Calculus I

Chris Johnson

Fall 2024

Contents

Contents ii

Introduction to the Course iv
- College-level mathematics . . . iv
- Calculus at WCU . . . viii
- Dr. Johnson's sections . . . xi
- How to do well in this course . . . xvii
- A warning . . . xxii

1 Limits & Continuity 1
- 1.1 Limits . . . 1
- 1.2 Properties of Limits . . . 18
- 1.3 Continuity . . . 32
- 1.4 The Sandwich Theorem . . . 42
- 1.5 Limits at Infinity and Horizontal Asymptotes . . . 50
- 1.6 The ε-δ definition . . . 58

2 Differentiation 72
- 2.1 Rates of Change . . . 72
- 2.2 Differentiability . . . 82
- 2.3 Derivative Rules . . . 94
- 2.4 Derivatives of Trig Functions . . . 105
- 2.5 The Chain Rule . . . 110

3 Applications of Derivatives 116
- 3.1 Implicit Differentiation . . . 116
- 3.2 Related Rates . . . 127
- 3.3 Linearization & Differentials . . . 132
- 3.4 The Mean Value Theorem . . . 145
- 3.5 Curve Sketching . . . 149
- 3.6 Optimization . . . 158

3.7 L'Hôpital's Rule 165

4 Integration **172**
4.1 Antiderivatives 172
4.2 Riemann Sums 178
4.3 Definite Integrals 191
4.4 The Fundamental Theorem of Calculus 206
4.5 Indefinite Integrals and Net Change 212
4.6 Substitution 219

Introduction to the Course

The reason a lot of people do not recognize opportunity is because it usually goes around wearing overalls looking like hard work.

THOMAS EDISON

Welcome to Math 153, the first semester of the calculus sequence at Western Carolina University. This course is meant to be a first introduction to calculus for students that are well-versed in the prerequisite background of algebra and trigonometry. That is, in this course we will assume that you are comfortable with topics such as solving an equation for a variable, completing the square, using trigonometric identities, and so forth. We *will not* assume you have seen any calculus before, and will take the time to carefully explain calculus topics and their applications in class.

College-level mathematics

As many students in this class will be first-semester freshmen, or upperclassmen taking their first mathematics course, it may be worthwhile to go ahead and discuss some of the general differences between courses in high-school and college, as well as some particular differences in mathematics courses.

General information about courses in college

College-level courses typically require more of students than the courses you have taken in high-school, even if the courses cover similar material. It is generally expected that once you are in college you are capable of some independent learning, and mature enough to realize that you need to set aside dedicated time to study for your courses. Courses in college typically move at a faster pace than courses in high-school and require that you regularly and consistently take notes in class, review your notes after class, read in your textbook, and work on problems and exercises from the textbook to practice the material even if those problems aren't explicitly assigned in class.

In addition, compared to teachers in high school, professors in college are usually much stricter with deadlines, more demanding on homework and exam problems, and less lenient about seemingly minor mistakes. Most professors feel very strongly that students should receive the grade they earned through their work over the semester, and as such *many college professors do not offer extra credit of any kind*! If you happen to have done poorly on an exam, including earning a failing grade, you'll usually have to live with that grade and simply try to do better on future assignments. While in high-school you may have had teachers that allowed you to turn homework assignments in late for some partial credit, or turn in corrections to incorrect problems after an exam, these opportunities are usually non-existent in college. There are many fewer, if any, opportunities to make up poor grades in college, and so you need to take all of your assignments seriously from the very beginning of the semester.

For some students, even students that did well in high-school, the lack of extra credit opportunities, the difficulty of the material, the high expectations of professors, and the very little leniency in grading can come as a very unwelcome surprise. Instead of being scared or worried about these things, though, you should simply be aware of them. Even though there is no denying college courses are more difficult than courses in high-school, there are a lot of resources available to help you. Here are a few things to keep in mind as you make the transition to college:

- If you're struggling in a class, you are almost certainly not alone. Many courses are technical and challenging for most students. However, these courses often have some sort of organized help session where students can work together to understand the material, or sometimes interact with students in more advanced courses who can help guide them through the material they are struggling with. The exact format and schedule of such help sessions varies depending on the course and the department, so you may have to ask around to determine if there are help sessions and when they take place. If in doubt, email your professor and ask them.

- Regardless of whether there are official help sessions for a course or not, interacting with other students in a course can greatly enhance your learning and help you to enjoy college. It's a good idea to try to make friends with other students in your courses so that you have someone else to study with, someone who can possibly explain something you missed in class, or let you copy their notes if you happen to be absent from class one day.

- Professors are typically very upfront and transparent about how they will grade your work, often describing very precisely what their expectations will be at the start of the semester. You may very well have information overload the first few days of class and feel like you won't be able to remember everything, but usually professors list everything you need to know either in the course syllabus, in some document online, or verbally in class.
- Professors want their students to succeed (even though sometimes that can be hard to believe if you have a course with very regular homework, required readings, and difficult exams), and will help you if you come to them for help. Many professors enjoy discussing the material from class with their students and will try to offer alternative explanations or descriptions if students will simply ask. Most professors have *office hours* where they simply wait in their office for students to come by and ask questions about the material from class. If a professor's office hours happen to conflict with your schedule, the professor will usually be willing to meet with you at an alternative time, though you will need to coordinate that time with the professor.

Just to emphasize, your professors are there to help you, but you often have to take the first step in asking for help. You can try to talk to your professor just before or after class, come to their office hours, set up an alternative meeting time, or simply send them an email with any comments or concerns about the course. Having said that, you *should not* expect your professor to do your work for you. For example, if you are working on a homework assignment and you have a problem where you don't even know how to get started, it's okay to ask your professor if they can explain what the question is asking. Usually your professor will then try to clarify anything that isn't clear about the problem and might try to nudge you in the right direction to getting started on the problem. However, the professor will not walk you through every step of solving the problem from start to finish: your professor wants you to figure out how to solve the problem, not solve the problem for you.

Information about mathematics courses

Mathematics courses in college are often more concerned with logic and conceptual understanding, and less concerned with rotely memorized formulas, than classes in high school. For example, professors usually want to see your line of reasoning in solving a problem and award more points for logical work than for the final numerical answer. In fact, many professors will award zero points on a homework or exam problem if the student writes down a correct numerical answer with no supporting work. This often results in students being shocked or upset, feeling they should have received credit for their answer. Professors, though, really want you to learn the concepts and learn how to think and communicate your thoughts. If you simply write down a number without any work, your professor does not know if you actually know how to solve the problem or if you just made a lucky guess. For this reason it is important that you write down all of the details of your calculations.

There is no denying that learning college-level mathematics, regardless of the course or the professor, can be difficult. To truly learn the material and do well in the course, you will need to invest serious time in studying outside of class. You may also need to learn *how* to study; study habits that were successful in high-school may not directly translate to success at college. In particular, simply memorizing formulas or working on problems similar to in-class examples but with a few altered numbers, is usually not sufficient. To study mathematics successfully you need to be incredulous: don't simply take statements of theorems or formulas for granted, but instead try to understand them. That is, try to understand why a term is defined a certain way; understand how each term or factor in a formula makes a contribution; try to find the underlying logic behind why a theorem is true. Learning mathematics this way is often time-consuming, confusing, and frustrating, but it can also be extremely rewarding: there's nothing quite like that *"Ah-ha!"* moment when a topic you struggled with suddenly starts to make sense.

Remark.
If the above worries you or makes you feel anxious, that just means you understand that you will be challenged when it comes to learning college-level mathematics – and that's a good thing! It's much better to be aware of the challenges ahead so that you can think about how to tackle them, than it is to be blindsided and not know what to do.

Calculus at WCU

Calculus is taught at all universities and at many high-schools as well; regardless of where you take calculus you will learn the same basic material, although in slightly different ways and at a slightly different pace depending on the school and the instructor. In this section we'll quickly discuss some of the specifics of calculus as it is taught at WCU.

Three-semester sequence

WCU, like most universities, breaks calculus up into a sequence of three one-semester long courses, each course building on the previous courses. This course, Math 153, is the first of these three courses and is primarily focused on *differentiation*, which is the measurement and computation of how things change. Before discussing differentiation we will discuss the more fundamental notion of a *limit*; everything in calculus is defined in terms of limits, and so we will spend a good bit of time getting comfortable with limits before moving to differentiation. Finally, towards the end of the course we will learn the basics of *integration* which can be thought of in two seemingly different, but closely related, ways: as a method of adding together infinitely-many quantities in a special way, and as a method of doing differentiation "in reverse."

Remark.
Don't worry too much if the terms mentioned above don't mean anything to you right now: the point of the course is to define and study those ideas. Right now we're just giving a very general road map for the topics we will ultimately discuss in this course.

The second calculus course, Math 255, is usually taken immediately after students complete Math 153. (E.g., if you take Math 153 in the fall, you'll probably take Math 255 the following spring.) Math 255 picks up where Math 153 left off by continuing the study of integration. Towards the end of the semester Math 255 switches gears and discusses sequences and series, which are two topics that are fundamental to many more advanced mathematics courses.

Finally, the third calculus course, Math 256, revisits the topics of Math 153 and Math 255, but using functions of multiple variables. This requires introducing a new mathematical object called a *vector* and studying various operations that can be performed with vectors. Many applications of calculus to science and engineering make heavy use of the material discussed in Math 256.

Textbook

There are many, many different calculus textbooks used at different schools across the United States. At WCU, however, the mathematics department has decided to use the OpenStax book *Calculus* by Strang & Herman for all three courses in the calculus sequence. This book contains three volumes, that roughly correspond to the three one-semester courses that are taught at Western. The book is available for free online, but students may purchase a hardcopy (about $30 for Volume 1 through Amazon) if they desire.

Prerequisites and resources for WCU students

As previously mentioned, we will assume that students in this class are comfortable with algebra and trigonometry. If, however, there are certain prerequisite topics you need to look up at some point in the semester (e.g., if you need a refresher on how to complete the square), you should always feel free to ask your professor either in-person during class, or office hours, or through email. If you would like to look up some additional information about prerequisite topics on your own, however, the following resources may be useful:

Stewart's Algebra Review
: http://www.stewartcalculus.com/data/default/upfiles/AlgebraReview.pdf
Intended as an appendix for a calculus textbook written by the late James Stewart, this short document has quick reviews of some of the algebra that is most often used in calculus.

Paul's Online Algebra Notes
: http://tutorial.math.lamar.edu/Classes/Alg/Alg.aspx
Written as notes for his students at Lamar University, Paul Dawkins' online notes contain many interesting examples and do a good job of explaining the concepts behind several topics in algebra.

The Wikibook for Algebra and Trigonometry
: https://en.wikibooks.org/wiki/Algebra

https://en.wikibooks.org/wiki/Trigonometry
Wikibooks are an attempt by the Wikipedia community to write high-quality textbooks about a variety of different topics. Like Wikipedia, these books are under perpetual revision and tend to get better with time as more people contribute to them.

KhanAcademy for Algebra I, II and Trigonometry
https://www.khanacademy.org/math/algebra
https://www.khanacademy.org/math/algebra2
https://www.khanacademy.org/math/trigonometry
KhanAcademy has several videos on precalculus topics, but these links contain the material that is most pertinent for our course.

The Math Tutoring Center

WCU has a Math Tutoring Center (MTC) which is dedicated to helping students succeed in their mathematics courses. Located on the fourth floor of Stillwell, the MTC employs several students that have completed mathematics courses (such as calculus) and were recommended by their professors to be tutors. The MTC offers free tutoring for students both as a drop-in and a scheduled service. That is, you can simply show up at the MTC any time they are open, without having made an appointment, and there should be a tutor available to help you. If there is a particular tutor you want to work with, you can also schedule one-on-one appointments with them. Be aware, though, that tutors are explicitly told to *not* do your work for you, but instead will try to explain concepts and work similar examples with you.

Differences between sections

College courses that have very high enrollment numbers, such as calculus, are often broken up into smaller sections. For example, if 150 students at WCU were enrolled in Math 153, those 150 students might be split up into five sections of thirty students each. When a course is broken up into several sections like this, it is very common that different sections will be taught by different professors. Even though all of the students across all sections are learning the same material, there will be small differences between sections taught by different professors. This can be a cause of confusion if one student is taking one professor's section and they have a friend in a different section with another professor. Even though both students are taking the same course, they may very well have different sets of homework, different exams, and their professors may cover the material in different orders or at different paces.

At WCU, professors have a lot of autonomy in deciding how they are going to teach their sections. Each professor writes their own syllabus, chooses their own homework assignments, quizzes, and exams, decides how those assignments will be graded, and so on. Despite these differences, by the end of the semester all Math 153 students – regardless of which section they were in or which professor they had – will have learned the material. It's just worthwhile to be aware of these differences in case you discuss the course with someone in a different section: don't be surprised if they are covering material we have not yet covered, or are currently covering things we already covered.

Dr. Johnson's sections

Now that we've discussed some of the commonalities across the Math 153 sections, but mentioned the fact there can be differences between sections, let's talk about some of the particularities of Dr. Johnson's sections.

Structure of the class

This semester (Fall 2023), I am teaching two sections of Math 153. Both sections meet three days a week (Monday, Tuesday, and Thursday), with the Monday classes being 50 minutes long and the Tuesday-Thursday classes being 75 minutes long.

Our in-class lectures will typically be very discussion based, where I'll introduce a topic, do some examples, and periodically ask students for their input. I often pause during lecture to ask students if they understand or have any questions, and I strongly encourage students to interact and ask questions as they arise. The class (I hope) will feel very comfortable and informal, and students should be ready to ask questions. I will also sometimes give the class a few minutes to try an exercise on their own, then ask for a volunteer to post their solution to the board, and then we'll discuss the solution. If there are mistakes in the solution, we'll try to use that as an opportunity to discuss any misconceptions or common errors.

My general opinion is that the best way to learn math is to do math, and so I'll plan to give you several exercises and problems to do each week. The in-class exercises just mentioned are not taken up and graded, but just meant to give you a chance to see how well you understand the material as we're going over it in class, and hopefully provide an opportunity for you to ask questions if you realize there's some particular part of the lecture you're not understanding. In addition, we will have short "labs" that are done on Canvas, as well as weekly written homework assignments, and sporadic pop quizzes.

The "labs" will usually be about ten questions long and will cover the topics we discussed in class that day, and are due by noon the next class day. For example, if we discuss the quotient rule in class on Tuesday, then there will be a lab about the quotient rule that's due by noon on Thursday; if we discuss derivatives of trig functions on Thursday, the corresponding lab is due by noon the following Monday (since our class meets Monday-Tuesday-Thursday). The point of the labs is to give us a way to be sure everyone is following along with the basic idea of the material from day-to-day. The questions are not meant to be overly complicated, but just to test your basic knowledge. (That's not to say the problems are always "easy;" some concepts we will learn in class are a little tricky the first time you see them, and so the corresponding lab might also appear a little bit tricky.) As the labs are on Canvas, you will know instantly once you submit them which problems you got correct, and which problems you got incorrect. After the due date of the lab, the correct answers to all of the problems will be visible for everyone.

While the Canvas labs are meant to be relatively straight-forward, the weekly homework assignments will typically be more difficult. The homework is an opportunity for us to dive into the topics from class a little deeper, and so the problems will often require multiple steps, or may be more conceptual and less computational in nature, or will require some more serious algebraic manipulations than what we usually have time for in our in-class exercises and the Canvas labs. Students often need to invest significant time working on the homeworks outside of class, and will very likely need to come to office hours to ask questions, or visit the Math Tutoring Center to receive help with the homeworks. (I'm not trying to be mean to you by giving you difficult homeworks, but the way: instead, I'm trying to push you a little bit to help strengthen your understanding of the material and build up your problem-solving skills.) Homework will usually be due by 11:59pm on Mondays, and you will upload your homework solutions to Canvas. You will need to upload a single PDF file with your solutions. You can create this file either by typing up your solutions, or by creating a scan of your hand-written solutions using an app such as Adobe Scan or Microsoft Lens. *Your scan must be readable to receive credit! If I can't read your work, whether because of poor handwriting or a low-quality scan, you will not receive credit!*

Homework assignments are graded out of ten points, where (roughly) five points will come from an effort grade regardless of whether your answers are correct or not, and the remaining points come from accuracy where I will grade one or two problems on your homework to see if you did the problem correctly. When I grade for accuracy, I'm not really concerned with your final numerical answer to a problem, but rather the process you used to solve the problem. If you have the right setup and idea for the solution, and just make a minor algebraic or arithmetic mistake, you'll get most of the credit. If you only have an answer and no supporting work, however, you'll receive zero credit.

After homework assignments are due, I'll plan to post a key to the assignment with detailed solutions on Canvas. *You should look at the key to the assignment after it's posted!* There is a very good chance that homework problems will reappear during a midterm exam, or the final, and so you want to be sure you know how to correctly solve the problems.

I do not plan to take attendance during class, except for the very beginning of the semester. However, *you really need to be coming to class when you are physically able*. To encourage this, I will sporadically give pop quizzes at the start of class sometimes. These will be very short (sometimes just a single question), and are not graded for accuracy. Instead, it serves as a way for me to see who is and isn't coming to class. It also serves as a starting point for our discussion for the day. Just to be clear: the pop quizzes are worth just one point each, and you will receive that point as long as you are in class for the pop quiz, even if your answers to the questions are incorrect. The only way to receive zero credit for a pop quiz is to not be in class for it, whether that's due to an absence or being late.

Remark.
During the COVID-19 pandemic, I started teaching this class as a *flipped classroom* where students were expected to watch lecture videos outside of class. This semester I am not using the flipped classroom model anymore, but have decided to make the previously recorded lecture videos available for you as supplementary study material. So, if you missed a class for whatever reason, you can go and watch the corresponding lecture videos to get an idea of what we did while you were away. This should also be helpful resource as you're studying for exams: if you realized there's a particular topic you don't understand as well you as you'd like, you can watch

the corresponding video and hopefully that will help to clear things up.

Midterm exams

During the semester we will have three midterm exams on Tuesdays or Thursdays. On these days students will have the entire 75 minutes to complete an exam. The problems on the midterm will range from being trivial (to make sure students know the basics) to being more involved (to see that students are able to apply the concepts from class).

Calculators and other technology will not be allowed on midterm exams, but the problems will be constructed in such a way that any required arithmetic will be simple enough that students can quickly do the calculation in their head or with pencil and paper. For example, you may need to add simple fractions (such as $\frac{1}{4} + \frac{2}{5}$) or multiply small-ish integers (e.g., 13×7). Sometimes you may need to leave your answers in terms of trigonometric functions, or as expressions involving roots or irrational numbers, *but that's okay*. It is totally fine to leave an answer as, say, $e^{\sqrt{\pi}}$ – you are not expected to be able to convert such a number to a decimal on an in-class assignment.

Midterm exams will always be graded out of 100 points, with each problem graded for accuracy. Partial credit is awarded when students begin solving a problem correctly but make mistakes or simply stop solving the problem. However, students *must* begin solving the problem correctly to receive partial credit. Students will not receive partial credit for completely erroneous or illogical work, or for solving a problem different from what is asked on the exam.

The work you present on a midterm exam is expected to be written legibly and easy to follow.

All students are expected to take each midterm exam. As discussed in the *Make-up exams and homework* section below, make-up exams will only be allowed in a few specific circumstances. Students should always prepare to take the midterm at the date and time announced in class. ***Test anxiety is not a legitimate reason to delay an exam.*** Students who miss an exam for an unexcused reason will receive a grade of zero.

Tentatively, our midterm exams will take place on Tuesday September 19, Tuesday October 24, and Thursday November 16. These dates are subject to change, however.

Final exam

We will have a cumulative final exam at the end of the semester, the exact date and time of which will be determined by the registrar. The structure and format of the final exam is very similar to that of the midterm exams, though the final will be somewhat longer and counts for a larger portion of the student's grade.

Make-up assignments

Generally speaking, no late work is accepted and no make-ups for missed assignments are allowed. Of course, there are exceptions to this. For example, if you are seriously ill or suddenly injured, then we will work together to find a reasonable solution to a missed assignment. Or, if you are student-athlete that will miss class because you are traveling with your team to a university-sanctioned event *and you notify me before you leave with documentation from your coach*, then we will find a reasonable solution to what you have missed. However, if you happen to miss class the day of an in-class assignment or when a written homework is due because you overslept, are hungover, or simply too anxious or feel unprepared, you **will not** be allowed to make up any missed assignments and ***you will receive a grade of zero on that assignment!***.

As labs and homeworks are taken up on Canvas, I do not plan to grant extensions or make-ups for these (even for university-sanctioned travel), except in very extreme circumstances. Similarly, missed pop quizzes will have a recorded grade of zero, even if you have an "excused" absence. To compensate for this, I will drop some to-be-determined number of pop quizzes for everyone at the end of the semester.

Extra credit

There is no extra credit of any form in this class.

Expectations

Students in this class are expected to be mature and conduct themselves in a professional manner. In terms of this classroom this means

- students are expected to come to class each day;
- be in class prepared with pencil and paper at the start of class

- students should have completed the assigned reading before coming to class;
- pay active attention during class and have any computers, phones, or tablets put away (students *may* take notes on a tablet, however);
- and be ready to participate in class by asking questions about examples from the previous lecture, problems from homework assignments, or any concepts discussed in class or the assigned reading.

Students are expected to spend a *minimum* of eight hours per week working on material for Math 153 (working on homework, reading the textbook, studying notes, etc.). Keep in mind eight hours is the minimum: each additional hour spent working outside of class will have been well-invested come exam time.

Students are strongly encouraged to take advantage of the various studying resources provided by the university and the mathematics department, such as the MTC.

Online notes

In addition to the textbook, I will be typing up my lecture notes for the course and posting them online in Canvas. Students are expected to read both the online lecture notes as well as the OpenStax textbook. The readings for each week will be posted to Canvas.

Theorems and proofs

Throughout the online notes and the textbook you will see several lemmas, theorems, and proofs. These first two terms simply refer to any true mathematical statement, although we usually use the term *lemma* to mean a relatively simple statement, and reserve *theorem* for a particularly important statement (e.g., the Pythagorean theorem). A *proof* is a logical argument detailing why that lemma or theorem is actually true.

We will see many lemmas and theorems throughout the semester, and it's important that you make an attempt to remember and understand any lemmas or theorems as you read them in the book, notes, or in lecture. Immediately after the statement of a lemma or theorem you will often see a proof. Proofs are very important in mathematics: they basically justify why the things we claim are true are actually true.

Despite the importance of proofs, for this particular course you will never be asked to recite a proof on an exam. (Although you may be asked to do a few simple derivations in a homework assignment from time-to-time.) Our goal in this class is to be able to apply the ideas of calculus to solve problems, and we can actually do that without necessarily knowing the proofs of all the tools we use. (This is analogous to being able to drive a car without understanding the mechanics of internal combustion engines.) I've included proofs in the online textbook mostly for the sake of completeness, and also so any particularly interested student that is curious why a given theorem is true can read the proof if they want to. You *do not* need to try to memorize the proofs that appear in the course, you don't really even need to read them when they appear in the text or the notes if you don't want to. (Having said that, if you're planning to major in mathematics you may want to at least make an effort to read through the proofs, even if you don't understand all of the details.)

If a proof of a theorem isn't particularly complicated, we may see the proof in class or a lecture video. This may seem a bit odd: if you aren't required to memorize the proof, why bother seeing it? Without a proof, you're basically just taking everything on faith. Some of the theorems we will see are not obvious, and without a proof they will feel like very mysterious statements out of the blue that somehow magically work. Discussing the proofs in class at least convinces us the theorem is true and is logical instead of magical. Even if we never repeat the proof again, it's somehow comforting to be aware that our theorems are true.

How to do well in this course

I firmly believe that every student is capable of succeeding in this course, but I also know that some students will struggle and so I want to mention a few concrete things that you can do to succeed in the course.

Recognize that this class is difficult

In the first few weeks of the semester this class may seem relatively straight-forward, but you should not let this lull you into thinking the entire semester will be easy. One of the difficult things about this course is that we will cover a lot of material, and each new topic will build off the previous ones. If you start to fall behind, you will have a very small window of time to get back up to speed before being behind will negatively affect your grade.

Study every day

To do well in this course you will need to invest a significant amount of time into studying outside of class. Sitting in lecture, even if you feel like you understand what is going on during the lecture, will not be enough. You should get into the habit of studying every day: not just the days the class meets, not just on weekdays, but every day. Something as simple as putting aside one hour for individual studying outside of class each day can have a huge impact on your grade and keep you from falling behind. Sometimes you will have other commitments that prevent you from getting an hour each day, but when at all possible, you should really try to study at least one hour each day.

When you study you should first review your notes from class; notice this implies that you need to be taking notes in class. If there is something from your notes that you don't understand, try to figure it out. It's best if you try to figure things out on your own first without having to look in a book or online: sometimes you just need to spend a few minutes thinking through the details of some calculation or the logic behind some argument before it starts to make sense.

Read your book and the online lecture notes. The book and the lecture notes cover the same material, but sometimes presented in different ways. By reading both you see the same ideas from two different points of view. This can be helpful because one point of view may "click" when the other does not.

Communicate

Communicate with other students in the class, with friends in another section of Math 153, with students that have already completed Math 153, with people interested in math online, and anyone else you can think of. Don't simply ask other people to do your homework, but discuss the material with them – even if you think you already know the material. Sometimes other people will have insights that you were unaware of and can shed light on something you didn't understand, or can help you see something you thought you understood from another angle.

You can use websites such as the Math Stackexchange, https://math.stackexchange.com or the Learn Math https://www.reddit.com/r/learnmath/ or Calculus https://www.reddit.com/r/calculus/ subreddits to talk about mathematics with other people. You are welcome to use sites such as this to help get clarification on topics from class, however if you use these sites to help you solve a problem on a written homework assignment you *must* cite the source by giving a URL to the site (or a description of how to get to that site) on your homework. Using these sources without proper citation will be considered a violation of the university's honor code.

Be incredulous

To do well in advanced math courses you should try to think like a mathematician. This means trying to understand the ins and outs of every argument, why each step in a computation was performed, what earlier results were used, etc. In general, you should be incredulous: you should not simply take it on faith that what we have learned in class is true (even though it is!), but you should instead always ask why it's true and try to figure out the reason. This one piece of advice, if taken to heart, subsumes everything else.

Practice, then practice some more

You know that you understand a concept well and are prepared for an exam when you have practiced so much that solving problems becomes mechanical. For example, think about solving for x in an equation such as $x^2 = x + 2$. The first time you started learning algebra this may have seemed very odd and difficult, but as you did more examples you started to notice the patterns and the tricks, and now you (hopefully) are able to solve for x in equations like the one above without any trouble.

Similarly, in this class you will probably find some computations and some logical arguments very difficult and time-consuming at first. If you do enough examples, however, then the things that at first seem difficult and confusing will over time become second nature. The only way for this to happen is to invest time in practicing. When you review your notes and see an example that we did in class, try to reproduce the result without looking at the notes and then look back at the notes if you get stuck or make a mistake. Pick and choose extra problems at the end of the sections in the book; make up your own problems; look for extra problems online. The more practice you do, the easier everything will be when you actually sit down to take an exam.

Prepare for exams

The biggest mistake you can make when an exam is coming up is to put off studying for it. The earlier you start preparing for an exam, the better. When an exam is coming up, start adding more time to your usual study sessions. Ideally you should add an extra hour each day for a week leading up to an exam. It's probably best to try to split this up into two one-hour study sessions each day instead of doing two hours at once.

When you're preparing for an exam, you should study as if the upcoming exam is the hardest one you have ever taken in your life. (This isn't to say that it necessarily will be the hardest exam you ever take, but it's better to over-prepare than to under-prepare.) If there is a topic you don't feel comfortable with or are worried about, don't ignore it! Study as if the problems you dislike and find hard will be on the exam: chances are at least a couple types of problems you dislike will make their way onto an exam at some point.

To prepare for an exam you should review all of the relevant notes, look over old homeworks and labs (as well as their keys), and try to understand any mistakes that you made. It is very likely that problems from homework or labs will reappear on an exam.

Come to class

For some strange reason there always seem to be people who think it's okay to skip class. You should come to class each and every day which you are physically able. In class you should be actively paying attention to the lecture and trying to think through examples as they are done on the board. If you have questions in class, then that's good! Having questions means that you're thinking, which is what you should be doing in class.

You *should not* be daydreaming, working on assignments for other classes, playing games on your phone, or checking Facebook, Twitter, Instagram, Snapchat, Tinder, etc.

Start assignments early and work on them often

You will usually have about a week to do a homework assignment, and that is for a reason. Some of the questions on these assignments will be difficult and you will have to spend some time thinking in order to do the assignment. You should really try to start on assignments early, meaning the day they are assigned, and try to do a few problems each day. You should also anticipate that some questions are going to require a lot of time – maybe an hour or more for the hardest questions. If you wait until the last minute to do an assignment, you won't have time to get it done.

Get help when you need help

There are a lot of resources to help you succeed in this class and it would be wise to take advantage of them. You can ask questions and discuss the course material with your classmates, come to office hours, or attend tutoring sessions in the MTC. You can also work with other students and look up resources online like Khan Academy and MIT's Open Course Ware. You can also email me (cjohnson@wcu.edu), drop in during my office hours, or schedule an appointment outside of office hours.

In general when you have a specific concern about an assignment or a topic from class, you should try to address your concern by taking advantage of resources in the following order:

1. Try to figure things out yourself. There will be plenty of times when you just need to spend a little bit of time thinking on your own and you can figure things out.

2. Check in the book or lecture notes. Many questions you have will be answered in the book or notes, you just have to take the time to look through the book/notes and find it.

3. Ask a classmate. Sometimes you may have a misconception or misheard something in lecture, and asking a friend might be all it takes for you to realize your mistake.

4. Make a one-on-one tutoring appointment with through the MTC. The MTC tutors know calculus extremely well and can probably answer any questions you have, or at least help you get started in answering the problem yourself.

5. Email me or come to office hours. I put this at the end of the list not because I'm trying to avoid seeing you or talking to you, but just as a matter of practicality. If everyone in the class came to me the instant they had a question I would spend my entire day answering their questions. I am fine with answering your questions or talking to you when you have concerns, but I also have other classes and other responsibilities that take up my time and so you can probably get your questions answered quicker through one of the options described above.

If you have more serious concerns about your standing in the class – not simply a homework problem you can't figure out – then by all means contact me first.

Don't stress out (too much)

There will be times when this class frustrates you: maybe there is a topic you can't seem to wrap your head around, or a problem that you feel like you're staring at and have no idea how to get started. This is completely normal and you shouldn't get too stressed out about it. This class is going to be hard and you are going to get confused and feel stuck sometimes, but that is just a normal part of learning difficult material. The important thing to remember is that you should persevere. If you're getting frustrated, take a break: go get something to eat, play a game, read a book, take a nap; do something you enjoy for a little while and then get back to work when you're ready.

A warning

It is relatively common that among first-year students taking a calculus class at a university, about half of the students took calculus in some form in high-school and the other half of the students are seeing calculus for the first time. **In this course we will make no assumptions that students have seen calculus before.** We will take the time to carefully motivate and define all of the calculus terms and ideas we use throughout the semester.

It is also very common that students who have had some sort of calculus before think they understand calculus much better than they really do. Our class will likely be much more thorough and go into much more depth than any high-school calculus course you may have taken, and as such you really need to take this class seriously from the very start of the semester. In order to make the class fair to everyone (whether they have taken calculus before or not), when we perform various calculations in class ***you will not be allowed to use any short-cuts you know from a previous class if we have not discussed those short-cuts in class!*** For example, students that have had calculus know there is a very simple trick for calculating the derivative of a polynomial such as $x^2 + 3x - 2$. Students learning calculus for the first time, however, will learn how to do this calculation using limits. And as such if we have a problem like this on an assignment all students will be expected to use the limit definition and not the shortcut, at least until we discuss the relevant shortcuts in class.

There is no denying that this course is difficult, and students will need to work hard to do well in this course. I firmly believe, however, that all students are capable of succeeding in this course *if they are willing to study regularly, start on assignments early, and take the course seriously from the very beginning*. You should be reading these lecture notes and the corresponding sections of the textbook *before* coming to class, and come to class with questions. No question is too simple or basic, and you should feel free to ask questions anytime you have them. You can ask questions during class time, or before or after class, or during office hours, or through email. I will always do my best to try to give you a complete answer to your question that you can understand. I also encourage you to work outside of class with other students. Sometimes simply bouncing an idea off of someone else can help you see how to start on a problem, and explaining a concept to someone else can help solidify your own understanding. *You are strongly encouraged to work on out-of-class assignments with other people!*

The lecture notes

The notes you are reading are in their eighth incarnation, having evolved from the handwritten examples I used when I first taught a version of this course as a graduate student at Clemson University. I have used a version of these notes at Clemson, Wake Forest, Indiana, Bucknell, and WCU. Despite this, there may be places where the notes are "rough around the edges" and may contain typos and mistakes (though hopefully those are all minor). If you see something in the notes you think is a mistake, it may very well be, and it would be greatly appreciated if you would email me (cjohnson@wcu.edu) to let me know about any mistakes. While these notes are my primary resource for the examples I use in the lectures, they should not be a substitute for the textbook. Besides the fact that your textbook has fewer mistakes than these notes (probably not mistake-free, but relatively few and minor mistakes) since it was professionally edited, the textbook also has lots of exercises and practice problems, which these notes do not. I hope these notes are helpful to you, but you should not use them as your only source of study material.

Chris Johnson
Fall 2024

Limits & Continuity 1

1.1 Limits

Pure mathematics is, in its way, the poetry of logical ideas.

ALBERT EINSTEIN
Obituary of Emmy Noether, New York Times

What we refer to today as "calculus" is a collection of mathematical tools, techniques, and ideas that are related by a common theme: if a quantity of interest can not be calculated directly, approximate it with something easier to calculate, and then try to systematically improve your approximations. While we usually think of calculus as being developed by Isaac Newton and Gottfried Leibniz in the 17th century, this basic idea goes back to antiquity. We'll begin our study of calculus by reviewing some of the history that lead to the development of the mathematics we will study throughout the semester.

The key idea: approximation

The crucial ideal behind all of calculus, approximate and then improve your approximation, is startlingly simple and goes back to at least the time of the ancient Greeks. One of the earliest examples of this principle is Archimedes' approximation of the mathematical constant π, which is the ratio of a circle's circumference to its diameter. You have heard many times before that π is $3.14159...$[1]. But if you had never been told that π was $3.14159...$, how could you go about determining this value?

Calculating this value, the ratio of a circle's circumference to its diameter, may seem intimidating at first glance. It was Archimedes who realized, however, that this value could be approximated by replacing the circle with something simpler: a polygon. It is relatively easy to compute the length of the perimeter of a polygon and compare this to the

[1]The decimal expression for π is infinitely long and contains no repetition. For this reason we will generally prefer to refer to this number as π and not write out the decimal expansion. Writing out the decimal expansion of π will always be slightly incorrect because no matter how many digits you include there are infinitely-many digits you have excluded.

furthest distance between two points on the polygon: this is very similar to comparing a circle's circumference to its diameter, but much easier to actually calculate. If we take an equilateral triangle whose *diameter* is 2, some simple trigonometry will tell us that the lengths of the sides of the triangle are each about 1.73205. Adding up the lengths of each of the three edges and then dividing by the diamter of 2, we'd calculate that π was approximately 2.59808. You may not consider this to be a good approximation to π, but we're only getting started.

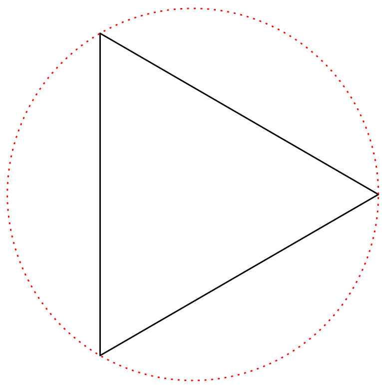


Figure 1.1: Approximating π using a triangle.

To get a better approximation we can replace the triangle with something closer to a circle: a polygon with more sides, like a square. Repeating the calculation above we would approximate π to be 2.828427. Again, this is only an estimate, but we can replace the square with a pentagon, hexagon, heptagon, etc. to obtain better and better approximations for π: as we look at polygons with more and more sides, our polygons begin to resemble a circle and we obtain better approximations for π.

Archimedes carried this idea out for a polygon with ninety-six sides and estimated that π was roughly 3.14103. Considering that Archimedes lived around 200BC and so all of his calculations would have been done by hand, this is a very respectable estimate!

The Idea of a Limit

The basic idea of calculus is to approximate quantities you can't compute directly, and then improve your approximations. The technical tool we use to understand what happens as we improve our approximations is called a *limit*, and the next few lectures will be devoted to understanding

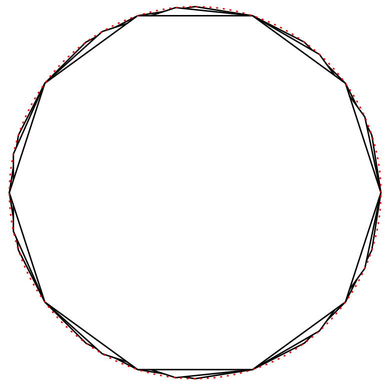


Figure 1.2: Approximating π using polygons with 10, 15, and 20 sides.

precisely what limits are, how to calculate limits, and some of the important properties of limits. Everything else in calculus is defined in terms of limits: limits are always lurking in the background in calculus, even though sometimes we'll look at things in a more abstract way and won't work with limits directly.

To ge an idea of what limits are, we will first limits of sequences and then move on to limits of functions.

Limits of sequences

A ***sequence*** is an infinitely long, ordered list of numbers. For example,

$$1, {}^1\!/_2, {}^1\!/_4, {}^1\!/_8, {}^1\!/_{16}, \ldots$$

The "..." above is used to indicate that the pattern continues. In this particular case the pattern is that each element in the sequence is half of the previous element. Notice that in this situation the numbers are getting smaller and smaller, but are always positive. In a situation such as this, the numbers in the sequence become *arbitrarily close* to zero. What we mean by this is that if you go far enough out in the sequence you can get as close to zero as you like. Zero never appears in the sequence, but the numbers are getting closer and closer and closer to zero. We say that zero is the ***limit*** of this sequence.

As another example, consider the sequence

$$2.1, 1.9, 2.01, 1.99, 2.001, 1.999, 2.0001, 1.9999, \ldots$$

In this sequence the numbers are getting arbitrarily close to two. Notice the numbers jump back and forth between being a little above two and a little below two, but are always getting closer and closer to two. Here we say the limit of the sequence is 2.

Let's consider one last example:

$$^1/_2,\ -^3/_4,\ ^7/_8,\ -^{15}/_{16},\ ^{31}/_{32},\ \ldots$$

What are the numbers in this example getting close to? Well, some of the numbers seem to be getting closer and closer to 1, while other numbers are getting closer and closer to -1. So, what should the limit of this sequence be? Since the numbers in the sequence aren't getting close to any one particular thing, we say the limit ***does not exist***.

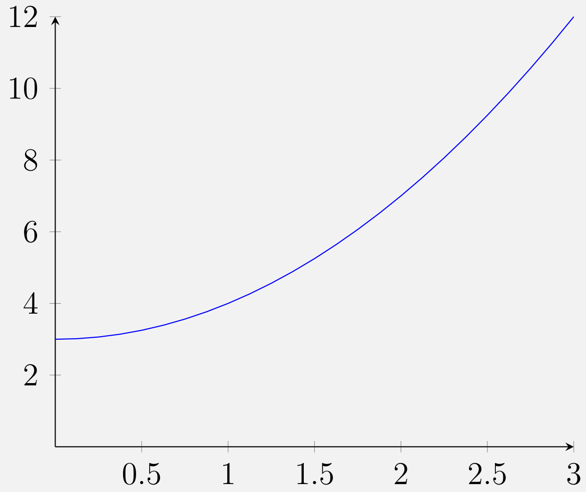
The very first question you should always ask yourself when dealing with limits is whether the limit exists or not. When a limit does not exist we usually write ***DNE***.

Limits of Functions

While limits of sequences are relatively straight-forward to understand (either everything in the sequence starts getting close to some particular number, or the limit doesn't exist), what we'll care about more in this class are limits of functions. The core idea is essentially the same: we want to see what values the output of a function gets close to, just as we saw how the numbers in a sequence were getting close to something.

To begin our study of limits of functions we'll start off just by looking at graphs of some simple functions.

Example 1.1.
Consider the function $f(x) = x^2 + 3$ which is graphed in the figure below.



What happens to the outputs of the function (the y-values) as the inputs (the x-values) get close to $x = 2$? As we can see from the graph, as the inputs get closer and closer to $x = 2$, the outputs get closer and closer to 7. We thus say the *limit of $f(x)$ as x approaches 2 is equal to* 7. Notationally we write this as

$$\lim_{x \to 2} f(x) = 7 \quad \text{or} \quad \lim_{x \to 2} \left(x^2 + 3\right) = 7.$$

All this means is that as the input x gets very close to 2, the output $f(x)$ gets very close 7.

In general, if the output of a function $f(x)$ gets arbitrarily close to a number L as the inputs to x gets arbitrarily close to some value a, we say ***the limit of $f(x)$ as x approaches a is equal to L*** and we write

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

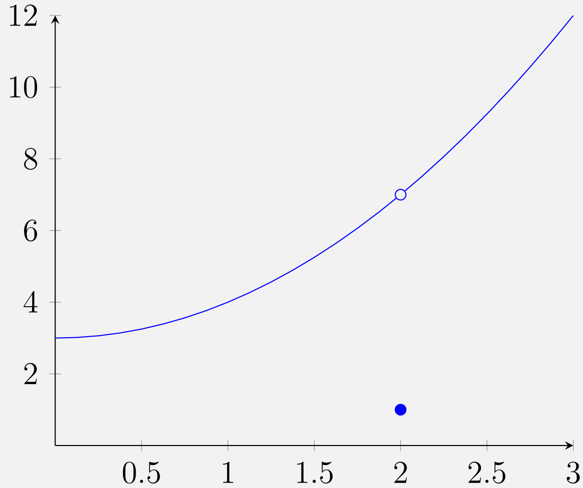
In the example above, when $f(x) = x^2+3$, you may think that the limit was 7 because the value of the function at 2 was equal to 7; $f(2) = 7$. The idea of a limit is a little bit more subtle than that, however, as the next example illustrates.

Example 1.2.
Let's modify the previous example just slightly replacing $f(x) =$

$x^2 + 3$ with the following piecewise function,

$$g(x) = \begin{cases} x^2 + 3 & \text{if } x \neq 2 \\ 1 & \text{if } x = 2 \end{cases}$$

The graph of this function appears below.



Notice that $g(x)$ above is almost exactly the same as the function $f(x)$ in the earlier example: the only difference between the functions is that $f(2) = 7$ while $g(2) = 1$. What do you think the limit should be in this situation?

In this example as the inputs get close to $x = 2$, the outputs get closer and closer to 7. Notice that 7 is not actually attained near 2, but the outputs $g(x)$ never the less get arbitrarily close to 7 as x gets close to 2. Here again, the limit of the function is equal to 7 as x approaches 2:

$$\lim_{x \to 2} g(x) = 7.$$

The above example should convince you that limits may not be quite as simple as your intuition would naïvely lead you to believe.

Example 1.3.
Consider the function $f(x) = \frac{x^2-4}{x-2}$. Before we look at the graph of this function, let's take a minute to do a little bit of algebra.

First notice that this function is not defined when $x = 2$: when x equals 2, the expression in the denominator is zero, but division by zero is undefined! Since division by zero is undefined, this function $f(x)$ is undefined whenever that denominator would be zero which occurs when $x = 2$. So whatever the graph of the function looks like, it must have a "hole" at $x = 2$ where the function is undefined.

Aside from the "hole" that must appear when $x = 2$, what else can we say about the graph of the function? Let's maybe see if we can simplify the expression $\frac{x^2-4}{x-2}$ that defines our function.

If you take an expression like $(x + a)(x - b)$ and expand it (by FOILing, for example) you get $x^2 + (a - b)x - ab$. If $a = b$ this simplifies to $x^2 - a^2$. That is, $(x + a)(x - a) = x^2 - a^2$. An expression of the form $x^2 - a^2$ is sometimes called a *difference of perfect squares*, and the above bit of algebra shows us that it's very easy factor a difference of perfect squares: $x^2 - a^2 = (x + a)(x - a)$.

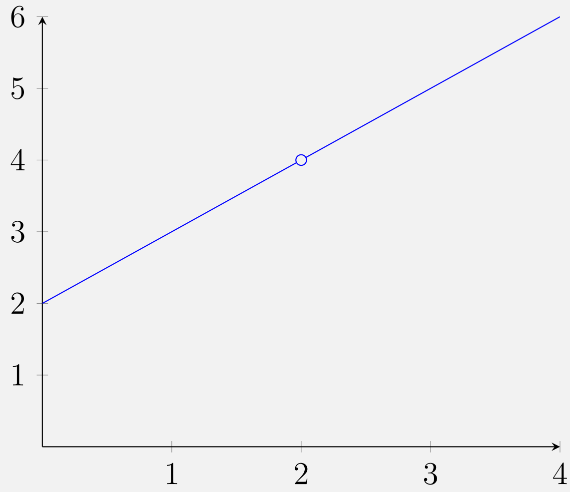
Noticing that the numerator in $f(x) = \frac{x^2-4}{x-2}$ is a difference of perfect squares, we can rewrite the numerator as $x^2 - 4 = (x + 2)(x - 2)$: that is,

$$f(x) = \frac{x^2 - 4}{x - 2} = \frac{(x + 2)(x - 2)}{x - 2}.$$

Again, our function is not defined at $x = 2$, but away from $x = 2$ we can cancel out the factor of $x - 2$ that appears in the numerator and denominator of our function. Hence

$$f(x) = x + 2 \text{ provided } x \neq 2.$$

So the graph of our function $f(x) = \frac{x^2-4}{x-2}$ is the same as the graph of $x + 2$, which is of course just a line, except with a hole at $x = 2$.



Now that we know what the graph of the function looks like, let's consider the limit of this function as x approaches 2.

As x approaches 2, the outputs $f(x)$ get arbitrarily close to 4 as we can see from the graph. Notice this function never equals 4 anywhere, but does get very, very, very close to 4 as x gets very, very, very close to 2. Thus the limit of $f(x)$ as x approaches 2 is equal to 4:

$$\lim_{x \to 2} f(x) = 4.$$

So far we've seen three different situations relating the limit of a function at a point to the function's value at that point:

1. The limit was equal to the value of the function.
2. The limit and the function took on different values.
3. The limit was defined, but the function was not.

Let's look at a few more examples of the different types of things that could happen with limits.

Example 1.4.
Let $f(x) = \frac{x}{|x|}$. Notice that this function is undefined at $x = 0$ as this would result in division by zero. As we saw in the last example, however, we may still be able to make sense of the limit of this

function as x approaches 0.

As before, we'd like to look at a graph of this function to help us reason about what the limit should be. So, what should the graph of this function look like?

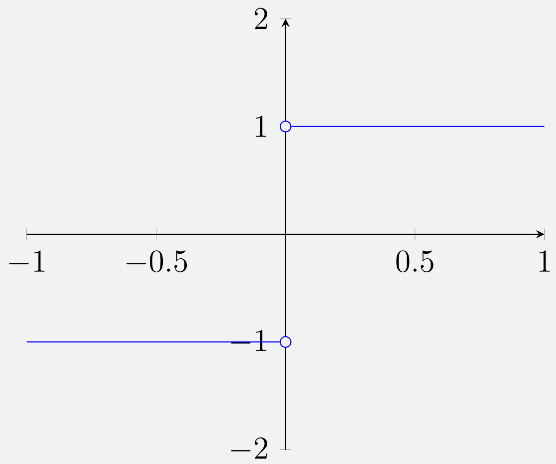
Notice that if $x > 0$, then $|x| = x$. Thus

$$f(x) = \frac{x}{|x|} = \frac{x}{x} = 1 \text{ provided } x > 0.$$

If $x < 0$, however, we'd have $|x| = -x$ and so

$$f(x) = \frac{x}{|x|} = \frac{x}{-x} = -1 \text{ provided } x < 0.$$

So $f(x)$ is 1 for positive values of x, -1 for negative values of x, and undefined when x equals zero.



Now that we have the graph to help us, what is the limit of $f(x)$ as x approaches 0? If the limit were some number L, whatever L happens to be, then what we'd like to see is that the outputs $f(x)$ get arbitrarily close to L as x gets close to 0. Notice that this *can not* happen for *any* value of L!

The obvious possible choices for the limit would be 1 or -1. The limit can't be 1, though, since there are values close to 0 (namely the negative values) where $f(x)$ does not get close to 1. Similarly, the limit can't be -1 since there are nearby values (the positive values of x) where $f(x)$ doesn't get close to -1.

We could make the same sort of argument for any other value someone claimed the limit should be: no matter what value you think the limit should be, there will be x-values near zero so that the outputs $f(x)$ don't get arbitrarily close to your choice of limit. For this reason the limit does not exist:

$$\lim_{x\to 0} f(x) \text{ DNE.}$$

Notice in the previous example where the limit did not exist we wrote

$$\lim_{x\to 0} f(x) \text{ DNE,}$$

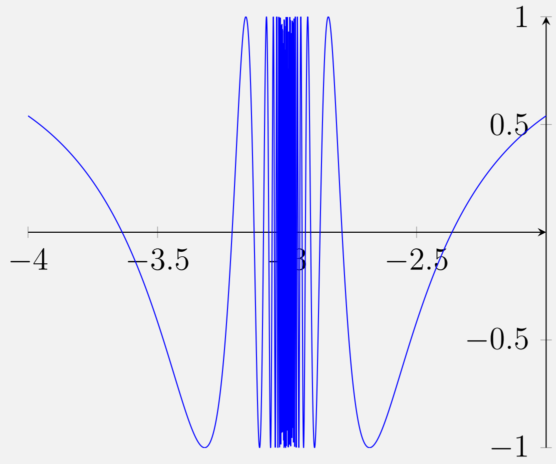
and we did not write $\lim_{x\to 0} f(x) = \text{DNE}$. Writing $= \text{DNE}$ is incorrect because DNE is not a value: it's short-hand for saying the value doesn't exist!

Example 1.5.
Consider the function $f(x) = \cos\left(\frac{1}{x+3}\right)$. Let's again think about what the graph of this function should be before actually viewing the graph. Recall that the outputs of the cosine function are always between positive and negative one: $-1 \leq \cos(x) \leq 1$ for all values of x. So the outputs of $\cos\left(\frac{1}{x+3}\right)$ have to be bound between -1 and 1.

Recall also that $\cos(x)$ oscillates between -1 and 1 infinitely-many times. In particular, $\cos(x)$ is 1 whenever x is an even multiple of π, and $\cos(x)$ is -1 whenever x is an odd multiple of π. Hence $\cos\left(\frac{1}{x+3}\right)$ is equal to 1 if $\frac{1}{x+3}$ is an even multiple of π, and equal to -1 if $\frac{1}{x+3}$ is an odd multiple of π.

Notice that as x gets very close to -3, $\frac{1}{x+3}$ gets very, very large (or very, very negative, depending on whether x is to the left or right of -3). In fact, $\frac{1}{x+3}$ is an even multiple of π for infinitely-many values of x which are close to -3! Similarly, $\frac{1}{x+3}$ is an odd multiple of π for infinitely-many values of x which are close to -3.

What this tells us is that $\cos\left(\frac{1}{x+3}\right)$ jumps back and forth between -1 and 1 more and more and more as x approaches -3.



Since the outputs of this function does not get close to any one value as x gets close to -3, again we have that the limit of the function does not exist as x approaches -3!

$$\lim_{x \to -3} \cos\left(\frac{1}{x+3}\right) \text{ DNE.}$$

Exercise 1.1.
Suppose that $f(x)$ is the following piecewise function:

$$f(x) = \begin{cases} 4 - x^2 & \text{if } x \leq 1 \\ -2x + 5 & \text{if } x > 1 \end{cases}$$

Plot the graph $y = f(x)$ and then determine the limit of the function as x approaches 1, or explain why the limit does not exist.

Exercise 1.2.
Let's slightly modify the function from the previous exercise: sup-

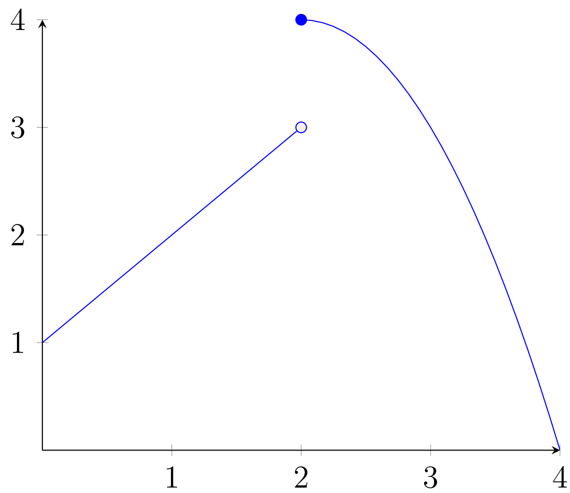
pose $f(x)$ is the following piecewise function:

$$f(x) = \begin{cases} 3 - x^2 & \text{if } x \leq 1 \\ -2x + 5 & \text{if } x > 1 \end{cases}$$

Plot the graph $y = f(x)$ and then determine the limit of the function as x approaches 1, or explain why the limit does not exist.

One-Sided Limits

In the examples above we saw that sometimes the limit of a function may not exist because the function oscillates infinitely-many times between two values, but other times the function seemed to approach two different values: one from the left, and one from the right. In this second situation, the graph of the function appeared to have a "break" in it that prevented the function form having a limit at that breakpoint, as in the figure below. This type of situation is common enough that it has a special name.



$$f(x) = \begin{cases} 4x - x^2 & \text{if } x \geq 2 \\ x + 1 & \text{if } x < 2 \end{cases}$$

When the values of $f(x)$ get closer and closer to a single value L as inputs x get closer and closer to a *from the right* (i.e., all values of x are larger than a), then we say ***the limit of $f(x)$ as x approaches a from the right*** is L. This is denoted

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

In the figure above, for example, $\lim_{x \to 2^+} f(x) = 4$.

Similarly, if the outputs $f(x)$ become arbitrarily close to M as the inputs x get closer to a *from the left* (i.e., all values of x are smaller than a), then we say ***the limit of $f(x)$ as x approaches a from the left*** is M, and we write

$$\lim_{x \to a^-} f(x) = M.$$

In the example above, $\lim_{x \to 2^-} f(x) = 3$.

We call these two types of limits, from the left and from the right, ***one-sided limit***. The limit we discussed earlier is sometimes called a ***two-sided limit***, although when people just say ***limit*** they usually mean the two-sided limit.

Two-sided limits and one-sided limits are related by the following important property:

Theorem 1.1.
The two sided limit $\lim_{x \to a} f(x)$ exists and equals L if and only if both one-sided limits of $f(x)$ at $x = a$ exist and are equal to L.

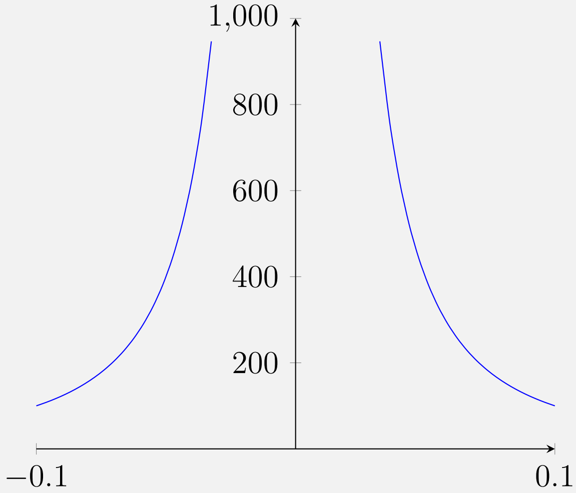
Intuitively, this theorem tells us that the only way that the outputs $f(x)$ can get arbitrarily close to L as the inputs get close to a, is if the outputs get close to L regardless of whether the nearby inputs are to the left or right of a. A simple consequence of this theorem is that if the one-sided limits disagree, then the two-sided limit can not exist.

Vertical Asymptotes

Limits, both one-sided and two-sided, mean that the outputs of a function $f(x)$ become arbitrarily close to some value as the inputs get close some particular number. Sometimes, though, instead of getting closer to a particular number, the inputs simply grow without bound. When this happens *the limit does not exist*, but we will sometimes say the limit is infinity. This language, and the notation we will introduce, can be a little bit confusing. The important thing to remember is that if the limit exists, the output is getting close to a particular number. If the limit is infinity, the outputs are not getting close to a number, but instead are growing and growing.

Example 1.6.
A simple example of such a situation is given by the function $f(x) = {}^1/_{x^2}$.

As x gets closer to zero, whether from the left or the right, we are dividing 1 by smaller and smaller positive numbers. This division yields increasingly larger and larger numbers that can be made arbitrarily large.



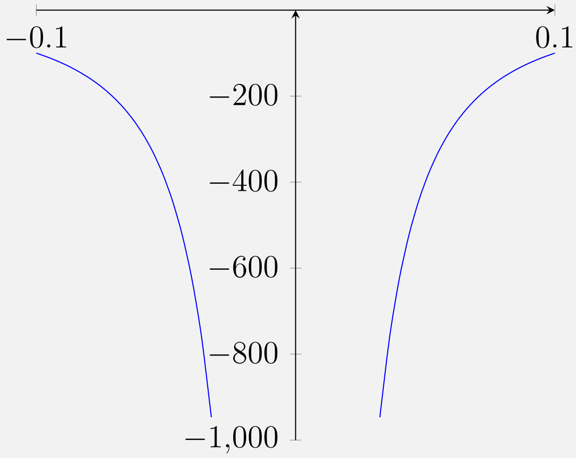
We say that the limit as x goes to 0 of ${}^1/_{x^2}$ is infinite and we write

$$\lim_{x\to 0}\frac{1}{x^2} = \infty.$$

Notice that even though we say the limit is infinity, the limit does not exist!

Instead of the outputs becoming larger and larger, it could happen that the outputs of $f(x)$ become more and more negative without bound. In this situation we say the limit is negative infinity, though as before the limit does not exist.

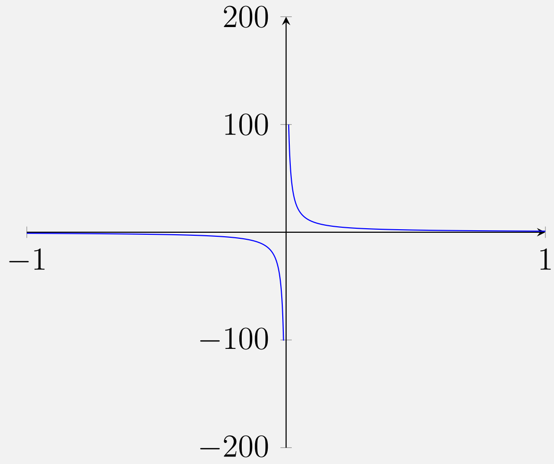
Example 1.7.
Consider the function $f(x) = -1/x^2$.



As x goes to zero, the outputs $-1/x^2$ become more negative. In this situation we write

$$\lim_{x\to 0} -\frac{1}{x^2} = -\infty$$

Example 1.8.
It could happen that a one of the one-sided limits goes to positive infinity, while the other goes to negative infinity:

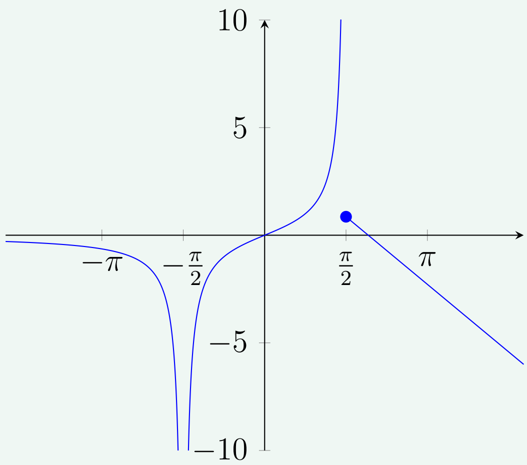


In such a situation we have

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

If either one sided limit as x approaches a of $f(x)$ is positive or negative infinity, we say that the vertical line $x = a$ is a ***vertical asymptote*** of the graph $y = f(x)$.

Exercise 1.3.
Find all of the vertical asymptotes of the function $f(x)$ graphed below.



Remark.
WeBWorK is extremely picky about how you enter limits that do not exist, or which are infinite. In particular, when a limit does not exist sometimes WeBWorK will want you to enter the text `DNE`, but sometimes it will want you to enter something else like `F`. Unfortunately it changes from problem to problem, so you have to read the statement of the problem to see which one WeBWorK wants.

Additionally, limits at infinity are special. By our definition of a limit, a limit at infinity does not exist (since infinity is not a number), however these limits do not exist for a particular reason. In those situations some WeBWorK problems will want you to enter `DNE` (or some other text) or infinity. Both of these *should* be correct, but most problems on WeBWorK will only accept one or the other. Luckily you have unlimited attempts on WeBWorK problems, so you can try again if you enter infinity and WeBWorK was expected `DNE`.

1.2 Properties of Limits

Algebra is the offer made by the devil to the mathematician.

MICHAEL ATIYAH

We saw in the last lecture that the limit of a function tells us what the function is doing near a given point. When the limit exists, the outputs of the function are getting arbitrarily close to some given value. Even if the normal, two-sided limit does not exist, the one-sided limits may, and these tell us what values the function approaches as the inputs get close to a particular value from the left or the right. We also saw that there are many things that may prevent a limit from existing. It could be that the function behaves wildly, jumping back and forth between several values with getting close to one particular value, or it could be that the values of the function "blow up" to infinity or negative infinity instead of getting close to a number.

In this lecture we will move from trying to reason about limits in terms of graphs, to calculating limits algebraically. This is predicated on a few assumptions: we need a few basic building blocks before we can do any serious calculations. We'll begin by first giving two very basic building blocks, and then looking at the myriad of ways these simple functions can be combined into more interesting things.

Our definition of limit is still "hand-wavy" for the time-being, meaning the outputs get "close" to some value L as the inputs get "close" to a – without saying precisely what "close" means. In the next lecture we'll come back and make the definition of limit precise, but until then the theorems in this lecture will have to be taken on faith. After the next lecture, however, we will be in a position to come back and rigorously prove everything from this lecture.

Two Building Blocks

Before we can really go anywhere with calculating limits, we need to start with a few simple things. Our two building blocks for more complicated functions will be the constant functions, $f(x) = c$ where c is some fixed number, and the identity function, $f(x) = x$.

The constant functions

By a ***constant function*** we mean any function which *always* spits out the same number. The graph of any such function is just a horizontal line.

Theorem 1.2.
If $f(x) = c$ is a constant function, then for every real number a the limit as x approaches a of $f(x)$ equals c.

$$\lim_{x \to a} c = c.$$

Here are a few simple examples just for the sake of illustration:

$$\lim_{x \to 3} 7 = 7 \qquad \lim_{x \to -2} \pi = \pi \qquad \lim_{x \to 13} -4 = -4.$$

The identity function

The ***identity function*** is the function whose output always equals its input:

$$f(x) = x$$

The graph of this function is a straight line through the origin with slope 1.

Theorem 1.3.
If $f(x) = x$, then for every real number a the limit as x approaches a of $f(x)$ equals a:

$$\lim_{x \to a} x = a.$$

Again, a few simple examples:

$$\lim_{x \to 7} x = 7 \qquad \lim_{x \to 19} x = 19 \qquad \lim_{x \to -3} x = -3.$$

The Limit Laws

Now that we have a few building blocks to play with, we can construct more interesting functions by adding, subtracting, multiplying, and dividing our basic building blocks. For example, a polynomial like $3x^5 - 7$ is built by multiplying the identity function with itself 5 times, then multiplying the result of that by the constant function 3, and then subtracting from that the constant function 7. We know how to calculate limits with our building blocks, so now we need to understand what happens to the limits when we put these building blocks together.

The following theorem applies for *all* functions $f(x)$ and $g(x)$ whose limit as $x \to a$ exists, but today we will specifically care about the case where the functions $f(x)$ and $g(x)$ are the identity function or a constant function.

Theorem 1.4.
Suppose that $f(x)$ and $g(x)$ are any two function whose limits as $x \to a$ exist (in particular, the limits are finite). Then the following limit laws hold:

1. *A limit of sums is a sum of limits:*
$$\lim_{x \to a} (f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x).$$
2. *A limit of differences is a difference of limits:*
$$\lim_{x \to a} (f(x) - g(x)) = \lim_{x \to a} f(x) - \lim_{x \to a} g(x).$$
3. *A limit of products is a product of limits:*
$$\lim_{x \to a} (f(x)g(x)) = \left[\lim_{x \to a} f(x)\right] \left[\lim_{x \to a} g(x)\right].$$
4. *If $\lim_{x \to a} g(x) \neq 0$, then*
$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}.$$

The above theorem basically says we can do arithmetic with limits the way we would hope to.

This theorem combined with our building blocks means that we can now algebraically evaluate limits of lots of different types of functions without bothering to figure out what the graph looks like.

Example 1.9.
Calculate the limit of $3x-4$ as x approaches -2.

Using our limit laws above, we can break up $\lim_{x\to -2}(3x-4)$ a bit at a time until we get to our building blocks:

$$\begin{aligned}
&\lim_{x\to -2}(3x-4)\\
=&\lim_{x\to -2}3x-\lim_{x\to -2}4\\
=&\left(\lim_{x\to -2}3\right)\left(\lim_{x\to -2}x\right)-\lim_{x\to -2}4\\
=&3\cdot(-2)-4\\
=&-6-4\\
=&-10.
\end{aligned}$$

Example 1.10.
Calculate the limit of $6x^3-x^2+2x+1$ as x approaches 3.

$$\begin{aligned}
&\lim_{x\to 3}\left(6x^3-x^2+2x+1\right)\\
=&\lim_{x\to 3}6x^3-\lim_{x\to 3}x^2+\lim_{x\to 3}2x+\lim_{x\to 3}1\\
=&\left(\lim_{x\to 3}6\right)\left(\lim_{x\to 3}x\right)\left(\lim_{x\to 3}x\right)\left(\lim_{x\to 3}x\right)-\left(\lim_{x\to 3}x\right)\left(\lim_{x\to 3}x\right)+\left(\lim_{x\to 3}2\right)\left(\lim_{x\to 3}x\right)+\left(\lim_{x\to 3}1\right)\\
=&6\cdot 3\cdot 3\cdot 3-3\cdot 3+2\cdot 3+1\\
=&162-9+6+1\\
=&160.
\end{aligned}$$

Polynomials and Rational Functions

A ***monomial*** is a function of the form $f(x) = cx^n$ where c is a constant, and n is a positive integer. Our theorem above tells us that the limit of a monomial is simply the monomial evaluated at a particular value.

Theorem 1.5.
If $f(x) = cx^n$ is a monomial, then for every real number a

$$\lim_{x \to a} f(x) = ax^n.$$

Proof.

$$\begin{aligned}
&\lim_{x \to a} cx^n \\
&= \left(\lim_{x \to a} c\right) \underbrace{\left(\lim_{x \to a} x\right) \cdots \left(\lim_{x \to a} x\right)}_{n \text{ times}} \\
&= c \cdot \underbrace{x \cdot x \cdots x}_{n \text{ times}} \\
&= cx^n.
\end{aligned}$$

□

Example 1.11.

$$\lim_{x \to -2} 7x^3 = 7(-2)^3 = -56$$

A ***polynomial*** is simply a sum of monomials: a function of the form

$$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$$

where the $c_0, c_1, ..., c_n$ are some constants, and n is a positive integer. For example,

$$19x^7 - 11x^3 + 4 \quad \text{and} \quad -x^3 + x^2 - x$$

are polynomials.

Since we know that we can evalute the limit of a monomial by evaluating it, and the limit of a sum is the sum of the limits, we now know that we can also evaluate polynomials simply by evaluating them:

Theorem 1.6.
Let $f(x)$ be a polynomial, say

$$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0.$$

Then for every real number a,

$$\lim_{x \to a} f(x) = f(a) = c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0.$$

Proof.
We simply break our polynomial up into monomials, and calculate the limit of each monomial by evaluating:

$$\begin{aligned} &\lim_{x \to a} \left(c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 \right) \\ = &\lim_{x \to a} c_n x^n + \lim_{x \to a} c_{n-1} x^{n-1} + \cdots + \lim_{x \to a} c_1 x + \lim_{x \to a} c_0 \\ = &c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0 \end{aligned}$$

□

For example, we are now justified in saying something like the following:

$$\begin{aligned}
&\lim_{x\to 4}\left(3x^5 - x^2 + 3\right)\\
=&3\left(4^5\right) - 4^2 + 3\\
=&3\cdot 1024 - 16 + 3\\
=&3072 - 13\\
=&3059.
\end{aligned}$$

A ***rational function*** is simply a ratio of polynomials:

$$f(x) = \frac{c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0}{d_m x^m + d_{m-1}x^{m-1} + \cdots + d_1 x + d_0}.$$

The fourth limit law above tells us that limits of quotients are quotients of limits *provided the denominator is not zero*! In terms of rational functions, we thus have an easy way of evaluating limits as long as the denominator does not go to zero.

Theorem 1.7.
Suppose that $f(x)$ is a rational function. That is, suppose that we can write $f(x)$ as a ratio

$$f(x) = \frac{g(x)}{h(x)}$$

where $g(x)$ and $h(x)$ are polynomials. Provided that $h(a) \neq 0$, then for any real number a we have

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

Proof.
Again, suppose that $f(x) = \frac{g(x)}{h(x)}$ where $g(x)$ and $h(x)$ are polynomi-

als and $h(a)$ is not zero.

$$\begin{aligned}
&\lim_{x\to a} f(x)\\
=&\lim_{x\to a}\frac{g(x)}{h(x)}\\
=&\frac{\lim_{x\to a} g(x)}{\lim_{x\to a} h(x)}\\
=&\frac{g(a)}{h(a)}\\
=&f(a).
\end{aligned}$$

□

Example 1.12.

$$\begin{aligned}
&\lim_{x\to 2}\frac{x^3-3x^2+1}{x-7}\\
=&\frac{2^3-3\left(2^2\right)+1}{2-7}\\
=&\frac{8-12+1}{-5}\\
=&\frac{-3}{-5}\\
=&\frac{3}{5}
\end{aligned}$$

Notice that the theorem tells us absolutely nothing if the denominator $h(x)$ does happen to give 0 at $x=a$. It could be that the limit exists, as in the case of

$$\lim_{x\to 2}\frac{x^2-4}{x-2}=2,$$

or it could be the case that the limit does not exist as in

$$\lim_{x\to 0}\frac{1}{x}\ \text{DNE};$$

it simply depends on which rational function you're dealing with.

The important thing to remember when working with limits, however, is that we only care about what the function does *near* a point, not at the point. The following theorem makes this precise:

Theorem 1.8.
Suppose that $f(x)$ and $g(x)$ are two functions with the following property: for all values of x "near" a, except possibly a itself, $f(x) = g(x)$. Then

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

That is, if one limit exists, then so does the other, and the two limits equal the same value.

This theorem gives us a useful tool for calculating limits at points not in the domain of the function: if we can do some algebra to "replace" the function we care about with another function whose limit we can actually compute, then we know the limit of the original function.

Example 1.13.
Let's compute the following limit:

$$\lim_{x\to 1} \frac{x^2+x-2}{x^2-5x+4}.$$

Notice that 1 is not in the domain of this function since we would get division by zero if we plugged in $x = 1$. However, the theorem above says that if we replace this function with another function that agrees nearby, then the limits will be the same.

To find such a function, let's try to do a little bit of algebra. The numerator x^2+x-2 factors as $(x-1)(x+2)$, and the denominator x^2-5x+4 factors as $(x-1)(x-4)$:

$$\frac{x^2+x-2}{x^2-5x+4} = \frac{(x-1)(x+2)}{(x-1)(x-4)}.$$

As long as we're not plugging 1 into the function (since this results in division by zero), we can cancel out the $x-1$ factors:

$$\frac{x^2+x-2}{x^2-5x+4} = \frac{x+2}{x-4} \text{ provided } x \neq 1.$$

So we have two functions,

$$\frac{x^2+x-2}{x^2-5x+4} \text{ and } \frac{x+2}{x-4}$$

which agree everywhere except 1. The theorem tells us these two functions have the same limit as $x \to 1$, and the function on the right is something whose limit we can calculate:

$$\lim_{x\to 1} \frac{x^2+x-2}{x^2-5x+4} = \lim_{x\to 1} \frac{x+2}{x-4} = \frac{3}{-3} = -1.$$

Example 1.14.

$$\begin{aligned} \lim_{x\to 3} \frac{x^2+x-12}{x^2-9} &= \lim_{x\to 3} \frac{(x-3)(x+4)}{(x-3)(x+3)} \\ &= \lim_{x\to 3} \frac{x+4}{x+3} \\ &= \frac{7}{6} \end{aligned}$$

Example 1.15.

$$\begin{aligned}\lim_{x\to 1}\frac{x^4-1}{x-1} &= \lim_{x\to 1}\frac{(x^2-1)(x^2+1)}{x-1}\\ &= \lim_{x\to 1}\frac{(x-1)(x+1)(x^2+1)}{x-1}\\ &= \lim_{x\to 1}(x+1)\left(x^2+1\right)\\ &= 4\end{aligned}$$

Powers and Roots

The next theorem we will have to take on faith until the precise definition of the limit is given in the next lecture.

Theorem 1.9.
Suppose $f(x)$ is a function whose limit as $x \to a$ exists. Then for any real number r,

$$\lim_{x\to a}(f(x))^r = \left(\lim_{x\to a} f(x)\right)^r$$

assuming $(\lim_{x\to a} f(x))^r$ exists.

Corollary 1.10.
Suppose $f(x)$ is a function whose limit as $x \to a$ exists. Then for any real number r,

$$\lim_{x\to a}\sqrt[r]{f(x)} = \sqrt[r]{\lim_{x\to a} f(x)}$$

assuming $\sqrt[r]{\lim_{x\to a} f(x)}$ exists.

Proof.

$$\begin{aligned}\lim_{x\to a} \sqrt[r]{f(x)} &= \lim_{x\to a} \left(f(x)\right)^{1/r} \\ &= \left(\lim_{x\to a} f(x)\right)^{1/r} \\ &= \sqrt[r]{\lim_{x\to a} f(x)}\end{aligned}$$

□

Example 1.16.
Calculate $\lim_{x\to 2} \sqrt{\frac{2x^2+1}{3x-2}}$.

$$\begin{aligned}\lim_{x\to 2} \sqrt{\frac{2x^2+1}{3x-2}} &= \sqrt{\lim_{x\to 2} \frac{2x^2+1}{3x-2}} \\ &= \sqrt{\frac{2\cdot 2^2+1}{3\cdot 2-2}} \\ &= \sqrt{\frac{9}{4}} \\ &= \frac{3}{2}\end{aligned}$$

Example 1.17.

$$\begin{aligned}\lim_{x\to 4} \frac{x-4}{\sqrt{x}-2} &= \lim_{x\to 4} \frac{(\sqrt{x}-2)(\sqrt{x}+2)}{\sqrt{x}-2} \\ &= \lim_{x\to 4} \left(\sqrt{x}+2\right) \\ &= 4\end{aligned}$$

Example 1.18.

$$
\begin{aligned}
&\lim_{x\to 2} \frac{\sqrt{x^2+x}-\sqrt{x^2+2}}{x-2} \\
&= \lim_{x\to 2} \frac{\sqrt{x^2+x}-\sqrt{x^2+2}}{x-2} \cdot \frac{\sqrt{x^2+x}+\sqrt{x^2+2}}{\sqrt{x^2+x}+\sqrt{x^2+2}} \\
&= \lim_{x\to 2} \frac{x^2+x-(x^2+2)}{(x-2)\sqrt{x^2+x}+\sqrt{x^2+2}} \\
&= \lim_{x\to 2} \frac{x-2}{(x-2)\sqrt{x^2+x}+\sqrt{x^2+2}} \\
&= \lim_{x\to 2} \frac{1}{\sqrt{x^2+x}+\sqrt{x^2+2}} \\
&= \frac{1}{2\sqrt{6}}
\end{aligned}
$$

Exercise 1.4.
Compute the following limit:

$$\lim_{x\to 9} \frac{x^2-8x-9}{\sqrt{x}-3}.$$

Properties of One-Sided Limits

All of the properties we've discussed still apply if we replace the two-sided limit with the right-hand or left-hand limits.

Example 1.19.

Calculate $\lim_{h\to 0^+} \frac{\sqrt{2+h}-\sqrt{2}}{h}$.

$$
\begin{aligned}
&\lim_{h\to 0^+} \frac{\sqrt{2+h}-\sqrt{2}}{h} \\
&= \lim_{h\to 0^+} \frac{\sqrt{2+h}-\sqrt{2}}{h} \cdot \frac{\sqrt{2+h}+\sqrt{2}}{\sqrt{2+h}+\sqrt{2}} \\
&= \lim_{h\to 0^+} \frac{(2+h)-2}{h\left(\sqrt{2+h}+\sqrt{2}\right)} \\
&= \lim_{h\to 0^+} \frac{h}{h\left(\sqrt{2+h}+\sqrt{2}\right)} \\
&= \lim_{h\to 0^+} \frac{1}{\sqrt{2+h}+\sqrt{2}} \\
&= \frac{1}{\lim_{h\to 0^+}(2+h)+\sqrt{2}} \\
&= \frac{1}{\sqrt{2}+\sqrt{2}} \\
&= \frac{1}{2\sqrt{2}}
\end{aligned}
$$

Exercise 1.5.
Compute the following limit:

$$\lim_{x\to 3^-} \frac{x^2-x-6}{|x-3|}$$

1.3 Continuity

Mathematics, rightly viewed, possesss not only truth, but supreme beauty.

BERTRAND RUSSELL

One of the nice things about functions like polynomials is that it is extremely to calculate limits of such functions: we simply plug in! In this lecture we want to discuss which other functions have this property, $\lim_{x\to a} f(x) = f(a)$, and what the consequences of this are.

The functions satisfying this property are called ***continuous***, and continuity is an extremely nice property for a function to have because it makes our lives much simpler. We'll give the formal definition of what it means for a function to be continuous in just a moment, but the idea is that *small changes in inputs result in small changes in outputs*. Many of the functions we actually care about (i.e., functions that appear "in nature") are in fact continuous: temperature, barometric pressure, etc. are all continuous function.

Definition and examples

Continuity at a Point

We say that a function $f(x)$ is ***continuous at a point*** $x = a$ if the following three conditions are satisfied:

1. f is defined at $x = a$ (i.e., $f(a)$ is defined),
2. $\lim_{x\to a} f(x)$ exists, and
3. $\lim_{x\to a} f(x) = f(a)$.

That is, for a function to be continuous at a point simply means we can take the limit of the function at the point by just plugging in $x = a$. Intuitively, this means that what the function does at $x = a$ is described by what the function does near $x = a$.

Example 1.20.

The function $f(x) = \frac{x^2-x-2}{x-3}$ is continuous at the point $x = 4$ since

$$\lim_{x\to 4} f(x) = \lim_{x\to 4} \frac{x^2 - x - 2}{x - 3} = \frac{16 - 4 - 2}{4 - 3} = 10 = f(4).$$

However, this function is not continuous at $x = 3$ since it fails the first of the three conditions above: the function is not defined at $x = 3$.

Example 1.21.
The function $f(x) = \frac{\sin(x)}{x}$ is can not be continuous at $x = 0$ since it is undefined at $x = 0$, even though the limit as $x \to 0$ does exist.

Example 1.22.
Consider the function

$$g(x) = \begin{cases} \frac{\sin(x)}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

This function is still not continuous at $x = 0$: even though it is defined at $x = 0$ and the limit exists, the limit of the function as $x \to 0$ is not the same as $g(0)$:

$$\lim_{x\to 0} g(x) = 1 \neq 0 = g(0).$$

Example 1.23.
Consider the function

$$h(x) = \begin{cases} \frac{\sin(x)}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$$

This function *is* continuous at $x = 0$: it is defined at $x = 0$ and now the value $h(0)$ agrees with the limit:

$$\lim_{x\to 0} h(x) = 1 = h(0).$$

Example 1.24.
For what values of x is the function $f(x) = \frac{x^2+17x-13}{(x^3-x+2x^2-2}$ continuous?

This function will be defined as long as its denominator is not zero. In fact, since this is a rational function, we know that the limit $\lim_{x\to a} f(x)$ will be equal to the function value $f(a)$ as long as the denominator is not zero at $x = a$. That is, the function will be continuous everywhere except where its denominator is zero.

By factoring the denominator,

$$x^3-x+2x^2-2 = x(x^2-1)+2(x^2-1) = (x+2)(x^2-1) = (x+2)(x-1)(x+1),$$

we see that the function is continuous everywhere except at $x = -2$, $x = -1$, and $x = 1$.

Example 1.25.
What value of b makes the function $f(x)$ below continuous at $x = -1$?

$$f(x) = \begin{cases} x^2 + bx & \text{if } x < -1 \\ 3x + 1 & \text{if } x \geq -1 \end{cases}$$

This function is defined at $x = -1$, and in fact $f(-1) = -2$. So in order to be continuous we need that $\lim_{x\to -1} f(x) = -2$. Currently we have the following:

$$\lim_{x\to -1^+} f(x) = \lim_{x\to -1^+} (3x+1) = -2$$
$$\lim_{x\to -1^-} f(x) = \lim_{x\to -1^-} \left(x^2 + bx\right) = 1 - b$$

Thus we require $-2 = 1 - b$, and so $b = 3$ will make this function continuous.

Continuity in an Interval

We say that a function is ***continuous in an interval*** (a, b) if it is continuous at every point in the interval. That is, for every value c satisfying $a < c < b$, the three conditions for continuity at a point $x = c$ above are satisfied:

1. $f(c)$ is defined,
2. $\lim_{x \to c} f(x)$ exists, and
3. $\lim_{x \to c} f(x) = f(c)$.

Example 1.26.
The function $\frac{x}{x+2}$ is continuous in the interval $(1, 3)$, but not in the interval $(-3, -1)$ since it is not continuous at $x = -2$.

Example 1.27.
Every polynomial is continuous in the interval $(-\infty, \infty)$.

More generally, if a function is continuous at every point where it's defined, we simply say that the function is ***continuous***.

Example 1.28.
Every polynomial is continuous.

This can lead to a little bit of confusion if we don't pay careful attention to the definition. For example, the function $\frac{1}{x}$ is continuous (it is continuous everywhere it's defined); but the function *is not* continous at $x = 0$ because it's not defined at $x = 0$. This distinction is subtle, but important.

The Intuition

The technical definition of continuity may seem a little odd, so it's important to have some intuition for what this definition is really saying. What it means for a function to be continuous is essentially this: *small changes in the input x result in small changes in the output $f(x)$*. That is, the function can't jump around wildly. Because of this, it's usually pretty easy to tell if a function is continuous or not by considering its graph.

Graphically, a function is (usually) continuous if its graph $y = f(x)$ doesn't have any breaks in it: that is, the function has to appear as a continuous curve. Sometimes it's described this way: a function is continuous if you can draw its graph without lifting your pencil. There are some technical reasons why you shouldn't take this as the definition of continuity, but it's a good bit of intuition to keep in mind.

Notice that most functions that occur in the natural world are continuous. The temperature in a room as a function of your position in the room, for example, is continuous: the temperature does not instantaneously jump from $72°$ to $83°$, but instead changes gradually with sufficiently small changes in position resulting in correspondingly small changes in temperature.

The velocity of a car as it travels down the highway is also a continuous function: the car's velocity does not instantly go from 35mph to 60mph, but gradually changes, hitting all values inbetween 35 and 60 at some point in time.

Examples of Continuous Functions

Almost all functions we will discuss in this class will be continuous, in fact most (but not all) of the functions that appear in science and engineering problems are continuous. Given that intuition that continuity means "small changes in inputs result in small changes in outputs," this shouldn't be terribly surprising. If we measure the temperature in Farhrenheit as a function of time, $F(t)$, then we would expect the temperature to change gradually as time passes: the temperature isn't $-18°F$ one moment and then instantaneously $192°F$: it has to heat up from $-18°F$ to $192°F$, along the way hitting $-17°F$, $-3°F$, $42°F$, and so on. This can happen relatively quickly (e.g., it's conceivable that in some sort of very hot industrial oven this could be happen in under a second), but it doesn't happen instantaneously.

Mathematically, *most* of the functions you'd write down if I asked you to write down a function are going to be continuous. Polynomials, rational functions, trig functions, logarithms, and so on are all continuous functions. Combining this with a few rules for combining continuous functions, it's easy to believe that if you were to start writing some expression of x that you wanted to use as a function, you'd write down a continuous function.

To make this precise, let's begin with a few basic functions that can be shown to be continuous.

Theorem 1.11.
All of the families of functions described below are continuous:

1. *Polynomials*
2. *Rational functions*
3. *Trigonometric functions (sine, cosine, tangent, secant, cosecant, cotangent)*
4. *Logarithmic functions (with any base!)*
5. *Exponential functions (e.g.,* 2^x*,* e^x*,* $\left(\frac{382}{4813}\right)^x$*)*

Proof.
Homework: think about this on your own. How could you formally justify that something like the claim that all polynomials are continuous? It's too time-consuming to do in class, but you have all of the tools at your disposal to prove this claim at least for polynomials. For the other functions things are a little trickier, so we'll have to take this on faith for the moment. □

Because of this theorem, we know that it's *extraordinarily* easy to calculate limits of the five types of functions above.

Example 1.29.

$$\begin{aligned}\lim_{x\to 0}\cos(x) &= \cos(0) = 1\\ \lim_{x\to \pi/4}\csc(x) &= \csc(\pi/4) = \sqrt{2}\\ \lim_{x\to 1}\ln(x) &= \ln(1) = 0\\ \lim_{x\to -1/3} 8^x &= 8^{-1/3} = \frac{1}{2}\end{aligned}$$

In fact, we can combine the continuous functions above to get even more continuous functions:

Theorem 1.12.
If $f(x)$ and $g(x)$ are continuous at $x = a$, then so are each of the following:

1. $f(x) + g(x)$
2. $f(x) - g(x)$
3. $f(x)\,g(x)$
4. $\frac{f(x)}{g(x)}$ *provided* $g(a) \neq 0$.

Furthermore, if $f(x)$ and $g(x)$ are continuous functions (so continuous at every point in their domain), then so are the four functions above.

Proof.
This is an immediate consequence of our limit laws. By assumption, $f(a)$ and $g(a)$ are defined; $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist; and $\lim_{x\to a} f(x) = f(a)$, $\lim_{x\to a} g(x) = g(a)$. Thus the limit laws give the following:

1. $\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = f(a) + g(a)$.

2. $\lim_{x\to a}(f(x)-g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = f(a)-g(a)$.
3. $\lim_{x\to a}(f(x)\,g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = f(a)\,g(a)$.
4. $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{f(a)}{g(a)}$ as long as $g(a) \neq 0$.

The second statement is a consequence of the definition of a continuous function. □

Example 1.30.
Each of the functions below is continuous:

- $x^3 + \cos(x)$
- $\frac{\tan(x)-x^3}{\sec(x)+\ln(x)}$
- $\ln(x) - \frac{x}{1+\sin(x)}$

The following theorem also allows us to take limits of compositions of continuous functions:

Theorem 1.13.
If $f(x)$ and $g(x)$ are functions satisfying the following two properties:

1. $\lim_{x\to a} g(x) = b$, *and*
2. $f(x)$ *is continuous at* b,

then

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

Sketch of Proof.

As x approaches a, $g(x)$ approaches b, and so $f(g(x))$ approaches $f(b)$. Another way to say this would be to introduce a new variable $z = g(x)$. Then because f is continuous at $z = b$,

$$f(b) = \lim_{z \to b} f(z) = f(\lim_{z \to b} z)$$

now notice that $g(x) \to b$ when $x \to a$, and we may write

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

□

Corollary 1.14.
A composition of two continuous functions is continuous where defined.

Example 1.31.
Each of the functions below is continuous:

- $\cos\left(\frac{x^3}{1-x^2}\right)$
- $\tan\left(\frac{e^{x^2+1}}{x+\sin(x^2)}\right)$
- $\sin(\cos(\tan(x)))$

The Intermediate Value Theorem

One important property of continuous functions which we've already hinted at is that they can't instantaneously jump between values. Instead, they have to gradually move between any two given values, hitting everything inbetween along the way. The way to make this precise is given by the intermediate value theorem.

Theorem 1.15 (The Intermediate Value Theorem)**.**
Suppose that $f(x)$ is continuous on the interval (a, b) and that $f(a) = \alpha$, $f(b) = \beta$ and $\alpha \leq \beta$. Then for every value γ between α and β, there exists a c in (a, b) such that $f(c) = \gamma$.

The intermediate value theorem has several interesting consequences. For example, since temperature is a continuous function of time, if the low last night was $68°F$ and the high today is $87°$. Then, because of the intermediate value theorem, there must exist a time when the temperature is exactly $78.9101112131415161718192021 22....°F$. If you're driving your car and you start at $0mph$ and accelerate to $35mph$, at some point you are driving at exactly $10\pi mph$.

Mathematically, we can use the intermediate value theorem to show that certain equations must have solutions: even if we are unable to calculate them!

Example 1.32.
Does there exist an x such that $x^3 + \cos(x) = -xe^{\sin(x)}$?

If you try to answer this by explicitly solving for x, you will quickly realize that algebraically this is exceptionally difficult: maybe even impossible. However, the IVT easily allows us to determine there must be a solution.

Notice the equation above may be rewritten as

$$x^3 + \cos(x) + xe^{\sin(x)} = 0.$$

The left-hand side of this expression is a continuous function. It's easy to show that if $x = -\pi$, then

$$(-\pi)^3 + \cos(-\pi) + (-\pi)e^{\sin(-\pi)} = -\pi^3 - 1 - \pi e^0 < 0.$$

Similarly, if $x = \pi$, then

$$\pi^3 + \cos(\pi) + \pi(e^{\sin(\pi)}) = \pi^3 - 1 + \pi > 0.$$

By the intermediate value theorem, there must exist an x between $-\pi$ and π such that x satisfies the above equation.

1.4 The Sandwich Theorem

"Obvious" is the most dangerous word in mathematics.

E. T. Bell

We'll now discuss one very important theorem that allows us to calculate some limits that we can't evaluate using continuity properties, or our algebraic rules described earlier. This theorem often goes by one of two names: ***the sandwich theorem*** or ***the squeeze theorem*** are both standard names for this result, and which one is used is really just personal preference.

The idea behind the sandwich theorem is that if we're given two functions, say $g(x)$ and $h(x)$ that both have the same limit at a given point, that is

$$\lim_{x\to a} g(x) = L = \lim_{x\to a} h(x),$$

then any function that's "sandwiched" between $g(x)$ and $h(x)$ must also have the same limit.

Before giving the theorem we give a preliminary result. These preliminary results are usually called ***lemmas*** in mathematics.

Lemma 1.16.
If $g(x) \le h(x)$ for all x "near" a, then

$$\lim_{x\to a} g(x) \le \lim_{x\to a} h(x)$$

assuming these limits exist.

Theorem 1.17 (The Sandwich Theorem)**.**
Suppose that $f(x)$, $g(x)$, and $h(x)$ are three function satisfying

$$g(x) \le f(x) \le h(x)$$

for all values of x "near" a and that

$$\lim_{x \to a} g(x) = L = \lim_{x \to a} h(x).$$

Then

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

as well.

Example 1.33.
Calculate

$$\lim_{x \to 2} \left(2x^3 - 16\right) \sin\left(\frac{x+2}{x-2}\right)$$

First notice that

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

Multiplying through by $2x^3 - 16$ gives

$$-\left(2x^3 - 16\right) \leq \left(2x^3 - 16\right) \sin\left(\frac{x+2}{x-2}\right) \leq \left(2x^3 - 16\right)$$

Now taking the limits of each function gives us

$$\lim_{x \to 2} -\left(2x^3 - 16\right) \leq \lim_{x \to 2} \left(2x^3 - 16\right) \sin\left(\frac{x+2}{x-2}\right) \leq \lim_{x \to 2} \left(2x^3 - 16\right)$$
$$\implies 0 \leq \lim_{x \to 2} \left(2x^3 - 16\right) \sin\left(\frac{x+2}{x-2}\right) \leq 0$$
$$\implies \lim_{x \to 2} \left(2x^3 - 16\right) \sin\left(\frac{x+2}{x-2}\right) = 0.$$

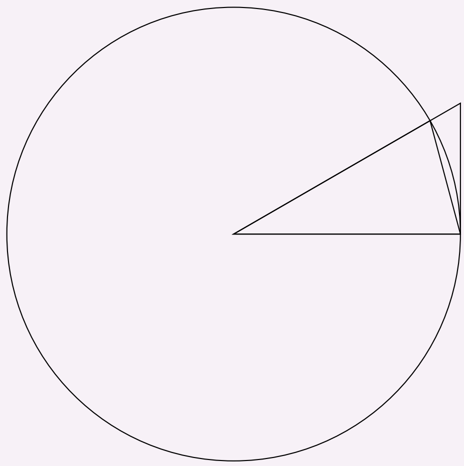
Let's now use the sandwich theorem to prove one useful fact:

Theorem 1.18.

$$\lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} = 1$$

Proof.
Proving this theorem will require that we use the sandwich theorem, but to get our inequality we'll have to review a little bit of trigonometry.

Given an angle θ less than $\pi/2$, consider the triangles shown in the image below.



The inner triangle has area $\frac{1}{2}\sin(\theta)$; the sector of the unit circle has area $\frac{1}{2}\theta$; and the outter triangle has area $\frac{1}{2}\tan(\theta)$. This gives us the inequalities:

$$\frac{1}{2}\sin(\theta) \leq \frac{1}{2}\theta \leq \frac{1}{2}\tan(\theta)$$
$$\implies \sin(\theta) \leq \theta \leq \tan(\theta)$$
$$\implies 1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{1}{\cos(\theta)}$$

Now recall that if $a \leq b$, then $1/b \leq 1/a$. Thus the above string of inequalities may be written as

$$\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1.$$

Now we simply take the limit as θ goes to zero:

$$\begin{aligned} &\cos\theta \leq \frac{\sin\theta}{\theta} \leq 1 \\ \implies &\lim_{\theta\to 0}\cos(\theta) \leq \lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} \leq \lim_{\theta\to 0} 1 \\ \implies &1 \leq \lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} \leq 1 \\ \implies &\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1. \end{aligned}$$

□

This particular limit is actually very useful for evaluating other types of limits involving trig functions because of the following ***change of variables principle***:

Theorem 1.19 (Change of variables for limits)**.**
Suppose $g(x)$ is a function such that $\lim_{x\to a} g(x) = b$. Then by introducing a new variable u which we define to be equal to $g(x)$, we have that for any function $f(x)$ we may compute the limit

$$\lim_{x\to a} f(g(x)) = \lim_{u\to b} f(u).$$

Notice the statement of Theorem 1.19 is very similar to the one which appears in Theorem 1.13; in fact, you could think of Theorem 1.13 as a consequence of Theorem 1.19.

Sketch of proof of Theorem 1.19.
As x gets "really, really" close to a, then u gets "really, really" close to b (as $u = g(x)$ and we are assuming $\lim_{x\to a} g(x) = b$.) We are then

plugging those values "really, really close" to b into the function $f(x)$ and seeing what values of that function are getting near, but this is exactly our (hand-wavy) definition of the limit of f as the input approaches b. □

It should be noticed that the change of variables (i.e., introducing a new variable u) above is just a convenience. What we actually call a variable (x, y, z, u, ☺, whatever...) doesn't really matter. The conclusion of the above theorem could just as easily be written as

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

However, this might be a little bit confusing and seem miraculous, so changing variables can help you understand what's happening.

To see why the change of variables is convenient, consider the following examples:

Example 1.34.

Compute $\lim_{x\to 0} \dfrac{\sin(3x)}{3x}$.

If we simply had x's above instead of $3x$, we could immediately apply Theorem 1.18. Using the change of variables principle, however, basically let's us do this. Letting $u = 3x$, notice that $\lim_{x\to 0} 3x = 0$ and so by the change of variables we have

$$\lim_{x\to 0} \frac{\sin(3x)}{3x} = \lim_{u\to 0} \frac{\sin(u)}{u} = 1.$$

Generalizing the previous example basically proves the following corollary:

Corollary 1.20.

If k is any non-zero constant, then

$$\lim_{x \to 0} \frac{\sin(kx)}{kx} = 1.$$

As another, similar example of the change of variables principle, consider

Example 1.35.

Compute $\lim\limits_{x \to 3} \dfrac{\sin(2x-6)}{2x-6}$.

Introducing the variable $u = 2x - 6$ and noting $\lim_{x \to 3}(2x-6) = 0$ we have

$$\lim_{x \to 3} \frac{\sin(2x-6)}{2x-6} = \lim_{u \to 0} \frac{\sin(u)}{u} = 1.$$

In the last two examples the expression in the denominator and inside the sine in the numerator were equal, but this need not be the case:

Example 1.36.

Compute $\lim\limits_{x \to 0} \dfrac{\sin(6x)}{2x}$.

Our trick here will be to manipulate this limit to turn it into one where the denominator matches up with the input to sine and then apply our change of variables principle again. To accomplish this, we can multiply the denominator by 3 to turn it into $6x$. If we just multiply the denominator by 3, however, we're changing the function and we don't want to do that, so let's compensate by also multiplying the numerator by 3 as well – effectively multiplying everything by 1:

$$\begin{aligned}\lim_{x\to 0}\frac{\sin(6x)}{2x} &= \lim_{x\to 0} 1\cdot\frac{\sin(6x)}{2x}\\ &= \lim_{x\to 0}\frac{3}{3}\cdot\frac{\sin(6x)}{2x}\\ &= 3\lim_{x\to 0}\frac{\sin(6x)}{6x}\\ &= 3\cdot 1\\ &= 3.\end{aligned}$$

Note that since other trig functions are ratios of sines and cosines, we can sometimes use the tricks above to evaluate other limits involving trig functions where we can't simply plug in.

Example 1.37.

Compute $\lim\limits_{x\to 0}\dfrac{\tan(3x)}{5x}$.

First we note that since $\tan(x) = \frac{\sin(x)}{\cos(x)}$ we may write this limit as

$$\begin{aligned}\lim_{x\to 0}\frac{\sin(3x)}{5x\cos(3x)} &= \lim_{x\to 0}\frac{\sin(3x)}{5x}\cdot\frac{1}{\cos(3x)}\\ &= \lim_{x\to 0}\frac{\sin(3x)}{5x}\cdot\lim_{x\to 0}\frac{1}{\cos(3x)}\\ &= \lim_{x\to 0}\frac{\sin(3x)}{5x}\cdot 1\end{aligned}$$

where in the last step we used the fact $\frac{1}{\cos(3x)}$ is continuous at $x = 0$ since $\cos(0) = 1$.

Now to compute the rest of the limit we must turn the $5x$ in the denominator into a $3x$, which we can do by multiplying the denominator by $\frac{3}{5}$. Of course, we must compensate by multiplying the numerator by $\frac{3}{5}$ as well. We then have

$$\begin{aligned}\lim_{x\to 0}\frac{\sin(3x)}{5x\cos(3x)} &= \lim_{x\to 0}\frac{\sin(3x)}{5x}\\ &= \lim_{x\to 0}\frac{3/5}{3/5}\frac{\sin(3x)}{5x}\\ &= \frac{3}{5}\lim_{x\to 0}\frac{\sin(3x)}{3x}\\ &= \frac{3}{5}\end{aligned}$$

Another useful limit is the following:

$$\lim_{\theta\to 0}\frac{1-\cos(\theta)}{\theta} = 0.$$

This limit can be proven using some basic identities from trigonometry and the limit of $\frac{\sin(\theta)}{\theta}$ above, and so we'll leave the proof as an exercise. Though we haven't proven this limit yet, it is a tool we can go ahead and use, as in the following example.

Example 1.38.

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

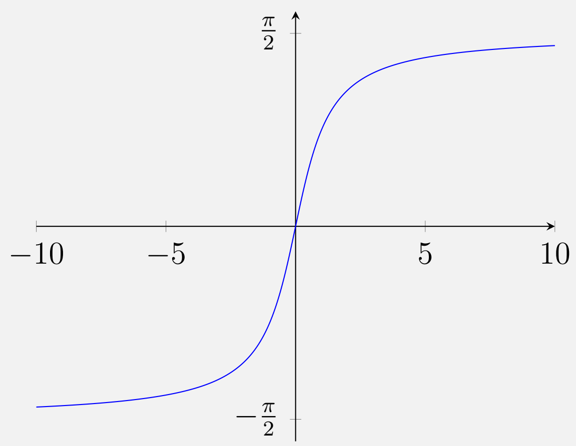
1.5 Limits at Infinity and Horizontal Asymptotes

> *Wahrlich es ist nicht das Wissen, sondern das Lernen, nicht das Besitzen sondern das Erwerben, nicht das Da-Seyn, sondern das Hinkommen, was den grössten Genuss gewährt.*
> It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.
>
> CARL FRIEDRICH GAUSS

We say that the line $y = L$ is a ***horizontal asymptote*** of $y = f(x)$ if the graph gets arbitrarily close to L as x gets arbitrarily large or arbitrarily negative. That is, as x gets "really big" (or "really negative"), we have that $f(x)$ gets "really close" to L. I.e., we're talking some sort of limit.

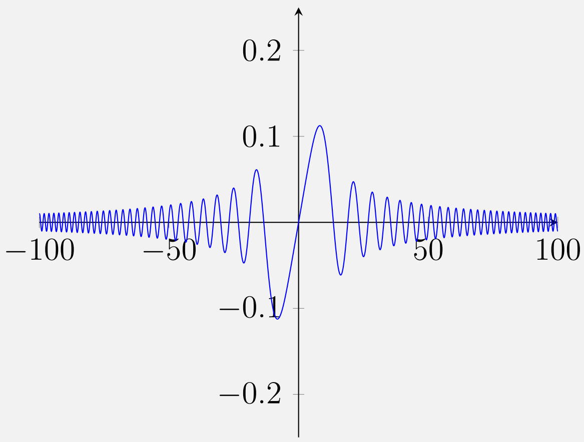
Intuitively, $\lim_{x\to\infty} f(x) = L$ if $f(x)$ is "very close" to L when x is "very large."

Example 1.39.
Consider the function $f(x) = \tan^{-1}(x)$:



$$\lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}$$
$$\lim_{x\to-\infty} \arctan(x) = -\frac{\pi}{2}$$

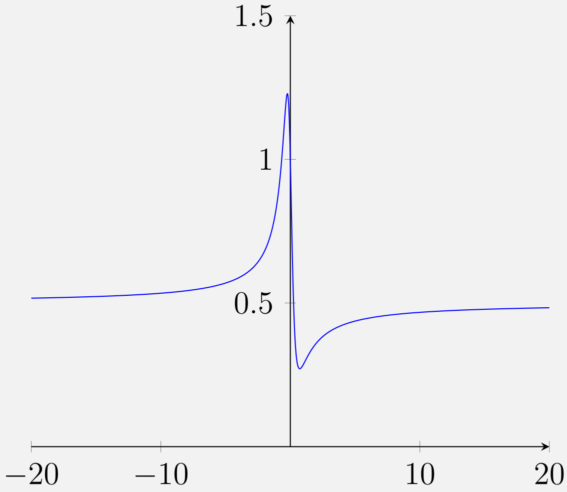
Example 1.40.
Consider the function $\frac{\sin(x^2)}{x}$



It may be hard to see from this image, but notice that the y-values on the graph bounce up and down less and less (less range in the y values) as x grows to $\pm\infty$.

$$\lim_{x\to\infty} \frac{\sin(x^2)}{x} = 0$$
$$\lim_{x\to-\infty} \frac{\sin(x^2)}{x} = 0$$

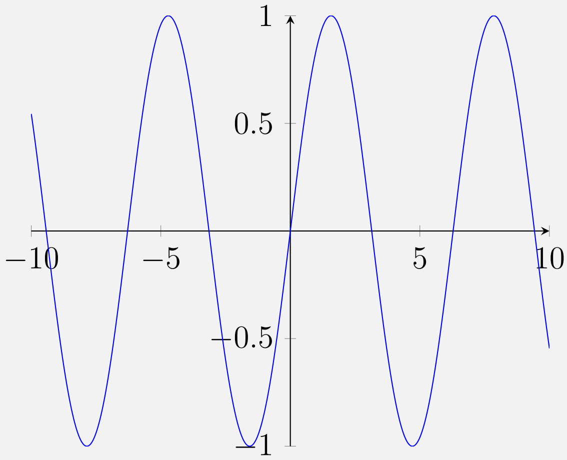
Example 1.41.
Consider the function $\frac{3x^2-2x+1}{6x^2+1}$



$$\lim_{x\to\pm\infty} \frac{3x^2 - 2x + 1}{6x^2 + 1} = \frac{1}{2}$$

It could happen, of course, that the limit as $x \to \pm\infty$ might not exist: the function may not get close to any particular number.

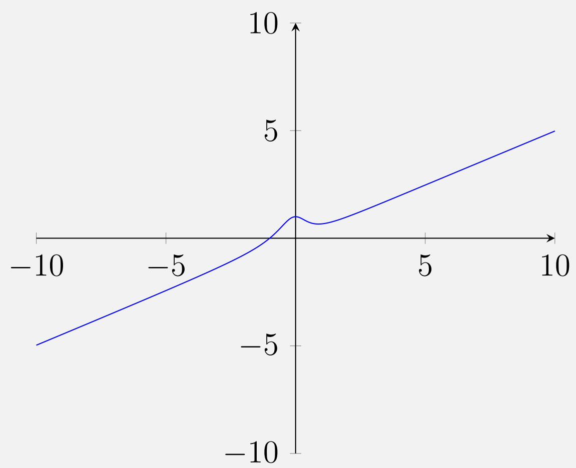
Example 1.42.
Consider $\sin(x)$:



$$\lim_{x \to \pm\infty} \sin(x) \text{ DNE}.$$

Or that the function grows without bound in either the positive or negative direction.

Example 1.43.
Consider $\frac{x^3+1}{2x^2+1}$



$$\lim_{x \to \infty} \frac{x^3+1}{2x^2+1} = \infty$$
$$\lim_{x \to -\infty} \frac{x^3+1}{2x^2+1} = -\infty$$

Theorem 1.21.
If $\lim_{x \to 0^+} f(x) = a$*, then* $\lim_{x \to \infty} f\left(\frac{1}{x}\right) = a$*.*

Proof.
Intuitively, as $x \to \infty$, the fraction $\frac{1}{x}$ approaches zero from the right, so the input to our function approaches zero, and we know the outputs of the function approach a as this happens.

To be precise, we know that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - a| < \varepsilon$ whenever $0 < x < \delta$. We need to show that $|f(1/x) - a| < \varepsilon$ whenever $x > N$ for some N.

If we take $N = \frac{1}{\delta}$, then $x > N$ implies $x > \frac{1}{\delta}$ and so $\delta > \frac{1}{x}$. Thus $|f(1/x) - a| < \varepsilon$. □

Example 1.44.
$\lim_{x\to\infty} \cos(1/x) = 1$.

Theorem 1.22.
If $n > 0$, *then* $\lim_{x\to\infty} \frac{1}{x^n} = 0$.

Proof.
Let $f(x) = x^n$. Then we know $\lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} x^n = 0$. Notice that $\frac{1}{x^n} = \left(\frac{1}{x}\right)^n = f(\frac{1}{x})$, and so by our earlier theorem, $\lim_{x\to\infty} \frac{1}{x^n} = 0$. □

We can combine this theorem with our limit laws to make calculating some infinite limits (particularly for rational functions) extremely easy.

Example 1.45.

$$\begin{aligned}\lim_{x\to\infty}\frac{3x^2-2x+1}{6x^2+1} &= \lim_{x\to\infty}\frac{3x^2-2x+1}{6x^2+1}\cdot\frac{1/x^2}{1/x^2}\\ &= \lim_{x\to\infty}\frac{3x^2/x^2-2x/x^2+1/x^2}{6x^2/x^2+1/x^2}\\ &= \lim_{x\to\infty}\frac{3-2/x+1/x^2}{6+1/x^2}\\ &= \frac{\lim_{x\to\infty}3-\lim_{x\to\infty}\frac{2}{x}+\lim_{x\to\infty}\frac{1}{x^2}}{\lim_{x\to\infty}6+\lim_{x\to\infty}1/x^2}\\ &= \frac{3-0+0}{6+0}\\ &= \frac{1}{2}.\end{aligned}$$

At this point it will be helpful if we recall some properties of arithmetic with infinity.

Arithmetic with Infinity

Infinity is not a number, it's more of a concept. Saying something "goes to infinity" really means it gets larger and larger and larger without bound. Just as we have limit laws for regular, finite limits, we also have limit laws for infinite limits.

Theorem 1.23.
In the following, let $f(x)$ be a function where $\lim_{x\to\infty} f(x) = \infty$ and $g(x)$ a function where $\lim_{x\to\infty} g(x) = C$ where C is a real number (i.e., a finite number).

1. $\lim_{x\to\infty}(f(x)\pm g(x)) = \infty$ – *i.e.,* $\infty \pm C = \infty$
2. $\lim_{x\to\infty}(g(x)+f(x)) = \infty$ – *i.e.,* $C+\infty = \infty$.
3. $\lim_{x\to\infty}(g(x)-f(x)) = -\infty$ – *i.e.,* $C-\infty = \infty$.
4. $\lim_{x\to\infty}\frac{g(x)}{f(x)} = 0$ – *i.e.,* $\frac{C}{\infty} = 0$.
5. *If* $C > 0$*, then* $\lim_{x\to\infty} g(x)f(x) = \infty$ – *i.e.,* $C\cdot\infty = \infty$ *if* $C > 0$.

6. *If* $C < 0$, *then* $\lim_{x\to\infty} g(x)f(x) = -\infty$ – *i.e.,* $C \cdot \infty = \infty$ *if* $C < 0$.

7. *If* $C > 0$, *then* $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$ – *i.e.,* $\frac{\infty}{C} = \infty$ *if* $C > 0$.

8. *If* $C < 0$, *then* $\lim_{x\to\infty} \frac{f(x)}{g(x)} = -\infty$ – *i.e.,* $\frac{\infty}{C} = -\infty$ *if* $C < 0$.

Now suppose that $\lim_{x\to\infty} g(x) = \infty$. *Then*

9. $\lim_{x\to\infty} f(x)g(x) = \infty$ – *i.e.,* $\infty \cdot \infty = \infty$.

10. $\lim_{x\to\infty} (f(x) + g(x)) = \infty$ – *i.e.,* $\infty + \infty = \infty$.

We can use these rules to help us evaluate some other infinite limits.

Example 1.46.

$$\begin{aligned}
\lim_{x\to\infty} \frac{6x-1}{x^2+1} &= \lim_{x\to\infty} \frac{6x-1}{x^2+1} \cdot \frac{1/x^2}{1/x^2} \\
&= \lim_{x\to\infty} \frac{6/x - \frac{1}{x^2}}{1 + \frac{1}{x^2}} \\
&= \frac{\lim_{x\to\infty} 6/x - \lim_{x\to\infty} 1/x^2}{\lim_{x\to\infty} 1 + \lim_{x\to\infty} \frac{1}{x^2}} \\
&= \frac{0-0}{1+0} \\
&= 0.
\end{aligned}$$

Notice there are a few items which are conspicuously absent from our list in the theorem above: we *do not* have rules telling us what to in a situation like $0 \cdot \infty$, $\frac{\infty}{\infty}$, or $\infty - \infty$. These situations are examples of ***indeterminate forms*** – things we can't assign a value to.

Example 1.47.

$$\lim_{x\to\infty} \frac{2}{x} \cdot \frac{x^2+1}{x-3}$$

Even though this naively looks like 0 times ∞, we have to actually do some algebra before we can turn this into something we can evaluate:

$$\begin{aligned}
&\lim_{x\to\infty} \frac{2}{x} \cdot \frac{x^2+1}{x-3} \\
=&\lim_{x\to\infty} \frac{2}{x} \cdot \frac{2x^2+2}{x^2-3x} \\
=&\lim_{x\to\infty} \frac{2x^2+2}{x^2-3x} \cdot \frac{1/x^2}{1/x^2} \\
=&\lim_{x\to\infty} \frac{2+2/x^2}{1-3/x} \\
=&2
\end{aligned}$$

Example 1.48.

$$\begin{aligned}
&\lim_{x\to\infty} \left(\sqrt{x^2-2} - \sqrt{x^2+x}\right) \\
=&\lim_{x\to\infty} \frac{\left(\sqrt{x^2-2} - \sqrt{x^2+x}\right)\left(\sqrt{x^2-2} + \sqrt{x^2+x}\right)}{\left(\sqrt{x^2-2} + \sqrt{x^2+x}\right)} \\
=&\lim_{x\to\infty} \frac{-2-x}{\sqrt{x^2-2} + \sqrt{x^2+x}} \\
=&\lim_{x\to\infty} \frac{-2-x}{\sqrt{x^2-2} + \sqrt{x^2+x}} \cdot \frac{1/x}{1/x} \\
=&\lim_{x\to\infty} \frac{-2-x}{\sqrt{x^2-2} + \sqrt{x^2+x}} \cdot \frac{1/x}{1/\sqrt{x^2}} \\
=&\lim_{x\to\infty} \frac{-2/x-1}{\sqrt{1-2/x} + \sqrt{1+1/x}} \\
=&-\frac{1}{2}
\end{aligned}$$

1.6 The ε-δ definition

> *The calculus was the first achievement of modern mathematics and it is difficult to overestimate its importance. I think it defines more unequivocally than anything else the inception of modern mathematics; and the system of mathematical analysis, which is its logical development, still constitutes the greatest technical advance in exact thinking.*
>
> JOHN VON NEUMANN

In this lecture we will make the idea of a limit precise: whereas we had originally defined $\lim_{x\to a} f(x) = L$ to mean that $f(x)$ gets "arbitrarily close to" L when x gets "arbitrarily close to" a, we never said what "arbitrarily close" meant, and for this reason we haven't really defined limits yet.

We will begin by quickly recalling some preliminaries that will be necessary for this lecture. After this we will give the precise definition of a limit and then consider several examples.

Preliminaries

Absolute Values

Recall that the ***absolute value*** of a number a, denoted $|a|$, is the number's distance from zero. If $a > 0$, then $|a|$ is simply a. If $a < 0$, however, then $|a|$ is $-a$. For example, $|-5| = -(-5) = 5$.

One important property of absolute values is that we can split up products inside of absolute values.

$$|ab| = |a|\,|b|$$

Notice that this property tells us that a number and its negative thus have the same absolute value:

$$|-a| = |(-1)\,a| = |-1|\,|a| = |a|.$$

Furthermore, we can factor expressions inside absolute values, but each factor remains in absolute values:

$$|x^2 - 25| = |(x+5)(x-5)| = |x+5|\,|x-5|.$$

Distances Between Numbers

Given any two numbers a, b on the real line, the ***distance*** between the numbers is defined as the absolute value of their difference:

$$|a-b|.$$

So 5 and 7 are two units apart:

$$|5-7|=|7-5|=2,$$

and 12 and -3 are 15 units apart:

$$|12-(-3)|=|(-3)-12|=15.$$

In order to define the limit of a function, we need to understand how the distance between two numbers is given like this.

Intervals

The set of all x satisfying the equation $|x-c|<r$ is an interval around c: $(c-r, c+r)$. We can see this by recalling the following basic fact:

$$|x|<r \implies -r<x<r.$$

Thus

$$\begin{aligned} &|x-c|<r \\ \implies &-r<x-c<r \\ \implies &c-r<x<c+r \end{aligned}$$

The Idea

To make precise sense of $\lim_{x\to a} f(x)=L$ we want $f(x)$ and L to be close when x and a are close. That is, we want the distance between $f(x)$ and L to be small whenever the distance between x and a is small:

$$\lim_{x\to a} f(x)=L \text{ means } |x-a| \text{ small } \implies |f(x)-L| \text{ small.}$$

This is a little bit closer to the true definition of the limit, but is still imprecise because the question is "how small is small enough?"

The short answer is that no one value is small enough: we need to make sure that $|f(x)-L|$ can be made as small as we'd like, provided we make $|x-a|$ small.

The ε-δ definition

Here is the precise definition of the limit: We say that $\lim_{x\to a} f(x) = L$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - L| < \varepsilon$ whenever $0 < |x - a| < \delta$.

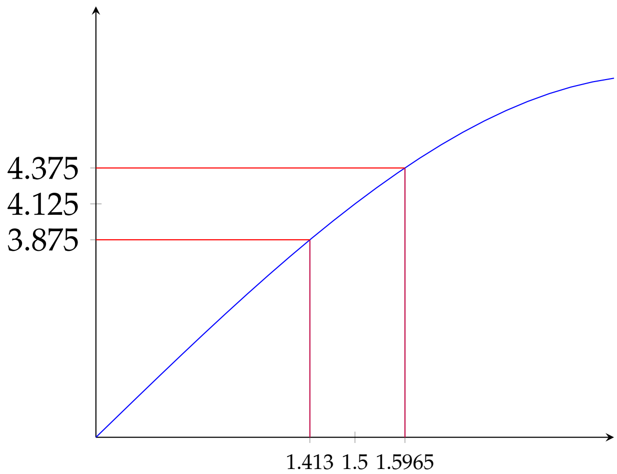
What this definition is saying is that we can make $f(x)$ as close to L as we'd like, provided we pick x to be close enough to a.

The order in which things are stated in the definition above is important. Notice that we say *for every* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that ...* That is, the ε is chosen first, and then the δ is determined. Generally this δ will depend on ε: how close you want $f(x)$ to be to L will influence how closely you need to choose x to be to a. However, the ε can *never* depend on δ!

The ε-δ definition graphically

If $\lim_{x\to a} f(x) = L$, then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$. That is, you want the outputs (the y-values on the graph $y = f(x)$) to be between $y = L - \varepsilon$ and $y = L + \varepsilon$, and I'm saying that is guaranteed to happen provided I pick x's close enough to a, namely $a - \delta < x < a + \delta$. For the limit to be L this has to work for *all* choices of ε: for any ε you choose, I can find a corresponding δ.

How can I go about finding the δ? If we think of drawing two horizontal lines, $y = L - \varepsilon$ and $y = L + \varepsilon$, then I want to see where the lines intersect the graph $y = f(x)$ and then follow those lines "down" to the x-axis. In the image below, we have the graph $y = -x^3 + 3x^2 + x/2$. If I claim the limit as x goes to 1.5 is $\frac{33}{8}$, then when I pick x values "close enough" to 1.5, I should get y-values that are within, say, $1/4$ of $33/8$. To figure out the appropriate y-values, I'll draw the lines $y = 33/8 \pm 1/4$ on the graph, see where those intersect the graph, and then follow those back down to the x-axis.



In this case this gives me x-values of about 1.413 and 1.5965. If I pick x's inbetween these values, then my outputs are with distance $1/4$ of my supposed limit, $33/8$.

Because the set of points satisfying $0 < |x - a| < \delta$ is "symmetric" (δ units to the left, and δ units to the right), we need to be slightly careful when we pick our δ. There are two choices for δ: the distance to the left, $1.5 - 1.413 = 0.087$, and the distance to the right, $1.5965 - 1.5 = 0.0965$. To be sure our outputs, the y-values, stay between $33/8 \pm 1/4$, we need to choose the smaller value and take $\delta = 0.0875$.

A non-example

Instead of jumping into examples of the above definition, it may be helpful to first start with something that doesn't satisfy the definition.

So let's suppose $f(x) = x$, and suppose it was claimed that $\lim_{x\to 3} x = 5$: where does the above definition of the limit cease to hold? Remember that you get to pick an ε first – whatever positive ε you want – and then I get to pick a δ. If the limit really was 5, then it would have to be the case that regardless of what ε you picked, we could guarantee $|x - 5| < \varepsilon$ provided $|x - 3| < \delta$. If you can choose an ε where this won't happen – where there's no choice of δ that I could make so that $|x - 3| < \delta \implies |x - 5| < \varepsilon$, then the limit can't be 5.

So why don't you choose $\varepsilon = 1$. If $\varepsilon = 1$, then $|x - 5| < \varepsilon$ can be rewritten as

$$\begin{aligned} &|x - 5| < 1 \\ \implies &-1 < x - 5 < 1 \\ \implies &4 < x < 6. \end{aligned}$$

So, I need to try to pick a δ so that if $|x-3|<\delta$, then it must be that $|x-5|<1$. Or, put another way, I need to pick a δ so that $|x-3|<\delta$ implies $-4<x<6$.

But if $|x-3|<\delta$, then

$$\begin{aligned} &|x-3|<\delta \\ \implies &-\delta<x-3<\delta \\ \implies &3-\delta<x<3+\delta, \end{aligned}$$

and now we have a problem.

We want that x in the interval $(3-\delta, 3+\delta$ implies that x is also in the interval $(4,6)$! But this is impossible! For any choice of $\delta>0$, there will be points in $(3-\delta, 3+\delta)$ which are not in $(4,6)$. Said another way: there is no choice of $\delta>0$ so that $|x-3|<\delta$ implies $|x-5|<1$. And so the limit can not be 5.

Linear Examples

Example 1.49.
Let's begin with the the "correct" limit for the function in the last example. That is, let's use the precise definition of the limit to show that $\lim_{x\to 3} x=3$.

We need to show that for any ε you choose, there is a δ that guarantees that $|f(x)-L|=|x-3|<\varepsilon$ whenever $|x-a|=|x-3|<\delta$. The right choice here is clearly to take $\delta=\varepsilon$.

Theorem 1.24.
For every real number a, $\lim_{x\to a} x=a$.

Proof.
Let $\varepsilon>0$ be given and take $\delta=\varepsilon$. If $|x-a|<\delta$, then $|f(x)-L|=$

$|x - a| < \varepsilon$. □

Intuitively what the above theorem and proof are saying is that for the function $f(x) = x$, if you want to guarantee that $f(x) = x$ is within ε-distance of the supposed limit $L = a$, then take x to be within $\delta = \varepsilon$ distance of a.

Theorem 1.25.
For every real number a and real number c, $\lim_{x\to a} c = c$.

Proof.
Let $\varepsilon > 0$ be given and take $\delta = 1$. (There's nothing special about 1 here – pick δ to be any positive number you like.) Then for all x within δ distance of a, clearly $|f(x) - c| = |c - c| = 0 < \varepsilon$. □

Example 1.50.
Use the precise definition of the limit to show that $\lim_{x\to 3} 4x = 12$.

Supposing that some $\varepsilon > 0$ has been given, we need to find a $\delta > 0$ which guarantees that if $|x - 3| < \delta$, then $|4x - 12| < \varepsilon$. To do this we'll work backwards: we'll start with what we want to have happen ($|4x - 12| < \varepsilon$), and try to manipulate this a little bit at a time until we get down to something like $|x - 3| <$ something, and then take that "something" to be δ.

This is basically just scratchwork to find δ:

$$\begin{aligned} &|4x - 12| < \varepsilon \\ \implies &4|x - 3| < \varepsilon \\ \implies &|x - 3| < \varepsilon/4. \end{aligned}$$

What our scratch work tells us is that we want to take $\delta = \varepsilon/4$. The "proper" presentation of our proof that $\lim_{x\to 3} 4x = 12$ is the following:

Let $\varepsilon > 0$ be given and take $\delta = \varepsilon/4$. Then for every $|x-3| < \delta$ we have the following:

$$\begin{aligned} & |x-3| < \delta \\ \Longrightarrow\ & |x-3| < \varepsilon/4 \\ \Longrightarrow\ & 4|x-3| < \varepsilon \\ \Longrightarrow\ & |4x-12| < \varepsilon. \end{aligned}$$

Hence $\lim_{x\to 3} 4x = 12$.

Unravelling the last example a little bit, what we're saying is that if you want $f(x) = 4x$ to be within ε-distance of 12, then you need to pick your x's to be within $\varepsilon/4$-distance of 3.

Example 1.51.

Use the precise definition of the limit to show that $\lim_{x\to 2}(5x-3) = 7$.

Again, we work backwards starting from what we want to have happen:

$$\begin{aligned} & |(5x-3)-7| < \varepsilon \\ \Longrightarrow\ & |5x-10| < \varepsilon \\ \Longrightarrow\ & 5|x-2| < \varepsilon \\ \Longrightarrow\ & |x-2| < \varepsilon/5 \end{aligned}$$

Thus we should take $\delta = \varepsilon/5$. The above was scratchwork to figure out δ: now we present the proof:

Let $\varepsilon > 0$ be given and take $\delta = \varepsilon/5$. If $|x-2| < \delta$, we have the following:

$$\begin{aligned} & |x-2| < \delta \\ \Longrightarrow\ & |x-2| < \varepsilon/5 \\ \Longrightarrow\ & 5|x-2| < \varepsilon \\ \Longrightarrow\ & |5x-10| < \varepsilon \\ \Longrightarrow\ & |(5x-3)-7| < \varepsilon \end{aligned}$$

Hence $\lim_{x\to 2}(5x-3)=7$.

We do one last example of linear functions before moving on to quadratics:

Example 1.52.
Use the precise definition of the limit to show that $\lim_{x\to -1}(-2x+3)=5$.

We do some scratchwork to find the right choice of δ:

$$\begin{aligned} &|-2x+3-5|<\varepsilon \\ \Longrightarrow\ &|-2x-2|<\varepsilon \\ \Longrightarrow\ &|(-2)(x+1)|<\varepsilon \\ \Longrightarrow\ &2|x+1|<\varepsilon \\ \Longrightarrow\ &2|x-(-1)|<\varepsilon \\ \Longrightarrow\ &|x-(-1)|<\varepsilon/2. \end{aligned}$$

Now the actual proof:

Let $\varepsilon>0$ be given and take $\delta=\varepsilon/2$. For all x within δ-distance of -1 we have the following:

$$\begin{aligned} &|x-(-1)|<\delta \\ \Longrightarrow\ &|x-(-1)|<\varepsilon/2 \\ \Longrightarrow\ &2|x+1|<\varepsilon \\ \Longrightarrow\ &|2x+2|<\varepsilon \\ \Longrightarrow\ &|(-1)(-2x-2)|<\varepsilon \\ \Longrightarrow\ &|-2x-2|<\varepsilon \\ \Longrightarrow\ &|-2x+3-5|<\varepsilon \end{aligned}$$

and so $\lim_{x\to -1}(-2x+3)=5$.

Quadratic Examples

The above examples were relatively easy. We now turn our attention to some harder examples:

Example 1.53.
Use the precise definition of the limit to show that $\lim_{x\to 2} x^2 = 4$.

We'll start off as before, working backwards:

$$\begin{aligned} &|x^2 - 4| < \varepsilon \\ \implies &|(x+2)(x-2)| < \varepsilon \\ \implies &|x+2||x-2| < \varepsilon \end{aligned}$$

Notice that we need to get something like $|x-2| < (...)$. It's tempting to do something like $|x-2| < \frac{\varepsilon}{|x+2|}$, but this presents us with a problem. We haven't made a choice of δ yet, and remember our argument needs to work for all x in the interval $(2-\delta, 2+\delta)$. Since we haven't picked a δ, we don't know how big or small this interval is going to be, so we don't have any bounds on what $|x+2|$ could be.

There's an easy way to fix this, however: let's go ahead and agree that our δ will be less than some fixed number, like 1. We may want δ to smaller than this, but let's agree that whatever δ we wind up using, it'll be smaller than 1. If that's the case, then

$$\begin{aligned} &|x-2| < \delta \\ \implies &|x-2| < 1 \\ \implies &-1 < x-2 < 1 \\ \implies &-1+4 < x-2+4 < 1+4 \\ \implies &3 < x+2 < 5 \\ \implies &3 < |x+2| < 5. \end{aligned}$$

So as long as we agree that $\delta < 1$, then we can so for sure that $|x+2|$ is no bigger than 5. Continuing from before we can now do the following:

$$\begin{aligned} &|x+2||x-2| < \varepsilon \\ \implies &5|x-2| < \varepsilon \\ \implies &|x-2| < \varepsilon/5. \end{aligned}$$

So it looks like we want to take $\delta = \varepsilon/5$ – but remember this little argument was contingent on $\delta < 1$. If you choose $\varepsilon = 10$, then $\delta = \varepsilon/5 = 2$ is not less than 1, and our logic above no longer applies.

To fix this, we'll take δ to be the smaller of $\varepsilon/5$ and 1:

$$\delta = \min\{1, \varepsilon/5\}.$$

Now here's the actual proof:

Let $\varepsilon > 0$ be given, and let $\delta = \min\{1, \varepsilon/5\}$. Notice that if $|x-2| < \delta$, then in particular $|x-2| < 1$ and so $1 < x < 3$. Hence $3 < x+2 < 5$, and $|x+2| < 5$.

$$\begin{aligned} & |x-2| < \delta \\ \implies & |x-2| < \varepsilon/5 \\ \implies & 5|x-2| < \varepsilon \\ \implies & |x+2||x-2| < \varepsilon \\ \implies & |x^2-4| < \varepsilon \end{aligned}$$

And so $\lim_{x\to 2} x^2 = 4$.

Example 1.54.

Use the precise definition of the limit to show that $\lim_{x\to -3}(4x^2+1) = 37$.

Again, we try to work backwards:

$$\begin{aligned} & |4x^2+1-37| < \varepsilon \\ \implies & |4x^2-36| < \varepsilon \\ \implies & 4|x^2-9| < \varepsilon \\ \implies & |x^2-9|\varepsilon/4 \\ \implies & |x+3||x-3| < \varepsilon/4 \\ \implies & |x-(-3)||x-3| < \varepsilon/4 \end{aligned}$$

Just as before we want to place some bounds on what $|x-3|$ could be by putting some restrictions on δ. Let's again choose δ to be less than 1. Then

$$\begin{aligned} & |x-(-3)|<1 \\ \implies & |x+3|<1 \\ \implies & -1<x+3<1 \\ \implies & -1-6<x-3<1-6 \\ \implies & -7<x-3<-5 \\ \implies & 5<|x-3|<7 \end{aligned}$$

Continuing from above,

$$\begin{aligned} \implies & |x-(-3)||x-3|<\varepsilon/4 \\ \implies & 7|x-(-3)|<\varepsilon/4 \\ \implies & |x-(-3)|<\varepsilon/28 \end{aligned}$$

So we want to take $\delta=\varepsilon/28$, but keep in mind we also need to guarantee that $\delta<1$. Hence we'll take δ to be the minimum of these two.

Now for the proof: Let $\varepsilon>0$ be given and take $\delta=\min\{1,\varepsilon/28\}$. Notice that if $|x-(-3)|<\delta$, then $|x+3|<1$, so $-1<x+3<1$, hence $-7<x-3<-5$, and so $|x-3|<7$. Now,

$$\begin{aligned} & |x-(-3)|<\delta \\ \implies & |x+3|<\varepsilon/28 \\ \implies & 28|x+3|<\varepsilon \\ \implies & 4|x-3||x+3|<\varepsilon \\ \implies & 4|x^2-9|<\varepsilon \\ \implies & |4x^2-36|<\varepsilon \\ \implies & |4x^2+1-36|<\varepsilon. \end{aligned}$$

One-Sided Limits

The formal definition of the right- and left-hand limits is very similar to the definition for the two-sided limit. The main distinction is that we'll be assuming that x is always greater than (or less than) the value a, and so we can drop the absolute values.

We write $\lim_{x \to a^+} f(x) = L$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ so that $|f(x) - L| < \varepsilon$ whenever $x - a < \delta$.

We write $\lim_{x \to a^-} f(x) = L$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ so that $|f(x) - L| < \varepsilon$ whenever $a - x < \delta$.

Example 1.55.
Let $f(x)$ be the piecewise function below:

$$f(x) = \begin{cases} 4x + 2 & \text{if } x > 1 \\ -x & \text{if } x < 1 \end{cases}$$

(a) Show that $\lim_{x \to 1^+} f(x) = 6$.

Since we're only concerned about values to the right of x, for all the values of x we'll consider, we use the $4x + 2$ rule for $f(x)$. Hence we want to show that $|4x + 2 - 6| < \varepsilon$ if $x - 1 < \delta$. We again work backwards to get an idea of what δ should be:

$$\begin{aligned} & |4x + 2 - 6| < \varepsilon \\ \implies & |4x - 4| < \varepsilon \\ \implies & 4|x - 1| < \varepsilon \\ \implies & |x - 1| < \varepsilon/4 \end{aligned}$$

Recalling that we're only looking at values of x which are greater than 1, so $x - 1 > 0$ and we can drop the absolute values above to get

$$x - 1 < \varepsilon/4$$

And of course this is the value we want to take for δ:

Let $\varepsilon > 0$ be given and let $\delta = \varepsilon/4$. Then

$$\begin{aligned} & x - 1 < \varepsilon/4 \\ \implies & |x - 1| < \varepsilon/4 \\ \implies & 4|x - 1| < \varepsilon \\ \implies & |4x - 4| < \varepsilon \\ \implies & |4x + 2 - 6| < \varepsilon \end{aligned}$$

and hence $\lim_{x \to 1^+} f(x) = 6$.

(b) Show that $\lim_{x\to 0^-} f(x) = 0$.
Homework.

Infinite Limits

When the limit of a function is infinity, what this really means is that the function grows without bound: it gets bigger than 10, and bigger than 100, bigger than $1,000$, and so on. Formally this means the following:

We say $\lim_{x\to a} f(x) = \infty$ if for every $N > 0$ there exists a $\delta > 0$ such that $f(x) > N$ whenever $0 < |x - a| < \delta$.

The case for a right-hand limit being infinite is the same except that $0 < x - a < \delta$; for a left-hand limit we have $0 < a - x < \delta$.

Example 1.56.
Show that $\lim_{x\to 0^+} \frac{1}{x} = \infty$.

Let $N > 0$ be given and let $\delta = \frac{1}{N}$. Then for all x satisfying $0 < x < \delta$ we have the following:

$$\begin{aligned} & 0 < x < \delta \\ \implies & x < {}^1\!/\!{}_N \\ \implies & \frac{x}{{}^1\!/\!{}_N} < 1 \\ \implies & \frac{1}{{}^1\!/\!{}_N} < \frac{1}{x} \\ \implies & N < \frac{1}{x} \end{aligned}$$

The limit going to negative infinity is similar, except we want $f(x)$ to get arbitrarily negative:

We say $\lim_{x\to a} f(x) = -\infty$ if for every $N < 0$ there exists a $\delta > 0$ such that $f(x) < N$ whenever $0 < |x - a| < \delta$.

The definitions for left and right hand limits are made by modifying the above definition exaclty as before.

Example 1.57.
Show $\lim_{x\to -2} \frac{-3}{(x+2)^2} = -\infty$.

We work backwards to find δ:

$$\begin{aligned} & \frac{-3}{(x+2)^2} < N \\ \implies & \frac{-3}{|x+2|^2} < N \\ \implies & \frac{-3}{N} > |x+2|^2 \\ \implies & |x-(-2)|^2 < \frac{-3}{N} \\ \implies & |x-(-2)| < \sqrt{-3/N}. \end{aligned}$$

Now here's the actual proof:

Let $N < 0$ be given and take $\delta = \sqrt{-3/N}$. For all x satisfying $0 < |x-(-2)| < \delta$ and then we have:

$$\begin{aligned} & |x-(-2)| < \sqrt{-3/N} \\ \implies & |x-(-2)|^2 < \frac{-3}{N} \\ \implies & \frac{-3}{N} > |x+2|^2 \\ \implies & \frac{-3}{|x+2|^2} < N \\ \implies & \frac{-3}{(x+2)^2} < N \end{aligned}$$

2 Differentiation

2.1 Rates of Change

The further a mathematical theory is developed, the more harmoniously and uniformly does its construction proceed, and unsuspected relations are disclosed between hitherto separated branches of the science.

DAVID HILBERT

Many quantities of interest in mathematics and the sciences are defined by how one quantity changes with respect to another: velocity describes a change in distance over a change in time; acceleration describes how a velocity changes over time; force is a change in potential energy over a distance; and so on. *Derivatives* are the mathematical formulation of this idea: a derivative tells us how one quantity changes as a function of some other quantity. We will discuss several physical examples, but there is also a very nice geometric interpretation of derivatives that we will see as well.

Preliminaries

Recall that the ***slope*** of a line is a number which tells us how steep the line is. If the slope of a line is m, then when you change the x-coordinate of a point on the line by going to the right one unit, you change the y-coordinate by m units. Given two points (x_0, y_0) and (x_1, y_1) on a line, we can easily determine the slope of the line:

$$m = \frac{y_1 - y_0}{x_1 - x_0}.$$

The line with slope m through (x_0, y_0) has equation

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

and this is called the ***point-slope form*** of the line. If we simplify this by distributing the m and moving y_0 to the other side to write the line as

$$y = mx + b$$

(so $b = -mx_0 + y_0$), then we have the ***slope-intercept form*** of the line.

Average rates of change

If you drive 120 miles over the course of two hours you might say that your speed as 60 miles per hour, which is given by taking 120 miles divided by 2 hours. If you measure the growth of a tree over the course of a year you may learn the tree grew 30 inches, you might say the tree grew 2.5 inches per month. Both of these quantities are examples of a rate of change: they describe how quickly one quantity changed (distance, height of the tree) as another quantity changed (time in both of these examples). Quantities like this are ubiquitous in mathematics and the sciences, and the next main topic we're about to discuss for the next several weeks is really concerned with studying these rates of changes, so we begin by first ...

First some notation: use the capital Greek letter delta, Δ, to mean the change in a quantity. So Δx represents the change in x; Δt means the change in t; and so on. We'll usually care about changes over an interval, say from $x = a$ to $x = b$. In this situation our Δx really refers to $b - a$. Sometimes we will use Δx to mean how much x changes without specifying explicity initial and final values of x.

If $f(x)$ is a function of x, then we will be interested in how much $f(x)$ changed as x changed, and we denote this by Δf. So if x changes by $\Delta x = b - a$, then $f(x)$ changes by $\Delta f = f(b) - f(a)$. The ratio of the change in f divided by the change in x is called the ***average rate of change*** of f:

$$\frac{\Delta f}{\Delta x} = \frac{f(b) - f(a)}{b - a}.$$

Velocity, for example, is the average rate of change in distance over time.

Example 2.1.
Consider a particle moving back and forth along a straight line. (This particle can represent any physical object you want: a person, a car, a rolling ball, a bacterium, an electron, etc.) Suppose that we begin recording the particle's position starting at some given moment in time, which we'll call $t = 0$. After t seconds, say the object has moved $f(t)$ centimetres from its initial position, with $f(t) > 0$ meaning motion to the right, and $f(t) < 0$ meaning motion to the left. In principle this $f(t)$ could be any weird function, but for simplicity let's suppose $f(t) = t^2 - 10t$. Thus after 1 second the particle is $f(1) = -9$ centimetres from its starting position; after 3.4 seconds

the particle is $f(3.4) = -22.44$ centimetres from its initial position, and so on.

The ***average velocity*** of the particle for the first six seconds of its journey is

$$\frac{\Delta f}{\Delta t} = \frac{f(6) - f(0)\text{ cm}}{6 - 0\text{ s}} = \frac{-24 - 0}{6 - 0} = -4\frac{\text{cm}}{\text{s}}.$$

This average velocity tells us that the end-result of the motion of the particle after 6 seconds is the same as if the particle was moving to the left at a constant speed of 4 centimetres per second.

The average velocity for the first two minutes of the particle's journey is

$$\frac{\Delta f}{\Delta t} = \frac{f(120) - f(0)}{120 - 0} = \frac{13,200}{120} = 110.$$

So the end result is that the particle's final position, after two minutes, is the same as if the particle was moving at a constant 110 cm/s.

The average veloctiy between times $t = 30$ and $t = 60$ is

$$\frac{\Delta f}{\Delta t} = \frac{f(60) - f(30)}{60 - 30} = \frac{2400}{30} = 80$$

The average velocity over the interval $[0, 10]$ is

$$\frac{\Delta f}{\Delta t} = \frac{f(10) - f(0)}{10 - 0} = \frac{0}{10} = 0.$$

(What's the physical meaning of an average velocity being zero? The only way a fraction – like the average velocity – is zero is if its numerator is zero. If $f(b) - f(a) = 0$, then that just means $f(b) = f(a)$. In terms of positions and velocities, this just means that our particle was at the same location at time $t = 0$ and time $t = 10$. The particle was definitely moving during this period, but at time $t = 10$ it was back to where it had started from.)

The average velocity thus tells us some information about how quickly the particle would be moving if it was always moving at a constant speed. Since the particle is not moving at a constant speed, it's clear that average velocities – while very easy to calculate – leave something to be desired: we'd like to have more detailed information about how the particle is moving. Before discussing how to get this more detailed information, let's think about what an average rate of change represents graphically.

Notice that in our example above, the average velocity corresponds to the slope of the line connecting two points on the graph. In the case of the average velocity over the time interval $[30, 60]$, for example, the formula for average velocity

$$\frac{f(60) - f(30)}{60 - 30}$$

is precisely the slope of the line connecting $(60, f(60))$ to $(30, f(30))$.

Thus, graphically, the average rate of change of a function $f(x)$ over an interval $[a, b]$ corresponds to the slope of the ***secant line*** of the curve $y = f(x)$ between $(a, f(a))$ and $(b, f(b))$ (that is, a line connecting two points on the curve).

Instantaneous rates of change

As the example above indicated, the average rate of change of a function gives us very rough information about what the function is doing. We saw that while the function was constantly changing at different rates, the average rate of change only tells us what the function would be doing if it was changing at a constant rate (i.e., if its graph was a line and not a more complicated curve). We can get more precise information how quickly things are changing by passing from the average rate of change to the instantaneous rate of change.

By ***instantaneous rate of change*** we mean how quickly the function is changing at a single point. In terms of positions and velocities, this means how quickly an object is moving at a single moment in time (ignoring any philosophical objections about what "moving at a single moment in time" should mean, a la Zeno's paradoxes).

The difference between average rate of change and instantaneous rate of change is like the difference between saying "I drove 90 miles to Greensboro in just under an hour" and "at 2:03pm I was driving exactly 63 miles per hour." One statement gives you very rough information about what happened over a long period of time, and the other tells you very precise information about what's happening at a given moment.

The question now, though, is how do we actually go about calculating these instantaneous rates of change. Calculating an average rate of change is very easy, so perhaps what we should do is estimate the instantaneous rate of change with average rates of change. That is, to see what's happening at a particular instant, say $x = a$, we should consider average rates of change over an interval $[a, b]$, but then see what happens as b gets closer and closer to a.

That is, to calculate an instantaneous rate of change of $f(x)$ at $x = a$, we should take the limit of the average rates of change over intervals $[a, b]$ as b approaches a:

$$\lim_{b \to a} \frac{f(b) - f(a)}{b - a}.$$

Example 2.2.
What's the instantaneous velocity of a particle whose position at time t is $f(t) = t^2 - 10t$ at time $t = 1$? What about time $t = 5$ and $t = 10$?

In general, the average velocity of our function over an interval $[a, b]$ is

$$\frac{f(b) - f(a)}{b - a} = \frac{b^2 - 10b - a^2 + 10a}{b - a}.$$

The instantaneous velocity when $t = 1$ is thus given by taking the limit of the average velocities over intervals $[1, b]$ as b approaches 1:

$$\begin{aligned} \lim_{b \to 1} \frac{b^2 - 10b - 1 + 10}{b - 1} &= \lim_{b \to 1} \frac{b^2 - 1 + 10 - 10b}{b - 1} \\ &= \lim_{b \to 1} \frac{(b-1)(b+1) + 10(1-b)}{b - 1} \\ &= \lim_{b \to 1} \frac{(b-1)(b+1) - 10(b-1)}{b - 1} \\ &= \lim_{b \to 1} (b + 1 - 10) \\ &= -8 \end{aligned}$$

We could repeat this same set of algebra for $t = 5$ and $t = 10$, but let's notice the procedure would be exactly the same: for a general

$t = a$, the instantaneous velocity at $t = a$ is

$$\begin{aligned}\lim_{b\to a} \frac{b^2 - 10b - a^2 + 10a}{b-a} =& \lim_{b\to 1} \frac{b^2 - a^2 + 10a - 10b}{b-a} \\ =& \lim_{b\to 1} \frac{(b-a)(b+a) + 10(a-b)}{b-a} \\ =& \lim_{b\to 1} \frac{(b-a)(b+a) - 10(b-a)}{b-a} \\ =& \lim_{b\to 1} (b + a - 10) \\ =& 2a - 10\end{aligned}$$

Thus the instantaneous velocity at $t = 5$ is 0, and the average velocity at $t = 10$ is 10.

(What does an instantaneous velocity of zero mean? It means for a brief moment – just an instant – the particle was not moving. Usually, but not always, this means that the particle transitioned from travelling in one direction to travelling in the opposite direction. For example, if you throw a ball straight up into the air there is a point where it transitions from having positive velocity (going up) to having negative velocity (going down): at that instant the velocity must be zero and the ball is actually not moving for just a single moment in time.)

Just as average rates of change have a nice graphical interpretation, so do instantaneous rates of change. The average rate of change represents the slope of a line connecting two points on a graph, and an instantaneous rate of change represents the limit of those slopes as our two points get closer together. When the points collide, we have the slope of the line tangent to the curve at a point.

(It's worth pointing out that many people have a misconception about what a "tangent line" is. You may have heard before that the line tangent to a curve can only touch the curve in one place, and this isn't quite true: it's actually very easy to come up with curves where tangent lines will intersect the curve multiple – even infinitely many! – times. The "right" way to think about a tangent line is that it's the line which "best approximates" the curve near a point. That is, of all the lines which pass through the point $(a, f(a))$ on the curve $y = f(x)$, the tangent line is the one which "hugs" the curve the most. We can make this idea precise, and we actually will do this later, but right now it's just important to realize that a "tangent line" can touch a curve multiple times.)

Example 2.3.
Determine the equation of the line tangent to the graph $y = 4x^2 + 2x - 1$ at the point $(2, 19)$.

By what we said above, the slope of the tangent line is the instantaneous rate of change of $f(x) = 4x^2 + 2x - 1$ at the point $x = 2$, and this is given by taking the limit of average rate of change (slope of the secant lines) over the interval $[2, b]$ as $b \to 2$. Thus the slope is

$$\begin{aligned}\lim_{b\to 2}\frac{f(b)-f(2)}{b-2} &= \lim_{b\to 2}\frac{4b^2+2b-1-4\cdot 2^2-2\cdot 2+1}{b-2}\\ &= \lim_{b\to 2}\frac{4b^2+2b-20}{b-2}\\ &= \lim_{b\to 2}\frac{(b-2)(4b+10)}{b-2}\\ &= \lim_{b\to 2}(4b+10)\\ &= 18\end{aligned}$$

So the slope of the line is $m = 18$. We know the line passes through $(2, 19)$, and so the line in point-slope form is

$$y - 19 = 18(x - 2),$$

or in slope-intercept form

$$y = 18x - 17$$

Derivatives

The instantaneous rates of change described above are called ***derivatives*** in mathematics. That is, ***instantaneous rate of change of $f(x)$ at $x = a$*** is synonymous with ***derivative of $f(x)$ at $x = a$***. We denote this derivative (aka, instantaneous rate of change) as $f'(a)$:

$$f'(a) = \lim_{b\to a}\frac{f(b)-f(a)}{b-a}.$$

Notice that since we're taking the limit as $b \to a$, we care about values of b which are very close to a: that is, b is equal to a plus a little bit more. We can thus write $b = a + h$. If we replace the b in our above equation for $f'(a)$ with $a + h$, what happens?

Well, if b is getting closer to a, that means that the h in $a + h$ is getting closer to 0, and so we can rewrite our limit as the limit as h goes to zero. In the numerator all we can change is $f(b) - f(a) = f(a + h) - f(a)$, but in the denominator we have $b - a = a + h - a = h$, and so we can rewrite the derivative $f'(a)$ as

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

Both definitions of the derivative are completely the same, just represented in slightly different ways. It will sometimes be slightly more convenient to use one definition over the other, however, so it's useful to be aware of both ways of writing $f'(a)$.

Example 2.4.
Calculate the derivative $f'(3)$ where $f(x) = x^2 + 2$.

$$\begin{aligned} f'(3) &= \lim_{h \to 0} \frac{f(3+h) - f(3)}{h} \\ &= \lim_{h \to 0} \frac{(3+h)^2 + 2 - 3^2 - 2}{h} \\ &= \lim_{h \to 0} \frac{9 + 6h + h^2 - 9}{h} \\ &= \lim_{h \to 0} \frac{6h + h^2}{h} \\ &= \lim_{h \to 0} (6 + h) \\ &= 6 \end{aligned}$$

Example 2.5.
Determine the equation of the line tangent to $y = x^2 + 2$ at the point $(3, 11)$.

We saw earlier that the slope of the tangent line is the instantaneous rate of change of the function, but this is exactly the derivative. Thus, by our calculation in the last example, the slope of the tangent line is 6. Hence our line in point-slope form is

$$y - 11 = 6(x - 3),$$

or in slope-intercept form

$$y = 6x - 7.$$

There are lots and lots of common notations for the derivative, one that we'll use in this class is $\left.\frac{df}{dx}\right|_a$. This notation might look odd, but it has the advantage of letting us write the derivative of a function without "naming" the function: for example,

$$\left.\frac{d}{dx}\right|_2 \left(6x^2 + x\right)$$

means the derivative of $6x^2 + x$ at the point $x = 2$.

Example 2.6.
Determine the equation of the line tangent to $y = \sqrt{x}$ when $x = 16$.

First notice that the y-value of the point on the curve $y = \sqrt{x}$ when $x = 16$ is $\sqrt{16} = 4$. So we know the line passes through the point $(16, 4)$. Now we need to find the slope of the line, but this is of course the derivative of $\sqrt{x}$ when $x = 16$:

$$\begin{aligned}
\left.\frac{d}{dx}\right|_{16} \sqrt{x} &= \lim_{h\to 0} \frac{\sqrt{16+h}-\sqrt{16}}{h} \\
&= \lim_{h\to 0} \frac{\sqrt{16+h}-4}{h} \\
&= \lim_{h\to 0} \frac{\sqrt{16+h}-4}{h} \cdot \frac{\sqrt{16+h}+4}{\sqrt{16+h}+4} \\
&= \lim_{h\to 0} \frac{16+h-16}{h\left(\sqrt{16+h}+4\right)} \\
&= \lim_{h\to 0} \frac{1}{\sqrt{16+h}+4} \\
&= \frac{1}{\lim_{h\to 0}\left(\sqrt{16+h}+4\right)} \\
&= \frac{1}{4+4} \\
&= \frac{1}{8}.
\end{aligned}$$

Thus the tangent line is

$$y - 4 = \frac{1}{8}(x-16)$$

2.2 Differentiability

> *La mathématique est l'art de donner le même nom à des choses différentes.*
> Mathematics is the art of giving the same name to different things.
>
> HENRI POINCARÉ
> *L'avenir des mathématiques*

An Alternative Definition

In the last lecture we defined the derivative of a function $f(x)$ at a point $x = a$ to be

$$f'(a) = \lim_{b \to a} \frac{f(b) - f(a)}{b - a}.$$

Notice that we can think of b as the number a "plus a little more." That is, if we define

$$h = b - a$$

then we can write

$$b = a + h.$$

Notice that as $b \to a$, the quantity h approaches 0. Thus we can rewrite the derivative $f'(a)$ as

$$\begin{aligned} f'(a) &= \lim_{b \to a} \frac{f(b) - f(a)}{b - a} \\ &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}. \end{aligned}$$

These two versions of the derivative are completely equivalent: they are exactly the same thing, but represented in a different way. However, sometimes one choice of the definition may be slightly easier to do algebra with than the other version.

Example 2.7.
Suppose the position of a particle at time t is $f(t) = t^3 + t^2 - 1$ feet from a starting point after t minutes. What is the particle's velocity at time $t = 3$?

We need to calculate $f'(3)$:

$$\begin{aligned} f'(3) &= \lim_{h\to 0} \frac{f(3+h)-f(3)}{h} \\ &= \lim_{h\to 0} \frac{(3+h)^3+(3+h)^2-1-35}{h} \\ &= \lim_{h\to 0} \frac{27+27h+9h^2+h^3+9+6h+h^2-36}{h} \\ &= \lim_{h\to 0} \frac{h^3+10h^2+33h}{h} \\ &= \lim_{h\to 0} \left(h^2+10h+33\right) \\ &= 33\frac{\text{feet}}{\text{min}} \end{aligned}$$

Differentiability

Regardless of which definition of the derivative $f'(a)$ we choose to use, the derivative is a limit. As we've seen before, however, limits don't have to exist: everytime we see a limit, we should always ask ourselves if the limit exists or not. If the limit $f'(a)$ exists, then we say that the function $f(x)$ is ***differentiable*** at $x = a$. The function $f(t) = t^3+t^2-1$, for example, is differentiable at $t = 0$.

Example 2.8.
Where is the absolute value function differentiable?

Recall that the absolute value function is defined piecewise:

$$|x| = \begin{cases} x & = \text{ if } x > 0 \\ -x & = \text{ if } x < 0 \\ 0 & = \text{ if } x = 0 \end{cases}$$

Since this function is defined by two rules, let's first consider what happens in the ranges where each of these rules applies.

If $a > 0$, then $|a| = a$. The derivative is then:

$$\begin{aligned}\left.\frac{d}{dx}\right|_{x=a}|x| &= \lim_{h\to 0}\frac{|a+h|-|a|}{h}\\ &= \lim_{h\to 0}\frac{a+h-a}{h}\\ &= \lim_{h\to 0} 1\\ &= 1\end{aligned}$$

If $a < 0$, then $|a| = -a$. The derivative is then:

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

So at this point we see that $|x|$ is differentiable at all points except possibly at $x = 0$. We still need to determine

$$\lim_{h\to 0}\frac{|0+h|-|0|}{h} = \lim_{h\to 0}\frac{|h|}{h}.$$

Let's try to calculate this limit by looking at the left- and right-hand limits:

$$\begin{aligned}\lim_{h\to 0^+}\frac{|h|}{h} &= \lim_{h\to 0^+}\frac{h}{h} = 1\\ \lim_{h\to 0^-}\frac{|h|}{h} &= \lim_{h\to 0^-}\frac{-h}{h} = -1\end{aligned}$$

Since the left- and right-hand limits disagree, the limit does not exist. Hence $|x|$ is *not* differentiable at $x = 0$.

Example 2.9.
Is the function below differentiable at $x = 2$?

$$f(x) = \begin{cases} x^2 & \text{if } x \geq 2 \\ -x & \text{if } x < 1 \end{cases}$$

We need to determine if the limit

$$\lim_{h \to 0} \frac{f(2+h) - f(2)}{h}$$

exists or not. Again, let's consider the left- and right-hand limits:

$$\begin{aligned} \lim_{h \to 0^+} \frac{f(2+h) - f(2)}{h} \\ &= \lim_{h \to 0^+} \frac{(2+h)^2 - 2^2}{h} \\ &= \lim_{h \to 0^+} \frac{4 + 4h + h^2 - 4}{h} \\ &= \lim_{h \to 0^+} (4 + h) \\ &= 4. \end{aligned}$$

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

Again, the one-sided limits disagree, and so the function is not differentiable at $x = 2$.

These two examples illustrate the two most common ways for a function to fail to be differentiable.

Recall that we said last time that the derivative $f'(a)$ represents the slope of the line tangent to $y = f(x)$ at the point $(a, f(a))$, and we said that a tangent line was the line that best approximated the function near the point. Saying a function is differentiable thus means the function can be approximated by a line. Two ways this can screw up is if the graph of the function has a sharp corner (as with $|x|$), or has a break in it (as with the second example).

Theorem 2.1.
If f is differentiable at $x = a$, then f is continuous at $x = a$.

Proof.
We need to show that $\lim_{x\to a} f(x) = f(a)$, or equivalently

$$\lim_{x\to a} (f(x) - f(a)) = 0.$$

It will be convenient to use our earlier definition of the derivative for this:

$$\begin{aligned}
\lim_{x\to a} (f(x) - f(a)) &= \lim_{x\to a} (f(x) - f(a)) \frac{x-a}{x-a} \\
&= \lim_{x\to a} \frac{f(x) - f(a)}{x-a} \cdot (x-a) \\
&= \left(\lim_{x\to a} \frac{f(x) - f(a)}{x-a}\right) \cdot \left(\lim_{x\to a} (x-a)\right) \\
&= f'(a) \cdot 0 \\
&= 0 \\
\implies \lim_{x\to a} f(x) &= f(a)
\end{aligned}$$

and thus the function is continuous. □

Corollary 2.2.
If f is not continuous at $x = a$, then f is not differentiable at $x = a$.

Example 2.10.
The function $f(x) = \frac{x}{x-2}$ is not differentiable at $x = 2$ since it is not continuous there.

The Derivative as a Function

If a function is differentiable at every point where it's defined, we say the function is ***differentiable***. If a function is differentiable, then for each x in the domain of $f(x)$ we can associate the number $f'(x)$. That is, we can we think of the derivative as a function called (surprise, surprise), the ***derivative*** of $f(x)$. This function is denoted in several ways:

$$f'(x), \quad \frac{df}{dx}, \quad \text{or } \frac{d}{dx}f(x)$$

Sometimes when discussing the graph $y = f(x)$ we will also denote the derivative as

$$y' \text{ or } \frac{dy}{dx}$$

Notice this function $f'(x)$ is defined as a limit at each point:

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

Example 2.11.
If $f(x) = 4x^2 + 2x + 1$, what is $f'(x)$?

$$\begin{aligned}
f'(x) =& \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\
=& \lim_{h\to 0} \frac{4(x+h)^2+2(x+h)+1-(4x^2+2x+1)}{h} \\
=& \lim_{h\to 0} \frac{4x^2+8xh+4h^2+2x+2h+1-4x^2-2x-1}{h} \\
=& \lim_{h\to 0} \frac{8xh+4h^2+2h}{h} \\
=& \lim_{h\to 0} (8x+4h+2) \\
=& 8x+2.
\end{aligned}$$

Example 2.12.
Differentiate $\sqrt{x}$:

$$\begin{aligned}
\frac{d}{dx}\sqrt{x} =& \lim_{h\to 0} \frac{\sqrt{x+h}-\sqrt{x}}{h} \\
=& \lim_{h\to 0} \frac{\sqrt{x+h}-\sqrt{x}}{h} \cdot \frac{\sqrt{x+h}+\sqrt{x}}{\sqrt{x+h}+\sqrt{x}} \\
=& \lim_{h\to 0} \frac{(x+h)-x}{h\left(\sqrt{x+h}+\sqrt{x}\right)} \\
=& \lim_{h\to 0} \frac{1}{\sqrt{x+h}+\sqrt{x}} \\
=& \frac{1}{2\sqrt{x}}
\end{aligned}$$

Example 2.13.
Differentiate the constant function $f(x) = c$:

$$
\begin{aligned}
\frac{d}{dx}c =& \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\\
=& \lim_{h\to 0}\frac{c-c}{h}\\
=& 0
\end{aligned}
$$

Higher-Order Derivatives

If $f(x)$ is differentiable, then we can consider the derivative function $f'(x)$. But since this is a function, we can ask about its derivative. This is called the ***second derivative*** of $f(x)$ and is denoted in a few different ways. The most common notations are

$$f''(x), \text{ and } \frac{d^2f}{dx^2}.$$

The derivative we talked about before, $f'(x)$, is sometimes called the ***first derivative***.

Example 2.14.
Let $f(x) = 6x^3 + 3x^2 + 2x + 1$ and calculate $f''(x)$.

The first thing we have to do is find the first derivative, and then we differentiate once more to get the second derivative.

$$
\begin{aligned}
f'(x) =& \lim_{h\to 0}\frac{6(x+h)^3+3(x+h)^2+2(x+h)+1-(6x^3+3x^2+2x+1)}{h}\\
=& \lim_{h\to 0}\frac{6x^3+18x^2h+18xh^2+6h^3+3x^2+6xh+3h^2+2x+2h+1-6x^3-3x^2-2x-}{h}\\
=& \lim_{h\to 0}\frac{18x^2h+18xh^2+6h^3+6xh+2h}{h}\\
=& \lim_{h\to 0}\left(18x^2+18xh+6h^2+6x+2\right)\\
=& 18x^2+6x+2
\end{aligned}
$$

Now we take the derivative of $f'(x)$ to calculate $f''(x)$:

$$\begin{aligned}f''(x) =& \lim_{h\to 0} \frac{18(x+h)^2 + 6(x+h) + 2 - (18x^2 + 6x + 2)}{h} \\ =& \lim_{h\to 0} \frac{18x^2 + 36xh + 18h^2 + 6x + 6h + 2 - 18x^2 - 6x - 2}{h} \\ =& \lim_{h\to 0} \frac{36xh + 18h^2 + 6h}{h} \\ =& \lim_{h\to 0} (36x + 18h + 6) \\ =& 36x + 6.\end{aligned}$$

We take the derivative of the derivative to determine the second derivative of $f(x)$. We could then differentiate $f''(x)$ to calculate the ***third derivative*** of $f(x)$, denoted

$$f'''(x) \text{ or } \frac{d^3 f}{dx^3}.$$

We could differentiate yet againf or the fourth derivative,

$$f''''(x) \text{ or } \frac{d^4 f}{dx^4},$$

and so on.

In general, the ***n-th derivative*** of $f(x)$ is obtained by differentiating $f(x)$ n times. We of course don't write out n primes if n is large. Instead we write

$$f^{(n)}(x) \text{ or } \frac{d^n f}{dx^n}.$$

Any of these derivatives (second, third, fourth, etc...) are called ***higher-order derivatives*** of $f(x)$.

Velocity and Acceleration

We discussed last time that the derivative of position is called ***velocity***. We can also talk about the derivative (instantaneous rate of change of velocity) which is called acceleration.

Example 2.15.
The Empire State Building is 381 metres tall. If a stone is dropped

from the top of the Empire State Building, its height above the ground t seconds after being dropped is

$$f(t) = -4.9t^2 + 381$$

When does the stone hit the ground? How fast is the stone going when it hits the ground? How quickly is the stone accelerating?

The first question is really an algebra question: we want to know when the height of the stone is zero:

$$\begin{aligned} &-4.9t^2 + 381 = 0 \\ \implies &4.9t^2 = 381 \\ \implies &t^2 = \frac{381}{4.9} \\ \implies &t = \sqrt{\frac{381}{4.9}} \approx 8.81\text{sec} \end{aligned}$$

The velocity of the stone at any given moment is $f'(t)$:

$$\begin{aligned} f'(t) &= \lim_{h\to 0} \frac{-4.9(t+h)^2 + 381 - 4.9t^2 - 381}{h} \\ &= \lim_{h\to 0} \frac{-4.9t^2 - 9.8th - 4.9h^2 + 381 - 4.9t^2 - 381}{h} \\ &= \lim_{h\to 0} \frac{-9.8th - 4.9h^2}{h} \\ &= \lim_{h\to 0} (-9.8t - 4.9h) \\ &= -9.8t\frac{\text{m}}{\text{s}} \end{aligned}$$

The velocity when the stone hits the ground is thus

$$f'(8.81) \approx -86.34\frac{\text{m}}{\text{s}}.$$

The stone's acceleration is

$$
\begin{aligned}
f''(t) &= \lim_{h\to 0} \frac{f'(t+h) - f'(t)}{h} \\
&= \lim_{h\to 0} \frac{-9.8(t+h) + 9.8t}{h} \\
&= -9.8 \frac{\text{m}}{\text{s}^2}
\end{aligned}
$$

Example 2.16.
Suppose the position of a particle at time t is $f(t) = t^3 + t^2$. What is the particle's acceleration?

First we find the velocity:

$$
\begin{aligned}
f'(t) &= \lim_{h\to 0} \frac{f(t+h) - f(t)}{h} \\
&= \lim_{h\to 0} \frac{(t+h)^3 + (t+h)^2 - t^3 - t^2}{h} \\
&= \lim_{h\to 0} \frac{t^3 + 3t^2h + 3th^2 + h^3 + t^2 + 2th + h^2 - t^3 - t^2}{h} \\
&= \lim_{h\to 0} \frac{3t^2h + 3th^2 + h^3 + 2th}{h} \\
&= \lim_{h\to 0} \left(3t^2 + 3th + h^2 + 2t\right) \\
&= 3t^2 + 2t.
\end{aligned}
$$

And we differentiate once more to get the acceleration:

$$\begin{aligned}
f''(t) &= \lim_{h \to 0} \frac{f'(t+h) - f'(t)}{h} \\
&= \lim_{h \to 0} \frac{3(t+h)^2 + 2(t+h) - 3t^2 - 2t}{h} \\
&= \lim_{h \to 0} \frac{3t^2 + 6th + 3h^2 + 2t + 2h - 3t^2 - 2t}{h} \\
&= \lim_{h \to 0} \frac{6th + 3h^2 + 2h}{h} \\
&= \lim_{h \to 0} (6t + 3h + 2) \\
&= 6t + 2.
\end{aligned}$$

2.3 Derivative Rules

"Reeling and Writhing, of course, to begin with," the Mock Turtle replied, "and then the different branches of Arithmetic: Ambition, Distraction, Uglification, and Derision."

LEWIS CARROLL
Alice's Adventures in Wonderland

In this lecture we start to develop a set of rules which make calculating derivatives much simpler.

Derivatives of Constants, Constant Multiples, and Sums

We begin by recalling an example from the last lecture.

Theorem 2.3.
The derivative of a constant is zero.

Proof.

$$\begin{aligned}\frac{d}{dx}c &= \lim_{h\to 0}\frac{c-c}{h}\\ &= 0.\end{aligned}$$

□

Example 2.17.

$$\frac{d}{dx}17 = 0$$

Theorem 2.4.
If c is a constant and f is a differentiable function, then

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

Proof.

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

□

Example 2.18.
Suppose that $f(x)$ is a function with the property that $f'(x) = 6x^2 + \sqrt{x-7}$. Then

$$\frac{d}{dx}3f(x) = 18x^2 + 3\sqrt{x-7}.$$

Theorem 2.5.
If $f(x)$ and $g(x)$ are differentiable functions, then so is their sum $f(x) + g(x)$ and

$$\frac{d}{dx}\left(f(x)+g(x)\right) = f'(x) + g'(x).$$

Proof.

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

□

Corollary 2.6.
If $f(x)$ and $g(x)$ are differentiable, then so is their difference $f(x)-g(x)$ and

$$\frac{d}{dx}\left(f(x)-g(x)\right)=f'(x)-g'(x)$$

Proof.

$$\begin{aligned}\frac{d}{dx}\left(f(x)-g(x)\right) &= \frac{d}{dx}\left(f(x)+(-1)\cdot g(x)\right)\\ &= \frac{d}{dx}f(x)+\frac{d}{dx}(-1)g(x)\\ &= \frac{d}{dx}f(x)+(-1)\frac{d}{dx}g(x)\\ &= f'(x)-g'(x).\end{aligned}$$

□

The Product Rule and Derivatives of Polynomials

Theorem 2.7 (The Product Rule).
If $f(x)$ and $g(x)$ are differentiable, then so is their product $f(x)\,g(x)$, and

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

Proof.

$$\begin{aligned}
\frac{d}{dx} f(x)\,g(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{(f(x+h) - f(x))\,g(x+h) + f(x)\,(g(x+h) - g(x))}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \cdot g(x+h) + \lim_{h\to 0} f(x) \cdot \frac{g(x+h) - g(x)}{h} \\
&= f'(x)g(x) + f(x)g'(x).
\end{aligned}$$

□

Theorem 2.8.
If n is a positive integer, then

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

Proof.

Consider the case when $n = 1$:

$$\begin{aligned}\frac{d}{dx}x =& \lim_{h\to 0}\frac{x+h-x}{h}\\ =&1\\ =&1\cdot x^0\end{aligned}$$

Now consider the case when $n = 2$:

$$\begin{aligned}\frac{d}{dx}x^2 =& \frac{d}{dx}x\cdot x\\ =&1\cdot x + x\cdot 1\\ =&2x\end{aligned}$$

When $n = 3$:

$$\begin{aligned}\frac{d}{dx}x^3 =& \frac{d}{dx}x\cdot x^2\\ =&1\cdot x^2 + x\cdot 2x\\ =&x^2+2x^2\\ =&3x^2\end{aligned}$$

More generally, if we've shown the rule applies for the integers 1, 2, 3, ..., $n-1$, then we can show the rule also applies for n:

$$\begin{aligned}\frac{d}{dx}x^n =& \frac{d}{dx}x\cdot x^{n-1}\\ =&1\cdot x^{n-1} + x\cdot (n-1)x^{n-2}\\ =&x^{n-1}+(n-1)x^{n-1}\\ =&nx^{n-1}\end{aligned}$$

□

We are now able to easily calculate the derivative of any polynomial.

Example 2.19.

$$\begin{aligned}\frac{d}{dx}\left(7x^3-2x^2+x-1\right)&\\ &=\frac{d}{dx}7x^3-\frac{d}{dx}2x^2+\frac{d}{dx}x-\frac{d}{dx}1\\ &=7\frac{d}{dx}x^3-2\frac{d}{dx}x^2+\frac{d}{dx}x-\frac{d}{dx}1\\ &=7\cdot 3x^2-2\cdot 2x+1-0\\ &=21x^2-4x+1\end{aligned}$$

Example 2.20.

$$\begin{aligned}\frac{d}{dx}\left(100x^{37}+11x^{33}-\pi x^2+e\right)&=\frac{d}{dx}100x^{37}+\frac{d}{dx}11x^{33}-\frac{d}{dx}\pi x^2+\frac{d}{dx}e\\ &=100\frac{d}{dx}x^{37}+11\frac{d}{dx}x^{33}-\pi\frac{d}{dx}x^2+\frac{d}{dx}e\\ &=3700x^{36}+363x^{32}-2\pi x\end{aligned}$$

This is important enough that we should record it as a theorem:

Theorem 2.9.

$$\begin{aligned}&\frac{d}{dx}\left(c_nx^n+c_{n-1}x^{n-1}+\cdots+c_2x^2+c_1x+c_0\right)\\ =&nc_nx^{n-1}+(n-1)c_{n-1}x^{n-2}+\cdots+2c_2x+c_1\end{aligned}$$

Example 2.21.

$$\frac{d}{dx}\left(17x^4 - 6x^2 + 3x\right) = 68x^3 - 12x + 3$$

$$\frac{d}{dx}\left(3x^{15} + 14x^{10} - 6x^9 + 7x^5 + x^2 - 4x - 2\right)$$
$$= 45x^{14} + 140x^9 - 54x^8 + 35x^4 + 2x - 4$$

Note too that the product rule makes it easy to calculate the derivative of a product of polynomials without first multiplying the polynomials out:

Example 2.22.

$$\frac{d}{dx}\left(3x^5 + 6x^2 + 2\right) \cdot \left(x^7 - x^4 + x^3 - 3x\right)$$
$$= \left[\frac{d}{dx}\left(3x^5 + 6x^2 + 2\right)\right] \cdot \left(x^7 - x^4 + x^3 - 3x\right) + \left(3x^5 + 6x^2 + 2\right) \cdot \left[\frac{d}{dx}\left(x^7 - x^4 + x^3 - 3x\right)\right]$$
$$= \left(15x^4 + 12x\right) \cdot \left(x^7 - x^4 + x^3 - 3x\right) + \left(3x^5 + 6x^2 + 2\right) \cdot \left(7x^6 - 4x^3 + 3x^2 - 3\right)$$

The Quotient Rule

Theorem 2.10 (The Quotient Rule)**.**
Suppose that $f(x)$ and $g(x)$ are differentiable, then so is $\frac{f(x)}{g(x)}$ and

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{g(x)f'(x) - f(x)g'(x)}{\left(g(x)\right)^2}.$$

Proof.

$$
\begin{aligned}
&\frac{d}{dx}\frac{f(x)}{g(x)} \\
&= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} \cdot \frac{g(x)}{g(x)} - \frac{f(x)}{g(x)} \cdot \frac{g(x+h)}{g(x+h)}}{h} \\
&= \lim_{h\to 0} \frac{\frac{f(x+h)g(x)-f(x)g(x+h)}{g(x)g(x+h)}}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x)g(x+h)} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{hg(x)g(x+h)} \\
&= \lim_{h\to 0} \frac{g(x)\left(f(x+h) - f(x)\right) + f(x)\left(g(x) - g(x+h)\right)}{hg(x)g(x+h)} \\
&= \lim_{h\to 0} \frac{g(x)\left(f(x+h) - f(x)\right) - f(x)\left(g(x+h) - g(x)\right)}{hg(x)g(x+h)} \\
&= \frac{\lim_{h\to 0}\left[g(x) \cdot \frac{f(x+h)-f(x)}{h} - f(x) \cdot \frac{g(x+h)-g(x)}{h}\right]}{\lim_{h\to 0} g(x)g(x+h)} \\
&= \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}
\end{aligned}
$$

□

There is a nice mnemonic for remember the quotient rule: “lo-de-hi minus hi-de-lo, all over low squared.”

Example 2.23.

$$\begin{aligned}\frac{d}{dx}\frac{x^2+2x-1}{3x+1} =& \frac{(3x+1)\frac{d}{dx}\left(x^2+2x-1\right)-\left(x^2+2x-1\right)\frac{d}{dx}(3x+1)}{(3x+1)^2}\\ =& \frac{(3x+1)\cdot(2x+2)-\left(x^2+2x-1\right)\cdot 3}{9x^2+6x+1}\\ =& \frac{6x^2+6x+2x+2-3x^2-6x+6}{9x^2+6x+1}\\ =& \frac{3x^2+2x+8}{9x^2+6x+1}\end{aligned}$$

We won't be able to prove this rule until we talk about the chain rule later, but it's good to go ahead and make note of it so that we can use it:

Theorem 2.11 (Power Rule)**.**
For any real number a,

$$\frac{d}{dx}x^a = ax^{a-1}.$$

Example 2.24.

$$\begin{aligned}\frac{d}{dx}\sqrt{x} =& \frac{d}{dx}x^{1/2}\\ =& \frac{1}{2}x^{-1/2}\\ =& \frac{1}{2\sqrt{x}}.\end{aligned}$$

Exponentials and Logs

The derivative of the exponential function, e^x, and the natural logarithm, $\ln(x)$, are given by the following:

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

Example 2.25.
Compute the derivative of e^{2x}.

We can compute this using the product rule since $e^{2x} = e^x \cdot e^x$,

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

(We'll see another way to compute this derivative later using the chain rule.)

Example 2.26.
Compute the derivative of $\log_2(x)$.

Recalling that we can compute logs of other bases in terms of the natural log using

$$\log_a(x) = \frac{\ln(x)}{\ln(a)},$$

we can compute the derivative as follows:

$$\begin{aligned}\frac{d}{dx}\log_2(x) &= \frac{d}{dx}\frac{\ln(x)}{\ln(2)} \\ &= \frac{1}{\ln(2)}\frac{d}{dx}\ln(x) \\ &= \frac{1}{x\ln(2)}.\end{aligned}$$

2.4 Derivatives of Trig Functions

Difficulties strengthen the mind, as labor does the body.

SENECA THE YOUNGER

The Six Trig Functions

There are six trigonometric functions, but the two most important ones are sine and cosine. Recall that sine and cosine are defined as follows. Given the *unit circle* (a circle or radius one centered at the origin), pick any number x. Draw a line segment from the origin, up at angle x from the positive x-axis until intersecting the unit circle. By definition, the x-coordinate of this point is $\cos(x)$, and the y-coordinate is $\sin(x)$.

Notice that $\sin(x)$ and $\cos(x)$ are the horizontal and vertical sides of a right triangle with hypotenuse 1. By the Pythagorean theorem we have

$$\sin^2(x) + \cos^2(x) = 1$$

which implies

$$\cos^2(x) = 1 - \sin^2(x) \quad \text{and} \quad \sin^2(x) = 1 - \cos^2(x).$$

The other four trig functions are defined as

$$\begin{aligned} \tan(x) &= \frac{\sin(x)}{\cos(x)} \\ \sec(x) &= \frac{1}{\cos(x)} \\ \csc(x) &= \frac{1}{\sin(x)} \\ \cot(x) &= \frac{\cos(x)}{\sin(x)} \end{aligned}$$

One helpful identity is the following:

Theorem 2.12.

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

A Useful Limit

Before differentiating the trig functions we need to justify one limit that will be helpful:

Theorem 2.13.

$$\lim_{x\to 0} \frac{\cos(x) - 1}{x} = 0$$

Proof.

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

□

Derivative of Sine, Cosine, and Tangent

We are now in a position to differentiate the trig functions.

Theorem 2.14.

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

Proof.

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

□

The proof of the following theorem is very similar to the previous proof and is left as an exercise:

Theorem 2.15.

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

Theorem 2.16.

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

Proof.

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

□

2.5 The Chain Rule

The calculus is the story this the Western world first told itself as it became the modern world.

DAVID BERLINSKI

Using our derivative rules we are now able to easily differentiate several different types of functions. However, there are still some functions we can't easily differentiate, such as

$$\sin\left(x^3 + 4x^2 + x\right) \quad \text{or} \quad \sqrt{\frac{x^2+1}{x^4+2}}.$$

Addition there are some functions that we can apply our rules to in principle, but doing so might be very difficult. For example, we could differentiate

$$(4x+2)^{39}$$

using our current rules, but we would first have to expand $(4x+2)^{39}$ which is certainly do-able, but extremely tedious.

Today we will discuss the chain rule which will allow us to easily calculate the derivatives above.

The Chain Rule

Theorem 2.17.
Suppose that $f(x)$ and $g(x)$ are differentiable. Then the composition $f(g(x))$ is also differentiable and

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

Before proving the chain rule, let's use it to differentiate the three functions mentioned above.

Example 2.27.
In the first example, $\sin\left(x^3+4x^2+x\right)$, our outer-most function is $\sin(x)$, and the inside function is x^3+4x^2+x. Using the notation above, $f(x)=\sin(x)$ and $g(x)=x^3+4x^2+x$. The chain rule then gives us the following:

$$\begin{aligned}\frac{d}{dx}\sin\left(x^3+4x^2+x\right) &= \cos\left(x^3+4x^2+x\right)\cdot\frac{d}{dx}\left(x^3+4x^2+x\right)\\ &= \cos\left(x^3+4x^2+x\right)\cdot\left(3x^2+8x+1\right)\end{aligned}$$

Example 2.28.
In our second example, $\sqrt{\frac{x^2+1}{x^4+2}}$, the outside function is $f(x)=\sqrt{x}$ and the inside function is $g(x)=\frac{x^2+1}{x^4+2}$.

$$\begin{aligned}\frac{d}{dx}\sqrt{\frac{x^2+1}{x^4+2}} &= \frac{1}{2\sqrt{\frac{x^2+1}{x^4+2}}}\cdot\frac{d}{dx}\frac{x^2+1}{x^4+2}\\ &= \frac{1}{2\sqrt{\frac{x^2+1}{x^4+2}}}\cdot\frac{(x^4+2)2x-(x^2+1)4x^3}{(x^4+2)^2}\end{aligned}$$

Example 2.29.
The polynomial $(4x+2)^{39}$ we could differentiate by expanding, but that's extremely annoying to do. Instead, we can take advantange of the chain rule where our outer-most function is $f(x)=x^{39}$ and the inner function is $g(x)=4x+2$.

$$\frac{d}{dx}(4x+2)^{39} = 39(4x+2)^{38}\cdot 4$$
$$= 156(4x+2)^{38}$$

Example 2.30.
Let's consider one more example where we will invoke the chain rule multiple times:

$$\frac{d}{dx}\cos\left(\sqrt[3]{7x^4+x^3-8x^2+2}\right)$$
$$= -\sin\left(\sqrt[3]{7x^4+x^3-8x^2+2}\right)\cdot\frac{d}{dx}\left(7x^4+x^3-8x^2+2\right)^{1/3}$$
$$= -\sin\left(\sqrt[3]{7x^4+x^3-8x^2+2}\right)\cdot\frac{1}{3}\left(7x^4+x^3-8x^2+2\right)^{-2/3}\cdot\frac{d}{dx}\left(7x^4+x^3-8x^2+2\right)$$
$$= -\sin\left(\sqrt[3]{7x^4+x^3-8x^2+2}\right)\cdot\frac{1}{3}\left(7x^4+x^3-8x^2+2\right)^{-2/3}\cdot\left(28x^3+3x^2-16x\right)$$

Now that we've seen how to use the chain rule, that's prove that this rule actually holds:

Proof of the chain rule.
Writing out the limit definition of the derivative of $f(g(x))$ at $x=a$

we have

$$\begin{aligned}
&\lim_{b\to a} \frac{f(g(b)) - f(g(a))}{b-a} \\
&= \lim_{b\to a} \frac{f(g(b)) - f(g(a))}{b-a} \cdot \frac{g(b)-g(a)}{g(b)-g(a)} \\
&= \lim_{b\to a} \frac{f(g(b)) - f(g(a))}{g(b)-g(a)} \cdot \frac{g(b)-g(a)}{b-a} \\
&= \left(\lim_{b\to a} \frac{f(g(b)) - f(g(a))}{g(b)-g(a)}\right) \cdot \left(\lim_{b\to a} \frac{g(b)-g(a)}{b-a}\right) \\
&= \left(\lim_{b\to a} \frac{f(g(b)) - f(g(a))}{g(b)-g(a)}\right) \cdot g'(a).
\end{aligned}$$

Now, as g is differentiable it must be continuous. That is, when b is near a, $g(b)$ is near $g(a)$. Write $h = g(b) - g(a)$, and notice that as $b \to a$ we have $h \to 0$ and we may write the above as

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

□

Derivative of an Inverse

We say that a function $f(x)$ is ***one-to-one*** if no two outputs of $f(x)$ are the same for two different inputs. That is, if $f(b) = f(a)$, then $b = a$. Graphically, this means the graph of the function $y = f(x)$ passes the horizontal line test. The function $f(x) = x^3$ is one-to-one:

$$\begin{aligned}
&f(b) = f(a) \\
\implies &b^3 = a^3 \\
\implies &b = a.
\end{aligned}$$

The function $f(x) = x^2$ is not one-to-one:

$$\begin{aligned}
&f(b) = f(a) \\
\implies &b^2 = a^2 \\
\implies &b = \pm a.
\end{aligned}$$

If a function is one-to-one then it has an ***inverse***: a way of converting outputs back into inputs. For example, $f(x) = 6x + 5$ has inverse $g(x) = \frac{x-5}{6}$. That $f(x)$ and $g(x)$ are inverses is expressed by the following equations:

$$f(g(x)) = x \quad \text{and} \quad g(f(x)) = x.$$

We can easily verify both of these equations in this example:

$$\begin{aligned} f(g(x)) =& 6g(x) + 5 \\ =& 6\frac{x-5}{6} + 5 \\ =& x - 5 + 5 \\ =& x \end{aligned}$$

$$\begin{aligned} g(f(x)) =& \frac{f(x)-5}{6} \\ =& \frac{6x+5-5}{6} \\ =& \frac{6x}{6} \\ =& x \end{aligned}$$

Theorem 2.18.
If a function has an inverse, its inverse must be unique: there can't be two different inverses for $f(x)$.

Proof.
Suppose $g_1(x)$ and $g_2(x)$ were two inverses for $f(x)$. We then have

$$\begin{aligned} f(g_1(x)) = x = f(g_2(x)) \text{ for all } x, \text{ and} \\ g_1(f(x)) = x = g_2(f(x)) \text{ for all} x. \end{aligned}$$

Thus

$$g_1(f(g_1(x))) = g_1(x) \text{ and } g_2(f(g_2(x))) = g_2(x)$$

but as $g_1(f(x)) = g_2(f(x))$, we must have $g_1(x) = g_2(x)$. $\square$

Since the inverse of a function must be unique, we sometimes write $f^{-1}(x)$ to denote the inverse of $f(x)$. *This is not* $1/f(x)$! For example, if $f(x) = \sqrt[3]{x}$, then $f^{-1}(x) = x^3$, not $\frac{1}{\sqrt[3]{x}}$.

Thus an inverse $f^{-1}(x)$ of $f(x)$ satisfies the following equations:

$$f(f^{-1}(x)) = x \text{ and } f^{-1}(f(x)) = x.$$

We can use these equations together with the chain rule to determine the derivative of an inverse.

Theorem 2.19.
If $f(x)$ is differentiable and one-to-one, then the inverse of $f(x)$ is differentiable and

$$\frac{d}{dx} f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}$$

Proof.

$$\begin{aligned} & f(f^{-1}(x)) = x \\ \implies & \frac{d}{dx} f(f^{-1}(x)) = \frac{d}{dx} x \\ \implies & \frac{d}{dx} f(f^{-1}(x)) = 1 \\ \implies & f'(f^{-1}(x)) \cdot \frac{d}{dx} f^{-1}(x) = 1 \\ \implies & \frac{d}{dx} f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}. \end{aligned}$$

$\square$

The derivative of $\arcsin(x)$ is thus given by $\frac{1}{\cos(\arcsin(x))}$.

3 Applications of Derivatives

3.1 Implicit Differentiation

Mathematics is less related to accounting than it is to philosophy.

LEONARD ADLEMAN

In this lecture we'll discuss a very useful tool for differentiating certain types of functions which are defined *implicitly*.

Implicit and Explicit Functions

Most of the functions we've seen in this class have been defined ***explicitly***: that is, we specify a "rule" for converting an input x into an output $f(x)$. All of the functions below are examples of such ***explicit*** functions:

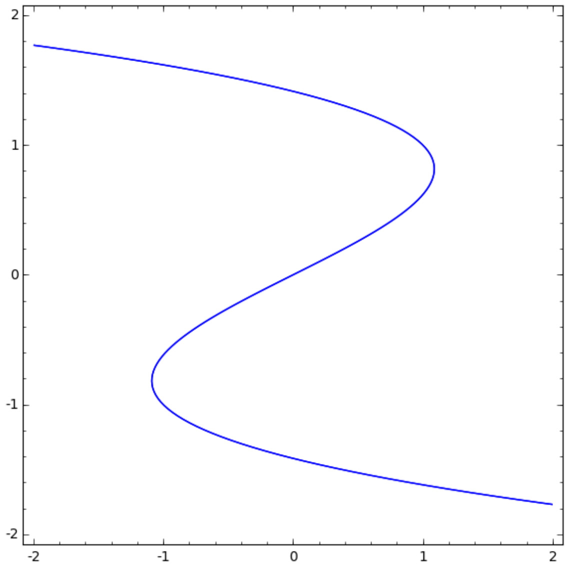
$$\begin{aligned} f(x) =& 6x^2 - 3x + 4 \\ g(x) =& \sin\left(x^2 + 2\right) \\ h(x) =& \frac{x^2 - \sqrt{x}}{\tan(x-2)} \end{aligned}$$

These are all examples of explicitly defined function because they tell us exactly how the function transforms its input x into an output $f(x)$, $g(x)$, or $h(x)$.

Functions which are defined ***implicitly*** are given by some property they must satisfy, instead of saying how to convert inputs into outputs. For example, consider the function $f(x)$ satisfying the following relationship:

$$f(x)^3 - f(x) = f(x) - x$$

When we write an expression like this we're defining $f(x)$ ***implicitly***. We're saying that $f(x)$ has to satisfy some particular equation, instead of saying how $f(x)$ should convert inputs into outputs.



As another example which might be more familiar, consider the function $f(x)$ which satisfies the following equation:

$$f(x)^2 + x^2 = 1.$$

In an example like this we might try to solve for $f(x)$ algebraically to determine the "rule" for converting inputs to outputs:

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

Notice that there are two different functions satisfying our rule: if we want a function $f(x)$ satisfies

$$f(x)^2 + x^2 = 1,$$

then we can take $f(x)$ be either $f(x) = \sqrt{1-x^2}$ or $f(x) = -\sqrt{1-x^2}$.

This is a general phenomenon when we're dealing with implicit functions: there may actual be several different functions satisfying the given equation.

Graphically, if we graph the set of solutions to an equation like

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

then what we get is *not* the graph of a function if the graph fails the vertical line test. However, if we only focus on a small portion of the graph at a time, then we may find that there are functions which give pieces of the graph. In the case of the circle, we just saw that we can think of the circle as being given by the graphs of two functions: one for the positive square root (the top half of the circle), and one for the negative square root (the bottom half of the circle).

In the case of our earlier example,

$$f(x)^3 - f(x) = f(x) - x$$

if we plot all of the solutions to

$$y^3 - y = y - x$$

we do not get the graph of a function. However, if we look at small enough pieces of this graph, then we do get the graph of a function. This indicates that there are actually three different functions that satisfy the equation

$$f(x)^3 - f(x) = f(x) - x$$

Solving for $f(x)$ algebraically in this example is not too difficult, but also not quite as easy as solving $f(x)^2 + x^2 = 1$ as before. In other more complicated examples, it may not even be possible. Despite this, we may still want to differentiate these implicitly defined functions – even if we can't solve for them explicitly. Graphically this corresponds to finding the slope of a tangent line to a curve such as

$$y^3 - y = y - x \quad \text{or} \quad y^2 + x^2 = 1.$$

To do this we use a technique called ***implicit differentiaton***.

Implicit Differentation

Suppose that we wanted to compute $f'(x)$ where $f(x)$ is given implicitly by the equation

$$f(x)^2 + x^2 = 1.$$

We could of course do this by first solving for $f(x)$ in this example, but let's consider another approach.

Given any equation (of differentiable functions), we could differentiate both sides of the equation to get a new equation. In the example at hand,

$$\begin{aligned} &f(x)^2 + x^2 = 1 \\ \implies &\frac{d}{dx}\left(f(x)^2 + x^2\right) = \frac{d}{dx}1 \end{aligned}$$

The right-hand side is easy to differentiate: it's just $\frac{d}{dx}1 = 0$. But how should we differentiate the right-hand side, in particular the $f(x)^2$ term? Well, by the chain rule, we know that the derivative of $f(x)^2$ must have the following form:

$$\frac{d}{dx}f(x)^2 = 2f(x) \cdot f'(x).$$

We don't know what $f(x)$ necessarily, and so we don't (yet) know what $f'(x)$ is, but that's okay. Continuing,

$$\begin{aligned} & f(x)^2 + x^2 = 1 \\ \implies & \frac{d}{dx}\left(f(x)^2 + x^2\right) = \frac{d}{dx}1 \\ \implies & 2f(x)f'(x) + 2x = 0. \end{aligned}$$

Notice we can now solve for $f'(x)$!

$$\begin{aligned} & 2f(x)f'(x) + 2x = 0 \\ \implies & f'(x) = \frac{-2x}{2f(x)} = -\frac{x}{f(x)}. \end{aligned}$$

So, if we have an implicitly defined function, its derivative is also implicitly defined!

This may not sound very useful at first: you may think "well, if we don't know what $f(x)$ is, we still don't know what $f'(x)$ is since $f'(x)$ depends on $f(x)$." This is true, in the sense that we don't have an explicit formula for $f'(x)$, however let's see an example of why this is useful.

Example 3.1.
Find the equation of the line tangent to $y^2 + x^2 = 1$ at the point $(\sqrt{3}/2, 1/2)$.

Writing $f(x)$ in place of y we have the equation

$$f(x)^2 + x^2 = 1.$$

As we just saw, the derivative of this function satisfies the following equation:

$$f'(x) = -\frac{x}{f(x)}.$$

Even though $f(x)^2 + x^2 = 1$ doesn't give us a single function, each point on $y^2 + x^2 = 1$ lives on the graph of *some* function, whatever it is. In particular, for some $f(x)$ satisfying the equation $f(x)^2 + x^2 = 1$, we have $x = \sqrt{3}/2$ and $f\left(\sqrt{3}/2\right) = 1/2$. Thus

$$f'\left(\sqrt{3}/2\right) = -\frac{\sqrt{3}/2}{f\left(\sqrt{3}/2\right)} = -\sqrt{3}.$$

This means the slope of the tangent line has to be $-\sqrt{3}$!

Thus the equation of the line tangent to the circle $y^2 + x^2 = 1$ at the point is

$$y - 1/2 = -\sqrt{3}\left(x - 3/2\right).$$

Usually when we do a problem like this, we don't bother to write $f(x)$: instead we just use y, and write $\frac{dy}{dx}$ in place of $f'(x)$. So to differentiate the equation $y^2 + x^2 = 1$ we have

$$\begin{aligned} & y^2 + x^2 = 1 \\ \implies & \frac{d}{dx}\left(y^2 + x^2\right) = \frac{d}{dx}1 \\ \implies & \left(\frac{d}{dx}y^2\right) + 2x = 0 \\ \implies & 2y\frac{dy}{dx} + 2x = 0 \\ \implies & 2y\frac{dy}{dx} = -2x \\ \implies & \frac{dy}{dx} = -\frac{x}{y}. \end{aligned}$$

The important thing to realize here is that we're just doing the chain rule.

Example 3.2.
If x and y satisfy the equation $y^3 - y = y - x$, find $\frac{dy}{dx}$.

$$
\begin{aligned}
& y^3 - y = y - x \\
\implies & 3y^2 \frac{dy}{dx} - \frac{dy}{dx} = \frac{dy}{dx} - 1 \\
\implies & 3y^2 \frac{dy}{dx} - 2\frac{dy}{dx} = -1 \\
\implies & \left(3y^2 - 2\right) \frac{dy}{dx} = -1 \\
\implies & \frac{dy}{dx} = -\frac{1}{3y^2 - 2}
\end{aligned}
$$

Example 3.3.
Find the equation of the line tangent to the curve $y^3 - y = y - x$ at the point $(0, \sqrt{2})$.

By our previous example, we know

$$\frac{dy}{dx} = -\frac{1}{3y^2 - 2}$$

If we plug in $x = 0$ and $y = \sqrt{2}$ in, we have

$$\frac{dy}{dx} = -\frac{1}{4}.$$

Thus the equation of the tangent line is

$$y - \sqrt{2} = -\frac{1}{4}x$$

Example 3.4.
Find the equation of the line tangent to the hyperbola $x^2 - y^2 = 1$ at the point $(3, -\sqrt{8})$.

Again, we first find $\frac{dy}{dx}$:

$$\begin{aligned} x^2 - y^2 &= 1 \\ \implies 2x - 2y\frac{dy}{dx} &= 1 \\ \implies -2y\frac{dy}{dx} &= 1 - 2x \\ \implies \frac{dy}{dx} &= \frac{2x-1}{y}. \end{aligned}$$

So at the point $(3, -\sqrt{8})$, the slope of the tangent line is $\frac{5}{-\sqrt{8}}$, and the equation of the line is

$$y + \sqrt{8} = \frac{5}{-\sqrt{8}}(x-3)$$

Example 3.5.
Find all the x-coordinates of points on the curve

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

where the tangent lines are horizontal.

The tangent line will be horizontal when its slope is zero, so we need to find the points where $\frac{dy}{dx} = 0$. First we find $\frac{dy}{dx}$:

$$\begin{aligned} 2y^3 + y^2 - y^5 &= x^4 - 2x^3 + x^2 \\ \implies 6y^2\frac{dy}{dx} + 2y\frac{dy}{dx} - 5y^4\frac{dy}{dx} &= 4x^3 - 6x^2 + x \\ \implies \frac{dy}{dx} &= \frac{4x^3 - 6x^2 + x}{6y^2 + 2y - 5y^4} \end{aligned}$$

If this is to equal zero, then the numerator must equal zero:

$$\begin{aligned} 4x^3 - 6x^2 + x &= 0 \\ \implies x(4x^2 - 6x + 1) &= 0 \end{aligned}$$

So either $x = 0$, or $4x^2 - 6x + 1 = 0$. To solve the second equation, we use the quadratic formula:

$$\begin{aligned} x =& \frac{6 \pm \sqrt{36-16}}{8} \\ =& \frac{6 \pm \sqrt{20}}{8} \\ =& \frac{6 \pm 2\sqrt{5}}{8} \\ =& \frac{3 \pm \sqrt{5}}{4} \end{aligned}$$

So this curve will have horizontal tangent lines when $x = 0$, $x = \frac{3-\sqrt{5}}{4}$, and $x = \frac{3+\sqrt{5}}{4}$.

Example 3.6.
Sometimes we can use implicit differentation in unexpected ways. For example, suppose we wanted to differentiate

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

We saw how to do this earlier in the semester using the chain rule, but if you had forgotten the rule, you could actually figure this out with implicit differentation as follows:

$$\begin{aligned} & y = \tan^{-1}(x) \\ \Longrightarrow\ & \tan(y) = x \\ \Longrightarrow\ & \frac{d}{dx}\tan(y) = \frac{d}{dx}x \\ \Longrightarrow\ & \sec^2(y)\frac{dy}{dx} = 1 \\ \Longrightarrow\ & \frac{dy}{dx} = \frac{1}{\sec^2(y)} = \frac{1}{\sec^2(\tan^{-1}(x))}. \end{aligned}$$

Logarithmic Differentiation

Sometimes we can use implicit differentiation to help us differentiate explicitly defined functions. In particular, derivatives of functions with lots of products and quotients can be made simpler by first taking the log of the function. This turns products into sums and quotients into differences which are much easier to differentiate.

Example 3.7.
Find the derivative of the following function:

$$\sqrt[3]{\frac{x\sin(x)}{x^2+2x-1}}.$$

We *can* differentiate this function using our previous differentiation rules, but we can also avoid having to apply the product and quotient rules by taking the log of the function. In particular, if we write

$$y = \sqrt[3]{\frac{x\sin(x)}{x^2+2x-1}}$$

then the derivative we want to calculate is $\frac{dy}{dx}$. Taking the log of both sides of the equation above gives us

$$\begin{aligned}
\ln(y) &= \ln\left(\sqrt[3]{\frac{x\sin(x)}{x^2+2x-1}}\right) \\
&= \ln\left(\left(\frac{x\sin(x)}{x^2+2x-1}\right)^{1/3}\right) \\
&= \frac{1}{3}\ln\left(\frac{x\sin(x)}{x^2+2x-1}\right) \\
&= \frac{1}{3}\big(\ln(x) + \ln(\sin(x)) - \ln(x^2+2x-1)\big)
\end{aligned}$$

Now we can differentiate both sides of this equation without using any product rules or quotient rules,

$$
\begin{aligned}
\ln(y) &= \frac{1}{3}\big(\ln(x) + \ln(\sin(x)) - \ln(x^2 + 2x - 1)\big) \\
\Longrightarrow \frac{d}{dx}\ln(y) &= \frac{d}{dx}\frac{1}{3}\big(\ln(x) + \ln(\sin(x)) - \ln(x^2 + 2x - 1)\big) \\
\Longrightarrow \frac{1}{y}\frac{dy}{dx} &= \frac{1}{3}\left(\frac{1}{x} + \frac{\cos(x)}{\sin(x)} - \frac{2x+2}{x^2+2x-1}\right) \\
\Longrightarrow \frac{dy}{dx} &= y\frac{1}{3}\left(\frac{1}{x} + \frac{\cos(x)}{\sin(x)} - \frac{2x+2}{x^2+2x-1}\right) \\
\Longrightarrow \frac{dy}{dx} &= \frac{1}{3}\sqrt[3]{\frac{x\sin(x)}{x^2+2x-1}}\left(\frac{1}{x} + \frac{\cos(x)}{\sin(x)} - \frac{2x+2}{x^2+2x-1}\right)
\end{aligned}
$$

As another example of using this trick, we can take logs to pull exponents down. This is helpful if the exponent is a function of x instead of just a constant:

Example 3.8.
Compute the following derivative:

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

We again write

$$y = \sin(x)^{\cos(x)}$$

and so we want to find $\frac{dy}{dx}$. Taking logs of both sides of the equation gives us

$$\ln(y) = \ln\left(\sin(x)^{\cos(x)}\right) = \cos(x)\ln(\sin(x))$$

we now differentiate both sides to obtain

$$\frac{1}{y}\frac{dy}{dx} = -\sin(x)\ln(\sin(x)) + \cos(x)\cdot\frac{\cos(x)}{\sin(x)}.$$

Solving for $\frac{dy}{dx}$, and keeping in mind $y = \sin(x)^{\cos(x)}$, gives

$$\frac{dy}{dx} = \sin(x)^{\cos(x)} \cdot (\cos(x)\cot(x) - \sin(x)\ln(\sin(x)))\,.$$

3.2 Related Rates

Experience has shown repeatedly that a mathematical theory with a rich internal structure generally turns out to have significant implications for the understanding of the real world, often in ways no one could have envisioned before the theory was developed.

WILLIAM THURSTON

Examples

Sometimes there will be a quantity whose derivative you're interested in, but you can't measure the quantity directly, though you might be able to measure a related quantity. We can then use the related quantity to calculate the derivative we care about – this trick is really just the chain rule, but in this context such problems are called *related rates*.

Example 3.9.
Suppose we blow air into a balloon at a rate of 50 cubic inches every second. How quickly is the radius of the balloon changing in general? (Assume the ballon is a perfect sphere.) How quickly is the radius changing when the radius is 3 inches? 4 inches? What about when the volume is $\frac{243\pi}{16}$ cubic inches?

We want to know one particular quantity, rate of change of the radius, but what we have is a different (though related) quantity: rate of change of the volume. That is, we know $\frac{dV}{dt} = 50$, but we want to find $\frac{dr}{dt}$. However, V can be expressed in terms of r:

$$V = \frac{4}{3}\pi r^3.$$

If we take $\frac{d}{dt}$ of both sides we have

$$\begin{aligned} V &= \frac{4}{3}\pi r^3 \\ \implies \frac{d}{dt}V &= \frac{d}{dt}\frac{4}{3}\pi r^3. \end{aligned}$$

Notice when we differentiate the right-hand side we'll do implicit differentiation: r is a function of time, but we don't know that function explicitly.

$$\begin{aligned} V &= \frac{4}{3}\pi r^3 \\ \implies \frac{d}{dt}V &= \frac{d}{dt}\frac{4}{3}\pi r^3 \\ \implies \frac{dV}{dt} &= \frac{4}{3}\pi 3r^2\frac{dr}{dt} \end{aligned}$$

Now we can simply solve for $\frac{dr}{dt}$:

$$\frac{dr}{dt} = \frac{dV/dt}{4\pi r^2},$$

and we already know that $\frac{dV}{dt}$ is 50, so

$$\frac{dr}{dt} = \frac{50}{4\pi r^2}\text{inches per second}$$

When $r = 3$, the radius is increasing at a rate of

$$\left.\frac{dr}{dt}\right|_{r=3} = \frac{50}{36\pi} \approx 0.442\text{inches per second.}$$

When $r = 4$, the radius is increasing at a rate of

$$\left.\frac{dr}{dt}\right|_{r=4} = \frac{50}{64\pi} \approx 0.249\text{inches per second.}$$

Our formula here depends on us knowing r. So to find $\frac{dr}{dt}$ when the volume is $V = \frac{243\pi}{16}$, we need to find the corresponding r. Going back to $V = {}^4\!/\!{}_3\pi r^3$, we have

$$\begin{aligned} \frac{243\pi}{16} &= \frac{4}{3}\pi r^3 \\ \Longrightarrow r^3 &= \frac{729}{64} \\ \Longrightarrow r &= \frac{9}{4}. \end{aligned}$$

So when the volume is $\frac{243\pi}{16}$, the radius is $\frac{9}{4}$, and the rate of change of the radius is

$$\left.\frac{dr}{dt}\right|_{r={}^9\!/\!{}_4} = \frac{50}{{}^{81}\!/\!{}_4\pi} \approx 0.789\text{inches per second}.$$

This example illustrates the key idea behind related rates problems: there's a rate of change (derivative) you care about, but you're given something that's different. Try to relate the two quantities (the one you care about, and the one you're given) with an equation, then use implicit differentiation to find the rate of change.

Example 3.10.
Suppose a particle moves along the curve $xy = 1$, and when it reaches the point $(2, {}^1\!/\!{}_2)$, suppose the x-coordinate is increasing at a rate of 3. How quickly is the y-coordinate increasing?

We know $\frac{dx}{dt}$ and want to find $\frac{dy}{dt}$, but we know the two quantities are related by $xy = 1$. Differentiating both sides gives

$$\begin{aligned} xy &= 1 \\ \Longrightarrow \frac{d}{dt}xy &= \frac{d}{dt}1 \\ \Longrightarrow \frac{dx}{dt}y + x\frac{dy}{dt} &= 0 \\ \Longrightarrow x\frac{dy}{dt} &= -y\frac{dx}{dt} \\ \Longrightarrow \frac{dy}{dt} &= \frac{-y}{x}\frac{dx}{dt}. \end{aligned}$$

We now plug in $x = 2$, $y = \frac{1}{2}$, and $\frac{dx}{dt} = 3$ to obtain

$$\left.\frac{dy}{dt}\right|_{(x,y)=(2,1/2)} = \frac{-1/2}{2} \cdot 3 = \frac{-3}{4}.$$

Example 3.11.
Consider a cylindrical tank of radius 7 feet which is being filled with water at a rate of 15 gallons per minute. How quickly is the height of the water rising?

We're told $\frac{dV}{dt} = 15$ and want to find $\frac{dh}{dt}$. So we need a formula relating the height of a cone to its volume, but the volume of a cone of height h and radius r is

$$V = \pi r^2 h.$$

In our situation $r = 7$, so we have

$$V = 49\pi h.$$

Differentiating both sides gives

$$\frac{dV}{dt} = 49\pi \frac{dh}{dt}$$

and we know $\frac{dV}{dt} = 15$, so

$$\frac{dh}{dt} = \frac{15}{49\pi}.$$

Example 3.12.
Consider a boat being pulled into a dock which is 3 feet above the surface of the water. If the rope used to pull the boat in is being reeled in at a rate of 2 feet per minute, how quickly is the boat approach the dock when 5 feet of rope is let out?

Consider the right triangle whose height represents the dock three feet above the water, and whose base is the distance of the boat from the dock – let's call this b. The hypotenuse of this triangle represents the length of rope let out, let's call this r.

By the Pythagorean theorem,

$$r = \sqrt{9 + b^2}.$$

We know $\frac{dr}{dt} = -2$ and want to find $\frac{db}{dt}$ when $r = 5$. So, we differentiate both sides of our equation above:

$$\begin{aligned} & r = \sqrt{9 + b^2} \\ \implies & \frac{dr}{dt} = \frac{1}{2\sqrt{9 + b^2}} \cdot 2b\frac{db}{dt} \implies \qquad \frac{db}{dt} = \frac{dr}{dt} \cdot \frac{2\sqrt{9 + b^2}}{2b}. \end{aligned}$$

We know that $\frac{dr}{dt} = -2$. When $r = 10$, the Pythagorean theorem tells us

$$\begin{aligned} & 5 = \sqrt{9 + b^2} \\ \implies & 25 = 9 + b^2 \\ \implies & b^2 = 16 \\ \implies & b = 4 \quad \text{(Note } b = -4 \text{ makes no sense in this problem.)} \end{aligned}$$

Now plugging this into the above we have

$$\begin{aligned} \frac{db}{dt} &= -2 \cdot \frac{2\sqrt{9 + 16}}{8} \\ &= -\frac{10}{4} = -\frac{5}{2}. \end{aligned}$$

3.3 Linearization & Differentials

The beauty of mathematics only shows itself to more patient followers.

MARYAM MIRZAKHANI

As we've said before in class, the whole idea behind calculus is to take hard problems and make them simpler by approximating with something that's easier to work with. One of the best examples of this occurs with *linearization*, which is the idea that we should approximate complicated functions with simpler ones and in particular with linear functions. This idea has had, and continues to have, a profound impact on applications of mathematics. If you've ever wondered how it is a computer or calculator is able to compute $\cos(22.091)$ or $\sqrt{16.29}$, then you may be surprised to learn that the answer relies on calculus.

Motivation

If you think about the mathematical procedures that you can "really" do, the things that you can in principle sit down and work out with pencil and paper, you may come to the realization that you only know how to do four things: add, subtract, multiply, and divide. Of course, you can do some other things like square or cube a number, but this is just multiplication applied several times. Other operations – even ones as simple as taking square roots! – are much, much harder to do by hand. Without a calculator there are probably only a small handful of numbers whose square roots you can actually calculate: things you can actually work out the answer to with just a pencil and paper.

A computer is no different. Computers are programmed by people, so if the computer is about to determine the square root of 384.193, then someone had to tell it how to do that. And computers don't posess some magical computational ability that you don't: when you get down to the nuts and bolts of it, a computer can also only add, subtract, multiply, and divide. This is meant quite literally, by the way. In terms of what the hardware of a computer is actually able to do there are special circuits that use combinations of logical operators (AND, OR, NOT) to do arithmetic with numbers represented in binary (base 2). In some sense you are actually much better at arithmetic than a computer: a computer only has a finite amount of space to store numbers, but in principle there's no actual limitation on what a person with pen and paper can do (even if there are serious practical limitations).

So, this still begs the question: if you can only add, substract, multiply, or divide, how is it that you're supposed to compute a quantity like $\tan^{-1}(\sqrt{17} + 3^{\pi})$? In terms of the functions you can build with the four arithmetic operations – i.e., the functions you can actually evaluate – all you have are rational functions: these are things built from addition, subtraction, multiplication, and division. In fact, for what we're going to do today we're going to replace complicated "transcendental" functions with about the simplest type of function of all: a linear function. A linear function, by the way, is just a function whose graph is a line. So it's a function that looks something like

$$f(x) = ax + b.$$

Notice that to evaluate such a linear function we only need to be able to multiply and add.

Linearization

The whole idea behind derivatives is that if a function is differentiable, then it can be well approximated by its tangent line – at least for a little while. The whole theory of derivatives is basically cooked up to make this precise, but this is the key idea: if a function is differentiable, then we can approximate it with its tangent line which is given by a linear function.

To approximate a function like $\sin(x)$, for example, what we should do is compute tangent lines near the x-value we want to evalute. If we wanted to approximate $\sin(0.123)$, then since 0.123 is close to 0 *and since we can evaluate both sine and its derivative exactly at* $x = 0$, we can use the tangent line to $y = \sin(x)$ at $x = 0$ to help us approximate $\sin(0.123)$.

Example 3.13.
Use the equation of the line tangent to $y = \sin(x)$ at $x = 0$ to approximate $\sin(0.123)$.

First let's calcualte the tangent line. Since the derivative of sine is cosine and $\cos(0) = 1$, the tangent line is

$$y = x.$$

That is, for x-values near 0 we have that

$$\sin(x) \approx x.$$

And so we compute that $\sin(0.123) \approx 0.123$.

For the sake of comparison, if you plug $\sin(0.123)$ into a calculator it'll come back with something like 0.12269. This means two things: our approximation to $\sin(0.123)$ was actually pretty good, and the calculator is using a slightly different method of approximating $\sin(0.123)$. If you take the second semester of calculus in the spring you'll actually learn exactly the approximation technique that the calculator is using, which is called a ***Taylor polynomial***, but you have to built up a little bit more calculus first. What we're doing now is basically the first step towards Taylor polynomials.

Example 3.14.

Use the equation of the line tangent to $y = \sqrt{x}$ at $x = 4$ to approximate $\sqrt{4.03}$.

Since $\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}$, the slope of our tangent line will be $\frac{1}{2\sqrt{4}} = \frac{1}{4}$. The actual equation of the tangent line is

$$y - 2 = \frac{1}{4}(x - 4).$$

Let's write this in slope-intercept form, just so it has the shape of $y = f(x)$:

$$y = \frac{1}{4}(x - 4) + 2.$$

Now, if we plug $x = 4.03$ into this we have an approximation to

$\sqrt{4.03}$:

$$\begin{aligned}\sqrt{4.03} \approx& \frac{1}{4}(4.03-4)+2\\ =& \frac{0.03}{4}+2\\ =& \frac{0.015}{2}+2\\ =& 0.0075+2\\ =& 2.0075.\end{aligned}$$

Again, pluggint $\sqrt{4.03}$ into a calculator just for the sake of comparison gives us $\sqrt{4.03} \approx 2.007486$. So a pretty good approximation!

In general, if we want to approximate $f(a)$ when a is close to some value a_0 where we can truly calculate $f(a_0)$ and $f'(a_0)$, we calculate the tangent line to $y = f(x)$ at the point $(a, f(a_0))$, which is

$$y - f(a_0) = f'(a_0)(x - a_0)$$

we then rewrite this in slope-intercept form,

$$y = f'(a_0)(x - a_0) + f(a_0)$$

and use the function on the right-hand side as our approximation.

This function on the right is called the ***linearization of $f(x)$ at $x = a_0$*** and is normally denoted $L(x)$:

$$y = \underbrace{f'(a_0)(x - a_0) + f(a_0)}_{L(x)},$$

and the idea is that for x-values near a_0, $f(x) \approx L(x)$.

Notice the linearization always takes the form

$$L(x) = f'(a_0)(x - a_0) + f(a_0)$$

Example 3.15.
Calculate the linearization of $f(x) = 3x^2 + 4x - 2$ at $x = 5$ and use this to approximate $f(5.122)$.

First we calculate the linearization:

$$L(x) = f'(5)(x-5) + f(5)$$

So we need to know $f(5) = 3 \cdot 5^2 + 4 \cdot 5 - 2 = 93$. To calculate $f'(5)$ we first calculate $f'(x)$ in general:

$$f'(x) = 6x + 4$$

and evaluating this at 5 gives $f'(5) = 34$. Hence our linearization is

$$L(x) = 34(x-5) + 93.$$

If we plug in $x = 5.122$ we have

$$\begin{aligned} f(5.122) \approx& L(5.122) \\ =& 34 \cdot 0.122 + 93 \\ =& 4.148 + 93 \\ =& 97.148. \end{aligned}$$

Again, for comparison purposes, let's double-check this with a calculator which tells us $f(5.122)$ is about 97.1926. So again we have an okay approximation. When you learn about Taylor polynomials later, howeve,r you'll see how to make this approximation even better.

Example 3.16.
Use a linearization to approximate $f(x) = \frac{x^2+1}{x}$ at the point $x = 1.95$.

We calculate the linearization at $x = 2$:

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

So $f'(2) = \frac{3}{4}$, and $f(2) = \frac{5}{2}$. The linearization is thus

$$L(x) = \frac{3}{4}(x-2) + \frac{5}{2}.$$

So

$$\begin{aligned} f(1.95) \approx & L(1.95) \\ = & \frac{3}{4}(1.95-2)+\frac{5}{2} \\ = & \frac{3}{4}0.05+\frac{5}{2} \\ = & \frac{3}{4}\cdot\frac{1}{20}+\frac{5}{2} \\ = & \frac{3}{80}+\frac{200}{80} \\ = & \frac{203}{80} \\ = & 2.5375 \end{aligned}$$

For comparison, a calculator tells us the answer is 2.4628.

Example 3.17.
Approximate $\sqrt[3]{65}+\sqrt{65}$.

Notice that we know $\sqrt[3]{64}+\sqrt{64}=4+8=12$, so let's base our linearization at $a_0=64$. Here $f(x)=\sqrt[3]{x}+\sqrt{x}=x^{1/3}+x^{1/2}$, and so

$$f'(x)=\frac{1}{3x^{2/3}}+\frac{1}{2\sqrt{x}}.$$

At $a_0=64$ we have

$$\begin{aligned} f'(64) = & \frac{1}{3\cdot 16}+\frac{1}{2\cdot 8} \\ = & \frac{1}{48}+\frac{1}{16} \\ = & \frac{1}{48}+\frac{3}{48} \\ = & \frac{4}{48} \\ = & \frac{1}{12}. \end{aligned}$$

Our linearization is thus

$$L(x) = \frac{1}{12}(x - 64) + 12.$$

Thus

$$\begin{aligned}
\sqrt[3]{65} + \sqrt{65} \approx& L(65) \\
=& \frac{1}{12}(65 - 64) + 12 \\
=& \frac{1}{12} + 12 \\
=& \frac{1}{12} + \frac{144}{12} \\
=& \frac{145}{12} \\
\approx& 12.0833...
\end{aligned}$$

For comparison, a calculator tells us $\sqrt[3]{65} + \sqrt{65}$ is about $12.08298...$.

Differentials

Differentials give us another point of view of linearizations: they are essentially the same thing as a linearization, but expressed differently.

Given a differentiable function $f(x)$ and a value a where we can compute the true values of $f(a)$ and $f'(a)$, we saw in the last lecture that for "nearby" x-values, $f(x)$ can be approximated by the ***linearization*** $L(x)$,

$$L(x) = f'(a)(x - a) + f(a).$$

If we're doing several calculations of nearby x-values, then we certainly don't need to re-compute the function $L(x)$ each time. For example, suppose we want to approximate $\sqrt{8.99}$ and 9.1. We can use the linearization of $f(x) = \sqrt{x}$ at $a = 9$:

$$L(x) = \frac{1}{6}(x - 9) + 3.$$

Plugging in 8.99 and 9.1 into this linearization, the only difference between the two expressions for $L(8.99)$ and $L(9.1)$ is the quantity $x - 9$:

$$L(8.99) = \frac{1}{6}(8.99 - 9) + 3 = \frac{1}{6}(-0.01) + 3 = 2.998333...$$
$$L(9.1) = \frac{1}{6}(9.1 - 9) + 3 = \frac{1}{6}(0.1) + 3 = 3.01666...$$

So, once we've determined the formula for our linearization, the only thing that can change when we approximate nearby points is the factor of $x - a$. This quantity represents how much our x-value has changed from a, and so we're justified in representing this quantity as Δx.

We could now reasonably write the linearization of our function as

$$L(x) = f'(a)\Delta x + f(a).$$

Since $f(a)$ is the same for any quantity we compute with this linearization, what we really care about is how much $L(x)$ differs from $f(a)$. If we notice that $L(a) = f(a)$, we could rewrite this as follows

$$\begin{aligned} L(x) &= f'(a)\Delta x + f(a) \\ \implies L(x) &= f'(a)\Delta x + L(a) \\ \implies L(x) - L(a) &= f'(a)\Delta x. \end{aligned}$$

This quantity, $L(x) - L(a)$ is really what we're interested in: it tells us how much our approximation changes as we change the x-value. Notice too that this quantity is a change in y-values, so you might be tempted to denote this quantity by Δy. However, the convention that has been adopted is that Δy should mean the change in the true value of the function and *not* the change in the approximation. Since the Greek letter Δ is already taken, let's instead use the Latin letter d to write

$$dy = L(x) - L(a).$$

We could then write $dy = f'(a)\Delta x$, but since we're adopting the convention that we reserve Δ to mean the "true" change and d means the "approximate" change, we will write dx in place of Δx – although this is exactly the same quantity. We thus arrive at the formula $dy = f'(a)dx$

So, in the example above where $f(x) = \sqrt{x}$ and $a = 9$, we have

$$dy = \frac{1}{6}dx.$$

This quantity is called ***the differential of*** $f(x) = \sqrt{x}$ ***at*** 9.

Of course, we could have calculated this dy quantity at another place – using something other than 9 – so we should really imagine that quantity is variable, which is the x-coordinate of our original function, so let's continue to call it x.

We then define ***the differential*** of $f(x)$ to be

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

(The notation $df = f'(x)dx$ is also common and is also called the differential.)

There are a few things to notice about these differentials we've defined. The first is that dy and dx here area actual numeric values: they're not just symbols. In fact, dy depends on dx: dy is a function of dx with dx being an independent variable.

An interesting byproduct of our definitions is that if we divide both sides of the equation $dy = f'(x)dx$ by dx, then we have $\frac{dy}{dx} = f'(x)$. This isn't simply a coincidence: by definition, $f'(x)$ is a limit of changes in y-values over changes in x-values. The intuition behind dy and dx is that they should represent "infinitesimal changes" in x or y, and this is how people like Newton and Leibniz originally thought about derivatives. (There is a way to make "infinitesimal changes" precise, but discussing it would take us very far afield.)

The other important thing to notice is that this dy quantity is the important part of a linear approximation. We can write the linearization of a function at a as

$$L(x) = dy + f(a).$$

We can use this to help us calculate linear approximations. In the case of $\sqrt{8.99}$ and $\sqrt{9.1}$, for example, where we took $a = 9$, we have $L(x) = dy + 3 = \frac{1}{6}dx + 3$. For 8.99, $dx = 8.99 - 9 = -0.01$, and so $\sqrt{8.99} \approx \frac{-0.01}{6} + 3 = 2.998333...$, just as before. So, again, differentials are really just linearizations from another point of view.

Examples

We will first do some examples where we calculate differentials "formally," and then do some examples where we apply differentials to help us solve some approximation problems.

Example 3.18.

(a) Compute dy where $y = 8x^2 + 6x$.

$$dy = \frac{d}{dx}\left(8x^2 + 6x\right) dx = (16x + 6)dx$$

(b) Compute dy where $y = \sin(\sqrt{x})\cos(x^2)$.

$$dy = \left(\frac{\sin(\sqrt{x})}{2\sqrt{x}}\cos(x^2) + 2x\sin(\sqrt{x})\cos(x^2)\right) dx$$

(c) Compute df where $f(x) = \frac{x^2+1}{x}$.

$$df = f'(x)dx = \frac{x^2 - 1}{x^2}dx.$$

Example 3.19.
Use differentials to approximate $\sqrt{24.9}$, $\sqrt{25.1}$, and $\sqrt{24.975}$.

Since these values are near $\sqrt{25} = 5$, we should expect they are all approximately $dy + 5$ where dy is the differential of $\sqrt{x}$ at $x = 25$.

$$y = \sqrt{x}$$
$$\implies dy = \frac{1}{2\sqrt{x}}dx$$

When $x = 25$ we have $dy = \frac{dx}{10}$.

For $\sqrt{24.9}$, $dx = -0.1$ and so $\sqrt{24.9} \approx \frac{-0.1}{10} + 5 = 4.99$.

For $\sqrt{25.1}$, $dx = 0.1$ and so $\sqrt{25.1} \approx \frac{0.1}{10} + 5 = 5.01$.

For $\sqrt{24.975}$, $dx = -0.025$, and so $\sqrt{24.975} \approx \frac{-0.025}{10} + 5 = 5 - 0.0025 = 4.9975$.

Example 3.20.
Use differentials to estimate the amount of paint needed to apply a coat of paint 0.05 cm thick on a hemispherical dome of diameter 50m.

The way to interpret this problem is that we want to know what volume of paint is needed. We basically want the volume between two spheres: one of radius 25, and one of radius 25.0005. We estimate this using differentials: we want to change in volume as we go from one sphere to the other. So our r in the problem will be 25 and $dr = 0.0005$.

The volume of a sphere of radius r is

$$V = \frac{4}{3}\pi r^3.$$

Since we have a hemisphere, we need half of this quantity,

$$V = \frac{4}{6}\pi r^3.$$

Then

$$dV = 2\pi r^2 dr.$$

When $r = 25$ and $dr = 0.0005$ we have

$$dV = 2\pi 25^2 \cdot 0.0005 = 1.964.$$

Let's notice the units here: r is in metres and squared, and dr is also in metres, so the units are cubic metres.

We need about 1.964 cubic metres of paint to paint the dome.

Example 3.21.
Suppose a circular disc was measured to have a radius of 30 inches, with a margin of error of $0.2in$. What is the maximum error in the calculation of the area of the disc?

Here we use different variables than before, but we know $A = \pi r^2$. The margin of error in our original measurement corresponds to dr (how far the actual value is from 30), and the error in the area of the disc is dA. We compute

$$dA = 2\pi r dr.$$

In our particular situation $r = 30$ and $dr = 0.2$ and so we have that the error in the area of the disc is

$$dA = 2\pi \cdot 30 \cdot 0.2 \approx 37.699$$

When we're doing a problem like this, sometimes it's helpful to have something to compare our answer to: 37.699 might seem like a large number, but how big is that in comparison to the area of the disc?

The true change in a quantity is sometimes called the ***absolute error***, and this corresponds to ΔA in the example above. Even though ΔA is relatively easy to calculate in this example, in "real world" situations that might not be the case, so it's important to know that we can approximate the absolute error ΔA with the differential dA.

In the example above, we've estimated that the absolute error in the radius of the disc is 37.699. (The true absolute error here is 37.8248.)

Again, the absolute error is not really a very helpful measurement: it's more helpful to look at the ***relative error***.

The true value of the relative error is $\frac{\Delta A}{A}$, but we can approximate the relative error as $\frac{dA}{A}$. In our example we have

$$\frac{dA}{A} = \frac{37.699}{\pi 30^2} = 0.01333...$$

The ***percentage error*** is thus 1.333%. So the error in the area of the disc resulting from a margin of error of 0.2 inches in the radius is pretty small: the area changes by about 1.33%.

Example 3.22.
By Ohm's law, if a current of I passes through a resistor with resistance R, then the voltage drop is $V = RI$. Suppose that V is constant and the resistance R is measured with a certain error. What is the relative error in calculating I?

Notice that $I = \frac{V}{R}$. We want to describe dI in terms of dR – dR is the error in our measurement of R, and dI is the corresponding (approximate) error in the I.

$$
\begin{aligned}
I &= \frac{V}{R} \\
\Longrightarrow dI &= \frac{d}{dR}\frac{V}{R}dR \\
\Longrightarrow dI &= -VR^{-2}dR
\end{aligned}
$$

The relative error is approximately $\frac{dI}{I}$ which is

$$\frac{-VR^{-2}dR}{VR^{-1}} = -\frac{dR}{R}.$$

And so the relative error in I is approximately the same as the (negative of) the relative error in R.

3.4 The Mean Value Theorem

I keep the subject constantly before me, and wait 'till the firrst dawnings open slowly, by little and little, into a full and clear light.

ISAAC NEWTON
Comment about how he made scientific discoveries, in Biographia Britannica.

In the next lecture we will discuss how the derivative of a function tells us information about the graph of the function. We will see, for example, that if a function is increasing, then its derivative must be positive; if a function is decreasing, its derivative must be negative. We will also discuss asymptotes and concavity of the graph. Before doing this we need one technical tool that is useful in many problems besides curve sketching.

In this lecture we will discuss Rolle's theorem and the mean value theorem, which is arguably the most important theorem in calculus.

Rolle's Theorem

Before stating the mean value theorem, we state a preliminary result.

Theorem 3.1 (Rolle's Theorem)**.**
Suppose that f is a function which is continuous on $[a, b]$, differentiable on (a, b), whose derivative is continuous on (a, b), and has the property that $f(a) = f(b)$. Then there exists a c between a and b with the property that $f'(c) = 0$.

Proof.
Suppose no such c existed: that is, $f'(x) \neq 0$ for any x in (a, b). Since $f'(x)$ is continuous, this means that $f'(x) > 0$ for all x in (a, b), or $f'(x) < 0$ for all x in (a, b).

If $f'(x) > 0$ for all x in (a, b), then we must have that

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} > 0.$$

which means in particular that

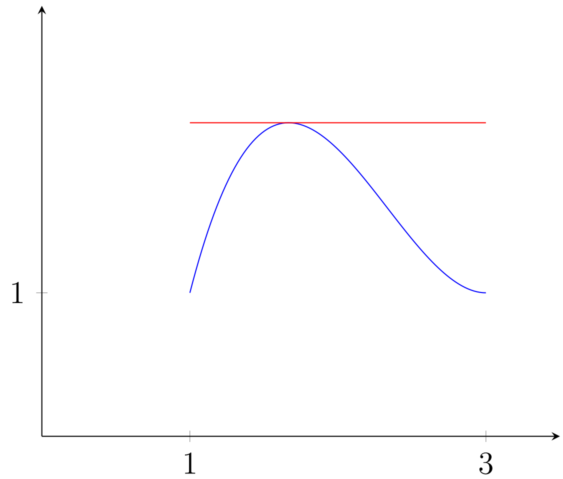
$$\lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h} > 0.$$

Since the h's appearing in the quotient above are positive (as we take the limit from the right), we must have that $f(x+h) - f(x) > 0$ for all (sufficiently small) h. That is $f(x+h) > f(x)$.

This must hold for all $x \in (a, b)$, but this implies that $f(x)$ is a strictly increasing function, and so $f(a) \neq f(b)$.

The case when $f'(x) < 0$ for all x is similar. □

Intuitively, Rolle's theorem says that if $f(x)$ "goes up" when you leave $x = a$, then it must come back down before you hit $x = b$.



The Mean Value Theorem

Rolle's theorem sounds almost obvious, and the mean value theorem we are about to state is essentially just a corollary of Rolle's theorem – i.e., it is an "almost obvious" result as well. Despite this, the mean value theorem is used throughout mathematics: many results in pure and applied mathematics require the mean value theorem, and so calling the mean value theorem the most important theorem in calculus is not unjustified.

Theorem 3.2.
Suppose that $f(x)$ is a function which is continuous on $[a, b]$, differentiable on (a, b), whose derivative is continuous. Then there exists a c between a and b such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof.
We want to just replace $f(x)$ with something where we can apply Rolle's theorem, so we need to bring $f(a)$ and $f(b)$ down to the same level.

Consider the function $g(x)$ defined by

$$g(x) = f(x) - f(a) - \frac{x - a}{b - a}(f(b) - f(a)).$$

Notice

$$\begin{aligned} g(a) =& f(a) - f(a) - \frac{a - a}{b - a}(f(b) - f(a)) = 0 \\ g(b) =& f(b) - f(a) - \frac{b - a}{b - a}(f(b) - f(a)) = 0. \end{aligned}$$

By Rolle's theorem there is some c between a and b with $g'(c) = 0$. But notice

$$g'(x) = f'(x) - \frac{1}{b - a}(f(b) - f(a)).$$

So there is some c making the following hold:

$$\begin{aligned} f'(c) - \frac{1}{b - a}(f(b) - f(a)) = 0 \\ \implies f'(c) = \frac{f(b) - f(a)}{b - a}. \end{aligned}$$

□

The mean value theorem only tells us the existence of such a c and does not tell us what a c is.

Despite this, the mean value theorem has some suprising applications. One application that we've already considered in class is velocity. If an object moving in a straight line starts and stops at the same point, by the mean value theorem there must be a place where the object's velocity is exactly zero.

Example 3.23.
If $f(0) = 3$ and $f'(x) \leq 4$ for all x. What is the largest possible value of $f(10)$?

Because of the mean value theorem, we know there must be a c between 0 and 10 such that

$$f'(c) = \frac{f(10) - f(0)}{10 - 0} = \frac{f(10) - 3}{10}.$$

However, we're told $f'(c) \leq 4$, and hence we know

$$\begin{aligned} &\frac{f(10) - 3}{10} \leq 4 \\ \implies &f(10) - 3 \leq 40 \\ \implies &f(10) \leq 43. \end{aligned}$$

3.5 Curve Sketching

The purpose of computation is insight, not numbers.

RICHARD HAMMING

In this lecture we discuss how to use knowledge about a function's first and second derivatives to describe the function's graph.

Increasing and Decreasing

We say that a function f is ***increasing*** on an interval (a, b) if for every $x_1 < x_2$ in (a, b), we have $f(x_1) \leq f(x_2)$. If in fact $f(x_1) < f(x_2)$ for every $x_1 < x_2$ in (a, b), then we say that f is ***strictly increasing***.

Similarly, we say that f is ***decreasing*** (respectively, ***strictly decreasing***) on (a, b) if for every $x_1 < x_2$ we have that $f(x_1) \geq f(x_2)$ (respectively, $f(x_1) > f(x_2)$.)

Graphically, being increasing means the function's graph goes "up" from left-to-right, and being decreasing means the function's graph goes "down."

These definitions do not rely on calculus at all, but we can still use calculus to help us determine when a function is increasing or decreasing thanks to the following theorems.

Theorem 3.3.
Suppose that f is differentiable on (a, b). If $f'(x) > 0$ for each x in (a, b), then f is strictly increasing.

Proof.
We need to show that for every $x_1 < x_2$ in (a, b), $f(x_1) < f(x_2)$. We simply apply the mean value theorem to f on the interval (x_1, x_2). By the mean value theorem there must exist a value of c between x_1

and x_2 with

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

But as $f'(c) > 0$ we have

$$\begin{aligned} & f'(c) > 0 \\ \implies & \frac{f(x_2) - f(x_1)}{x_2 - x_1} > 0 \\ \implies & f(x_2) - f(x_1) > 0 \\ \implies & f(x_2) > f(x_1). \end{aligned}$$

Thus f is strictly increasing. □

Theorem 3.4.
Suppose that f is differentiable on (a, b). If $f'(x) < 0$ for each x in (a, b), then f is strictly decreasing.

Exercise 3.1.
Prove Theorem 3.4.

So positive derivative means a function is increasing, negative derivative means a function is decreasing. What if a function's derivative is zero?

Theorem 3.5.
If $f'(x) = 0$ on (a, b), then $f(x)$ is a constant on (a, b).

Proof.
Pick any two $x_1 < x_2$ in (a, b). By the mean value theorem, there is some c in (x_1, x_2) such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c) = 0.$$

But this implies $f(x_2) = f(x_1)$. Since this is true for any pair of points we pick, the function must be constant. □

Example 3.24.
Where is the function $f(x) = -2x^3 + 15x^2 - 24x + 3$ increasing? Where is it decreasing?

We need to find where $f'(x)$ is positive and where it's negative.

$$f'(x) = -6x^2 + 30x - 24.$$

To figure out where this function is positive and negative, we first find where it's zero.

$$\begin{aligned} & f'(x) = 0 \\ \implies & -6x^2 + 30x - 24 = 0 \\ \implies & -x^2 + 5x - 4 = 0 \end{aligned}$$

This factors as $(1-x)(x-4)$ (equivalently, you could use the quadratic formula). Hence the derivative $f'(x)$ is zero when $x = 1$ and when $x = 4$.

So we have three intervals to consider: $(-\infty, 1)$, $(1, 4)$, and $(4, \infty)$. On each of these $f'(x)$ is either positive or negative, and so $f(x)$ is increasing or decreasing on each one. We can simply plug in a value in each interval to determine if the derivative is positive or negative.

For $(-\infty, 1)$, let's plug in $x = 0$ to get $f'(0) = -24$, so the derivative is negative on this interval.

For $(1, 4)$, we plug in $x = 2$ to get $f'(2) = -24 + 60 - 24 = 12$, so the derivative is positive here.

For $(4, \infty)$ let's plug in 5 to get $f'(5) = -24$, so the derivative is negative here.

Putting all of this together we have that $f(x)$ is increasing on the interval $(1, 4)$, but decreasing on $(-\infty 1) \cup (4, \infty)$.

Concavity

We say that a function f is ***concave down*** on the interval (a, b) if for each c in (a, b) there is a small interval I around c where

$$f(x_0) < f'(x_0)(x_0 - c) + f(c)$$

for each x_0 in I. That is, the graph of $f(x)$ lies below the tangent line at $x = c$.

We say that a function f is ***concave up*** on the interval (a, b) if for each c in (a, b) we have that

$$f(x) > f'(c)(x - c) + f(c)$$

for each x in (a, b). That is, the graph of $f(x)$ lies above the tangent line at $x = c$.

Graphs which are concave up are shaped like a bowl; concave down graphs are shaped like an umbrella.

Points where the concavity changes, where the graph goes from being concave up to being concave down (or vice versa), are called ***inflection points***.

Theorem 3.6.
If $f''(x) > 0$ on (a, b), then f is concave up on (a, b).

Proof.
We need to show that for each c in (a, b), the graph of $f(x)$ is lies above the tangent line at $x = c$. That is, we need to show that

$$f(x) > f'(c)(x - c) + f(c).$$

We consider two cases: first suppose that $x > c$ and apply the mean value theorem on the interval (c, x). Then there must exist a d such that

$$f'(d) = \frac{f(x) - f(c)}{x - c}.$$

and so $f(x) - f(c) = f'(d)(x - c)$, meaning $f(x) = f'(d)(x - c) + f(c)$. That is, we need to show

$$f'(d)(x - c) + f(c) > f'(c)(x - c) + f(c).$$

Notice that $d > c$ and that f' is increasing since f'' is positive. Thus $f'(d) > f'(c)$ and this proves that $f(x) > f'(c)(x - c) + f(c)$.

In the other case, suppose that $x < c$. We apply MVT on (x, c) to get that there must exist a $x < d < c$ with

$$f'(d) = \frac{f(x) - f(c)}{x - c}$$

and so $f(x) = f'(d)(x - c) + f(c)$, so we need to show

$$f'(d)(x - c) + f(c) > f'(c)(x - c) + f(c)$$

Equivalently,

$$f'(d)(c - x) < f'(c)(c - x).$$

But again, $f'' > 0$, so f' is increasing and $f'(c) > f'(d)$. □

Theorem 3.7.
If $f''(x) < 0$ on (a, b), then f is concave down on (a, b).

Proof.
Execise. □

Example 3.25.
Where is the function $f(x) = -2x^3 + 15x^2 - 24x + 3$ concave up? Concave down? Where are the inflection points?

We simply check to see where the second derivative is positive or negative. Notice $f''(x) = -12x + 30$. To find where this is positive and negative, we first find where it's zero: $x = \frac{5}{2}$. So we have two intervals to consider: $(-\infty, \frac{5}{2})$ and $(\frac{5}{2}, \infty)$. We again pick some point inside each of these intervals to evaluate our function (say $x = 0$ and $x = 10$) and we see that the second derivative is negative on $(-\infty, \frac{5}{2})$ and positive on $(\frac{5}{2}, \infty)$.

Thus, f is concave up on $(-\infty, \frac{5}{2})$, and concave down on $(\frac{5}{2}, \infty)$. The point $\frac{5}{2}$ is our inflection point.

It's worth noticing that points where $f''(x) = 0$ are *candidates* for inflection points and are not necessarily inflection points! An easy example is $f(x) = x^4$. Here $f''(x) = 12x^2$ which is zero at $x = 0$, but this isn't an inflection point because the concavity doesn't change!

Examples

Example 3.26.
Sketch the curve $y = \frac{2x^2-12}{x^2-9}$.

Domain: $(-\infty, -3) \cup (-3, 3) \cup (3, \infty)$. So we should expect to have vertical asymptotes at $x = \pm 3$, but we need to double-check this:

$$\begin{aligned}
\lim_{x\to -3^-} \frac{2x^2-8}{x^2-9} &= \lim_{x\to -3^-} \frac{2(x^2-4)}{x^2-9} \\
&= \lim_{x\to -3^-} \frac{2(x-2)(x+2)}{(x-3)(x+3)} \\
&= \lim_{x\to -3^-} \frac{2(x-2)(x+2)}{x-3} \cdot \lim_{x\to -3^-} \frac{1}{x+3} \\
&= \frac{10}{-6} \cdot \lim_{x\to -3^-} \frac{1}{x+3} \\
&= \infty
\end{aligned}$$

$$\begin{aligned}
\lim_{x\to -3^-} \frac{2x^2-8}{x^2-9} &= \lim_{x\to -3^-} \frac{2(x^2-4)}{x^2-9} \\
&= \frac{10}{-6} \cdot \lim_{x\to -3^+} \frac{1}{x+3} \\
&= -\infty
\end{aligned}$$

$$\begin{aligned}
\lim_{x\to 3^-} \frac{2x^2-8}{x^2-9} &= \lim_{x\to 3^-} \frac{2x^2-8}{x+3} \cdot \lim_{x\to 3^-} \frac{1}{x-3} \\
&= \frac{10}{6} \cdot \lim_{x\to 3^-} \frac{1}{x-3} \\
&= -\infty
\end{aligned}$$

$$\begin{aligned}
\lim_{x\to 3^+} \frac{2x^2-8}{x^2-9} &= \lim_{x\to 3+-} \frac{2x^2-8}{x+3} \cdot \lim_{x\to 3^+} \frac{1}{x-3} \\
&= \frac{10}{6} \cdot \lim_{x\to 3^+} \frac{1}{x-3} \\
&= \infty
\end{aligned}$$

Now we find the places where our function is increasing by seeing where the derivative is positive and where it's negative. First we find the "transition points" between increasing/decreasing: i.e., the places where the derivative is zero.

$$\frac{dy}{dx} = \frac{(x^2-9)(4x)-(2x^2-8)(2x)}{(x^2-9)^2} = 0$$
$$\implies 4x^3 - 36x - (4x^3 - 16x) = -20x = 0.$$

We need to be a little bit careful here. While $x = 0$ is a place where our function could transition from being increasing to being decreasing, we also have to consider the points $x = \pm 3$ as transition points as well.

That is, we have four intervals we need to consider:

$$(-\infty, -3),\ (-3, 0),\ (0, 3),\ (3, \infty).$$

We plug a point from each of these intervals into the first derivative.

At $x = -4$, $\frac{dy}{dx} = \frac{-20(-4)}{((-4)-9)^2} = \frac{80}{169} > 0$, so the function is increasing on $(-\infty, 3)$.

At $x = -1$, $\frac{dy}{dx} = \frac{-20(-1)}{((-1)-9)^2} = \frac{20}{100} > 0$, so the function is increasing on $(-3, 0)$.

At $x = 1$, $\frac{dy}{dx} = \frac{-20(1)}{(1-9)^2} = \frac{-20}{64} < 0$, so the function is decreasing on $(0, 3)$.

At $x = 4$, $\frac{dy}{dx} = \frac{-20(4)}{(4-9)^2} = \frac{-80}{25} < 0$, so the function is decreasing on $(3, \infty)$.

Now we have to consider the concavity.

$$\frac{d^2y}{dx^2} = \frac{(-20)(x^2-9)^2 - (-20x)(2(x^2-9)2x)}{(x^2-9)^4} = 0$$
$$\implies \frac{60(x^2+3)}{(x^2-9)^3} = 0$$

This equation has no solutions, so there are no inflection points.

We thus have three intervals whose concavity we need to check:

$$(-\infty, -3),\ (-3, 3),\ (3, \infty).$$

At $x = -4$,
$$\frac{d^2y}{dx^2} = \frac{60 \cdot 19}{(16-9)^3} = \frac{1140}{7^3} > 0.$$

So the function is concave up on $(-\infty, -3)$.

At $x = 0$,

$$\frac{d^2y}{dx^2} = \frac{180}{-9^3} < 0.$$

So the function is concave down on $(-3, 3)$.

At $x = 4$,

$$\frac{d^2y}{dx^2} = \frac{60 \cdot 19}{(16-9)^3} = \frac{1140}{7^3} > 0.$$

So the function is concave up on $(3, \infty)$.

3.6 Optimization

> *All such problems can be formulated as mathematical programming problems. Naturally, we can propose many sophisticated algorithms and a theory, but the final test of a theory is its capacity to solve the problems which originated it.*
>
> GEORGE DANTZIG

One of the most important applications of calculus to real-world problems concerns *optimization*. That is, many real-world problems often have multiple possible solutions, but some solutions are more desirable than others. Our goal in this chapter is to determine the "best" solution to a given problem. For us, "best" will usually mean that we want to find the input to a function which makes the function as large as it can possibly be, or as small as it can possibly be. For example, if our function represents the profit a company makes as a function of the number of items the company produces, we may want to determine what number of items we should make to maximize the profit. Conversely, if our function represents the cost to a company, we may want to minimize that cost.

The material in this chapter is necessarily just the starting point for a very active, and very applied, area of mathematics called *operations research* which seeks to develop methods for solving the complicated optimization problems that often arise in commercial and industrial settings. We will only be scratching the surface of how to solve optimization problems, but it is nice to know the material we are learning has practical, real-world applications.

Definitions and main ideas

We begin by defining precisely what we mean when we talk about maximizing or minimizing a function. Usually what we want to do is to find the value for the input to a function $f(x)$ that makes the output as large or as small as it can possibly be. The inputs where this occurs are called the global maxima and global minima of the function. To be precise, we say c is a ***global maximum*** (or ***absolute maximum***) of a function f if $f(c) \geq f(x)$ for all x in the domain of f. Similarly, we say c is a ***global***

minimum (or ***absolute minimum***) of a function f if $f(c) \leq f(x)$ for all x in the domain of f.

In much of what we will do, the process fo finding a global maximum or a global minimum is the same, and so to save ourselves some energy from saying "global maxima or global minima" all the time, we will simply refer to ***global extrema*** if what we are discussing applies to both the maxima and minima.

Though we wish to find global extrema, it will turn out that it is much easier to find something related called the ***local extrema***. These are points which may not make the output of the function the absolute largest or smallest it could be of all the values, but instead makes it the largest or smallest among all "nearby" values. As we will see soon, the tools of calculus will allow us to find local extrema relatively easily, and our goal will be to winnow the list of local extrema to down to global extrema. First, though, we need to carefully define what we mean by these local extrema.

We say c is a ***local maximum*** of f if there is a small neighborhood $(c-\delta, c+\delta)$ if $f(c) \geq f(x)$ for all x in $(c-\delta, c+\delta)$. Similarly, we say c is a ***local minimum*** of f if there is a small neighborhood $(c-\delta, c+\delta)$ if $f(c) \leq f(x)$ for all x in $(c-\delta, c+\delta)$.

Notice that global extrema are also local extrema: if a point c is a global maximum, for instance, $f(c) \geq f(x)$ for all x in the domain of the function, and so in particular $f(c) \geq f(x)$ for all x in any neighborhood around c.

Using calculus to find extrema

Our goal is to locate the maxima and minima of a given function. The following theorem is key: it gives us possible candidates for maxes and mins.

Theorem 3.8 (Fermat's Theorem)**.**
If c is a local extremum of f, then either $f'(c) = 0$ or f is not differentiable at c.

Proof.
Suppose that c is a local maximum (the case when c is a minimum is similar) and $f'(c)$ is defined. Suppose that $f(c) \geq f(x)$ for all x in $(c-\delta, c+\delta)$. Consider the limit definition of the derivative:

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}.$$

For all sufficiently small h, $f(c+h) \leq f(c)$. Hence each $f(c+h) - f(c) \leq 0$. Thus

$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0.$$

On the other hand,

$$\lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq 0.$$

Since we're assuming $f'(c)$ exists we have

$$0 \leq \lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0.$$

But the only way this can happen is if

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = 0.$$

□

This theorem tells us that in order to search for local extrema, we should find the places where the derivative of our function is zero or undefined. These places are candidates for local extrema (they are not guaranteed to be local extrema, however), and have a special name: we call points c where $f'(c)$ is zero or undefined ***critical points*** (or ***critical numbers***).

Example 3.27.
Find the critical points of $f(x) = \frac{x-1}{x^2-x+1}$.

We simply solve the equation $f'(x) = 0$:

$$\begin{aligned} f'(x) =& 0 \\ \implies \frac{(x^2 - x + 1) \cdot 1 - (x-1)(2x-1)}{(x^2 - x + 1)^2} =& 0 \\ \implies x^2 - x + 1 - (2x^2 - 3x + 1) =& 0 \\ \implies -x^2 + 2x =& 0 \\ \implies x(-x+2) =& 0 \\ \implies x = 0 \text{ or } x = 2.& \end{aligned}$$

We also have to worry about the places where our derivative is undefined: this occurs when the denominator $x^2 - x + 1$ is zero:

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

Notice this equation has no solutions since the parabola $y = x^2 + 1$ and the line $y = x$ have no points of intersection.

So the critical points of $f(x)$ are $x = 0$ and $x = 2$.

Notice that we've already discussed critical points this semester: they are precisely the places where the the function can transition from being increasing to decreasing. This motivates the next theorem for determining when a critical point is in fact a local extremum.

Theorem 3.9 (The first derivative test)**.**
Suppose that c is a critical point of f. If f is decreasing to the left of c and increasing to the right, then c is a local minimum. If f is increasing to the left of c and decreasing to the right, then c is a local maximum. In all other cases (where there's no transition from increasing-to-decreasing or vice versa), c is not a local extremum.

Example 3.28.

What are the local maxima and minima of $f(x) = \frac{x-1}{x^2-x+1}$?

We already have our candidates: $x = 0$ and $x = 2$. We now need to see if the function is increasing or decreasing on $(-\infty, 0)$, $(0, 2)$ and $(2, \infty)$.

At $x = -1$, $\frac{dy}{dx}$ is $\frac{2(-1)-(-1)^2}{((-1)^2-(-1)+1)^2} = \frac{-2-1}{(1+1+1)^2} < 0$, so the function is decreasing on $(-\infty, 0)$.

At $x = 1$, $\frac{dy}{dx}$ is $\frac{2-1}{(1-1+1)^2} > 0$, so the function is increasing on $(0, 2)$.

By the first derivative test, the point $x = 0$ is a local minimum of the function.

At $x = 3$, $\frac{dy}{dx}$ is $\frac{2(3)-3^2}{(3^2-3+1)^2} < 0$, so the function is decreasing on $(2, \infty)$.

Thus, by the first derivative test, the point $x = 2$ is a local maximum of the function.

We can also use concavity to help us determine if a critical point is a local maximum or a local minimum.

Theorem 3.10 (The second derivative test)**.**
Suppose that c is a critical point of f and that $f''(c)$ is defined. If $f''(c) > 0$, then c is a local minimum. If $f''(c) < 0$, then c is a local maximum. If $f''(c) = 0$, then the test is inconclusive (c could be a local max, a local min, or neither).

Example 3.29.
Use the second derivative test to find all of the local maxes and local mins of the function $f(x) = \frac{x^3}{6} - \frac{7}{4}x^2 + 5x + 4$.

First we have to find our critical points (candidates for maxes and mins) by solving the equation $f'(x) = 0$.

$$
\begin{aligned}
& f'(x) = 0 \\
\implies & \frac{1}{2}x^2 - \frac{7}{2}x + 5 = 0 \\
\implies & x^2 - 7x + 10 = 0 \\
\implies & (x-2)(x-5) = 0
\end{aligned}
$$

So our critical points are $x = 2$ and $x = 5$.

To use the second derivative test we need to determine if $f''(2)$ and $f''(5)$ are positive or negative. Notice

$$f''(x) = x - \frac{7}{2} = x - 3.5.$$

Thus $f''(2) = -1.5 < 0$ and $f''(5) = 1.5 > 0$, and $x = 2$ is a local max while $x = 5$ is a local min.

Global Extrema

We now know how to find local extrema. We also know that global extrema are local extrema. It would be reasonable to assume, then, that to find the global extrema we first find the local extrema, and then find which of the local maxes is a global max by plugging the local maxes into $f(x)$ and seeing which gives us the largest value. Similarly, to find the global mins we might like to take the local mins, plug them into the function and see which gives the smallest value.

The problem with this strategy is that our function may not have any global maxes or global mins. Consider the function $f(x) = \frac{2x^2-12}{x^2-9}$. This function has only one local extremum: $x = 0$ is a local max. However the function has no global extrema!

So the strategy above doesn't work in general, but there is one situation when this naïve strategy does work: if our function is continuous and is defined on a closed interval.

Theorem 3.11 (Extreme value theorem)**.**
If f is defined and continuous on the closed, bounded interval $[a, b]$, then f must have a global maximum and a global minimum in $[a, b]$.

Most of the theorems in this class we've proved, or simply omitted for the sake of time. This theorem, however, is considerably more difficult to prove than any other theorem we've seen in this class and for that reason we will not attempt to prove it.

The extreme value theorem basically says that we can find the global maxes and mins of a function if it's continuous on a closed, bounded interval by taking the local extrema which give us the largest and smallest possible outputs of the function. The one thing we need to keep in mind is that the endpoints $x = a$ and $x = b$ could also be extrema.

In short, we can find the global extrema of a continuous function f defined on a closed interval $[a, b]$ by doing the following:

1. Find the critical points in $[a, b]$ – i.e., the values of x in $[a, b]$ where $f'(x) = 0$ or $f'(x)$ DNE.
2. Determine which critical points are local maxes and local mins by using the first or second derivative test.
3. Determine which local max gives the largest output of $f(x)$, and which local min gives the smallest output of $f(x)$.
4. Plug $x = a$ and $x = b$ into the function as well.
5. The largest local max, or $f(a)$ or $f(b)$ is the global max. The smallest is the local min.

Example 3.30.
Find the global maximum and minimum of the function $f(x) = \sqrt[3]{x}(8 - x)$ on the interval $[0, 8]$.

Example 3.31.
Find the global maximum and minimum of the function $f(x) = x^3 - 6x^2 - 36x + 41$ on the interval $[-4, 10]$.

3.7 L'Hôpital's Rule

Mathematics is no more computation than typing is literature.

JOHN ALLEN PAULOS

Statement

We end our discussion of derivatives by mentioning *l'Hôpital's rule* which allows us to calculate certain special types of limits.

Theorem 3.12 (L'Hôpital's Rule).
Suppose that f and g are differentiable functions satisfying one of the following conditions:

1. $\lim_{x\to a} f(x) = 0$ *and* $\lim_{x\to a} g(x) = 0$, *or*
2. $\lim_{x\to a} f(x) = \pm\infty$ *and* $\lim_{x\to a} g(x) = \pm\infty$.

Then

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

That is, l'Hôpital's rule says that if we want to take the limit of a fraction which has the indeterminate form $\frac{\infty}{\infty}$ or $\frac{0}{0}$, then we can replace the fraction withe fraction of the derivatives of the numerator and denominator. This trick also works if we're taking a left- or right-hand limit, or limits as x goes to $\pm\infty$.

Simple examples

Example 3.32.

Use l'Hôpital's rule to calculate the following limit:

$$\lim_{x \to 3} \frac{x-3}{x^2-9}.$$

$$\lim_{x \to 3} \frac{x-3}{x^2-9} = \lim_{x \to 3} \frac{1}{2x}$$
$$= \frac{1}{6}$$

Example 3.33.
Use l'Hôpital's rule to calculate the following limit

$$\lim_{x \to 0} \frac{\tan(3x)}{\sin(2x)}$$

$$\lim_{x \to 0} \frac{\tan(3x)}{\sin(2x)} = \lim_{x \to 0} \frac{3\sec^2(3x)}{2\cos(2x)}$$
$$= \frac{3}{2}$$

Example 3.34.
Use l'Hôpital's rule to calculate the following limit

$$\lim_{x \to \infty} \frac{x+x^2}{1-2x^2}$$

$$\begin{aligned}\lim_{x\to\infty}\frac{x+x^2}{1-2x^2} &= \lim_{x\to\infty}\frac{1+2x}{-4x}\\ &= \lim_{x\to\infty}\frac{2}{-4}\\ &= -\frac{1}{2}\end{aligned}$$

Notice in the last example we actually used l'Hôpital's rule twice in the process of calculating the limit as one application still gave us a $\frac{\infty}{\infty}$ indeterminate form.

Product examples

We can also use l'Hôpital's rule to evaluate limits of products by turning the product into a quotient.

Example 3.35.
Calculate the limit

$$\lim_{x\to\infty} x^3 e^{-x^2}.$$

The "trick" here is to rewrite e^{-x^2} as $\frac{1}{e^{x^2}}$ and then apply l'Hôpital's rule.

$$\begin{aligned}\lim_{x\to\infty} x^3 e^{-x^2} &= \lim_{x\to\infty}\frac{x^3}{e^{x^2}}\\ &= \lim_{x\to\infty}\frac{3x^2}{2xe^{x^2}}\\ &= \lim_{x\to\infty}\frac{3x}{2e^{x^2}}\\ &= \lim_{x\to\infty}\frac{3}{4xe^{x^2}}\\ &= 0\end{aligned}$$

Example 3.36.

Calculate the limit

$$\lim_{x\to 0^+} x\ln(x).$$

We turn multiplication by x into division by $1/x$ so that we can apply l'Hôpital's rule:

$$\begin{aligned}
\lim_{x\to 0^+} x\ln(x) &= \lim_{x\to 0^+} \frac{\ln(x)}{1/x} \\
&= \lim_{x\to 0^+} \frac{\ln(x)}{x^{-1}} \\
&= \lim_{x\to 0^+} \frac{1/x}{-x^{-2}} \\
&= \lim_{x\to 0^+} \frac{-x^2}{x} \\
&= \lim_{x\to 0^+} -x \\
&= 0
\end{aligned}$$

It is not uncommon that we have to apply l'Hôpital's rule multiple times in calculating a limit, but sometimes we can run into an "infinite loop" of taking l'Hôpital's infinitely-many times. For example, if we tried to evaluate the limit

$$\lim_{x\to -\infty} xe^x$$

by rewriting the limit as

$$\lim_{x\to -\infty} \frac{e^x}{x^{-1}}$$

we would have the following:

$$\begin{aligned}
\lim_{x\to -\infty} xe^x &= \lim_{x\to -\infty} \frac{e^x}{x^{-1}} \\
&= \lim_{x\to -\infty} \frac{e^x}{-x^{-2}} \\
&= \lim_{x\to -\infty} \frac{e^x}{2x^{-3}} \\
&= \lim_{x\to -\infty} \frac{e^x}{-6x^{-4}} \\
&= \lim_{x\to -\infty} \frac{e^x}{24x^{-5}} \\
&\vdots
\end{aligned}$$

This process would never stop. Notice that we made a choice above to rewrite xe^x as ${}^{e^x}/_{x^{-1}}$. If we make the opposite choice, however, and write xe^x as ${}^{x}/_{e^{-x}}$ we have the following:

$$\begin{aligned}
&\lim_{x\to-\infty} xe^x \\
&\qquad = \lim_{x\to-\infty} \frac{x}{e^{-x}} \\
&\qquad = \lim_{x\to-\infty} \frac{1}{-e^{-x}} \\
&\qquad = -\infty
\end{aligned}$$

Difference examples

We can also sometimes rewrite differences as quotients to apply l'Hôpital's.

Example 3.37.
Calculate the limit

$$\lim_{x\to1^+} \left(\ln\left(x^2-1\right) - \ln\left(x^3-1\right)\right)$$

To turn this into a fraction we can use properties of logarithms, and then use the fact that logarithms are continuous to bring the limit inside of the log.

$$\begin{aligned}
\lim_{x\to1^+} \left(\ln\left(x^2-1\right) - \ln\left(x^3-1\right)\right) &= \lim_{x\to1^+} \ln\left(\frac{x^2-1}{x^3-1}\right) \\
&= \ln\left(\lim_{x\to1^+} \frac{x^2-1}{x^3-1}\right) \\
&= \ln\left(\lim_{x\to1^+} \frac{2x}{3x^2}\right) \\
&= \ln\left(\lim_{x\to1^+} \frac{2}{3x}\right) \\
&= \ln\left(\frac{2}{3}\right)
\end{aligned}$$

Example 3.38.
Calculate the limit

$$\lim_{x\to 0^+}\left(\frac{1}{x}-\frac{1}{e^x-1}\right)$$

To turn this difference into a quotient, we'll combine the fractions with a common denominator.

$$\begin{aligned}\lim_{x\to 0^+}\left(\frac{1}{x}-\frac{1}{e^x-1}\right) &= \lim_{x\to 0^+}\frac{e^x-1-x}{x(e^x-1)}\\ &= \lim_{x\to 0^+}\frac{e^x-1}{e^x+xe^x-1}\\ &= \lim_{x\to 0^+}\frac{e^x}{e^x+e^x+xe^x}\\ &= \frac{1}{2}\end{aligned}$$

Example 3.39.
Calculate the limit

$$\lim_{x\to\infty}(x-\ln(x))$$

This time we'll turn the difference into a product by factoring out an x, and then take the limit of each factor, which will require l'Hôpital's rule.

$$\begin{aligned}\lim_{x\to\infty}(x-\ln(x)) &= \lim_{x\to\infty} x\left(1-\frac{\ln(x)}{x}\right)\\ &= \left(\lim_{x\to\infty} x\right)\cdot\lim_{x\to\infty}\left(1-\frac{\ln(x)}{x}\right)\\ &= \left(\lim_{x\to\infty} x\right)\cdot\left(1-\lim_{x\to\infty}\frac{\ln(x)}{x}\right)\\ &= \infty\cdot\left(1-\lim_{x\to\infty}\frac{1}{x}\right)\\ &= \infty\cdot 1\\ &= \infty\end{aligned}$$

Exponential example

We can even use l'Hôpital's rule to help us evaluate indeterminate powers, like 0^0. Here we require that since $\ln(x)$ and e^x are inverses, we write any function $f(x)$ as $e^{\ln(f(x))}$. This is helpful because we can then use properties of logs to pull down any exponents.

Example 3.40.
Calculate the limit

$$\lim_{x\to 0^+} x^{\sqrt{x}}$$

$$\begin{aligned}\lim_{x\to 0^+} x^{\sqrt{x}} &= \lim_{x\to 0^+} e^{\ln(x^{\sqrt{x}})} \\ &= \lim_{x\to 0^+} e^{\sqrt{x}\ln(x)} \\ &= e^{\lim_{x\to 0^+}\sqrt{x}\ln(x)}\end{aligned}$$

Now we compute the limit in the exponent,

$$\begin{aligned}\lim_{x\to 0^+} \sqrt{x}\ln(x) &= \lim_{x\to 0^+} \frac{\ln(x)}{1/\sqrt{x}} \\ &= \lim_{x\to 0^+} \frac{\ln(x)}{x^{-1/2}} \\ &= \lim_{x\to 0^+} \frac{1/x}{-1/2x^{-3/2}} \\ &= \lim_{x\to 0^+} -2x^{-sfrac12} \\ &= 0\end{aligned}$$

Plugging this back into the above we have

$$\lim_{x\to 0^+} x^{\sqrt{x}} = e^{\lim_{x\to 0^+}\sqrt{x}\ln(x)} = e^0 = 1.$$

Integration 4

4.1 Antiderivatives

The study of mathematics, like the Nile, begins in minuteness but ends in magnificence.

CHARLES COLTON

Motivation

We have seen that velocity is the derivative of position, and acceleration is the derivative of velocity. So if we know position, it's now relatively easy for us to calculate velocity, and once we know velocity it's easy to calculate acceleration.

But what if we want to go in the other direction? What if we knew acceleration and wanted to find position? This is not a far-fetched scenario. A consequence of Newton's theory of gravity is that the acceleration due to gravity (near the surface of the Earth) is constant: regardless of whether you drop a golf ball or a bowling ball from the top of Wait Chapel, its acceleration towards the ground will be -32 feet per second squared (or -9.8 metres per second squared).

Given this and the height of Wait Chapel, we can determine the ball's height above the ground at any moment in time. The height of Wait chapel is 213 feet. We know

$$a(t) = -32.$$

So the velocity, $v(t)$, must be a function with the property that $v'(t) = a(t)$. Since $a(t) = -32$, this means that $v(t)$ must have the form $v(t) = -32t + C$ where C is some constant. What should this constant be? Notice that $v(0) = C$. If we drop the ball (as opposed to throwing it downward), then 0 seconds after dropping it, its velocity is zero. So $C = 0$.

Now that we know $v(t) = -32t$ we can find position $s(t)$. We know that $s'(t) = v(t) = -32t$, which means $s(t) = -16t^2 + D$ where again D is a constant. Noticing $s(0) = D$, D must be the height of the ball when we first drop it. If we're dropping it from the rop of Wait Chapel, which is 213 feet tall, then we must have $D = 213$.

Thus, just knowing the acceleration due to gravity, and the height of Wait Chapel, we are able to determine that a ball dropped from Wait Chapel will have a height of

$$s(t) = -16t^2 + 213$$

feet above the ground, t seconds after being dropped.

Antiderivatives

In general, if $F(x)$ is a function whose derivative is little $f(x)$, then we call $F(x)$ an ***antiderivative*** of $f(x)$.

Example 4.1.

1. If $f(x) = 16x + 6 + \cos(x)$, then $F(x) = 8x^2 + 6x + \sin(x)$ is an antiderivative of $f(x)$.
2. If $f(x) = \frac{1}{2\sqrt{x}} - \frac{1}{\sqrt[3]{x^2}}$, then $F(x) = \sqrt{x} + \sqrt[3]{-x}$ is an antiderivative of $f(x)$.
3. If $f(x) = \frac{-x^2-1}{x^4-2x^2+1}$, then $F(x) = \frac{x}{x^2-1}$ is an antiderivative of $f(x)$.

If you are given a function $F(x)$ that you think is an antiderivative of $f(x)$, it's very easy to check: just see if $F'(x) = f(x)$. If we're given $f(x)$ and need to determine an antiderivative $F(x)$, however, that's usually more difficult. Before we see how to do this, let's make an observation about antiderivatives: if a function has an antiderivative (not all functions do!), then it has infinitely-many. That is, if $f(x)$ has antiderivative $F(x)$, then $F(x) + 1$ and $F(x) - 17$ and $F(x) + 32874832^{\pi}$ are all also antiderivatives of $f(x)$. In fact, all of the antiderivatives have to have this form.

First we need one quick result:

Theorem 4.1.
If $f(x) = 0$, then the only antiderivatives of $f(x)$ are the constant functions.

Proof.
Suppose $F(x)$ is an antiderivative of $f(x)$. By the mean value theorem, for any interval $[a, b]$ there exists a c inbetween such that

$$F'(c) = f(c) = \frac{F(b) - F(a)}{b - a}.$$

But we're assuming $f(c) = 0$, so

$$\frac{F(b) - F(a)}{b - a} = 0 \implies F(b) - F(a) = 0 \implies F(b) = F(a).$$

This is true for all a and all b, so F must be constant. □

Theorem 4.2.
If $F(x)$ and $G(x)$ are two antiderivatives of $f(x)$, then $G(x) = F(x) + C$ for some constant C.

Proof.
Notice that $\frac{d}{dx}(F(x) - G(x)) = f(x) - f(x) = 0$. By the above, this means the only thing whose derivative is that $F(x) - G(x)$ – since its an antiderivative of $\frac{d}{dx}(F(x) - G(x)) = 0$ – must be constant:

$$F(x) - G(x) = C \implies F(x) = G(x) + C$$

□

What this tells us is that the ***general antiderivative*** of a function $f(x)$ has the form $F(x) + C$ where $F'(x) = f(x)$ and C is any constant. We usually call the $F(x)$ a ***particular antiderivative***. In some applications we may want a particular antiderivative, and sometimes we may want a general antiderivative: it just depends on the problem.

Rules for Calculating Antiderivatives

Just as we have rules for calculating derivatives, we have some rules for calculating antiderivatives.

Theorem 4.3.
If k is a constant and $F(x)$ is an antiderivative of $f(x)$, then the general antiderivative of $kf(x)$ is $kF(x) + C$.

Proof.

$$\frac{d}{dx}(kF(x) + C) = kF'(x) + 0 = kf(x).$$

□

Theorem 4.4.
If $F(x)$ is an antiderivative of $f(x)$ and $G(x)$ is an antiderivative of $g(x)$, then $F(x) \pm G(x) + C$ is the general antiderivative of $f(x) \pm g(x)$.

Theorem 4.5.
If $f(x) = x^n$ and $n \neq -1$, then the antiderivative of $f(x)$ is $F(x) =$

$\frac{x^{n+1}}{n+1} + C$.

We're now able to find antiderivatives of any polynomial:

Example 4.2.
Find the general antiderivative of $6x^4 + x^3 - 7x^2 + 5$.

Notice the antiderivative of $6x^4$ is $\frac{6x^5}{5}$; the antiderivative of x^3 is $\frac{x^4}{4}$; the antiderivative of $-7x^2$ is $\frac{-7x^3}{3}$; and the antiderivative of 5 is $5x$. Putting all of this together, the antiderivative of $6x^4+x^3-7x^2+5$ is

$$\frac{6}{5}x^5 + \frac{1}{4}x^4 - \frac{7}{3}x^3 + 5x + C.$$

We can also find antiderivatives of some, *but not all* rational functions:

Example 4.3.
Find the antiderivative of

$$\frac{6x^7 - 3x^4 + 2}{x^3}.$$

Notice that

$$\frac{6x^7 - 3x^4 + 2}{x^3} = \frac{6x^7}{x^3} - \frac{3x^4}{x^3} + \frac{2}{x^3} = 6x^4 - 3x + 2x^{-3}.$$

We know the antiderivative of $6x^4$ is $\frac{6}{5}x^5$; the antiderivative of $3x$ is $\frac{3}{2}x^2$; and the antiderivative of $2x^{-3}$ is $-x^{-2}$; hence the antiderivative of $\frac{6x^7-3x^4+2}{x^3}$ is

$$\frac{6}{5}x^5 - 2x^{-3} - x^{-2}.$$

We can also easily determine the antiderivatives of some trig functions, applying our rules for trig derivatives "in reverse."

Theorem 4.6.

- *The antiderivative of* $\cos(x)$ *is* $\sin(x) + C$.
- *The antiderivative of* $\sin(x)$ *is* $-\cos(x) + C$.
- *The antiderivative of* $\sec^2(x)$ *is* $\tan(x) + C$.
- *The antiderivative of* $\sec(x)\tan(x)$ *is* $\sec(x) + C$.
- *The antiderivative of* $\csc(x)\cot(x)$ *is* $-\csc(x) + C$.
- *The antiderivative of* $\csc^2(x)$ *is* $-\cot(x) + C$.

Example 4.4.
Find the general antiderivative of $3\sec(x) \cdot (\sec(x) - 2\tan(x))$.
Notice that we can distribute the $3\sec(x)$ to get

$$3\sec(x) \cdot (\sec(x) - 2\tan(x)) = 3\sec^2(x) - 6\sec(x)\tan(x).$$

As we know the antiderivative of $\sec^2(x)$ is $\tan(x)$, and the antiderivative of $\sec(x)\tan(x)$ is $\sec(x)$, we have that the antiderivative of $3\sec(x) \cdot (\sec(x) - 2\tan(x))$ is

$$3\tan(x) - 6\sec(x).$$

4.2 Riemann Sums

In mathematics the art of proposing a question must be held of higher value than solving it.

GEORG CANTOR

Motivating Examples

Suppose that you ride a bicycle down a straight path at speed of 10 feet per second for ten seconds. How far did you ride the bike? Of course we know that you rode the bike for 100 feet; if you rode the bike for 30 seconds, then you instead rode a total of 300 feet. We find this distance travelled by multiplying the time we travel with the speed at which we are travelling.

That is, if speed is constant, distance is speed times time.

$$\text{Distance} = \text{Speed} \times \text{Time}.$$

But suppose you were not travelling at a constant speed: maybe you started riding your bike slowly and gradually accelerated. Then how would we find the distance travelled?

For example, let's say that you again ride your bike for 10 seconds, but t seconds after you start riding your speed is $\frac{t^2}{4}$ feet per second. So after one second your speed is $\frac{1}{4}$-foot per second, but after four seconds your speed is $\frac{25}{4} = 6.25$ feet per second, and by the time you get to 10 seconds, you're traveling at a speed of 25 feet per second.

Given that we know how to calculate the distance traveled when the speed is constant, how can we determine the distance traveled when the speed is not constant. This seems like it might be a hard problem.

To answer this problem we do what we always do in calculus: if something seems hard, estimate it with somthing simpler. So let's suppose that instead of having a continuously changing speed, let's suppose our speed was constant over little chunks of time.

For example, just to get an estimate, imagine we break up our 10 second bike ride into four pieces of 2.5 seconds each, and maybe on each of those little intervals of time we use the fastest speed we actually achieve during that time.

Time Interval	**Fastest Speed**
$[0, 2.5]$	$\frac{2.5^2}{4} = \frac{6.25}{4} = 1.5625$
$[2.5, 5]$	$\frac{5^2}{4} = 6.25$
$[5, 7.5]$	$\frac{7.5^2}{4} = 14.0625$
$[7.5, 10]$	25

If our speed was constant on each of these intervals, then we can easily determine how far we traveled during those intervals:

Time Interval	**Distance Traveled**
$[0, 2.5]$	3.90625
$[2.5, 5]$	15.625
$[5, 7.5]$	35.15625
$[7.5, 10]$	62.5

Adding these together, we'd might estimate that we travelled a total of

$$3.90625 + 15.625 + 35.15625 + 62.5 = 117.1875$$

feet during those ten seconds.

Notice that this estimate is definitely an upper estimate of the true distance traveled since we used the highest speed attained. If we instead used the lowest speed attained we'd have a lower estimate.

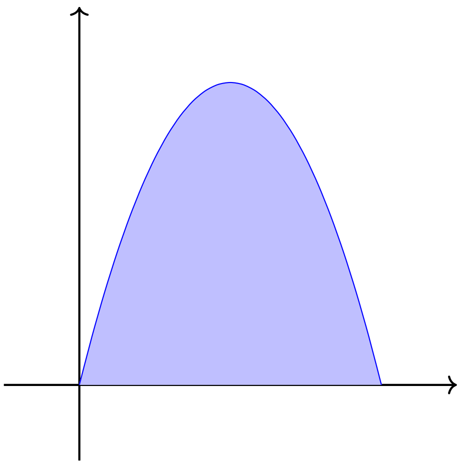
Time Interval	**Slowest Speed**	**Distance Travelled**
$[0, 2.5]$	0	0
$[2.5, 5]$	1.5625	3.90625
$[5, 7.5]$	6.25	35.15625
$[7.5, 10]$	14.0625	62.5

Adding these together, we get that

$$0 + 3.90625 + 15.625 + 35.15625 = 54.6875$$

is a lower estimatefor the distance traveled.

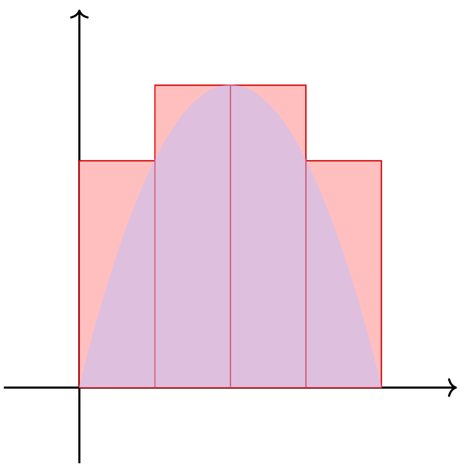
Now consider another problem which may look different at first, but is in fact very similar. Suppose we wanted to determine the area that lives below the parabola $y = -x^2 + 4x$ but above the x-axis between $[0, 4]$.



This looks like it might be a hard problem: we know how to calculate the area of some simple objects like rectangles, but how do we calculate the area under a parabola. Again, we do what we always do in calculus: if a problem seems hard, estimate it with something simpler. In this case what we'll do is estimate the area by covering the area we care about with rectangles.

Imagine that we take four rectangles whose bases are on the x-axis, and whose height is the largest value of $-x^2 + 4x$ for x-values in the base of the rectangle.

Rectangle's base	**Height**
$[0, 1]$	3
$[1, 2]$	4
$[2, 3]$	4
$[3, 4]$	3

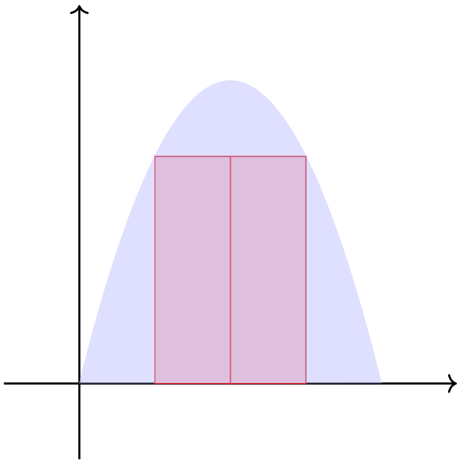


Notice that we can easily determine the area of each rectangle, and sum them up to get an (over) estimate of the area underneath the parabola:

$$1 \cdot 3 + 1 \cdot 4 + 1 \cdot 4 + 1 \cdot 3 = 14.$$

We could also have use the lowest point on the curve $y = -x^2 + 4x$ as the height of our rectangle to get a lower estimate:

Rectangle's base	**Height**
$[0, 1]$	0
$[1, 2]$	3
$[2, 3]$	3
$[3, 4]$	0



Adding the areas of the rectangles we get the lower estimate:

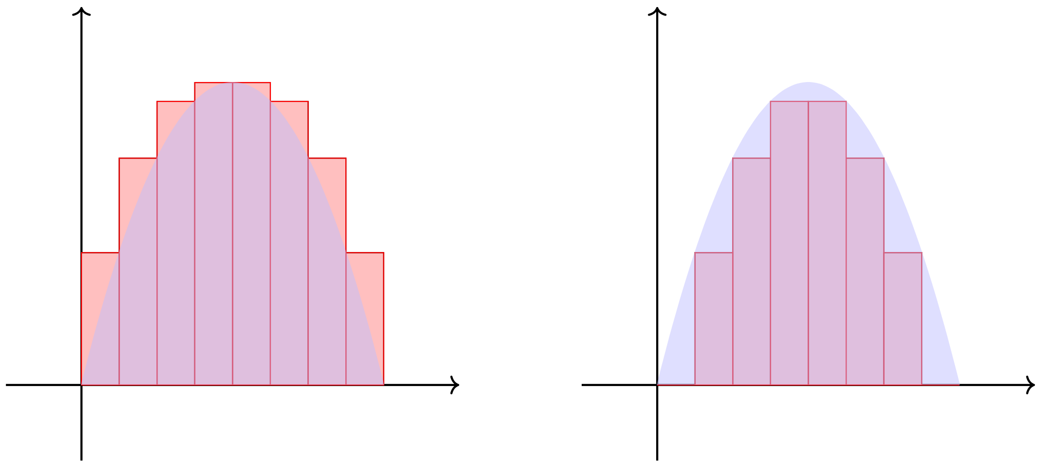
$$0 + 3 + 3 + 0 = 6.$$

So the true area is somewhere between 6 and 14.

It should be clear that we can improve our estimates, both the lower and upper estimates by using more, skinnier rectangles.

If we used eight rectangles instead of four, we'd have

Rectangle's base	**Highest Height**	**Lowest Height**
$[0, 1/2]$	1.75	0
$[1/2, 1]$	3	1.75
$[1, 3/2]$	3.75	3
$[3/2, 2]$	4	3.75
$[2, 5/2]$	4	3.75
$[5/2, 3]$	3.75	3
$[6, 7/2]$	3	1.75
$[7/2, 8]$	1.75	0



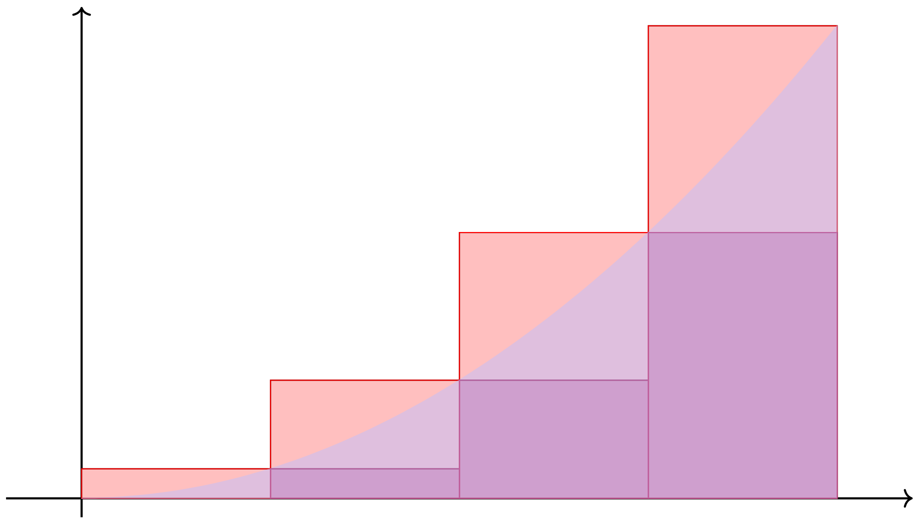
We can then find the upper and lower estimates of the of the area under the parabola to be

$$\begin{aligned}\textbf{Upper Estimate} =& \frac{1}{2}\cdot 1.75+\frac{1}{2}\cdot 3+\frac{1}{2}\cdot 3.75+\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 3.75+\frac{1}{2}\cdot 3+\frac{1}{2}\cdot 1.75\\ =&12.25\end{aligned}$$

$$\begin{aligned}\textbf{Lower Estimate} =& \frac{1}{2}\cdot 0+\frac{1}{2}\cdot 1.75+\frac{1}{2}\cdot 3+\frac{1}{2}\cdot 3.75+\frac{1}{2}\cdot 3.75+\frac{1}{2}\cdot 3+\frac{1}{2}\cdot 1.75+\frac{1}{2}\cdot 0\\ =&8.5\end{aligned}$$

We now know that the true area is between 8.5 and 12.25. Continuing to find upper and lower estimates like this we can "zoom in" on the true area.

Notice that this example of estimating areas is exactly the same as estimating distance in our first example. In fact, if you were to graph the speed function $v(t) = \frac{t^2}{4}$, then what we were doing before could be interepted as estimating the area between $y = \frac{t^2}{4}$ and the portion of the x-axis in the interval $[0, 10]$ by rectangles.



We can in fact do this sort of calculation for any (continuous) function $f(x)$, and the interpretation of the numbers we calculate changes depending on what our function is supposed to represent. So it's important to realize the process we are about to describe has many different interpretations and you shouldn't get hung up on any one particular one (e.g., area under a curve).

But based on what we've just done, we aren't too surprised by the following:

Theorem 4.7.
If $f(x)$ is a continuous, positive function on the interval $[a, b]$, then the area under the curve $y = f(x)$ and above the x-axis is the limit of the sums of the areas of rectangles which approximate the curve, as the width of the areas becomes arbitrarily small:

$$\textit{Area} = \lim_{n\to\infty} \left(f(x_1)\Delta x_1 + f(x_2)\Delta x_2 + \cdots + f(x_n)\Delta x_n\right).$$

Here $f(x_i)\Delta x$ is the area of the i-th rectangle: $f(x_i)$ is the height which we get by evaluating f at some point, for example the right-most point in each interval, inside the base, and Δx_n is the length of the base.

Theorem 4.8.
If $v(t)$ is a continuous, positive function on the interval $[a,b]$, then the distance travelled by a particle with velocity $v(t)$ at time t from $t=a$ to $t=b$ is the limit of distances obtained by assuming the velocity is constant on intervals that become arbitrarily small:

$$\textit{Distance} = \lim_{n\to\infty} \left(v(t_1)\Delta t_1 + v(t_2)\Delta t_2 + \cdots + v(t_n)\Delta t_n\right).$$

In the case of our example where $f(x) = -x^2+4x$ over the interval $[0,4]$, we cut the interval $[0,4]$ up into n intervals of equal width. The width of each interval is $\frac{4}{n}$, so each Δx_i is $\frac{4}{n}$. Notice that the interval $[0,4]$ gets cut up into the following subintervals:

$$[0, {}^4\!/\!_n],\ [{}^4\!/\!_n, {}^8\!/\!_n],\ [{}^8\!/\!_n, {}^{12}\!/\!_n],\ \cdots [4-\frac{4}{n}, 4].$$

The right-hand endpoint of the i-th interval is thus

$$x_i = \frac{4i}{n},$$

and the left-hand endpoint is

$$x_{i-1} = \frac{4(i-1)}{n}.$$

This means the area under the parabola is

$$\lim_{n\to\infty}\left(f\left(\frac{4}{n}\right)\frac{4}{n} + f\left(\frac{8}{n}\right)\frac{4}{n} + f\left(\frac{12}{n}\right)\frac{4}{n} + \cdots + f(4)\frac{4}{n}\right)$$

where $f(x) = -x^2+4x$.

Sigma Notation

We will be dealing with summations a lot in the next few lectures, and so now is a good time to introduce a useful type of notation.

Suppose that f is a function and we want to sum the function evaluated at several different integers. For example, we may want to evaluate

$$f(1)+f(2)+f(3)+\cdots+f(50).$$

We can use the following ***sigma notation*** to express this sum. We write a large capital Greek Σ, and below the Σ we say where we want the sum to start, above the Σ we say where we want the sum to stop, and to the right of Σ we say what we're summing up.

For example,

$$\sum_{i=1}^{50} f(i) = f(1) + f(2) + \cdots f(3)$$
$$\sum_{i=-3}^{7} f(i) = f(-3) + f(-2) + f(-1) + f(0) + \cdots + f(7)$$

This is simply a convenient way to write out certain sums.

Sometimes we the function we're trying to sum up won't have a given name: we may simply give an expression in place of $f(i)$:

$$\sum_{i=0}^{5} i^2 = 0^2 + 1^2 + 2^2 + 3^2 + 4^2 + 5^2$$
$$\sum_{i=-2}^{2} (2i-1) = (2\cdot(-2)-1)) + (2\cdot(-1)-1) + (2\cdot 0 - 1) + (2\cdot 1 - 1) + (2\cdot 2 - 1) +$$

Example 4.5.
The area under the curve $y = -x^2 + 4x$, over $[0,4]$ can be written as

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left(-\left(\frac{4i}{n}\right)^2 + 4\frac{4i}{n} \right) \frac{4}{n}$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} \left(-\frac{64i^2}{n^3} + \frac{64i}{n^2} \right)$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} 64 \left(\frac{i}{n^2} - \frac{i^2}{n^3} \right)$$

Notice that we're chopping $[0,4]$ into n intervals of equal length (the length is $\frac{4}{n}$), and the height of the i-th interval. We're getting the height of i-th rectangle by plugging the right-hand endpoint of the i-th interval,

$$\left[\frac{4(i-1)}{n}, \frac{4i}{n} \right]$$

into the function $f(x) = -x^2 + 4x$.

Example 4.6.
The distance travelled by a particle with velocity $v(t) = \frac{t^2}{4}$ over $[0, 10]$ may be written as

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\left({}^{10i}/_{n}\right)^2}{4} \cdot \frac{10}{n} = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{250i^2}{n^3}$$

Properties of Sums

Let's go ahead and notice a few simple properties of sums written in this Σ-notation:

Theorem 4.9.
For any two functions f and g, and for any constant k we have

$$\sum_{i=a}^{b} (f(i) + g(i)) = \sum_{i=a}^{b} f(i) + \sum_{i=a}^{b} g(i)$$
$$\sum_{i=a}^{b} kf(i) = k \sum_{i=a}^{b} f(i)$$

Proof.

$$\begin{aligned}\sum_{i=a}^{b}(f(i)+g(i)) =& f(a)+g(a)+f(a+1)+g(a+1)+\cdots+f(b)+g(b)\\ =& f(a)+f(a+1)+\cdots f(b)+g(a)+g(a+1)+\cdots g(b)\\ =& \sum_{i=a}^{b} f(i)+\sum_{i=a}^{b} g(i)\end{aligned}$$

$$\begin{aligned}\sum_{i=a}^{b} kf(i) =& kf(a)+kf(a+1)+\cdots+kf(b)\\ =& k\left(f(a)+f(a+1)+\cdots+f(b)\right)\\ =& k\sum_{i=a}^{b} f(a).\end{aligned}$$

□

The "constant" k in the above simply can not depend on i, but it could depend on some other quantity. For example,

$$\lim_{n\to\infty}\sum_{i=1}^{n} n^2 f(i) = \lim_{n\to\infty} n^2\sum_{i=1}^{n} f(i).$$

Using our properties above, we can rewrite the area under the parabola $y=-x^2+4x$ as

$$\begin{aligned}A =& \lim_{n\to\infty}\sum_{i=1}^{n} 64\left(\frac{i}{n^2}-\frac{i^2}{n^3}\right)\\ =& 64\lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n} i-\frac{1}{n^3}\sum_{i=1}^{n} i^2\right)\end{aligned}$$

Similarly, the distance travelled by a particle with velocity $v(t)=\frac{t^2}{4}$

over $[0, 10]$ may be rewritten as

$$\begin{aligned} D &= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\left({}^{10i}/_{n}\right)^2}{4} \cdot \frac{10}{n} \\ &= \lim_{n\to\infty} \frac{250}{n^3} \sum_{i=1} i^2 \end{aligned}$$

Some Helpful Formulas

We will be dealing with sums like

$$\sum_{i=1}^{n} i \quad \text{and} \quad \sum_{i=1}^{n} i^2$$

a lot, so it would be helpful if we had some formula for calculating these sums.

Theorem 4.10.

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Proof.
Let $S = 1 + 2 + \cdots + n$. Notice we can write this as $S = n + (n-1) + \cdots + 1$. Adding S to itself we have

$$2S = \underbrace{(n+1) + (n+1) + \cdots + (n+1)}_{n \text{ times}} = n(n+1)$$

and so dividing by 2 gives the sum. □

Theorem 4.11.

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

With these formulas at our disposal we can now calculate the area above and distance above.

Example 4.7.
The area under the curve $y = -x^2 + 4x$ above the interval $[0, 4]$ on the x-axis is

$$\begin{aligned}
A =& 64 \lim_{n\to\infty} \left(\frac{1}{n^2} \sum_{i=1}^{n} i - \frac{1}{n^3} \sum_{i=1}^{n} i^2 \right) \\
=& 64 \lim_{n\to\infty} \left(\frac{1}{n^2} \cdot \frac{n(n+1)}{2} - \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} \right) \\
=& 64 \lim_{n\to\infty} \left(\frac{n^2+n}{2n^2} - \frac{n(2n^2+n+2n+1)}{6n^3} \right) \\
=& 64 \lim_{n\to\infty} \left(\frac{n^2+n}{2n^2} - \frac{2n^3+3n^2+n)}{6n^3} \right) \\
=& 64 \left(\frac{1}{2} - \frac{1}{3} \right) \\
=& 64 \frac{3-2}{6} \\
=& \frac{64}{6} = \frac{32}{3}.
\end{aligned}$$

Example 4.8.
The distance traveled by a particle with velocity $v(t) = \frac{t^2}{4}$ from $t = 0$

to $t = 10$ is

$$\begin{aligned} D &= \lim_{n\to\infty} \frac{250}{n^3} \sum_{i=1}^{n} i^2 \\ &= \lim_{n\to\infty} \frac{250}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} \\ &= \lim_{n\to\infty} \frac{250(2n^3+3n^2+n)}{6n^3} \\ &= \lim_{n\to\infty} \frac{500n^3+750n^2+250n}{6n^3} \\ &= \frac{500}{6} = \frac{250}{3} \end{aligned}$$

4.3 Definite Integrals

> *Mathematics compares the most diverse phenomena and discovers the secret analogies that unite them.*
>
> JOSEPH FOURIER

In the last lecture we saw how to calculate the area under a curve, and the distance traveled by a particle, by considering limits of sums. The sums from the last lecture are examples of ***Riemann sums***, and the limits of these Riemann sums are examples of ***definite (Riemann) integrals***.

The Setup

To define Riemann sums and integrals in general, we need a few preliminary definitions.

Let $[a, b]$ be an interval. A ***partition*** $\mathcal{P}$ of $[a, b]$ is simply a finite ordered list of numbers in the interval $[a, b]$ which starts with a and ends with b. We will let x_0 denote a, x_1 the next value in the partition, x_2 the one after that and so on, until we get to the end when $x_n = b$. So a partition is just a list of numbers x_0, x_1, x_2, ..., x_n where

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

For example, one possible partition of $[3, 7]$ is

$$3, 4, 5.25, 5.75, 6.1, 7.$$

Another partition of $[3, 7]$ is

$$3, 3.1, 3.2, 3.9, 4, 6, 6.999999, 7.$$

This is all a partition is: just a finite list of numbers in order.

We use the numbers in the partition $\mathcal{P}$ to break the interval $[a, b]$ up into pieces called subintervals. The partition $\mathcal{P}$ given by

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

cuts $[a, b]$ into n subintervals:

$$[x_0, x_1], \quad [x_1, x_2], \quad [x_2, x_3], \quad \cdots, \quad [x_{n-2}, x_{n-1}], \quad [x_{n-1}, x_n].$$

We will let Δx_i denote the width of the i-th interval. That is,

$$\Delta x_i = x_i - x_{i-1}.$$

With this setup, we can now define a Riemann sum. A ***Riemann sum*** of a continuous function $f(x)$ over an interval $[a, b]$ is a sum of the form

$$\sum_{i=1}^{n} f\left(x_i^*\right) \Delta x_i$$

where $\mathcal{P}$ is some partition

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

of $[a, b]$, and where x_i^* is a *any* point inside the interval $[x_{i-1}, x_i]$.

The x_i^* points could be chosen in a number of different ways. We could always take x_i^* to be the left-hand endpoint of the interval (so $x_i^* = x_{i-1}$), or the right-hand endpoint ($x_i^* = x_i$), or the midpoint of the interval ($x_i^* = \frac{x_{i+1}+x_i}{2}$), or any other point. Different choices of x_i^* will of course give us different values for the Riemann sum, but this actually won't matter for what we're about to do next.

Given a partition $\mathcal{P}$ of $[a, b]$ we define the ***norm*** of the partition, denoted $|\mathcal{P}|$, to be the width of the largest subinterval: i.e., the maximum value of Δx_i:

$$|\mathcal{P}| = \max\left\{\Delta x_1, \Delta x_2, \cdots, \Delta x_n\right\}.$$

We then define the ***definite integral of a function*** $f(x)$ ***over*** $[a, b]$ to be the following limit, if the limit exists:

$$\lim_{|\mathcal{P}| \to 0} \sum_{i=1}^{n} f(x_i^*) \Delta x_i.$$

So the definite integral is the limit of Riemann sums. If this limit exists, we denote the value of the limit by $\int_a^b f(x)\,dx$:

$$\int_a^b f(x)\,dx = \lim_{|\mathcal{P}| \to 0} \sum_{i=1}^{n} f(x_i^*) \Delta x_i.$$

and we say the function $f(x)$ is ***integrable*** over $[a, b]$.

The above definitions apply for any partition $\mathcal{P}$, so we can simply things a little bit if we choose our partition to be the one that cuts $[a, b]$ up into n subintervals of equal width. In this case $\Delta x = \frac{b-a}{n}$, and we can take

x_i^* to be the right-hand endpoint of the i-th interval, $x_i^* = x_i = a + i\frac{b-a}{n}$. We then may write

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f\left(a + i\frac{b-a}{n}\right) \cdot \frac{b-a}{n}.$$

Example 4.9.
Express the following limit as an integral:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left(3 + \frac{4i}{n}\right)^2 \cdot \frac{4}{n}.$$

Here $f(x) = x^2$ and we're integrating over the region $[3, 7]$, so this is the integral

$$\int_3^7 x^2\,dx.$$

We will later see that there's a much simpler way to evaluate most integrals, but for the time being integrals for us mean limits of Riemann sums. In the last lecture we calculated some of these limits of Riemann sums: we found

$$\int_0^4 \left(-x^2 + 4x\right)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} \left(-\left(\frac{4i}{n}\right)^2 + 4\frac{4i}{n}\right)\frac{4}{n} = \frac{32}{3}$$

$$\int_0^{10} \frac{t^2}{4}\,dt = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\left(10i/n\right)^2}{4} \cdot \frac{10}{n} = \frac{250}{3}.$$

In evaluating these limits we used some helpful formulas:

$$\sum_{i=1}^{n}(f(i)\pm g(i)) = \sum_{i=1}^{n} f(i) \pm \sum_{i=1}^{n} g(i)$$
$$\sum_{i=1}^{n} kf(i) = k\sum_{i=1}^{n} f(i)$$
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

We add two more helpful formulas to this list.

Theorem 4.12.

$$\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$$
$$\sum_{i=1}^{n} 1 = n$$

With these at our disposal, let's calculate some more integrals.

Example 4.10.
Integrate $f(x) = x^3 - 2x^2 + x + 3$ over $[2, 5]$.

If we partition $[2, 5]$ into n intervals of equal length, then each interval has length $\Delta x = \frac{5-2}{n} = \frac{3}{n}$, and the right-hand endpoint of the i-th interval is $x_i = 2 + \frac{3i}{n}$. Thus

$$
\begin{aligned}
&\int_2^5 \left(x^3 - 2x^2 + x + 3\right)\, dx \\
=& \lim_{n\to\infty} \sum_{i=1}^{n} \left[\left(2 + \frac{3i}{n}\right)^3 - 2\left(2 + \frac{3i}{n}\right)^2 + \left(2 + \frac{3i}{n}\right) + 3 \right] \frac{3}{n} \\
=& \lim_{n\to\infty} \sum_{i=1}^{n} \left[8 + \frac{36i}{n} + \frac{54i^2}{n^2} + \frac{27i^3}{n^3} - 2\left(4 + \frac{12i}{n} + \frac{9i^2}{n^2}\right) + 5 + \frac{3i}{n} \right] \frac{3}{n} \\
=& \lim_{n\to\infty} \sum_{i=1}^{n} \left[8 + \frac{36i}{n} + \frac{54i^2}{n^2} + \frac{27i^3}{n^3} - 8 - \frac{24i}{n} - \frac{18i^2}{n^2} + 5 + \frac{3i}{n} \right] \frac{3}{n} \\
=& \lim_{n\to\infty} \sum_{i=1}^{n} \left[5 + \frac{15i}{n} + \frac{36i^2}{n^2} + \frac{27i^3}{n^3} \right] \frac{3}{n} \\
=& \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{15}{n} + \frac{45i}{n^2} + \frac{108i^2}{n^3} + \frac{81i^3}{n^4} \right) \\
=& \lim_{n\to\infty} \left(\frac{15}{n} \sum_{i=1}^{n} 1 + \frac{45}{n^2} \sum_{i=1}^{n} i + \frac{108}{n^3} \sum_{i=1}^{n} i^2 + \frac{81}{n^4} \sum_{i=1}^{n} i^3 \right) \\
=& \lim_{n\to\infty} \left(\frac{15}{n} \cdot n + \frac{45}{n^2} \cdot \frac{n(n+1)}{2} + \frac{108}{n^3} \frac{n(n+1)(2n+1)}{6} + \frac{81}{n^4} \left(\frac{n(n+1)}{2} \right)^2 \right) \\
=& \lim_{n\to\infty} \left(15 + \frac{45}{2} \cdot \frac{n^2 + n}{n^2} + \frac{108}{6} \cdot \frac{2n^3 + 3n^2 + n}{n^3} + \frac{81}{4} \frac{n^4 + 2n^3 + n^2}{n^4} \right) \\
=& 15 + \frac{45}{2} + \frac{108}{3} + \frac{81}{4} \\
=& \frac{180 + 270 + 432 + 243}{12} \\
=& \frac{1125}{12} \\
=& \frac{375}{4}
\end{aligned}
$$

Since we're talking about limits, we always have to ask ourselves if the limit exists or not. It turns out, however, the limit will exist for nice functions $f(x)$ – in particular, every continuous function is integrable.

Theorem 4.13.
If $f(x)$ is continuous on the interval $[a, b]$, then the definite integral $\int_a^b f(x)\,dx$ exists.

The following properties of integrals follow immediately from the definition of the integral as a limit of sums:

Theorem 4.14.
Let $f(x)$ and $g(x)$ be two integrable functions, and let k be any constant. Then for any interval $[a, b]$ we have

$$\int_a^b (f(x) \pm g(x))\; dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$$
$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$$
$$\int_a^b 1\,dx = b - a$$

Example 4.11.
Consider the function $f(x) = x + \sqrt{1-(x-1)^2}$. By the above we have

$$\int_0^2 \left(x + \sqrt{1-(x-1)^2}\right)\; dx = \int_0^2 x\,dx + \int_0^2 \sqrt{1-(x-1)^2}\,dx.$$

If we recognize that $y = \sqrt{1-(x-1)^2}$ is the top half of a circle of radius 1 centered at $x = 1$, and notice the area under the graph $y = x$ over $[0, 2]$ is a triangle with height 2 and base 2, then this integral becomes very easy:

$$\int_0^2 \left(x + \sqrt{1-(x-1)^2}\right)\, dx = \int_0^2 x\, dx + \int_0^2 \sqrt{1-(x-1)^2}\, dx$$
$$= \frac{1}{2}\cdot 2\cdot 2 + \frac{1}{2}\pi$$
$$= 1 + \frac{\pi}{2}.$$

Two less obvious properties are the following:

Theorem 4.15.
Let f be any integrable function and $[a, b]$ any interval.

$$\int_b^a f(x)\, dx = -\int_a^b f(x)\, dx$$
$$\int_a^a f(x)\, dx = 0.$$

The second property in the above theorem is easy to understand as $\Delta x = \frac{x-x}{n} = 0$, so each Riemann sum is zero. The first property may seem a little bit strange right now, but it will be useful later when we talk about substitutions.

One very helpful property of integrals is the following:

Theorem 4.16.
Let $a < b < c$ and let f be any integrable function. Then

$$\int_a^c f(x)\, dx = \int_a^b f(x)\, dx + \int_b^c f(x)\, dx$$

Example 4.12.
Let $f(x)$ be the function below:

$$f(x) = \begin{cases} 2x & \text{if } 1 \leq x < 4 \\ \frac{x^2}{2} & \text{if } 4 \leq x \leq 9 \end{cases}$$

Evaluate $\int_1^9 f(x)\,dx$.

Notice that by the above property we have

$$\begin{aligned} \int_1^9 f(x)\,dx &= \int_1^4 f(x)\,dx + \int_4^9 f(x)\,dx \\ &= \int_1^4 2x\,dx + \int_4^9 \frac{x^2}{2}\,dx \\ &= 2\int_1^4 x\,dx + \frac{1}{2}\int_4^9 x^2\,dx \end{aligned}$$

We can calculate each of these integrals separately:

$$\begin{aligned} &2\int_1^4 x\,dx \\ =&2\lim_{n\to\infty}\sum_{i=1}^{n}\left(1+\frac{3i}{n}\right)\frac{3}{n} \\ =&2\lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{3}{n}+\frac{9i}{n^2}\right) \\ =&2\lim_{n\to\infty}\left(\frac{3}{n}\sum_{i=1}^{n}1+\frac{9}{n^2}\sum_{i=1}^{n}i\right) \\ =&2\lim_{n\to\infty}\left(\frac{3}{n}\cdot n+\frac{9}{n^2}\cdot\frac{n(n+1)}{2}\right) \\ =&2\lim_{n\to\infty}\left(3+\frac{9}{2}\cdot\frac{n^2+n}{2}\right) \\ =&2\left(3+\frac{9}{2}\right) \\ =&6+9 \\ =&15 \end{aligned}$$

$$
\begin{aligned}
&\frac{1}{2}\int_4^9 x^2\,dx\\
=&\frac{1}{2}\lim_{n\to\infty}\sum_{i=1}^{n}\left(4+\frac{5i}{n}\right)^2\frac{5}{n}\\
=&\frac{1}{2}\lim_{n\to\infty}\sum_{i=1}^{n}\left(16+\frac{40i}{n}+\frac{25i^2}{n^2}\right)\frac{5}{n}\\
=&\frac{1}{2}\lim_{n\to\infty}\left(\frac{80}{n}\sum_{i=1}^{n}1+\frac{200}{n^2}\sum_{i=1}^{n}i+\frac{125}{n^3}\sum_{i=1}^{n}i^2\right)\\
=&\frac{1}{2}\lim_{n\to\infty}\left(\frac{80n}{n}+\frac{200(n^2+n)}{2n^2}+\frac{125(2n^3+3n^2+n)}{6n^3}\right)\\
=&\frac{1}{2}\left(80+100+\frac{250}{6}\right)\\
=&40+50+\frac{125}{6}\\
=&\frac{240+300+125}{6}\\
=&\frac{665}{6}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\int_1^9 f(x)\,dx =&2\int_1^4 x\,dx+\frac{1}{2}\int_4^9 x^2\,dx\\
=&15+\frac{665}{6}\\
=&\frac{90+665}{6}\\
=&\frac{755}{6}
\end{aligned}
$$

Inequalities & Total Area

It is sometimes necessary to relate the integrals of different functions over the same interval.

Theorem 4.17.
If $f(x)$ and $g(x)$ are integrable functions defined over $[a, b]$ and if $f(x) \geq g(x)$ for every x in $[a, b]$, then

$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$

Proof.
This follows immediately from the definition of the integral as a limit of Riemann sums and the corresponding statements for limits and summations. □

Corollary 4.18.
If f is integrable over $[a, b]$ and if there exist constants m and M with

$$m \leq f(x) \leq M$$

for every x in $[a, b]$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

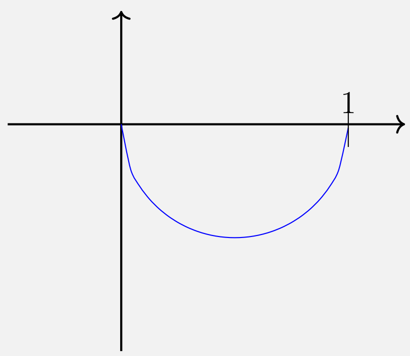
Notice this theorem and corollary imply that if $f(x) \leq 0$ for all x in $[a, b]$, then $\int_a^b f(x)\,dx \leq 0$ as well. Graphically, whenever the graph $y = f(x)$ drops below the x-axis, the integral will be negative over that interval.

Example 4.13.

$$\begin{aligned}
&\int_{-1}^{0} x\,dx \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \left(-1 + \frac{i}{n}\right) \frac{1}{n} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \left(-\frac{1}{n} + \frac{i}{n^2}\right) \\
&= \lim_{n\to\infty} \left(-\frac{1}{n} \sum_{i=1}^{n} 1 + \frac{1}{n^2} \sum_{i=1}^{n} i\right) \\
&= \lim_{n\to\infty} \left(-\frac{1}{n} \cdot n + \frac{1}{n^2} \cdot \frac{n^2+n}{2}\right) \\
&= -1 + \frac{1}{2} \\
&= -\frac{1}{2}
\end{aligned}$$

Notice that if $\int_a^b -f(x)\,dx = -\int_a^b f(x)\,dx$. Thus when our integral dips below the x-axis, we can still think about the integral in terms of areas to figure out the integral – we just have to remember to negate at the end.

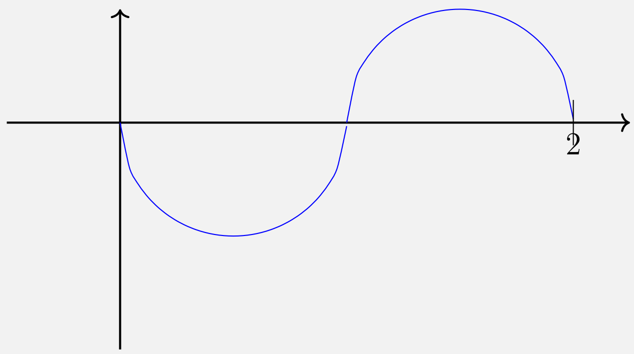
Example 4.14.
Suppose $f(x)$ is the function whose graph is given below. What is $\int_0^1 f(x)\,dx$?



Notice that the graph is half of a circle of radius $\frac{1}{2}$. Thus we know the integral $\int_0^1 -f(x)\,dx$ must be $\frac{\pi}{8}$, so

$$\int_0^1 f(x)\,dx = -\frac{\pi}{8}$$

Example 4.15.
What is the integral over $[0, 2]$ of the function whose graph is given below?



$$\begin{aligned}\int_0^2 f(x)\,dx &= \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx \\ &= -\frac{\pi}{8} + \frac{\pi}{8} \\ &= 0.\end{aligned}$$

Different Interpretations

Definite integrals are ubiquitous in mathematics and have many different interpretations, depending on what the function $f(x)$ is supposed to represent.

1. If $f(x) > 0$ on $[a, b]$, then $\int_a^b f(x)\,dx$ is the area under the curve $y = f(x)$ but above the interval $[a, b]$ on the x-axis.

2. If $v(t)$ represents the velocity of a particle at time t, then $\int_a^b v(t)\,dt$ is the displacement of the particle from time $t = a$ to time $t = b$.

3. If $\rho(x)$ represents the density of a wire at a point x in the interval $[a, b]$, then $\int_a^b \rho(x)\,dx$ is the mass of the wire.

4. If $I(x)$ is the electric current in a wire at a point x in $[a, b]$, then $\int_a^b I(x)\,dx$ is the change in the charge across the wire.

5. If $f(x)$ is the ***probability density function*** of a random variable, then $\int_a^b f(x)\,dx$ is the probability the value of the random variable is between a and b.

6. If $F(x)$ is the force exerted on an object at a point x in-between a and b, then $\int_a^b F(x)\,dx$ is the work done by that force as the object moves from a to b.

Example 4.16.
If you start walking up and down a straight path so that your velocity t minutes after you start walking is

$$v(t) = 4 - \frac{3}{2}x^2.$$

How far are you from your original position after 1 minute? What about after two minutes?

We find our displacement by integrating the velocity:

$$\begin{aligned}\int_0^1 \left(4-\frac{3}{2}x^2\right)\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(4-\frac{3}{2}\left(\frac{i}{n}\right)^2\right)\cdot\frac{1}{n}\\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{4}{n}-\frac{3i^2}{2n^3}\right)\\ &= \lim_{n\to\infty}\left(\frac{4}{n}\sum_{i=1}^{n}1-\frac{3}{2n^3}\sum_{i=1}^{n}i^2\right)\\ &= \lim_{n\to\infty}\left(4-\frac{3(2n^3+3n^2+n)}{12n^3}\right)\\ &=4-\frac{1}{2}\\ &=\frac{7}{2}.\end{aligned}$$

So after one minute we're $\frac{7}{2}$ feet to one direction of where we've started.

If we repeat the calculation up to time $t=2$ we have:

$$\begin{aligned}\int_0^2 \left(4-\frac{3}{2}x^2\right)\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(4-\frac{3}{2}\left(\frac{2i}{n}\right)^2\right)\cdot\frac{2}{n}\\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{4}{n}-\frac{12i^2}{2n^3}\right)\\ &= \lim_{n\to\infty}\left(\frac{4}{n}\sum_{i=1}^{n}1-\frac{12}{2n^3}\sum_{i=1}^{n}i^2\right)\\ &= \lim_{n\to\infty}\left(4-\frac{12(2n^3+3n^2+n)}{2n^2}\right)\\ &=4-12\\ &=-8.\end{aligned}$$

Example 4.17.
Given a spring, we can measure the amount of force needed to dis-

tort the spring by either stretching or compressing it. If it takes k Newtons of force to stretch the spring 1 meter, then we will say the spring has ***spring constant*** k. Then, in general, the force needed to distort the spring by x metres is kx (this is Hooke's law).

If we compress the string by ℓ metres, integrating the force at each point as we compress gives us

$$\begin{aligned}\int_0^\ell kx\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n} k\cdot\frac{\ell i}{n}\cdot\frac{\ell}{n}\\ &= \lim_{n\to\infty}\frac{k\ell^2}{n^2}\sum_{i=1}^{n} i\\ &= \lim_{n\to\infty}\frac{k\ell^2}{n^2}\frac{n(n+1)}{2}\\ &= \lim_{n\to\infty}\frac{k\ell^2}{2}\frac{n^2+n)}{n^2}\\ &=\frac{k\ell^2}{2}\end{aligned}$$

This is the elastic potential energy in the spring after we compress it.

4.4 The Fundamental Theorem of Calculus

The only way to learn mathematics is to do mathematics.

PAUL HALMOS

As we have seen, integrals have many interpretations and uses. However, they are extremely tedious to calculate as a limit of Riemann sums. In this lecture we introduce the fundamental theorem of calculus which makes computing integrals much easier.

The fundamental theorem of calculus is really two theorems. The first part promises us that antiderivatives of continuous functions exist, and the second part uses antiderivatives to give us a tool for computing integrals.

Fundamental Theorem of Calculus, Part I

We have seen in class some rules for computing antiderivatives of some simple functions, and it's easy to believe that we can extend our list of rules to find antiderivatives of more interesting functions. However, we have until now completely side stepped the question of whether antiderivatives exist in general or not.

Theorem 4.19 (The fundamental theorem of calculus, pt. 1)**.**
Let f be a continuous function on $[a, b]$, and consider the function F defined on $[a, b]$ as follows:

$$F(x) = \int_a^x f(t)\, dt.$$

Then F is continuous on $[a, b]$ and is an antiderivative of f; that is, $\frac{dF}{dx} = f$.

Proof.
First we prove continuity. Let $x_0 \in [a,b]$. We need to show

$$\lim_{x \to x_0} \int_a^x f(t)\,dt = \int_a^{x_0} f(t)\,dt.$$

Notice that because f is continuous on $[a,b]$, so is $|f(t)|$, so it has a global max and a global min. Let M be the max, and m be the min.

Notice

$$\begin{aligned} |F(x) - F(x_0)| &= \left| \int_a^x f(t)\,dt - \int_a^{x_0} f(t)\,dt \right| \\ &= \left| \int_x^{x_0} f(t)\,dt \right| \\ &\leq M\,|x_0 - x| \end{aligned}$$

Let $\varepsilon > 0$ be given, and choose $\delta = \frac{\varepsilon}{M}$.

$$\begin{aligned} &|x - x_0| < \delta \\ \implies &|x - x_0| < \frac{\varepsilon}{M} \\ \implies &M|x - x_0| < \varepsilon \\ \implies &|F(x) - F(x_0)| < \varepsilon. \end{aligned}$$

Thus F is continuous.

Now we compute the derivative of F:

$$\begin{aligned} &\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \\ =&\lim_{h \to 0} \frac{\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt}{h} \\ =&\lim_{h \to 0} \frac{\int_x^{x+h} f(t)\,dt}{h}. \end{aligned}$$

By the extreme value theorem, there is some $\alpha(h)$ in $[x, x+h]$ so that $f(\alpha(h))$ is the global max over $[x, x+h]$; and some $\omega(h)$ so that

$f(\omega(h))$ is the global min over $[x, x+h]$. Notice that $\lim_{h\to 0} \omega(h) = \lim_{h\to 0} \alpha(h) = x$ and

$$f(\omega(h))h \leq \int_x^{x+h} f(t)\,dt \leq f(\alpha(h))h,$$

so

$$f(\omega(h)) \leq \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\,dt}{h} \leq f(\alpha(h))$$

By the sandwich theorem, and continuity of f, taking the limit as h goes to zero gives

$$f(x) \leq \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\,dt}{h} \leq f(x)$$

and so

$$\begin{aligned} F'(x) &= \lim_{h\to 0} \frac{\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt}{h} \\ &= \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\,dt}{h} \\ &= f(x). \end{aligned}$$

□

The function defined in the theorem above looks very strange, but if we are interpreting the integral as area under a curve, all this function is doing is getting the under a portion of the curve, namely the portion over the interval $[a, x]$.

(Note to self: make a picture to go here.)

Though this function is written down in a weird way, the fundamental theorem of calculus tells us it is extremely easy to differentiate.

Example 4.18.

(a) The derivative of $\int_{-2}^{x} \cos(\sqrt{t})\,dt$ is $\cos(\sqrt{x})$.

(b) The derivative of $\int_4^x t^3\,dt$ is x^3.

(c) The derivative of $\int_x^4 t^3\,dt$ is $-x^3$. Notice that as the limits of integration have the variable x on the bottom, we have to switch the order to make the function agree with the form given in the theorem, and this makes us pick up a negative.

(d) The derivative of $\int_2^{x^2} \tan(t)\,dt$ is $2x\tan(x^2)$. Here we actually have to use the chain rule. If $F(x) = \int_2^x \tan(t)$, then our given function is $F(x^2)$, and so its derivative is $F'(x^2)2x$.

Exercise 4.1.
Suppose $f(x)$ is continuous at every point of the real line. What is the derivative of $\int_{-x}^x f(t)dt$?

Fundamental Theorem of Calculus, Part II

With the existence of antiderivatives out of the way, we now turn to the second part of the fundamental theorem of calculus, which gives us a tool for computing integrals *without* having to resort to limits of Riemann sums – at least provided we can compute the antiderivative of the function we want to integrate in some other way.

Theorem 4.20 (The fundamental theorem of calculus, pt. 2)**.**
If f is continuous on $[a, b]$ and F is an antiderivative of f, then

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

Proof.
We already know that $\int_a^x f(t)\,dt$ is an antiderivative of f, and since

F is another antiderivative we must have

$$F(x) = \int_a^x f(t)\,dt + C.$$

Simply notice

$$\begin{aligned} F(b) - F(a) &= \left(\int_a^b f(t)\,dt + C\right) - \left(\int_a^a f(t)\,dt + C\right) \\ &= \int_a^b f(t)\,dt + C - 0 - C \\ &= \int_a^b f(t)\,dt. \end{aligned}$$

□

Notice that the $+C$ that appears in the antiderivative we used to evaluate the integral above simply cancels out in the $F(b) - F(a)$ expression. For this reason we often don't bother to write the $+C$ when computing antiderivatives to evaluate integrals in this way.

We will adopt the following notation: we will write $F(x)\Big|_a^b$ as short hand for $F(b) - F(a)$.

Example 4.19.

(a)

$$\begin{aligned} \int_0^4 \left(-x^2 + 4x\right)\,dx &= \left(-\frac{x^3}{3} + \frac{4x^2}{2}\right)\Bigg|_0^4 \\ &= \left(-\frac{4^3}{3} + 2 \cdot 4^2\right) - (-0 + 0) \\ &= 32 - \frac{64}{3} \\ &= \frac{96 - 64}{3} \\ &= \frac{32}{3} \end{aligned}$$

(b)

$$\int_0^{10} \frac{t^2}{4}\,dt = \left.\frac{t^3}{12}\right|_0^{10}$$
$$= \frac{10^3}{12} - \frac{0}{12}$$
$$= \frac{1000}{12} = \frac{500}{6}$$
$$= \frac{250}{3}$$

(c) Consider the function

$$f(x) = \begin{cases} 2x & \text{if } 1 \leq x \leq 4 \\ \frac{x^2}{2} \text{ if } 4 \leq x \leq 9 \end{cases}$$

$$\int_1^9 f(x)\,dx = \int_1^4 f(x)\,dx + \int_4^9 f(x)\,dx$$
$$= \int_1^4 2x\,dx + \int_4^9 \frac{x^2}{2}\,dx$$
$$= \left.\frac{2x^2}{2}\right|_1^4 + \left.\frac{x^3}{6}\right|_4^9$$
$$= \left(4^2 - 1^2\right) + \left(\frac{9^3}{6} - \frac{4^3}{6}\right)$$
$$= 15 + \frac{729 - 64}{6}$$
$$= 15 + \frac{665}{6}$$

4.5 Indefinite Integrals and Net Change

Mathematics reveals its secrets only to those who approach it with pure love, for its own beauty.

ARCHIMEDES

In the last lecture we described the fundamental theorem of calculus which tells us that integration is the opposite of differentiation, in the sense that we can integrate a function to find its antiderivative: if f is continuous on $[a, b]$, then the function

$$F(x) = \int_a^x f(t)\,dt$$

is an antiderivative of f. This observation, combined with the fact that all antiderivatives of a function differ by a constant, tells us that derivatives and integrals *undo* one another, and that integrals are very closely related to antiderivatives.

Indefinite Integrals

A ***definite integral*** is by definition a limit of Riemann sums, and we can evalute these integrals using the fundamental theorem of calculus.

The indefinite integral is just another way of saying ***the general antiderivative***. Notice the key distinction: a definite integral is a number, while an indefinite integral is a function (really, an infinite family of functions because we can always add on an arbitrary constant). We thus denote the general antiderivative of $f(x)$ using integral notation: if $F(x)+C$ is the general antiderivative of $f(x)$, then we write

$$\int f(x)\,dx = F(x) + C.$$

For example,

$$\begin{aligned}
\int \left(x^2+4x-3\right)\, dx =& \frac{x^3}{3} + \frac{4x^2}{2} - 3x + C \\
\int \left(\cos(x)+\sqrt{x}\right)\, dx =& \int \left(\cos(x) + x^{1/2}\right)\, dx \\
=& \sin(x) + \frac{x^{3/2}}{3/2} + C \\
\int \frac{\sin\theta}{\cos^2\theta}\, d\theta =& \int \frac{1}{\cos\theta}\frac{\sin\theta}{\cos\theta}\, d\theta \\
=& \int \sec\theta\tan\theta\, d\theta \\
=& \sec\theta + C
\end{aligned}$$

Let's make a list of all the things we know about indefinite integrals (aka antiderivatives) so that you have a list of these in your notes:

$$\int \left(f(x)+g(x)\right)\, dx = \int f(x)\, dx + \int g(x)\, dx \qquad \int kf(x)\, dx = k\int f(x)\, dx$$

$$\int x^n\, dx = \frac{x^{n+1}}{n+1} + C \text{ provided } n \neq -1 \qquad \int \frac{dx}{x^n} = \frac{-1}{(n-1)x^{n-1}} + C$$

$$\int \sin(x)\, dx = -\cos(x) + C \qquad \int \cos(x)\, dx = \sin(x) + C$$

$$\int \sec^2(x)\, dx = \tan(x) + C \qquad \int \csc^2(x)\, dx = -\cot(x) + C$$

$$\int \sec(x)\tan(x)\, dx = \sec(x) + C \qquad \int \csc(x)\cot(x)\, dx = -\csc(x) + C$$

There are many things we don't yet know how to integrate: the even something like $\tan(x)$ is outside of what we know how to integrate right now, but you'll learn more techniques for integrating functions / finding antiderivatives next semester in 112.

Net Change

The second version of the fundamental theorem of calculus told us that if $F(x)$ is an antiderivative of $f(x)$, then

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

Notice that this quantity represents the change in F from $x = a$ to $x = b$; thus one way to interpret the fundamental theorem of calculus is that integrating a function's derivative tells us the ***net change*** in the function.

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

One example of this that we've already alluded to is *displacement*: if a particle moves up and down a straight line, its ***displacement*** is the difference between where the particle started and where it stopped – regardless of what happens inbetween. We can thus calculate displacement by integrating the derivative of position – velocity.

If $s(t)$ is the position of a particle at time t and $v(t) = s'(t)$ is its velocity, then

$$\int_a^b v(t)\,dt = s(b) - s(a).$$

Example 4.20.
Suppose you walk up and down a sidewalk so that your velocity t minutes into the walk is given by

$$v(t) = -100\sin(t)$$

measured in metres per second, where we will adopt the convention that positive velocity means travelling to the right, and negative velocity means travelling to the left. How far are you from your original starting position at time $t = \pi$. What about at time $t = 2\pi$?

To find the displacement we want to integrate velocity, thus at time $t = \pi$ our displacement is

$$\begin{aligned}\int_0^{\pi} -100\sin(t)\,dt =& -100\cos(t)\Big|_0^{\pi} \\ =& -100\cos(\pi) - (-100\cos(0)) \\ =& -100(-1) + 100\cdot 1 \\ =& 200.\end{aligned}$$

So you are 200 metres to the right of where you started at time $t = \pi$.

At time $t = 2\pi$ the displacement is

$$\begin{aligned}\int_0^{2\pi} -100\sin(t)\,dt =& -100\cos(t)\Big|_0^{2\pi} \\ =& -100\cos(2\pi) - (-100\cos(0)) \\ =& -100(1) + 100\cdot 1 \\ =& 0.\end{aligned}$$

so we are back to where we started.

Notice that this displacement *is not* the same as distance travelled – unless your velocity is always positive. For example, when the velocity is negative (e.g., we're moving to the left) we'd calculate the displacement to be the negative of the distance. To fix this we need to break the function up into intervals where the velocity is positive and intervals where the velocity is negative, calculate the displacement on each interval (i.e., integrate), but remember to multiply the negative parts by -1 to make them positive.

Notice that this is the same thing as integrating $|v(t)|$!

Example 4.21.
Find the total distance traveled if velocity is $v(t) = -100\sin(t)$ and t goes from $t = 0$ to $t = 2\pi$.

$$\begin{aligned}
\int_0^{2\pi} |v(t)|\,dt &= \int_0^{2\pi} |-100\sin(t)|\,dt \\
&= \int_0^{\pi} |-100\sin(t)|\,dt + \int_{\pi}^{2\pi} |-100\sin(t)|\,dt \\
&= \int_0^{\pi} -(-100\sin(t))\,dt + \int_{\pi}^{2\pi} -100\sin(t)\,dt \\
&= \int_0^{\pi} 100\sin(t)\,dt + \int_{\pi}^{2\pi} -100\sin(t)\,dt \\
&= -100\cos(t)\Big|_0^{\pi} + -(-100\cos(t))\Big|_{\pi}^{2\pi} \\
&= (-100\cos(\pi) + 100\cos(0)) + (100\cos(2\pi) - 100\cos(\pi)) \\
&= 100 + 100 + 100 + 100 \\
&= 400.
\end{aligned}$$

Example 4.22.
If we have a wire whose density at a point x is $\rho(x)$, then the integral

$$\int_a^b \rho(x)\,dx$$

represents the mass of the portion of the wire between a and b.

Suppose the density of a wire of length four metres is given by $\rho(x) = 9 + 2\sqrt{x}$, measured in kilograms per meter, where x is measured in metres from one end of the rod. What is the rod's mass? What is the mass of the middle third of the rod?

$$\begin{aligned}\int_0^4 \rho(x)\,dx &= \int_0^4 \left(9 + 2\sqrt{x}\right)\,dx \\ &= \left(9x + \frac{4}{3}x^{3/2}\right)\Bigg|_0^4 \\ &= 9 \cdot 4 + \frac{4}{3}4^{3/2} \\ &= 36 + \frac{4}{3} \cdot 8 \\ &= 36 + \frac{32}{3} \\ &= \frac{108 + 32}{3} \\ &= \frac{140}{3}\end{aligned}$$

Example 4.23.
Imagine a large, empty tank which we begin pouring water into. Say t minutes after we've started pouring water into the tank, the rate at which water is coming into the tank is $3t^2$ gallons per minute. How much water is in the tank after 10 minutes?

We want to find $V(10)$, and we're told $V(0) = 0$ and $\frac{dV}{dt} = 3t^2$. Thus

$$\begin{aligned}V(10) &= V(10) - V(0) \\ &= \int_0^{10} \frac{dV}{dt}\,dt \\ &= \int_0^{10} 3t^2\,dt \\ &= t^3\Big|_0^{10} \\ &= 1000.\end{aligned}$$

Example 4.24.
Imagine a large, tank which initially contains 250 gallons of water. We begin pouring water into the tank so that t minutes after we've started pouring water into the tank, the rate at which water is coming into the tank is $\frac{t^2+\sqrt{t}}{t}$ gallons per minute. How much water is in the tank after 20 minutes?

We want to find $V(20)$, and we're told $V(0) = 250$ and $\frac{dV}{dt} = \frac{t^2+\sqrt{t}}{t}$. Thus

$$\begin{aligned} V(20) =& V(20) - V(0) + V(0) \\ =& \int_0^{20} \frac{dV}{dt}\, dt + V(0) \\ =& \int_0^{20} \frac{t^2+\sqrt{t}}{t}\, dt + 250 \\ =& \int_0^{20} \left(t + t^{-1/2}\right)\, dt + 250 \\ =& \left(\frac{t^2}{2} + 2\sqrt{t}\right)\Bigg|_0^{20} + 250 \\ =& \frac{20^2}{2} + 2\sqrt{20} + 250 \\ =& 450 + 2\sqrt{20} \\ \approx& 458.94. \end{aligned}$$

So after twenty minutes there's about 458.94 gallons in the tank: 250 of those gallons were already there, and $\int_0^{20} \frac{dV}{dt}\, dt \approx 258.94$ were added.

4.6 Substitution

Education isn't something you can finish.

ISAAC ASIMOV

Substitution in Indefinite Integrals

Recall that the chain rule tells us how to differentiate a composition of two functions:

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

Thus we know that the antiderivative of $f'(g(x))g'(x)$ is $f(g(x)) + C$:

$$\int f'(g(x))g'(x)\, dx = f(g(x)) + C.$$

We can take advantage of this help us evalute certain types of integrals. For example,

$$\int \sin(x^2)2x\, dx.$$

Here $f'(g(x))g'(x) = \sin(x^2)2x$, so $f'(x) = \sin(x)$, $g(x) = x^2$, and $g'(x) = 2x$. We thus have

$$\int \sin(x^2)2x\, dx = -\cos(x^2) + C.$$

We can easily double-check that this is in fact the right antiderivative by differentiating:

$$\frac{d}{dx}\big(-\cos(x^2) + C\big) = \sin(x^2)2x.$$

As another example, consider

$$\int \sqrt{6x^2 - 2x + 3}\,(12x - 2)\ dx.$$

If $f'(g(x))g'(x) = \sqrt{6x^2 - 2x + 3}(12x - 2)$, then we see that $f'(x) = \sqrt{x}$, $g(x) = 6x^2 - 2x + 3$ and $g'(x) = 12x - 2$. Notice if $f'(x) = \sqrt{x} = x^{1/2}$, then $f(x) = \frac{2}{3}x^{3/2} + C$. We thus have

$$\int \sqrt{6x^2 - 2x + 3}\,(12x - 2)\ dx = \frac{2}{3}\left(6x^2 - 2x + 3\right)^{3/2} + C.$$

Notice that it would be okay if our integral was a multiplicative constant away from being of the form $f'(g(x))g'(x)$. For example, if we had

$$\int \sqrt{6x^2 - 2x + 3}\,(6x - 1)\ dx$$

then we don't have "the right" $g'(x)$ as part of our integrand. But this is easy to fix:

$$\begin{aligned}
&\int \sqrt{6x^2 - 2x + 3}\,(6x - 1)\ dx \\
=&\frac{2}{2}\int \sqrt{6x^2 - 2x + 3}\,(6x - 1)\ dx \\
=&\frac{1}{2}\int \sqrt{6x^2 - 2x + 3}\,(12x - 2)\ dx \\
=&\frac{1}{2}\cdot\frac{2}{3}\left(6x^2 - 2x + 3\right)^{3/2} + C \\
=&\frac{1}{3}\left(6x^2 - 2x + 3\right)^{3/2} + C
\end{aligned}$$

Usually the way we solve problems using this sort of backwards chain rule is to introduce a new variable, normally called u, which makes the integral simpler. For example, in the integral

$$\int \sqrt{6x^2 - 2x + 3}\,(12x - 2)\ dx$$

we may write $u = 6x^2 - 2x + 3$. Notice that u is then a function of x, and so we may consider its differential:

$$\begin{aligned}
u =& 6x^2 - 2x + 3 \\
\implies du =& \frac{d}{dx}\left(6x^2 - 2x + 3\right)\ dx \\
=& (12x - 2)\ dx
\end{aligned}$$

So in the original integral we can replace $6x^2 - 2x + 3$ with u, and $(12x - 2)dx$ with du. Thus we have

$$\int \sqrt{6x^2 - 2x + 3}\,(12x - 2)\ dx = \int \sqrt{u}\,du.$$

This is much simpler to integrate:

$$\int \sqrt{u}\,du = \frac{2}{3}u^{3/2} + C.$$

Now recalling that $u = 6x^2 - 2x + 3$ we have

$$\frac{2}{3}\left(6x^2 - 2x + 3\right)^{3/2}$$

just as before.

As another example, consider the integral

$$\int \cos(x) \cdot \sin^2(x)\, dx.$$

If we introduce the variable $u = \sin(x)$, then $du = \cos(x)dx$ and we have

$$\int \cos(x) \cdot \sin^2(x)\, dx = \int \sin^2(x) \cdot \cos(x)\, dx = \int u^2\, du$$

Now it's very easy for us to integrate $\int u^2\, du$:

$$\int u^2\, du = \frac{u^3}{3} + C$$

But recalling that $u = \sin(x)$ we have

$$\frac{u^3}{3} + C = \frac{\sin^3(x)}{3} + C$$

and so the antiderivative of $\cos(x)\sin^2(x)$ is $\frac{1}{3}\sin^3(x) + C$.

Because we are introducing a new variable in place of a more complicated equation, this trick is called ***substitution***, and since u is usually the variable that gets used it is commonly referred to as ***u-substitution***. There's nothing magical about using the letter u here, though – it's just a standard convention. We can just as easily substitute using a different variable.

Example 4.25.
Calculate the following indefinite integral:

$$\int \tan(\theta)\sec^2(\theta)d\theta$$

Let's introduce the variable $\varphi = \tan(\theta)$, then $d\varphi = \sec^2(\theta)d\theta$, and so

$$\begin{aligned}\int \tan(\theta)\sec^2(\theta)d\theta &= \int \varphi d\varphi \\ &= \frac{\varphi^2}{2} + C \\ &= \frac{\tan^2(\theta)}{2} + C.\end{aligned}$$

Example 4.26.
Calculate the following indefinite integral:

$$\int \frac{dy}{y^2 + 2y + 1}.$$

At first it's not really clear what we should replace. If we try to replace $y^2 + 2y + 1$, then we'd need a $2y + 2$, which we don't have. However, if we notice that $y^2 + 2y + 1$ factors as $(y + 1)^2$,

$$\int \frac{dy}{y^2 + 2y + 1} = \int \frac{dy}{(y + 1)^2}$$

we might notice that we could easily replace $y + 1$. Let's write $x = y + 1$, so $dx = dy$. We then have

$$\begin{aligned}\int \frac{dy}{y^2 + 2y + 1} &= \int \frac{dy}{(y + 1)^2} \\ &= \int \frac{dx}{x^2} \\ &= \int x^{-2} dx \\ &= -x^{-1} + C \\ &= -(y + 1)^{-1} + C \\ &= \frac{-1}{y + 1} + C.\end{aligned}$$

Again, if we're off by a constant factor, this is easy for us to compensate for in the substitution:

Example 4.27.
Calculate the following:

$$\int \frac{4x^2}{\sqrt{1+8x^3}}\,dx$$

Let's introduce the variable $u = 1+8x^3$. Notice that $du = 24x^2\,dx$. In our integral we don't have $24x^2\,dx$, we only have $4x^2\,dx$. But we can easily fix this if we multiply by 6, however we then have to divide by 6 as well:

$$\begin{aligned}
\int \frac{4x^2}{\sqrt{1+8x^3}}\,dx &= \frac{6}{6}\int \frac{4x^2}{\sqrt{1+8x^3}}\,dx \\
&= \frac{1}{6}\int \frac{24x^2}{\sqrt{1+8x^3}}\,dx \\
&= \frac{1}{6}\int \frac{1}{\sqrt{u}}\,du \\
&= \frac{1}{6}\int u^{-1/2}\,du \\
&= \frac{1}{6}\cdot 2u^{1/2} + C \\
&= \frac{1}{3}\sqrt{1+8x^3} + C
\end{aligned}$$

Example 4.28.
Calculate the following:

$$\int \frac{x}{\sqrt{1+2x}}\,dx.$$

Let's try to do the substitution $u = 1+2x$ and see what happens. If $u = 1+2x$ then $du = 2dx$. Notice that our x in the numerator doesn't get replaced. This is a problem, but one we can get around.

$$u = 1 + 2x \implies u - 1 = 2x \implies x = \frac{u-1}{2}.$$

So when we replace u by $1 + 2x$, we replace x by $\frac{u-1}{2}$. Notice also that dx gets replaced by $\frac{1}{2}du$.

Thus

$$\begin{aligned}
\int \frac{x}{\sqrt{1+2x}}\,dx = &\int \frac{\left(\frac{u-1}{2}\right)}{\sqrt{u}} \cdot \frac{1}{2}du \\
&\frac{1}{4}\int \frac{u-1}{\sqrt{u}}\,du \\
=&\frac{1}{4}\int \left(\frac{u-1}{u^{1/2}}\right)\,du \\
=&\frac{1}{4}\int \left(\frac{u}{u^{1/2}} - \frac{1}{u^{1/2}}\right)\,du \\
=&\frac{1}{4}\int \left(u^{1/2} - u^{-1/2}\right)\,du \\
=&\frac{1}{4}\left(\frac{2}{3}u^{3/2} - 2u^{1/2}\right) + C \\
=&\frac{1}{4}\left(\frac{2}{3}\left(1+2x\right)^{3/2} - 2(1+2x)^{1/2}\right) + C \\
=&\frac{1}{6}(1+2x)^{3/2} - \frac{1}{2}\sqrt{1+2x} + C
\end{aligned}$$

Substitution for Definite Integrals

Just as we can do substitution in indefinite integrals, we can also perform substitution in definite integrals. There is only one small thing we have to wrry about: the limits of integration. When we have an integral of the form

$$\int f(g(x))g'(x)\,dx$$

we do the substitution $u = g(x)$, $du = g'(x)dx$ to rewrite the integral as

$$\int f(u)\,du.$$

If we want to put limits of integration on this, then our second integral is in terms of u, so its limits of integration should be in terms of u. That is,

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

Example 4.29.

$$\begin{aligned}\int_0^1 (x+1)^2 \, dx &= \int_1^2 u^2\,du \\ &= \frac{u^3}{3}\bigg|_1^2 \\ &= \frac{8}{3} - \frac{1}{3} \\ &= \frac{7}{3}\end{aligned}$$

Let's double-check that this is the right answer:

$$\begin{aligned}\int_0^1 (x+1)^2 \, dx &= \int_0^1 \left(x^2+2x+1\right) \, dx \\ &= \left(\frac{x^3}{3} + x^2 + x\right)\bigg|_0^1 \\ &= \frac{1}{3} + 1 + 1 \\ &= \frac{1+3+3}{3} \\ &= \frac{7}{3}\end{aligned}$$

Notice in this example that before doing the substitution, x is between 0 and 1. When we do the substitution we replace $x+1$ with u. So as x goes from 0 to 1, we have that u goes from 1 to 2.

Example 4.30.

$$\int_1^2 x\sqrt{x-1}\,dx$$

Let's do the substitution $u = x - 1$. Then $du = dx$. Notice that if $u = x - 1$, then $x = u + 1$. The integral is thus

$$\begin{aligned}
\int_1^2 x\sqrt{x-1}\,dx &= \int_0^1 (u+1)\sqrt{u}\,du \\
&= \int_0^1 \left(u^{3/2} + u^{1/2}\right)\,du \\
&= \left(\frac{2}{5}u^{5/2} + \frac{2}{3}u^{3/2}\right)\Bigg|_0^1 \\
&= \frac{2}{5} + \frac{2}{3} \\
&= \frac{6+10}{15} \\
&= \frac{16}{15}
\end{aligned}$$

Example 4.31.

$$\int_0^2 (6x-4)^{17}\,dx$$

Let's do the substitution $u = 6x - 4$, so $du = 6dx$, and $dx = \frac{1}{6}du$.

Notice that when $x = 0$, $u = -4$; when $x = 2$, $u = 8$.

$$\begin{aligned}\int_0^2 (6x-4)\ dx =& \int_{-4}^8 u\frac{du}{6}\\ =&\frac{1}{6}\cdot\frac{u^2}{2}\bigg|_{-4}^8\\ =&\frac{1}{12}\left(8^2-(-4)^2\right)\\ =&\frac{1}{12}(64-16)\\ =&\frac{1}{12}48\\ =&4\end{aligned}$$

Example 4.32.
Integrate

$$\int_{-2}^2 (x+3)\sqrt{4-x^2}\,dx.$$

Let's first split up our integral as

$$\int_{-2}^2 x\sqrt{4-x^2}\,dx + \int_{-2}^2 3\sqrt{4-x^2}\,dx.$$

For the integral on the left we can do the substitution $u = 4 - x^2$, $du = -2x\,dx$ and so

$$\begin{aligned}\int_{-2}^2 x\sqrt{4-x^2}\,dx &= -\frac{1}{2}\int_0^0 \sqrt{u}\,du\\ &= 0.\end{aligned}$$

For the integral on the right let's notice

$$\int_{-2}^2 3\sqrt{4-x^2}\,dx = 3\int_{-2}^2 \sqrt{4-x^2}\,dx$$

We can determine $\int -2^2\sqrt{4-x^2}\,dx$ geometrically: this is the area of a semicircle of radius 2 centered at the origin, so

$$3\int_{-2}^{2}\sqrt{4-x^2}\,dx = 3\frac{4\pi}{2} = 6\pi$$

Example 4.33.
Suppose a tank containing 100 gallons of oil begins leaking and t hours after the leak begins the oil is leaking out at a rate of

$$100t^3\sqrt{5t^2+4}$$

gallons per hour. How much oil is left in the tank after one hour?

By the net change theorem, we know the change in the volume of the oil in the tank is the integral of the rate of change. Here that rate of change is

$$\frac{dV}{dt} = -100t^3\sqrt{5t^2+4}$$

since the volume is geting smaller (oil is leaking out). So the volume after one hour is $V(1)$ which we can express as

$$\begin{aligned} V(1) &= V(1) - V(0) + V(0) \\ &= \int_0^1 \frac{dV}{dt}\,dt + V(0) \\ &= \int_0^1 -100t^3\sqrt{5t^2+4}\,dt + 100. \end{aligned}$$

Thus our main goal is to compute the integral

$$\int_0^1 -100t^3\sqrt{5t^2+4}\,dt.$$

Let's perform the substitution

$$\begin{aligned} u &= 5t^2+4 \\ du &= 10t\,dt. \end{aligned}$$

There *is* a $10t$ hiding in our integral:

$$\int_0^1 -100t^3\sqrt{5t^2+4}\,dt = \int_0^1 -10t^2\sqrt{5t^2+4}\,10t\,dt.$$

So our u substitution turns "most" of the integral into $\sqrt{u}\,du$. But we need to take care of that $-10t^2$ out front.

Let's notice that if $u = 5t^2+4$, then $5t^2 = u-4$, so $-10t^2 = -2(u-4) = 8-2u$.

Now we can rewrite our integrand as $(8-2u)\sqrt{u}$. Notice that our limits of integration change to $u=4$ and $u=9$.

The integral is thus

$$\begin{aligned}\int_4^9 (8-2u)\sqrt{u}\,du &= \int_4^9 \left(8u^{1/2} - 2u^{3/2}\right)\,du \\ &= \left(\frac{16}{3}u^{3/2} - \frac{4}{5}u^{5/2}\right)\Bigg|_4^9 \\ &= \left(\frac{16}{3}9^{3/2} - \frac{4}{5}9^{5/2}\right) - \left(\frac{16}{3}4^{3/2} - \frac{4}{5}4^{5/2}\right) \\ &\approx -67.47\end{aligned}$$

So the tank lost about 67.47 gallons over the course of an hour, and 32.53 gallons remain in the tank.

Example 4.34.
Suppose that $f(x)$ is a continuous function and $\int_0^9 f(x)\,dx = 4$. What is $\int_0^3 xf(x^2)\,dx$?

Let's perform the substitution $u = x^2$, $du = 2x\,dx$. Then our integral may be rewritten as

$$\begin{aligned}\int_0^3 xf(x^2)\,dx &= \frac{1}{2}\int_0^9 f(u)\,du \\ &= \frac{4}{2} \\ &= 2\end{aligned}$$

Example 4.35.
Compute $\int_0^{1/2} \frac{\sin^{-1}(x)}{\sqrt{1-x^2}}\,dx$.

If $u = \sin^{-1}(x)$, then $du = \sqrt{1-x^2}$ and our integral becomes

$$\int_0^{\pi/6} u\,du = \frac{u^2}{2}\bigg|_0^{\pi/6} = \frac{\pi^2}{72}.$$

Lemma 4.21.
If f is a continuous even function (i.e., $f(-x) = f(x)$), then for every a,

$$\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx.$$

Proof.
Let's first split our integral up as

$$\int_{-a}^{0} f(x)\,dx + \int_0^a f(x)\,dx$$

Now the first integral we could rewrite as

$$\int_{-a}^{0} f(x)\,dx = -\int_0^{-a} f(x)\,dx$$

Now let's perform a (not so obvious) substitution, $u = -x$, $du = -dx$. Then

$$-\int_0^{-a} f(x)\,dx = \int_0^a f(u)\,du = \int_0^a f(x)\,dx$$

and combining this with the other integral above we have the result.

□

Exercise 4.2.
Use substitution to show that if f is a continuous odd function (i.e., $f(-x) = -f(x)$, then for every a,

$$\int_{-a}^{a} f(x)\,dx = 0.$$