, few students understand geometry well enough to appropriately modify it for the solution of nonstandard problems. That said, I will invoke a geometry-based substitution in the final example that follows.

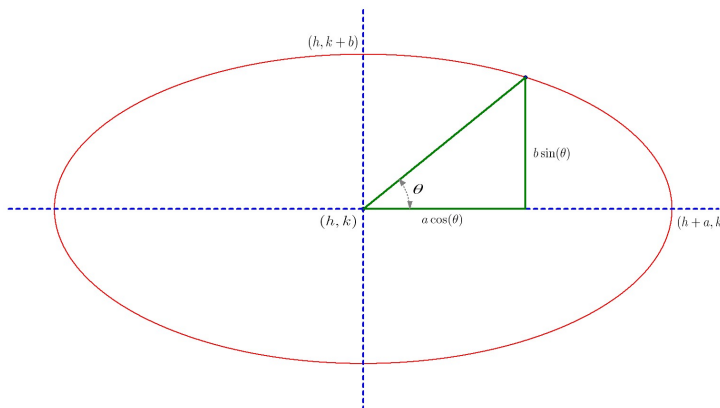
Example 9.3.13. Find the area bounded by $b^2(x-h)^2 + a^2(y-k)^2 = a^2b^2$ where $a, b, h, k \in \mathbb{R}$ and it is given that $a, b > 0$. This is an ellipse. Note if (x, y) is a solution to the given equation then it is also a solution to

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$$

and we identify that the area is an ellipse centered at (h, k) . We can view this ellipse as two graphs pasted together; note $(y-k)^2 = b^2[1 - \frac{(x-h)^2}{a^2}] = \frac{b^2}{a^2}[a^2 - (x-h)^2]$ has solutions

$$y_T = k + \frac{b}{a}\sqrt{a^2 - (x-h)^2} \quad \text{and} \quad y_B = k - \frac{b}{a}\sqrt{a^2 - (x-h)^2}.$$

These solutions have domains which are governed by the inequality $a^2 - (x-h)^2 \geq 0$ hence $a^2 \geq |x-h|^2$ and as $a > 0$ we find $|x-h| \leq a$ which is equivalent to $h-a \leq x \leq h+a$.



At each $x \in [h-a, h+a]$ we can envision an infinitesimal rectangle of width dx and height $y_T - y_B = 2\frac{b}{a}\sqrt{a^2 - (x-h)^2}$ hence the typical infinitesimal area is given by $dA = 2\frac{b}{a}\sqrt{a^2 - (x-h)^2} dx$ (I'll let you draw in the dA) and we can calculate the total area by integration:

$$A = \int_{\text{ellipse}} dA = \int_{h-a}^{h+a} 2\frac{b}{a}\sqrt{a^2 - (x-h)^2} dx$$

This is a nontrivial integration. However, we can do it. We let $x = h + a \cos(\theta)$ hence $x-h = a \cos(\theta)$ and $a^2 - (x-h)^2 = a^2 - a^2 \cos^2(\theta) = a^2 \sin^2(\theta)$ and $dx = -a \sin(\theta) d\theta$. Consider the bounds:

1. if $x = h + a = h + a \cos(\theta)$ then we find $\cos(\theta) = 1$ hence $\theta = 0$ (our choice)
2. if $x = h - a = h + a \cos(\theta)$ then we find $\cos(\theta) = -1$ hence $\theta = \pi$ (our choice)

I say "our choice" because once I state these bounds I am clarifying that we made the substitution for $0 \leq \theta \leq \pi$. Other intervals are possible, however this one will suffice. We calculate,

$$A = \int_{\pi}^0 2 \frac{b}{a} \sqrt{a^2 \sin^2(\theta)} (-a \sin(\theta) d\theta)$$

Now, our choice of $0 \leq \theta \leq \pi$ indicates the sine function is non-negative and as $a > 0$ was given it follows $\sqrt{a^2 \sin^2(\theta)} = a \sin(\theta)$. Flipping the bounds to remove the minus we find,

$$A = \int_0^{\pi} 2ab \sin^2(\theta) d\theta = \int_0^{\pi} ab[1 - \cos(2\theta)] d\theta = ab \left[\theta - \frac{1}{2} \sin(2\theta) \right] \Big|_0^{\pi} = \boxed{\pi ab.}$$

In the case $a = b$ we usually say $a = r$ and we find $A = \pi r^2$. Only in that special case is the "angle" θ technically an angle in the geometric sense. For $a \neq b$ the quantity θ is simply a parameter which does not match the polar angle as measured from the center of the ellipse.

There are more exotic trigonometric substitutions. For example, there is a well-known substitution that allows one to integrate a fraction of sines and cosines: $\int \frac{a \cos(x) + b \sin(x) + c}{d \cos(x) + e \sin(x) + g} dx$. The fact that this is solvable means that there are corresponding algebraic functions which become solvable under a substitution which brings the integral to the fraction of sines and cosines. I may assign such a problem in homework as a challenge.

9.4 partial fractions

If we can find a formula in terms of finitely many elementary functions for $\int f(x) dx$ then I say the integral is *solvable*. In contrast, I would say f is integrable if there exists an antiderivative function $F'(x) = f(x)$ for all $x \in \text{dom}(f)$. This criteria of *integrable* is much weaker than solvable. You should recall that the FTC part I proves that any continuous function is integrable. However, the FTC just provides the existence of the antiderivative, it does not explain how to find its formula. While it is not at all an easy thing to prove, there are many continuous functions which have insolvable integrals. For example, $\int e^{-x^2} dx$ or $\int \frac{\sin(x)}{x} dx$ cannot be solved in terms of elementary functions alone³. Given all of this, you might be surprised by the following claim:

THE INTEGRAL OF ANY RATIONAL FUNCTION IS SOLVABLE.

I will attempt to illustrate a partial proof of this claim in this section. After many examples I'll outline the general procedure which proves the claim above.

Example 9.4.1. The calculation of $\int \frac{dx}{x^2+6x+5}$ is greatly simplified by the algebra below:

$$\begin{aligned} \frac{1}{x^2+6x+5} &= \frac{1}{(x+1)(x+5)} = \frac{A}{x+1} + \frac{B}{x+5} \\ \Rightarrow 1 &= A(x+5) + B(x+1) \\ \text{evaluating at } x = -1 &\Rightarrow 1 = -4B \Rightarrow B = -1/4 \\ \text{evaluating at } x = -5 &\Rightarrow 1 = 4A \Rightarrow A = 1/4 \\ \Rightarrow \frac{1}{x^2+6x+5} &= \frac{1/4}{x+1} - \frac{1/4}{x+5} \end{aligned}$$

We say the algebraic technique above is performing the partial fractal decomposition of the rational function. In other words, I take the given rational function and try to write it as a sum of more basic rational functions which happen to have nice integrals. Observe that

$$\int \frac{dx}{x^2+6x+5} = \frac{1}{4} \int \frac{dx}{x+1} - \frac{1}{4} \int \frac{dx}{x+5}$$

Therefore,

$$\int \frac{dx}{x^2+6x+5} = \frac{1}{4} \ln|x+1| - \frac{1}{4} \ln|x+5| + c.$$

Remark 9.4.2.

Why is it ok to plug in $x = -1$ and $x = -5$? Shouldn't this lead to an error since the integrand is not even defined at those values? But, there is no error, the algebra is correct, why?

A rational function f is by definition the quotient of two polynomials p and q ; f is rational iff there exist polynomials p, q such that $f = p/q$. Moreover, we say $f = \frac{p}{q}$ is proper iff $\deg(p) < \deg(q)$. On the other hand, we say $f = \frac{p}{q}$ is improper iff $\deg(p) \geq \deg(q)$. In the case f is improper we can always find a proper rational function r and a polynomial g such that $f = \frac{p}{q} = g + \frac{r}{q}$. The polynomial r is sometimes called the remainder and it is calculated in general by long-division of polynomials. If you can do long-division for

³there are techniques to find approximations for the antiderivatives, this is one of the reasons we cover power series expansions later in this course.

integers then you should be able to do it for polynomials. I do not use synthetic division because it is neither necessary or general. The algorithm and logic for polynomial long-division is essentially the same as the process we were taught as children. We systematically subtract multiples of the divisor from the numerator until we can no longer subtract whole copy from what remains. For example,

$$\begin{array}{r}
 2x^2 - 5x + 18 \\
 x+3 \overline{) 2x^3 + x^2 + 3x + 5} \\
 \underline{-2x^3 - 6x^2} \\
 -5x^2 + 3x \\
 \underline{5x^2 + 15x} \\
 18x + 5 \\
 \underline{-18x - 54} \\
 -49
 \end{array}$$

The example below makes use of the division above.

Example 9.4.3.

$$\int \frac{2x^3 + x^2 + 3x + 5}{x+3} dx = \int \left(2x^2 - 5x + 18 - \frac{49}{x+3} \right) dx = \boxed{\frac{2}{3}x^3 - \frac{5}{2}x^2 + 18x - 49 \ln |x+3| + c.}$$

We can have higher degree polynomials in the denominator and still perform long-division.

$$\begin{array}{r}
 2x - 7 \\
 x^2 + 4x + 3 \overline{) 2x^3 + x^2 + 3x + 5} \\
 \underline{-2x^3 - 8x^2 - 6x} \\
 -7x^2 - 3x + 5 \\
 \underline{7x^2 + 28x + 21} \\
 25x + 26
 \end{array}$$

Again, I use this long-division to begin the next example.

Example 9.4.4.

$$\int \frac{2x^3 + x^2 + 3x + 5}{x^2 + 4x + 3} dx = \int \left(2x - 7 + \frac{25x + 26}{x^2 + 4x + 3} \right) dx = x^2 - 7x + \int \frac{25x + 26}{x^2 + 4x + 3} dx.$$

Calculation of the remaining integral is accomplished with the partial fractions idea once more:

$$\frac{25x + 26}{x^2 + 4x + 3} = \frac{25x + 26}{(x+1)(x+3)} = \frac{A}{x+1} + \frac{B}{x+3}$$

$$\Rightarrow 25x + 26 = A(x+3) + B(x+1)$$

$$\text{evaluating at } x = -3 \Rightarrow -75 + 26 = -2B \Rightarrow B = 49/2$$

$$\text{evaluating at } x = -1 \Rightarrow -25 + 26 = 2A \Rightarrow A = 1/2$$

$$\Rightarrow \frac{25x + 26}{x^2 + 4x + 3} = \frac{1/2}{x+1} + \frac{49/2}{x+3}$$

The denominator factors into three linear factors $x^3 + 4x^2 + 3x = x(x+1)(x+3)$. This factoring is what leads us to the decomposition beneath:

$$\frac{113x^2 + 116x}{x^3 + 4x^2 + 3x} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{x+3}$$

Multiply by the denominator, nice cancellations occur:

$$113x^2 + 116x = A(x+1)(x+3) + Bx(x+3) + Cx(x+1)$$

Now, since all three linear factors correspond to distinct roots we can plug in the roots to obtain solutions for A, B, C , if we had repeated roots then we'll see this system breaks down, but let's finish this example for now,

$$x = 0 \Rightarrow 0 = 3A$$

$$x = -1 \Rightarrow 113 - 116 = -2B$$

$$x = -3 \Rightarrow 113(9) - 3(116) = 6C$$

Hence, $A = 0$, $B = 3/2$ and $C = 223/2$ (the fact that $A = 0$ was obvious from the outset)

$$\begin{aligned} \int \frac{x^6 + 2x^3 + 2x + 7}{x^3 + 4x^2 + 3x} dx &= \frac{1}{4}x^4 + \frac{4}{3}x^3 + \frac{13}{2}x^2 - 38x + \int \frac{113x^2 + 116x}{x^3 + 4x^2 + 3x} dx \\ &= \frac{1}{4}x^4 + \frac{4}{3}x^3 + \frac{13}{2}x^2 - 38x + \frac{3}{2} \int \frac{dx}{x+1} + \frac{223}{2} \int \frac{dx}{x+3} \\ &= \boxed{\frac{1}{4}x^4 + \frac{4}{3}x^3 + \frac{13}{2}x^2 - 38x + \frac{3}{2} \ln|x+1| + \frac{223}{2} \ln|x+3| + c.} \end{aligned}$$

Ok, I think we've seen enough long-division examples for this section. Keep in mind that we could add that wrinkle to the examples that follow, but it just adds one extra step at the beginning of the calculation.

Example 9.4.8. We wish to calculate the integral below:

$$\int \frac{2x^2 + 3x - 9}{x^4 + 4x^3 + 5x^2 + 9x^2 + 36x + 45} dx.$$

This is a proper rational function so we don't have to do any long division. Our first task is to factor the denominator. Note $x^4 + 4x^3 + 5x^2 + 9x^2 + 36x + 45 = x^2(x^2 + 4x + 5) + 9(x^2 + 4x + 5)$ hence

$$\int \frac{2x^2 + 3x - 9}{x^4 + 4x^3 + 5x^2 + 9x^2 + 36x + 45} dx = \int \frac{2x^2 + 3x - 9}{(x^2 + 9)(x^2 + 4x + 5)} dx.$$

I cannot factor further over the real numbers because these quadratic factors have complex roots. The partial fractions requires two unknown coefficients for each quadratic factor:

$$\frac{2x^2 + 3x - 9}{(x^2 + 9)(x^2 + 4x + 5)} = \frac{Ax + B}{x^2 + 9} + \frac{Cx + D}{x^2 + 4x + 5}$$

$$\Rightarrow 2x^2 + 3x - 9 = (Ax + B)(x^2 + 4x + 5) + (Cx + D)(x^2 + 9)$$

$$\Rightarrow 2x^2 + 3x - 9 = Ax^3 + Bx^2 + 4Ax^2 + 4Bx + 5Ax + 5B + Cx^3 + Dx^2 + 9Cx + 9D$$

$$\Rightarrow 2x^2 + 3x - 9 = x^3(A + C) + x^2(B + 4A + D) + x(4B + 5A + 9C) + 5B + 9D$$

Two polynomials are equal iff they have matching coefficients. We equate coefficients of x^3, x^2, x^1 and x^0 :

$$(i.) 0 = A + C \quad (ii.) 2 = B + 4A + D \quad (iii.) 3 = 4B + 5A + 9C \quad (iv.) -9 = 5B + 9D$$

We have four linear equations and four unknowns. We proceed by elimination: using *i.* and *ii.*,

$$A = -C, \quad A = (2 - B - D)/4, \Rightarrow (2 - B - D)/4 = -C \Rightarrow \underline{2 - B - D + 4C = 0} \quad (v.)$$

Next eliminate A from *ii.* and *iii.*

$$A = (2 - B - D)/4, \quad A = (3 - 4B - 9C)/5 \Rightarrow (2 - B - D)/4 = (3 - 4B - 9C)/5 \Rightarrow 5(2 - B - D) = 4(3 - 4B - 9C)$$

Simplifying gives $11B + 36C - 5D - 2 = 0$ (*vi.*) so equations *iv.*, *v* and *vi.* involve only the variables B, C, D .

We have reduced the problem to 3 equations and 3 unknowns:

$$5B + 9D = -9, \quad -B + 4C - D = -2, \quad 11B + 36C - 5D = 2$$

We can finish by elimination, but for a change of pace I'll use Kramer's rule from high-school algebra II, we write the system as a matrix problem:

$$\begin{bmatrix} 5 & 0 & 9 \\ -1 & 4 & -1 \\ 11 & 36 & -5 \end{bmatrix} \begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{bmatrix} -9 \\ -2 \\ 2 \end{bmatrix}$$

Then, Kramer's rule says:

$$B = \frac{\det \begin{bmatrix} -9 & 0 & 9 \\ -2 & 4 & -1 \\ 2 & 36 & -5 \end{bmatrix}}{\det \begin{bmatrix} 5 & 0 & 9 \\ -1 & 4 & -1 \\ 11 & 36 & -5 \end{bmatrix}} = \frac{-9(-20 + 36) + 9(-72 - 8)}{5(-20 + 36) + 9(-36 - 44)} = \frac{-864}{-640} = \frac{27}{20}.$$

At which point finding D is easy from *iv.*,

$$D = -1 - \frac{5}{9}B = -1 - \frac{5}{9} \cdot \frac{27}{20} = \frac{-7}{4}$$

Find C with ease from *v.* next,

$$C = \frac{1}{4}(-2 + B + D) = \frac{1}{4}(-2 + \frac{27}{20} + \frac{-7}{4}) = \frac{-3}{5}$$

Hence, returning to *i.*, clearly $A = \frac{3}{5}$. As a check on the algebra, I'll solve the problem again from the start using technology: the row reduced echelon form of the augmented coefficient matrix will reveal the solution (if it exists) of any linear system: the first row comes from $A + C = 0$ then equations *ii.*, *iii.* and *iv.* follow in their matrix notation:

$$\text{rref} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 5 & 0 & 9 & -9 \\ 0 & -1 & 4 & -1 & -2 \\ 0 & 11 & 36 & -5 & 2 \end{array} \right] = \text{rref} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 3/5 \\ 0 & 1 & 0 & 0 & 27/20 \\ 0 & 0 & 1 & 2 & -3/5 \\ 0 & 0 & 0 & 1 & -7/4 \end{array} \right]$$

Which confirms the algebra of the last page. Pragmatically speaking, if you are not taking a test and don't need to practice your algebra then the reduced row echelon form (a.k.a. Gaussian elimination) is hard to beat. We examine the logical underpinnings of both Kramers rule and Gaussian elimination in the linear algebra course⁴ (Math 321). Let's collect our thoughts. We have shown that

$$\frac{2x^2 + 3x - 9}{x^4 + 4x^3 + 5x^2 + 9x^2 + 36x + 45} = \frac{\frac{3}{5}x + \frac{27}{20}}{x^2 + 9} + \frac{\frac{-3}{5}x - \frac{7}{4}}{x^2 + 4x + 5}$$

Observe we can integrate using u -substitution on the A -term and trig. substitution on the B -term⁵

$$\begin{aligned} \int \frac{\frac{3}{5}x + \frac{27}{20}}{x^2 + 9} dx &= \frac{3}{10} \int \frac{2xdx}{x^2 + 9} + \frac{27}{20} \int \frac{dx}{x^2 + 9} \\ &= \frac{3}{10} \int \frac{du}{u} + \frac{27}{20} \int \frac{3 \sec^2(\theta) d\theta}{9 \sec^2(\theta)} \\ &= \frac{3}{10} \ln|u| + \frac{9}{20} \theta + c_1 \quad (\text{where } u = x^2 + 9 \text{ and } x = 3 \tan(\theta)) \\ &= \frac{3}{10} \ln|x^2 + 9| + \frac{9}{20} \tan^{-1}\left(\frac{x}{3}\right) + c_1 \end{aligned} \quad \star$$

Next, turn to the C and D terms, a little creative algebra from the outset makes life easier,

$$\begin{aligned} \int \frac{\frac{-3}{5}x - \frac{7}{4}}{x^2 + 4x + 5} dx &= \int \frac{\frac{-3}{5}(x+2) + \frac{6}{5} - \frac{7}{4}}{x^2 + 4x + 5} dx \\ &= \frac{-3}{5} \int \frac{(x+2)dx}{(x+2)^2 + 1} - \frac{11}{20} \int \frac{dx}{(x+2)^2 + 1} \\ &= \frac{-3}{10} \int \frac{du}{u} - \frac{11}{20} \int \frac{dw}{w^2 + 1} \\ &= \frac{-3}{10} \ln|u| - \frac{11}{20} \tan^{-1}(w) + c_2 \quad (\text{where } u = (x+2)^2 + 1 \text{ and } w = x+2) \\ &= \frac{-3}{10} \ln|x^2 + 4x + 5| - \frac{11}{20} \tan^{-1}(x+2) + c_2 \end{aligned} \quad \star\star$$

Therefore, using \star and $\star\star$ we find $\int \frac{2x^2+3x-9}{x^4+4x^3+5x^2+9x^2+36x+45} dx$ is

$$\boxed{\frac{3}{10} \ln|x^2 + 9| + \frac{-3}{10} \ln|x^2 + 4x + 5| - \frac{11}{20} \tan^{-1}(x+2) + \frac{9}{20} \tan^{-1}\left(\frac{x}{3}\right) + c.}$$

At this point, we have seen examples of how to integrate rational functions with denominators that consist of a linear factor, a pair of linear factors, a triple of linear factors or a pair irreducible quadratic factors. I know consider an example where we have a mixture of linear and irreducible quadratic terms. Again, the heart of the solution is the partial fractal decomposition.

⁴however, linear algebra is about a lot more than just solving linear equations, you are really supposed to know what I've shown here from highschool. Granted many of you have forgotten. Now is time to remember. As you work the homework you should not use technology to solve all the algebra, instead, use technology to check your answers. I have links on my website to free online calculators which will do partial fractions and/or the rref calculation. Make use of them wisely.

⁵I do not expect you to memorize these formulas, I expect you to be able to derive them!

Example 9.4.9. We wish to calculate $\int \frac{3x^2-3x-8}{x^3-3x^2+x-3} dx$. Note that $x^3-3x^2+x-3 = (x-3)(x^2+1)$ hence we propose the following decomposition:

$$\frac{3x^2-3x-8}{x^3-3x^2+x-3} = \frac{A}{x-3} + \frac{Bx+C}{x^2+1}$$

Multiply by the denominator, nice cancellations occur:

$$3x^2-3x-8 = A(x^2+1) + (Bx+C)(x-3)$$

I'll plug in the real root and one of the complex roots⁶ to obtain:

$$x=3 \Rightarrow 3(9)-3(3)-8=10A \Rightarrow A=1.$$

$$x=i=\sqrt{-1} \Rightarrow 3i^2-3i-8=(iB+C)(i-3) \Rightarrow -11-3i=-B-3C+i(C-3B)$$

The beautiful thing about complex equations is they do twice the work of real equations. This is due to the fact that the real and imaginary parts of a complex equation must separately balance. In particular,

$$-11-3i=-B-3C+i(C-3B) \Rightarrow -11=-B-3C \text{ and } -3=C-3B$$

Multiply $-11=-B-3C$ by -3 to find $33=3B+9C$ then add this to $-3=C-3B$ to obtain $30=10C$ hence $C=3$ and it follows $B=11-3C=2$. In summary, $A=1$, $B=2$ and $C=3$. Finish it.

$$\begin{aligned} \int \frac{3x^2-3x-8}{x^3-3x^2+x-3} dx &= \int \frac{dx}{x-3} + \int \frac{2xdx}{x^2+1} + 3 \int \frac{dx}{x^2+1} \\ &= \boxed{\ln|x-3| + \ln|x^2+1| + \tan^{-1}(x) + c.} \end{aligned}$$

Example 9.4.10. We wish to calculate $\int \frac{2}{x^4-1} dx$. Well, this is easy to factor: $x^4-1 = (x^2+1)(x^2-1)$. The partial fractions decomposition with respect to this factoring is

$$\frac{2}{x^4-1} = \frac{Ax+B}{x^2+1} + \frac{Cx+D}{x^2-1}$$

Multiply by the denominator, nice cancellations occur:

$$2 = (Ax+B)(x^2-1) + (Cx+D)(x^2+1)$$

I'll plug in the real roots and one of the complex roots to obtain:

$$x=1 \Rightarrow 2=2C+2D.$$

$$x=-1 \Rightarrow 2=-2C+2D.$$

$$x=i=\sqrt{-1} \Rightarrow 2=(Ai+B)(i^2-1) \Rightarrow 2=-2B-2Ai$$

The real and imaginary parts of a complex equation must separately balance. In particular,

$$2=-2B-2Ai \Rightarrow 2=-2B \text{ and } 0=-2A$$

Therefore, $A=0$ and $B=-1$. Adding and subtracting the $x=\pm 1$ equations yields $C=0$ and $D=1$.

$$\boxed{\int \frac{2}{x^4-1} dx = \int \frac{dx}{x^2-1} - \int \frac{dx}{x^2+1} = -\tanh^{-1}(x) - \tan^{-1}(x) + c.}$$

⁶yes, you could also multiply the polynomials out and equate coefficients as we did in a previous example.

In the preceding example I made use of the inverse hyperbolic tangent integral. If you don't like the inverse hyperbolic function then you could just as well break the $1/(x^2 - 1)$ term into a sum of reciprocals of $(x - 1)$ and $(x + 1)$. You can show that

$$\frac{1}{x^2 - 1} = \frac{1}{2} \left[\frac{1}{x - 1} - \frac{1}{x + 1} \right].$$

Or, you could at least check that I am correct by making a common denominator of the RHS to recover the LHS. In any event, integration will yield natural logs from this algebraic approach:

$$\begin{aligned} \int \frac{dx}{x^2 - 1} &= \frac{1}{2} \int \frac{dx}{x - 1} - \frac{1}{2} \int \frac{dx}{x + 1} \\ &= \frac{1}{2} \ln|x - 1| - \frac{1}{2} \ln|x + 1| + c. \end{aligned}$$

There are identities which connect this expression and $-\tanh^{-1}(x)$. If nothing else, we know they have to be different formulas for the same function since

1. $\frac{d}{dx}[-\tanh^{-1}(x)] = \frac{d}{dx}[\frac{1}{2} \ln|x - 1| - \frac{1}{2} \ln|x + 1|] = \frac{1}{x^2 - 1}$
2. $\tanh^{-1}(0) = 0$ and $\frac{1}{2} \ln|0 - 1| - \frac{1}{2} \ln|0 + 1| = 0$

Recall that if $f'(x) = g'(x)$ for all $x \in J$ (a connected interval) then $f(x) = g(x) + c$ on J . If we also know $f(0) = g(0)$ then it follows $c = 0$ hence $f = g$, provided the domain of both functions is J . Now perhaps my argument is less than satisfying, after all I didn't really explain how these rather different looking formulas are the same. See Stewart around page 466 for more on how to connect these algebraically.

Example 9.4.11. We wish to calculate $\int \frac{x^2 + 5}{x^4 + 8x^2 + 16} dx$. This integral contains a new difficulty, it has a irreducible quadratic repeated in the denominator since $x^4 + 8x^2 + 16 = (x^2 + 4)^2$. The partial fractions decomposition for a repeated quadratic is as follows:

$$\frac{x^2 + 5}{x^4 + 8x^2 + 16} = \frac{Ax + B}{x^2 + 4} + \frac{Cx + D}{(x^2 + 4)^2}.$$

I invite the reader to verify that the usual algebra yields $A = 0, B = 1, C = 0$ and $D = 1$. I know this will happen since the following algebra leads to the same conclusion:

$$\frac{x^2 + 5}{x^4 + 8x^2 + 16} = \frac{x^2 + 4 + 1}{(x^2 + 4)^2} = \frac{1}{x^2 + 4} + \frac{1}{(x^2 + 4)^2}$$

If $x = 2 \tan(\theta)$ then $x^2 + 4 = 4 \tan^2(\theta) + 4 = 4 \sec^2(\theta)$ and $dx = 2 \sec^2(\theta) d\theta$. We will make use of this substitution in the integral below:

$$\begin{aligned} \int \frac{x^2 + 5}{x^4 + 8x^2 + 16} dx &= \int \left[\frac{1}{x^2 + 4} + \frac{1}{(x^2 + 4)^2} \right] dx \\ &= \int \left[\frac{1}{4 \sec^2(\theta)} + \frac{1}{(4 \sec^2(\theta))^2} \right] (2 \sec^2(\theta) d\theta) \\ &= \int \left[\frac{1}{2} + \frac{1}{8} \cos^2(\theta) \right] d\theta \\ &= \int \left[\frac{1}{2} + \frac{1}{16} + \frac{1}{16} \cos(2\theta) \right] d\theta \\ &= \frac{9\theta}{16} + \frac{1}{32} \sin(2\theta) + c \end{aligned}$$

Recall that $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ and note $\tan(\theta) = \frac{x}{2}$ hence $\sin(\theta) = \frac{x}{\sqrt{x^2+4}}$ and $\cos(\theta) = \frac{2}{\sqrt{x^2+4}}$ thus, $\sin(2\theta) = \frac{4x}{x^2+4}$. Therefore,

$$\int \frac{x^2 + 5}{x^4 + 8x^2 + 16} dx = \frac{9}{16} \tan^{-1}\left(\frac{x}{2}\right) + \frac{x}{8(x^2 + 4)} + c.$$

Example 9.4.12. We wish to calculate $\int \frac{2x^3+3x^2-x+1}{(x-2)(x+1)^3} dx$. To begin we need to work out the partial fractions decomposition. Given the factoring of the denominator we propose:

$$\frac{2x^3 + 3x^2 - x + 1}{(x-2)(x+1)^3} = \frac{A}{x-2} + \frac{B}{x+1} + \frac{C}{(x+1)^2} + \frac{D}{(x+1)^3}$$

Multiply by the denominator to find:

$$2x^3 + 3x^2 - x + 1 = A(x+1)^3 + B(x-2)(x+1)^2 + C(x-2)(x+1) + D(x-2)$$

We can evaluate at $x = -1$ to find that $-2 + 3 + 1 + 1 = -3D$ hence $\boxed{D = -1}$. Likewise, evaluate at $x = 2$ to obtain that $16 + 12 - 2 + 1 = 27A$ hence $\boxed{A = 1}$. To find B and C we need additional equations. One easy choice is to evaluate at $x = 0$ which gives $1 = A - 2B - 2C - 2D = 1 - 2B - 2C + 2$ hence $\underline{-2 = -2B - 2C}$. Next, I evaluate⁷ at $x = 1$ to obtain $2 + 3 - 1 + 1 = 8A - 4B - 2C - D$ hence $5 = 8 - 4B - 2C + 1$ thus $4 = 4B + 2C$ and dividing by two gives $\underline{2 = 2B + C}$. Collecting our thoughts,

$$-2 = -2B - 2C \quad \text{added to} \quad 2 = 2B + C \quad \Rightarrow \quad 0 = -C \quad \Rightarrow \quad \boxed{C = 0} \quad \text{and} \quad \boxed{B = 1}.$$

Therefore,

$$\begin{aligned} \int \frac{2x^3 + 3x^2 - x + 1}{(x-2)(x+1)^3} dx &= \int \left[\frac{1}{x-2} + \frac{1}{x+1} - \frac{1}{(x+1)^3} \right] dx \\ &= \int \frac{du}{u} + \int \frac{dw}{w} - \int \frac{dw}{w^3} \quad \text{let } u = x-2 \text{ and } w = x+1 \\ &= \boxed{\ln|x-2| + \ln|x+1| + \frac{1}{2(x+1)^2} + c.} \end{aligned}$$

Example 9.4.13. We wish to calculate $\int \frac{dx}{(x^2+6x+13)^3}$. Note that $x^2 + 6x + 13 = (x+3)^2 + 4$ hence the denominator cannot be further reduced. This is already a basic rational function. We make a $x+3 = 2\tan(\theta)$ substitution hence $(x+3)^2 + 4 = 4\sec^2(\theta)$ and $dx = 2\sec^2(\theta)d\theta$. Observe

$$\begin{aligned} \int \frac{dx}{(x^2 + 6x + 13)^3} &= \int \frac{2\sec^2(\theta)d\theta}{64\sec^6(\theta)} \\ &= \frac{1}{32} \int \cos^4(\theta)d\theta \quad (\text{the next step is nontrivial}) \\ &= \frac{1}{32} \left[\frac{3\theta}{8} + \frac{1}{8} \sin(\theta) \cos^3(\theta) - \frac{1}{8} \sin^3(\theta) \cos(\theta) + \frac{1}{2} \sin(\theta) \cos(\theta) \right] + c \end{aligned}$$

⁷because the arithmetic is easy and multiplying out to equate coefficients looks like more work here

We know $\tan(\theta) = \frac{x+3}{2}$ hence $\sin(\theta) = \frac{x+3}{\sqrt{(x+3)^2+4}}$ and $\cos(\theta) = \frac{2}{\sqrt{(x+3)^2+4}}$. Thus,

$$\begin{aligned} \int \frac{dx}{(x^2+6x+13)^3} &= \frac{1}{32} \left[\frac{3\theta}{8} + \frac{1}{8} \sin(\theta) \cos^3(\theta) - \frac{1}{8} \sin^3(\theta) \cos(\theta) + \frac{1}{2} \sin(\theta) \cos(\theta) \right] + c \\ &= \frac{1}{32} \left(\frac{3}{8} \tan^{-1} \left(\frac{x+3}{2} \right) \right. \\ &\quad + \frac{1}{8} \frac{x+3}{\sqrt{(x+3)^2+4}} \left[\frac{2}{\sqrt{(x+3)^2+4}} \right]^3 \\ &\quad - \frac{1}{8} \left[\frac{x+3}{\sqrt{(x+3)^2+4}} \right]^3 \frac{2}{\sqrt{(x+3)^2+4}} \\ &\quad \left. + \frac{1}{2} \frac{x+3}{\sqrt{(x+3)^2+4}} \frac{2}{\sqrt{(x+3)^2+4}} \right) + c \end{aligned}$$

After a some algebra the expressions above ought to simply to the following:

$$\int \frac{dx}{(x^2+6x+13)^3} = \frac{282+202x+54x^2+6x^3}{(13+6x+x^2)^2} + \frac{3}{256} \tan^{-1} \left(\frac{3+x}{2} \right) + c.$$

Remark 9.4.14.

I think we've seen enough. At this point we can make a few observations:

1. when integrating a rational function we can always reduce the problem to integrating a proper rational function and a polynomial because we know long division for polynomials.
2. any proper rational function can be decomposed into a sum of basic rational functions which either have the form $\frac{1}{(x-a)^k}$ or $\frac{1}{[(x-\alpha)^2+\beta^2]^k}$ for $k = 1, 2, \dots$
3. we can integrate any basic rational function, we have the methods, we just need paper and perservance.
4. the partial fractions decomposition of a proper rational function with denominator of degree N will have N -unknown coefficients in its decomposition.

I seek to sketch a proof of (2.) and (4.) in the section that follows. The proofs of (1.) and (3.) are left to the reader. .

Beyond these comments, we also observed that the inverse hyperbolic functions allow us to treat reducible reciprocal quadratic factors in much the same way as we treat the irreducible reciprocals. This led to a great diversity in the appearance of answers.

9.4.1 existence of partial fractions decomposition

Remark 9.4.15.

Read if interested. Ask me questions if want further detail. Pragmatically speaking, if you understand the examples before this subsection you should do just fine. This section does contain useful comments about algebra in general and a few complex calculations that all electrical engineers should be eager to master.

Suppose we have a rational function $f(x) = \frac{P(x)}{Q(x)}$. Suppose further that $\deg(Q) = N$ and $\deg(P) = n$ and $n < N$ so this rational function is proper. A fundamental theorem of algebra states that any polynomial with real coefficients can be written as a product of linear factors which may be repeated and/or correspond to complex roots. However, the complex roots must come in complex conjugate pairs which once multiplied give irreducible quadratic factors. That said, I will give an argument which allows complex notation. We know that Q factors as

$$Q(x) = A(x - a_1)^{k_1}(x - a_2)^{k_2} \cdots (x - a_r)^{k_r}$$

where $A \neq 0$ and $a_j \in \mathbb{C}$ and $k_1 + k_2 + \cdots + k_r = N$. We claim that f can be written as a sum of the reciprocals $(x - a_1)^{k_1}, (x - a_2)^{k_2}, \dots, (x - a_r)^{k_r}$. That is, we claim that the following equation has a solution for some $A_1, A_2, \dots, A_r \in \mathbb{C}$,

$$f(x) = \frac{P(x)}{Q(x)} = \frac{A_1}{x - a_1} + \frac{A_2}{(x - a_1)^2} + \cdots + \frac{A_{k_1}}{(x - a_1)^{k_1}} + \cdots + \frac{A_{N-k_r}}{x - a_r} + \frac{A_{N-k_r+1}}{(x - a_r)^2} + \cdots + \frac{A_N}{(x - a_r)^{k_r}}$$

Multiply both sides by $Q(x)$ to obtain \star :

$$\begin{aligned} P(x) &= A_1[(x - a_1)^{k_1-1}(x - a_2)^{k_2} \cdots (x - a_r)^{k_r}] + \cdots \\ &\quad \cdots + A_{k_1}[(x - a_2)^{k_2} \cdots (x - a_r)^{k_r}] + \cdots \\ &\quad \cdots + A_{N-k_r}[(x - a_1)^{k_1-1}(x - a_2)^{k_2} \cdots (x - a_r)^{k_r-1}] + \cdots \\ &\quad \cdots + A_N[(x - a_1)^{k_1-1}(x - a_2)^{k_2} \cdots (x - a_{r-1})^{k_{r-1}}] \end{aligned}$$

Observe that if we plug in $x = a_1$ into \star then we find

$$P(a_1) = A_{k_1}(a_1 - a_2)^{k_2} \cdots (a_1 - a_r)^{k_r} \quad \Rightarrow \quad A_{k_1} = \frac{P(a_1)}{(a_1 - a_2)^{k_2} \cdots (a_1 - a_r)^{k_r}}$$

Next, we can plug in $x = a_2$ into \star and all terms drop to zero except one

$$P(a_2) = A_{k_2}(a_2 - a_1)^{k_1}(a_2 - a_3)^{k_3} \cdots (a_2 - a_r)^{k_r} \quad \Rightarrow \quad A_{k_2} = \frac{P(a_2)}{(a_2 - a_1)^{k_1}(a_2 - a_3)^{k_3} \cdots (a_2 - a_r)^{k_r}}$$

Continue in this fashion and we find solutions for A_{k_j} for $j = 3, 4, \dots, r$. However, this is only a subset of the coefficients since we allow $k_j > 1$ since we want to treat the case of repeated roots. Differentiate \star and call it $\frac{d\star}{dx}$. Plug $x = a_1$ into $\frac{d\star}{dx}$ and note that just one term remains. The derivative of the term with coefficient A_{k_1-1} produces many terms, however only one remains after the evaluation. The term that remains is the one which $(x - a_1)$ was differentiated hence

$$P'(a_1) = A_{k_1-1}(a_1 - a_2)^{k_2} \cdots (a_1 - a_r)^{k_r} \quad \Rightarrow \quad A_{k_1-1} = \frac{P'(a_1)}{(a_1 - a_2)^{k_2} \cdots (a_1 - a_r)^{k_r}}$$

Then, if we differentiate twice then evaluate more we find a formula for A_{k_1-2} . Continuing in this fashion we find formulas for A_1, A_2, \dots, A_{k_1} . To complete the argument simply apply the procedure for the first root to

the remaining roots a_2, a_3, \dots, a_r . Finally, note in the case of complex roots we must have a conjugate pair. Observe that for such a pair we can reproduce the form we've used in this section:

$$\frac{A}{x - \alpha - i\beta} + \frac{B}{x - \alpha + i\beta} = \frac{A(x - \alpha + i\beta) + B(x - \alpha - i\beta)}{(x - \alpha - i\beta)(x - \alpha + i\beta)} = \frac{(A + B)x - \alpha(A + B) + i\beta(A - B)}{(x - \alpha)^2 + \beta^2}$$

This is precisely the form we used, just written with ugly coefficients. You might worry that the rational function is complex so that will not match the manifestly real partial fractions decomposition. However, the fact that $f = P/Q$ is real implies certain reality conditions on A and B hence the imaginary term will vanish. Rather than work this out in general let me illustrate how the real and complex form of the partial fractions work out in a specific case:

$$\frac{2x + 6}{x^2 + 4x + 5} = \frac{x + 3}{(x + 2)^2 - i^2} = \frac{A}{x + 2 - i} + \frac{B}{x + 2 + i}$$

Multiply both sides by the denominator to find $2x + 6 = A(x + 2 + i) + B(x + 2 - i)$. Thus,

$$2x + 6 = (A + B)x + 2(A + B) + i(A - B)$$

Equating coefficients gives $2 = A + B$ and $6 = 2(A + B) + i(A - B)$ which yields $6 = 4 + i(A - B)$. It follows that $2 = i(A - B)$ thus $A - B = 2/i = -2i$. Add $A - B = -2i$ and $A + B = 2$ to find $2A = 2 - 2i$ thus $A = 1 - i$. Likewise, subtract $A - B = -2i$ from $A + B = 2$ to find $2B = 2 + 2i$ thus $B = 1 + i$. We have derived the complex partial fractions decomposition

$$\frac{2x + 6}{x^2 + 4x + 5} = \frac{1 - i}{x + 2 - i} + \frac{1 + i}{x + 2 + i}.$$

In general, one can prove that if $A + B \in \mathbb{R}$ and $A - B \in i\mathbb{R}$ and it follows that expression on the far right below is manifestly real once the imaginary terms are cancelled

$$\frac{A}{x - \alpha - i\beta} + \frac{B}{x - \alpha + i\beta} = \frac{(A + B)x - \alpha(A + B) + i\beta(A - B)}{(x - \alpha)^2 + \beta^2}.$$

In any event, I include these general comments for those students who wonder about why this method works in general. The algebraic technique of partial fractions extends beyond calculus, but it is seldom proved in any course. It is doubtful I lecture on this stretch of the notes.

Lemma 9.4.16. *equating coefficients.*

Suppose F, G and Q are continuous functions and $Q(x) = 0$ has finitely many solutions. If $\frac{F(x)}{Q(x)} = \frac{G(x)}{Q(x)}$ for all $x \in \mathbb{R}$ such that $Q(x) \neq 0$ then $F(x) = G(x)$.

Proof: Note that if $x \in \mathbb{R}$ such that $Q(x) \neq 0$ then it is clear that $F(x) = G(x)$. Suppose $x_o \in \mathbb{R}$ such that $Q(x_o) = 0$ and consider that $F(x) = G(x)$ for all $x \in B_\delta(x_o)_o$ (a deleted nbhd of radius $\delta > 0$ must exist on which $Q(x) \neq 0$ throughout since the zeros of Q are finite in number so we can separate them) but then

$$\lim_{x \rightarrow x_o} F(x) = \lim_{x \rightarrow x_o} G(x)$$

since the limit considers only points near the limit point. But, F and G are continuous hence

$$\lim_{x \rightarrow x_o} G(x) = G(x_o) \quad \text{and} \quad \lim_{x \rightarrow x_o} F(x) = F(x_o).$$

Therefore, $F(x_o) = G(x_o)$ and as this was an arbitrary zero of Q the lemma follows. \square

This was \approx exercise 69 of page 519 of Stewart's Ed. 6, however I proved it in more generality. Note that polynomials have finitely many zeros so they are covered by the lemma.

Chapter 10

Numerical Integration

In this chapter we examine a few basic numerical techniques to approximate a definite integral. You may recall some of this from Calculus I where we discussed the left, right and midpoint rules to motivate the definition of the Riemann integral.

10.1 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.

Definition 10.1.1. *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, $x_0 = a$ and $x_n = b$.

The closed interval $[a, b]$ is a union of n -subintervals of length Δx . Note that the closed interval $[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup \dots \cup [x_{n-1}, x_n]$. The following rule is an intuitively obvious way to calculate the signed-area.

Definition 10.1.2. *left endpoint rule (L_n).*

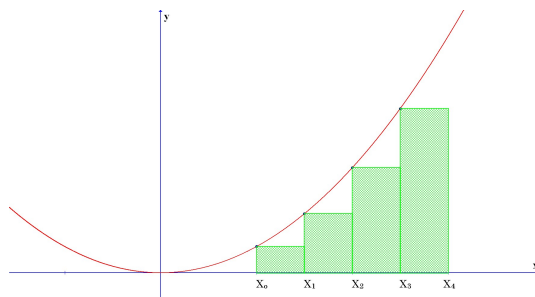
Suppose that $[a, b] \subseteq \text{dom}(f)$ then we define

$$L_n = \sum_{j=0}^{n-1} f(x_j)\Delta x = [f(x_0) + f(x_1) + \dots + f(x_{n-1})]\Delta x.$$

Example 10.1.3. 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 $x_0 = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ 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 10.1.4. *right endpoint rule (R_n).*

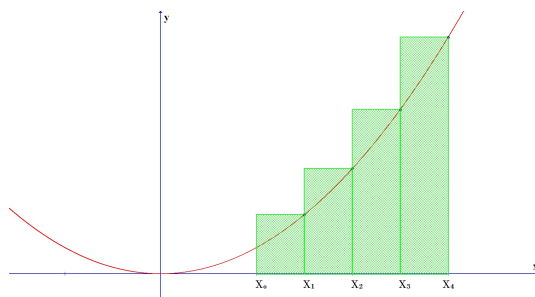
Suppose that $[a, b] \subseteq \text{dom}(f)$ then we define

$$R_n = \sum_{j=1}^n f(x_j) \Delta x = [f(x_1) + f(x_2) + \cdots + f(x_n)] \Delta x.$$

Example 10.1.5. 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 $x_0 = 1, x_1 = 1.5, x_2 = 2, x_3 = 2.5$ 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 10.1.6. *midpoint rule (M_n).*

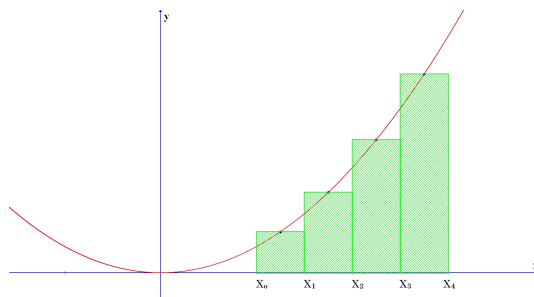
Suppose that $[a, b] \subseteq \text{dom}(f)$ and denote the midpoints by $\bar{x}_k = \frac{1}{2}(x_k + x_{k-1})$ and define

$$M_n = \sum_{j=1}^n f(\bar{x}_j) \Delta x = [f(\bar{x}_1) + f(\bar{x}_2) + \cdots + f(\bar{x}_n)] \Delta x.$$

Example 10.1.7. 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 $n \rightarrow \infty$ 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 $n \rightarrow \infty$. Careful proof of this seemingly obvious claim is beyond the scope of this course. In any event, we did prove the FTC part II in calculus I and as such we can calculate the area exactly with ease for this function:

Example 10.1.8.

$$\int_1^3 x^2 dx = \left. \frac{1}{3}x^3 \right|_1^3 = \frac{1}{3}(27 - 1) = \frac{26}{3}$$

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. However, I do hope you know terms such as increasing, decreasing, concave up/down and are able to discuss the relative errors in these methods for a specific function qualitatively, at least you should be able to rank the rules based on a geometric understanding of how the approximation schemes are constructed. Stewart discusses such issues in §8.7. Qualitatively, if the function is monotonic then we should expect that the area is bounded between L_n and R_n .

Remark 10.1.9.

You might ask why in the world we would calculate L_n , R_n or M_n when we could just use the FTC part II? Isn't it easier to evaluate $F(b) - F(a)$ rather than adding all those terms? Well, yes, of course it's easier. However, it's hard to use the FTC when you can't find a formula for the antiderivative. For example, $\int_0^{10} e^{-x^2} dx$ or $\int_0^{10} \sqrt{1+x^3} dx$.

Example 10.1.10. Calculate M_5 to approximate $\int_0^{10} e^{-x^2} dx$. Note $\delta x = 2$ and $\bar{x}_j = 1, 3, 5, 7, 9$ for $j = 1, 2, 3, 4, 5$.

$$M_5 = 2(e^{-1} + e^{-9} + e^{-25} + e^{-49} + e^{-81}) \approx 0.7360057$$

In contrast, the numerical integrator in Wolfram Alpha returns:

$$\int_0^{10} e^{-x^2} \approx 0.8862269$$

If you study the graph $y = e^{-x^2}$ the fact that M_5 is an underestimate is not surprising. Next, I'll calculate M_{10} , note $\bar{x}_j = j - 0.5$ for $j = 1, 2, \dots, 10$ and $\delta x = 1$.

$$M_{10} = \sum_{j=1}^{10} e^{-(j-0.5)^2} \approx 0.8861352$$

Now, we agree with Wolfram Alpha to the first 3 decimals. Calculate M_{100} , this makes $\delta x = 0.1$ and $\bar{x}_j = 0.1j - 0.05$ for $j = 1, 2, \dots, 100$

$$M_{100} = \sum_{j=1}^{100} e^{-(0.1j-0.05)^2} \approx 0.8862269.$$

Obviously I used technology to compute this sum. I expect you will do the same in your homework. The methods used to compute integrals numerically by mathematical software are most likely built off the Simpson's rule which we consider at the end of this chapter.

10.2 trapezoid rule

The idea of the trapezoid rule is simple enough, approximate a signed area by replacing the function with a trapezoidal approximation. Then the signed-area under the trapezoids are easy to calculate. We still use the same partition notation and our goal is to approximate once more $\int_a^b f(x)dx$. Let's think about a subinterval $[x_{j-1}, x_j]$. We can replace $f(x)$ with the secant line from $(x_{j-1}, f(x_{j-1}))$ to $(x_j, f(x_j))$. This secant line has equation

$$y = s_j(x) = f(x_j) + \frac{f(x_j) - f(x_{j-1})}{\Delta x}(x - x_j)$$

Therefore, we can approximate the integral on this subinterval as follows:

$$\begin{aligned} \int_{x_{j-1}}^{x_j} f(x)dx &\approx \int_{x_{j-1}}^{x_j} \left(f(x_j) + \frac{f(x_j) - f(x_{j-1})}{\Delta x}(x - x_j) \right) dx \\ &= f(x_j)(x_j - x_{j-1}) - \frac{f(x_j) - f(x_{j-1})}{2\Delta x}(x_{j-1} - x_j)^2 \\ &= f(x_j)\Delta x - \frac{f(x_j) - f(x_{j-1})}{2\Delta x}(-\Delta x)^2 \\ &= \frac{1}{2} \left[f(x_j) + f(x_{j-1}) \right] \Delta x. \end{aligned}$$

Consider that

$$\begin{aligned} \int_a^b f(x)dx &= \int_a^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \dots + \int_{x_{n-2}}^{x_{n-1}} f(x)dx + \int_{x_{n-1}}^b f(x)dx \\ &\approx \frac{1}{2} \left(f(x_1) + f(a) \right) \Delta x + \frac{1}{2} \left(f(x_2) + f(x_1) \right) \Delta x + \dots \\ &\quad \dots + \frac{1}{2} \left(f(x_{n-1}) + f(x_{n-2}) \right) \Delta x + \frac{1}{2} \left(f(b) + f(x_{n-1}) \right) \Delta x \\ &= \frac{1}{2} \left(f(a) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-2}) + 2f(x_{n-1}) + f(b) \right) \Delta x \end{aligned}$$

Definition 10.2.1. Trapezoid rule (T_n).

Suppose that $[a, b] \subseteq \text{dom}(f)$ then we define

$$T_n = \frac{1}{2} (f(a) + f(b)) + \sum_{j=1}^{n-1} f(x_j) \Delta x = \frac{1}{2} [f(a) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(b)] \Delta x.$$

Example 10.2.2. Once more I return to $\int_0^{10} e^{-x^2}$. We'll examine T_{10} ,

$$T_{10} = \frac{1}{2}(1 + 2e^{-1} + 2e^{-4} + \cdots + 2e^{-81} + e^{-100}) \approx 0.8863186.$$

Wolfram Alpha yields $\int_0^{10} e^{-x^2} \approx 0.8862269$ so you can see we got the first 3 decimals, just like M_{10} . This is not surprising (see the error bounds for T_n and M_n they are very much the same order, page 535).

Again, I should emphasize that if you must do these with technology. I've used a mixture of Wolfram Alpha and plain old figuring to arrive at my numbers. You are free to use Mathematica, Matlab, etc... it matters not to me which CAS¹ you use. Just learn something.

10.3 Simpson's Rule

The idea of Simpson's rule is simple enough: approximate the signed area by replacing the given function with a parabolic splice on each subinterval. We can easily integrate the signed area under parabolas so calculating the approximate total signed area should be fairly easy. In this case we must suppose that the number of approximating regions is even. After a couple pages of algebra one comes to the following formula:

$$\int_a^b f(x)dx \approx \frac{\Delta x}{3} \left[f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-3}) + 2f(x_{n-2}) + 4f(x_{n-1}) + f(b) \right]$$

This is known as Simpson's Rule.

Definition 10.3.1. *Simpson's rule (S_n).*

Suppose that $[a, b] \subseteq \text{dom}(f)$ and n is even then we define

$$S_n = \frac{\Delta x}{3} \left[f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \cdots + 4f(x_{n-3}) + 2f(x_{n-2}) + 4f(x_{n-1}) + f(b) \right].$$

Let us once more return to my favorite example.

Example 10.3.2. Once more I return to $\int_0^{10} e^{-x^2} dx$. We'll examine S_{10} ,

$$S_{10} = \frac{1}{3}(1 + 4e^{-1} + 2e^{-4} + 4e^{-9} + 2e^{-16} + 4e^{-25} + 2e^{-36} + 4e^{-49} + 2e^{-64} + 4e^{-81} + e^{-100})$$

$$S_{10} \approx 0.8362143$$

Recall $T_{10} \approx 0.8863186$ and Wolfram Alpha yields $\int_0^{10} e^{-x^2} \approx 0.8862269$ so you can see we only correctly calculated the first decimal. This is actually somewhat interesting since it is usually the case that Simpson's rule gives a more accurate result for a given number of approximating regions (see the error bounds for S_n , page 539 in Stewart's 6-th Ed.). One may calculate that $S_{20} \approx 0.886196$ so at least by $n = 20$ the rule of Simpson has arrived at the same accuracy as the trapezoid or midpoint rules of order 10.

¹Computer Algebra System

Example 10.3.3. *I feel bad for Simpson's rule so I'm going to give it another chance at a new function. Let's calculate $\int_0^1 \sqrt{1+x^5} dx$ via all three methods and by the numerical integrator in Wolfram Alpha:*

$$\int_0^1 \sqrt{1+x^5} dx \approx 1.074669188848 \quad \text{by Wolfram Alpha.}$$

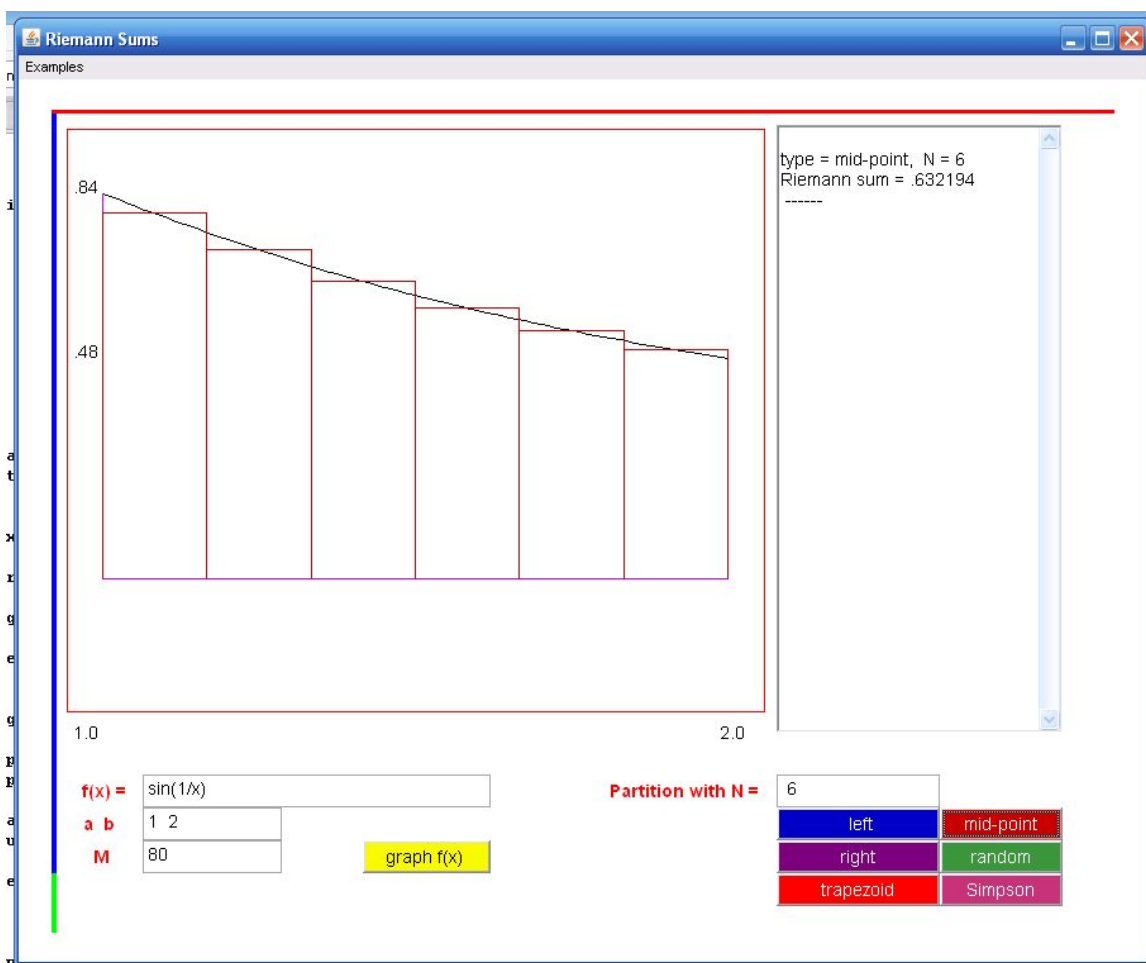
$$M_{10} \approx 1.073933$$

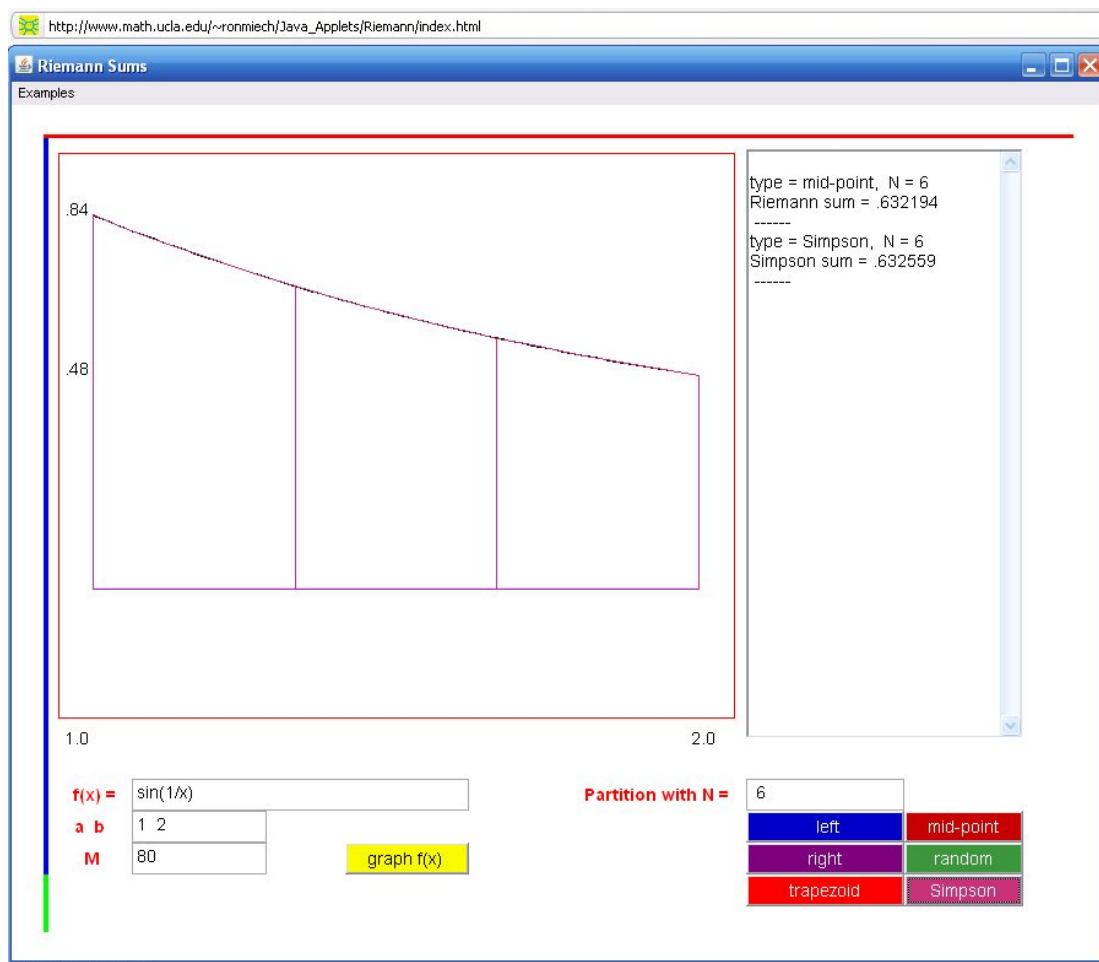
$$T_{10} \approx 1.076142$$

$$S_{10} \approx 1.074671$$

You can see that Simpson's rule has 5 correct digits whereas the midpoint and trapezoid have only 3 correct digits. You can find a link to an online calculator at my website which both calculates and graphs these rules.

Example 10.3.4. *Below are a couple screen-shots of a Java Applet which calculates and graphs the left, right, midpoint, trapezoid and Simpson's rule. Note the website of the author is visible at the top of one of the pictures.*





Finally, Wolfram Alpha weighs in with an estimate of

$$\int_1^2 \sin(1/x) dx = -Ci(1/2) + Ci(1) + 2 \sin(1/2) - \sin(1) \approx 0.632568$$

where $Ci(x)$ is the *cosine integral* function. This is one of the aspects of Wolfram Alpha I especially enjoy, it leads you to discover new mathematics while you are solving a standard problem. Welll, I hope you've seen enough that you could numerically approximate a definite integral. The general and careful study of this problem is beyond this course.

Chapter 11

Improper Integrals

Improper integrals are important to many applications where the concept of ∞ is used to implement some physical idealization. Or in probability, if you have a continuous random variable then the natural condition for the probability density function ϕ is $\int_{-\infty}^{\infty} \phi(x)dx = 1$. In physics the divergence of functions often represents a physically interesting quantity. For example, if $E(x)$ is the electric field in a one-dimensional system and $E(x) \rightarrow \infty$ as $x \rightarrow x_o^+$ then there is a positive electric charge at $x = x_o$. The mathematics of improper integrals are all made by combining the concept of the integral with the concept of a limit at a point or infinity or both. Once you understand the definitions in this Chapter they are entirely natural, with perhaps the exception of $\int_{-\infty}^{\infty} f(x)dx$.

Beyond infinity, or perhaps before it, we can use the concept of improper integration to carefully define integration of piecewise-continuous functions. The reason is subtle. Riemann integration is only defined for closed intervals. If $f(x)$ is piecewise defined to be $f_1(x)$ for $x \in [a, c)$ and on the other hand $f_2(x)$ for $x \in (c, b]$ then many calculus I instructors state that

$$\int_a^b f(x)dx = \int_a^c f_1(x)dx + \int_c^b f_2(x)dx.$$

However, it is cannot just mean the Riemann integral¹ of f_1 and f_2 . At least without some qualification this equation is nonsense² because I cannot set up a partition which allows us to evaluate $f_1(c)$ or $f_2(c)$. These points need not be in the domain of f and they certainly cannot be defined for both f_1 and f_2 and yet be distinct (if they were not equal then f would not be function). So, how should we define such an integral? It's not hard, we just have to introduce variable bounds and allow those bounds to tend to the point c rather than actually reach it. In practice, I don't force students to write these limits in calculus I because in the case of finite-jump discontinuities there is no danger of getting the wrong answer³. On the other hand, once the integrand has a point of discontinuity where there is a vertical asymptote the limit is an essential detail as application of L'Hospital's rule and other subtle limit calculations are totally possible.

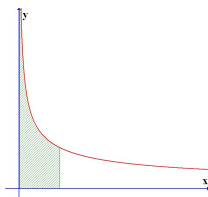
¹not as we have defined it at least

²it will make perfect sense soon, we just need a simple definition to fix this hole. Actually, depending on the finer points of your definition of integration this may already be meaningful. The point I raise here may be too picky for the integration theory given in calculus I and II.

³I show we were correct later in this chapter, I show the limits are not really necessary for the bounded piecewise continuous case.

11.1 divergent integrands

Let me begin with a question: what is the meaning of $\int_0^1 \frac{dx}{\sqrt{x}}$? Notice it cannot just be the limit of a Riemann sum since the function is not even defined at the first partition point $x_o = a = 0$ and that point gives division by zero for the integrand $1/\sqrt{x}$. On the other hand, if we consider the graph then we'll note there might be a chance that the area bounded by $x = 0$, $x = 1$, $y = 0$ and $y = 1/\sqrt{x}$ is finite since the shape gets very very narrow where it gets very very tall.



Furthermore, if $t > 0$ then clearly the integral $\int_t^1 \frac{dx}{\sqrt{x}}$ is meaningful for each such value t . I propose the natural definition for the integral in question is simply as follows:

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{\sqrt{x}}$$

The expression above defines the *improper* integral on the LHS in terms of concepts on the RHS which we have already given explicit and rigorous meaning.

Example 11.1.1.

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{\sqrt{x}} = \lim_{t \rightarrow 0^+} \left(2\sqrt{x} \Big|_t^1 \right) = \lim_{t \rightarrow 0^+} (2\sqrt{1} - 2\sqrt{t}) = \boxed{2}.$$

I'll give a careful definition that covers the preceding example. Again, this is the natural definition because it extends our concept in a way such that the integral still has the same essential meaning.

Definition 11.1.2. *Improper integral with divergence at edge of integration interval.*

Assume $a < b$ and $(a, b) \subseteq \text{dom}(f)$. The integral $\int_a^b f(x)dx$ is generalized to mean the following in the cases that either $a \notin \text{dom}(f)$ or $b \notin \text{dom}(f)$.

1. If $a \notin \text{dom}(f)$ however $b \in \text{dom}(f)$ then $\int_a^b f(x)dx = \lim_{t \rightarrow a^+} \int_t^b f(x)dx$.
2. If $b \notin \text{dom}(f)$ however $a \in \text{dom}(f)$ then $\int_a^b f(x)dx = \lim_{t \rightarrow b^-} \int_a^t f(x)dx$.
3. If both $a, b \notin \text{dom}(f)$ then we choose $c \in (a, b)$ and define $\int_a^b f(x)dx = \lim_{t \rightarrow a^+} \int_t^c f(x)dx + \lim_{s \rightarrow b^-} \int_c^s f(x)dx$.

If any of the limits exist then we say the integral converges to the value of the limit. Moreover, if the integral converges to a value we say it is **convergent**. However, if any of the limits above do not exist then we say the integral is **divergent**. In the case the the limits diverge to $\pm\infty$ we will denote the divergence by $\int_a^b f(x)dx = \pm\infty$ in cases (1.) and (2.). However, in case (3.) we say that $\int_a^b f(x)dx = d.n.e.$ if one of the limits tends to ∞ while the other tends to $-\infty$. Finally, if both limits tend ∞ ($-\infty$) then we say $\int_a^b f(x)dx$ diverges to ∞ ($-\infty$).

You might expect that ∞ and $-\infty$ can cancel. While that is possible, it is not possible in the context of case (3.) above. The divergence between the upper and lower bound are not connected in this definition.

Example 11.1.3.

$$\int_0^1 \frac{dx}{x} = \lim_{t \rightarrow 0^+} \int_t^1 \frac{dx}{x} = \lim_{t \rightarrow 0^+} \left(\ln |x| \Big|_t^1 \right) = \lim_{t \rightarrow 0^+} \left(2 \ln 1 - 2 \ln t \right) = \boxed{\infty}.$$

This integral diverged to ∞ .

Example 11.1.4.

$$\int_0^1 \frac{dx}{x-1} = \lim_{t \rightarrow 1^-} \int_0^t \frac{dx}{x-1} = \lim_{t \rightarrow 1^-} \left(\ln |x-1| \Big|_0^t \right) = \lim_{t \rightarrow 0^+} \left(2 \ln |t-1| - 2 \ln |0-1| \right) = \boxed{-\infty}.$$

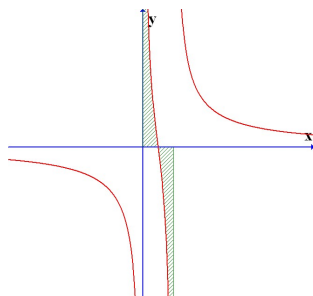
This integral diverged to ∞ .

Example 11.1.5.

$$\begin{aligned} \int_0^1 \left(\frac{1}{x} + \frac{1}{x-1} \right) dx &= \lim_{t \rightarrow 0^+} \int_t^{0.5} \left(\frac{1}{x} + \frac{1}{x-1} \right) dx + \lim_{s \rightarrow 1^-} \int_{0.5}^s \left(\frac{1}{x} + \frac{1}{x-1} \right) dx \\ &= \lim_{t \rightarrow 0^+} \left(\ln |x| + \ln |x-1| \Big|_t^{0.5} \right) + \lim_{s \rightarrow 1^-} \left(\ln |x| + \ln |x-1| \Big|_{0.5}^s \right) \\ &= \lim_{t \rightarrow 0^+} \left(2 \ln |0.5| - \ln |t| + \ln |t-1| \right) + \lim_{s \rightarrow 1^-} \left(\ln |s| + \ln |s-1| - 2 \ln |0.5| \right) \\ &= \infty - \infty \\ &= \boxed{d.n.e.} \end{aligned}$$

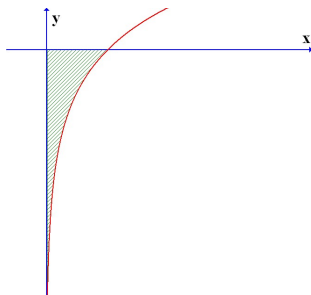
Notice that if we used some other middle point besides 0.5 the answer would not change. Moreover, if you notice the pattern of the terms following from the 0.5-evaluations it is clear this cancellation will also occur in the case of integrals which are not divergent. If this were not the case, if the integral depended on the choice of the middle point, then the definition we gave would be meaningless. I leave the proof that the ambiguity in the choice of c does not lead to ambiguity in the integral to the reader.

The area of the shaded region was shown to diverge in the last example.



Example 11.1.6. Recall that IBP yields that $\int \ln(x)dx = x \ln(x) - x + c$ and consider:

$$\begin{aligned}
 \int_0^1 \ln(x)dx &= \lim_{t \rightarrow 0^+} \int_t^1 \ln(x)dx \\
 &= \lim_{t \rightarrow 0^+} \left(1 \ln(1) - 1 - t \ln(t) - t \right) \\
 &= -1 + \lim_{t \rightarrow 0^+} \left(-t \ln(t) \right) \quad \text{note this is type } 0\infty \\
 &= -1 + \lim_{t \rightarrow 0^+} \left(\frac{\ln(t)}{\frac{-1}{t}} \right) \quad \text{now is type } \infty/\infty \text{ thus apply L'Hospital's Rule.} \\
 &= -1 + \lim_{t \rightarrow 0^+} \left(\frac{\frac{1}{t}}{\frac{1}{t^2}} \right) \\
 &= -1 + \lim_{t \rightarrow 0^+} (t) \\
 &= \boxed{-1}.
 \end{aligned}$$



Next we consider the case that the function has an isolated discontinuity in the domain of integration.

Definition 11.1.7. *Improper integral with discontinuities in the interior of integration interval.*

Assume $a < c < b$ and $\lim_{x \rightarrow c} f(x) \neq f(c)$. If both $\int_c^b f(x)dx \in \mathbb{R}$ and $\int_a^c f(x)dx \in \mathbb{R}$ then the integral $\int_a^b f(x)dx$ is generalized to mean $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$. If either $\int_a^c f(x)dx$ or $\int_c^b f(x)dx$ diverge then we say $\int_a^b f(x)dx$ diverges. We say that $\int_a^b f(x)dx = d.n.e.$ if one of the improper integrals diverges to ∞ while the other diverges to $-\infty$. Finally, if both diverge to ∞ ($-\infty$) then we say $\int_a^b f(x)dx$ diverges to ∞ ($-\infty$).

Suppose $\{a_1, a_2, \dots, a_n\}$ is a finite set of discontinuities for an otherwise continuous function f and $a < a_1 < a_2 < \dots < a_n < b$. are points of discontinuities for an otherwise continuous function f on $[a, b]$ then we define $\int_a^b f(x) = \int_a^{a_1} f(x)dx + \int_{a_1}^{a_2} f(x)dx + \dots + \int_{a_n}^b f(x)dx$.

Example 11.1.8.

$$\begin{aligned}
\int_{-2}^3 \frac{1}{x} dx &= \int_{-2}^0 \frac{dx}{x} + \int_0^3 \frac{dx}{x} \\
&= \lim_{t \rightarrow 0^-} \int_{-2}^t \frac{dx}{x} + \lim_{s \rightarrow 0^+} \int_s^3 \frac{dx}{x} \\
&= \lim_{t \rightarrow 0^-} (\ln |t| - \ln |-2|) + \lim_{s \rightarrow 0^+} (\ln |3| - \ln |s|) \\
&= -\infty + \infty \\
&= \boxed{d.n.e.}
\end{aligned}$$

Some students incorrectly think $\int_{-2}^3 \frac{1}{x} dx$ is calculated by $\lim_{t \rightarrow 0^+} \left(\int_{-2}^{-t} \frac{dx}{x} + \int_t^3 \frac{dx}{x} \right)$. Note:

$$\lim_{t \rightarrow 0^+} \left(\int_{-2}^{-t} \frac{dx}{x} + \int_t^3 \frac{dx}{x} \right) = \lim_{t \rightarrow 0^+} \left(\ln |-t| - \ln |-2| + \ln |3| - \ln |t| \right) = \ln 3 - \ln 2$$

However, while this calculation is interesting, it is not the improper integral at zero. It is incorrect to assume that we approach zero in the same way from the left and right. These are independent limits.

Example 11.1.9.

$$\begin{aligned}
\int_0^3 \frac{dx}{(x-1)^2} &= \int_0^1 \frac{dx}{(x-1)^2} + \int_1^3 \frac{dx}{(x-1)^2} \\
&= \lim_{t \rightarrow 1^-} \int_0^t \frac{dx}{(x-1)^2} + \lim_{s \rightarrow 1^+} \int_s^3 \frac{dx}{(x-1)^2} \\
&= \lim_{t \rightarrow 1^-} \left(\frac{-1}{t-1} + \frac{1}{0-1} \right) + \lim_{s \rightarrow 1^+} \left(\frac{-1}{3-1} + \frac{1}{s-1} \right) \\
&= -(-\infty) + \infty \\
&= \boxed{\infty}
\end{aligned}$$

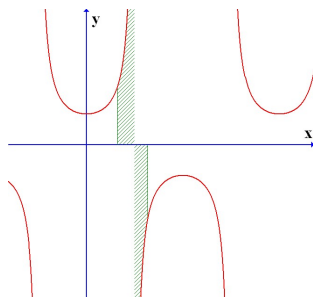
Example 11.1.10.

$$\int_0^3 \frac{x^2 + 3x - 2}{x(x^2 - 4)} dx = \lim_{t_1 \rightarrow 0^+} \int_{t_1}^1 \frac{x^2 + 3x - 2}{x(x^2 - 4)} dx + \lim_{t_2 \rightarrow 2^-} \int_1^{t_2} \frac{x^2 + 3x - 2}{x(x^2 - 4)} dx + \lim_{t_3 \rightarrow 2^+} \int_{t_3}^3 \frac{x^2 + 3x - 2}{x(x^2 - 4)} dx$$

Note that $x = -2, 2, 0$ are all points where the integrand is not defined due to division by zero hence we can only approach these points. However, only 0 and 2 are within the integration region.

Example 11.1.11. Recall $\pi/2 \approx 1.57$ and $\sec(x) = 1/\cos(x)$ to understand what follows:

$$\int_1^2 \sec(x) dx = \int_1^{\pi/2} \sec(x) dx + \int_{\pi/2}^2 \sec(x) dx.$$



11.1.1 on the integration of finite jump discontinuous functions

Example 11.1.12. Notice $f(x) = x/|x|$ is discontinuous at $x = 0$ since $f(x) = x/-x = -1$ for $x < 0$ and $f(x) = x/x = 1$ for $x > 0$.

$$\int_{-2}^3 f(x) dx = \int_{-2}^0 \frac{x dx}{|x|} + \int_0^3 \frac{x dx}{|x|} = \lim_{t \rightarrow 0^+} \int_{-2}^t (-dx) + \lim_{s \rightarrow 0^-} \int_s^3 dx = \lim_{t \rightarrow 0^+} (-2 - t) + \lim_{s \rightarrow 0^-} (3 - s) = 1.$$

Compare the example above to the following:

$$\int_{-2}^3 f(x) dx = \int_{-2}^0 \frac{x dx}{|x|} + \int_0^3 \frac{x dx}{|x|} = \int_{-2}^0 (-dx) + \int_0^3 dx = -x \Big|_{-2}^0 + x \Big|_0^3 = 0 - 2 + 3 - 0 = 1.$$

The calculation following the example is what is more common to find in calculus texts and courses, we ought to wonder if both of these procedures always yield the same result. The answer, in the case the function is bounded piecewise continuous, is yes. I explain why below: Suppose f has a finite jump-discontinuity at $x = c$ and is otherwise continuous on $[a, b]$ where $c \in (a, b)$. This means that $\lim_{x \rightarrow c^-} f(x) = L_1$ and $\lim_{x \rightarrow c^+} f(x) = L_2$ where $L_1, L_2 \in \mathbb{R}$ and $L_1 \neq L_2$. Define g by $g(x) = f(x)$ for $x \in [a, c)$ and $g(c) = L_1$. Observe that the function g is continuous on $[a, c]$. Apply FTC part I to obtain the existence of an antiderivative G for g then apply FTC part II to find $\int_a^t g(x) dx = G(t) - G(a)$ for any $t \in (a, c]$.

Next, define h by $h(x) = f(x)$ for $x \in (c, b]$ and define $h(c) = L_2$. Observe that the function h is continuous on $[c, b]$. Apply FTC part I to obtain the existence of an antiderivative H for h then apply FTC part II to find $\int_s^b h(x) dx = H(b) - H(s)$ for any $s \in [c, b)$.

Consider,

$$\begin{aligned}
 \int_a^b f(x)dx &= \int_a^c f(x)dx + \int_c^b f(x)dx \\
 &= \lim_{t \rightarrow c^+} \int_a^t f(x)dx + \lim_{s \rightarrow c^-} \int_s^b f(x)dx \\
 &= \lim_{t \rightarrow c^+} \int_a^t g(x)dx + \lim_{s \rightarrow c^-} \int_s^b h(x)dx \\
 &= \lim_{t \rightarrow c^+} (G(t) - G(a)) + \lim_{s \rightarrow c^-} (H(b) - H(s)) \\
 &= G(c) - G(a) + H(b) - H(c) \\
 &= \int_a^c g(x)dx + \int_c^b h(x)dx.
 \end{aligned}$$

Therefore, to integrate a piecewise-defined function with simple jump-discontinuities we can break the integral into a sum of integrals where each integral is not of the function itself, but rather the continuous extension of the function onto the closure of the subinterval. Pragmatically, this means we need not write limits in the case of improper integrals of the finite-jump-discontinuity type. We can calculate the antiderivative of each piece, and simply sum the sub-integrals.

The especially picky reader will question how I made the first equality in the argument above. In truth, to really understand such questions we need a more careful theory of integration. The question I raised on this page probably is better left to a real analysis course.

The overall lesson I'd like you to take here is that we can ignore the contribution of a few points if the points are merely finite-jump-discontinuities. On the other hand, if the points are genuine divergences for the integrand then we have seen that more care is required.

Dr. Skoumbourdis sometimes emphasized "you can only integrate over a closed interval" and that is what got me started on the last page. I thought, when I break the piecewise integral into pieces those pieces are not closed intervals so what I am calling an integral must not quite be what I usually call an integral. Or at least, something subtle is being done with the points of discontinuity.

Those points of discontinuity are on the edge of the subintervals so it also occurred to me that if we did not use the edges for our integral then we might be able to get around the problem. For example, the midpoint rule would never sample the endpoints so it would be well-defined even if the endpoints were not given for the integrand. It turns out that this is precisely the reason that the midpoint rule is interesting for numerical methods. If you want to calculate an improper integral then the midpoint rule is a nice choice because it does not sample endpoints where the function is often not even defined. In contrast, left, right, trapezoid and Simpsons all use boundary points of the integration region to approximate the integral.

11.2 divergent bounds

The other type of improper integral is where the bounds are infinite. The question we begin with is what should we mean by $\int_a^\infty f(x)dx$? The natural interpretation is to use this to denote the signed area bounded by $y = f(x)$ for $x \geq a$. We can calculate the area out to $x = t$ by $\int_a^t f(x)dx$. We want to let $t \rightarrow \infty$ to catch all the signed-area bounded by $y = f(x)$. This brings us to the definition below:

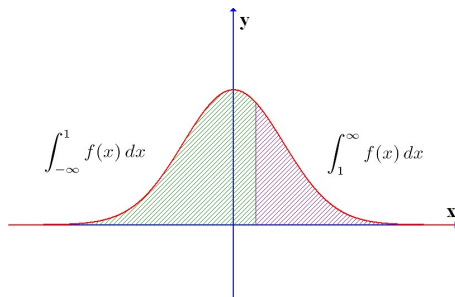
Definition 11.2.1. *Integrals with infinite bounds.*

1. If f is continuous on $[a, \infty)$ then we define $\int_a^\infty f(x)dx = \lim_{t \rightarrow \infty} \int_a^t f(x)dx$.
2. If f is continuous on $(-\infty, a]$ then we define $\int_{-\infty}^a f(x)dx = \lim_{t \rightarrow -\infty} \int_t^a f(x)dx$.
3. If f is continuous on \mathbb{R} then we define $\int_{-\infty}^\infty f(x)dx = \lim_{t \rightarrow -\infty} \int_t^a f(x)dx + \lim_{s \rightarrow \infty} \int_a^s f(x)dx$.

In each case if the limit exists then we say the integral is convergent, whereas if the limit does not exist then we say it diverges. If the limit tends to $\pm\infty$ in cases (1.) or (2.) then we say the integral diverges to $\pm\infty$. In case (3.) if both limits tend to ∞ ($-\infty$) then we say the integral diverges to ∞ ($-\infty$). In case (3.) if one of the limits is divergent and the other is convergent then the integral diverges. Finally, if the function in question is not continuous at some finite number of points we combine this definition with the previous definition for improper integrals at points of discontinuity.

Sorry for the abundance of words, I'm simply trying to cover all the bases. One of the concepts I'm trying to impress on you is that the words "convergence" and "divergence" are now also applicable to improper integrals. Moreover, the actual calculation of these improper integrals amounts to performing an integral with a variable bound(s) followed by taking the limit as that bound tends to $\pm\infty$. Graphically the idea is simple enough: note below that the entire area, if it is finite, is found from adding the two areas together

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^1 f(x)dx + \int_1^\infty f(x)dx$$



Note the choice of $x = 1$ as the dividing line was simply a choice. Fortunately, the result will not change for different choices since an alternate choice amounts to cutting the same area into two different sized pieces⁴.

Example 11.2.2.

$$\int_0^\infty x dx = \lim_{t \rightarrow \infty} \int_0^t x dx = \lim_{t \rightarrow \infty} \left[\frac{1}{2} t^2 - 0 \right] = \boxed{\infty}$$

⁴a more careful proof of this claim is left to the interested reader.

Example 11.2.3.

$$\int_{-\infty}^0 e^x dx = \lim_{t \rightarrow -\infty} \int_t^0 e^x dx = \lim_{t \rightarrow \infty} [e^0 - e^t] = \boxed{1}.$$

Example 11.2.4.

$$\int_0^{\infty} \frac{dx}{1+x^2} = \lim_{t \rightarrow \infty} \int_0^t \frac{dx}{1+x^2} = \lim_{t \rightarrow \infty} [\tan^{-1}(t) - \tan^{-1}(0)] = \boxed{\frac{\pi}{2}}.$$

See Example 2.7.11 for the background to make the last example meaningful. Notice that the integrand $\frac{1}{1+x^2} \rightarrow 0$ as $x \rightarrow \infty$. One is tempted to suppose this is enough for such an improper integral to exist. However, the next example shows that life is not so easy.

Example 11.2.5.

$$\int_1^{\infty} \frac{dx}{x} = \lim_{t \rightarrow \infty} \int_1^t \frac{dx}{x} = \lim_{t \rightarrow \infty} [\ln(t) - \ln(1)] = \boxed{\infty}.$$

Example 11.2.6. Let $p > 1$ and consider:

$$\int_1^{\infty} \frac{dx}{x^p} = \lim_{t \rightarrow \infty} \int_1^t \frac{dx}{x^p} = \lim_{t \rightarrow \infty} \left[\frac{1}{(-p+1)t^{p-1}} + \frac{1}{p-1} \right] = \boxed{\frac{1}{p-1}}.$$

The preceding pair of examples prove the nontrivial portion of the proposition below:

Theorem 11.2.7. *p-test for improper integrals.*

1. If $p \geq 1$ then $\int_1^{\infty} \frac{dx}{x^p}$ diverges.
2. If $p < 1$ then $\int_1^{\infty} \frac{dx}{x^p} = \frac{1}{p-1}$, hence it converges.

The case $p = 1$ is important and we will see it again.

Example 11.2.8. Refer to Example 9.3.7 for the details of how to calculate $\int \frac{dx}{4(x+1)^2+5}$.

$$\begin{aligned} \int_0^{\infty} \frac{dx}{4x^2+8x+9} &= \lim_{t \rightarrow \infty} \int_0^t \frac{dx}{4x^2+8x+9} \\ &= \lim_{t \rightarrow \infty} \left[\frac{1}{2\sqrt{5}} \tan^{-1} \left(\frac{2(t+1)}{\sqrt{5}} \right) - \frac{1}{2\sqrt{5}} \tan^{-1} \left(\frac{2}{\sqrt{5}} \right) \right] \\ &= \boxed{\frac{1}{2\sqrt{5}} \left[\frac{\pi}{2} - \tan^{-1} \left(\frac{2}{\sqrt{5}} \right) \right]}. \end{aligned}$$

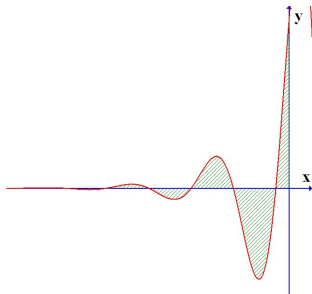
Example 11.2.9. Refer to Example 9.1.9 for details on why

$$\int e^x \sin(3x+1) dx = \frac{1}{10} e^x [\sin(3x+1) - \cos(3x+1)] + c.$$

I use the indefinite integral above to calculate the improper integral below:

$$\begin{aligned}
 \int_{-\infty}^0 e^x \sin(3x+1) dx &= \lim_{t \rightarrow -\infty} \int_t^0 e^x \sin(3x+1) dx \\
 &= \lim_{t \rightarrow -\infty} \left[\frac{1}{10} e^0 [\sin(1) - \cos(1)] - \frac{1}{10} e^t [\sin(3t+1) - \cos(3t+1)] \right] \\
 &= \boxed{\frac{\sin(1) - \cos(1)}{10}}.
 \end{aligned}$$

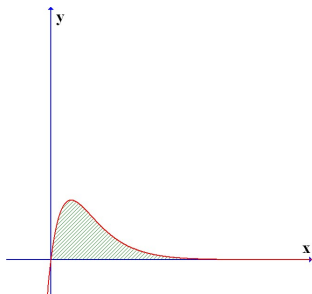
The area illustrated below is calculated by the integral above.



Example 11.2.10. Note that $\int x e^{-x} dx = -x e^{-x} + \int e^{-x} dx = -e^{-x}(x+1) + c$ by IBP. Thus,

$$\begin{aligned}
 \int_0^{\infty} x e^{-x} dx &= \lim_{t \rightarrow \infty} \int_0^t x e^{-x} dx \\
 &= \lim_{t \rightarrow \infty} \left[-e^{-t}(t+1) + e^{-0}(0+1) \right] \\
 &= \lim_{t \rightarrow \infty} \left[-e^{-t}(t+1) \right] + 1 \\
 &= \lim_{t \rightarrow \infty} \left[-\frac{t+1}{e^t} \right] + 1 \\
 &= \lim_{t \rightarrow \infty} \left[-\frac{1}{e^t} \right] + 1 \\
 &= \boxed{1}.
 \end{aligned}$$

We used L'Hospital's rule to simplify the limit of type ∞/∞ going from the 4th to 5th line. The area illustrated below is calculated by the integral above.



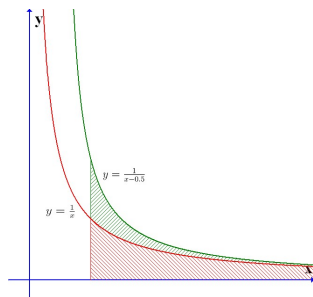
Theorem 11.2.11. *Comparison test for improper integrals.*

Suppose that f and g are continuous functions on $[a, \infty)$ such that $0 \leq f(x) \leq g(x)$ for all $x \geq a$.

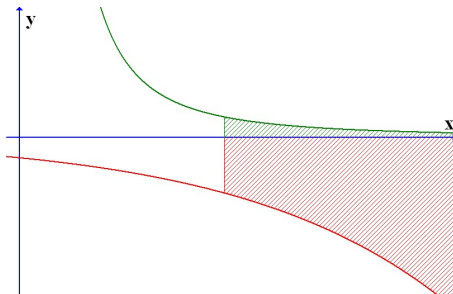
1. If $\int_a^\infty g(x)dx$ converges then $\int_a^\infty f(x)dx$ converges. Moreover, $\int_a^\infty f(x)dx \leq \int_a^\infty g(x)dx$.
2. If $\int_a^\infty f(x)dx$ diverges then $\int_a^\infty g(x)dx$ diverges.

Proof: To prove (1.) we remind the reader that a property of definite integrals is that they preserve inequalities. In particular, $0 \leq f(x) \leq g(x)$ implies $0 \leq \int_a^t f(x)dx \leq \int_a^t g(x)dx$. Then, the limiting process likewise preserves this inequality hence $0 \leq \lim_{t \rightarrow \infty} \int_a^t f(x)dx \leq \lim_{t \rightarrow \infty} \int_a^t g(x)dx$. Note further that $\int_a^t f(x)dx$ is an increasing function of t since $\frac{d}{dt} \int_a^t f(x)dx = f(t) \geq 0$ for $t \geq a$. We have that $\int_a^t f(x)dx$ is an increasing function of t which is bounded between 0 and $\int_a^\infty g(x)dx \in \mathbb{R}$. It follows that $\lim_{t \rightarrow \infty} \int_a^t f(x)dx$ exists. Therefore, the convergence of $\int_a^\infty g(x)dx$ implies the convergence of $\int_a^\infty f(x)dx$. I leave the proof of (2.) to the reader. In fact, the reader may also need to supply the lemma which proves the limit of a bounded increasing function exists. I don't believe we have proved that result in these notes. \square

Example 11.2.12. Does $\int_1^\infty \frac{dx}{x-0.5}$ converge or diverge? Notice that $0 \leq \frac{1}{x} \leq \frac{1}{x-0.5}$ for all $x \geq 1$. Furthermore $f(x) = \frac{1}{x}$ and $g(x) = \frac{1}{x-0.5}$ are clearly continuous on $[1, \infty)$. Observe that $\int_1^\infty \frac{dx}{x} = \infty$ by the $p = 1$ test hence by the comparison theorem we find $\int_1^\infty \frac{dx}{x-0.5}$ diverges. The area illustrates the comparison theorem in action, the integral of the lower function diverges by the p -test for improper integrals thus the integral of the larger function must likewise diverge.



Example 11.2.13. Notice that $-e^x < \frac{1}{x^2}$ for $x \geq 1$. Furthermore, note that $\int_1^\infty \frac{dx}{x^2}$ converges by the $p = 2$ improper integral test. Therefore, the integral of a function larger than $-e^x$ converges hence by the comparison test $\int_1^\infty -e^x dx$ converges. Yet this is clearly false: note the area is obviously unbounded for the graph $y = -e^x$ for $x \geq 1$. What is the flaw in my logic? Fix this argument. Here is a picture of the problem:



Theorem 11.2.14. *Tail of integral is key for improper integrals.*

Suppose that f is a continuous function on $[a, \infty)$ and $b > a$ then

1. $\int_a^\infty f(x)dx$ converges iff $\int_b^\infty f(x)dx$ converges.
2. $\int_a^\infty f(x)dx$ diverges iff $\int_b^\infty f(x)dx$ diverges.

Notice that the assumption of continuity is important since for a discontinuous functions we can find divergence at finite points. For example, $\int_1^\infty \frac{dx}{(x-4)^2}$ diverges whereas $\int_5^\infty \frac{dx}{(x-4)^2}$ converges. We cannot just discard the low values of x in the convergence analysis for $\int_1^\infty \frac{dx}{(x-4)^2}$. Ok, enough about what we can't do let's see what we can do⁵:

Example 11.2.15. *Does $\int_{10}^\infty \frac{dx}{x^3}$ converge or diverge? Notice*

$$\underbrace{\int_1^\infty \frac{dx}{x^3}}_{1/2} = \underbrace{\int_1^{10} \frac{dx}{x^3}}_{99/200} + \int_{10}^\infty \frac{dx}{x^3}$$

Furthermore, it is clear that $1/x^3$ is continuous on $[1, \infty)$ hence we conclude that this integral is just the tail of the $p = 3$ integral which converges hence $\int_{10}^\infty \frac{dx}{x^3}$ converges. Moreover, in a case such as this we can even calculate its value indirectly by $\int_{10}^\infty \frac{dx}{x^3} = 1/2 - 99/200 = 1/200$.

Remark 11.2.16. *alternate solutions possible.*

When analyzing convergence and divergence it is often possible to argue a point many different ways. In this section, I do not always seek an optimal solution. Instead, I am trying to illustrate the theorems with examples that are not overly difficult. One side-effect of this approach is that currently we have no need to use the Tail theorem or the comparison theorem for the examples above. We could have just straight calculated the improper integrals and came to the same conclusion with less thinking. However, we will soon consider examples where the integral is simply not calculatable in closed form. Instead our approach will be to see how it is like something we can calculate then use the tail, comparison and/or the linearity theorem (given later in this section) to circumvent direct calculation of the improper integral.

⁵proof of the tail theorem is left to the reader, I don't think this is a hard one

Example 11.2.17. Does $\int_a^\infty \frac{dx}{x^4+8x^2+3}$ converge for a sufficiently large choice of a ? Note, $x^4+8x^2+3 > 0$ thus we can choose $a = 0$ in this case. The integrand $\frac{1}{x^4+8x^2+3}$ is clearly continuous on $[0, \infty)$. Moreover, notice that for $x \geq 0$

$$\frac{1}{x^4+8x^2+3} \leq \frac{1}{x^4} \quad [\text{think about it, I made the denominator smaller}].$$

Observe $\int_0^\infty \frac{dx}{x^4}$ converges because it is the tail of the $p = 4$ integral which we know converges. Furthermore, $0 \leq \frac{1}{x^4+8x^2+3} \leq \frac{1}{x^4}$ hence the comparison theorem applies and we find $\int_0^\infty \frac{dx}{x^4+8x^2+3}$ converges.

The inequality in the last example was pretty easy to see. The next example, which concerns $\int_a^\infty \frac{dx}{x^4-8x^2-3}$ is much harder due to the change of sign. We would like to compare this integral to the $p = 4$ integral. However, the approach used in the last example doesn't quite work because we have the wrong inequality. Notice that for $x \geq a$

$$\frac{1}{x^4} \leq \frac{1}{x^4-8x^2-3} \quad [\text{think about it, I made the denominator smaller on the RHS}].$$

This inequality is correct, it's just not what we need for our comparison theorem argument. The fact that the integral of the LHS converges doesn't say anything about integral of the RHS. We need to find a different argument for the example below.

Example 11.2.18. Does $\int_a^\infty \frac{dx}{x^4-8x^2-3}$ converge for a sufficiently large choice of a ? Consider that $x^4-8x^2-3 = 0$ is a quadratic equation in x^2 and the solutions are

$$x^2 = \frac{8 + \sqrt{64+12}}{2} \quad \text{or} \quad x^2 = \frac{8 - \sqrt{64+12}}{2}$$

Note that $\frac{8+\sqrt{76}}{2} > 0$ whereas $\frac{8-\sqrt{76}}{2} < 0$ thus we find only two real solutions to the equation $x^4-8x^2-3 = 0$. I'll give them labels for our convenience:

$$\lambda_1 = \sqrt{\frac{8 + \sqrt{76}}{2}} \quad \lambda_2 = -\sqrt{\frac{8 + \sqrt{76}}{2}}.$$

We can calculate $\lambda_1 < 3$. Thus, we know there is no zero for the denominator of $\frac{1}{x^4-8x^2-3}$ on $[3, \infty)$ hence the function is continuous and positive. Moreover, if $x > 3$ then clearly $x^2 - \lambda_1 > 1$ thus

$$\frac{1}{x^4-8x^2-3} = \frac{1}{(x^2-\lambda_1)(x^2-\lambda_2)} < \frac{1}{x^2-\lambda_2} < \frac{1}{x^2}.$$

We apply the comparison theorem to argue that $\int_3^\infty \frac{dx}{x^4-8x^2-3}$ converges since $\int_3^\infty \frac{dx}{x^2}$ is the tail of the $p = 2$ integral which is known to converge. We can apply the comparison theorem because we have shown the integrands are continuous and positive on $[3, \infty)$ and satisfy the needed inequality of $\frac{1}{x^4-8x^2-3} < \frac{1}{x^2}$ for all $x \in [3, \infty)$.

There are many techniques to use here. Certainly factoring polynomials is a great tool and sign charts help to unravel the non-linear inequalities. The idea that, for a positive fraction, increasing the denominator makes a fraction smaller whereas decreasing the denominator makes it larger is very important to keep in mind. Finally, the goal to reduce to a p -type integral for the purpose of comparison saves us the trouble of integration. However, there are other examples where an integration is likely the only way to go. For example:

Example 11.2.19. Does the integral below converge or diverge?

$$\int_1^{\infty} \frac{e^x dx}{\sin(x) + e^{2x}}$$

The sine function is bounded by $-1 \leq \sin(x) \leq 1$ and $e^{2x} \geq e^2$ for $x \geq 1$ since the exponential function is increasing. We find that $\sin(x) + e^{2x} \leq 1 + e^{2x}$ on the other hand the function $\frac{e^x}{\sin(x) + e^{2x}}$ clearly positive since the $\sin(x) + e^{2x} \geq -1 + e^{2x} > -1 + e^2 > 0$. We argue that for all $x \geq 1$

$$0 \leq \frac{e^x}{1 + e^{2x}} \leq \frac{e^x}{\sin(x) + e^{2x}}.$$

Notice that we can integrate the function above by making a $u = e^x$ substitution:

$$\int \frac{e^x dx}{1 + e^{2x}} = \int \frac{du}{1 + u^2} = \tan^{-1}(e^x) + c.$$

Hence we calculate

$$\begin{aligned} \int_1^{\infty} \frac{e^x dx}{1 + e^{2x}} &= \lim_{t \rightarrow \infty} \int_1^t \frac{e^x dx}{1 + e^{2x}} \\ &= \lim_{t \rightarrow \infty} \left[\tan^{-1}(e^t) - \tan^{-1}(e) \right] \\ &= \frac{\pi}{2} - \tan^{-1}(e). \end{aligned}$$

Therefore, $\int_1^{\infty} \frac{e^x dx}{1 + e^{2x}}$ converges. It follows by the comparison theorem that $\int_1^{\infty} \frac{e^x dx}{\sin(x) + e^{2x}}$ converges. Moreover, due to our calculations here we can even reason that $0 \leq \int_1^{\infty} \frac{e^x dx}{\sin(x) + e^{2x}} \leq \frac{\pi}{2} - \tan^{-1}(e)$.

Theorem 11.2.20. Linearity for improper integrals.

If $\int_a^{\infty} f(x)dx$ and $\int_a^{\infty} g(x)dx$ converge then for each $c \in \mathbb{R}$ we find $\int_a^{\infty} [f(x) + cg(x)]dx$ converges and

$$\int_a^{\infty} [f(x) + cg(x)]dx = \int_a^{\infty} f(x)dx + c \int_a^{\infty} g(x)dx.$$

On the other hand, if $\int_a^{\infty} f(x)dx$ diverges then $\int_a^{\infty} cf(x)dx$ also diverges for any $c \in \mathbb{R}$ such that $c \neq 0$. If $c > 0$ and $\int_a^{\infty} f(x)dx = \pm\infty$ then we likewise find $\int_a^{\infty} f(x)dx = \pm\infty$. If $c < 0$ and $\int_a^{\infty} f(x)dx = \pm\infty$ then we find $\int_a^{\infty} cf(x)dx = \mp\infty$.

I just made use of the matching \pm with \mp convention. As an alternate example of its use note you can express $\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$ and $\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B)$ with just a single equation in the matching \pm notation: $\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$. Notations aside, I hope the theorem above is entirely reasonable in your opinion. The proof stems primarily from the corresponding linearity rules for limits at ∞ and the details are left to the reader. This is a very useful theorem. We could take any pair of examples and splice them together with this theorem. I conclude with just one such example:

Example 11.2.21. Does the integral below converge or diverge?

$$\int_1^{\infty} \left[\frac{1}{3x^2} + \frac{2e^x}{\sin(x) + e^{2x}} \right] dx.$$

We know this converges because this is just a linear combination of the example above the theorem and a simple $p = 2$ integral multiplied by $1/3$. In particular, using our work in the previous example,

$$\begin{aligned}\int_1^\infty \left[\frac{1}{3x^2} + \frac{2e^x}{\sin(x) + e^{2x}} \right] dx &= \frac{1}{3} \int_1^\infty \frac{dx}{x^2} + 2 \int_1^\infty \frac{e^x}{\sin(x) + e^{2x}} dx \\ &= \boxed{\frac{1}{9} + \pi - 2 \tan^{-1}(e)}.\end{aligned}$$

The misuse the linearity theorem is often a source of logical weakness for students. For a bad example:

$$\int_1^\infty \left(\frac{1}{x} - x \right) dx = \int_1^\infty \frac{dx}{x} - \int_1^\infty x dx = -\infty + \infty$$

At which point some students will choose 0 and others will choose $-\infty$, yes others $-\infty/2$. There are many wrong choices, but they all stem from the assumption that you can apply the linearity theorem in the divergent case and it is simply false. We can only split an improper integral into a sum of new improper integrals when the new integrals actually converge. Otherwise, we beg all sorts of questions about how we should go about defining arithmetic for ∞ . We will not engage in such a discussion. Well, I think this is a good start. I hope you learn more as you work the homework.

Example 11.2.22. *By the way, here's a problem I don't know how to solve with our current methods:*

$$\int_1^\infty \frac{dx}{x \sin(x)}.$$

I want to say it is convergent, but, the nonlinearity of the $\frac{1}{x}$ threatens the symmetry of the $\sin(x)$ -induced cancellation. On the other hand, what happens with

$$\int_1^\infty \frac{\sin(x) dx}{x}?$$

Both of these are in some sense on the bubble of the $p = 1$ case which is where many of the most fascinating and challenging examples emerge.

Chapter 12

Curves

In this short chapter we analyze the concept of a curve from the three major viewpoints prevalent in analytic geometry. For now, we use only Cartesian coordinates¹. The three views of a curve are:

1. a graph of a function; $\text{graph}(f) = \{(x, y) \mid y = f(x), x \in \text{dom}(f)\}$
2. a level curve; $C_k = \{(x, y) \mid F(x, y) = k\}$
3. a parametrized curve; $C = \{(x(t), y(t)) \mid t \in J \subseteq \mathbb{R}\}$

These views are not mutually exclusive and each has their advantage and disadvantage. We desire you understand all three in this course. Experiment and question is key, you have to discover these concepts for yourself. I'll tell you what I think, but don't stop with my comments. Think. Ask your own questions.

12.1 graphs

Let's begin by reminding ourselves of the definition of a graph:

Definition 12.1.1. *Graph of a function.*

Let $f : \text{dom}(f) \rightarrow \mathbb{R}$ be a function then

$$\text{graph}(f) = \{(x, f(x)) \mid x \in \text{dom}(f)\}.$$

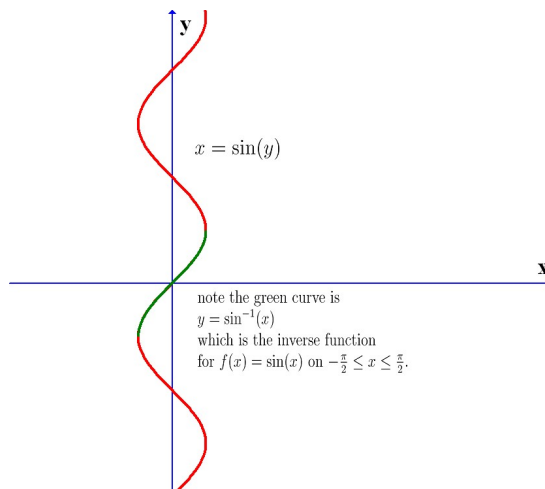
We know this is quite restrictive. We must satisfy the vertical line test if we say our curve is the graph of a function.

Example 12.1.2. *To form a circle centered at the origin of radius R we need to glue together two graphs. In particular we solve the equation $x^2 + y^2 = R^2$ for $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$. Let $f(x) = \sqrt{R^2 - x^2}$ and $g(x) = -\sqrt{R^2 - x^2}$ then we find $\text{graph}(f) \cup \text{graph}(g)$ gives us the whole circle.*

Example 12.1.3. *On the other hand, if we wish to describe the set of all points such that $\sin(y) = x$ we also face a similar difficulty if we insist on functions having independent variable x . Naturally, if we allow for functions with y as the independent variable then $f(y) = \sin(y)$ has graph $\text{graph}(f) = \{(f(y), y) \mid y \in$*

¹all of these constructions find parallel versions when other coordinates such as polar, skew-linear or hyperbolics are used to describe \mathbb{R}^2 .

$\text{dom}(f)\}$. You might wonder, is this correct? I would say a better question is, "is this allowed?". Different books are more or less imaginative about what is permissible as a function. This much we can say, if a shape fails both the vertical and horizontal line tests then it is not the graph of a single function of x or y .



Example 12.1.4. Let $f(x) = mx + b$ for some constants m, b then $y = f(x)$ is the line with slope m and y -intercept b .

12.2 level curves

Level curves are amazing. The full calculus of level curves is only partially appreciated even in calculus III, but trust me, this viewpoint has many advantages as you learn more. For now it's simple enough:

Definition 12.2.1. *Level Curve.*

A level curve is given by a function of two variables $F : \text{dom}(F) \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ and a constant k . In particular, the set of all $(x, y) \in \mathbb{R}^2$ such that $F(x, y) = k$ is called the level-set of F , but more commonly we just say $F(x, y) = k$ is a level curve.

In an algebra class you might have called this the "graph of an equation", but that terminology is dead to us now. For us, it is a level curve. Moreover, for a particular set of points $C \subseteq \mathbb{R}^2$ we can find more than one function F which produces C as a level set. Unlike functions, for a particular curve there is not just one function which returns that curve. This means that it might be important to give both the level-function F and the level k to specify a level curve $F(x, y) = k$.

Example 12.2.2. A circle of radius R centered at the origin is a level curve $F(x, y) = R^2$ where $F(x, y) = x^2 + y^2$. We call F the level function (of two variables).

Example 12.2.3. To describe $\sin(y) = x$ as a level curve we simply write $\sin(y) - x = 0$ and identify the level function is $F(x, y) = \sin(y) - x$ and in this case $k = 0$. Notice, we could just as well say it is the level curve $G(x, y) = 1$ where $G(x, y) = x - \sin(y) + 1$.

Note once more this type of ambiguity is one distinction of the level curve language, in contrast, the graph $\text{graph}(f)$ of a function $y = f(x)$ and the function f are interchangeable. Some mathematicians insist the

rule $x \mapsto f(x)$ defines a function whereas others insist that a function is a set of pairs $(x, f(x))$. I prefer the mapping rule because it's how I think about functions in general whereas the idea of a graph is much less useful in general.

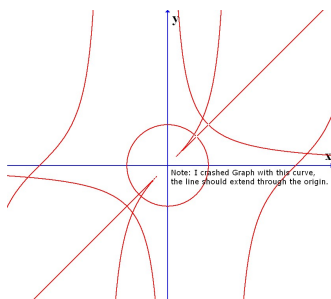
Example 12.2.4. A line with slope m and y -intercept b can be described by $F(x, y) = mx + b - y = 0$. Alternatively, a line with x -intercept x_o and y -intercept y_o can be described as the level curve $G(x, y) = \frac{x}{x_o} + \frac{y}{y_o} = 1$.

Example 12.2.5. Level curves need not be simple things. They can be lots of simple things glued together in one grand equation:

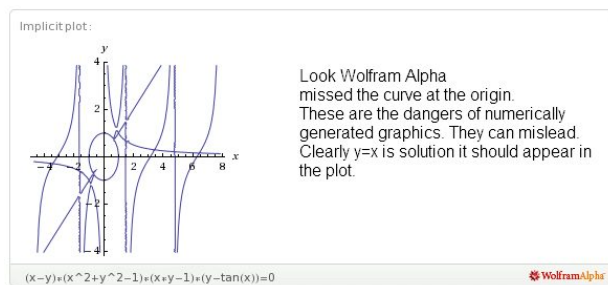
$$F(x, y) = (x - y)(x^2 + y^2 - 1)(xy - 1)(y - \tan(x)) = 0.$$

Solutions to the equation above include the line $y = x$, the unit circle $x^2 + y^2 = 1$, the tilted-hyperbola known more commonly as the reciprocal function $y = \frac{1}{x}$ and finally the graph of the tangent. Some of these intersect, others are disconnected from each other.

It is sometimes helpful to use software to plot equations. However, we must be careful since they are not as reliable as you might suppose. The example above is not too complicated but look what happens with Graph:



Wolfram Alpha shares the same fate:



I hope Mathematica proper fairs better...

Theorem 12.2.6. any graph of a function can be written as a level curve.

If $y = f(x)$ is the graph of a function then we can write $F(x, y) = f(x) - y = 0$ hence the graph $y = f(x)$ is also a level curve.

Not much of a theorem. But, it's true. The converse is not true without a lot of qualification. It's a little harder. Basically, the theorem below gives a criteria for when we can undo a level curve and rewrite it as a single function of x .

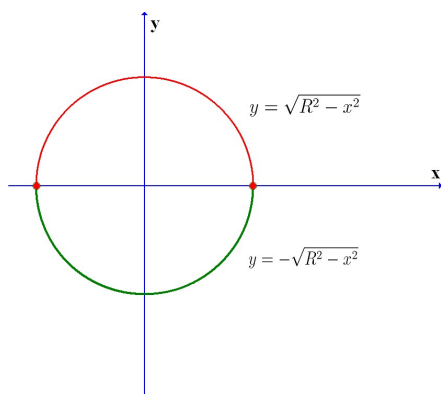
Theorem 12.2.7. *sometimes a level curve can be locally represented as the graph of a function.*

Suppose (x_o, y_o) is a point on the level curve $F(x, y) = k$ hence $F(x_o, y_o) = k$. We say the level curve $F(x, y) = k$ is **locally represented by a function** $y = f(x)$ at (x_o, y_o) iff $F(x, f(x)) = k$ for all $x \in B_\delta(x_o)$ for some $\delta > 0$. Claim: if

$$\frac{\partial F}{\partial y}(x_o, y_o) = \left(\frac{d}{dy} F(x_o, y) \right) \Big|_{y=y_o} \neq 0$$

and the $\frac{\partial F}{\partial y}$ is continuous near (x_o, y_o) then $F(x, y) = k$ is locally represented by some function near (x_o, y_o) .

The theorem above is called the **implicit function theorem** and its proof is nontrivial. Its proper statement is given in Advanced Calculus (Math 332). I'll just illustrate with the circle: $F(x, y) = x^2 + y^2 = R^2$ has $\frac{\partial F}{\partial y} = 2y$ which is continuous everywhere, however at $y = 0$ we have $\frac{\partial F}{\partial y} = 0$ which means the implicit function theorem might fail. On the circle, $y = 0$ when $x = \pm R$ which are precisely the points where we cannot write $y = f(x)$ for just one function. For any other point we may write either $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$ as a local solution of the level curve.



Remark 12.2.8. *finding the formula for a local solution generally a difficult problem.*

The implicit function theorem is an existence theorem. It merely says there exists a solution given a certain criteria, however it does not tell us how to solve the equation to find the formula for the local function. Sometimes we are just content to have an equation which implicitly defines a function of x . For example, most sane creatures do not try to solve $y^5 + y^4 + 3y^2 - y + 1 = x$ for y . Or $\sin(xy) = 3 + y$. When faced with such curves we prefer the level curve description. We've already done some work on this in Calculus I when we did implicit differentiation. The idea of implicit differentiation is only logical when we can write $y = f(x)$, so we must rely on the implicit function theorem as a backdrop to any implicit differentiation problem (see §4.10 if you forgot about implicit differentiation).

12.3 parametrized curves

The idea of a parametrized curve is probably simpler than the definition below appears. In short, we want to take the real number line, or some subset, and paste it into the plane. Think of taking a string and placing it on a table. You can place the string in a great variety of patterns. Imagine the string has little markers placed along it which say $t = 1$, $t = 2$ etc... the value of this label will tell us at which point we are on the string². That label is called the "parameter".

Definition 12.3.1. *Parametrization of a curve.*

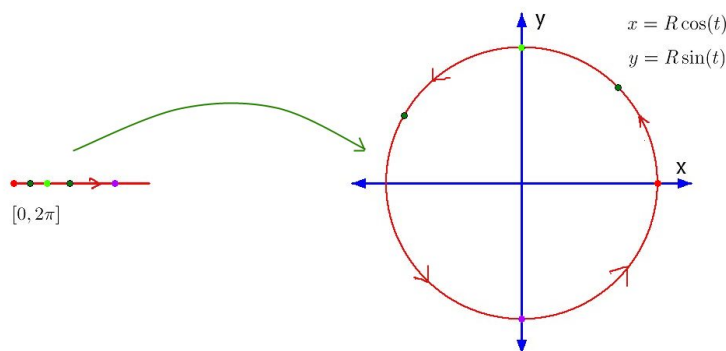
Let C be some curve in the plane. A parametrization of the curve C is a pair of functions $f, g : J \subseteq \mathbb{R} \rightarrow \mathbb{R}$ such that $C = \{(f(t), g(t)) \mid t \in J\}$. In other words, a parametric curve is a mapping from $J \subseteq \mathbb{R} \rightarrow \mathbb{R}^2$ given by the rule $t \mapsto (f(t), g(t))$ for each $t \in J$. We say that t is the parameter and that the parametric equations for the curve are $x = f(t)$ and $y = g(t)$. In the case $J = [a, b]$ we say that $(f(a), g(a))$ is the initial point and $(f(b), g(b))$ is the terminal point. We often use the notation x for f and y for g when it is convenient. Furthermore, the notation t is just one choice for the label of the parameter, we also may use s or λ or other demarcations.

Finding the parametric equations for a curve does require a certain amount of creativity. However, it's almost always some slight twist on the examples I give in this section.

Example 12.3.2. Let $x = R \cos(t)$ and $y = R \sin(t)$ for $t \in [0, 2\pi]$. This is a parametrization of the circle of radius R centered at the origin. We can check this by substituting the equations back into our standard Cartesian equation for the circle:

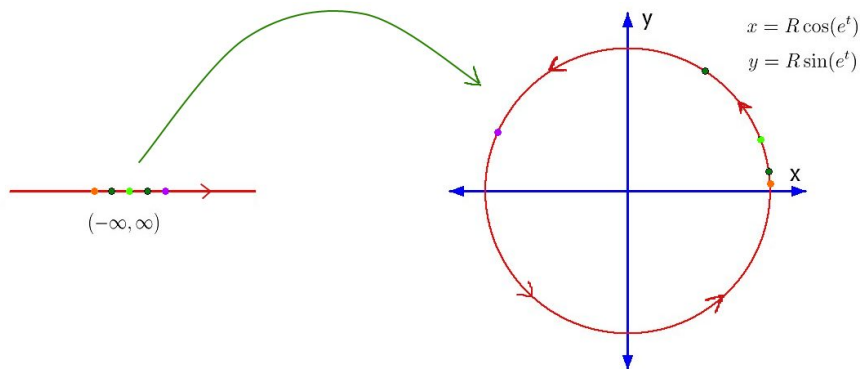
$$x^2 + y^2 = (R \cos(t))^2 + (R \sin(t))^2 = R^2(\cos^2(t) + \sin^2(t))$$

Recall that $\cos^2(t) + \sin^2(t) = 1$ therefore, $x(t)^2 + y(t)^2 = R^2$ for each $t \in [0, 2\pi]$. This shows that the parametric equations do return the set of points which we call a circle of radius R . Moreover, we can identify the parameter in this case as the standard angle from standard polar coordinates.



²the parameter can also be thought of as time as the next chapter discusses at length

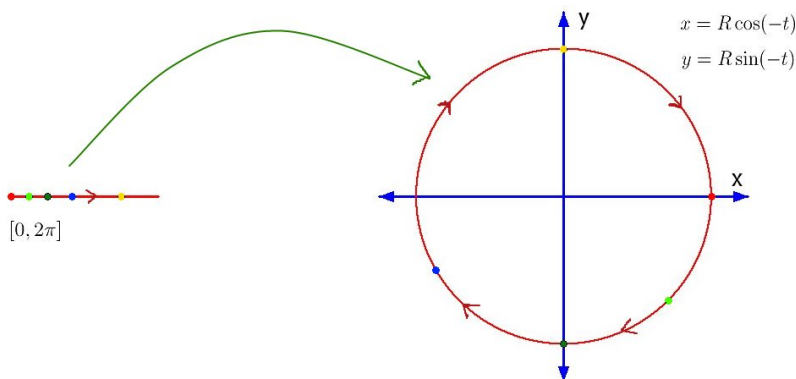
Example 12.3.3. Let $x = R \cos(e^t)$ and $y = R \sin(e^t)$ for $t \in \mathbb{R}$. We again cover the circle at t varies since it is still true that $(R \cos(e^t))^2 + (R \sin(e^t))^2 = R^2(\cos^2(e^t) + \sin^2(e^t)) = R^2$. However, since $\text{range}(e^t) = [1, \infty)$ it is clear that we will actually wrap around the circle infinitely many times. The parametrizations from this example and the last do cover the same set, but they are radically different parametrizations: the last example winds around the circle just once whereas this example winds around the circle ∞ -ly many times.



Example 12.3.4. Let $x = R \cos(-t)$ and $y = R \sin(-t)$ for $t \in [0, 2\pi]$. This is a parametrization of the circle of radius R centered at the origin. We can check this by substituting the equations back into our standard Cartesian equation for the circle:

$$x^2 + y^2 = (R \cos(-t))^2 + (R \sin(-t))^2 = R^2(\cos^2(-t) + \sin^2(-t))$$

Recall that $\cos^2(-t) + \sin^2(-t) = 1$ therefore, $x(t)^2 + y(t)^2 = R^2$ for each $t \in [0, 2\pi]$. This shows that the parametric equations do return the set of points which we call a circle of radius R . Moreover, we can identify the parameter an angle measured CW^3 from the positive x -axis. In contrast, the standard polar coordinate angle is measured CCW from the positive x -axis. Note that in this example we cover the circle just once, but the direction of the curve is opposite that of Example 12.3.2.



The idea of directionality is not at all evident from Cartesian equations for a curve. Given a graph $y = f(x)$ or a level-curve $F(x, y) = k$ there is no intrinsic concept of direction ascribed to the curve. For example, if I

³CW is an abbreviation for ClockWise, whereas CCW is an abbreviation for CounterClockWise.

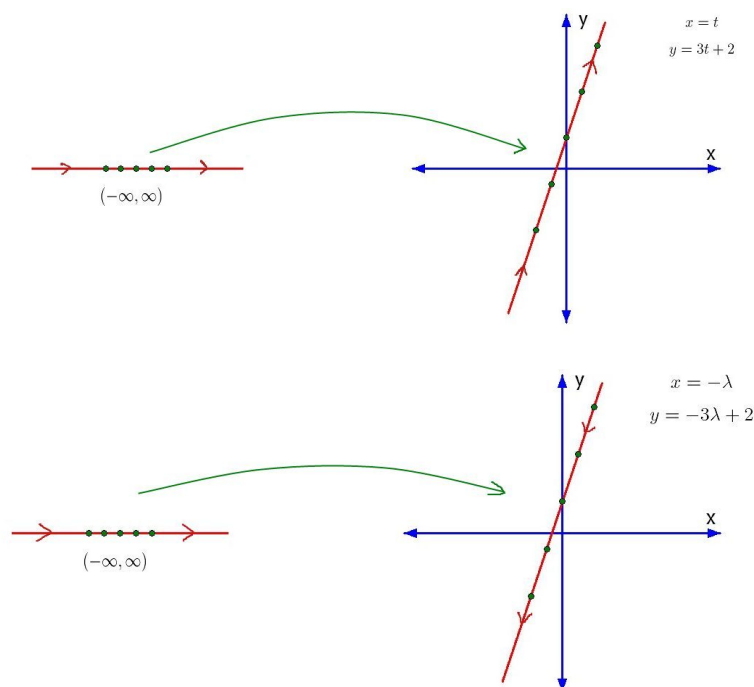
ask you whether $x^2 + y^2 = R^2$ goes CW or CCW then you ought not have an answer. I suppose you could ad-hoc pick a direction, but it wouldn't be natural. This means that if we care about giving a direction to a curve we need the concept of the parametrized curve. We can use the ordering of the real line to induce an ordering on the curve.

Definition 12.3.5. *oriented curve.*

Suppose $f, g : J \subseteq \mathbb{R} \rightarrow \mathbb{R}$ are 1-1 functions. We say the set $\{(f(t), g(t)) \mid t \in J\}$ is an **oriented curve** and say $t \rightarrow (f(t), g(t))$ is a consistently oriented **path** which covers C . If $J = [a, b]$ and $(f(a), g(a)) = p$ and $(f(b), g(b)) = q$ then we can say that C is a curve from p to q .

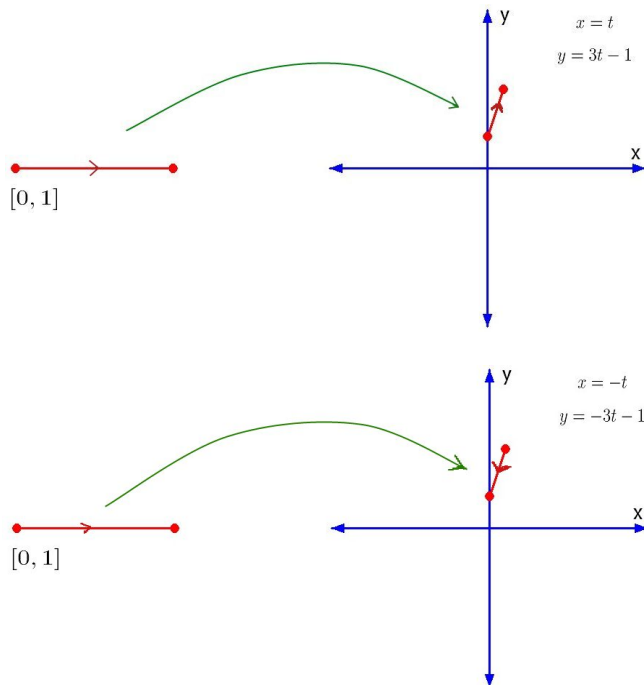
I often illustrate the orientation of a curve by drawing little arrows along the curve to indicate the direction. Furthermore, in my previous definition of parametrization I did not insist the parametric functions were 1-1, this means that those parametrizations could reverse direction and go back and forth along a given curve. What is meant by the terms "path", "curve" and "parametric equations" may differ from text to text so you have to keep a bit of an open mind and try to let context be your guide when ambiguity occurs. I will try to be uniform in my language within this course.

Example 12.3.6. *The line $y = 3x + 2$ can be parametrized by $x = t$ and $y = 3t + 2$ for $t \in \mathbb{R}$. This induces an orientation which goes from left to right for the line. On the other hand, if we use $x = -\lambda$ and $y = -3\lambda + 2$ then as λ increases we travel from right to left on the curve. So the λ -equations give the line the opposite orientation.*



To reverse orientation for $x = f(t), y = g(t)$ for $t \in J = [a, b]$ one may simply replace t by $-t$ in the parametric equations, this gives new equations which will cover the same curve via $x = f(-t), y = g(-t)$ for $t \in [-a, -b]$.

Example 12.3.7. The line-segment from $(0, -1)$ to $(1, 2)$ can be parametrized by $x = t$ and $y = 3t - 1$ for $0 \leq t \leq 1$. On the other hand, the line-segment from $(1, 2)$ to $(0, -1)$ is parametrized by $x = -t, y = -3t - 1$ for $-1 \leq t \leq 0$.



The other method to graph parametric curves is simply to start plugging in values for the parameter and assemble a table of values to plot. I have illustrated that in part by plotting the green dots in the domain of the parameter together with their images on the curve. Those dots are the results of plugging in the parameter to find corresponding values for x, y . I don't find that is a very reliable approach in the same way I find plugging in values to $f(x)$ provides a very good plot of $y = f(x)$. That sort of brute-force approach is more appropriate for a CAS system. There are many excellent tools for plotting parametric curves, hopefully I will have some posted on the course website. In addition, the possibility of animation gives us an even more exciting method for visualization of the time-evolution of a parametric curve. In the next chapter we connect the parametric viewpoint with physics and such an animation actually represents the physical motion of some object. My focus in the remainder of this chapter is almost purely algebraic, I could draw pictures to explain, but I wanted the notes to show you that the pictures are not necessary when you understand the algebraic process. That said, the best approach is to do some combination of algebraic manipulation/figuring and graphical reasoning.

Remark 12.3.8. *Ahhhhhh! what is t ???*

If you are at all like me when I first learned about parametric curves you're probably wondering what is t ? You probably, like me, suppose incorrectly that t should be just like x or y . There is a crucial difference between x and y and t . The notations x and y are actually shorthands for the Cartesian coordinate maps $x : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $y : \mathbb{R}^2 \rightarrow \mathbb{R}$ where $x(a, b) = a$ and $y(a, b) = b$. When I use the notation $x = 3$ then you know what I mean, you know that I'm focusing on the vertical line with first coordinate 3. On the other hand, if I say $t = 3$ and ask where is it? Then you should say, your question doesn't make sense. The concept of t is tied to the curve for which it is the parameter. There are infinitely many geometric meanings for t . In other words, if you try to find t in the xy -plane without regard to a curve then you'll never find an answer. It's a meaningless question.

On the other hand if we are given a curve and ask what the meaning of t is for that curve then we ask a meaningful question. There are two popular meanings.

1. the parameter s measures the arclength from some base point on the given curve.
2. the parameter t gives the time along the curve.

In case (1.) for an oriented curve this actually is uniquely specified if we have a starting point. Such a parameterization is called the **arclength parametrization** or **unit-speed** parametrization of a curve. These play a fundamental role in the study of the differential geometry of curves. In case (2.) we have in mind that the curve represents the physical trajectory of some object, as t increases, time goes on and the object moves. I tend to use (2.) as my conceptual backdrop. But, keep in mind that these are just applications of parametric curves. In general, the parameter need not be time or arclength. It might just be what is suggested by algebraic convenience: that is my primary motivator in this chapter.

12.4 converting to and from the parametric viewpoint

Let's change gears a bit, we've seen that parametric equations for curves give us a new method to describe particular geometric concepts such as orientability or multiple covering. Without the introduction of the parametric concept these geometric ideas are not so easy to describe. That said, I now turn to the question of how to connect parametric descriptions with Cartesian descriptions of a curve. We'd like to understand how to go both ways if possible:

1. how can we find the Cartesian form for a given parametric curve?
2. how can we find a parametrization of a given Cartesian curve?

In case (2.) we mean to include the ideas of level curves and graphs. It turns out that both questions can be quite challenging for certain examples. However, in other cases, not so much: for example any graph $y = f(x)$ is easily recast as the set of parametric equations $x = t$ and $y = f(t)$ for $t \in \text{dom}(f)$. For the standard graph of a function we use x as the parameter.

12.4.1 how can we find the Cartesian form for a given parametric curve?

Example 12.4.1. What curve has parametric equations $x = t$ for $y = t^2$ for $t \in \mathbb{R}$? To find Cartesian equation we eliminate the parameter (when possible)

$$t^2 = x^2 = y \Rightarrow y = x^2$$

Thus the Cartesian form of the given parametrized curve is simply $y = x^2$.

Example 12.4.2. Example 15.2.2: Find parametric equations to describe the graph $y = \sqrt{x+3}$ for $0 \leq x < \infty$. We can use $x = t^2$ and $y = \sqrt{t^2+3}$ for $t \in \mathbb{R}$. Or, we could use $x = \lambda$ and $y = \sqrt{\lambda+3}$ for $\lambda \in [0, \infty)$.

Example 12.4.3. What curve has parametric equations $x = t$ for $y = t^2$ for $t \in [0, 1]$? To find Cartesian equation we eliminate the parameter (when possible)

$$t^2 = x^2 = y \Rightarrow y = x^2$$

Thus the Cartesian form of the given parametrized curve is simply $y = x^2$, however, given that $0 \leq t \leq 1$ and $x = t$ it follows we do not have the whole parabola, instead just $y = x^2$ for $0 \leq x \leq 1$.

Example 12.4.4. Identify what curve has parametric equations $x = \tan^{-1}(t)$ and $y = \tan^{-1}(t)$ for $t \in \mathbb{R}$. Recall that $\text{range}(\tan^{-1}(t)) = (-\pi/2, \pi/2)$. It follows that $-\pi/2 < x < \pi/2$. Naturally we just equate inverse tangent to obtain $\tan^{-1}(t) = y = x$. The curve is the open line-segment with equation $y = x$ for $-\pi/2 < x < \pi/2$. This is an interesting parameterization, notice that as $t \rightarrow \infty$ we approach the point $(\pi/2, \pi/2)$, but we never quite get there.

Example 12.4.5. Consider $x = \ln(t)$ and $y = e^t - 1$ for $t \geq 1$. We can solve both for t to obtain

$$t = e^x = \ln(y+1) \Rightarrow y = -1 + \exp(\exp(x)).$$

The domain for the expression above is revealed by analyzing $x = \ln(t)$ for $t \geq 1$, the image of $[1, \infty)$ under natural log is precisely $[0, \infty)$; $\ln[1, \infty) = [0, \infty)$.

Example 12.4.6. Suppose $x = \cosh(t) - 1$ and $y = 2\sinh(t) + 3$ for $t \in \mathbb{R}$. To eliminate t it helps to take an indirect approach. We recall the most important identity for the hyperbolic sine and cosine: $\cosh^2(t) - \sinh^2(t) = 1$. Solve for hyperbolic cosine; $\cosh(t) = x + 1$. Solve for hyperbolic sine; $\sinh(t) = \frac{y-3}{2}$. Now put these together via the identity:

$$\cosh^2(t) - \sinh^2(t) = 1 \Rightarrow (x+1)^2 - \frac{(y-3)^2}{4} = 1.$$

Note that $\cosh(t) \geq 1$ hence $x+1 \geq 1$ thus $x \geq 0$ for the curve described above. On the other hand y is free to range over all of \mathbb{R} since hyperbolic sine has range \mathbb{R} . You should⁴ recognize the equation as a hyperbola centered at $(-1, 3)$.

⁴many students need to review these at this point, we use circles, ellipses and hyperbolas as examples in this course. I'll give examples of each in this chapter.

12.4.2 how can we find a parametrization of a given Cartesian curve?

I like this topic more, the preceding bunch of examples, while needed, are boring. The art of parameterizing level curves is much more fun.

Example 12.4.7. Find parametric equations for the circle centered at (h, k) with radius R .

To begin recall the equation for such a circle is $(x - h)^2 + (y - k)^2 = R^2$. Our inspiration is the identity $\cos^2(t) + \sin^2(t) = 1$. Let $x - h = R \cos(t)$ and $y - k = R \sin(t)$ thus

$$\boxed{x = h + R \cos(t)} \quad \text{and} \quad \boxed{y = k + R \sin(t)}.$$

I invite the reader to verify these do indeed parametrize the circle by explicitly plugging in the equations into the circle equation. Notice, if we want the whole circle then we simply choose any interval for t of length 2π or longer. On the other hand, if you want to select just a part of the circle you need to think about where sine and cosine are positive and negative. For example, if I want to parametrize just the part of the circle for which $x > h$ then I would choose $t \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

The reason I choose that intuitively is that the parametrization given for the circle above is basically built from polar coordinates⁵ centered at (h, k) . That said, to be sure about my choice of parameter domain I like to actually plug in some of my proposed domain and make sure it matches the desired criteria. I think about the graphs of sine and cosine to double check my logic. I know that $\cos(-\frac{\pi}{2}, \frac{\pi}{2}) = (0, 1]$ whereas $\sin(-\frac{\pi}{2}, \frac{\pi}{2}) = (-1, 1)$, I see it in my mind. Then I think about the parametric equations in view of those facts,

$$x = h + R \cos(t) \quad \text{and} \quad y = k + R \sin(t).$$

I see that x will range over $(h, h + R]$ and y will range over $(k - R, k + R)$. This is exactly what I should expect geometrically for half of the circle. Visualize that $x = h$ is a vertical line which cuts our circle in half. These are the thoughts I think to make certain my creative leaps are correct. I would encourage you to think about these matters. Don't try to just memorize everything, it will not work for you, there are simply too many cases. It's actually way easier to just understand these as a consequence of trigonometry, algebra and analytic geometry.

Example 12.4.8. Find parametric equations for the level curve $x^2 + 2x + \frac{1}{4}y^2 = 0$ which give the ellipse a CW orientation.

To begin we complete the square to understand the equation:

$$x^2 + 2x + \frac{1}{4}y^2 = 0 \Rightarrow (x + 1)^2 + \frac{1}{4}y^2 = 1.$$

We identify this is an ellipse centered at $(-1, 0)$. Again, I use the pythagorean trig. identity as my guide: I want $(x + 1)^2 = \cos^2(t)$ and $\frac{1}{4}y^2 = \sin^2(t)$ because that will force the parametric equations to solve the ellipse equation. However, I would like for the equations to describe CW direction so I replace the t with $-t$ and propose:

$$\boxed{x = -1 + \cos(-t)} \quad \text{and} \quad \boxed{y = 2 \sin(-t)}$$

If we choose $t \in [0, 2\pi)$ then the whole ellipse will be covered. I could simplify $\cos(-t) = \cos(t)$ and $\sin(-t) = -\sin(t)$ but I have left the minus to emphasize the idea about reversing the orientation. In the preceding example we gave the circle a CCW orientation.

⁵we will discuss further in a later section, but this should have been covered in at least your precalculus course.

Example 12.4.9. Find parametric equations for the part of the level curve $x^2 - y^2 = 1$ which is found in the first quadrant.

We recognize this is a hyperbola which opens horizontally since $x = 0$ gives us $-y^2 = 1$ which has no real solutions. Hyperbolic trig. functions are built for a problem just such as this: recall $\cosh^2(t) - \sinh^2(t) = 1$ thus we choose $x = \cosh(t)$ and $y = \sinh(t)$. Furthermore, the hyperbolic sine function $\sinh(t) = \frac{1}{2}(e^t - e^{-t})$ is everywhere increasing since it has derivative $\cosh(t)$ which is everywhere positive. Moreover, since $\sinh(0) = 0$ we see that $\sinh(t) \geq 0$ for $t \geq 0$. Choose non-negative t for the domain of the parametrization:

$$\boxed{x = \cosh(t), \quad y = \sinh(t), \quad t \in [0, \infty).}$$

Example 12.4.10. Find parametric equations for the part of the level curve $x^2 - y^2 = 1$ which is found in the third quadrant.

Based on our thinking from the last example we just need to modify the solution a bit:

$$\boxed{x = -\cosh(t), \quad y = \sinh(t), \quad t \in (-\infty, 0].}$$

Note that if $t \in (-\infty, 0]$ then $-\cosh(t) \leq -1$ and $\sinh(t) \leq 0$, this puts us in the third quadrant. It is also clear that these parametric equations solve the hyperbola equation since

$$(-\cosh(t))^2 - (\sinh(t))^2 = \cosh^2(t) - \sinh^2(t) = 1.$$

The examples thus far are rather specialized, and in general there is no method to find parametric equations. This is why I said it is an art.

Example 12.4.11. Find parametric equations for the level curve $x^2 y^2 = x - 2$.

This example is actually pretty easy because we can solve for $y^2 = \frac{x-2}{x^2}$ hence $y = \pm \sqrt{\frac{x-2}{x^2}}$. We can choose x as parameter so the parametric equations are just

$$x = t \quad \text{and} \quad y = \sqrt{\frac{t-2}{t^2}}$$

for $t \geq 2$. Or, we could give parametric equations

$$x = t \quad \text{and} \quad y = -\sqrt{\frac{t-2}{t^2}}$$

for $t \geq 2$. These parametrizations simply cover different parts of the same level curve.

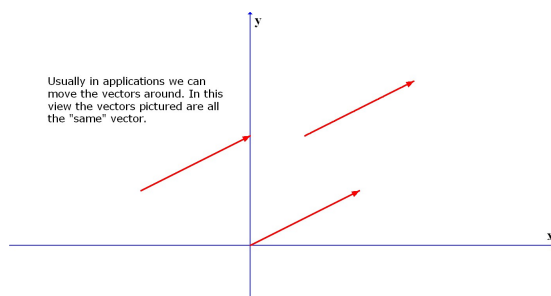
Chapter 13

Calculus of Parametrized Curves

In the preceding chapter my intention was to introduce you to the basic mathematical structure of parametrized curves and argue for why they are mathematically necessary. In this chapter we take the next step: we learn what we can do with parametrized curves in the plane. In particular, we study the physics of motion from this viewpoint. Parametric equations are the natural choice for physics because in physics we gain a meaning for the parameter. In classical nonrelativistic physics the parameter is time. The nice thing about the parametric viewpoint is that there is no preference between x or y . Motion in the x -direction and motion in the y -direction are both covered by the same calculus. To fully appreciate this formalism it helps to think of the x and y equations together as a single object. This is precisely what the vector formalism does for us. In addition, vectors play an important role in both physics and calculus III. In this course we will work primarily with two-dimensional vectors. The notation I'll use is as follows:

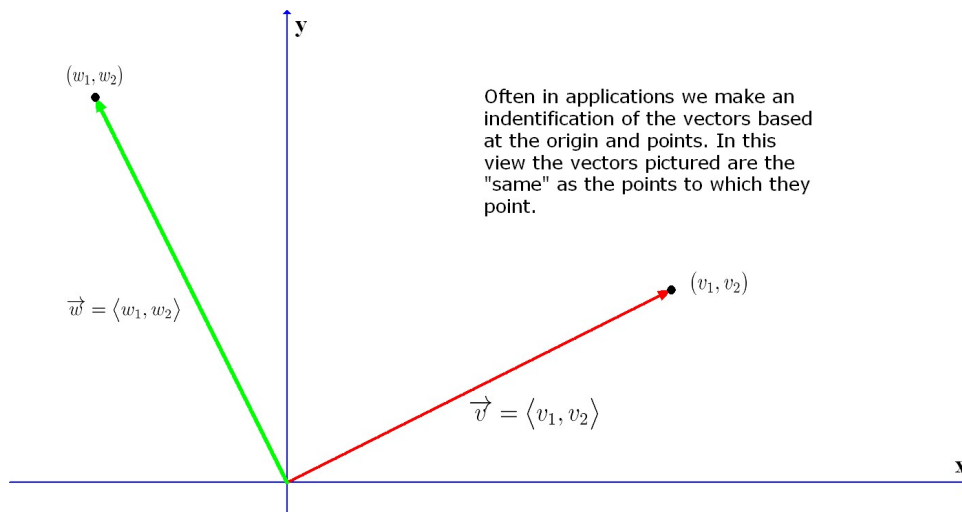
1. (a, b) is a point in \mathbb{R}^2
2. $\langle a, b \rangle$ is a vector in \mathbb{R}^2

A vector is a directed line segment whereas a point is a point. Vectors you can move around the plane whereas a point is stuck.



These are our **conventions**¹.

¹often points and vectors are identified since they are in 1-1 correspondance. Moreover, you could think about vectors which are stuck in place. The conventions I adopt here are simply to avoid a multiplication of notation.



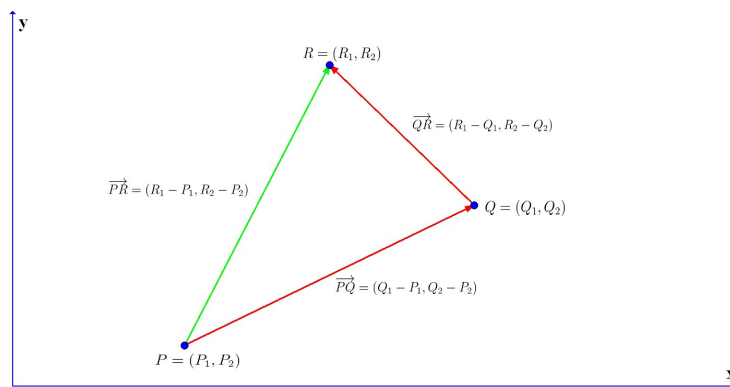
13.1 vectors in a nutshell

Definition 13.1.1. *vector terminology.*

Given two points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ we can construct the vector from p to q by

$$\vec{pq} = \langle q_1 - p_1, q_2 - p_2 \rangle.$$

We call directed line segments **vectors** and denote the set of all such objects by V , we also write a little arrow over such objects to emphasize that they are vectors. If $\vec{v} \in V$ then we can write the vector in Cartesian form $\vec{v} = \langle v_1, v_2 \rangle$. We say v_1 is the x -component of \vec{v} and v_2 is the y -component of \vec{v} . Two vectors are equal iff both the x and y components are equal.



The main thing to remember is that a vector has both a length and a direction. However, it is often more convenient to think of the vector in terms of its Cartesian components. The definition that follows tells us how to work with vectors. They're really quite simple:

Definition 13.1.2. *vector addition, subtraction, lengths and standard angles.*

Suppose $\langle v_1, v_2 \rangle, \langle w_1, w_2 \rangle \in V$ and $c \in \mathbb{R}$ we define

1. addition of vectors is done component-wise; $\langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle = \langle v_1 + w_1, v_2 + w_2 \rangle$.
2. subtraction of vectors is done component-wise; $\langle v_1, v_2 \rangle - \langle w_1, w_2 \rangle = \langle v_1 - w_1, v_2 - w_2 \rangle$.
3. scalar multiplication is defined by; $c \langle v_1, v_2 \rangle = \langle cv_1, cv_2 \rangle$.
4. the length of a vector is defined to be

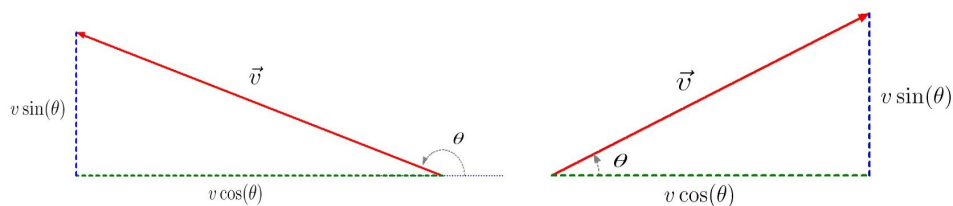
$$\| \langle v_1, v_2 \rangle \| = \sqrt{(v_1)^2 + (v_2)^2}.$$

Moreover, if $\vec{v} = \langle v_1, v_2 \rangle$ then we also write $v = \|\vec{v}\|$. If $\vec{v} \neq \langle 0, 0 \rangle$ then the unit vector in the \vec{v} -direction is a vector of length one denoted with a hat: $\hat{v} = \frac{1}{v}\vec{v}$. We can always write $\vec{v} = v\hat{v}$ for a nonzero vector. In words, v is the **magnitude** and \hat{v} is the **direction** vector.

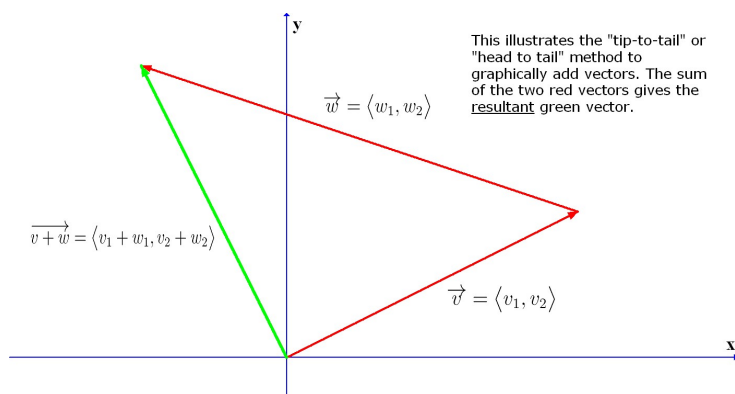
5. the standard angle of a nonzero vector $\langle v_1, v_2 \rangle$ is an angle θ such that

$$v_1 = v \cos(\theta) \quad \text{and} \quad v_2 = v \sin(\theta).$$

I have made a point in (5.) to help you avoid the pitfalls of carelessly solving $\tan(\theta) = \frac{v_2}{v_1}$ for $\theta = \tan^{-1}\left(\frac{v_2}{v_1}\right)$. This formula is only correct for angles that point in quadrants I and IV whereas (5.) needs no qualification.



The definition for vector addition is also very natural if you look at the picture below:



Example 13.1.3. Let $\vec{v} = \langle 1, 2 \rangle$ and $\vec{w} = \langle 3, 4 \rangle$ then we can calculate:

$$\begin{aligned}\vec{v} + \vec{w} &= \langle 1 + 3, 2 + 4 \rangle = \langle 4, 6 \rangle \\ \vec{v} - \vec{w} &= \langle 1 - 3, 2 - 4 \rangle = \langle -2, -2 \rangle = -2 \langle 1, 1 \rangle \\ 2\vec{v} - 3\vec{w} &= 2 \langle 1, 2 \rangle - 3 \langle 3, 4 \rangle = \langle 2, 4 \rangle - \langle 9, 12 \rangle = \langle -7, -8 \rangle\end{aligned}$$

Note that $v = \sqrt{1+4} = \sqrt{5}$ and $w = \sqrt{9+16} = 5$. However, the length of $\vec{v} + \vec{w}$ is not $v + w$. Instead, $\|\vec{v} + \vec{w}\| = \|\langle 4, 6 \rangle\| = \sqrt{16+36} = \sqrt{52}$.

Example 13.1.4. Find a vector of length 6 in the same direction as $\vec{v} = \langle 1, 1 \rangle$. Let's call the vector we're trying to find \vec{w} . If \vec{w} points in the same direction as \vec{v} then there exists some constant $k > 0$ such that $\vec{w} = k\vec{v} = \langle k, k \rangle$. We also want $w = 6$ thus $6 = \sqrt{k^2 + k^2} = k\sqrt{2}$. Therefore, $\vec{w} = \frac{6}{\sqrt{2}} \langle 1, 1 \rangle$.

Example 13.1.5. The standard angle for $\langle -1, -1 \rangle$ is an angle θ such that $-1 = \sqrt{2}\cos(\theta)$ and $-1 = \sqrt{2}\sin(\theta)$. We must solve $\cos(\theta) = -1/\sqrt{2}$ and $\sin(\theta) = -1/\sqrt{2}$. The solution is $\theta = \frac{3\pi}{4}$. (in quadrant III)

Note $\tan^{-1}(-1/-1) = \frac{\pi}{4}$ is not the correct angle in the preceding example. You have to think about both components when determining standard angle. The inverse tangent will mislead you unless you really understand what you are doing.

Example 13.1.6. The standard angle for $\langle -\sqrt{3}, 1 \rangle$ is an angle θ such that $-\sqrt{3} = 2\cos(\theta)$ and $1 = 2\sin(\theta)$. We must solve $\cos(\theta) = -2/\sqrt{3}$ and $\sin(\theta) = 1/2$. The solution is $\theta = \frac{5\pi}{6}$. (in quadrant II)

Once more, the inverse tangent fails us since $\tan^{-1}(-1/\sqrt{3}) = \frac{-\pi}{6}$. To use inverse tangent correctly you must adjust answers when they ought to be in quadrants II and III.

13.2 calculus for vector-valued functions

A vector valued function of a real variable is an assignment of of vector for each real number in some domain. It's a mapping $t \mapsto \langle f(t), g(t) \rangle$ for each $t \in J \subset \mathbb{R}$.

Definition 13.2.1. *vector-valued functions.*

We say \vec{v} is a vector-valued function of a real variable if $\vec{v} : J \subseteq \mathbb{R} \rightarrow V$ is a function. In addition if $\vec{v}(t) = \langle v_1(t), v_2(t) \rangle$ then we say v_1 is the x -component function of \vec{v} and v_2 is the y -component function of \vec{v} .

If our vectors are all based from the origin outward then we can write $V = \mathbb{R}^2$. Think about it: each point is naturally identified with the vector which points from the origin out to the point. The identification amounts to saying that $\langle a, b \rangle = (a, b)$ in our current notation. With this identification in mind we can write parametric curves in nice vector notation:

Example 13.2.2. Recall the circle centered at the origin of radius R had parametric equations $x = R\cos(t)$ and $y = R\sin(t)$ for $t \in [0, 2\pi]$. We can group these together and simply say that the circle is parametrized by $\vec{r}(t) = \langle R\cos(t), R\sin(t) \rangle$.

One is often asked to find the parametrization of a line-segment from a point p to a point q . I recommend the following approach: for $0 \leq t \leq 1$ let

$$\vec{r}(t) = p(1 - t) + tq.$$

It's easy to calculate $\vec{r}(0) = p$ and $\vec{r}(1) = q$. This formula can also be written as

$$\vec{r}(t) = p + t(q - p) = p + t [\vec{pq}].$$

Example 13.2.3. Find the parametrization of a line segment which goes from $(1, 3)$ to $(5, 2)$. We use the comment preceding this example and construct:

$$\vec{r}(t) = (1, 3) + t[(5, 2) - (1, 3)] = \langle 1 + 4t, 3 - t \rangle$$

And now for the calculus. In short, we just do everything component by component.

Definition 13.2.4. calculus of vector-valued functions.

We say \vec{v} is a vector-valued function of a real variable where $\vec{v}(t) = \langle v_1(t), v_2(t) \rangle$ then

1. If v_1 and v_2 are both differentiable functions we define

$$\frac{d\vec{v}}{dt} = \frac{d}{dt} \langle v_1, v_2 \rangle = \left\langle \frac{dv_1}{dt}, \frac{dv_2}{dt} \right\rangle.$$

2. If v_1 and v_2 are both integrable functions on $[a, b]$ then we define

$$\int_a^b \vec{v} dt = \int_a^b \langle v_1, v_2 \rangle dt = \left\langle \int_a^b v_1(t) dt, \int_a^b v_2(t) dt \right\rangle.$$

3. We write $\int \vec{v}(t) dt = \vec{V} + \vec{c}$ iff $\vec{V} = \langle V_1, V_2 \rangle$ and $\vec{c} = \langle c_1, c_2 \rangle$ where $\int v_1(t) dt = V_1 + c_1$ and $\int v_2(t) dt = V_2 + c_2$.

We also use the prime notation for differentiation of vector valued functions if it is convenient; this means $\vec{r}'(t) = d\vec{r}/dt = \frac{d\vec{r}}{dt}$.

Example 13.2.5. Let $\vec{r}(t) = \langle t, \cos(t) \rangle$.

$$\frac{d\vec{r}}{dt} = \left\langle \frac{d}{dt}(t), \frac{d}{dt}(\cos(t)) \right\rangle = \langle 1, -\sin(t) \rangle.$$

$$\int \vec{r}(t) dt = \left\langle \int t dt, \int \cos(t) dt \right\rangle = \left\langle \frac{1}{2}t^2 + c_1, \sin(t) + c_2 \right\rangle.$$

$$\int_0^1 \vec{r}(t) dt = \left\langle \int_0^1 t dt, \int_0^1 \cos(t) dt \right\rangle = \left\langle \frac{1}{2}, \sin(1) \right\rangle.$$

Often a vector-valued function will be called a space curve since its image is a curve in space.

Theorem 13.2.6. *fundamental theorems of calculus for space curves.*

$$(I.) \quad \frac{d}{dt} \int_a^t \vec{r}(\tau) d\tau = \vec{r}(t)$$

$$(II.) \quad \int_a^b \frac{d\vec{r}}{dt} dt = \vec{r}(b) - \vec{r}(a)$$

Proof: Apply the FTC part I and II componentwise to arrive at the corresponding theorems above. \square

Theorem 13.2.7. *rules of calculus for space curves.*

Let $\vec{v}, \vec{w} : J \subseteq \mathbb{R} \rightarrow \mathbb{R}^2$ and $f : J \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be differentiable functions and let $c \in \mathbb{R}$

$$(1.) \quad \frac{d}{dt} [\vec{v} + \vec{w}] = \frac{d\vec{v}}{dt} + \frac{d\vec{w}}{dt}$$

$$(2.) \quad \frac{d}{dt} [f\vec{v}] = \frac{df}{dt} \vec{v} + f \frac{d\vec{v}}{dt}$$

$$(3.) \quad \frac{d}{dt} [c\vec{v}] = c \frac{d\vec{v}}{dt}$$

$$(4.) \quad \int [\vec{v} + \vec{w}] dt = \int \vec{v} dt + \int \vec{w} dt$$

$$(5.) \quad \int c\vec{v} dt = c \int \vec{v} dt$$

The proof of the theorem above is easily derived by simply expanding what the vector notation means and borrowing the corresponding theorems from calculus I to simplify the component expressions. I might ask for this in homework so I'll not offer details here.

13.3 position, velocity and acceleration in two dimensions

In this section the notations \vec{r} , \vec{v} and \vec{a} is special and set-apart. I don't use these as abstract variables here with no set meaning. Instead, these are connected as is described in the definition that follows:

Definition 13.3.1. *position, velocity and acceleration.*

The position, velocity and acceleration of an object are vector-valued functions of time and we define them as follows:

1. $\vec{r}(t)$ is the position at time t .
2. $\vec{v}(t) = \frac{d\vec{r}}{dt}$ is the velocity at time t .
3. $\vec{a}(t) = \frac{d\vec{v}}{dt}$ is the acceleration at time t .

If we are given the position vector as a function of time then we need only differentiate to find the velocity and acceleration. On the other hand, if we are given the acceleration then we need to integrate and apply initial conditions to obtain the velocity and position.

Example 13.3.2. Suppose R and ω are positive constants and the motion of an object is observed to follow the path $\vec{r}(t) = \langle R \cos(\omega t), R \sin(\omega t) \rangle = R \langle \cos(\omega t), \sin(\omega t) \rangle$. We wish to calculate the velocity and acceleration as functions of time.

Differentiate to obtain the velocity

$$\vec{v}(t) = R\omega \langle -\sin(\omega t), \cos(\omega t) \rangle .$$

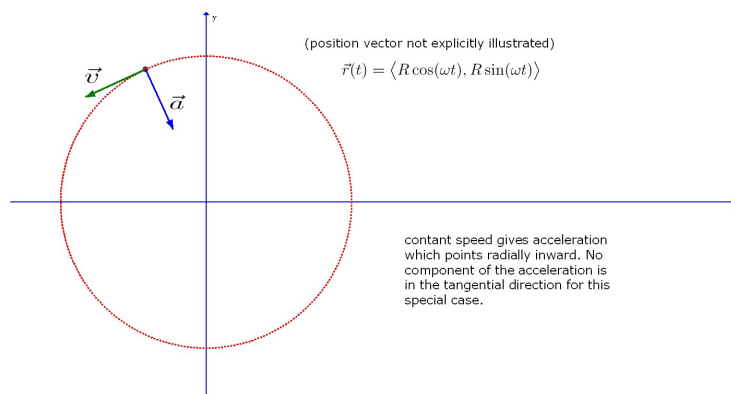
Differentiate once more to obtain the acceleration:

$$\vec{a}(t) = R\omega \langle -\omega \cos(\omega t), -\omega \sin(\omega t) \rangle = -R\omega^2 \langle \cos(\omega t), \sin(\omega t) \rangle .$$

Notice we can write that $\vec{a}(t) = -\omega^2 \vec{r}(t)$ in this very special example. This means the acceleration is opposite the direction of the position. Furthermore, we can calculate

$$r = R, \quad v = R\omega, \quad a = R\omega^2$$

Thus the magnitudes of the position, velocity and acceleration are all constant. However, their directions are always changing. Perhaps you recognize these equations as the foundational equations describing constant speed circular motion. This acceleration is called the **centripetal** or center-seeking acceleration since it points towards the center. Here we imagine attaching the acceleration vector to the object which is traveling in the circle.



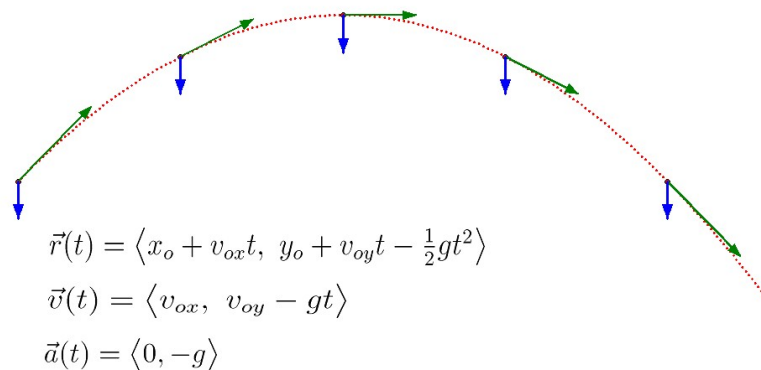
Example 13.3.3. Suppose that the acceleration of an object is known to be $\vec{a} = \langle 0, -g \rangle$ where g is a positive constant. Furthermore, suppose that initially the object is at \vec{r}_o and has velocity \vec{v}_o . We wish to calculate the position and velocity as functions of time.

Integrate the acceleration from 0 to t ,

$$\int_0^t \frac{d\vec{v}}{d\tau} d\tau = \int_0^t \vec{a}(\tau) d\tau \Rightarrow \vec{v}(t) - \vec{v}(0) = \int_0^t \langle 0, -g \rangle d\tau \Rightarrow \boxed{\vec{v}(t) = \vec{v}_o + \langle 0, -gt \rangle}$$

Integrate the velocity from 0 to t ,

$$\int_0^t \frac{d\vec{r}}{d\tau} d\tau = \int_0^t \vec{v}(\tau) d\tau \Rightarrow \vec{r}(t) - \vec{r}(0) = \int_0^t (\vec{v}_o + \langle 0, -gt \rangle) d\tau \Rightarrow \boxed{\vec{r}(t) = \vec{r}_o + t\vec{v}_o + \langle 0, -\frac{1}{2}gt^2 \rangle}$$



The acceleration is constant for this parabolic trajectory.
The velocity is changing in the vertical direction, but is constant in the x-direction.

The best understanding of Newtonian Mechanics is given by a combination of both vectors and calculus. We need vectors to phrase the geometry of force addition whereas we need calculus to understand how the position, velocity and acceleration variables change in concert. In calculus II we don't delve too deeply into the geometry of vectors. To do it properly I would also need to discuss dot and cross products and three dimensional geometry. In addition, we will consider a few problems of physics, but the set-up will be fairly

predictable and the problems we work on tests will be like those we have already solved. In contrast, I make no such promise in a physics course. In physics we must understand the concepts of physical law. Here our focus is more towards how calculus is used in physics.

13.3.1 position vs. displacement vs. distance traveled

The position of an object is simply the (x, y) coordinates of the object. Usually it is convenient to think of the position as a vector-valued function of time which we denote $\vec{r}(t)$. The displacement is also a vector, however it compares two possibly distinct positions:

Definition 13.3.4. *displacement and distance traveled.*

Suppose $\vec{r}(t)$ is the position at time t of some object.

1. The **displacement** from position \vec{r}_1 to position \vec{r}_2 is the vector $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$.
2. The **distance travelled** during the interval $[t_1, t_2]$ along the curve $t \mapsto \vec{r}(t)$ is given by

$$s_{12} = \int_{t_1}^{t_2} v(t) dt = \int_{t_1}^{t_2} \sqrt{\frac{dx^2}{dt} + \frac{dy^2}{dt}} dt$$

where $v(t) = ||d\vec{r}/dt||$.

Note that the position is the displacement from the origin. Distance travelled is a **scalar** quantity which means it is just a number or if we think of an endpoint as variable it could be a function.

Definition 13.3.5. *arclength function and speed.*

We define

$$s(t) = \int_{t_1}^t v(\tau) d\tau = \int_{t_1}^t \sqrt{\frac{dx^2}{d\tau} + \frac{dy^2}{d\tau}} d\tau$$

to be the arclength travelled from time t_1 to t along the parametrized curve $t \mapsto \vec{r}(t)$. Furthermore, we define the **speed** to be the instantaneous rate of change in the arclength; speed is ds/dt .

Notice it is simple to show that the speed is also equal to the magnitude of the velocity; $ds/dt = v$. We'll explore the geometric foundation of the arclength² formula in the next section. For now let's examine a few examples:

Example 13.3.6. Let $\omega, R > 0$. Suppose $\vec{r}(t) = \langle R \cos(\omega t), R \sin(\omega t) \rangle$ for $t \geq 0$. We can calculate that

$$\frac{d\vec{r}}{dt} = \langle -R\omega \sin(\omega t), R\omega \cos(\omega t) \rangle \Rightarrow v(t) = \sqrt{(-R\omega \sin(\omega t))^2 + (R\omega \cos(\omega t))^2} = \sqrt{R^2\omega^2} = R\omega.$$

Now use this to help calculate the distance travelled during the interval $[0, t]$

$$s(t) = \int_0^t v(\tau) d\tau = \int_0^t R\omega d\tau = R\omega\tau \Big|_0^t = R\omega t.$$

In other words, $\Delta s = R\omega\Delta t$. On a circle the arclength subtended Δs divided by the radius R is defined to be the radian measure of that arc which we typically denote $\Delta\theta$. We find that $\Delta\theta = \omega\Delta t$ or if you prefer $\omega = \Delta\theta/\Delta t$.

²or distance travelled formula if you insist the parameter be time

Circular motion which is not at a constant speed can be obtained mathematically by replacing the constant ω with a function of time. Let's examine such an example.

Example 13.3.7. Suppose $\vec{r}(t) = \langle R \cos(\theta), R \sin(\theta) \rangle$ for $t \geq 0$ where $\theta_o, \omega_o, \alpha$ are constants and $\theta = \theta_o + \omega_o t + \frac{1}{2} \alpha t^2$. To calculate the distance travelled it helps to first calculate the velocity:

$$\frac{d\vec{r}}{dt} = \langle -R(\omega_o + \alpha t) \sin(\theta), R(\omega_o + \alpha t) \cos(\theta) \rangle$$

Next, the speed is the length of the velocity vector,

$$v = \sqrt{[-R(\omega_o + \alpha t) \sin(\theta)]^2 + [R(\omega_o + \alpha t) \cos(\theta)]^2} = R\sqrt{(\omega_o + \alpha t)^2} = R|\omega_o + \alpha t|.$$

Therefore, the distance travelled is given by the integral below:

$$s(t) = \int_0^t R|\omega_o + \alpha \tau| d\tau$$

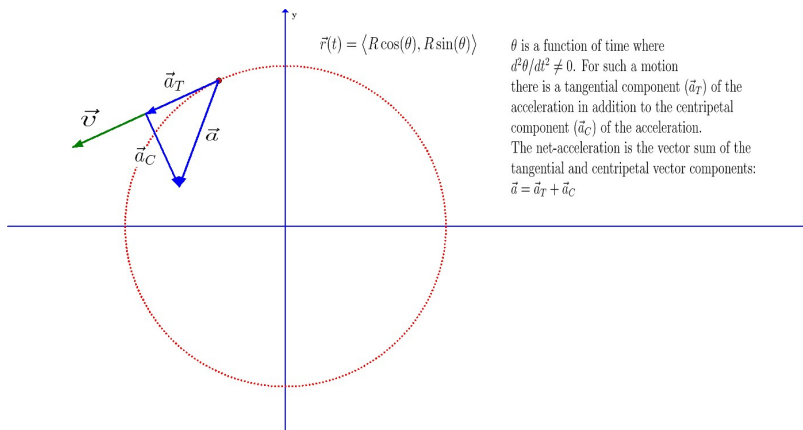
To keep things simple, let's suppose that ω_o, α are given such that $\omega_o + \alpha t \geq 0$ hence $v = R\omega_o + R\alpha t$. To suppose otherwise would indicate the motion came to a stopping point and reversed direction, which is interesting, just not to us here.

$$s(t) = R \int_0^t (\omega_o + \alpha \tau) d\tau = R\omega_o t + \frac{1}{2} R\alpha t^2.$$

Observe that $\theta(t) - \theta_o = (s(t) - s(0))/R$ thus we find that $\Delta\theta = \omega_o t + \frac{1}{2} \alpha t^2$ which is the formula for the angle subtended due to motion at a constant **angular acceleration** α . In invite the reader to differentiate the position twice and show that

$$\vec{a}(t) = \frac{d^2 \vec{r}}{dt^2} = \underbrace{-R\omega^2 \langle \cos(\theta(t)), \sin(\theta(t)) \rangle}_{\text{centripetal}} + \underbrace{R\alpha \langle -\sin(\theta(t)), \cos(\theta(t)) \rangle}_{\text{tangential}}$$

where $\omega = \omega_o + \alpha t$.



Distance travelled is not always something we can calculate in closed form. Sometimes we need to relegate the calculation of the arclength integral to a numerical method. However, the example that follows is still calculable without numerical assistance. It did require some thought.

Example 13.3.8. We found that $\vec{a} = \langle 0, -g \rangle$ twice integrated yields a position of $\vec{r}(t) = \vec{r}_o + t\vec{v}_o + \langle 0, -\frac{1}{2}gt^2 \rangle$ for some constant vectors $\vec{r}_o = \langle x_o, y_o \rangle$ and $\vec{v}_o = \langle v_{ox}, v_{oy} \rangle$. Thus,

$$\vec{r}(t) = \langle x_o + v_{ox}t, y_o + v_{oy}t - \frac{1}{2}gt^2 \rangle$$

From which we can differentiate to derive the velocity,

$$\vec{v}(t) = \langle v_{ox}, v_{oy} - gt \rangle.$$

If you've had any course in physics, or just a proper science education, you should be happy to observe that the zero-acceleration in the x -direction gives rise to constant-velocity motion in the x -direction whereas the gravitational acceleration in the y -direction makes the object fall back down as a consequence of gravity. If you think about $v_{oy} - gt$ it will be negative for some $t > 0$ whatever the initial velocity v_{oy} happens to be, this point where $v_{oy} - gt = 0$ is the turning point in the flight of the object and it gives the top of the parabolic³ trajectory which is parametrized by $t \rightarrow \vec{r}(t)$. Suppose $x_o = y_o = 0$ and calculate the distance travelled from time $t = 0$ to time $t_1 = v_{oy}/g$. Additionally, let us assume $v_{ox}, v_{oy} \geq 0$.

$$\begin{aligned} s &= \int_0^{t_1} v(t) dt = \int_0^{t_1} \sqrt{(v_{ox})^2 + (v_{oy} - gt)^2} dt \\ &= \int_{v_{oy}}^0 \sqrt{(v_{ox})^2 + (u)^2} \left(\frac{du}{-g} \right) \quad u = v_{oy} - gt \\ &= \frac{1}{g} \int_0^{v_{oy}} \sqrt{(v_{ox})^2 + (u)^2} du \end{aligned}$$

Recall that a nice substitution for an integral such as this is provided by the $\sinh(z)$ since $1 + \sinh^2(z) = \cosh^2(z)$ hence a $u = v_{ox} \sinh(z)$ substitution will give

$$(v_{ox})^2 + (u)^2 = (v_{ox})^2 + (v_{ox} \sinh(z))^2 = v_{ox}^2 \cosh^2(z)$$

and $du = v_{ox} \cosh(z) dz$ thus, $\int \sqrt{(v_{ox})^2 + (u)^2} du = \int \sqrt{v_{ox}^2 \cosh^2(z)} v_{ox} \cosh(z) dz = \int v_{ox}^2 \cosh^2(z) dz$. Furthermore, $\cosh^2(z) = \frac{1}{2}(1 + \cosh(2z))$ hence

$$\int \sqrt{(v_{ox})^2 + (u)^2} du = \frac{v_{ox}^2}{2} \left[z + \frac{1}{2} \sinh(2z) \right] + c = \frac{v_{ox}^2}{2} \left[z + \sinh(z) \cosh(z) \right] + c$$

Note $u = v_{ox} \sinh(z)$ and $v_{ox} \cosh(z) = \sqrt{(v_{ox})^2 + (u)^2}$ hence substituting,

$$\int \sqrt{(v_{ox})^2 + (u)^2} du = \frac{1}{2} \left[v_{ox}^2 \sinh^{-1} \left(\frac{u}{v_{ox}} \right) + u \sqrt{v_{ox}^2 + u^2} \right] + c$$

Well, I didn't think that was actually solvable, but there it is. Returning to the definite integral to calculate s we can use the antiderivative just calculated together with FTC part II to conclude: (provided $v_{ox} \neq 0$)

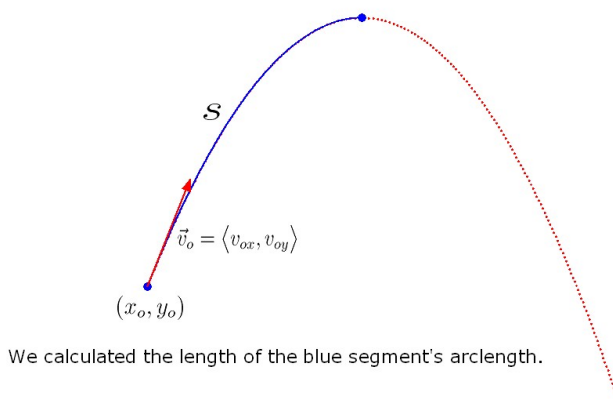
$$s = \frac{1}{2g} \left[v_{ox}^2 \sinh^{-1} \left(\frac{v_{oy}}{v_{ox}} \right) + v_{oy} \sqrt{v_{ox}^2 + v_{oy}^2} \right]$$

³no, we have not shown this is a parabola, I invite the reader to verify this claim. That is find A, B, C such that the graph $y = Ax^2 + Bx + C$ is the same set of points as $\vec{r}(\mathbb{R})$.

If $v_{ox} = 0$ then the problem is much easier since $v(t) = |v_{oy} - gt| = v_{oy} - gt$ for $0 \leq t \leq t_1 = v_{oy}/g$ hence

$$s = \int_0^{t_1} v(t)dt = \int_0^{t_1} (v_{oy} - gt)dt = \left[v_{oy}t - \frac{1}{2}gt^2 \right]_0^{v_{oy}/g} = \boxed{\frac{v_{oy}^2}{2g}}$$

Interestingly, this is the formula for the height of the parabola even if $v_{ox} \neq 0$. The initial x -velocity simply determines the horizontal displacement as the object is accelerated vertically by gravity.



Remark 13.3.9. *physics, grrr...*

I have made a few physical comments in this section, however, you should not confuse these comments with your responsibilities. I intend for you to understand the mathematics. This means that if you are given \vec{r} or \vec{v} or \vec{a} then you can calculate whatever you are not given provided sufficient initial data. Also, if the integration isn't crazy then you can calculate the distance travelled. Notice that while these calculations do use some physical terminology, like position, velocity, speed or acceleration, they are hardly physics problems. We have made no effort to understand what a force is or how it is applied. In short, the calculations we explore here are easy physics problems. I say they are easy because we know exactly what to calculate once the problem is stated. It's just a matter of applying the rules of calculus which we already know and love.

13.4 geometric analysis in parametric setting

In this section we seek to understand tangent lines and arclength. We begin with the problem of tangent lines. Since we have taken a little time to introduce vectors this problem is simple.

Definition 13.4.1. *tangent vector to a path $t \mapsto \vec{r}(t)$.*

The tangent vector at $\vec{r}(t_o)$ is the vector \vec{v} defined below:

$$\vec{v}_o = \left. \frac{d\vec{r}}{dt} \right|_{t=t_o}.$$

If $\vec{v}_o \neq 0$ then the tangent line to $t \mapsto \vec{r}(t)$ at $\vec{r}(t_o) = \vec{r}_o$ is given parametrically by

$$\lambda \mapsto \vec{L}(\lambda) = \vec{r}_o + \lambda \vec{v}_o$$

where we call the **direction vector** for the line \vec{L} .

Example 13.4.2. Find the tangent line to the curve $\vec{r}(t) = \langle 1+t, 2+t^2 \rangle$ at the point $(3, 6)$.

To begin we find if this point is actually on the path; do the equations $3 = 1+t$ and $6 = 2+t^2$ have a simultaneous solution? Answer: yes, $t = 2$ is the solution which solves both equations ($t = -2$ is also a solution for $6 = 2+t^2$ however it does not solve $3 = 1+t$). Calculate:

$$\vec{v}(t) = \frac{d\vec{r}}{dt} = \langle 1, 2t \rangle \Rightarrow \text{hence, } \vec{v}(2) = \langle 1, 4 \rangle$$

and we find the tangent line $\vec{L}(\lambda) = (3, 6) + \lambda \langle 1, 4 \rangle$.

We could use t as the parameter for the tangent line, however I have not done this because I don't want you to think the parameter for the curve and its tangent have to be the same. In the preceding example $\lambda = 0$ matches up with $t = 2$; $\vec{r}(2) = \vec{L}(0)$. Finding the direction of the tangent line is very similar to finding the slope of the tangent line. When we found the slope we first calculated the derivative function which gives us the slope at many points beside the point in question. We had to evaluate $\left. \frac{dy}{dx} \right|_{x=x_o}$ in order to calculate the slope to the tangent line to $y = f(x)$ at $x = x_o$. In the same way, we calculate the *velocity vector field* $\frac{d\vec{r}}{dt}$ which once evaluated $\left. \frac{d\vec{r}}{dt} \right|_{t=t_o}$ gives us a direction vector for the tangent line at $\vec{r}(t_o)$. Again, we find the tangent vector for many points beside the point in question, but evaluation singles out the direction vector we desire. It is called a *vector field* because we envision the velocity vectors glued to the points of tangency. The tangent vector $\frac{d\vec{r}}{dt}$ is most naturally based at the point $\vec{r}(t)$.⁴

⁴In calculus III when we study space curves in \mathbb{R}^3 it is natural to study unit tangent vector field (\vec{T}) along with the unit normal (\vec{N}) and unit binormal vector fields (\vec{B}) of a space curve. For a regular smooth curve the $\{\vec{T}, \vec{N}, \vec{B}\}$ vectors together form a moving coordinate system which is called a moving frame. The classical differential geometry tied to the moving frame is beautifully encapsulated by the Frenet-Serret equations which explain how $\{\vec{T}, \vec{N}, \vec{B}\}$ change along the curve.

Example 13.4.3. Let a be a constant and suppose $x = e^{-t} \cos(t)$ and $y = e^{-t} \sin(t)$ then $\vec{r}(t) = e^{-t} \langle \cos(t), \sin(t) \rangle$ and we can calculate from the product rule for space curves,

$$\begin{aligned} \frac{d\vec{r}}{dt} &= -e^{-t} \langle \cos(t), \sin(t) \rangle + e^{-t} \langle -\sin(t), \cos(t) \rangle \\ &= e^{-t} \langle -\cos(t) - \sin(t), -\sin(t) + \cos(t) \rangle \end{aligned}$$

The mapping $t \mapsto \frac{d\vec{r}}{dt}$ gives us the tangent vector field to $\vec{r}(t) = e^{-t} \langle \cos(t), \sin(t) \rangle$.

We can ask a question: **is there a point which has a tangent vector pointing in the $\langle 0, 1 \rangle$ direction ?**

Solution: We search for solutions to $\frac{d\vec{r}}{dt} = k \langle 0, 1 \rangle$ for some $k \neq 0$. This means we need to solve:

$$e^{-t}(-\cos(t) - \sin(t)) = 0 \quad \text{and} \quad e^{-t}(-\sin(t) + \cos(t)) = k.$$

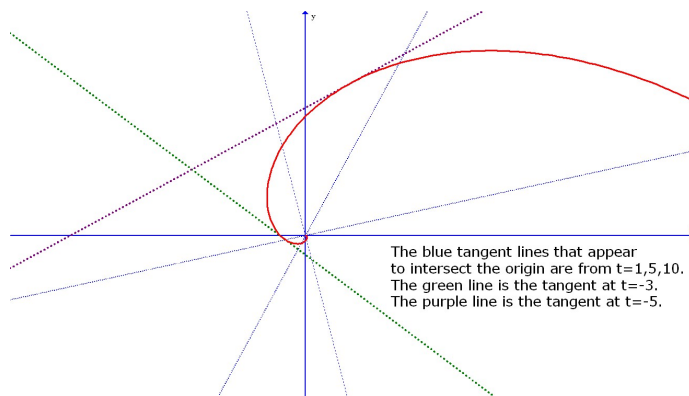
Since $e^{-t} \neq 0$ we must have $-\cos(t) - \sin(t) = 0$ hence $\tan(t) = -1$ which has solutions $t = -\frac{\pi}{4} + n\pi$ for $n \in \mathbb{Z}$. Which is just a fancy way of saying that $t = \dots, -\frac{9\pi}{4}, -\frac{5\pi}{4}, -\frac{\pi}{4}, \frac{3\pi}{4}, \frac{7\pi}{4}, \frac{11\pi}{4}, \dots$. When $t = -\frac{\pi}{4} + 2m\pi$ then

$$-\sin(t) + \cos(t) = -\left(\frac{-\sqrt{2}}{2}\right) + \frac{\sqrt{2}}{2} = \sqrt{2}$$

and we find that $k = \sqrt{2}e^{-\frac{\pi}{4}+2m\pi}$ for any $m \in \mathbb{Z}$ yields the desired direction. On the other hand when $t = -\frac{\pi}{4} + (2m+1)\pi$ we find $k = -\sqrt{2}e^{-\frac{\pi}{4}+(2m+1)\pi}$ for any $m \in \mathbb{Z}$ which means the tangent vector points opposite the desired direction. The tangent lines at $\vec{r}(-\frac{\pi}{4} + 2m\pi)$ are

$$\vec{L}_m(t) = e^{-\frac{\pi}{4}+2m\pi} \langle \sqrt{2}/2, -\sqrt{2}/2 \rangle + t\sqrt{2}e^{-\frac{\pi}{4}+2m\pi} \langle 1, 0 \rangle = \frac{e^{-\frac{\pi}{4}+2m\pi}}{\sqrt{2}} \langle 1 + 2t, -1 \rangle.$$

The curve in the last example winds very tightly around the origin as $t \rightarrow \infty$. It's hard to picture the curve directly but the tangent lines pictured give us some idea.



The family of tangent lines found in the last example were all vertical tangent lines. A vertical tangent line is not covered as the linearization of a function of x . This is one of the geometric deficiencies of using graphs of the form $y = f(x)$. Perhaps you recall that the slope of a line is defined in every case except it is a vertical line. We face no such exception for the parametric case. However, if the velocity vector is zero at a point the analysis does require some thought. There might not be a well-defined tangent line at such a point thus many discussions of the geometry of curves rule out such a case by insisting the curve is *smooth*.

Definition 13.4.4. *smooth curve $t \mapsto \vec{r}(t)$.*

If $\frac{d\vec{r}}{dt}$ exists and $\frac{d\vec{r}}{dt} \neq \langle 0, 0 \rangle$ for all $t \in \text{dom}(\vec{r})$ then we say $t \mapsto \vec{r}(t)$ is **smooth or non-stop**. A point for which the tangent vector either is zero or fails to exist is called a **critical point** for the path.

When the curve is smooth we can sometimes locally express the parametric curve as the graph of a function. We test if $y = f(x)$ is the same curve as $t \mapsto \vec{r}(t) = \langle x(t), y(t) \rangle$ by substituting the parametric formulas for $x(t)$ into f and checking if $y(t)$ is the result. We insist that $y(t) = f(x(t))$ for all t in at least some subset $J \subseteq \text{dom}(\vec{r})$. Differentiate implicitly to find:

$$\frac{dy}{dt} = \frac{df}{dx} \bigg|_{x=x(t)} \frac{dx}{dt}$$

Thus, if $y = f(x)$ and $\vec{r}(t_o) = \langle x(t_o), y(t_o) \rangle = \langle x_o, f(x_o) \rangle$ then

$$\frac{dy}{dx} \bigg|_{x_o} = \frac{\frac{dy}{dt} \big|_{t_o}}{\frac{dx}{dt} \big|_{t_o}}$$

Example 13.4.5. Suppose $x = \cos(t)$ and $y = \sin(t)$ then we have $\vec{r}(t) = \langle \cos(t), \sin(t) \rangle$ hence $\frac{d\vec{r}}{dt} = \langle -\sin(t), \cos(t) \rangle$. We find many critical points: if $t = n\pi$ for $n \in \mathbb{Z}$ then $\frac{d\vec{r}}{dt} = 0$. However, for $t \neq n\pi$ for $n \in \mathbb{Z}$ we find the curve is smooth. Let t_o be some point where this curve is smooth, we calculate

$$\frac{dy}{dx} \bigg|_{x=\cos(t_o)} = \frac{\sin(t_o)}{\cos(t_o)} = \tan(t_o).$$

This is no surprise since we can easily eliminate t by equating $\cos(t) = x$. Note that the parametric equations, while giving the same set of points as $y = \sqrt{1-x^2}$ for $-1 \leq x \leq 1$, is quite different as it covers this line segment infinitely many times as it oscillated between the points $(-1, 0)$ and $(1, 0)$.

Example 13.4.6. Suppose $x = \cosh(t)$ and $y = \sinh(t)$ we can calculate the slope of the corresponding Cartesian curve $y = f(x)$ at $(1, 0)$. Note $t = 0$ yields $x = 1$ and $y = 0$ thus calculate

$$\frac{dy}{dx} \bigg|_{x=1} = \frac{\sinh(t)}{\cosh(t)} \bigg|_{t=0} = \frac{0}{1} = 0.$$

This would be less easy to arrive at through implicit differentiation of the Cartesian equation defining this curve. Just for fun, note $t = \ln(\sqrt{y+1})$ hence $x = \cosh(\ln(\sqrt{y+1})) + \sinh(\ln(\sqrt{y+1}))$. Thus,

$$1 = [\cosh(\ln(\sqrt{y+1})) + \sinh(\ln(\sqrt{y+1}))] \frac{1}{2y} \frac{dy}{dx}$$

Plug in $y = 0$ to obtain $1 = [\cosh(0) + \sinh(0)] \frac{1}{2} \frac{dy}{dx} = \frac{1}{2} \frac{dy}{dx}$. Moral of story? Ignore parametric techniques at your own peril.

If $\frac{d\vec{r}}{dt}$ does not exist at a particular point it could signify the path has abruptly changed direction so the past-pointing and future-pointing tangents differ⁵. However, if the velocity vector $\frac{d\vec{r}}{dt} \big|_{t=t_o} = \langle 0, 0 \rangle$ then it may just be a point like that on the top of a hill where the velocity might tend to zero just as the object barely makes it over the hill. There might still be a well-defined direction for the curve which can be detected by examining how the curve approaches the point of zero-tangent.

⁵this happened in Example 13.4.5 at points $(1, 0)$ and $(-1, 0)$

Example 13.4.7. Let $\vec{r}(t) = \langle t^3, t^3 \rangle$. Calculate $\frac{d\vec{r}}{dt} = \langle 3t^2, 3t^2 \rangle$ thus observe $t = 0$ yields the critical point $\vec{r}(0) = \langle 0, 0 \rangle$. In this case notice:

$$\lim_{t \rightarrow 0} \frac{dy/dt}{dx/dt} = \lim_{t \rightarrow 0} \frac{3t^2}{3t^2} = 1$$

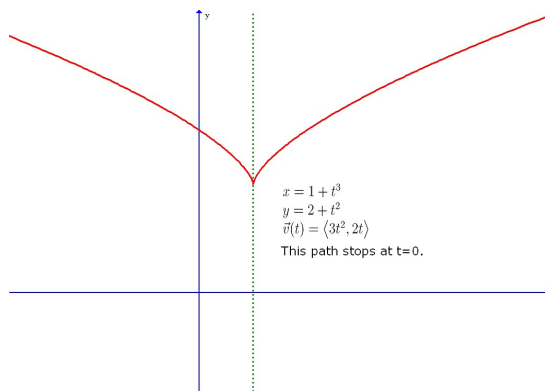
Thus we identify that $dy/dx = 1$ at $(0, 0)$.

However, critical points are much like limits of indeterminate form. There are many possible types of behaviour.

Example 13.4.8. Let $\vec{r}(t) = \langle 1 + t^3, 2 + t^2 \rangle$ hence $\frac{d\vec{r}}{dt} = \langle 3t^2, 2t \rangle$ thus observe $t = 0$ yields the critical point $\vec{r}(0) = \langle 1, 2 \rangle$. Notice:

$$\lim_{t \rightarrow 0^+} \frac{dy/dt}{dx/dt} = \lim_{t \rightarrow 0^+} \frac{2t}{3t^2} = \lim_{t \rightarrow 0^+} \frac{2}{3t} = \infty \quad \text{and} \quad \lim_{t \rightarrow 0^-} \frac{dy/dt}{dx/dt} = \lim_{t \rightarrow 0^-} \frac{2t}{3t^2} = \lim_{t \rightarrow 0^-} \frac{2}{3t} = -\infty$$

Thus we identify that dy/dx is not defined at $(0, 0)$. This is a vertical tangent.



In any event, we are most interested in smooth curves in this course so don't worry about finding other possible cases. For now you should mainly be interested in being able to find the parametric equations for the tangent line for a given smooth curve. In addition, you should work on developing a picture in your mind for our calculations. Draw pictures. Test your hunches. Of course Mathematica or another CAS will produce nice plots of parametric curves and their tangents but you must test your understanding on simple examples as to make sure you understand what the CAS is telling you if nothing else. I wouldn't ask a question which required the full assistance of a CAS on the test. Obviously the problem I give will be algebraically accessible and mainly test your competence in understanding the definitions we have explored in this section. Some of your homework is probably deliberately messy so the CAS is the only way to go. That is for breadth not for the test.

Remark 13.4.9. *higher derivatives for parametric curves*

To calculate $y''(x)$ in terms of $x(t)$ and $y(t)$ we can just differentiate (omitting explicit notation for the t -dependence, let it be understood all of the objects below depend appropriately on t)

$$\frac{dy}{dt} = \frac{df}{dx} \frac{dx}{dt}$$

implicitly a second time to obtain by the product and chain rules:

$$\frac{d^2y}{dt^2} = \frac{d^2f}{dx^2} \frac{dx}{dt}^2 + \frac{df}{dx} \frac{d^2x}{dt^2}$$

Hence substitute $\frac{\frac{dy}{dt}}{\frac{dx}{dt}}$ for $\frac{df}{dx}$ to find

$$\frac{d^2y}{dt^2} = \frac{d^2f}{dx^2} \left(\frac{dx}{dt} \right)^2 + \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \frac{d^2x}{dt^2}$$

Therefore, if $y = f(x(t))$ then

$$\frac{d^2y}{dx^2} = \frac{\frac{d^2y}{dt^2} - \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \frac{d^2x}{dt^2}}{\left(\frac{dx}{dt} \right)^2} = \frac{\frac{dx}{dt} \frac{d^2y}{dt^2} - \frac{dy}{dt} \frac{d^2x}{dt^2}}{\left(\frac{dx}{dt} \right)^3}$$

Example 13.4.10. Let $x = t + 1$ and $y = t^3$ calculate $\frac{dx}{dt} = 1$ and $\frac{d^2x}{dt^2} = 0$ whereas $\frac{dy}{dt} = 3t^2$ and $\frac{d^2y}{dt^2} = 6t$ it follows that

$$\frac{d^2y}{dx^2} = \frac{\frac{dx}{dt} \frac{d^2y}{dt^2} - \frac{dy}{dt} \frac{d^2x}{dt^2}}{\left(\frac{dx}{dt} \right)^3} = 6t$$

Hence for $t < 0$ the curve is concave down whereas for $t > 0$ the curve is concave up. From $x = t + 1$ we find CD for $x < 1$ and CU for $x > 1$.

If you eliminate t to find $y = (x - 1)^3$ and it's easy to calculate that the result we just found should be agreeable from your previous work on concavity.

13.4.1 motivating the arclength formula

To begin we assume the curve $t \mapsto \vec{r}(t) = \langle x(t), y(t) \rangle$ is smooth and $t \in [a, b]$. Let us begin by partitioning the closed interval as we did in the integration theory:

$$t_0 = a, t_1 = a + \Delta t, t_2 = a + 2\Delta t, \dots, t_{n-1} = a + (n-1)\Delta t, t_n = b$$

where $\Delta t = (b-a)/n$ and $n \in \mathbb{N}$ is the number of subintervals in the partition. Let us define $\vec{r}_k = \vec{r}(t_k) = \langle x_k, y_k \rangle$ for $k = 0, 1, 2, \dots, n$. For each subinterval $[t_{k-1}, t_k]$ we can construct the tangent line from \vec{r}_{k-1} to \vec{r}_k by the usual construction for a line segment: however, we set-up the parameter of each line segment as to make the segment's parameter match the parameter of the given curve:

$$\vec{L}_k(t) = \vec{r}_{k-1} + \left(\frac{t - t_{k-1}}{\Delta t} \right) (\vec{r}_k - \vec{r}_{k-1}) \quad \text{for } t \in [t_{k-1}, t_k].$$

As a check on the formula above, note $\vec{L}_k(t_{k-1}) = \vec{r}_{k-1}$ whereas $\vec{L}_k(t_k) = \vec{r}_k$. We calculate the length of the line segment $t \mapsto \vec{L}_k(t)$ for $t \in [t_{k-1}, t_k]$ by calculating the length of the displacement from \vec{r}_{k-1} to \vec{r}_k :

$$\Delta s_k = \|\vec{r}_k - \vec{r}_{k-1}\| = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}$$

The total length of this piecewise-polygonal approximation to the smooth path is

$$\Delta s_n = \sum_{k=1}^n \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}.$$

If we allow $n \rightarrow \infty$ then the approximation should converge to the arclength.

$$\begin{aligned} \Delta s &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2} \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \sqrt{\left(\frac{x_k - x_{k-1}}{\Delta t} \right)^2 + \left(\frac{y_k - y_{k-1}}{\Delta t} \right)^2} \Delta t. \end{aligned}$$

By construction, $t_k = t_{k-1} + \Delta t$ so we may identify that as $n \rightarrow \infty$ and $\Delta t \rightarrow 0$ since $\Delta t = (b-a)/n$ the difference quotients tend to derivatives:

$$\frac{x_k - x_{k-1}}{\Delta t} = \frac{x(t_{k-1} + \Delta t) - x(t_{k-1})}{\Delta t} \longrightarrow \frac{dx}{dt}(t_{k-1}).$$

Likewise,

$$\frac{y_k - y_{k-1}}{\Delta t} = \frac{y(t_{k-1} + \Delta t) - y(t_{k-1})}{\Delta t} \longrightarrow \frac{dy}{dt}(t_{k-1})$$

Hence,

$$\begin{aligned} \Delta s &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \sqrt{\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2} \Delta t \\ &= \int_a^b \sqrt{\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2} dt \end{aligned}$$

where in the last step we identified the Riemann sum as the arclength integral we sought from the outset.

Notice that in the case of a graph $y = f(x)$ we can take $x = t$ and we derive $\Delta s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx} \right)^2} dx$. \square

Remark 13.4.11. *heuristic metric calculation*

The infinitesimal argument should be understood as a shorthand for the technical argument we covered before this remark:

1. $ds^2 = dx^2 + dy^2$ since a tiny triangle with sides dx and dy has a hypotenuse which is really close to the arclength.
2. Thus $ds = \sqrt{dx^2 + dy^2}$
3. Factor out a dt to obtain $ds = \sqrt{(dx/dt)^2 + (dy/dt)^2} dt$.
4. add up all the little ds 's by integrating.

This infinitesimal argument should be thought of as a nice notation to summarize a more technically correct summation argument (which is what follows this remark). I usually emphasize the infinitesimal approach in calculus because it gets to the heart of the matter quickly and provides a very general framework to solve a myriad of applied problems. The infinitesimal ds^2 is called the **metric**. In general relativity the metric is sometimes used as the central object of interest in the theory. You can calculate the influence of gravity in curved space through the mathematics of the metric. The calculus for curved space involves **tensors** and **manifolds**. The calculations in tensor calculus are very very fun.

In both notations there is a question we have neglected. In particular, how do we know our calculation is independent of parametrization? If the arclength calculated depends on the choice of parameter then it is not what we would want to call arclength. Geometrically, it is clear that the arclength is an intrinsic geometric property of a given curve. I prove independence of the arclength formula from choice of parameter in calculus III if time permits.

13.5 arclength examples

It might be helpful to note Stewart discusses arclength for graphs and parametric curves in separate sections: see §9.1 and page 669-670 for his arclength examples. The distinction between the examples in this section and those in the previous section on distance travelled is that we make no assumption that the parameter is time. The calculations here are merely geometrical.

Example 13.5.1. .

Consider the following parametric curve where $0 \leq t \leq 3$, find arclength.

$$\begin{aligned} x &= e^t + e^{-t} & \dot{x} &= e^t - e^{-t} & \dot{x}^2 + \dot{y}^2 &= (e^t - e^{-t})^2 + 4 \\ y &= 5 - 2t & \dot{y} &= -2 \end{aligned}$$

$$s = \int_0^3 \sqrt{e^{2t} - 2 + e^{-2t} + 4} dt = \int_0^3 \sqrt{e^{2t} + 2 + e^{-2t}} dt = \int_0^3 (e^t + e^{-t}) dt$$

$$= (e^t - e^{-t}) \Big|_0^3 = \boxed{e^3 - 1/e^3}$$

Example 13.5.2. .

Let $\alpha(t) = (r \cos t, r \sin t)$ $0 \leq t \leq 2\pi$

$x = r \cos t$ $\frac{dx}{dt} = -r \sin t$ $\therefore \left(\frac{dx}{dt}\right)^2 = r^2 \sin^2 t$

$y = r \sin t$ $\frac{dy}{dt} = r \cos t$ $\therefore \left(\frac{dy}{dt}\right)^2 = r^2 \cos^2 t$

$$S = \int_0^{2\pi} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt = \int_0^{2\pi} \sqrt{r^2 \sin^2 t + r^2 \cos^2 t} dt = \int_0^{2\pi} r dt = \boxed{2\pi r}$$

Which is good since $x^2 + y^2 = r^2$, is a circle of radius r .

Example 13.5.3. .

Let $\alpha(x) = (x, \ln(\cos(x)))$ $0 \leq x \leq \pi/4$ $\begin{matrix} X = x \\ Y = \ln(\cos(x)) \end{matrix}$

$Y = \ln(\cos(x))$ for $0 \leq x \leq \pi/4$

$$S = \int_0^{\pi/4} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

$$= \int_0^{\pi/4} \sqrt{1 + \left(\frac{-\sin(x)}{\cos(x)}\right)^2} dx$$

$$= \int_0^{\pi/4} \sqrt{1 + \tan^2(x)} dx \quad ; \quad \tan^2(x) + 1 = \sec^2(x)$$

$$= \int_0^{\pi/4} \sec(x) dx$$

$$= \int_1^{1+\sqrt{2}} \frac{du}{u}$$

$$= \left(\ln |u| \right) \Big|_1^{1+\sqrt{2}}$$

$$= \ln(1+\sqrt{2}) - \ln(1)$$

$$= \boxed{\ln(1+\sqrt{2})}$$

$u = \sec(x) + \tan(x)$

$\frac{du}{u} = \sec(x) dx$

$u(0) = \sec(0) + \tan(0) = 1$

$u(\pi/4) = \sec(\pi/4) + \tan(\pi/4) = \sqrt{2} + 1$

Example 13.5.4. .

Find circumference of an ellipse: $X = a \cos t$; $\left(\frac{X}{a}\right)^2 + \left(\frac{Y}{b}\right)^2 = \sin^2 t + \cos^2 t = 1$
 $Y = b \sin t$ where we are given $0 \leq t \leq 2\pi$.
 $\frac{dx}{dt} = -a \sin t$ and $\frac{dy}{dt} = b \cos t$

$$\begin{aligned} S &= \int_0^{2\pi} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt \\ &= \int_0^{2\pi} a \sqrt{\sin^2 t + \left(\frac{b \cos t}{a}\right)^2} dt \\ &= \int_0^{2\pi} a \sqrt{1 - \cos^2 t + \left(\frac{b \cos t}{a}\right)^2} dt \\ &= \int_0^{2\pi} a \sqrt{1 - (1 - (b/a)^2) \cos^2 t} dt \quad : \text{define } \beta = \sqrt{1 - (b/a)^2} \\ &= \int_0^{2\pi} a \sqrt{1 - \beta^2 \cos^2 t} dt \end{aligned}$$

This integral is not elementary. We'll need a specific a and b in order to proceed. Note that when $a = b$ then $\beta = 0$

So $S = 2\pi a$ happily. Consider $a = 1$, $b = \sqrt{2}$ then $\beta = 1$

$$\begin{aligned} \int_0^{2\pi} \sqrt{1 - \cos^2 t} dt &= \int_0^{2\pi} |\sin t| dt \\ &= \int_0^{\pi} \sin t dt - \int_{\pi}^{2\pi} \sin t dt = 4 \end{aligned}$$

For most choices of a & b we would need to do this integral numerically

Example 13.5.5. .

$y = \frac{x^3}{6} + \frac{1}{2x}$ with $1/2 \leq x \leq 1$ find arclength.

$$\frac{dy}{dx} = \frac{x^2}{2} - \frac{1}{2x^3} \quad \therefore \left(\frac{dy}{dx}\right)^2 = \frac{1}{4} \left(x^4 - 2 + \frac{1}{x^4}\right)$$

$$\begin{aligned} S &= \int_{1/2}^1 \sqrt{1 + \frac{1}{4} \left(x^4 - 2 + \frac{1}{x^4}\right)} dx = \int_{1/2}^1 \sqrt{\frac{1}{4} \left(x^4 + 2 + \frac{1}{x^4}\right)} dx = \int_{1/2}^1 \sqrt{\frac{1}{4} \left(x^2 + \frac{1}{x^2}\right)^2} dx \\ &= \int_{1/2}^1 \frac{1}{2} \left(x^2 + \frac{1}{x^2}\right) dx = \frac{1}{2} \left[\frac{x^3}{3} - \frac{1}{x} \right]_{1/2}^1 = \frac{31}{48} = \boxed{0.6458} \end{aligned}$$

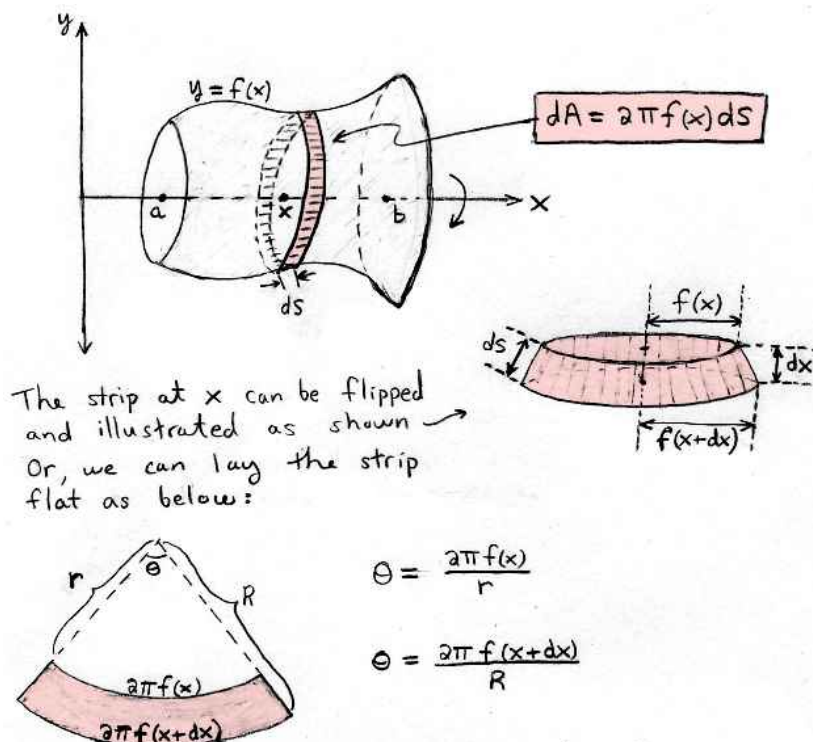
Chapter 14

Further Applications of Integral Calculus

In this short chapter we examine two applications of integral calculus. I will utilize the infinitesimal method to motivate both sections. The surface area problem we consider in this chapter is just the simple case of a surface of revolution. The general problem of surface area is dealt with properly in calculus III. Physics provides a wealth of applied problems, we content ourselves with the problems of work, hydrostatic force on a dam and the center of mass or centroid problem. Our emphasis is once more on mathematical set-up as opposed to physical concept. I will attempt to provide the conceptual framework and then you simply need to work out various geometries in my framework.

14.1 surface area

Imagine we take a graph $y = f(x)$ for $a \leq x \leq b$ and rotate it around the x -axis. We suppose $f(x) \geq 0$ for the purposes of this discussion. This creates a surface of revolution. You may recall from calculus I that we calculated the volume contained inside such surfaces for a variety of cases. For now we will just focus on the case of a surface revolved around the x -axis. Let us focus on just a small portion of the surface. In particular the bit from x to $x + dx$. Suppose ds is the arclength of the graph from $(x, f(x))$ to $(x + dx, f(x + dx))$. The straight-line distance from $(x, f(x))$ to $(x + dx, f(x + dx))$ is identical to ds in this infinitesimal limit. Notice then we can calculate the area of this conical ribbon which has radii $f(x)$ and $f(x + dx)$ at its edges and a length of ds along the edge as described below¹:



The area of this partial annulus is given by

$$\begin{aligned}
 dA &= \pi R^2 \left(\frac{\theta}{2\pi} \right) - \pi r^2 \left(\frac{\theta}{2\pi} \right) \\
 &= \frac{1}{2} \theta (R^2 - r^2) \\
 &= \frac{1}{2} \theta (R+r)(R-r) : \left[\begin{array}{l} \text{observe } \theta \approx \pi \text{ since} \\ f(x+dx) \approx f(x) \text{ as} \\ dx \approx 0. \end{array} \right] \\
 &= \frac{1}{2} (\pi) (2f(x)) ds \\
 &= \underline{2\pi f(x) ds}.
 \end{aligned}$$

Or, in terms of dx , $dA = 2\pi y \sqrt{1 + \left(\frac{dy}{dx} \right)^2} dx$ where $y = f(x)$.

¹thanks to Ginny for this idea

Furthermore, while we assumed an increasing function for the ease of visualization this formula holds for the case that f is decreasing. Note that ds is positive and we assumed from the outset that $f(x) \geq 0$. Recall $ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + (dy/dx)^2}dx$ hence the total surface area is thus found from the following integration:

$$A = \int_a^b 2\pi f(x) \sqrt{1 + \frac{df}{dx}^2} dx$$

All of this said, we can state a more general formula for parametric curves around an arbitrary axis in the plane. Suppose that $t \mapsto \langle x(t), y(t) \rangle$ is a parametric curve and \mathcal{L} is a line in the plane. Suppose this parametric curve does not cross the axis and any perpendicular bisector of the axis crosses the curve in at most one point. Let $r(t)$ be the distance from the curve to the axis then we can by the same argument as given for $y = f(x)$ derive that the area of the surface of revolution formed by rotation the parameterized curve for $a \leq t \leq b$ is simply:

$$A = \int_a^b 2\pi r(t) \sqrt{\frac{dx}{dt}^2 + \frac{dy}{dt}^2} dt$$

In the case of the graph we had $t = x$ and $r(t) = f(x)$. For an arbitrary example the real problem is geometrically determining the formula for $r(t)$.

Example 14.1.1. Problem: Let $R > 0$ and $x = R \cos(t)$ and $y = R \sin(t)$ for $-\frac{\pi}{2} \leq t \leq \frac{\pi}{2}$. Find the surface area of the surface formed by revolving the given curve around the y -axis.

Solution: the curve we consider is a half-circle centered at the origin with radius R . Since the axis \mathcal{L} is the y -axis the distance² to a point $(x(t), y(t))$ on the curve is clearly $x(t)$. Recall $ds = \sqrt{(dx/dt)^2 + (dy/dt)^2} dt$ hence we calculate $ds = \sqrt{R^2 \sin^2(t) + R^2 \cos^2(t)} dt = R dt$. The area of a typical infinitesimal ribbon is $dA = 2\pi r(t) ds = 2\pi(R \cos(t))(R dt)$ for each $t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Add together all the little dA 's by integration to find the total surface area:

$$A = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 2\pi R^2 \cos(t) dt = 2\pi R^2 \sin(t) \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \boxed{4\pi R^2}.$$

Therefore, the surface area of a sphere of radius R is $4\pi R^2$. It is interesting to note that the integral of this formula with respect to R yields the volume of a sphere; $V = \frac{4}{3}\pi R^3$. Likewise, the circumference of the disk of radius R is $2\pi R$ which once integrated yields the area πR^2 . This pattern does not hold for all solids. For example, if you think about a cube of side length x then the $V = x^3$ whereas the surface area is $A = 6x^2$. The symmetry of the sphere or circle is very special and the pattern continues for higher dimensional spheres³

²distance from a point to a set is by definition the distance from the point to the closest point in the set

³ $x^2 + y^2 = 1$ gives unit circle or S_1 , $x^2 + y^2 + z^2 = 1$ gives unit spherical shell or S_2 , $x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1$ gives the unit n -sphere S_n . You can find a derivation of the hypervolume of the higher dimensional spheres in Apostol if you're interested.

Example 14.1.2. Problem: Find the surface area of the surface of revolution formed by rotating $y = \sqrt{x}$ around the x -axis for $0 \leq x \leq 1$.

Solution: We have

$$\begin{aligned} dA &= 2\pi\sqrt{x}\sqrt{1 + \left[\frac{1}{2\sqrt{x}}\right]^2} dx \\ &= 2\pi\sqrt{x\left[1 + \frac{1}{4x}\right]} dx \\ &= 2\pi\sqrt{x + \frac{1}{4}} dx \\ &= \pi\sqrt{4x + 1} dx \end{aligned}$$

We can integrate the expression above with the help of a $u = 4x + 1$ substitution. Note

$$\int \sqrt{4x + 1} dx = \int \sqrt{u} \frac{du}{4} = \frac{2u^{\frac{3}{2}}}{12} + c = \frac{1}{6}(4x + 1)^{\frac{3}{2}} + c.$$

Thus,

$$A = \int_0^1 \pi\sqrt{4x + 1} dx = \frac{\pi}{6}(4x + 1)^{\frac{3}{2}} \Big|_0^1 = \boxed{\frac{\pi}{6}(\sqrt{125} - 1)}.$$

Example 14.1.3. Problem: Find the surface area of an open right-circular cone of height h and radius R .

Solution: we can view this as a surface of revolution. Take the line $y = Rx/h$ for $0 \leq x \leq h$ and rotate it around the x -axis. Observe that

$$dA = 2\pi\left(\frac{Rx}{h}\right)\sqrt{1 + \frac{R^2}{h^2}} dx = \frac{2\pi Rx}{h^2}\sqrt{h^2 + R^2} dx.$$

Now integrate over $0 \leq x \leq h$ to find the total area:

$$A = \int_0^h \frac{2\pi Rx}{h^2}\sqrt{h^2 + R^2} dx = \frac{2\pi Rx^2}{2h^2}\sqrt{h^2 + R^2} \Big|_0^h = \boxed{\pi R\sqrt{h^2 + R^2}}.$$

Notice that the formula above checks nicely in the limits $h \rightarrow 0$ and $R \rightarrow 0$ where we find $A \rightarrow \pi R^2$ and $A \rightarrow 0$ respective. Can you see why this makes sense?

14.2 physics

In this section we examine a few variable force work problems, variable pressure hydrostatic force problems and finally the center of mass problem for a homogeneous laminate in the plane.

14.2.1 work and force with calculus

The basic physical concepts used here are as follows:

1. work W due to a force F over a displacement Δx is defined to be $W = F\Delta x$ provided the force is exerted in the direction of the displacement and is constant.
2. the force F exerted over an area A by a pressure P is defined to be $F = PA$ provided the pressure is constant over the area A .

In the examples we consider in this section we cannot simply multiply as described above because the requisite idealizations are not met in our examples in the finite case. In other words, the forces are variable and the pressures are not constant. However, if we instead consider an infinitesimal displacement dx or an infinitesimal area dA we can in fact realize the idealized physical laws. It is true that $dW = Fdx$ because the F does not change over the tiny displacement dx . For the problem of the dam, we can say $dF = PdA$ if our dA is a horizontal strip since the pressure is constant over a certain depth. I'll leave the rest of the details for the examples. Mainly we need the following physical equations to complete the examples:

1. $F = mg$, near the surface of the earth this is the force of gravity on a mass m .
2. $P = \rho gd$, is the pressure due to water at a depth d where $\rho \cong 1000\text{kg}/\text{m}^3$ is the density of water.

Example 14.2.1. . This is a variable work due to variable mass problem.

Find magnitude of work required to lift a 10m cable with a uniformly distributed mass of 10 kg.

Lets define l to be the length of the cable being lifted, this will vary from $l = 10\text{ m} \rightarrow l = 0\text{ m}$ as the cable is lifted. We should find the work dW done as we lift length l a distance dl ,

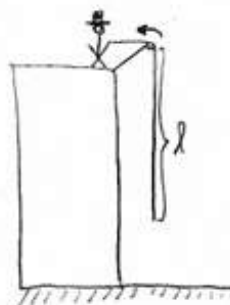
$$\begin{aligned} dW &= Fdl \\ &= -mg\,dl \\ &= -\lambda g\,dl \end{aligned}$$

Now add up the work

$$\begin{aligned} W &= \int_{10}^0 -\lambda g\,l\,dl \\ &= -\lambda g \left(\frac{1}{2} l^2 \right) \Big|_{10}^0 \\ &= \lambda g \frac{1}{2} (10)^2 \\ &= -1 \cdot 9.8 \cdot 50 \\ &= \boxed{490\text{ J}} \end{aligned}$$

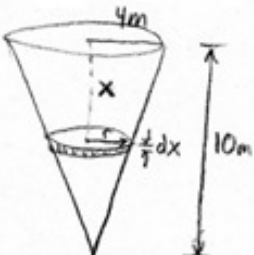
$$\lambda = \frac{\text{mass}}{\text{length}} = \frac{10\text{ kg}}{10\text{ m}} = 1\text{ kg/m}$$

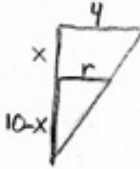
" λ is linear mass density"
 $m = \lambda l$.



Example 14.2.2. . This is a variable work due to variable mass problem.

Find minimum work to pump water out of a full cone.



$$dm = \rho dV = \rho \pi r^2 dx$$


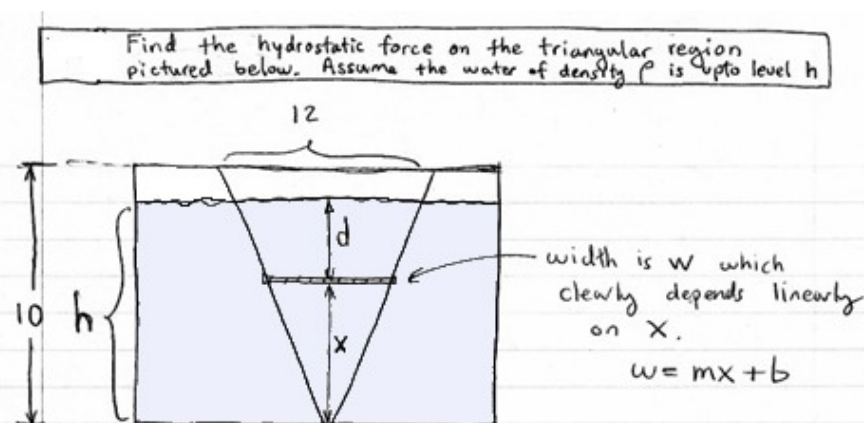
$$\frac{4}{10} = \frac{r}{10-x}$$

$$r = \frac{2}{5}(10-x)$$

$$\begin{aligned}
 dW &= (dm) g \cdot x \\
 &= (\rho \pi r^2 dx) g \cdot x \\
 &= \rho g \pi \frac{4(10-x)^2}{25} x dx \\
 &= \frac{4\pi \rho g}{25} (100x - 20x^2 + x^3) dx \\
 W &= \int_0^{10} \frac{4\pi \rho g}{25} (100x - 20x^2 + x^3) dx \\
 &= \frac{4\pi \rho g}{25} \left(50 \times 10^2 - \frac{20}{3} \times 10^3 + \frac{10^4}{4} \right) \\
 &= \frac{4\pi (1000)(9.8)}{25} \left(50 \times 10^2 - \frac{20}{3} \times 10^3 + \frac{10^4}{4} \right) \\
 &= \boxed{4.11 \times 10^6 \text{ J}}
 \end{aligned}$$

Remark: this is a magic cone, we cannot just tip it over or drill a hole in the bottom. We are forced to pump out the top 😊.

Example 14.2.3. . The triangular dam problem.



$$w(0) = m(0) + b = 0 \quad \therefore b = 0$$

$$w(10) = m(10) + 0 = 12 \quad \therefore m = \frac{12}{10} = \frac{6}{5} \quad \therefore \boxed{w = \frac{6}{5}x}$$

This formula for w checks because $w(10) = \frac{6}{5}(10) = 12$ as it should.

Then the area of strip is $dA = w dx = \frac{6}{5}x dx$.

Now set up the pressure $P = \rho g d$,

$$x + d = h \Rightarrow d = h - x \Rightarrow P = \rho g (h - x)$$

Again the connection between force & pressure is $P = \frac{dF}{dA}$
 so $dF = P dA$, thus

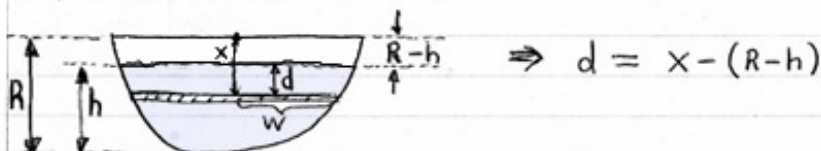
$$dF = \rho g (h - x) \frac{6}{5} x dx$$

Now sum the forces for the strips at x in $0 \leq x \leq h$,

$$\begin{aligned} F &= \int_0^h \frac{6\rho g}{5} (hx - x^2) dx \\ &= \frac{6\rho g}{5} \left(\frac{1}{2}hx^2 - \frac{1}{3}x^3 \right) \Big|_0^h \\ &= \frac{6\rho g}{5} \left(\frac{1}{2}h^3 - \frac{1}{3}h^3 \right) \\ &= \boxed{\frac{1}{5}\rho g h^3} \end{aligned}$$

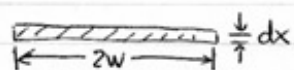
Example 14.2.4. . The hemispherical dam problem.

Find hydrostatic force on half-barrel pictured below, well just the end piece. The radius of barrel is R and the water of density ρ is filled to height h . Lets find force dF on a strip of area dA at position x and depth d below the surface.

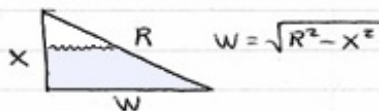


From picture above we find $P = \rho g d = \rho g (x - R + h)$.

(I set up the pressure wrong in notes, it is the depth that should determine the pressure, specifically $P = \rho g d$.) Next;



$$dA = 2w dx = 2\sqrt{R^2 - x^2} dx$$



Ok so our choice of x makes dA relatively pretty, (you can try defining x differently but it'll make the square root nasty...) Ok, we know $P = \frac{dF}{dA}$ so $dF = P dA$

$$dF = \rho g (x - R + h) 2\sqrt{R^2 - x^2} dx$$

Now we just need to add-up the forces, $R-h \leq x \leq R$

$$\begin{aligned} F &= \int_{R-h}^R 2\rho g (x - R + h) \sqrt{R^2 - x^2} dx \\ &= \int_{R-h}^R 2\rho g x \sqrt{R^2 - x^2} dx + \int_{R-h}^R 2\rho g (h - R) \sqrt{R^2 - x^2} dx \\ &= 2\rho g \underbrace{\int_{R-h}^R x \sqrt{R^2 - x^2} dx}_{u\text{-substitute}} + 2\rho g (h - R) \underbrace{\int_{R-h}^R \sqrt{R^2 - x^2} dx}_{\text{trig-substitute}}. \end{aligned}$$

The hemispherical dam problem continued:

$$\begin{aligned} \int x \sqrt{R^2 - x^2} dx &= \int \sqrt{W} \frac{dW}{-2} & W &= R^2 - x^2 \\ &= -\frac{1}{2} \int W^{\frac{1}{2}} dW & dW &= -2x dx \\ &= -\frac{1}{2} \cdot \frac{2}{3} W^{\frac{3}{2}} + C & x dx &= -\frac{dW}{2} \\ &= -\frac{1}{3} (R^2 - x^2)^{3/2} + C \end{aligned}$$

$$\begin{aligned} \int \sqrt{R^2 - x^2} dx &= \int (R \cos \theta)(R \cos \theta d\theta) & X &= R \sin \theta \\ &= \int R^2 \cos^2 \theta d\theta & dx &= R \cos \theta d\theta \\ &= \int \frac{R^2}{2} (1 + \cos(2\theta)) d\theta & R^2 - x^2 &= R^2 \cos^2 \theta \\ &= \frac{R^2}{2} \int d\theta + \frac{R^2}{2} \int \cos(2\theta) d\theta & \sqrt{R^2 - x^2} &= R \cos \theta \\ &= \frac{R^2}{2} \left(\theta + \frac{1}{2} \sin(2\theta) \right) + C \end{aligned}$$

$$\begin{aligned} F &= 2\rho g \int_{R-h}^R x \sqrt{R^2 - x^2} dx + 2\rho g (h-R) \int_{R-h}^R \sqrt{R^2 - x^2} dx \\ &= -\frac{2\rho g}{3} (R^2 - x^2)^{3/2} \Big|_{R-h}^R + \rho g (h-R) R^2 \left(\theta + \frac{1}{2} \sin(2\theta) \right) \Big|_{X=R-h}^{X=R} \\ &= +\frac{2}{3} \rho g (R^2 - (R-h)^2)^{3/2} + \rho g (h-R) R^2 \left(\sin^{-1}\left(\frac{x}{R}\right) + \frac{1}{2} \sin\left(2 \sin^{-1}\left(\frac{x}{R}\right)\right) \right) \Big|_{R-h}^R \\ &= \frac{2}{3} \rho g (R^2 - (R-h)^2)^{3/2} + \rho g (h-R) R^2 \left(\sin^{-1}\left(\frac{R}{R}\right) + \frac{1}{2} \sin\left(2 \sin^{-1}\left(\frac{R}{R}\right)\right) \right) \\ &\quad - \rho g (h-R) R^2 \left(\sin^{-1}\left(\frac{R-h}{R}\right) + \frac{1}{2} \sin\left(2 \sin^{-1}\left(\frac{R-h}{R}\right)\right) \right) \\ &= \boxed{\frac{2}{3} \rho g (R^2 - (R-h)^2)^{3/2} + \rho g (h-R) R^2 \left[\frac{\pi}{2} + \frac{1}{2} \sin\left(2 \frac{\pi}{2}\right) \right.} \\ &\quad \left. - \sin^{-1}\left(\frac{R-h}{R}\right) + \frac{1}{2} \sin\left(2 \sin^{-1}\left(\frac{R-h}{R}\right)\right) \right]} \end{aligned}$$

14.2.2 center of mass

The concept of center of mass is a ubiquitous topic in mechanics. In a nutshell it allows us to idealize shapes with finite size as if they were just a point mass. This is a tremendous simplification as it allows us to think of just one particle at a time rather than the infinity of atoms that make up a solid. I'll prove this idealization is reasonable in physics, but for here we just want to see how the center of mass is calculate via calculus.

Moments & Centers of Mass (c.o.m)

In the discrete case we define $\vec{r}_{cm} = \frac{\sum m_i \vec{r}_i}{\sum m_i}$ the sum is over all the particles.
For 2 dimensional case we have

$$\vec{r}_{cm} = (x_{cm}, y_{cm})$$

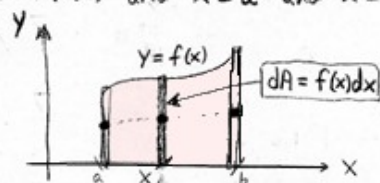
$$x_{cm} = (\sum m_i x_i) / (\sum m_i) = M_x / M$$

M_x = moment of inertia w.r.t x-axis.

$$y_{cm} = (\sum m_i y_i) / (\sum m_i) = M_y / M$$

M = total mass.

What then is the generalization of this to a continuous region. Lets see how to find the c.o.m. of the region bounded by $y=0$, $y=f(x)$ and $x=a$ and $x=b$. Assume uniform density.



Each strip has its c.o.m. at $(x, \frac{1}{2} f(x))$. We can treat it like a bunch of particles with dm each

$$\rho = \frac{dm}{dA} \rightarrow dm = \rho dA$$

So we find the c.o.m. of this system in the natural way,

$$\bar{X} = \frac{\int_a^b \rho x f(x) dx}{\int_a^b \rho f(x) dx} \quad \text{AND} \quad \bar{Y} = \frac{\int_a^b \rho \frac{1}{2} f(x) f(x) dx}{\int_a^b \rho f(x) dx}$$

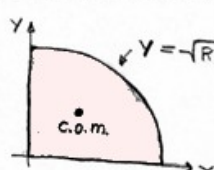
Or in terms of moments & mass we have

$M_y \equiv \rho \int x f(x) dx$	$\bar{X} = \frac{M_y}{M}$
$M_x \equiv \rho \int \frac{1}{2} [f(x)]^2 dx$	$\bar{Y} = \frac{M_x}{M}$
$M = \rho \int f(x) dx = \rho (\text{AREA})$	

We assume ρ to be a constant.

Example 14.2.5. . Center of mass problem.

Find the moments of inertia and c.o.m for a quarter-circle of uniform density ρ with radius R



$$M = \int_0^R \rho f(x) dx$$

$$= \rho \int_0^R f(x) dx$$

$$= \boxed{\frac{\pi}{4} \rho R^2 = M}$$

← mass is simply area times density.

$$M_y = \rho \int_0^R x \sqrt{R^2 - x^2} dx$$

$$= \rho \int_{R^2}^0 \frac{-1}{2} \sqrt{u} du$$

$$= \frac{\rho}{2} \frac{2}{3} u^{3/2} \Big|_0^{R^2}$$

$$= \frac{\rho}{3} (R^2)^{3/2}$$

$$= \boxed{\frac{1}{3} \rho R^3 = M_x}$$

← $u = R^2 - x^2 \quad u(R) = 0$
 $du = -2x dx \quad u(0) = R^2$

$$M_x = \rho \int_0^R \frac{1}{2} (R^2 - x^2) dx$$

$$= \frac{\rho}{2} \left(R^2 x - \frac{1}{3} x^3 \right) \Big|_0^R$$

$$= \boxed{\frac{1}{3} \rho R^3 = M_y}$$

Now we can find the center of mass,

$$x_{cm} = \frac{M_y}{M} = \frac{\frac{1}{3} \rho R^3}{\frac{\pi}{4} \rho R^2} = \frac{4}{3\pi} R \approx 0.4244 R$$

$$y_{cm} = \frac{M_x}{M} = \frac{\frac{1}{3} \rho R^3}{\frac{\pi}{4} \rho R^2} = \frac{4}{3\pi} R \approx 0.4244 R$$

From the symmetry of the region we could have anticipated $x_{cm} = y_{cm}$.

Notice we could work problems where the density depended on x without much trouble, however if the density depended on both x and y at once then it would not be easy given our current tools. In calculus III we can treat problems which allow both x and y to vary so we relegate that more interesting class of problems to that course.

Remark 14.2.6. *note format.*

Beyond this point the notes will change format. The notes to follow are from previous years however, I have numbered the equations and sections as to be consistent with the numbering up to this point. The page numbering ceases to be meaningful past this point. You can still refer to the section number without ambiguity provided you clarify if it is Stewart or my notes in question. Sorry for the change in format, I didn't have enough time over break to complete the notes to the level of the notes up to this point. Some of you will rejoice in the sudden reduction in proofy-ness.

Chapter 15

Geometry and Coordinates

I have posted a separate pdf of Chapter 15 notes. This particular Chapter of the notes is probably the most disorganized of the course. It is therefore extremely important you take good notes in class and study Stewart in addition to my notes. I will probably add a lot of material in lecture. My apologies this is not better prepared, sadly winter break is far too short.

15.1 polar coordinates

15.2 parametric curves in polar coordinates

15.3 conic sections

15.4 rotated coordinates

15.5 other coordinate systems

Chapter 16

Introduction to Differential Equations

I have posted a separate pdf of Chapter 16 notes. These notes are fairly well prepared. There are many examples I will not have time to cover in lecture. The only thing missing here is Euler's method and direction fields which I will likely cover in addition to these notes. Those topics are heavily numerical so probably I will just do some show and tell with a CAS worksheet or Java applet.

16.1 what is a solution?

16.2 separation of variables

16.3 exponential growth and decay

16.4 integrating factor method

16.5 constant coefficient second order DEqns

Chapter 17

Sequences and Series

I have posted a separate pdf of Chapter 17 notes.

17.1 sequence examples

17.2 convergence of sequences

17.3 series and convergence tests

17.4 error tests

17.5 conditional and absolute convergence

Chapter 18

Basics of Power Series

I have posted a separate pdf of Chapter 18 notes.

18.1 what is a power series?

18.2 power series via geometric series

18.3 power series representations

Chapter 19

Taylor Series and Techniques

I have posted a separate pdf of Chapter 19 notes. I will also cover the proof of Taylor's theorem this semester. That proof was included in my calculus I notes. See §6.5 of my notes.

19.1 Taylor Polynomials

19.2 Taylor's Theorem

19.3 binomial series

19.4 numerical applications of Taylor series

19.5 calculus applications of Taylor series

19.6 physical applications of Taylor series