These are merged lecture notes from several courses I taught at UBC in Vancouver, specifically MATH180 in 2008, MATH100 in 2009, and MATH105 in 2011. These are courses for engineers and other non-math-majors, so the notes are more calculation-oriented than proof-oriented, and not very rigorous.
Chapter 0

Introduction

The following is a rough overview of the course, and is intended to give an impression of what the main concepts are. To keep this outline short and simple, I left out a number of smaller topics, as well as some more complicated ones. For a complete list of topics, see the website, or see the contents of chapters 1-4 in the textbook.

0.1 Background • 0.2 Limits & Derivatives • 0.3 Applications

0.1 Background

Algebra. Algebra is the art of correctly manipulating formulas and equations. A large portion of the mistakes that are made on exams are simple algebra errors. Almost every problem or exercise you do will come down to some algebraic manipulations, and the smallest algebra mistake will ruin the final answer.

Algebra includes manipulation of sums, products, fractions, roots, exponentiation, equations, inequalities and many more things like that. You should already know how to do these things, but I’ll quickly mention two of them here that show up all the time. First is the good old quadratic formula

\[ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \]

which gives the two solutions to a quadratic equation \( ax^2 + bx + c = 0 \) (or you can tell from the square root that there’s only one or no solutions). Second is the factorization

\[ a^2 - b^2 = (a - b)(a + b), \]

so simple yet so often forgotten.

Functions. In my view there are 3 important perspectives on functions in calculus.

- The most concrete way to give a function is as a formula of a variable \( x \), for instance

\[ 2x + 3, \ x^2, \ \sin x, \ e^x, \ \log x, \ \frac{x^3 + 2}{x^2 + 1}, \ \sqrt{x - 2}, \ |x|, \ldots \]

Such a formula then ‘functions’ if we take a real number, plug it into the formula and get back another real number.

- Abstractly, a function is a rule that for any real number gives you another real number. Because ‘something given by a formula’ is not a good definition, we will need to understand this more general definition, although we won’t go very deep.
• Finally, a function can be given by a graph. This is probably the most intuitive way to think about functions, and it opens the door for studying a function through the geometry of its graph. We will see soon that that leads to the first main idea of calculus.

As a quick example, in the two graphs below we can see that \( f(x) = e^x \) is an increasing function, and that \( f(x) = x^2 + 1 \) has a minimum at \( x = 0 \), properties that are less easy to see from a formula.

Calculus is all about functions, so in this course we will build up a whole catalog of functions, we will study all kinds of properties of them, and we’ll learn how to sketch their graphs. Calculus has two main tools for analyzing functions, and this course is about the first, differentiation. The second, integration, you will see in Math 101.

Geometry. Another main ingredient of calculus is geometry. In this course, we’ll mostly see coordinate geometry (also called analytic geometry), and it will mostly be in 2 dimensions. In short, we study the geometry of the plane by laying a rectangular grid over it, so that we can describe points as coordinate pairs \((x, y)\), and we can describe objects like lines or circles by equations in terms of those coordinates.

Lines are the most basic objects in geometry, because they have the simplest equations. Given a line and two points \((x_1, y_1)\) and \((x_2, y_2)\) on it, we can find its equation by first computing the slope

\[
m = \frac{y_2 - y_1}{x_2 - x_1},
\]

which measures how steep the line is. Then the equation of the line can be written in the form

\[
y = mx + b,
\]

where we can solve for \(b\) if we plug in one of the points.

With lines we can describe lots of geometric objects, like squares, rectangles and triangles. We will also see some equations of other geometric objects, like a circle, which has equation

\[(x - a)^2 + (y - b)^2 = r^2,
\]

where \((a, b)\) is its center and \(r\) its radius.

But the geometric object we care about most in calculus is the graph of a function. If the function is \(f(x)\), then the equation of its graph is \(y = f(x)\). The whole idea of differentiation is that we can analyze a function by considering its graph as a geometric object and then analyzing it by approximating it with the simplest objects we have, which are lines.

### 0.2 Limits & Derivatives

We’ll first look at what a derivative is, and then we’ll see that we need limits to compute derivatives.

Consider the following graph \( y = f(x) \) (it doesn’t matter what \( f \) really is), with two points \( P_1 = (x_1, f(x_1)) \), \( P_2 = (x_2, f(x_2)) \) on it.
For reasons which will become clear later on, we want to approximate the graph at each point by a line, i.e. we want to find the line that at that point is most 'similar' to the graph. The line that does this is the 'tangent line', which goes through the point in the same direction as the graph. That direction is the slope of the line. In this picture, let's say the tangent line through $P_1$ has slope $m_1 = \frac{1}{2}$ and the one through $P_2$ has slope $m_2 = -1$. The derivative is a new function that keeps track of all these slopes.

\[
\text{The derivative of } f \text{ is the function } f' \text{ that for any } x_0 \text{ gives the slope of the tangent line to the graph } y = f(x) \text{ at } (x_0, f(x_0)).
\]

Suppose the two lines above have equations $y = m_1 x + b_1$ and $y = m_2 x + b_2$. Then we have

\[
f'(x_1) = m_1, \quad f'(x_2) = m_2.
\]

This definition is a bit tricky but it’s crucial. Students often mistakenly say things like 'the derivative is the tangent line' or 'the derivative is the slope of the tangent’, which are wrong and lead to confusion. A derivative is a function, and its value for a given $x$ is the slope of the tangent line to the graph above that $x$. And the graph we’re talking about is the graph of the original function $f$, whose derivative is the function $f'$.

As a first example of why this might be useful, consider the picture on the right. This should illustrate that at points where the graph has a maximum or minimum, the tangent lines are horizontal. We’ll see later that this gives an easy way to detect a maximum or minimum.

In many applications, that’s exactly what we want: the function might give the cost of something, and we would like to minimize that cost.

But this wouldn’t be useful unless we had some (easy) way of computing the derivative of a function. We do, of course, and that’s what this course is about. First we’ll see how to compute the slope of the tangent line to a graph at some point. To do this we introduce secant lines, which is just a fancy name for a line through two points on a graph.
For such a line, the slope is easy to compute, if we know the two points $P$ and $Q$ that determine the secant line. Let’s name the points $P = (x_P, f(x_P))$ and $Q = (x_Q, f(x_Q))$. Then the slope formula $\frac{y_2 - y_1}{x_2 - x_1}$ we saw above gives us the slope $m_{PQ}$ of the secant line,

$$m_{PQ} = \frac{f(x_Q) - f(x_P)}{x_Q - x_P}.$$  

We need these secant lines because we can’t get the slope of a tangent line directly, since we only know one point on the tangent line. The trick is that we can approximate the tangent line by secant lines, and then the slopes of those secant lines will approximate the slope of the tangent line. This is illustrated in the following picture.

We see that to find the tangent line at the point $Q$, we can take secant lines through $Q$ and some other point $P$. As we take $P$ closer and closer to $Q$, the slope $m_{PQ}$ of the secant line approaches the slope $m_Q$ of the tangent line. The technical notation for this is

$$m_Q = \lim_{P \to Q} m_{PQ}.$$  

The 'lim' here stands for limit, and this notation means that as we let $P$ approach $Q$, $m_{PQ}$ approaches $m_Q$. This may be a bit hard at first, but the idea is just what happens in the picture above. To be able to understand tangent lines (and then derivatives), we will have to understand these limits in a lot of detail.

The typical way of using limits is to denote the value of a function $f(x)$ as $x$ approaches some number $a$, as in

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

For example,

$$\lim_{x \to 2} (x^2 - x + 2) = 4.$$  

We can illustrate this by a graph or a table. For simplicity both only show an approach from the left, but in fact the approach in a limit should be from both sides. This will become clear when we treat limits in more detail in the course.
Let’s go through a complete example of computing the slope of a tangent line to a graph at a point. Let the graph be \( y = x^2 + x \) and the point \( Q = (x_Q, x_Q^2 + x_Q) = (0, 0) \). Then the equation of the tangent line looks like \( y = m_Q x \) (because plugging \( x = 0, \ y = 0 \) into \( y = m_Q x + b \) gives \( b = 0 \)), and we wish to compute \( m_Q \).

Let \( P = (x_P, x_P^2 + x_P) \) be the point that will approach \( Q \). Then by the formula above the slope of the secant line is

\[
m_{PQ} = \frac{f(x_P) - f(x_Q)}{x_P - x_Q} = \frac{(x_P^2 + x_P) - 0}{x_P - 0} = \frac{x_P(x_P + 1)}{x_P} = x_P + 1.
\]

Now we can write down the slope \( m_Q \) of the tangent line as a limit with \( P \) approaching \( Q \). Note that actually we only need to let \( x_P \) approach \( x_Q = 0 \). So

\[
m_Q = \lim_{P \to Q} m_{PQ} = \lim_{x_P \to 0} (x_P + 1) = 1.
\]

That’s all. Now we know that the tangent line at the origin has slope 1, and the equation of the tangent line is \( y = 1 \cdot x = x \). Most of what we will do in this course is building up the machinery that we need to do this systematically for much more complicated functions.

We can also write down the result above in terms of the derivative \( f' \) of \( f(x) = x^2 + x \),

\[
f'(0) = 1.
\]

By itself that’s not very useful notation. But we will learn how to compute that

\[
f'(x) = 2x + 1
\]

as a function, and then the above computation can be replaced by just plugging in \( f'(0) = 2 \cdot 0 + 1 = 1 \). That will save a lot of time.

The main skill we’ll have to learn is computing derivatives of functions, starting with the definition

\[
f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}.
\]

This is just what we did above: inside the limit is the slope of a secant line between two points on the graph, we take the limit as one point approaches the other, so that the secant lines approach the tangent line, and we get out of it the slope of the tangent line.

This might still seem pretty complicated, but we’ll learn how to take any function, break it into simpler pieces, take the derivatives of those pieces, and then put them back together. That way we will be able to differentiate (a different word for ‘computing the derivative of’) just about
any reasonable function, as long as we know the derivatives of a number of basic functions. The table below displays some of these basic derivatives. You don’t need to understand them yet, but you might already see some sort of pattern.

As a simple example of how we can break functions up into smaller pieces, consider \( f(x) = x^3 + x^2 + 5x \) from the first row of the table. It splits up like this:

\[
 f'(x) = (x^3 + x^2 + 5x)' = (x^3)' + (x^2)' + (5x)' = 3x^2 + 2x + 5.
\]

## 0.3 Applications

Finally, let’s take a quick look at one application of differential calculus, optimization. This is only one of the many applications we will see in this course, but it’s a good example, because it’s easy to see the use of it, as well as why calculus is necessary for it.

By **optimization** we mean finding the input value for which a certain function returns the best possible output. ‘Best’ means either a maximum or a minimum: if your function tells you the money you make from an investment, then you might want to maximize that, or if the function gives the distance of the route that you’re taking somewhere, then you want to minimize that. The observation that we made before about a maximum or minimum having a horizontal tangent line is what allows us to find it using the derivative.

Let’s look at a very simple example, the function \( f(x) = 2x - x^2 \). Here’s what its graph looks like:

![Graph of f(x) = 2x - x^2](image)

With your future differentiating skills you will easily see that the derivative of \( f \) is \( f'(x) = 2 - 2x \). So for any number \( a \) on the \( x \)-axis, we know that \( f'(a) = 2 - 2a \) is the slope of the tangent line to the graph of \( f \) at the point \((a, f(a))\). And to find the maximum we need to find the point that has a horizontal tangent line, in other words a point where the slope of the tangent line is zero. So we just have to solve the equation

\[
 f'(a) = 2a - 2 = 0 \quad \Rightarrow \quad a = 1.
\]

That means that the only place the graph can have a horizontal tangent line is at \( a = 1 \), so if it has a maximum or a minimum anywhere, it must be there. Now the value of \( f \) there is \( f(1) = 1 \), while at for instance \( x = 0 \), it is \( f(0) = 0 \), so it can’t be a minimum, since \( f(0) \) is smaller. Hence \( f(1) \) must be a maximum.
Indeed, this corresponds to what we see in the picture, the question mark is exactly in between 0 and 2. Of course, we could have seen that from the graph right away, but the point is that if we can do it without looking at the graph, we will also be able to do it for much harder functions, whose graphs may be much harder to draw.

This is still a pretty abstract application, but here’s a more concrete situation where we could use it. Imagine a lifeguard standing on the beach at the water’s edge. He sees a person in the water screaming for help, 100m further down the beach and 100m into the water. Suppose he can swim 1m/s and can run 2m/s. Of course he wants to get there as soon as possible, but should he jump straight in and swim diagonally (that’s $\sqrt{2} \cdot 100 \approx 141m$), or should he run the 100m along the water first and then swim the 100m? Well, swimming diagonally would take $\frac{141m}{1m/s} = 141$ seconds, while the other way would take $\frac{100m}{2m/s} + \frac{100m}{1m/s} = 150$ seconds. So should he jump right in?

Well, he can also run along the water some of the way, and swim from there. Suppose he runs 20m, then he will have to swim $\sqrt{80^2 + 100^2} = 128m$, and he will get there in $\frac{20m}{2m/s} + \frac{128m}{1m/s} = 138$ seconds, that’s a bit faster! But can he do any better?

We won’t do it now, because it’s a bit too hard, but you can write down a function for the time, dependent on how far the lifeguard runs along the water. Then if like above you look for horizontal tangent lines, you’ll find that it has a maximum somewhere in between, which corresponds approximately to running 42m along the water. Doing this will take him 137 seconds, and he can be certain that there’s no faster way.

Of course, it’s hard to imagine a lifeguard getting his notebook out to save a few seconds, and doing calculations for probably a few minutes (depending on how well he paid attention in his calculus class). But hopefully you see that this way we can solve a problem in an exact way, that we couldn’t really have found an answer for in any other way.
Chapter 1
Functions

1.1 Functions

Abstractly, a function is a rule that for each element of one set $D$ gives an element of another set $E$. The set $D$ is called domain of the function, and $E$ is the range. The notation for a function is

$$f : D \to R.$$ 

The way to think of an abstract function is as a group of arrows between two blobs (the sets): each arrow then goes from one element in the domain to one element in the range, and from each element, there is only one arrow going to the range, while each element of the range can be pointed to by zero, one, or more elements from the domain.

In this course, the domains will always be 'intervals' of 'real numbers'. I’ll have to sidetrack here to explain what those things mean.

A real number is a number that you can write down in (possibly infinite) decimal notation. This includes whole numbers (also called integers), like 0, 1, 2, 13, -5, it includes rational numbers like $\frac{1}{3} = 0.333\cdots$, $\frac{2}{7} = 0.714\cdots$, $\frac{-12}{5} = -2.4$, and then there are the irrational numbers. Some irrational numbers have familiar forms like $\sqrt{2}, \pi, e$, but others only exist in decimal notation, like 0.123456789101112\cdots.

The collection of all of the real numbers is denoted by $\mathbb{R}$. 

9
Another, more intuitive, way to think of \( \mathbb{R} \) is as ‘the real line’: imagine a line, infinite in both directions, with marks for 0 and 1. Then every real number has a place on this line, and every spot on this line is a real number.

Mathematicians have more precise definitions of \( \mathbb{R} \), but we don’t need those in calculus. An interval is a segment of the real line, or a combination of several (or even infinitely many) segments. Examples of basic intervals are \([1, 2]\), \((5, 7)\), \([0, 2\pi)\), \((-1, 3]\).

So an interval is described by its start and end point, and the type of bracket (‘[’ or ‘(’) tells you whether the start/end point is in the interval (if ‘[’), or not in it (if ‘(‘). In general, an interval is a combination of basic intervals like those above, for instance \([1, 3] \cup [4, 5) \cup (5, 6]\).

Note that 5 is not in this interval.

We will also have to work with infinite intervals like \([0, \infty)\), \((-\infty, 2]\) and \((-\infty, \infty) = \mathbb{R}\). But be warned that this is just notation, \(\infty\) is not a number!

Finally, note that this interval notation is closely related to inequalities: \([2, 3)\) is the set of real numbers \(x\) such that \(x \geq 2\) and \(x < 3\).

So in this course, the definition of function will be:

A function (with domain \(D\) and range \(E\), both intervals) is a rule that for each number in \(D\) gives a number in \(E\).

Functions as Formulas

This abstract definition is good for understanding functions, and setting the boundaries of what they can be, but in practice it’s not so useful. No one wants to describe a function by drawing a diagram with infinitely many arrows! So within this universe of abstract functions, we really only deal with the small world of functions that can be concretely described by some kind of formula. Some basic examples are

\[
2x + 3, \ x^2, \ \sqrt{x - 2}, \ \frac{1}{x^2 + 1}, \ |x|, \ \sin x, \ e^x, ...
\]

We will catalog some of these basic kinds of functions in section 3. And then there are several ways to combine functions to create new ones, see section 4.

Functions given by formulas are easier to deal with, because we can manipulate their formulas using algebra. But we would still be kind of helpless if we didn’t have a more intuitive or visual way to think about functions.

1.2 Graphs of Functions

The graph of a function is the set of points in the \(xy\)-plane given by the coordinates \((x, f(x))\), where \(x\) can be any number in the domain of \(f\). We think of a graph as the picture we get when we take the domain on the \(x\)-axis in the \(xy\)-plane, and then straight above each \(x\) in the domain we draw a point at height \(f(x)\). But note that a graph can be infinite, and the picture we draw is only a finite snapshot of it.

Another way of describing the graph of \(f\) is as the set of points \((x, y)\) that satisfy the equation \(y = f(x)\). For instance, the graph of the function \(f(x) = x^2\) is a parabola, and is also given by the equation \(y = x^2\). There are other curves that are given by an equation but that are not the graph of some function. For example, the equation \(x^2 + y^2 = 1\) gives a circle, which is not the graph of a function. We know that it’s not the graph of a function because for some \(x\)-values there are more than 1 points above (or below) it, for example with \(x = 0\) the circle contains \((0, 1)\) and \((0, -1)\). We can summarize this as the Graph Test, which we can use to check if a given curve is the graph of some function:
A curve is the graph of a function if for each \( x \) there is at most 1 point on the curve with that \( x \)-coordinate.

Note that for some \( x \) there might be no point on the graph with that \( x \)-coordinate; then that \( x \) is not in the domain.

### 1.3 Catalog of Functions

We will take a quick tour through the zoo of functions, focusing on the formulas and domains of the most important functions. For more detail (and pictures), see Stewart 1.2.

When we talk about the domain of a function, we really mean the biggest domain that that kind of function can have: if you have a function on a certain domain, you can always restrict the function to a smaller domain. So when I say 'the domain of a constant function is \( \mathbb{R} \)', it means that the biggest domain a constant function can have is \( \mathbb{R} \), even though it could also have a smaller domain.

On the last page you can see graphs of some of the basic functions; it’s useful to know these by heart.

- **Constant Functions** The simplest kind of function just always gives the same value, no matter what the input:
  \[
  f(x) = c
  \]
  for some real number \( c \). These functions are so simple they’re boring, but sometimes it is important to remember that they’re still functions. The domain is \( \mathbb{R} \), and the graph is just a horizontal line.

- **Linear Functions** A linear function looks like
  \[
  f(x) = ax + b
  \]
  for two real numbers \( a \) and \( b \). They’re the next simplest kind, but already much more interesting. In fact, calculus is based on using linear functions to analyze any function. The domain is \( \mathbb{R} \) and the graph is a line, but it cannot be a vertical line: a vertical line cannot be the graph of a function because it has more than one point above one value of \( x \).

- **Polynomials** A polynomial is a function that is built up out of addition and multiplication, for instance
  \[
  x^2 + 1 \quad \text{or} \quad 2x^7 - 3x^4 - 5.
  \]
  Constant and linear functions are of course also polynomials. Polynomials have domain \( \mathbb{R} \).

- **Rational Functions** Rational functions are built up out of addition, multiplication and division. Using basic algebra we can always rearrange such a function into a fraction of two polynomials, like
  \[
  \frac{x^5 + 2}{x^2 - 2} \quad \text{or} \quad \frac{x - 1}{x^2 + 7}.
  \]
  Because we cannot divide by zero, a rational function is not defined at the \( x \) for which its denominator is zero. The first example above is not defined at 2 or \(-2\), so its domain consists of all real numbers except 2 and \(-2\). On the other hand, the second example is
defined for all real numbers, because \( x^2 + 7 \) is always greater than 7. But sometimes we have to be careful. The rational function

\[
\frac{x^3 - 1}{x^2 - 1}
\]

looks like it might not be defined at 1, but if we simplify

\[
\frac{x^3 - 1}{x^2 - 1} = \frac{(x - 1)(x^2 + x + 1)}{(x - 1)(x + 1)} = \frac{x^2 + x + 1}{x + 1},
\]

we see that \( x = 1 \) actually is in the domain. The easy way to check if this might happen is to see if the numerator is zero for the \( x \) that you think is not in the domain. For instance, for \( x = 1 \) the numerator was \( 1^3 - 1 = 0 \), but for \( x = -1 \) the numerator is \((-1)^3 - 1 = 2 \neq 0\), so \(-1\) is indeed not in the domain.

- **Algebraic Functions** Algebraic functions are built up out of addition, multiplication, division and roots (\( \sqrt[n]{\cdot} \), \( \sqrt[3]{\cdot} \), \( \sqrt[4]{\cdot} \),...). For example,

\[
\sqrt{x + 1} \quad \text{and} \quad \frac{x^2 + \sqrt{x} + 1}{\sqrt{1 - x^2}}.
\]

Domains of algebraic functions are trickier. Aside from avoiding division by zero, we must also avoid taking even roots of negative numbers. The ‘even roots’ are \( \sqrt{\cdot} \), \( \sqrt[3]{\cdot} \), \( \sqrt[4]{\cdot} \),..., and we cannot take an even root of a negative number, because if for instance \( \sqrt{-8} = a \), then \( a^2 = -8 \), which is not possible because squares are always positive. On the other hand, there is no problem with \( \sqrt{-8} = -2 \), since \((-2)^3 = -8\).

So the domain of the first example above is where \( x + 1 \geq 0 \), i.e. \([ -1, \infty )\). The domain of the second example is more complicated: because of the \( \sqrt{x} \), we must have \( x \geq 0 \), and because of the \( \sqrt{x^2 - 1} \) we must have \( x \leq 1 \). But we also cannot have \( x = 1 \), because then we would divide by zero. Hence the domain is \([ 0, 1)\).

- **Trigonometric, Exponential & Logarithmic** We will devote more time to these later. Trigonometric functions are \( \sin x \), \( \cos x \) and all their family members, exponential functions are of the form \( a^x \), and logarithmic functions are of the form \( \log_a x \), where \( a > 0 \) is the base of the logarithm. The only one of these with an unusual domain is \( \log_a x \), which has domain \(( 0, \infty )\).

- **Many More...** Just so you don’t think that these are somehow all the functions we know, I should say that there are many more functions out there. Below you’ll see that we can combine any known functions to get new ones, which sometimes have completely new behavior. Later we’ll also see inverse functions like \( \arcsin = \sin^{-1} \). At the end of the course we’ll see Taylor Series, which open the door for defining lots of new functions, although we won’t really go into that in this course. And there’s many, many more.

### 1.4 Combining Functions

**Simple combinations**

There are several ways to combine functions to build new ones. For instance, we can **add**, **multiply** or **divide** two functions. For instance, let \( f(x) = \sqrt{x + 1} \) and \( g(x) = 2x - 4 \). Then the...
new functions \( f + g, f \cdot g = fg, \frac{f}{g} = f/g \) are determined by

\[
(f + g)(x) = f(x) + g(x) = \sqrt{x+1} + 2x - 4,
\]

\[
(fg)(x) = f(x) \cdot g(x) = \left(\sqrt{x+1}\right) \cdot (2x - 4),
\]

\[
\frac{f}{g}(x) = \frac{f(x)}{g(x)} = \frac{\sqrt{x+1}}{2x - 4}.
\]

Think about the domains of these new functions: a point is only in the domain of the new function if it was in the domain of both the old functions. BUT there is one exception for \( f/g \): because dividing by zero is never allowed, the denominator function shouldn’t take the value 0. In this example, we don’t want \( 2x - 4 = 0 \), which solves to \( x = 2 \). So \( x = 2 \) is not in the domain of \( f/g \), even though it is in the domain of \( f \) and \( g \).

**Composition**

The second most important way of making new functions is *composition*. The composition of two functions \( f \) and \( g \) is the function \( f \circ g \), defined by

\[
(f \circ g)(x) = f(g(x)).
\]

For example, if \( f(x) = \sqrt{x} \) and \( g(x) = x + 3 \), then

\[
(f \circ g)(x) = \sqrt{g(x)} = \sqrt{x+3}.
\]

So \( f \circ g \) is the function that first does \( g \) to \( x \), and then does \( f \) to the result of that.

As you may guess, determining the domain of \( f \circ g \) is a bit trickier: it consists of all \( x \) in the domain of \( g \) such that \( g(x) \) is in the domain of \( f \). In the example the domain of \( g \) is all of \( \mathbb{R} \), but the only \( x \) from \( \mathbb{R} \) that are in the domain of \( f \circ g \) are the ones for which \( g(x) = x + 3 \) is in the domain of \( f \), which is \([0, \infty)\). So we need the \( x \) for which \( x + 3 \) is in \([0, \infty)\), which happens for \( x \geq -3 \), so the domain of \( f \circ g \) is \([-3, \infty)\).

**Glueing**

Another way to create new functions is take two (or more...) functions, restrict their domains so that they don’t overlap, and then glue the two functions together.

For example, take \( f(x) = -x \) and \( f(x) = x \). Restrict the domain of \( f \) to \((-\infty, 0]\) and that of \( f \) to \((0, \infty)\), so they don’t overlap. Then gluing them gives this the familiar absolute value function \(|x|\). This is called a *piecewise defined function*, and we describe it with curly bracket notation:

\[
|x| = \begin{cases} 
 x & \text{if } x \geq 0 \\
 -x & \text{if } x < 0.
\end{cases}
\]
2.1 Definition

Our main definition is the following.

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

means that however \( x \) approaches the fixed number \( a \), the value of the function at \( x \), \( f(x) \), approaches the number \( L \). If such an \( L \) does not exist, we say that the limit does not exist.

Note that in this definition the value of \( f \) at \( a \) itself does not matter.
For example, the following picture illustrates this for \( \lim_{x \to 2} (x^2 - x + 2) = 4 \):

The tricky part of this definition is the different ways that \( x \) can approach \( a \). For instance, \( x \) can approach \( a \) from the left or the right, and the result for \( f(x) \) might be different. Consider the following graph, which uses the graphing convention that an open circle means that that point is not on the graph; here \( f(1) = 1 \), not 2.
Here as $x$ goes to 1 from the left, $f(x)$ goes to 1, but if $x$ comes from the right, $f(x)$ goes to 2. So in this case, the definition above tells us that $\lim_{x \to 1} f(x)$ does not exist, because two different ways of approach give different approached values. Nevertheless, we would still like to have notation for one-sided limits, which only involve the $x$ on one side of $a$.

We write

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

if as $x$ approaches $a$ from the left, $f(x)$ approaches $L$.

Similarly, we write

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

if as $x$ approaches $a$ from the right, $f(x)$ approaches $L$.

So be warned, a tiny change of notation ($a^-$ instead of $a$) makes a big difference in meaning. In the second graph above, where the two-sided limit did not exist, we have

$$\lim_{x \to 1^-} f(x) = 1 \quad \text{and} \quad \lim_{x \to 1^+} f(x) = 2.$$ 

It is still possible for a one-sided limit to not exist. For example, $\lim_{x \to 0} \frac{1}{x^2}$ does not exist, because as $x$ gets smaller and smaller, $\frac{1}{x^2}$ gets bigger and bigger, so does not approach any number.

This gives a way to check if a two-sided limit exists, by first checking if the two one-sided limits exist, and then seeing if their values are equal.

We have

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

if both

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

If the left and right limit are not equal, or if one of the one-sided limits does not exist, then the (two-sided) limit does not exist.

So in the first graph above, we have $\lim_{x \to 2} f(x) = 4$ since the left and right limits are both equal to 4. In the second graph, $\lim_{x \to 1} f(x)$ does not exist, since the left and right limits are different.

This left/right approach is especially convenient (and necessary) for piecewise defined functions, when the number $a$ is on the border between two of the pieces. The most common example of this is the absolute value function, so let’s look at

$$\lim_{x \to 0} |x|.$$ 

To determine the existence and value of this limit, we need to look at the one-sided limits first. For the left-hand limit, we only need to consider $x$ with $x < 0$, for which the definition of $|x|$ says $|x| = -x$. Hence

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

The last step is true because as $x$ approaches 0, so does $-x$.

Similarly,

$$\lim_{x \to 0^+} |x| = \lim_{x \to 0^+} x = 0.$$ 

So indeed the left and right limit are equal, and we can conclude that $\lim_{x \to 0} |x| = 0$. Of course the above limit would have been easy to see from the graph of $|x|$, but we want to be able to determine limits without looking at the graph of the function, especially because later
on we will have to use limits to draw graphs.
As another example, consider
\[ f(x) = \begin{cases} 
2x - 1 & \text{if } x \leq 1 \\
x^2 & \text{if } x > 1.
\end{cases} \]

Then to the left of \( x = 1 \), the function is given by \( 2x - 1 \), and on the right by \( x^2 \), so if we calculated the two one-sided limits separately, we get
\[
\lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} (2x - 1) = 2 \left( \lim_{x \to 1^-} x \right) - 1 = 2 \cdot 1 - 1 = 1,
\]
\[
\lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} x^2 = \left( \lim_{x \to 1^+} x \right)^2 = 1^2 = 1.
\]
So the left and right limits exist and both equal, hence we can conclude that \( \lim_{x \to 1} f(x) = 1 \).

Remark that when we say '\( x \) approaches \( a \)', we don’t care about what happens when \( x \) equals \( a \), since that is not part of the approach. If instead of \( |x| \) we considered the (weird) function
\[ h(x) = \begin{cases} 
-x & \text{if } x < 0 \\
13 & \text{if } x = 0 \\
x & \text{if } x > 0,
\end{cases} \]
the left and right limits would still be 0, so the limit exists and \( \lim_{x \to 0} h(x) = 0 \), even though \( h(0) = 13 \).

### 2.2 Computing Limits

The following example illustrates the basic approach to computing limits of functions given by nice formulas.
\[
\lim_{x \to 5} (2x^2 - 3x + 4) = \lim_{x \to 5} (2x^2) + \lim_{x \to 5} (-3x) + \lim_{x \to 5} 4 \\
= 2(\lim_{x \to 5} x)^2 - 3(\lim_{x \to 5} x) + (\lim_{x \to 5} 4) \\
= 2 \cdot 5^2 - 3 \cdot 5 + 4 = 50 - 15 + 4 = 39.
\]
The idea is to reduce the big limit to several small limits which are easy to determine, usually one of the types
\[ \lim_{x \to a} x = a \quad \text{or} \quad \lim_{x \to a} b = b. \]
Since the left/right issue is not a problem for these two easy limits, we do not have to worry about that in this approach.
The question is if all the steps above are actually valid: is it true for instance that
\[ \lim_{x \to 5} x^2 = \left( \lim_{x \to 5} x \right)^2 ? \]
We call this 'interchanging the order of the operations': on the left you square \( x \) first and then consider the limit of that, on the right you first take the limit and then square that.
This is what the 'limit laws' (Stewart 2.3) answer: can we interchange the limit operation with the basic arithmetic operations? I like to summarize these laws as follows:

*Limits interchange perfectly with addition, multiplication, division and root taking, except that we can’t divide by 0, and we can’t take even roots of anything negative.*
So for polynomials like the one above, everything works out. The same is true for rational and algebraic functions, as long as neither of the exceptions occurs. For instance
\[
\lim_{x \to 2} \frac{\sqrt{x} + 6}{x^2 + 1} = \frac{\sqrt{\lim_{x \to 2} x} + \lim_{x \to 2} 6}{(\lim_{x \to 2} x)^2 + \lim_{x \to 2} 1} = \frac{\sqrt{2} + 6}{4 + 1} = \frac{1}{5}.
\]

Two examples that do not work are
\[
\lim_{x \to 2} \frac{x^2 + 1}{x - 2} \quad \text{and} \quad \lim_{x \to -1} 3\sqrt{x} + x^2,
\]
in the first we would be dividing by \(\lim_{x \to 2} x - 2 = 0\), in the second we’d get a \(\sqrt{-1}\).

So now we can compute any limit of an algebraic function (or find that the limit doesn’t exist), by interchanging limits and operations until we have an easy limit, or until we run into one of the exceptions.

Writing this down step-by-step will get annoying quickly, as you’ll realize that you’re basically plugging in the number that \(x\) approaches. We can simplify this, but we still have to keep in mind when it’s allowed and when not. For that we will need the notion of \textit{continuity}.

\section{2.3 Continuity}

A function is \textit{continuous at} \(a\) if
\[
\lim_{x \to a} f(x) = f(a),
\]
and it is \textit{continuous on the interval} \(I\) if it is continuous at all numbers \(a\) in the interval \(I\). If we just say that a function is continuous, it means that it is continuous on all of its domain.

This means that for \(x\) close to \(a\), the function value \(f(x)\) is also close to \(f(x)\). An intuitive way to think of it is that you can draw the graph of \(f\) (above \(I\)) without lifting your pen from the paper (though this is incorrect in some cases, e.g. if there is a gap in the domain).

Yet another way to think of it is that ‘applying \(f\)’ is an operation that we can interchange limits with, because another way to write the equation above is (just plug in \(a = \lim_{x \to a} x\)):
\[
\lim_{x \to a} f(x) = f(\lim_{x \to a} x).
\]

So if \(f\) is continuous, applying \(f\) then the limit is the same as the other way around. For example, when we wondered if \(\lim_{x \to 5} x^2 = (\lim_{x \to 5} x)^2\), we could just as well have asked if the function \(f(x) = x^2\) is continuous at \(x = 5\). The limit laws then said that

\textit{Polynomials, rational and algebraic functions are continuous, \textbf{except} when there is division by zero, or an even root of a negative number.}

To check from the definition if a given function is continuous, we need to check the three things that are implicit in the definition:

- \(f(a)\) is defined (that is, \(a\) is in the domain of \(f\))
- \(\lim_{x \to a} f(x)\) exists
- \(\lim_{x \to a} f(x) = f(a)\).

If any of these fails, we say that the function is \textit{discontinuous} at \(a\), or has a \textit{discontinuity}.

To see some examples, it’s best to look at the examples of discontinuities on p.120 of Stewart.
The game is now to determine which functions are continuous where on their domains, and especially if the basic functions are continuous on all of their domain. Of course, outside of its domain a function is not defined, so the first of the three conditions above fails, and the function is certainly not continuous. By the limit laws, all algebraic functions are continuous on their domains; the two exceptions correspond to points outside of the domains. For instance, the function $\frac{1}{x}$ is continuous on its domain, which consists of all real numbers except for 0. Since the function is not defined at 0, it doesn’t make much sense to say that it’s continuous there or not. The same is true for exponential and trigonometric functions, and their inverses (which we’ll meet later): they are all continuous on all of their domain. But note that some of these functions do not have all of $\mathbb{R}$ as their domain; for instance, integer multiples of $\pi/2$ are not in the domain of $\tan x = \sin x / \cos x$, since there $\cos x$ is 0.

So in summary, all the elementary functions above are continuous on their domains. Of the three conditions, only the first one ever fails for these functions, mostly because of division by zero or even roots of negative numbers. When the second and third condition fail, it is most often for piecewise defined functions (including $|x|$), which is why most continuity questions in the book or the homework are about such functions.

The final thing to know is:

*A composition of two continuous functions is continuous as well.*

With that fact we can figure out if any function given by some sort of formula is continuous or not, by breaking it up as a composition of simple functions, and using what we know about those. The only hard part may be when piecewise defined functions are involved, in which case we will have to analyze the left and right limits. For example, the function

$$f(x) = \frac{1}{2 + \sin x}$$

is continuous on its domain (which is $\mathbb{R}$: we never divide by zero because $-1 \leq \sin x \leq 1$), because it is a composition $f = g \circ h$ of the functions $g(x) = \frac{1}{2+x}$ and $h(x) = \sin x$, a rational and a trigonometric function, both of which we know to be continuous.

On the other hand, the function

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

is not continuous at $x = 1$. We check the three conditions for continuity:

- The function is clearly defined at $x = 1$.
- To find $\lim_{x \to 1} f(x)$, we need to check the left and right limits (because it’s piecewise defined):

  $$\lim_{x \to 1^-} f(x) = \lim_{x \to 1^-} x^2 + 1 = 2, \quad \lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} \frac{2}{x} = 2.$$

  Since the two one-sided limits are equal, the two-sided limit exists, and $\lim_{x \to 1} f(x) = 2$.
- But $\lim_{x \to 1} f(x) = 2 \neq f(1)$!

So one of the conditions fails, and the function is not continuous at $x = 1$. At all other $x$, however, the function is given by $x^2 + 1$ or $\frac{2}{x}$, which we know to be continuous, except for $\frac{2}{x}$ at $x = 0$, but that doesn’t happen. So this $f(x)$ is continuous at all $x$ except 1.
2.4 Examples of Limits

Right now we have two main methods for computing limits:

- We can interchange limits and continuous functions, until we end up with easy limits, or we run into one of the exceptions. In practice, we do this by saying why the function is continuous (if it is), and then just plugging in.

- We can compute the left and right limits, and see if they are equal. Especially useful for piecewise defined functions.

In later chapters we will add to this two harder approaches, the Squeeze Theorem and Taylor series.

Many of the limit questions on exams are however of a slightly different nature: they require some algebraic trick before one of the methods above can be applied. We give some examples of these.

The most common kind is the following type of limit calculation:

\[
\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = \lim_{x \to 1} \frac{(x - 1)(x + 1)}{x - 1} = \lim_{x \to 1} x + 1 = 2.
\]

Here the function is a rational function that seems to be undefined at the number that \(x\) approaches, so we couldn’t just plug in. But after a cancellation there is no problem at all. What makes it hard is that the factors that cancel each other may be hidden, and you will have to factor a polynomial first. Another example is

\[
\lim_{x \to -4} \frac{x^2 + 5x + 4}{x^2 + 3x - 4} = \lim_{x \to -4} \frac{(x + 4)(x + 1)}{(x + 4)(x - 1)} = \lim_{x \to -4} \frac{x + 1}{x - 1} = \frac{3}{5}.
\]

Another very common kind looks like this:

\[
\lim_{x \to 9} \frac{\sqrt{x} - 3}{x - 9} = \lim_{x \to 9} \frac{\sqrt{x} - 3}{x - 9} \cdot \frac{\sqrt{x} + 3}{\sqrt{x} + 3} = \lim_{x \to 9} \frac{x - 9}{(\sqrt{x} - 3)(\sqrt{x} + 3)} = \lim_{x \to 9} \frac{1}{\sqrt{x} + 3} = \frac{1}{6}.
\]

The trick is that when you see something of the form \(a + \sqrt{b}\) in a rational function, multiplying top and bottom by the conjugate \(a - \sqrt{b}\) will lead to \((a + \sqrt{b})(a - \sqrt{b}) = a^2 - b^2\), which might clear up some cancellation.

But note that the second kind is actually a lot like the first kind, because for instance we could have used \(x - 9 = (\sqrt{x} - 3)(\sqrt{x} + 3)\), although such a factorization may be harder to spot.

Another algebraic step that you might have to take is clearing denominators, like in

\[
\lim_{x \to a} \frac{x - a}{x - a} = \lim_{x \to a} \frac{\frac{x - a}{ax}}{\frac{x - a}{ax}} = \lim_{x \to a} \frac{x - a}{ax(x - a)} = \lim_{x \to a} \frac{1}{ax} = \frac{1}{a^2}.
\]
Chapter 3

More Limits

3.1 Infinite Limits

• Infinite limits are limits that actually don’t exist, but in a specific way, namely as $x$ approaches $a$, $f(x)$ keeps growing, for example
\[
\lim_{x \to 0} \frac{1}{x^2} = \infty, \quad \text{or} \quad \lim_{x \to 0} \frac{-1}{x^2} = -\infty.
\]

• Limits at infinity are limits where $x$ keeps getting larger (either positive or negative; they look like
\[
\lim_{x \to \infty} \frac{1}{x^2} = 0, \quad \text{or} \quad \lim_{x \to -\infty} \frac{1}{x^2} = 0.\]

Subtleties

What we mean precisely by 'grows larger than any positive number' is that no matter how large a number $N$ we take, if we take $x$ close enough to $a$, $f(x)$ will be larger than $N$. For instance, we have
\[
\lim_{x \to 0^+} \frac{1}{x} = \infty
\]
because for any $N$, if we take $x = \frac{1}{N+1}$, which is very close to $a = 0$, then $f\left(\frac{1}{N+1}\right) = \frac{1}{1/(N+1)} = N + 1 > N$. For example, if our big number is $N = 1000$, then $f\left(\frac{1}{1001}\right) = 1001$, so $\frac{1}{x}$ grows

3.2 Limits at Infinity

• Limits at infinity are limits where $x$ keeps getting larger (either positive or negative; they look like
\[
\lim_{x \to \infty} \frac{1}{x^2} = 0, \quad \text{or} \quad \lim_{x \to -\infty} \frac{1}{x^2} = 0.\]

3.3 Squeeze Theorem

We will introduce two new kinds of limits:

We can make the same definitions for left and right limits, but we won’t spell that out here.

For example, $\lim_{x \to \infty} \frac{1}{x^2} = 0$ implies that $y = 0$ is a horizontal asymptote of $y = \frac{1}{x^2}$.

3.1 Infinite Limits

Definition: If however $x$ approaches $a$, $f(x)$ grows larger than any positive number, we write
\[
\lim_{x \to a} f(x) = \infty,
\]

instead of saying that the limit does not exist. If $f(x)$ grows below any negative number, we write $\lim_{x \to a} f(x) = -\infty$.

We can make the same definitions for left and right limits, but we won’t spell that out here.

For example, $\lim_{x \to a^-} f(x) = \infty$ means that if $x$ approaches $a$ from the left, $f(x)$ grows larger than any positive number (and it doesn’t matter what happens to the right of $a$).

Subtleties

What we mean precisely by 'grows larger than any positive number' is that no matter how large a number $N$ we take, if we take $x$ close enough to $a$, $f(x)$ will be larger than $N$. For instance, we have
\[
\lim_{x \to 0^+} \frac{1}{x} = \infty
\]
because for any $N$, if we take $x = \frac{1}{N+1}$, which is very close to $a = 0$, then $f\left(\frac{1}{N+1}\right) = \frac{1}{1/(N+1)} = N + 1 > N$. For example, if our big number is $N = 1000$, then $f\left(\frac{1}{1001}\right) = 1001$, so $\frac{1}{x}$ grows
larger than 1000.
In the same way, \( \lim_{x \to 0^-} \frac{1}{x} = -\infty \), because \( \frac{1}{x} \) grows below any large negative number.

Just like with regular limits, a two-sided infinite limit exists if the corresponding one-sided limits both exist and are equal (so both \( \infty \) or both \(-\infty \)). For example, we just saw that \( \lim_{x \to 0^-} \frac{1}{x} \neq \lim_{x \to 0^+} \frac{1}{x} \), so the two-sided limit \( \lim_{x \to 0} \frac{1}{x} \) does not exist, but is not \( \infty \) or \(-\infty \)!

On the other hand,
\[
\lim_{x \to 0^-} \frac{1}{x^2} = \left( \lim_{x \to 0^-} \frac{1}{x} \right)^2 = \infty = \lim_{x \to 0^+} \frac{1}{x^2},
\]
so \( \lim_{x \to 0} \frac{1}{x^2} = \infty \) (but we could still say that it doesn’t exist).

So beware of the difference between a limit not existing, and a limit being infinite. All infinite limits are limits that do not exist, but not all non-existing limits are infinite. Above we see an example (\( \lim_{x \to 0} \frac{1}{x} \)) of a two-sided limit that does not exist, but is not infinite either.

It’s also possible to have a one-sided limit that doesn’t exist, but isn’t infinite. Right now the only example of a function with a non-existing one-sided limit that we have seen is \( \lim_{x \to 0} \frac{1}{x} \), which is indeed an infinite one-sided limit, but at the end of this chapter we will see an example of a function with a non-existing one-sided limit that is not infinite. (It’s normal if you have to read this paragraph a couple of times...)

Whenever one of \( \lim_{x \to a} f(x) \), \( \lim_{x \to a^+} f(x) \) or \( \lim_{x \to a} f(x) \) is \( \infty \) or \(-\infty \), we say that the vertical line \( x = a \) is a vertical asymptote of the graph of \( f(x) \). See the end of section 2.2 in Stewart for lots of pictures of this.

**Examples**

We’ve now seen one way of showing that a limit is infinite, by arguing from the definition, like we did for \( \lim_{x \to 0^+} \frac{1}{x} \) above. From that we can handle \( \frac{1}{x^a} \) for any \( a > 0 \), as follows:

\[
\lim_{x \to 0^+} \frac{1}{x^a} = \left( \lim_{x \to 0^+} \frac{1}{x} \right)^a = \infty^a = \infty,
\]

where \( \infty^a \) is actually not allowed, and we should instead argue that if you have something that goes to \( \infty \), and then in the process you take a positive power of it, it will still go to infinity. For \( x \to 0^- \), it’s a bit trickier, because you could have trouble with even roots of negative numbers.

Fortunately, now that we’ve settled \( \frac{1}{x^a} \), we don’t need to argue from the definition anymore; we can do most other examples by reducing to one of those. For example, a limit like \( \lim_{x \to 1} \frac{1}{(x-1)^2} \)
we can do with the substitution trick: set \( u = x - 1 \). Then \( \frac{1}{(x-1)^2} = \frac{1}{u^2} \), and \( x \to 1 \) is the same as \( u \to 0 \). Hence

\[
\lim_{x \to 1} \frac{1}{(x-1)^2} = \lim_{u \to 0} \frac{1}{u^2} = \infty.
\]

A more complicated example, now with the substitution \( u = x - 2 \) (so \( x = u + 2 \)):

\[
\lim_{x \to 2^+} \frac{x - 3}{x^2 - 4} = \lim_{x \to 2^+} \frac{x - 3}{(x - 2)(x + 2)} = \lim_{u \to 0^+} \frac{(u + 2) - 3}{u((u + 2) + 2)} = \lim_{u \to 0^+} \frac{u - 1}{u(u + 4)} = \left( \lim_{u \to 0^+} \frac{u - 1}{u + 4} \right) \left( \lim_{u \to 0^+} \frac{1}{u} \right) = \frac{-1}{4} \cdot \infty = \infty.
\]

Of course, there’s plenty of ways to make this more complicated. For one, you first need to see that you might be dealing with an infinite limit; usually you can tell that when you plug in, you get division by zero, but you don’t get \( \frac{0}{0} \) (in which case there would probably be some kind of cancellation). This works for rational and algebraic functions, but there are other functions that have vertical asymptotes. For example, in \( \lim_{x \to 0} \frac{1}{\sin^2(x)} \) we would get division by zero, but
it’s not an algebraic function. Still, we can make the substitution \( u = \sin(x) \), and use the fact that as \( x \to 0 \), also \( \sin(x) \to 0 \) (in other words, \( \lim_{x \to 0} \sin(x) = 0 \)), and we get

\[
\lim_{x \to 0} \frac{1}{\sin^2(x)} = \lim_{u \to 0} \frac{1}{u^2} = \infty.
\]

Later we’ll also see that \( \lim_{x \to 0^+} \ln(x) = -\infty \), and that one you can’t recognize as division by zero.

### 3.2 Limits at Infinity

**Definition:** We say

\[
\lim_{x \to \infty} f(x) = L
\]

if as \( x \) gets bigger and bigger, \( f(x) \) approaches \( L \).

We can make the same definition for \( \lim_{x \to -\infty} \). Note again that \( \infty \) is not a number, just notation.

The basic examples are \( \lim_{x \to \infty} \frac{1}{x} = 0 \) and \( \lim_{x \to -\infty} \frac{1}{x} = 0 \), which say that as \( x \) gets bigger and bigger, \( \frac{1}{x} \) gets closer and closer to 0. More generally, for any number \( a > 0 \) we have

\[
\lim_{x \to \infty} \frac{1}{x^a} = 0,
\]

because as \( x \) grows large, \( x^a \) grows large as well, so \( \frac{1}{x^a} \) becomes small.

Almost the same thing is true for \( x \to -\infty \), except that we would have to avoid taking even roots of negative numbers.

If a limit at infinity exists and equals some number \( L \), then we say that the line \( y = L \) is a horizontal asymptote of the graph of the function. See Stewart 2.6 for several examples and pictures.

The same rules for interchanging apply, except that now we do not want to end up with \( \lim_{x \to \pm \infty} x \), but with \( \lim_{x \to \pm \infty} \frac{1}{x} \), because for those we can just write 0. For instance,

\[
\lim_{x \to \infty} \left( \frac{2}{x^2} + \frac{1}{x} \right) = 2 \left( \lim_{x \to \infty} \frac{1}{x} \right)^2 + \left( \lim_{x \to \infty} \frac{1}{x} \right) = 0^2 + 0 = 0.
\]

However, for limits at infinity we cannot use continuity to justify plugging in, since after all \( \infty \) is not a number, so plugging it in does not even make sense. All that we can do is interchange and then evaluate \( \lim_{x \to \infty} \frac{1}{x} \) to 0 everywhere it occurs (although you could still think of it as ‘plugging in \( \infty \), but you’re not allowed to say that out loud’). In practice, you won’t have to write down the interchanging steps, but you should write down how you get to the \( \frac{1}{x} \)-form. Actually, it’s enough to get to \( \frac{1}{x} \) with \( a > 0 \) everywhere, since we know those evaluate to 0 as well.

For rational functions, we do this by dividing top and bottom by the highest power of \( x \) that occurs, like in

\[
\lim_{x \to \infty} \frac{2x^2 - x + 3}{3x^2 + 5} = \lim_{x \to \infty} \frac{\frac{1}{x^2} \cdot 2x^2 - x + 3}{\frac{1}{x^2} \cdot 3x^2 + 5} = \lim_{x \to \infty} \frac{2 - \frac{1}{x} + \frac{3}{x^2}}{3 + \frac{5}{x^2}}
\]

\[
= \frac{2 - \left( \lim_{x \to \infty} \frac{1}{x} \right) + \left( \lim_{x \to \infty} \frac{3}{x^2} \right)}{3 + \left( \lim_{x \to \infty} \frac{5}{x^2} \right)}
\]

\[
= \frac{2 - 0 + 0}{3 + 0} = \frac{2}{3}.
\]
In general, we see that when the top and bottom polynomials have the same degree, we get
the quotient of the leading coefficients. On the other hand, when the degrees are not the same,
two kinds of things can happen:
\[
\lim_{x \to \infty} \frac{5x^2 + 2}{2x^3 - 1} = \lim_{x \to \infty} \frac{\frac{5}{2} + \frac{2}{x^2}}{2 - \frac{1}{x^3}} = \frac{0 + 0}{2 - 0} = 0,
\]
\[
\lim_{x \to \infty} \frac{2x^3 - 1}{5x + 2} \left( = \frac{2 - 0}{0 + 0} \right) \text{ does not exist (=} \infty \).
\]

We also need this trick of dividing by a power of \( x \) for some limits of algebraic functions, but
it becomes a bit trickier. For example, (recall \( x^{5/2} = x^2 \sqrt{x} \) and \( x^{-5/2} \cdot \sqrt{y} = \sqrt{x^{-5} \cdot y} \))
\[
\lim_{x \to \infty} \frac{x^2 + \sqrt{x + 2x^5 + 2}}{x^2 \sqrt{x + 3x^2}} = \lim_{x \to \infty} \frac{\frac{1}{x^{5/2}} \cdot x^2 + \sqrt{x + 2x^5 + 2}}{x^2 \sqrt{x + 3x^2}} = \lim_{x \to \infty} \frac{\frac{1}{\sqrt{x}} + \sqrt{\frac{1}{x^2} (x + 2x^5 + 2)}}{1 + \frac{3}{\sqrt{x}}}
\]
\[
= \lim_{x \to \infty} \frac{0 + \sqrt{\frac{1}{x} + 2 + \frac{2}{x}}}{1 + 0} = \frac{\sqrt{0 + 2 + 0}}{1} = \sqrt{2}.
\]

As before, we may also have to rationalize or do some other algebra, before we can do the
interchanging.
Another kind of example is
\[
\lim_{x \to \infty} \left( \sqrt{x^2 + 1} - x \right).
\]
The way to think of it is that as \( x \) becomes very large, the +1 becomes insignificant, and we
sort of approach \( \sqrt{x^2} - x \), which should be zero. Working this out formally requires the good
old conjugate trick:
\[
\lim_{x \to \infty} \left( \sqrt{x^2 + 1} - x \right) \cdot \left( \frac{\sqrt{x^2 + 1} + x}{\sqrt{x^2 + 1} + x} \right) = \lim_{x \to \infty} \frac{\sqrt{x^2 + 1}^2 - x^2}{\sqrt{x^2 + 1} + x} = \lim_{x \to \infty} \frac{1}{\sqrt{x^2 + 1} + x} \cdot \frac{\frac{1}{2}}{x}
\]
\[
= \lim_{x \to \infty} \frac{\frac{1}{x}}{\sqrt{1 + \frac{1}{x^2}} + 1} = \frac{0}{\sqrt{1 + 0 + 1}} = 0.
\]

### 3.3 A nasty example and the Squeeze Theorem

We will give an example of a function whose limit at 0 does not exist, and whose left and right
limit at 0 do not exist and are not infinite one-sided limits either. Then we will see an example
of a limit that does exist but for which we need a new method, the Squeeze Theorem.
First let’s take another look at the graph of the function \( \sin x \):

![Graph of sin(x)](image)

What is \( \lim_{x \to \infty} \sin x \)? It depends on how \( x \) approaches \( \infty \): if for instance we let \( x \) go to \( \infty \)
over the peaks, where the value is always 1, the limit would be 1. But if \( x \) goes over the valleys,
where the value is always \(-1\), the limit would be \(-1\). So different approaches of \( x \) give different
values for the limit, hence the limit does not exist, and is not \( \infty \) or \(-\infty \) either. But because
this is a limit at infinity, there is no left/right distinction.
Now consider $\sin\left(\frac{1}{x}\right)$, whose graph to the right of zero looks like

What is $\lim_{x \to 0^+} \sin\left(\frac{1}{x}\right)$? In the same way as for $\sin x$, different approaches of $x$ give different limits, so this right limit does not exist, and it is not an infinite limit, either.

Let’s go one step further and consider $x \cdot \sin\left(\frac{1}{x}\right)$:

It seems that $\lim_{x \to 0} \left(x \cdot \sin\left(\frac{1}{x}\right)\right) = 0$, but how do we prove that? Interchanging does not work: since $\lim_{x \to 0} \sin\left(\frac{1}{x}\right)$ does not exist, we can’t write $\lim \left(x \sin\left(\frac{1}{x}\right)\right) = (\lim x) \cdot (\lim \sin\left(\frac{1}{x}\right))$. We also cannot simplify it algebraically.

Right now we need to introduce the third method for determining limits, which will help us handle $x \sin(1/x)$:

**Squeeze Theorem:** Suppose we have $g(x) \leq f(x) \leq h(x)$ in some interval around $a$, and

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

Then we can conclude that

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

Note that this is not so much a method for calculating a limit, as it is a method of proving that a certain limit really is what we suspect it is.

- For $f(x) = x \sin(\frac{1}{x})$ we can apply the theorem with $g(x) = -|x|$ and $h(x) = |x|$. Then since $-|x| \leq x \leq |x|$ and $-1 \leq \sin\left(\frac{1}{x}\right) \leq 1$ we have

$$-|x| \leq x \cdot \sin\left(\frac{1}{x}\right) \leq |x|$$

on all of $\mathbb{R}$, so the condition of the theorem is satisfied, with $L = 0 = \lim_{x \to 0} -|x| = \lim_{x \to 0} |x|$. Hence we conclude that $\lim_{x \to 0} \left(x \cdot \sin\left(\frac{1}{x}\right)\right) = 0$.

- Another example is

$$\lim_{x \to 1} \sqrt{x^2 - 1} \cos\left(\frac{2}{x - 1}\right).$$

We can take $g(x) = \sqrt{x^2 - 1}$ and $h(x) = -\sqrt{x^2 - 1}$, then we have $g(x) \leq f(x) \leq h(x)$ since $-1 \leq \cos\left(\frac{2}{x - 1}\right) \leq 1$. Since $\lim_{x \to 1} g(x) = \lim_{x \to 1} h(x) = 0$, we get that

$$\lim_{x \to 1} \sqrt{x^2 - 1} \cos\left(\frac{2}{x - 1}\right) = 0.$$
In all honesty, in this course we won’t see other applications of the Squeeze Theorem than functions involving \( \sin\left(\frac{1}{x}\right) \), \( \cos\left(\frac{1}{x}\right) \) or some variation like \( \cos\left(\frac{2}{x-1}\right) \). What’s most important is to get a sense of such functions, because they provide our only examples of certain possibilities. For instance, \( \lim_{x \to 0^+} \sin(1/x) \) is a one-sided limit that doesn’t exist but doesn’t equal \( \infty \) or \( -\infty \). Or \( x^2 \sin(1/x) \) is a differentiable function whose derivative is not continuous; I haven’t shown that but you can check for yourself.
Chapter 4

Tangent Lines

4.1 Tangent Lines

Definition. The tangent line to a graph \( y = f(x) \) at the point \((a, f(a))\) is the line through that point with slope

\[
m = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},
\]

if that limit exists.

Recall from the Introduction where this formula came from. Inside the limit is the slope of the secant line through \((x, f(x))\) and \((a, f(a))\), and as \(x \to a\), the points come closer and closer, and the slope of the secant line approaches the slope of the tangent line.

Example. To find the equation of the tangent line to \( y = \sqrt{x + 1} \) at \((3, 2)\), we compute its slope

\[
m = \lim_{x \to 3} \frac{\sqrt{x + 1} - \sqrt{3 + 1}}{x - 3} = \lim_{x \to 3} \frac{\sqrt{x + 1} - 2}{x - 3} \cdot \frac{x + 2}{\sqrt{x + 1} + 2} = \lim_{x \to 3} \frac{1}{\sqrt{x + 1} + 2} = \frac{1}{\sqrt{3 + 1} + 2} = \frac{1}{4}.
\]

So the tangent line has equation \( y = \frac{1}{4}x + b \), and plugging in \((3, 2)\) and solving for \(b\) gives \(2 = \frac{1}{4}(3) + b \Rightarrow b = 2 - \frac{3}{4} = \frac{5}{4}\). Hence the equation of the tangent line is \( y = \frac{1}{4}x + \frac{5}{4}\).

Example. Now for \( y = x^3 + 8 \), at the point \((-2, 0)\). We’ll have to use polynomial division to factor \(x^3 + 8 = (x + 2)(x^2 - 2x + 4)\). Then

\[
m = \lim_{x \to -2} \frac{(x^3 + 8) - ((-2)^3 + 8)}{x - (-2)} = \lim_{x \to -2} \frac{x^3 + 8}{x + 2} = \lim_{x \to -2} \frac{(x + 2)(x^2 - 2x + 4)}{x + 2} = \lim_{x \to -2} (x^2 - 2x + 4) = 12.
\]

So the tangent line has equation \( y = 12x + b \), and plugging in \((-2, 0)\) and solving for \(b\) gives \(0 = 12(-2) + b \Rightarrow b = 24\). Hence the equation of the tangent line is \( y = 12x + 24\).
Example. Let’s find the tangent line at a general point \((a, a^2 + 1)\) on the graph of \(f(x) = x^2 + 1\). The slope is

\[
m_a = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{(x^2 + 1) - (a^2 + 1)}{x - a} = \lim_{x \to a} \frac{x^2 - a^2}{x - a} = \lim_{x \to a} \frac{(x - a)(x + a)}{x - a} = \lim_{x \to a} (x + a) = 2a.
\]

So now we know that for any point on the graph, the slope of the tangent line is \(2a\). Then we can also find the equation by plugging \((a, a^2 + 1)\) into \(y = 2ax + b\) and solving for \(b\):

\[
a^2 + 1 = 2a \cdot a + b \Rightarrow b = a^2 + 1 - 2a^2 = 1 - a^2.
\]

Therefore the equation of the tangent line is \(y = 2ax + (1 - a^2)\).

4.2 The Derivative

In the last example above, we saw that it might not be too hard to give a formula \((2a)\) for the slope of the tangent line at all points \(a\) of the graph at once. We can consider that formula as a function \((2x)\), which is the main object of study in differential calculus.

**Definition.** The derivative of a function \(f\) is the function \(f'\) that gives the slope \(f'(a)\) of the tangent line to the graph \(y = f(x)\) at a point \((a, f(a))\), for each \(a\) in the domain of \(f\) at which the tangent line exists.

Moreover precisely, the derivative is given by the limit that defines the slope of the tangent line (if it exists):

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

The second formula is new, so let us check that they are the same. In the first, we let \(x\) approach \(a\), in the second \(h\) approaches 0, which means \(a + h\) approaches \(a\). So we replaced \(x\) by \(a + h\), \(x - a\) by \((a + h) - a = h\), and \(x \to a\) by \(a + h \to h\), which is the same as \(a \to 0\). There are several different notations for the derivative, see p.157 of Stewart. The only one that I often use is

\[
\frac{dy}{dx} = \frac{d}{dx} f(x) = f'(x), \quad \text{if} \quad y = f(x).
\]

There are (sometimes) two advantages to this notation: you do not need to name the function, which is convenient when you’re talking about a graph \(y = x^2 + 1\) without defining \(f(x) = x^2 + 1\); and it emphasizes the variable \(x\) with respect to which you’re taking the derivative, which is useful when there are parameters \((a, b, ...)\) floating around, or if the variable has an unusual name (like in \(\frac{dy}{dt}\)).

See Stewart p.161 for several more notations, including for higher derivatives like

\[
f''(x) = \frac{d}{dx} \left( \frac{d}{dx} f(x) \right) = \frac{d^2 y}{dx^2}.
\]

Finally, instead of saying *take the derivative*, we use the word *differentiate*.

In the third example above we saw that for the function \(f(x) = x^2\), the slope of the tangent line at \((a, a^2)\) is \(2a\), which we can now rephrase as

\[
f'(x) = 2x, \quad \text{or} \quad \frac{d}{dx} x^2 = 2x.
\]
What makes this notation so useful is that there is a kind of *calculation procedure* (which is where the word *calculus* comes from), with which we can differentiate most functions using straightforward rules, without having to compute all those limits. Next week we will learn how to do this.

For now let’s see two more examples of determining derivatives from the limit definition. Take the functions \( f(x) = \frac{1}{x} \) and \( g(x) = 1 + \sqrt{x} \). Then we work out the limits much like in the examples above:

\[
f'(a) = \lim_{x \to a} \frac{\frac{1}{x} - \frac{1}{a}}{x - a} = \lim_{x \to a} \frac{\frac{1}{x} - \frac{1}{a}}{x - a} = \lim_{x \to a} \frac{1}{xa} = \lim_{x \to a} \frac{1}{xa} = \frac{-1}{a^2},
\]

\[
g'(a) = \lim_{x \to a} \frac{(1 + \sqrt{x}) - (1 + \sqrt{a})}{x - a} = \lim_{x \to a} \frac{\sqrt{x} - \sqrt{a}}{x - a} \cdot \frac{\sqrt{x} + \sqrt{a}}{\sqrt{x} + \sqrt{a}}
\]

\[
= \lim_{x \to a} \frac{x - a}{(x - a)(\sqrt{x} + \sqrt{a})} = \lim_{x \to a} \frac{1}{\sqrt{x} + \sqrt{a}} = \frac{1}{2\sqrt{a}}.
\]

To understand derivatives better, it is useful to consider what the graph of the derivative \( f' \) looks like, given the graph of the original function \( f \). I will not draw pictures, but summarize which properties of the graph of \( f \) translate into which properties of the graph of \( f' \). All of these can be justified from the interpretation of the derivative as the slope of the tangent line, the last few are clarified in the next section.

- If \( f \) is at a maximum or minimum (peak or valley), then \( f' = 0 \) (intersects the \( x \)-axis).
- If \( f \) is increasing (going up), then \( f' \) is positive (above the \( x \)-axis).
- If \( f \) is decreasing, then \( f' \) is negative.
- The steeper \( f \) (up or down), the larger \( f' \) (positive or negative).
- If \( f \) has a discontinuity, then \( f' \) does not exist.
- If \( f \) has a corner, then \( f' \) does not exist and has a jump discontinuity there.
- If \( f \) has a vertical tangent line, then \( f' \) has a vertical asymptote.

### 4.3 Differentiability

Above we defined the tangent line by a limit, but only if that limit exists. It can of course happen that that limit does not exist, and we introduce the following word for that.

**Definition.** We say that a function \( f \) is **differentiable** at \( a \) if the tangent line at \( a \) exists, i.e. if the limit defining the slope of the tangent line exists. We say it is differentiable on an interval \( I \) if it is differentiable at all numbers \( a \) in \( I \).

There are 3 basic examples of functions that are not differentiable at some point.

- **A corner:** the typical example of non-differentiability is the function \( f(x) = |x| \) at \( x = 0 \), for which the slope of the tangent line would be given by

\[
\lim_{x \to 0} \frac{|x| - |0|}{x - 0} = \lim_{x \to 0} \frac{|x|}{x},
\]

which does not exist because the left and right limit are different (−1 and 1). In the graph of \( |x| \), this is easy to recognize: at \( x = 0 \) there is a 'corner' where there couldn’t be a tangent line.
• **A vertical asymptote:** another example is \( g(x) = \sqrt[3]{x} = x^{\frac{1}{3}} \) at \( x = 0 \). Here the slope of the tangent line would be

\[
\lim_{x \to 0} \frac{x^{\frac{1}{3}} - 0}{x - 0} = \lim_{x \to 0} \frac{1}{x^{\frac{2}{3}}} = \left( \lim_{x \to 0} \frac{1}{x} \right)^{\frac{2}{3}} = \infty,
\]

which means the limit does not exist, hence neither does the tangent line. Again we can see in the graph what this means: if there was a tangent line, it would have to be vertical. With our current definition, we do not allow vertical tangent lines, because we do not want to have to deal with the slope being infinite.

• **A discontinuity:** the last kind of example is any function with a discontinuity. Take for example the function

\[
h(x) = \begin{cases} 
    x & \text{if } x < 1 \\
    x + 1 & \text{if } x \geq 1
\end{cases},
\]

which has a jump discontinuity at \( x = 1 \). Now if we try to compute the left limit (so we use \( h(x) = x \), but \( h(1) = 2 \)) of the slope of the tangent line at \( x = 1 \), we get

\[
\lim_{x \to 1^-} \frac{h(x) - h(1)}{x - 1} = \lim_{x \to 1^-} \frac{x - 2}{x - 1} = -\infty.
\]

What happens is that as \( x \) approaches 1 from the left, the secant line doesn’t approach a tangent line at all, it approaches a vertical line which is clearly not a tangent line. But note that the limit from the right does exist, and equals 1.

The last example leads to a more general point. For the secant/tangent picture to work at all, the function has to be continuous. We can state and prove this formally: (this is something mathematicians like to do a lot, and although in this course we try to spare you, you should still see some examples)

**Fact:** If \( f \) is differentiable at \( a \), then \( f \) is continuous at \( a \).

**Proof:** If \( f \) is differentiable at \( a \), then \( \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \) exists. But as \( x \to a \), the denominator \( x - a \) goes to zero. Then the numerator \( f(x) - f(a) \) must go to zero as well, otherwise the limit would be infinite. So as \( x \to a \) we have \( f(x) - f(a) \to 0 \), which is equivalent to \( f(x) \to f(a) \). This exactly means \( f \) is continuous at \( a \).

Turning this statement around, when \( f \) is not continuous at \( a \), then it also is not differentiable at \( a \). But it is not true that if a function is continuous, it must also be differentiable: \(|x|\) for instance is continuous, but not differentiable.

### 4.4 Rates of Change and Velocities

There are different ways of looking at the slope of the tangent line, which I will quickly introduce here. For many more examples, see Stewart 2.7 and the suggested exercises.

Think of \( f(x) \) as some quantity for which we care about how it changes as a function of \( x \). There are two different ways of measuring the rate of change: we could measure the average rate of change on some interval \([x_1, x_2]\):

\[
\frac{f(x_2) - f(x_1)}{x_2 - x_1},
\]

or we could measure the instantaneous rate of change at some \( x_1 \):

\[
\lim_{x \to x_1} \frac{f(x) - f(x_1)}{x - x_1}.
\]
For instance, \( p(t) \) could be the position function of a car, a running athlete, or a moving particle, as a function of time. Then the \textit{average velocity} between \( t_2 \) and \( t_1 \) is given by \( \frac{p(t_2) - p(t_1)}{t_2 - t_1} \), the distance travelled divided by the time spent. On the other hand, the \textit{velocity} (which is the same as instantaneous velocity) at time \( t_1 \) is given by

\[
v(t_1) = \lim_{t \to t_1} \frac{p(t) - p(t_1)}{t - t_1}.
\]

### 4.5 Tangent Line Problems

There is a large variety of trickier problems involving tangent lines, and usually one of these shows up on the exam. I will give examples of the most typical ones here, although I’ll only work out one of them, since they will become easier once we can use the differentiation rules.

- Find the line through \((1, 0)\) that is tangent to \(y = \frac{1}{x}\).
- Find the line through \((0, 0)\) that is normal to \(y = \frac{1}{x}\).
- Find the line that is tangent to both \(y = x^2\) and \(y = x^2 - 2x + 2\).
- Find the line that is tangent to \(y = x^4 - 2x^2 - x\) in two different points.

Note that a graph has lots of tangent lines, one for each point. But in these questions, we are asked for a tangent line satisfying some extra condition, like going through some other point not on the graph.

I’ll do the first one. Let the line be given by

\[
y = mx + b.
\]

Then since it goes through the point \((1, 0)\), we get \(0 = m \cdot 1 + b\), so \(b = -m\) and we can assume the equation is

\[
y = mx - m.
\]

Now we know the line is tangent to \(y = \frac{1}{x}\) at some point, but we don’t know which one yet, so let’s name it \((a, \frac{1}{a})\) for now. Above we computed the derivative \(f'(x) = -\frac{1}{x^2}\), and that gives us the slope of the tangent line at the point \((a, \frac{1}{a})\), namely \(m = \frac{-1}{a^2}\). Hence the equation of our line looks like

\[
y = -\frac{1}{a^2} x + \frac{1}{a^2}.
\]

Finally, the line has to actually go through the point \((a, \frac{1}{a})\), so we have to have \(\frac{1}{a} = -\frac{1}{a^2} a + \frac{1}{a^2}\). Multiplying by \(a^2\) gives \(a = -a + 1\), so \(a = \frac{1}{2}\). Hence the tangent line we’re looking for is

\[
y = -4x + 4.
\]

So we have to determine the line by using all the information we’re given step-by-step. In this case, it has to pass through \((1, 0)\), meet the graph in some undetermined point \((a, f(a))\), and it has to be tangent to the graph at that point, so have slope \(f'(a)\).

The other questions are similar, but a bit more complicated. For the second, we have to know what \textit{normal} means: given a line with slope \(m\), the line normal to it is the line with slope \(\frac{1}{m}\). And the line normal to a graph (at a certain point) is the line that is normal to the tangent line to the graph (at that point). So the normal line to the graph \(y = f(x)\) at the point \((a, f(a))\) has slope \(\frac{1}{f'(a)}\), and the equation of the normal line can then be found just like for the tangent line.

The last two questions are quite a bit harder, since they involve two undetermined points, which we’ll both have to name.
Chapter 5
Differentiation

5.1 Polynomials

We will introduce the differentiation rules, which will help us compute the derivative of just about any function, without having to do the limit calculations. However, to justify these rules, and to understand them, we will have to show the limit calculations that are behind them, which we do in the proofs. You won’t have to be able to reproduce these proofs, or much less come up with them yourselves. But you can still learn a lot from trying to understand them.

Note that these rules are only valid at values of $x$ where the functions involved are differentiable, but we will not mention that every time.

5.1 Polynomials

We start with the mother of all differential rules:

**Power Rule:** \((x^n)' = nx^{n-1}\) for any real number $n$

We could also write this as \(\frac{d}{dx}(x^n) = nx^{n-1}\), or \(f(x) = x^n \Rightarrow f'(x) = nx^{n-1}\). Here are a few examples:

\((x^3)' = 3x^2, \quad (x^{17})' = 17x^{16}, \quad (\sqrt{x})' = x^{\frac{1}{2}} = \frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2\sqrt{x}},\)

\((\frac{1}{x})' = (x^{-1})' = (-1) \cdot x^{-1-1} = -\frac{1}{x^2}, \quad \left(\frac{1}{x^{3/2}}\right)' = -\frac{3}{2} \cdot \frac{1}{x^{5/2}}, \quad (x^\pi)' = \pi x^{\pi-1}.

Proving this rule, for any real number, is not so easy (we’ll do it in Ch. 8), but here we will do it for $n$ a positive integer. We already saw this for $x^2$, and this is what it looks like for $f(x) = x^3$:

\[f'(a) = \lim_{x \to a} \frac{x^3 - a^3}{x - a} = \lim_{x \to a} \frac{(x - a)(x^2 + ax + a^2)}{x - a} = \lim_{x \to a} (x^2 + ax + a^2) = 3a^2.\]

Here we used the factorization \((x^3 - a^3) = (x - a)(x^2 + ax + a^2)\), which may not be easy to come up with, but is easy to check:

\[x(x^2 + ax + a^2) - a(x^2 + ax + a^2) = (x^3 + ax^2 + a^2x) - (ax^2 + a^2x + a^3) = x^3 - a^3.\]
Proofs:

For the Constant Rule, let

\[
(f - g)' = \lim_{x \to a} \frac{f(x) - g(x)}{x - a}
\]

Of course the Sum Rule, together with the Constant Multiple Rule, implies a similar rule for subtractions: \((f - g)' = (f + (-g))' = f' + (-g)' = f' - g'\).

Proofs: For the Constant Rule, let \(f(x) = c\). Then

\[
f'(a) = \lim_{x \to a} \frac{c - c}{x - a} = \lim_{x \to a} 0 = 0.
\]

Note that the limit is not \(0\), because \(c - c\) is always 0, while \(x - a\) is just approaching 0.

The Constant Multiple rule is pretty easy:

\[
(cf)'(a) = \lim_{x \to a} \frac{cf(x) - cf(a)}{x - a} = \lim_{x \to a} c \frac{f(a) - f(a)}{x - a} = cf'(a).
\]

Finally, the Sum Rule is a consequence of the fact that we can interchange limits and sums:

\[
(f + g)'(a) = \lim_{x \to a} \frac{(f(x) + g(x)) - (f(a) + g(a))}{x - a}
= \lim_{x \to a} \left( \frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a} \right)
= \left( \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \right) + \left( \lim_{x \to a} \frac{g(x) - g(a)}{x - a} \right)
= f'(a) + g'(a).
\]
With these rules we can differentiate any polynomial:

\[(x^3 - 2x + 5)' = (x^3)' + (2x)' + (5)'
\]
\[= 3x^2 + 2(x)' + 0
\]
\[= 3x^2 + 2,
\]
\[(5x^7 + 3x^6 - x^2) = 5(x^7)' + 3(x^6)' - (x^2)'
\]
\[= 35x^6 + 18x^5 - 2x
\]

In fact, we can already do more than polynomials:

\[(3\sqrt{x} + \frac{5}{x^2} + 7)' = 3 \cdot \frac{1}{2\sqrt{x}} + 5 \cdot \frac{-2}{x^3} + 0 = \frac{3}{2\sqrt{x}} - \frac{10}{x^3}.
\]

5.2 The Product Rule

Let’s see what happens when we differentiate the product of two functions. When we saw the sum rule, it said that we could interchange sums with differentiation, just like with limits. For products, that is not at all true: \((fg)' \neq f'g'\)! Instead, we have:

**Product Rule:**  \((fg)' = f'g + fg'\)

To see why the derivative of products takes this surprising form, we will go through the proof, which is a step harder than the previous ones:

**Proof:** The trick is to add \(0 = (f(a)g(x) - f(a)g(x))\) in the middle of the numerator; again this is not easy to come up with, but we see that it works.

\[
(fg)'(a) = \lim_{x \to a} \frac{f(x)g(x) - f(a)g(a)}{x - a}
\]
\[= \lim_{x \to a} \frac{f(x)g(x) - (f(x)g(x) - f(a)g(x)) - (f(a)g(x) - f(a)g(a))}{x - a}
\]
\[= \lim_{x \to a} \frac{f(x)g(x) - f(a)g(x)}{x - a} + \frac{(f(a)g(x) - f(a)g(a))}{x - a}
\]
\[= \left( \lim_{x \to a} f(x) \cdot \frac{g(x) - f(a)}{x - a} \right) + \left( \lim_{x \to a} g(x) \cdot \frac{f(x) - f(a)}{x - a} \right)
\]
\[= g(a)f'(a) + f(a)g'(a).
\]

Note that in the second to last step we used the fact that we can interchange limits and products, and in the last step we used the continuity of \(g\) to get \(\lim_{x \to a} g(x) = g(a)\). We know that \(g\) is continuous because we’re assuming it to be differentiable (see section 4.2).

Now we can use the product rule to compute for instance

\[
((x^2 - 2) \cdot (x^3 - 2))' = (x^2 - 2)' \cdot (x^3 - 2) + (x^2 - 2) \cdot (x^3 - 2)'
\]
\[= 2x \cdot (x^3 - 2) + (x^2 - 2) \cdot 3x^2
\]
\[= 2x^4 - 4x + 3x^4 - 6x^2 = 5x^4 - 6x^2 - 4x.
\]
We can actually check this, by first multiplying out and then differentiating the polynomial:

\[
\left( (x^2 - 2) \cdot (x^3 - 2) \right)' = (x^5 - 2x^3 - 2x^2 + 4)'
= 5x^4 - 6x^2 - 4x.
\]

In this case, the second way was actually easier, but later on we’ll see examples where we can’t do without the product rule.

5.3 The Quotient Rule

Now let’s turn to division. As a starter, let’s differentiate \( f(x) = \frac{1}{x} \) without using the power rule:

\[
f'(a) = \lim_{x \to a} \frac{\frac{1}{x} - \frac{1}{a}}{x - a} = \lim_{x \to a} \frac{\frac{a-x}{xa}}{x - a} = \lim_{x \to a} \frac{1}{xa} \frac{- (x - a)}{x - a} = \lim_{x \to a} \frac{-1}{xa} = \frac{-1}{a^2},
\]

which is the same as we deduced from the power rule before. In a similar way, we can get our next rule:

**Reciprocal Rule:**

\[
\left( \frac{1}{f} \right)' = - \frac{f'}{f^2}
\]

So this says that \( \frac{d}{dx} \frac{1}{f(x)} = - \frac{f'(x)}{f(x)^2} \). Let’s see the proof.

**Proof:**

\[
\left( \frac{1}{f} \right)'(a) = \lim_{x \to a} \frac{\frac{1}{f(x)} - \frac{1}{f(a)}}{x - a} = \lim_{x \to a} \frac{\frac{f(a)-f(x)}{f(x)f(a)}}{x - a}
= \lim_{x \to a} \frac{1}{f(x)f(a)} \frac{- (f(x) - f(a))}{x - a}
= \left( \lim_{x \to a} \frac{1}{f(x)f(a)} \right) \cdot \left( - \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \right)
= \frac{1}{f(a)^2} \cdot (- f'(a)).
\]

Here’s an example, with \( f(x) = x^n \):

\[
\left( \frac{1}{x^n} \right)' = \frac{(x^n)'}{(x^n)^2} = - \frac{nx^{n-1}}{x^{2n}} = - \frac{n}{x^{n+1}},
\]

which we can check by using the Power Rule

\[
\left( \frac{1}{x^n} \right)' = (x^{-n})' = -nx^{-n-1} = - \frac{n}{x^{n+1}}.
\]

In this case the power rule turned out to be easier, but in the next example it would be helpless:

\[
\left( \frac{1}{x^2 + 1} \right)' = - \frac{(x^2 + 1)'}{(x^2 + 1)^2} = - \frac{2x}{(x^2 + 1)^2}.
\]
Now we can do any rational function \( f(x) / g(x) \), by writing it as \( f(x) \cdot \frac{1}{g(x)} \) and applying the product rule and the reciprocal rule. For instance:

\[
\left( \frac{x^3 + 1}{x^2 + 2} \right)' = \left( (x^3 + 1) \cdot \frac{1}{x^2 + 2} \right)' = (x^3 + 1)' \cdot \frac{1}{x^2 + 2} + (x^3 + 1) \cdot \left( \frac{1}{x^2 + 2} \right)'
\]

\[
= \frac{3x^2}{x^2 + 2} + (x^3 + 1) \cdot \left( \frac{-2x}{(x^2 + 2)^2} \right)
\]

\[
= \frac{3x^2(x^2 + 2)}{(x^2 + 2)^2} - \frac{2x(x^3 + 1)}{(x^2 + 2)^2}
\]

\[
= \frac{3x^4 + 6x^2 - 2x^4 - 2x}{(x^2 + 2)^2}
\]

\[
= \frac{x^4 + 6x^2 - 2x}{(x^2 + 2)^2}.
\]

Doing this in general we get the next rule, whose proof we won’t write out here:

\[
\text{Quotient Rule:} \quad \left( \frac{f}{g} \right)' = \frac{f'g - fg'}{g^2}
\]

Let’s do the example above again, but now straight from the Quotient Rule:

\[
\left( \frac{x^3 + 1}{x^2 + 2} \right)' = \frac{(x^3 + 1)'(x^2 + 2) - (x^3 + 1)(x^2 + 2)'}{(x^2 + 2)^2}
\]

\[
= \frac{3x^2 \cdot (x^2 + 2) - (x^3 + 1) \cdot 2x}{(x^2 + 2)^2}
\]

\[
= \frac{x^4 + 6x^2 - 2x}{(x^2 + 2)^2}.
\]

It’s also good to see that the Reciprocal Rule is just a special case of the Quotient Rule:

\[
\left( \frac{1}{g(x)} \right)' = \frac{1' \cdot g(x) - 1 \cdot g'(x)}{g(x)^2} = \frac{0 \cdot g(x) - g'(x)}{g(x)^2} = \frac{-g'(x)}{g(x)^2}.
\]

### 5.4 The Chain Rule

So we know how to do rational functions and roots (since they’re powers), does that mean that we can do all algebraic functions? No, because of functions like

\[
F(x) = \sqrt{x^2 + 1}, \quad \sqrt{x^2 + \sqrt{x}}.
\]

They are compositions of functions \((F = f \circ g, \ f = \sqrt{x}, \ g(x) = x^2 + 1)\), and even though we know the derivatives of those \((f' = \frac{1}{2x}, \ g'(x) = 2x)\), we don’t know the derivative of the composition (warning: it is not the composition of the derivatives). The solution is the next rule, which is a bit more complicated than the others, and whose proof we will not do:
**Chain Rule:** If \( F(x) = f(g(x)) \), then \( F'(x) = f'(g(x)) \cdot g'(x) \)

The function \( F(x) = \sqrt{x^2 + 1} \) we can write as a composition \( f \circ g \) with \( f(x) = \sqrt{x} \), \( g(x) = x^2 + 1 \), whose derivatives are \( f'(x) = \frac{1}{2\sqrt{x}} \) and \( g'(x) = 2x \), so that applying the Chain Rule gives:

\[
(\sqrt{x^2 + 1})' = f'(g(x)) \cdot g'(x) = \frac{1}{2\sqrt{g(x)}} \cdot 2x = \frac{x}{\sqrt{x^2 + 1}}.
\]

For \( F(x) = \sqrt{x^2 + \sqrt{x}} \) we have \( F = f \circ g \) with \( f(x) = \sqrt{x} \), \( g(x) = x^2 + \sqrt{x} \), so \( f'(x) = \frac{1}{2\sqrt{x}} \) and \( g'(x) = 2x + \frac{1}{2\sqrt{x}} \). Then

\[
F'(x) = \frac{1}{2\sqrt{g(x)}} \cdot g'(x) = \frac{1}{2\sqrt{x^2 + \sqrt{x}}} \cdot (2x + \frac{1}{2\sqrt{x}}) = \frac{(2x + \frac{1}{\sqrt{x}})}{2\sqrt{x^2 + \sqrt{x}}}.
\]

We can also prove the Reciprocal Rule using the Chain Rule and the Power Rule. Let \( F(x) = \frac{1}{g(x)} = f(g(x)) \), with \( f(x) = \frac{1}{x} \), \( f'(x) = \frac{-1}{x^2} \). Then

\[
\left(\frac{1}{g(x)}\right)' = F'(x) = f'(g(x)) \cdot g'(x) = -\frac{1}{g(x)^2} \cdot g'(x) = \frac{-g'(x)}{g(x)^2},
\]

just as we derived before from the limit definition.

In the \( \frac{dy}{dx} \) notation, the Chain Rule is especially easy to remember. Let \( y = F(x) = f(g(x)) \), and write \( y = f(u) \), \( u = g(x) \). Then \( f'(u) = \frac{dy}{du} \) and \( g'(x) = \frac{du}{dx} \), so we could write the Chain Rule as

\[
\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.
\]

In fact this looks a little bit like a proof: just cancel the \( du \)'s. But this is not possible, because the \( du \)'s are not separate objects and these are not really fractions, that’s just notation, as you can see when you look at it in terms of \( f \) and \( g \).
Chapter 6

Trigonometric Functions

6.1 Definitions

From high school, you probably know the functions \( \sin(t) \) and \( \cos(t) \) as defined by

\[
\cos(t) = \frac{\text{adj}}{\text{hyp}}, \quad \sin(t) = \frac{\text{opp}}{\text{hyp}},
\]

where adj, hyp and opp are the lengths of the adjacent, hypotenuse and opposite sides of a right-angled triangle with one angle named \( t \):

Here we do things a bit differently. First of all, the input of the functions \( \sin(t) \) and \( \cos(t) \) will always be in radians, although we may still describe angles in degrees. Recall that to change between radians and degrees, all you need to know is

\[
\pi \text{ radians} = 180^\circ.
\]

For instance, a right angle is \( 90^\circ = \frac{1}{2} \cdot 180^\circ = \frac{\pi}{2} \) radians.

Next, in the triangle above the angle \( t \) has to be less than \( \frac{\pi}{2} \), otherwise the hypotenuse wouldn’t intersect the opposite side anymore. But there is a natural way of defining the trigonometric functions that allows the \( t \) to be any number. Consider a point \( P_t \) on a circle with radius 1, such that the line through \( P \) and the origin makes an angle \( t \) with the \( x \)-axis.
This picture is where radians come from: the length along the circle from (1, 0) to the point \( P_t \) is exactly the number of radians we associate to the angle \( t \). This often involves \( \pi \) because the circumference of the circle is \( 2\pi \), which is why an angle of 180° (halfway around the circle) equals \( \pi \).

Now we define the trigonometric functions using this picture:

\[
\begin{align*}
\cos(t) &= \text{the } x\text{-coordinate of } P_t \\
\sin(t) &= \text{the } y\text{-coordinate of } P_t.
\end{align*}
\]

If you draw the line straight down from \( P_t \) to the \( x \)-axis, then you can see the same triangle as above, but now we can also make the angle more than \( \pi/2 \), or even negative.

The main way of understanding these functions is by their graphs, which look like this:

From these graphs we can read off several properties (we’ll switch to writing the variable as \( x \) from now on):

<table>
<thead>
<tr>
<th>Property</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boundedness:</strong></td>
<td>(-1 \leq \sin(x) \leq 1, \quad -1 \leq \cos(x) \leq 1)</td>
</tr>
<tr>
<td><strong>Periodicity:</strong></td>
<td>( \cos(x + 2\pi) = \cos(x), \quad \sin(x + 2\pi) = \sin(x) )</td>
</tr>
<tr>
<td>( \cos(t) \text{ is even} )</td>
<td>( \cos(-x) = \cos(x) )</td>
</tr>
<tr>
<td>( \sin(t) \text{ is odd} )</td>
<td>( \sin(-x) = -\sin(x) )</td>
</tr>
<tr>
<td><strong>Shift Property:</strong></td>
<td>( \cos(x) = \sin(x + \frac{\pi}{2}), \quad \sin(x) = \cos(x - \frac{\pi}{2}) )</td>
</tr>
</tbody>
</table>

Certain values of these functions have easy forms, which are worth remembering:

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>( \frac{\pi}{6} )</th>
<th>( \frac{\pi}{4} )</th>
<th>( \frac{\pi}{3} )</th>
<th>( \frac{\pi}{2} )</th>
<th>( \frac{5\pi}{6} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sin(x) )</td>
<td>0</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{\sqrt{2}}{2} )</td>
<td>( \frac{\sqrt{3}}{2} )</td>
<td>1</td>
<td>( -\frac{1}{2} )</td>
</tr>
<tr>
<td>( \cos(x) )</td>
<td>1</td>
<td>( \frac{\sqrt{3}}{2} )</td>
<td>( \frac{\sqrt{2}}{2} )</td>
<td>( \frac{1}{2} )</td>
<td>0</td>
<td>( -\frac{\sqrt{2}}{2} )</td>
</tr>
</tbody>
</table>

Of course, \( \frac{\sqrt{1}}{2} = \frac{1}{2} \) and \( \frac{\sqrt{3}}{2} = \frac{1}{\sqrt{2}} \), but I think it’s easier to remember this way. Note that also \( \frac{\sqrt{0}}{2} = 0 \) and \( \frac{\sqrt{4}}{2} = 1 \).

There is another important trigonometric function:

\[
\tan(x) = \frac{\sin(x)}{\cos(x)}
\]
This function is convenient because it equals \( \frac{\text{opp}}{\text{adj}} \), which means it relates the angle \( t \) with the slope of the hypotenuse. To see this, suppose we have any line \( y = mx + b \). It intersects the \( x \)-axis at \( A = (-\frac{b}{m}, 0) \), and the \( y \)-axis at \( B = (0, b) \). Then \( A \), \( B \) and \( O = (0, 0) \) form a right-angled triangle, so if we let \( t \) be the angle between the line and the \( x \)-axis, we get (with \( |BO| = b \) the length of the line segment between \( B \) and \( O \), and \( |AO| = b/m \))

\[
\tan(t) = \frac{\text{opp}}{\text{adj}} = \frac{|BO|}{|AO|} = \frac{b}{b/m} = m,
\]
or in other words,

slope of a line = \( \tan(\text{angle with } x\text{-axis}) \).

Finally there are three other trigonometric functions, which are just reciprocals of the three above. Personally, I think they are useless, but they might occur on the final exam. Just try to remember which functions they are reciprocals of, and when you encounter one in a question, rewrite them in terms of \( \cos(x) \), \( \sin(x) \) and \( \tan(x) \), and never look back.

\[
\sec(x) = \frac{1}{\cos(x)}, \quad \csc(x) = \frac{1}{\sin(x)}, \quad \cot(x) = \frac{1}{\tan(x)}.
\]
The first one is the \textit{secant} function, the second the \textit{cosecant}, and the third is the \textit{cotangent}.

### 6.2 Identities

Because of their symmetric definition using the circle, the trigonometric functions satisfy numerous identities. We introduce the most important ones.

The most basic trigonometric identity is

\[
\sin^2(x) + \cos^2(x) = 1,
\]

which is nothing but the Pythagorean Theorem for the triangle with sides \( \sin(t) \), \( \cos(t) \) and hypotenuse 1 (the radius of the circle defining the functions).

The most important identities are probably the \textbf{addition formulas}:

\[
\sin(x + y) = \sin(x) \cos(y) + \cos(x) \sin(y)
\]

\[
\cos(x + y) = \cos(x) \cos(y) - \sin(x) \sin(y)
\]

Proving these requires quite a bit of geometry, which we’ll avoid here.

The three formulas above, and the properties from the previous section, are the basis for most trigonometric identities. For instance, putting \( x = y \) in the addition formulas gives the \textbf{double-angle formulas}:

\[
\sin(2x) = 2 \sin(x) \cos(x), \quad \cos(2x) = \cos^2(x) - \sin^2(x).
\]

We can get two different formulas for \( \cos(2x) \) by substituting either \( \sin^2(x) = 1 - \cos^2(x) \) or \( \cos^2(x) = 1 - \sin^2(x) \) into \( \cos(2x) = \cos^2(x) - \sin^2(x) \):

\[
\cos(2x) = 2 \cos^2(x) - 1, \quad \cos(2x) = 1 - 2 \sin^2(x).
\]
From these we can get the often useful **half-angle formulas**:

\[
\cos^2(x) = \frac{1 + \cos(2x)}{2}, \quad \sin^2(x) = \frac{1 - \cos(2x)}{2}.
\]

With these we can for instance derive one of the special values from the table above (using \(\cos(\pi) = 0\), which is easy to see from the circle definition):

\[
\cos\left(\frac{\pi}{4}\right) = \sqrt{\frac{1 + \cos(2 \cdot \frac{\pi}{4})}{2}} = \sqrt{\frac{1 + 0}{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}.
\]

### 6.3 Continuity and limits

We will not prove it here, but from the graphs it is plausible that

\[
\text{sin}(x) \quad \text{and} \quad \text{cos}(x) \quad \text{are continuous functions}.
\]

By the definition of continuity, this implies the basic limits

\[
\lim_{x \to 0} \sin(x) = \sin(0) = 0, \quad \lim_{x \to 0} \cos(x) = \cos(0) = 1.
\]

The following limits are less easy to prove, but are very important, because we will need them to find the derivatives of \(\sin(x)\) and \(\cos(x)\). To prove them, we will have to use the Squeeze Theorem.

\[
\lim_{x \to 0} \frac{\sin(x)}{x} = 1, \quad \lim_{x \to 0} \frac{\cos(x) - 1}{x} = 0.
\]

**Proof:**

We will first prove the one with \(\frac{\sin(x)}{x}\), and then deduce the other one from that. We’ll do this by sandwiching \(\frac{\sin(x)}{x}\) between two other functions whose limits we know, and then using the Squeeze Theorem.

On the right is a part of the picture that defined sin and cos at the beginning of this chapter. There we see that \(\sin t\) is the distance from \(P_t\) to the \(x\)-axis, and the angle \(t\) is also the length of the curve from \((1, 0)\) to \(P_t\) (because the circle has radius 1). We’ve added the tangent line to the circle at \(P_t\), and the distance along that line from \(P_t\) to the \(x\)-axis is \(\tan(t)\). That’s because the tangent line makes a right angle with the radius at \(P_t\), so we get a right triangle with adjacent side 1, so the opposite side has length \(\tan(t)\).

From this picture we can see that

\[
\tan(x) > x > \sin(x),
\]

at least when \(0 < x < \pi/2\), otherwise the picture would be different. If we divide this by \(\sin(x)\), we get (using \(\tan(x) = \frac{\sin(x)}{\cos(x)}\)):

\[
\frac{1}{\cos(x)} > \frac{x}{\sin(x)} > 1.
\]

40
Now if we take the reciprocal (so flip the inequalities), we get:

\[ \cos(x) < \frac{\sin(x)}{x} < 1, \]

for \( 0 < x < \frac{\pi}{2} \). Similarly, we can get that \( \cos(x) < \frac{\sin(x)}{x} < 1 \) when \(-\pi/2 < x < 0\), by observing that these three functions are all even (i.e. \( \cos(x) = \cos(-x) \) and \( \frac{\sin(-x)}{-x} = \frac{\sin(x)}{x} \)), so what is true for \( x \) is also true for \(-x\).

So we have our function \( \frac{\sin(x)}{x} \) squeezed in between two other functions (whose limits we know) on an interval around 0, as the Squeeze Theorem requires. Since we have \( \lim_{x \to 0} 1 = 1 \) and \( \lim_{x \to 0} \cos(x) = \cos(0) = 1 \), we can conclude by the Squeeze Theorem that \( \lim_{x \to 0} \frac{\sin(x)}{x} = 1 \).

For the second limit, we could do something similar, but it’s easier to use the half-angle formula \( \cos(x) = 1 - 2\sin^2(x/2) \) to rewrite it in terms of the first limit:

\[
\lim_{x \to 0} \frac{\cos(x) - 1}{x} = \lim_{x \to 0} \frac{(1 - 2\sin^2(x/2)) - 1}{x} = \lim_{x \to 0} \frac{-2\sin^2(x/2)}{x} = \lim_{x \to 0} \frac{-\sin(x/2)}{x/2} \cdot \lim_{x \to 0} \sin(x/2) = -1 \cdot 0 = 0
\]

**Trig limits**

Because of all the trigonometric identities, limits involving trig functions come with their own set of tricks. To write the second limit above in terms of the first one, we used \( \cos(x) = 1 - 2\sin^2(x/2) \), but we could also have used the following ‘conjugate’ trick, which comes down to \( (1 - \cos(x)) \cdot (1 + \cos(x)) = 1 - \cos^2(x) = \sin^2(x) \):

\[
\lim_{x \to 0} \frac{\cos(x) - 1}{x} \cdot \frac{\cos(x) + 1}{\cos(x) + 1} = \lim_{x \to 0} \frac{\cos^2(x) - 1}{x \cdot (\cos(x) + 1)} = \lim_{x \to 0} \frac{-\sin^2(x)}{x \cdot (\cos(x) + 1)}
\]

\[
= \lim_{x \to 0} \left( \frac{\sin(x)}{x} \cdot \frac{-\sin(x)}{\cos(x) + 1} \right) = \left( \lim_{x \to 0} \frac{\sin(x)}{x} \right) \cdot \left( \lim_{x \to 0} \frac{-\sin(x)}{\cos(x) + 1} \right)
\]

\[
= 1 \cdot 0 = 0
\]

Another common trick is forcing a limit into the form \( \lim_{x \to 0} \frac{\sin(x)}{x} \) by multiplying by a constant:

\[
\lim_{x \to 0} \frac{\sin(5x)}{x} = 5 \cdot \lim_{x \to 0} \frac{\sin(5x)}{5x} = 5 \cdot \lim_{u \to 0} \frac{\sin(u)}{u} = 5 \cdot 1 = 5.
\]

Here we used the substitution \( u = 5x \), but it’s okay to leave that step out.

A more complicated example is

\[
\lim_{x \to 0} \frac{\sin(mx)}{\sin(nx)} = \frac{m}{n} \lim_{x \to 0} \frac{\sin(mx)}{mx} \cdot \frac{nx}{\sin(nx)} = \frac{m}{n} \left( \lim_{x \to 0} \frac{\sin(mx)}{mx} \right) \cdot \left( \lim_{x \to 0} \frac{\sin(nx)}{nx} \right)^{-1} = \frac{m}{n}.
\]

### 6.4 Derivatives

We can now determine the derivative of \( \sin(x) \), using the two limits above and the addition formula for \( \sin(x + y) \):
\[ \frac{d}{dx} \sin(x) = \cos(x) \]

**Proof:** It turns out to be more convenient to use the second formula for the derivative, \( f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \):

\[
\begin{align*}
\frac{d}{dx} \sin(x) &= \lim_{h \to 0} \frac{\sin(x + h) - \sin(x)}{h} \\
&= \lim_{h \to 0} \frac{(\sin(x) \cos(h) + \cos(x) \sin(h)) - \sin(x)}{h} \\
&= \lim_{h \to 0} \frac{\sin(x) \cos(h) - 1 + \cos(x) \sin(h)}{h} \\
&= \left( \lim_{h \to 0} \sin(x) \right) \cdot \left( \lim_{h \to 0} \frac{\cos(h) - 1}{h} \right) + \left( \lim_{h \to 0} \cos(x) \right) \cdot \left( \lim_{h \to 0} \frac{\sin(h)}{h} \right) \\
&= \sin(x) \cdot 0 + \cos(x) \cdot 1 = \cos(x).
\end{align*}
\]

We could find the derivative of \( \cos(x) \) in a similar way, but it is easier to write \( \cos(x) \) in terms of \( \sin(x) \), and vice versa, using the shift property and the evenness and oddness:

\[
\begin{align*}
\cos(x) &= \cos(-x) = \sin \left( -x + \frac{\pi}{2} \right), \\
-\sin(x) &= -\cos \left( x - \frac{\pi}{2} \right) = -\cos \left( -x + \frac{\pi}{2} \right).
\end{align*}
\]

Using these identities we get (using the Chain Rule)

\[
\begin{align*}
\frac{d}{dx} \cos(x) &= \frac{d}{dx} \sin \left( -x + \frac{\pi}{2} \right) = \cos \left( -x + \frac{\pi}{2} \right) \cdot \frac{d}{dx} \left( -x + \frac{\pi}{2} \right) \\
&= -\cos \left( -x + \frac{\pi}{2} \right) = -\sin(x).
\end{align*}
\]

So we have

\[ \frac{d}{dx} \cos(x) = -\sin(x) \]

Now we can also get the derivative of \( \tan(x) \), by using the quotient rule:

\[
\begin{align*}
\frac{d}{dx} \tan(x) &= \left( \frac{\sin(x)}{\cos(x)} \right)' \\
&= \left( \frac{\sin(x)'}{\cos(x)} \right) + \left( \frac{\sin(x) \cdot \cos(x)'}{\cos^2(x)} \right) \\
&= \frac{\cos^2(x)}{\cos(x) \cos(x) - \sin(x) (-\sin(x))} \\
&= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} \\
&= \frac{1}{\cos^2(x)}.
\end{align*}
\]
With these new functions, we get lots of new examples using the Chain Rule, like

\[
\left( \sin(\sqrt{x}) \right)' = \cos(\sqrt{x}) \cdot \left( \sqrt{x} \right)' = \frac{\cos(\sqrt{x})}{2\sqrt{x}},
\]

\[
\left( \sqrt{\cos^2(x) + 1} \right)' = \frac{1}{2\sqrt{\cos^2(x) + 1}} \cdot (\cos^2(x) + 1)' = \frac{-2\cos(x)\sin(x)}{2\sqrt{\cos^2(x) + 1}}.
\]
Chapter 7

Exponential and Logarithmic Functions

7.1 Exponential functions

An exponential function is a function of the form

\[ f(x) = a^x, \]

where \( a \) is a positive real number. Often we also refer to \( C \cdot a^x \) as an exponential function, for \( C \) any real number.

This definition is not entirely straightforward. If \( x \) is an integer, we know what this means, e.g. \( a^5 = a \cdot a \cdot a \cdot a \cdot a \), \( a^{-2} = \frac{1}{a^2} \). Also if \( x \) is a rational number \( \frac{m}{n} \), then \( a^{\frac{m}{n}} = (a^{\frac{1}{n}})^m \), and \( a^{\frac{1}{n}} \) is that number that raised to the \( n \)-th power gives \( a \). But what if \( x \) is irrational, like \( \sqrt{2} \) or \( \pi \)? Basically, we define the function at those numbers to be whatever it has to be to make it continuous. For instance, we would define \( a^{\sqrt{2}} = \lim_{r \to \sqrt{2}} a^r \), where we only let the \( r \)'s in the limit be rational numbers (see Stewart p.53 for more detail). Still, that’s a tricky definition, but to define these functions (as well as the logarithmic functions) rigorously, requires more advanced mathematics that we want to avoid in this course.

The graph of an exponential function looks like one of these:

When \( a > 1 \), it is an increasing function, when \( 0 < a < 1 \) it is decreasing, and when \( a = 1 \) it is of course constant. No matter what \( a \) is, we always have

\[ a^0 = 1. \]
The most important properties (called exponent laws) of exponential functions are the following, which hold for any $a$:

\[
\begin{align*}
    a^{x+y} &= a^x \cdot a^y \\
    a^{xy} &= (a^x)^y \\
    (ab)^x &= a^x \cdot b^x
\end{align*}
\]

### 7.2 Logarithmic functions

A logarithmic function is a function of the form $f(x) = \log_a(x)$, for $a$ a positive real number. Here $\log_a(x)$ is the logarithm of $x$ to the base $a$, defined by

\[
y = \log_a(x) \iff x = a^y.
\]

In words, $\log_a(x)$ is the power to which you have to raise $a$ to get $x$. For example,

\[
\begin{align*}
    \log_2(8) &= 3, & \log_3(9) &= 2, & \log_5\left(\frac{1}{5}\right) &= -1, & \log_{\frac{1}{5}}(5) &= -1.
\end{align*}
\]

The function $\log_a(x)$ is the inverse of the function $a^x$. We will see later on what that means in more detail, but for now we just need to know that they cancel each other out:

\[
\log_a(a^x) = x, \quad a^{\log_a(x)} = x
\]

Both of these are tautological after carefully reading the definitions: $\log_a(a^x)$ is the power to which you have to raise $a$ to get $a^x$, which is of course $x$, and $a^{\log_a(x)}$ is $a$ raised to the power to which you have to raise $a$ to get $x$, which is $x$. Because of this, $\log_a(x)$ has properties corresponding exactly to the properties that $a^x$ has. For instance,

\[
\begin{align*}
    \log_a(1) &= 0, & \log_a\left(\frac{1}{x}\right) &= -\log_a(x),
\end{align*}
\]

which correspond to $a^0 = 1$ and $a^{-y} = \frac{1}{a^y}$. More generally, the main properties (logarithm laws) of $\log_a(x)$ are
\[
\log_a(x) + \log_a(y) = \log_a(xy)
\]
\[
\log_a \left( \frac{x}{y} \right) = \log_a(x) - \log_a(y)
\]
\[
\log_a(x^y) = y \log_a(x)
\]

Here are two examples of how these rules work:
\[
\log_2(10) + \log_2(12) - \log_2(15) = \log_2 \left( \frac{10 \cdot 12}{15} \right) = \log_2(8) = 3,
\]

\[
\log_{a^2}(a^3) = 3 \log_{a^2}(a) = \frac{3}{2} \log_{a^2}(a^2) = \frac{3}{2}.
\]

which indeed makes sense because \((a^2)^{\frac{3}{2}} = a^{2 \cdot \frac{3}{2}} = a^3\).

Another important rule tells us how to change from a logarithm in one base to a logarithm in a different base:

\[
\log_a(x) = \frac{\log_b(x)}{\log_b(a)}
\]

To justify this, we observe that

\[
b^{\log_a(x) \log_b(a)} = (b^{\log_b(a)})^{\log_a(x)} = a^{\log_a(x)} = x,
\]

which by definition means \(\log_a(x) \log_b(a) = \log_b(x)\).

As an example,

\[
3^{\log_3(4)} = 3^{\frac{\log_3(4)}{\log_3(3)}} = 3^{\frac{4}{2}} = (3^{\log_3(4)})^{1/2} = 4^{1/2} = 2.
\]

### 7.3  \(e^x\), \(\ln(x)\) and derivatives

Let’s try to compute the derivative of the exponential function \(f(x) = a^x\):

\[
f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} = \lim_{h \to 0} \frac{a^x \cdot a^h - a^x}{h}
\]

\[
= \lim_{h \to 0} a^x \cdot \frac{a^h - 1}{h} = a^x \cdot \lim_{h \to 0} \frac{a^h - 1}{h}
\]

\[
= a^x \cdot f'(0),
\]

where we can take \(a^x\) outside of the limit because it is independent of \(h\), and we have used \(f'(0) = \lim_{h \to 0} \frac{a^{h} - a^{0}}{h - 0}\).

We do not know the derivative of \(a^x\) yet, but we do know a lot about it: the derivative of an exponential function is the function itself, multiplied by a constant, \(f'(0)\).

We can’t determine the actual value of this constant yet, though. To do this we have to do things backwards: we’ll define a function \(g(x)\) such that \(g'(0) = 1\), and then for any other \(f(x)\) we can find \(f'(0)\) in terms of the function \(g(x)\). More precisely, we define

\[
f(x) = e^x\text{ is the exponential function such that } f'(0) = \lim_{h \to 0} \frac{e^h - 1}{h} = 1.
\]
In other words, $e$ is the real number such that \( \lim_{h \to 0} \frac{e^h - 1}{h} = 1 \). Such a definition only makes sense if we know that there is a number for which this limit actually exists, but right now we do not have the tools to prove this. On p.179 of Stewart, you can find some numerical evidence, including the fact that $e = 2.71828 \cdots$.

The function $e^x$ is often called the \textit{natural exponential function} or even \textit{the} exponential function. Because $f'(x) = f'(0) \cdot a^x$ for any exponential function, and because for $f(x) = e^x$ we have $f'(0)$ by definition, we know its derivative right away:

\[
(e^x)' = e^x
\]

So $e^x$ is a very special function, because it is the only function that is its own derivative (well, there’s $f(x) = 0...$).

Because this number is so special, we also give a special name to the logarithm to the base $e$:

\[
\ln(x) = \log_e(x)
\]

We call $\ln(x)$ the \textit{natural logarithm}, and the letters stand for \textit{logarithmus naturalis}. Note that most mathematicians actually refer to this function as $\log(x)$, while in high school and in old books $\log(x)$ is used to refer to $\log_{10}(x)$. We’ll avoid this confusion by never using $\log(x)$ without specifying the base.

From the basic properties of logarithms and exponential we now get

\[\ln(e^x) = x, \quad e^{\ln(x)} = x, \quad \ln(e) = 1, \quad \ln(x) = y \iff e^y = x.\]

Now, using the Chain Rule, we can also find the derivative of any $a^x$, using the trick

\[a = e^{\ln(a)} \implies a^x = a^x = (e^{\ln(a)})^x = e^{(\ln(a)) \cdot x},\]

so that we can apply the Chain Rule like so:

\[
\frac{d}{dx}a^x = \frac{d}{dx}e^{(\ln(a)) \cdot x} = e^{(\ln(a)) \cdot x} \cdot \frac{d}{dx}((\ln(a)) \cdot x) = e^{(\ln(a)) \cdot x} \cdot \ln(a) = a^x \cdot \ln(a).
\]

We can add another derivative to our inventory:
\[(a^x)' = a^x \cdot \ln(a)\]

Of course, since \(\ln(e) = 1\), this contains the fact that \((e^x)' = e^x\).

We have one more function that we need to know the derivative of, namely \(f(x) = \ln(x)\) (or more generally \(\log_a(x)\)). To prove what the derivative is we will have to wait till we have learned implicit differentiation, but we can use it already:

\[\left(\ln(x)\right)' = \frac{1}{x} \quad \text{with domain } (0, \infty).\]

So if \(f(x) = \ln(x)\), then \(f'(x)\) is the function \(\frac{1}{x}\), not with its whole domain, but with domain consisting of all \(x > 0\). This is because \(\ln(x)\) only exists for \(x > 0\), so it cannot have tangent lines at any \(x \leq 0\), therefore its derivative cannot be defined there. Note that we do have

\[\frac{d}{dx} \ln(|x|) = \frac{1}{x},\]

with full domain (all \(x \neq 0\)), because \(\ln(|x|)\) is defined for all \(x \neq 0\) (and you will see that the derivative is this if you write out the piecewise definition).

Finally, we can also get the derivative of \(\log_a(x)\) using

\[\log_a(x) = \frac{\log_e(x)}{\log_e(a)} = \frac{\ln(x)}{\ln(a)},\]

so

\[\frac{d}{dx} \log_a(x) = \frac{d}{dx} \frac{\ln(x)}{\ln(a)} = \frac{1}{\ln(a)} \frac{d}{dx} \ln(x) = \frac{1}{x \cdot \ln(a)},\]

again with domain restricted to \((0, \infty)\).

\[\frac{d}{dx} \log_a(x) = \frac{1}{x \ln(a)} \quad \text{with domain } (0, \infty).\]

### 7.4 Limits

As you might expect, all exponential and logarithmic functions are continuous. So most limits involving them can be evaluated by plugging in. We won’t see many limits with these functions where we need to do some algebra before we can evaluate, but we will need some important limits at infinity and infinite limits. The proofs here get trickier, and you only need to read them if you want to challenge your limit skills. But you should at least be aware of these limits, and that proving them requires the Squeeze Theorem. The first basic ones are

\[\lim_{x \to \infty} e^x = \infty, \quad \lim_{x \to -\infty} e^x = 0.\]
These say what the graph of $e^x$ looks like (so we can’t just read them off from the graph, because we need them to be able to draw the graph...).

The first one is easy because we know that $e^x > x$ and $\lim_{x\to\infty} x = \infty$, so then $e^x$ must go to $\infty$ as well. The second one we get using the Squeeze Theorem, because for negative $x$ (so $-\frac{1}{x}$ is positive),

$$0 < e^x < -\frac{1}{x},$$

and $\lim_{x\to-\infty} 0 = 0$ and $\lim_{x\to-\infty} -\frac{1}{x} = 0$. So $e^x$ is squeezed in between and also goes to zero.

The corresponding limits for $\ln(x)$ are

$$\lim_{x\to\infty} \ln(x) = \infty, \quad \lim_{x\to0^+} \ln(x) = -\infty.$$ 

To prove the first one, observe that the graph $y = \ln(x)$ contains the points $(2^n, \ln(2^n))$ for $n$ an integer, which clearly go to $\infty$ (in the $y$-direction) as we go to $\infty$ in the $x$-direction. But the derivative is $\frac{1}{x} > 0$, so the function is increasing, hence it must go to $\infty$ as a whole.

Now we can deduce the second from the first, using the substitution $u = \frac{1}{x}$; then as $x \to 0^+$, $u \to \infty$, so plugging in $x = \frac{1}{u}$ gives

$$\lim_{x\to0^+} \ln(x) = \lim_{u\to\infty} \ln\left(\frac{1}{u}\right) = -\lim_{u\to\infty} \ln(u) = -\infty.$$ 

Another pair of limits that we’ll need later on is, with $a > 0$ any real number,

$$\lim_{x\to\infty} x^a e^x = 0, \quad \lim_{x\to\infty} \frac{\ln x}{x^a} = 0.$$ 

What these say is basically that $e^x$ beats any power $x^a$ (no matter how large $a$), and any power $x^a$ beats $\ln x$ (no matter how small $a$). We’ll first prove the second using the Squeeze Theorem, and the inequality

$$0 < \frac{\ln(x)}{x} < \frac{\ln(\sqrt{x})}{\sqrt{x}} = \frac{\ln(\sqrt{x})}{\sqrt{x}} \cdot \frac{2}{\sqrt{x}} < \frac{2}{\sqrt{x}},$$

where for the last step we used that $\ln(\sqrt{x}) < \sqrt{x}$ so $\frac{\ln(\sqrt{x})}{\sqrt{x}} < 1$. So $\frac{\ln(x)}{x}$ is squeezed in between 0 and $\frac{2}{\sqrt{x}}$, and $\lim_{x\to\infty} \frac{2}{\sqrt{x}} = 0$, so by the Squeeze Theorem we have $\lim_{x\to\infty} \frac{\ln x}{x} = 0$. But then also

$$0 = \lim_{x\to\infty} \frac{\ln x^a}{x^a} = \lim_{x\to\infty} \frac{a \ln x}{x^a} = a \lim_{x\to\infty} \frac{\ln x}{x^a},$$

which implies the second limit since $a \neq 0$.

Finally, to prove the first we use the substitution $u = e^x$, so plugging in $x = \ln u$ gives

$$\lim_{x\to\infty} \frac{x^a}{e^x} = \lim_{u\to\infty} \frac{(\ln u)^a}{e^{\ln u}} = \lim_{u\to\infty} \frac{(\ln u)^a}{u} = \lim_{u\to\infty} \left(\frac{\ln u}{u^{1/a}}\right)^a = \left(\lim_{u\to\infty} \frac{\ln u}{u^{1/a}}\right)^a = 0^a = 0,$$

where $\lim_{u\to\infty} \frac{\ln u}{u^{1/a}} = 0$ is the second limit, with $1/a$ instead of $a$. 

49
Chapter 8

Implicit & Logarithmic Differentiation, Inverse Functions

8.1 Implicit differentiation

First, I should point out the difference between a graph and a curve. A graph is the set of points \((x, y)\) in the \(xy\)-plane that satisfy an equation \(y = f(x)\), where \(f\) is some function. A curve is the set of points in the \(xy\)-plane that satisfy any kind of equation, like \(3x^2y^3 + 5x^5 - 7 = 0\) or \(e^x + \tan(x) = 0\).

Every graph is a curve, but not every curve is a graph. You can see if a given curve is a graph using the Vertical Line Test: if it intersects any vertical line in no more than one point, then it is a graph. This is because a graph can only have one \(y\)-value above every \(x\)-value. Whether you can actually find the function that the curve is the graph of, is a different matter.

How do we find the tangent lines to a curve like the circle \(x^2 + y^2 = 1\)? We’ve seen how to find tangent lines to graphs \(y = f(x)\), but the circle is not the graph of a function, because it has two points on some vertical lines. So we need a new trick. Recall that when \(y = f(x)\), \(\frac{dy}{dx}\) is the same as \(f'(x)\). Then the Chain Rule tells us for instance that

\[
\frac{d}{dx} y^2 = \frac{d}{dx} f^2(x) = 2f(x) \cdot f'(x) = 2y \cdot \frac{dy}{dx}.
\]

With this notation we can take an equation like \(x^2 + y^2 = 1\) and (although we can’t write it as \(y = f(x)\)) we can differentiate it on both sides:

\[
\frac{d}{dx} (x^2 + y^2) = \frac{d}{dx} 1
\]

\[
2x + 2y \frac{dy}{dx} = 0.
\]

Then we can solve for \(\frac{dy}{dx}\) and we get

\[
\frac{dy}{dx} = \frac{-2x}{2y} = -\frac{x}{y}.
\]

So the slope of the tangent line to the circle at a point \((a, b)\) on the circle is \(-\frac{a}{b}\). For instance, at the point \((0, 1)\) the slope is \(-\frac{0}{1} = 0\) and at \((\sqrt{1/2}, \sqrt{1/2})\) it is \(-1\), both of which make sense.
if you draw the picture.
For the circle, we can actually check this result, because although it is not a graph, we can split it up into the two graphs \( y = \sqrt{1-x^2} \) and \( y = -\sqrt{1-x^2} \). For each of those we can find the slope of the tangent line with regular differentiation:

\[
\begin{align*}
y = \sqrt{1-x^2} & \Rightarrow \frac{dy}{dx} = \frac{-x}{\sqrt{1-x^2}} = \frac{-x}{y}, \\
y = -\sqrt{1-x^2} & \Rightarrow \frac{dy}{dx} = \frac{x}{\sqrt{1-x^2}} = \frac{x}{-y} = \frac{-x}{y},
\end{align*}
\]

just like we found using implicit differentiation.

We can now also use \( \frac{dy}{dx} \) to find horizontal tangent lines (which we’ll see later have special importance), by putting \( \frac{dy}{dx} \) equal to 0:

\[
\frac{dy}{dx} = 0 \Rightarrow -\frac{x}{y} = 0 \Rightarrow x = 0,
\]

so the points with horizontal tangent are \((0, 1)\) and \((0, -1)\), as we can obviously see in a drawing.

We can also find the vertical tangent lines, which occur when \( \frac{dy}{dx} = \infty \) (actually, we should say ‘does not exist’), so when the denominator \( y = 0 \). And indeed, the corresponding points are \((1, 0)\) and \((-1, 0)\).

So implicit differentiation is a way of finding \( \frac{dy}{dx} \) when \( y \) is given implicitly by an equation like \( x^2 + y^2 = 1 \), instead of explicitly as \( y = f(x) \).

In the case of the circle we could circumvent it, but for many other curves it is by far the easiest way to find the tangent lines. Take for instance the curve

\[
6x^2 = y^3 - y,
\]

for which we don’t know how to write \( y \) as a function of \( x \) (maybe it’s possible, but it won’t be as easy as using implicit differentiation).

Then just as above

\[
\frac{d}{dx} 6x^2 = \frac{d}{dx} (y^3 - y) \\
12x = 3y^2 \frac{dy}{dx} - \frac{dy}{dx} = (3y^2 - 1) \frac{dy}{dx},
\]

\[
\Rightarrow \frac{dy}{dx} = \frac{12x}{3y^2 - 1}.
\]

Then the slope of the tangent line at the point \((1, 2)\) (check that it’s on the curve; you don’t have to know how to find such a point, just like you don’t have to know how to draw such a curve) is \( \frac{12 \cdot 1}{3 \cdot 2^2 - 1} = \frac{12}{11} \), which makes sense in the picture.

Putting \( \frac{12x}{3y^2 - 1} = 0 \) gives \( x = 0 \), hence from the original equation \( 0 = y^3 - y = y(y^2 - 1) \), so \( y = 0 \) or \( y = \pm 1 \). Therefore the points with horizontal tangent are \((0, 0)\), \((0, 1)\) and \((0, -1)\), which again makes sense in the picture.

More difficult are the vertical tangent lines, for which \( \frac{dy}{dx} = \infty \), so the denominator is \( 3y^2 - 1 = 0 \). Solving gives \( y = \pm \frac{1}{\sqrt{3}} = \pm 3^{-1/2} \). Only \( y = -3^{-1/2} \) gives points on the curve, because then \( 0 < y^3 - y = -3^{-3/2} + 3^{-1/2} \), so \( x = \pm \sqrt[3]{\frac{1}{6}(y^3 - y)} = \pm \sqrt[3]{\frac{1}{6}(3^{-1/2} - 3^{-3/2})} \) exists. So we’ve identified the two points on the left and right of the egg: \( \left( \pm \sqrt[3]{\frac{1}{6}(3^{-1/2} - 3^{-3/2})}, -3^{-1/2} \right) \).
Now let’s look at a more difficult example, the *lemniscate* (see Stewart # 29 on p. 213), given by

\[ 2(x^2 + y^2)^2 = 25(x^2 - y^2). \]

Maybe it is possible to write \( y \) as a function of \( x \), but we just don’t want to try. But still we can find \( \frac{dy}{dx} \) just like above:

\[
\frac{d}{dx} \left( 2(x^2 + y^2)^2 \right) = 4(x^2 + y^2) \cdot \frac{d}{dx} (x^2 + y^2) = 25 \left( 2x - 2y \frac{dy}{dx} \right)
\]

\[ 4(x^2 + y^2) \cdot \left( 2x + 2y \frac{dy}{dx} \right) = 50x - 50y \frac{dy}{dx} \]

\[ \implies 8x(x^2 + y^2) + 8y(x^2 + y^2) \frac{dy}{dx} = 50x - 50y \frac{dy}{dx} \]

\[ \implies (8y(x^2 + y^2) + 50y) \frac{dy}{dx} = 50x - 8x(x^2 + y^2) \]

\[ \implies \frac{dy}{dx} = \frac{50x - 8x(x^2 + y^2)}{8y(x^2 + y^2) + 50y}. \]

So at for instance the point \((3,1)\) (check that it’s on the curve; you don’t have to know how to find such a point, other than by guessing) the tangent line has slope

\[
\frac{50 \cdot 3 - 8 \cdot 3(9 + 1)}{8(9 + 1) + 50} = \frac{150 - 240}{130} = \frac{9}{13},
\]

which makes sense in the picture.

With implicit differentiation, we can now give the **proof** of

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

By definition of the logarithm, if \( y = \ln(x) \) then \( x = e^y \), so differentiating on both sides gives

\[
\frac{d}{dx} x = \frac{d}{dx} e^y
\]

\[ 1 = e^y \cdot \frac{dy}{dx} \]

\[ \implies \frac{dy}{dx} = \frac{1}{e^y} = \frac{1}{e^{\ln(x)}} = \frac{1}{x}. \]

### 8.2 Logarithmic Differentiation

Logarithmic differentiation is when you simplify the differentiation process by using the equation

\[
\frac{d}{dx} \ln(f(x)) = \frac{f'(x)}{f(x)}
\]

which you get from \( \frac{d}{dx} \ln(x) = \frac{1}{x} \) and the Chain Rule.

For instance, take \( f(x) = x^a \). You can’t apply the Power Rule, and it’s not quite an exponential function \( a^x \). But \( \ln(f(x)) = \ln(x^a) = x \ln(x) \), so

\[
\frac{f'(x)}{f(x)} = \frac{d}{dx} \ln(f(x)) = \frac{d}{dx} (x \ln(x)) = x' \ln(x) + x(\ln(x))' = \ln(x) + 1,
\]

\[ 52 \]
\[ (x^x)' = f'(x) = f(x) \cdot (\ln(x) + 1) = x^x(\ln(x) + 1). \]

Actually, we already had a way to do this derivative, using \( e^{\ln(x)} = x \) and the Chain Rule:
\[
x^x = (x)^x = (e^{\ln(x)})^x = e^{x\ln(x)},
\]
\[ \Rightarrow (x^x)' = (e^{x\ln(x)})' = e^{x\ln(x)} \cdot (\ln(x))' = e^{x\ln(x)} \cdot (\ln(x) + 1) = x^x(\ln(x) + 1). \]

But logarithmic differentiation can do more. For instance, we could differentiate
\[ f(x) = \sqrt{x e^{x^2}} \]
the old-fashioned way using the Quotient Rule and Product Rule, but that would be quite a bit of work. Now we can use
\[
\ln \left( \frac{\sqrt{x e^{x^2}}}{(x^2 + 1)^{10}} \right) = \ln(\sqrt{x}) + \ln(e^{x^2}) - \ln((x^2 + 1)^{10}) = \frac{1}{2} \ln(x) + x^2 - 10 \ln(x^2 + 1)
\]
to get
\[
f'(x) = f(x) \cdot \frac{d}{dx} \ln \left( \frac{\sqrt{x e^{x^2}}}{(x^2 + 1)^{10}} \right) = f(x) \cdot \frac{d}{dx} \left( \frac{1}{2} \ln(x) + x^2 - 10 \ln(x^2 + 1) \right)
\]
\[
= \frac{\sqrt{x e^{x^2}}}{(x^2 + 1)^{10}} \cdot \left( \frac{1}{2x} + 2x - \frac{20x}{x^2 + 1} \right).
\]
This trick is useful whenever you have a function that is a product of several factors which themselves have exponents, because then applying the logarithm changes products into sums and exponents into multiplications, so that instead of the cumbersome Product and Quotient Rules, you deal with the Sum Rule, which is much easier.

We can also use logarithmic differentiation to finally give the proof of the Power Rule \((x^a)' = ax^{a-1}\) with exponent \(a\) an arbitrary number (recall that before we only proved this for integer exponents). Let \( f(x) = x^a \), \( a \) any real number. Then
\[
\frac{f'(x)}{f(x)} = \frac{d}{dx} \ln(x^a) = \frac{d}{dx} a \ln(x) = \frac{a}{x},
\]
\[ \Rightarrow f'(x) = f(x) \cdot \frac{a}{x} = x^a \cdot \frac{a}{x} = ax^{a-1}. \]

### 8.3 Inverse Functions

Given a function \( f \) with domain \( D \) and range \( R \), so
\[ f : D \to R, \]
an inverse function for \( f \) is defined to be a function \( f^{-1} \) with domain \( R \) and range \( D \), so
\[ f^{-1} : R \to D, \]
with the two properties
\[ f^{-1}(f(x)) = x \text{ and } f(f^{-1}(x)) = x. \]

Equivalently, these properties can be expressed as: \( y = f(x) \iff f^{-1}(y) = x \).

Do not confuse the notation \( f^{-1}(x) \) with \((f(x))^{-1} = \frac{1}{f(x)}\).

Let’s see some examples; pictures of the first two are below.
• $f(x) = e^x$ and $f^{-1}(x) = \ln(x)$: then the properties are $f^{-1}(f(x)) = \ln(e^x) = x$ and $f(f^{-1}(x)) = e^{\ln(x)} = x$, which we saw before.

• $f(x) = x^3 + 1$ and $f^{-1}(x) = \sqrt[3]{x-1}$: then $f^{-1}(f(x)) = \sqrt[3]{(x^3+1)-1} = x$ and $f(f^{-1}(x)) = (\sqrt[3]{x-1})^3 + 1 = x$.

• $f(x) = ax + b$ for some numbers $a$ and $b$: then we can find $f^{-1}$ by putting $y = ax + b$, and solving for $x$ in terms of $y$, which gives $f^{-1}(x) = \frac{1}{a}(x-b)$. Then

\[
f^{-1}(f(x)) = \frac{1}{a}((ax + b) - b) = \frac{1}{a}(ax) = x,
\]

\[
f(f^{-1}(x)) = a\left(\frac{1}{a}(x - b)\right) + b = (x - b) + b = x.
\]

The last example is typical: given $y = f(x)$, if we solve for $x$ in terms of $y$, and get a unique solution, that that gives an inverse function $x = f^{-1}(y)$ (things can get a little confusing when we then write $f^{-1}$ as a function of $x$ again, i.e. we swap $x$ and $y$).

The operation of going from a function $f$ to its inverse $f^{-1}$ is easily captured in terms of the graph: [PICTURE MISSING]

We can get the graph of $f^{-1}$ from that of $f$ by reflecting the graph of $f$ in the line $y = x$.

Unfortunately, things are not always this easy. Consider and $f(x) = x^2$, with domain $\mathbb{R}$; can we invert this function? No: if we reflect its graph, then we get a curve that cannot be a graph, because it fails the Vertical Line Test: there are vertical lines that intersect the curve in more than one point. In other words, there are values of $x$ for which there is more than one corresponding value of $y$, which is not possible for the graph of a function.

So $f(x) = x^2$ does not have an inverse function. In general, when the reflection of a graph $y = f(x)$ in the line $y = x$ is not the graph of any function, then the function does not have an inverse. We can phrase this as the

**Horizontal Line Test**: if the graph of a function intersects any horizontal line in more than two points, then the function does not have an inverse.

This is true because its reflection would fail the Vertical Line Test, so could not be the graph of a function.

For instance, the function $f(x) = |x|$ does not have an inverse, because it fails the Horizontal Line Test: for instance the line $y = 1$ intersects its graph in the points $(1,1)$ and $(-1,1)$. On the other hand, the function $f(x) = x^5$ passes the Horizontal Line Test, because it intersects a line $y = b$ only in the point $(\sqrt[5]{b}, b)$. Of course, the reason that there is only one such point is that the function has an inverse, $f^{-1}(x) = \sqrt[5]{x}$.

But all is not lost for functions that fail the Horizontal Line Test. If we restrict the domain of the function, then we might be able to find an inverse for the function, just on that domain; we call that a partial inverse.

For example, take $f(x) = x^2$ but now restrict its domain to $D = [0, \infty)$. Then if we draw the graph only above that domain, it does pass the Horizontal Line Test. And indeed, the function has an inverse, $f^{-1}(x) = \sqrt{x}$. Similarly, we can restrict the domain of $f$ to $D = (-\infty, 0]$, and there it has the inverse $f^{-1}(x) = -\sqrt{x}$. So what we’ve done is split the graph into two pieces, each of which passes the Horizontal Line Test by itself, and then we found the partial inverse for each of those pieces. If we then draw the two partial inverses in one picture, we in fact get the graph of $f$ reflected in $y = x$.  

54
A very important source of examples is trigonometry, for instance \( f(x) = \sin(x) \). This function fails the Horizontal Line Test miserably: \( y = \sin(x) \) intersects the line \( y = 0 \) in infinitely many points! But again we can restrict the domain; the standard choice (which could have been anything) is to take \( D = \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right] \), and then the range is \( R = [-1, 1] \). Then we define

\[
\text{arcsin}(x) = \sin^{-1}(x) \quad \text{with domain} \quad [-1, 1] \quad \text{and range} \quad \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right].
\]

This function does not have any simpler formula, we only know it as the inverse of \( \sin(x) \). So if we want to evaluate for instance \( \arcsin \frac{1}{2} \), then we need to find an \( x \) between \( -\frac{\pi}{2} \) and \( \frac{\pi}{2} \) that gives \( \sin(x) = 1/2 \): if we remember the table right, we know that \( \sin(\pi/6) = 1/2 \), so \( \arcsin \frac{1}{2} = \frac{\pi}{6} \).

See Stewart, 1.6 for pictures of these inverse trigonometric functions.

Similarly, if we restrict \( \cos(x) \) to the domain \( D = [0, \pi] \), then we can define the partial inverse

\[
\text{arccos}(x) = \cos^{-1}(x) \quad \text{with domain} \quad [-1, 1] \quad \text{and range} \quad [0, \pi],
\]

and if we restrict \( \tan(x) \) to \( D = [-\pi/2, \pi/2] \) (i.e., we take the piece between two asymptotes), then it’s range is \( R = (-\infty, \infty) \), then we define

\[
\text{arctan}(x) = \tan^{-1}(x) \quad \text{with domain} \quad (-\infty, \infty) \quad \text{and range} \quad \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right].
\]

When differentiating these functions in the next section, we’ll need the following identities:

\[
\cos(\arcsin(x)) = \sqrt{1 - x^2}
\]

This is \( \cos(\theta) \) where \( \theta \) is the angle such that \( \sin(\theta) = x \). Since \( \sin(\theta) = \frac{\text{opp}}{\text{hyp}} \), draw a right triangle with angle \( \theta \), hypotenuse 1 and opposite side \( x \). Then since \( \cos(\theta) = \frac{\text{adj}}{\text{hyp}} = \text{adj} \), we get \( \cos(\theta) = \sqrt{1 - x^2} \), using Pythagoras.

Similarly, we get

\[
\sin(\arccos(x)) = \sqrt{1 - x^2}
\]

by drawing a triangle with angle \( \theta = \arccos(x) \), hypotenuse 1 and adjacent side \( x \). Then \( \sin(\theta) = \text{opp} = \sqrt{1 - x^2} \).

Finally, we also need

\[
\cos(\arctan(x)) = \frac{1}{\sqrt{1 + x^2}}
\]

In this case, \( \theta = \arctan(x) \) so \( x = \tan(\theta) = \frac{\text{opp}}{\text{adj}} \). So draw the triangle with opposite side \( x \) and adjacent side 1, then the hypotenuse is \( 1 + x^2 \) by Pythagoras, so \( \cos(\theta) = \frac{\text{adj}}{\text{hyp}} = \frac{1}{\sqrt{1+x^2}} \).

### 8.4 Derivatives of inverse functions

Having learned these new functions, we’d like to know how to differentiate them, so to know

\[
\frac{d}{dx} f^{-1}(x),
\]

if we know \( f \). This involves implicit differentiation.

Write \( y = f^{-1}(x) \), which by definition is equivalent to \( x = f(y) \) (a little confusing, but \( f^{-1} \)
is the function we’re differentiating, so we take its variable to be $x$). Then differentiating $x = f(y)$ on both sides gives

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

$$1 = f'(y) \cdot \frac{dy}{dx}$$

$$\implies \frac{dy}{dx} = \frac{1}{f'(y)} = f'(f^{-1}(x))^{-1},$$

so we have what me might call the Inverse Function Rule:

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

Do not misuse this as $\frac{d}{dx} f^{-1}(x) = \frac{1}{f'(x)}$, that wouldn’t make sense because $x$ is from the domain of $f^{-1}(x)$, so might not be in the domain of $f$ or $f'$ at all, hence you shouldn’t be plugging $x$ into $f'$.

Let’s quickly test this on $f^{-1}(x) = \sqrt{x}$, so $f(x) = x^2$ with domain $[0, \infty)$, and $f'(x) = 2x$. Then the rule gives

$$\frac{d}{dx} \sqrt{x} = \frac{1}{f'(\sqrt{x})} = \frac{1}{2\sqrt{x}}$$

as we already learned. We can also test it on $f^{-1}(x) = \ln(x)$, so $f(x) = e^x$:

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

For the example $f(x) = x^3 + 1$, $f^{-1}(x) = \sqrt[3]{x - 1}$ that we saw above we have

$$\frac{d}{dx} \sqrt[3]{x - 1} = \frac{1}{3(f^{-1}(x))^2} = \frac{1}{3(\sqrt[3]{x - 1})^2} = \frac{1}{3}(x - 1)^{-2/3}.$$  

All of these we could of course already do, but for the inverse trigonometric functions we really need it (unless we go through the implicit differentiation again, of course). With $f(x) = \sin(x)$, $f'(x) = \cos(x)$ we get

$$f'(f^{-1}(x)) = \cos(\arcsin(x)) = \sqrt{1 - x^2},$$

using the identity $\cos(\arcsin(x)) = \sqrt{1 - x^2}$, which we saw at the end of the last section. So

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

Similarly when $f(x) = \cos(x)$ we get

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

Finally, when $f(x) = \tan(x)$ and $f'(x) = \frac{1}{\cos^2(x)}$ we have

$$f'(f^{-1}(x)) = \frac{1}{\cos(\arctan(x))^2} = \frac{1}{\left(\frac{1}{1+x^2}\right)^2} = \frac{1}{1 + x^2},$$

so

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

56
Chapter 9

The Intermediate Value Theorem and the Mean Value Theorem

9.1 The Intermediate Value Theorem

The following theorem captures one of the most basic properties of continuous functions.

\[
\text{If } f \text{ is a continuous function on } [a, b], \text{ and } d \text{ is a number between } f(a) \text{ and } f(b), \text{ then there is a number } c \text{ in } [a, b] \text{ such that } f(c) = d.
\]

In terms of the graph of \( f \), this property is almost too simple: if \((a, f(a))\) is below the line \( y = N \), and \((b, f(b))\) is above it, then if you draw the graph from one point to the other, you must cross the line \( y = N \), i.e. there is some point \((c, f(c))\) on that line.

This theorem is important to mathematicians because it allows us to formally prove lots of statements about function that seem intuitively true. The point is that the formal definition of continuity (which, remember, is just about limits and values of the function, not about drawing the graph) is so general that continuous functions can do a lot of things that do not match our intuitive idea of 'something you can draw without taking your pen off the paper'. For instance we saw that functions involving \( \sin \left( \frac{1}{x} \right) \) can have weird properties, like having infinitely many squiggles in a tiny piece of the graph, which is not something you do with a pen.

Should you take more advanced math classes, you will see how theorems like this (and the MVT below) are used to build up a logically strict theory of functions, and how that will help you understand issues that don’t make sense intuitively. For now, though, we will merely use it to prove a few simple statements. Questions like these do show up on the exam for this course.

Note that the conclusion of the theorem only says that some such number \( c \) exists, it gives you no clue about what it really is (other than the interval it is in). Nor does it say that this \( c \) is the only such number, there could easily be several such \( c \).

**Roots of functions.** A root of a function \( f(x) \) is a number \( c \) such that \( f(c) = 0 \).

For instance, let \( f(x) = x^4 + x - 3 \), which is certainly continuous. We can’t explicitly find a root of this function (like we could for simpler functions like \( x^2 - 4 \)). Suppose we take \( a = 0 \) and \( b = 2 \). Then \( f(0) = -3 \) and \( f(2) = 16 + 2 - 3 = 15 \). So \( N = 0 \) is between \( f(0) \) and \( f(2) \), hence by the Intermediate Value Theorem, there is a number \( c \) between 0 and 2 such that \( f(c) = N = 0 \), i.e. a root of the function.
In general, to show that a function has a root in some interval, it is enough to find two numbers \(a\) and \(b\) in that interval such that \(f(a) < 0\) and \(f(b) > 0\) (or the other way around). We can take the above example further. Since \(f(1) = 1 + 1 - 3 < 0\), and \(f(2) > 0\), there is in fact a root of \(f\) between 1 and 2. Similarly, \(f\left(1\frac{1}{2}\right) = \left(\frac{3}{2}\right)^4 + \frac{3}{2} - 3 = \left(\frac{3}{2}\right)^4 - \frac{3}{2} > 0\), so there is a root between 1 and \(1\frac{1}{2}\). Next \(f\left(\frac{5}{4}\right) = \frac{625}{256} + \frac{5}{4} - 3 = \frac{625}{256} - \frac{7}{4} = \frac{625 - 4 \cdot 64}{4} > 0\), so the root is between 1 and \(1\frac{1}{4}\).

You see that we can keep going like this to get an approximation to a root of \(f\) (which in fact is 1.164 \cdots). This is called the **bisection method**, because we keep cutting in half the interval that the root is in. It is nice in theory, but actually very inefficient; soon we will see that using information from the derivative, we can approximate roots much, much faster (with Taylor polynomials or Newton’s method).

Of course, \(f\) might have more roots than just this one: indeed, since \(f(-2) = 16 - 2 - 3 > 0\) and \(f(0) < 0\), there is another root somewhere in the interval \([-2, 0]\). And actually, we are not at all sure that there isn’t more than one root in these intervals.

### Solutions to equations.

A **solution** to an equation \(f(x) = g(x)\) is a number \(c\) such that \(f(c) = g(c)\). Of course, this is nothing but a root of the function \(f(x) - g(x)\), so this is really the same thing, but it is good to know the distinction.

Take for instance the equation \(\cos(x) = x\). Since for \(a = 0\), we have \(\cos(a) = 1 > 0 = a\) and for \(b = \pi/2\), we have \(\cos(b) = 0 < \pi/2 = b\), there must be a solution to this equation between 0 and \(\pi/2\). That follows from applying the Intermediate Value Theorem to the function \(f(x) = \cos(x) - x\), which is continuous (on any interval) and has \(f(0) > 0\) and \(f(\pi/2) < 0\). Again we could approximate this solution by looking at \(\cos(\pi/4) = 1/\sqrt{2} < \pi/4\), so the solution is between 0 and \(\pi/4\). And we can go on like this, although evaluating \(\cos\) and comparing the values gets harder (without a calculator).

### 9.2 The Mean Value Theorem

The mean value theorem is the big brother of the intermediate value theorem. In a similar way, it asserts the existence of some number \(c\), but now instead of just continuous, the function has to be differentiable.

\[
\text{If } f \text{ is a differentiable function on } [a, b], \text{ then there is a number } c \text{ in } [a, b] \text{ such that } f'(c) = \frac{f(b) - f(a)}{b - a}.
\]

Again, try to think of what this means in terms of the graph of \(f\). The expression \(\frac{f(b) - f(a)}{b - a}\) (called a **difference quotient**) is the slope of the line between the two points \((a, f(a))\) and \((b, f(b))\), which are the endpoints of the graph above the interval \([a, b]\). Then the theorem says there is a point \((c, f(c))\) on the graph where the slope \(f'(c)\) is exactly that difference quotient. In other words, a point on the graph where the slope of the graph is equal to the slope of the line between the endpoints of the graph. Or to look at it in another way, \(\frac{f(b) - f(a)}{b - a}\) is the **average rate of change** over the interval \([a, b]\), and the theorem says there is a point where the **instantaneous** rate of change equals the average rate of change.

If you try some pictures, this should be pretty intuitive (though not as much as the IVT), but it may not be clear why this conclusion is useful. As it turns out, it is very useful, at least to mathematicians. Recall that the derivative \(f'(a)\) is defined as the limit, as \(b\) goes to \(a\), of the difference quotient \(\frac{f(b) - f(a)}{b - a}\), but that doesn’t give us any kind of equation between values of
the derivative and of the function itself. That’s exactly what the MVT does: it gives a relation between the values of \( f \) on the endpoints of an interval \([a, b]\) and the value of the derivative at some number within that interval. Even if we don’t know what that number is exactly, we can still deduce a lot of information about the function from its derivative.

Let’s see a simple example of how this works. Let \( f(x) = \sqrt{x} \), and consider the interval \([1, 2]\). Then \( f \) is certainly differentiable there, so the Mean Value Theorem says there is a number \( c \) with \( 1 \leq c \leq 2 \) such that

\[
 f'(c) = \frac{f(2) - f(1)}{2 - 1} = \sqrt{2} - 1.
\]

So the average slope over \([1, 2]\) is \( \sqrt{2} - 1 \), and at some number in between, the slope \( f'(c) \) of the graph equals that graph. In fact, in this case we can find that \( c \) explicitly: since \( f'(x) = \frac{1}{2\sqrt{x}} \), we are looking for \( c \) such that

\[
 \frac{1}{2\sqrt{c}} = \sqrt{2} - 1 \Rightarrow \sqrt{c} = \frac{1}{2(\sqrt{2} - 1)} \Rightarrow c = \frac{1}{4(\sqrt{2} - 1)^2} \approx 1.457,
\]

so indeed there is such a \( c \) in that interval. This was possible because the function is pretty simple, and there is only one number \( c \) overall where the graph has this slope.

I should also show you an example where the conclusion fails. Take for instance \( f(x) = |x| \) on the interval \([-2, 2]\), where it is not differentiable at all values: there is a corner at \( x = 0 \). Then

\[
 \frac{f(2) - f(-2)}{2 - (-2)} = \frac{[2] - [-2]}{4} = 0,
\]

but there is no \( c \) anywhere with \( f'(c) = 0 \), because \( f'(x) = -1 \) to the left of \( 0 \) and \( f'(x) = 1 \) to the right.

So the differentiability condition is crucial.

The MVT has many more different applications than the IVT, and we will only see a few. First of all, it can provide formal proofs of many facts about differentiable functions that we have been using throughout this course, or that we’re about to see:

- If \( f'(x) > 0 \), then \( f \) is increasing, and if \( f'(x) < 0 \), then \( f \) is decreasing.
- If \( f''(x) > 0 \), then \( f \) is concave up, and if \( f''(x) < 0 \), then \( f \) is concave down.
- If \( F'(x) = f(x) \), then all antiderivatives of \( f \) are of the form \( F(x) + c \) for a constant \( c \).
- Taylor’s inequality

You can find these proofs in the book, but let me illustrate the first one. It seems easy to explain when you point at a picture (and that’s what I did earlier on in the class): if \( f'(x) > 0 \), then the tangent line has positive slope, so is going up, hence the graph must be going up, right? Well, mathematicians are never happy when an argument about functions involves ‘pointing at a picture’, because that picture can never represent all the possibilities. It clearly works for the typical graph that you would draw, but does it also work for graphs involving \( \sin \left( \frac{1}{x} \right) \)? And even if it does, how do we know that there are no other functions where crazy things happen?

This is why we want to use the MVT to give a formal proof. So, suppose \( f'(x) > 0 \) on some interval \([a, b]\). Then by the MVT, there is a \( c \) in \([a, b]\) such that \( f'(c) = \frac{f(b) - f(a)}{b - a} \). Therefore \( f(b) - f(a) = f'(c)(b - a) \), and \( f'(c)(b - a) > 0 \) since both \( f'(c) > 0 \) and \( b - a > 0 \). So \( f(b) - f(a) > 0 \), i.e. \( f(b) > f(a) \), which exactly means that \( f \) is increasing from \( a \) to \( b \).

I will list the different types of questions involving the MVT that you may encounter.
**Roots of functions.** We can often use the MVT to show that some root is unique. For instance, let \( f(x) = x^3 - x - 1 \). Then \( f(1) = -1 \) and \( f(2) = 5 \), so by the IVT \( f \) has at least one root in \([1, 2]\). But to show that there isn’t more than one root in that interval, we need the MVT.

Suppose there were at least 2 roots, let’s say \( f(a) = 0 \) and \( f(b) = 0 \), with \( a \) and \( b \) in \([1, 2]\) and \( a < b \). Then the MVT says there is \( c \) in \([a, b]\) such that (with \( f'(x) = 3x^2 - 1 \))

\[
3c^2 - 1 = \frac{f(b) - f(a)}{b - a} = 0,
\]

so \( 3c^2 = 1 \) and \( c = \sqrt{1/3} \). But \( \sqrt{1/3} < 1 \) is not in the interval \([1, 2]\), so something must be wrong: the assumption that there were at least two roots must be false, hence there is at most one.

**Bounded derivative ⇒ bounded function.** Suppose we know that \( 3 \leq f'(x) \leq 5 \) for all values of \( x \). Then we can bound, for instance, the difference between \( f(2) \) and \( f(8) \): by the MVT, there is a \( c \) such that (since \( d\frac{d}{dx} \sin(x) = \cos(x) \))

\[
\cos(c) = \frac{\sin(x) - \sin(0)}{x - 0} = \frac{\sin(x)}{x},
\]

so since \( \cos(c) < 1 \) (for \( 0 < x \leq 1 \)) we have \( \frac{\sin(x)}{x} < 1 \), hence \( \sin(x) < x \).

Another example is

\[
e^x > x + 1 \quad \text{for all } x \neq 0.
\]

Let \( x \) be any nonzero number, then the MVT gives \( c \) in the interval \([0, x]\) or \([x, 0]\) (whichever makes sense) such that

\[
e^c = \frac{e^x - e^0}{x - 0} = \frac{e^x - 1}{x}.
\]

If \( x > 0 \), then \( e^c > 1 \), and multiplying through by \( x \) gives \( e^x - 1 = e^c \cdot x > x \), so we get \( e^x > x + 1 \).

If \( x < 0 \), then \( e^c < 1 \), so \( e^x - 1 = e^c \cdot x > x \), since \( x \) is negative and \( e^c < 1 \). Hence we get \( e^x > x + 1 \) for all \( x \neq 0 \).
Chapter 10

Maclaurin Series and Taylor Series

10.1 Polynomial approximation

Recall that the linear approximation to a function \( f(x) \) around \( a = 0 \) is given by

\[
T_1(x) = f(0) + f'(0)x
\]

Why I renamed \( L(x) \) to \( T_1(x) \) will become clear soon. Because this is the function whose graph \( y = T_1(x) \) is the tangent line to \( y = f(x) \) at \( x = 0 \), this seems to be the best approximation to \( f(x) \) by a linear function. But what if we try to approximate \( f(x) \) using a quadratic function \( ax^2 + bx + c \), or using a polynomial of higher degree? That’s what this chapter is about.

Another way to look at it is that \( T_1(x) \) is the easiest function that has \( T_1(0) = f(0) \) and \( T_1'(0) = f'(0) \). This makes sense because \( f(0) \) and \( f'(0) \) are the most basic pieces of information about \( f \) around \( 0 \), so if we want to approximate \( f \) there, why not try to imitate \( f \) by making a new function with just those two pieces of information.

Then the next step would be to find the easiest function that has \( T_2(0) = f(0) \) and \( T_2'(0) = f'(0) \), and also \( T_2''(0) = f''(0) \). That should look like

\[
T_2(x) = f(0) + f'(0)x + ax^2,
\]

since then \( T_2(0) = f(0) \) and \( T_2'(0) = f'(0) \) already work; the quadratic term doesn’t change that since it and its derivative are zero at \( x = 0 \). Then we can determine \( a \) by taking the second derivative:

\[
T_2'(x) = f'(0) + 2ax,
\]

so if we want \( T_2''(0) = f''(0) \), we should take \( a = \frac{f''(0)}{2} \), which gives us the quadratic approximation to \( f(x) \) at \( a = 0 \):

\[
T_2(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2
\]

For instance, let \( f(x) = e^x \). If you use the linear approximation \( T_1(x) = e^0 + e^0x = 1 + x \) to approximate \( e^{0.1} = 1.10517 \ldots \), you get \( T_1(0.1) = 1.1 \). But if we use the quadratic approximation \( T_2(x) = e^0 + e^0x + \frac{e^0}{2}x^2 = 1 + x + \frac{x^2}{2} \), we get

\[
T_2(0.1) = 1 + 0.1 + \frac{0.01}{2} = 1.105,
\]
which is a better approximation. Similarly, to approximate $\sqrt{99.7} = 9.8498873 \cdots$ we can take the function $f(x) = \sqrt{100 + x}$ and use the linear approximation $T_1(x) = 10 + \frac{1}{20} x$, which gives $T_1(-0.3) = 10 - 3/200 = 9.985$, pretty good. But using $f''(x) = \frac{d}{dx} \left( \frac{1}{2}(x+100)^{-1/2} \right) = \frac{-1}{4} (x+100)^{-3/2}$, so $f''(0) = \frac{-1}{4} \cdot \frac{1}{\sqrt{100}} = -\frac{1}{4000}$, we get the quadratic approximation

$$T_2(x) = 10 + \frac{1}{20} x - \frac{1}{8000} x^2,$$

which gives the approximation $T_2(-0.3) = 10 - 3/200 - \frac{1}{8000} \frac{9}{100} = 9.98498875$; that’s much better.

Before we go even further, a little notation: we define $f^{(n)}(x)$ to be the $n$-th derivative of $f(x)$. So $f^{(1)} = f'$, $f^{(2)} = f''$, $f^{(3)} = f''', etc, and also $f^{(0)} = f$ itself.

Now let’s find the cubic approximation (a cubic is a polynomial of degree 3). It should look like

$$T_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2} x^2 + bx^3,$$

and we want to have $T_3^{(3)}(0) = T_3'''(0) = f^{(3)}(0)$. Since

$$T_3^{(1)}(x) = f'(0) + f''(0)x + 3bx^2, \quad T_3^{(2)}(x) = f''(0) + 6bx, \quad T_3^{(3)} = 6b,$$

to get $T_3^{(3)}(0) = f^{(3)}(0)$, we should choose $b = \frac{f^{(3)}(0)}{6}$, so the cubic approximation is

$$T_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2} x^2 + \frac{f^{(3)}(0)}{6} x^3.$$

Let’s apply this cubic approximation to the two examples above. For $f(x) = e^x$ we get $T_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$, which gives an approximation to $e^{0.1} = 1.10517 \cdots$ of $T_3(0.1) = 1 + 0.1 + 0.01/2 + 0.001/6 \approx 1.105167$, again an improvement.

For $f(x) = \sqrt{x}$ we get $f^{(3)}(x) = \frac{d}{dx} \frac{1}{4} (x+100)^{-3/2} = \frac{3}{8} (x+100)^{-5/2}$, and $f^{(3)}(0) = \frac{3}{800000}$, so the cubic approximation is

$$T_3(x) = 10 + \frac{1}{20} x - \frac{1}{8000} x^2 + \frac{3}{6 \cdot 800000} x^3,$$

which gives $T_3(-0.3) = 9.84988733125$, which again is closer to $\sqrt{99.7} = 9.8498873309 \cdots$.

### 10.2 Maclaurin polynomials

Just like we did above for $n = 1, 2, 3$, we can approximate a function around 0 by a polynomial $T_n(x)$ of any degree $n$. These polynomials are called Maclaurin polynomials. They are a special case of Taylor polynomials, which are approximations around some number $a$, which doesn’t have to be 0; we will get to those later, but I will focus on Maclaurin polynomials because they are simpler, and if you understand those, Taylor polynomials are only one step further.

The formulas we found for $T_2(x)$ and $T_3(x)$ generalize to the following formula for the Maclaurin polynomial of degree $n$:

$$T_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2!} x^2 + \frac{f^{(3)}(0)}{3!} x^3 + \frac{f^{(4)}(0)}{4!} x^4 + \cdots + \frac{f^{(n)}(0)}{n!} x^n$$
Here \( n! \) stands for \( n \cdot (n - 1) \cdot (n - 2) \cdot \ldots \cdot 2 \cdot 1 \), so \( 1! = 1 \), \( 2! = 2 \cdot 1 = 2 \), \( 3! = 3 \cdot 2 \cdot 1 = 6 \), \( 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24 \), etc. These factorials show up here because if you take \( g(x) = x^n \) and you differentiate it \( n \) times, you get \( g^{(n)}(x) = n \cdot x^{n-1} \), \( g^{(n)}(x) = n \cdot (n - 1) \cdot x^{n-2} \), etc.

So \( T_n(x) \) is constructed in such a way that \( T_n(0) = 0 \), \( T'_n(0) = f'(0) \), \( T''_n(0) = f''(0) \), and so on, up to \( T^{(n)}_n(0) = f^{(n)}(0) \), and that makes it a good approximation to \( f(x) \) around \( x = 0 \).

The easiest example to find the Maclaurin polynomials of is \( f(x) = e^x \), because \( f^{(n)}(x) = e^x \), so \( f^{(n)}(0) = 1 \) for all \( n \). Therefore:

\[
T_n(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots + \frac{1}{n!}x^n.
\]

We could use this to approximate numbers like \( e^{0.1} \) as closely as we want (although right now, to know how close we are, we need to know the actual value; soon we will see that Taylor’s inequality solves this). Next week we will see Newton’s method, which is actually even more efficient at approximating values than these polynomials. What’s special about these polynomials is that they approximate the whole function on an interval; with Newton’s method, to get approximations to different values, you need to do the same thing all over again, but once you have computed a Maclaurin polynomial, you can plug in any number (near 0), and you get your approximation.

What’s more, they also have lots of theoretical applications. For instance, we do not yet know how to evaluate limits like \( \lim_{x \to 0} \frac{e^x - 1}{x} \), where plugging in gives an undefined \( \frac{0}{0} \), and there is no algebraic simplification possible. But what if we approximate \( e^x \) in this limit by, say, \( T_2(x) = 1 + x + \frac{1}{2}x^2 \)? Then we get

\[
\lim_{x \to 0} \frac{(1 + x + \frac{1}{2}x^2) - 1}{x} = \lim_{x \to 0} \frac{x + \frac{1}{2}x^2}{x} = \lim_{x \to 0} \frac{1 + \frac{1}{2}x}{1} = 1.
\]

In fact, it is not hard to see that if you take \( T_3(x) \) instead, or any higher degree \( T_n(x) \), then the same happens. So if you believe me when I say that as you take larger and larger \( n \), you can approximate \( e^x \) as closely as you want by a \( T_n(x) \), then we must have

\[
\lim_{x \to 0} \frac{e^x - 1}{x} = 1.
\]

If you don’t believe me, wait till we learn about Taylor’s Inequality, then we will be able to prove this.

This approach in fact lets you compute pretty much any limit you could think of, except maybe ones with \( |x| \) or some other non-differentiable function (makes sense because then you can’t compute \( f^{(n)}(0) \)). Before we do more of this, I’ll go one step further and introduce Maclaurin series, which will allow us to skip over this approximation argument when doing limits this way, and that will make the computations a bit easier.

### 10.3 Maclaurin series

This section may seem a bit mysterious, and with good reason, because it uses some mathematics that will not be covered in this course at all. You can tell from the fact that in the book, this material comes at the end of chapter 11, and we don’t treat anything from the first 9 sections of that chapter. Nevertheless, if you’re able to suspend your disbelief, this is a very
useful simplification of the ideas around Maclaurin (and Taylor) polynomials, and it is a sneak peak at some deeper mathematics.

Above we saw that we could approximate functions and their values with Maclaurin polynomials, and higher degree polynomials give better approximations. What if we could instead use ‘polynomials’ of infinite degree? Would they be ‘infinitely close’, i.e. exactly right? Amazingly, yes. We call that the Maclaurin series of a function:

\[
    f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \frac{f^{(n+1)}(0)}{(n+1)!}x^{n+1} + \cdots
\]

This says that around 0, the function \( f \) is equal, not just approximately equal, but equal, to the thing on the right. The thing on the right is a ‘series’, an infinite sum of functions.

To see that an infinite sum can even make sense, think of
\[
    2 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots.
\]

So an infinite sum of numbers can be equal to something finite. Apparently, it is even possible for an infinite sum of functions to be finite for every number you plug in (from some interval), since that’s what the equation for the Maclaurin series above says.

Another way to think of it is that \( \lim_{n \to \infty} T_n(x) = f(x) \), i.e. if you ‘let the \( n \) in \( T_n(x) \) go to infinity’, you get exactly \( f(x) \).

There are two important aspects that I’ve ignored here:

• First of all, this only makes sense if \( f \) is infinitely differentiable at 0, i.e. if \( f^{(n)}(0) \) exists for any \( n \). Luckily, this is the case for many functions we know, like \( e^x \) and \( \sin(x) \).

• Second, it may not work for all \( x \), but possibly only for an interval around 0, and then outside that interval, the series does not equal a finite number at all. Whether or not it does will again depend on Taylor’s Inequality. The first example we will see that only works on a restricted interval is \( \frac{1}{1-x} \).

The basic example is again \( e^x \), whose Maclaurin series is

\[
    e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots
\]

We could also describe this series by saying that its ‘general term’ is \( \frac{1}{n!}x^n \). This one is in fact correct for any \( x \), so for instance

\[
    e = e^1 = 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots.
\]

Let’s find the Maclaurin series of \( \sin(x) \). Then

\[
    f'(x) = \cos(x), \quad f''(x) = -\sin(x), \quad f'''(x) = -\cos(x), \quad f^{(4)}(x) = \sin(x), \ldots,
\]

so

\[
    f'(0) = \cos(0) = 1, \quad f''(x) = -\sin(0) = 0, \quad f'''(x) = -\cos(0) = -1, \quad f^{(4)}(x) = \sin(0) = 0, \ldots,
\]

and if we wanted to, we could keep going with this list, because it repeats itself every four steps. Hence in the Maclaurin series for \( \sin(x) \), the even powers of \( x \) all drop out, and the odd ones alternately have a minus or not:

\[
    \sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \frac{1}{9!}x^9 + \cdots.
\]
This one also works for all \( x \), so we magically get all kinds of series, like
\[
\frac{1}{\sqrt{2}} = \sin(\pi/4) = \frac{\pi}{4} - \frac{1}{3!}\frac{\pi^3}{4^3} + \frac{1}{5!}\frac{\pi^5}{4^5} - \cdots
\]
Who knew \( \sqrt{2} \) and \( \pi \) were related?

To get the Maclaurin series of \( \cos(x) \), we could do similar computations, but instead we will use a nice trick. Since \( \frac{d}{dx} \sin(x) = \cos(x) \), the Maclaurin series for \( \cos(x) \) should be the derivative of the Maclaurin series for \( \sin(x) \), and we can compute that term by term. So
\[
\frac{d}{dx} x = 1, \quad \frac{d}{dx} \frac{-1}{3!} x^3 = -\frac{3}{3 \cdot 2 \cdot 1} x^2 = -\frac{1}{2!} x^2, \quad \frac{d}{dx} \frac{1}{5!} x^5 = \frac{1}{4!} x^4, \text{ etc,}
\]
which gives us
\[
\cos(x) = 1 - \frac{1}{2!} x^2 + \frac{1}{4!} x^4 - \frac{1}{6!} x^6 + \frac{1}{8!} x^8 + \cdots.
\]
Note that differentiating this again would give the Maclaurin series for \( -\sin(x) \) (and look back at \( e^x \): you can see that its Maclaurin series is its own derivative!).

Finally let’s do one that doesn’t work for all \( x \), but only on some interval around 0. I’ll just give it, and you could check the derivatives \( f^{(n)}(0) \) yourself. It is
\[
\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + \cdots, \quad \text{only on } (-1, 1).
\]
It makes sense that this doesn’t hold for all \( x \), because for instance for \( x = 1, \frac{1}{1-x} \) is not defined. What’s more, if you plug in \( x = 2 \), then \( f(2) = -1 \), but the series would be \( 1+2+4+8+16+\cdots \), which surely can’t be finite, because each term is larger than the last. In fact, for any \( x \) with \( |x| \geq 1 \), we would get a series with each term larger than the previous one, which can never give a finite number.

As a little trick to see this equality, try multiplying the right side by \( 1 - x \), then we get
\[
(1-x)(1+x+x^2+\cdots) = (1-x) + (1-x) \cdot x + (1-x) \cdot x^2 + \cdots = 1-x+x-x^2+x^2-x^3+\cdots,
\]
and since every term other than 1 gets cancelled, this equals 1.

### 10.4 Computing with Maclaurin series

We’ve seen how to compute the coefficients in Maclaurin series using repeated derivatives, but that is often a lot of work. Fortunately, we can compute a lot of them by reducing them to the four basic ones we saw above (in this course, those four are the only ones that you should know by heart). Here are some examples:
\[
e^{3x} = 1 + (3x) + \frac{1}{2!}(3x)^2 + \frac{1}{3!}(3x)^3 + \cdots = 1 + 3x + \frac{9}{2!}x^2 + \frac{27}{3!}x^3 + \cdots
\]
\[
\frac{\sin(x^2)}{x^2} = \frac{1}{x^2} \cdot (x^2 - \frac{1}{3!}(x^2)^3 + \frac{1}{5!}(x^2)^5 - \cdots) = 1 - \frac{1}{3!}x^4 + \frac{1}{5!}x^8 - \cdots
\]
\[
\frac{1}{1+x} = \frac{1}{1-(-x)} = 1 + (-x) + (-x)^2 + (-x)^3 + \cdots = 1 - x + x^2 - x^3 + \cdots
\]
\[
\frac{x}{1+x^2} = x \cdot \frac{1}{1-(-x^2)} = x \cdot (1 + (-x^2) + (-x^2)^2 + (-x^2)^3 + \cdots) = x - x^3 + x^5 - x^7 + \cdots
\]
With some more clever trickery, we can find the Maclaurin series of \( \ln(1 + x) \). Note that \( \ln(x) \) does not exist at 0, so \( \ln(x) \) has no Maclaurin series itself; the Maclaurin series for \( \ln(1 + x) \) is actually a Taylor series for \( \ln(x) \) around \( x = 1 \) in disguise. We can use that 
\[
\frac{d}{dx} \ln(1 + x) = \frac{1}{1 + x} = 1 - x + x^2 - x^3 + \cdots .
\]
Then it’s not hard to see what Maclaurin series has this as a derivative: it must be 
\[
\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots .
\]
Because the Maclaurin series for \( \frac{1}{1-x} \) only works on \((-1, 1)\), the same is true of this one.

### 10.5 Limits with Maclaurin series

Now that we know about Maclaurin series, limit computations like we did above become a lot easier:

\[
\lim_{x \to 0} \frac{e^x - 1}{x} = \lim_{x \to 0} \frac{(1 + x + x^2 + \cdots) - 1}{x} = \lim_{x \to 0} \frac{x + x^2 + \cdots}{x} = \lim_{x \to 0} \frac{1 + x + \cdots}{1} = 1.
\]

We only have to write down as many of the terms in the Maclaurin series as we need, since we know what happens to the rest.

I will give a few examples that hopefully speak for themselves. You should check first that plugging in actually gives \( \frac{\pi}{6} \! \)

\[
\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{x - \frac{1}{3!}x^3 + \cdots}{x} = \lim_{x \to 0} \frac{1 - \frac{1}{3!}x^2 + \cdots}{1} = 1
\]

\[
\lim_{x \to 0} \frac{\sin(mx)}{\sin(nx)} = \lim_{x \to 0} \frac{(mx) - \frac{1}{3!(mx)^3} + \cdots}{(nx) - \frac{1}{3!(nx)^3} + \cdots} = \lim_{x \to 0} \frac{m - \frac{1}{3!}(mx)^2 + \cdots}{n - \frac{1}{3!}(nx)^2 + \cdots} = \frac{m}{n}
\]

\[
\lim_{x \to 0} \frac{\cos(2x) - e^{x^2}}{x^2} = \lim_{x \to 0} \frac{(1 - \frac{1}{2!(2x)^2} + \frac{1}{4!}(2x)^4 - \cdots) - (1 + (x)^2 + (x^2)^2 + \cdots)}{x^2}
\]

\[
= \lim_{x \to 0} \left(-\frac{1}{2}4 - 1\right)x^2 + \left(\frac{1}{4!}2^4 - 1\right)x^4 + \cdots = -\frac{1}{2}4 - 1 = -3.
\]

### 10.6 Taylor Polynomials and series

We saw that Maclaurin polynomials give approximations to functions around \( 0 \). Taylor polynomials are the more general version of Maclaurin polynomials, and they approximate functions around \( a \), for some number \( a \).

The formulas for Taylor polynomials are basically the same as for Maclaurin polynomials, except that instead of writing them as a polynomials in \( x \), we write them as polynomials in \( x - a \), because then it will be easier to plug in values close to \( a \). For example, the Taylor polynomial of degree 2 for \( f(x) = \sqrt{x} \) around \( a = 25 \) looks like

\[
T_2(x) = 5 + \frac{1}{10}(x - 25) - \frac{1}{1000}(x - 25)^2.
\]

This gives an approximation to \( \sqrt{26} \) of \( 5 + \frac{1}{10} \cdot 1 - \frac{1}{1000} \cdot 1^2 = 5.099 \). If we had multiplied out this polynomial, plugging in 26 would have been much more work.
The general formula for the Taylor polynomial for \( f(x) \) around \( x = a \) is
\[
T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n
\]

Note that if we take \( x = a \), then we would get exactly the formula for the Maclaurin polynomials that we saw above. Also, the Taylor polynomial \( T_1(x) = f(a) + f'(a)(x-a) \) is nothing but the linear approximation.

As an example, let’s find the Taylor polynomial of degree 3 for \( e^x \) around \( a = 2 \). Then \( f^{(n)}(a) = e^2 \) for any \( n \), so we get
\[
T_3(x) = e^2 + e^2(x-2) + \frac{e^2}{2}(x-2)^2 + \frac{e^2}{6}(x-2)^3.
\]

If we used this to approximate \( e^{2.1} \), we would get
\[
e^{2.1} \approx e^2 + e^2 \cdot \frac{1}{10} + \frac{e^2}{2} \cdot \frac{1}{100} + \frac{e^2}{6} \cdot \frac{1}{1000} = e^2 \cdot \left( 1 + \frac{1}{10} + \frac{1}{200} + \frac{1}{6000} \right) = 8.166138,
\]
which is a pretty good approximation to \( e^{2.1} = 8.166170 \cdots \).

The point of using the approximation around 2 is that if we had used the Maclaurin polynomial of the same degree, which is \( T_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} \), we would have gotten the approximation \( e^{2.1} \approx 1 + 2.1 + \frac{(2.1)^2}{2} + \frac{(2.1)^3}{6} = 16.771 \), which is terrible.

### 10.7 Taylor’s Inequality and error estimation

So far we have seen lots of approximations, but to see if they were any good, we had to compare them to the actual value that we got from a calculator. Taylor’s inequality provides a way around this, by letting us bound the error \( |f(x) - T_n(x)| \), which is the difference between the actual function and the Taylor polynomial at \( x \). Of course, we have seen that a Taylor approximation is only any good near the number \( x = a \) that we use in the approximation, so such an error bound will only work on an interval around \( x = a \). So let’s suppose we want to use the Taylor approximation on some interval \([c, d]\) that contains \( a \). If for example we wanted to approximate \( e^{2.1} \) using \( a = 2 \), we would use the smallest possible interval that contains both 2 and 2.1, which is [2, 2.1].

**Taylor’s Inequality**

If \( T_n(x) \) is the Taylor polynomial for \( f(x) \) on the interval \([c, d]\) around \( x = a \), and if we know a number \( M \) such that \( |f^{(n+1)}(x)| \leq M \) on that interval \([c, d]\), then
\[
|f(x) - T_n(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1}.
\]

Note that this almost says that the error is bounded by what would be the next term in the Taylor polynomial, \( \frac{f^{(n+1)}(a)}{(n+1)!}|x-a|^{n+1} \), except that we have to replace \( f^{(n+1)}(a) \) by the upper bound \( M \) for \( |f^{(n+1)}(x)| \) over the interval.

Above we used \( T_2(x) \) for \( f(x) = \sqrt{x} \) around \( a = 25 \) to approximate \( \sqrt{26} \). To bound the error in this approximation, we need to compute \( f''(x) = \frac{3}{8}x^{-5/2} \). On the interval [25, 26], this is bounded by
\[
M = f''(25) = \frac{3}{8} \cdot \frac{1}{25^{5/2}} = \frac{3}{8} \cdot \frac{1}{5^5} = \frac{3}{2500}.
\]
this is because the function $x^{-5/2}$ is decreasing, so the maximum on $[25, 26]$ occurs at 25. Hence Taylor’s Inequality with $a = 25$, $x = 26$ and $n = 2$ gives

$$|\sqrt{26} - T_2(26)| \leq \frac{3/25000}{3!} |26 - 25|^3 = \frac{1}{50000} = 0.00002.$$  

And indeed, the true error is

$$\sqrt{26} - T_2(26) = 5.0990195 - 5.099 = 0.0000195,$$

which is less than 0.00002.

We can also use Taylor’s inequality to estimate the error in a linear approximation. For instance, if we approximate $e^{-0.1}$ with the linear approximation $T_1(x) = 1 + x$ we got $T_1(-0.1) = 0.9$. We use Taylor’s inequality with $a = 0$, $n = 1$ on the interval $[-0.1, 0]$, so we can take $M = e^0 = 1$ (now since the function is increasing, we plug in the right endpoint of the interval). Then

$$|e^{-0.1} - T_1(-0.1)| \leq \frac{M}{2!} |-0.1 - 0|^2 = \frac{1}{2} (0.1)^2 = \frac{1}{200} = 0.005.$$  

Therefore we know that $e^{-0.1}$ is between $0.9 - 0.005 = 0.895$ and $0.905$, without having used a calculator (with a calculator we would get $0.9048\cdots$).

Another typical approach is to ask for an approximation that has error less than some number. For instance, we could ask for an approximation to $e^{-0.1}$ that has an error of less than $0.0000001$ (so the linear approximation is not good enough). Then we have to find the $n$ such that the error bound for $T_n$ is less than $0.0000001$, so with $M = 1$ as above we need

$$0.0000001 \geq \frac{M}{n!} |-0.1 - 0|^n + 1 = \frac{1}{n! \cdot 10^{n+1}}.$$  

For $n = 2$, this is $\frac{1}{6\cdot1000}$, not good enough yet, for $n = 3$ this is $\frac{1}{24\cdot10^5} = 0.000004\cdots$, still not quite good enough, and finally for $n = 4$ we get $\frac{1}{120\cdot10^{10}} = 0.00000008\cdots$, which will do. So we have to use $T_4(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4$, which gives

$$T_4(-0.1) = 1 - \frac{1}{10} + \frac{1}{20} - \frac{1}{600} + \frac{1}{240000} = 0.9048375,$$

and with a calculator we would get $e^{-0.1} = 0.904837418036\cdots$, so the true error is indeed

$$0.000000082\cdots < 0.0000001.$$  

68
Chapter 11
Antiderivatives & Substitution

11.1 Antiderivatives

• Reminder about Derivatives
First I’ll remind you of a few derivatives that we’ll be using a lot:

\[(x^n)' = nx^{n-1},\]
\[(\sin x)' = \cos x, \quad (\cos x)' = -\sin x, \quad (\tan x)' = \sec^2 x,\]
\[(\sec x)' = \sec x \tan x, \quad (\sin^{-1} x)' = \frac{1}{\sqrt{1-x^2}}, \quad (\tan^{-1} x)' = \frac{1}{x^2 + 1},\]
\[(e^x)' = e^x, \quad (\ln |x|)' = \frac{1}{x}.\]

• About \((\ln |x|)' = \frac{1}{x}\)
The last one deserves some extra attention, since this will show up again and again in integration.
You may only remember the derivative \((\ln x)' = \frac{1}{x}\); this is of course correct, but is only true for
\(x\) in the domain of \(\ln x\), so for \(x > 0\). For \(x \leq 0\), \(\ln x\) doesn’t exist, and so a derivative wouldn’t make sense.
We can (and should) extend this differentiation formula to negative \(x\) by taking \(\ln |x|\) instead.
Think of the graphs: \(\ln x\) only has a graph to the right of the \(y\)-axis, while \(\ln |x|\) also has the mirror image of the same graph to the left of the \(y\)-axis. We do have to actually check that the
derivative is still \(1/x\) for \(x < 0\), which we do as follows:

\[x < 0 \Rightarrow |x| = -x \Rightarrow (\ln |x|)' = (\ln(-x))' = \frac{1}{-x} \cdot (-1) = \frac{1}{x}.\]

• Antidifferentiation
Antidifferentiation (or ‘taking the derivative’) is doing differentiation in reverse: given a function
\(f(x)\), find a function \(F(x)\) such that \(F'(x) = f(x)\). This \(F(x)\) is then called an antiderivative
of \(f(x)\).
Here are a few examples of important antiderivatives that you can obtain by reversing the
differentiation formulas above:

\[f(x) = x^n \rightarrow F(x) = \frac{1}{n+1}x^{n+1} \quad (\text{for } n \neq -1),\]
A bit of notation: we will write

\[ f(x) = \sin x \quad \rightarrow \quad F(x) = -\cos x, \quad f(x) = \cos x \quad \rightarrow \quad F(x) = \sin x, \]
\[ f(x) = \frac{1}{x^2+1} \quad \rightarrow \quad F(x) = \tan^{-1} x, \quad f(x) = \frac{1}{\sqrt{1-x^2}} \quad \rightarrow \quad F(x) = \sin^{-1} x \]
\[ f(x) = e^x \quad \rightarrow \quad F(x) = e^x, \quad f(x) = \frac{1}{x} \quad \rightarrow \quad F(x) = \ln |x| \]

**Examples**

Just like with differentiation, we can combine these basic antiderivatives into more complex ones. I will give a few examples first, and then explain some of the steps. Note that an antiderivative is easily checked: just differentiate it and see if you get the orginal function back. But I won’t do that in this writeup.

\[ f(x) = 3x^2 - 5 \sin(x) \quad \rightarrow \quad F(x) = 3 \cdot \frac{1}{8} \cdot x^8 - 5 \cdot (-\cos(x)) = \frac{3}{8} x^8 + 5 \cos(x), \]
\[ f(x) = e^{3x} + \frac{2}{\sqrt{1-x^2}} + 1 \quad \rightarrow \quad F(x) = \frac{1}{3} e^{3x} + 2 \sec(x) + x, \]
\[ f(x) = \sin(2x - 3) \quad \rightarrow \quad F(x) = -\frac{1}{2} \cos(2x - 3), \]
\[ f(x) = (3x + 7)^5 + \frac{1}{x+1} \quad \rightarrow \quad F(x) = \frac{1}{6} (3x + 7)^6 + \frac{1}{3} + \ln |x + 1| = \frac{1}{18} (3x + 7)^6 + \ln |x + 1| \]

**‘Compensating’**

There is one informal trick that I used here several times; let me explain it for \( \sin(2x - 3) \). We can guess that an antiderivative should involve \(-\cos(2x - 3)\), but if we differentiate that, we would get \(2 \sin(2x - 3)\) (with the 2 coming from the chain piece \((2x - 3)'\)), which is not quite what we want. However, we can fix it by ‘compensating’ for the 2 by putting a \(\frac{1}{2}\) in front of \(-\cos(2x - 3)\). And indeed, then it works:

\[
\left(-\frac{1}{2} \cos(2x - 3)\right)' = -\frac{1}{2} \cdot (-\sin(2x - 3) \cdot 2) = \sin(2x - 3).
\]

In general, if you know that \(F(x)\) is an antiderivative for \(f(x)\), then it’s easy to get one for \(f(ax + b)\): you compensate for the \(a\) with \(\frac{1}{a}\), and you get \(\frac{1}{a} F(ax + b)\).

In the next section we will see a more systematic method that also includes this trick, but often it might be easier to do it this way.

**A function has infinitely many antiderivatives**

Above I intentionally said ‘an’ antiderivative instead of ‘the’ antiderivative, because any function has many, in fact infinitely many, antiderivatives.

To see this, consider \(f(x) = 2x\). We can easily guess one antiderivative, \(F(x) = x^2\). But \(G(x) = x^2 + 1\) is also a derivative: \((x^2 + 1)' = 2x + 0 = 2x\). And so is \(x^2 - 17\), etc.

In general, if \(F(x)\) is an antiderivative of \(f(x)\), then there are infinitely many antiderivatives of the form \(F(x) + C\), for some constant \(C\), and there are no others.

**Indefinite Integral**

A bit of notation: we will write \(\int f(x) \, dx\) for the ‘general antiderivative’ \(F(x) + C\) of \(f(x)\), where \(C\) is a constant.

For example:

\[
\int 4x^2 \, dx = \frac{4}{3} x^3 + C, \quad \int e^{-5x} \, dx = -\frac{1}{5} e^{-5x} + C, \quad \int \frac{1}{x-2} \, dx = \ln |x-2| + C.
\]

We call \(\int f(x) \, dx\) the *indefinite integral* of \(f(x)\), and then we call \(f(x)\) the *integrand* of the integral. Evaluating an integral is called *integration*.

**Antidifferentiation Rules**

In differentiation, we had lots of rules that let us differentiate combinations of functions. In
antidifferentiation, there aren’t nearly as many rules; the ones we’ve used above are: the sum rule, the constant multiple rule, and a rule for compositions of functions where the inner function is a linear function \((ax + b)\):

\[
\begin{align*}
\int (f(x) + g(x)) \, dx &= \int f(x) \, dx + \int g(x) \, dx \\
\int af(x) \, dx &= a \int f(x) \, dx \\
\int f(ax + b) \, dx &= \frac{1}{a} F(ax + b) + C, \quad \text{if } F(x) \text{ is an antiderivative of } f.
\end{align*}
\]

However, for integrals there is in general no product rule for \(\int f(x)g(x) \, dx\), no quotient rule for \(\int \frac{f(x)}{g(x)} \, dx\), and no chain rule for \(\int f(g(x)) \, dx\).

Unlike for differentiation, where every function has a derivative and we have the rules to find it, for antidifferentiation there will be many functions that we won’t be able to find an antiderivative for.

### 11.2 Substitution

- **Turning the chain rule around**
  As I said, for integration there is no analog to the chain rule, in the sense of a formula for \(\int f(g(x)) \, dx\). However, we do get a formula by simply turning the chain rule around:

  \[
  \int f'(g(x)) \cdot g'(x) \, dx = f(g(x)).
  \]

  Of course, this won’t be nearly as useful as a formula for \(\int f(g(x)) \, dx\) would have been, but it will have to do.

  To apply to this to an integral, we’ll have to identify a \(g\) and an \(f\) such that the integrand looks like \(f'(g(x)) \cdot g'(x)\). This may involve some guessing.

  - For instance, to evaluate \(\int 2x \cos(x^2) \, dx\) we can guess \(g'(x) = 2x\), and then deduce the rest:

    \[
    \begin{align*}
    g'(x) &= 2x \\
    g(x) &= x^2 \\
    f'(g(x)) &= \cos(g(x)) \\
    f'(x) &= \cos(x) \\
    f(x) &= \sin(x),
    \end{align*}
    \]

    \[
    \Rightarrow \int 2x \cos(x^2) \, dx = f(g(x)) = \sin(x^2) + C.
    \]

  - Here’s a harder one:

    \[
    \int \frac{\cos(x)}{\sin(x)} \, dx.
    \]

    We can see the integrand as \(\frac{1}{\sin(x)} \cdot \cos(x)\), and then guess \(g'(x) = \cos(x)\), and deduce:

    \[
    \begin{align*}
    g'(x) &= \cos(x) \\
    g(x) &= \sin(x) \\
    f'(g(x)) &= \frac{1}{g(x)} \\
    f'(x) &= \frac{1}{x} \\
    f(x) &= \ln |x|,
    \end{align*}
    \]

    \[
    \Rightarrow \int \frac{\cos(x)}{\sin(x)} \, dx = f(g(x)) = \ln |\sin(x)| + C.
    \]

- **Substitution**  (aka change of variables)
  We can do the above trick more systematically, in several steps. Basically, we guess what the inner function \(g(x)\) is, then change variables to \(u = g(x)\), turning the integral into one in terms of \(u\), which will hopefully be simpler.

  I will show the steps for the example \(\int 2x \cos(x^2) \, dx\):
• Substitute $u = x^2$.
• Replace $dx$ by $du$, using $\frac{du}{dx} = 2x \Rightarrow du = 2xdx$.
• Solve the resulting $\int f(u)du$.
• Write in terms of $x$ again, using $u = x^2$.

It will look like this:

$$\int 2x \cos(x^2)dx \ \overset{u = x^2}{=} \ \int \cos(u) \cdot (2xdx) \ \overset{du = 2xdx}{=} \ \int \cos(u)du = \sin(u) + C = \sin(x^2) + C.$$

**Examples**

◦ Let’s do

$$\int x^2(x^3 + 1)^7 dx.$$

If we pick $u = x^3$, then $\frac{du}{dx} = 3x^2$, so $du = 3x^2dx$. There’s no 3 in the integrand, so it will be easier to use $\frac{1}{3}du = x^2dx$. Then

$$\int x^2(x^3 + 1)^7 dx \ \overset{u = x^3}{=} \ \left(\frac{1}{3}\right) \int (u+1)^7 u^{\frac{1}{3}}du = \frac{1}{3} \cdot \frac{1}{8} (u+1)^8 + C = \frac{1}{24} (x^3 + 1)^8 + C.$$

◦ Let’s do a slightly harder one:

$$\int \frac{e^x}{(e^x)^2 + 1} dx.$$

If we pick $u = e^x$, then $du = e^x dx$, and we get:

$$\int \frac{e^x}{(e^x)^2 + 1} dx = \int \frac{1}{u^2 + 1} \cdot (e^x dx) = \int \frac{1}{u^2 + 1} du = \arctan(u) + C = \arctan(e^x) + C.$$

**How to choose $u$?**

To apply the substitution method, first the integrand should look like a product of two functions. Then the 'smaller' (or 'simpler') of those two might be your $g'(x) = u'$, and if you know an antiderivative of that one, then that will be your $g(x) = u$. If you recognize this $u$ inside the 'bigger' function, that’s a good sign. But you’ll have to try and see what comes out.

I’m not saying this always works; there are plenty of exceptions, though for the examples we’ve seen so far it works. When this does fail, you’ll have to try choosing $u$ to be any smaller function that you see inside another one. For example, if the 'bigger' function is $\tan(\sqrt{x+2})$, you can try $u = x + 2$ or $u = \sqrt{x+2}$; either might work.

So to summarize:

• Write your integrand as a product of a 'smaller' and a 'bigger' function;
• Let $u'$ be the smaller function, antidifferentiate it to get $u$, and try substituting that;
• If the above fails, try choosing $u$ to be any inner function.

There is often more than one choice for $u$. In the first of the two examples above, we could have used $u = x^3 + 1$, then $\frac{1}{3}du = x^2dx$ again, and you get

$$\int x^2(x^3 + 1)^7 dx = \frac{1}{3} \int u^7 du = \frac{1}{3} \cdot \frac{1}{8} u^8 + C = \frac{1}{24} (x^3 + 1)^8 + C,$$
just like before.
It’s important to make sure that after you substituted \( u \) and \( du \), the \( x \) is really gone. Otherwise the substitution is useless, like in this example:

\[
\int 2x \sin(x^2 + x + 1) \, dx = \int \sin(u + x + 1) \, du = \text{??？}
\]

This is a common mistake. (In this example, I don’t think there is an antiderivative.)

• More examples
  
  ◦ Let’s apply this to

  \[
  \int \sqrt{\tan(x) - 3} \sec^2(x) \, dx.
  \]

  The integrand is a product of \( \sqrt{\tan(x) - 3} \) and \( \sec^2(x) \), the ’smaller’ one is \( \sec^2(x) \), and we know (from page 1) an antiderivative for that: \( \tan(x) \), which does indeed show up in the ’bigger’ factor. So we choose \( u = \tan(x) \), which implies \( du = \sec^2(x) \, dx \), and we can work it out:

  \[
  \int \sqrt{\tan(x) - 3} \cdot (\sec^2(x) \, dx) = \int \sqrt{u - 3} \, du = \frac{1}{3/2} (u - 3)^{3/2} + C = \frac{2}{3} (\tan(x) - 3)^{3/2} + C.
  \]

  ◦ Let’s try

  \[
  \int \frac{\cos(\ln |x|)}{x} \, dx.
  \]

  To see it as a product of a ’smaller’ and a ’bigger’ function, we should write the integrand as \( \cos(\ln |x|) \cdot \frac{1}{2} \). That should lead us to try \( u' = \frac{1}{2} \), which implies \( u = \ln |x| \). In the ’bigger’ function we can recognize this \( u \) inside \( \cos(x) \). So let’s substitute \( u = \ln |x| \), with \( du = \frac{1}{x} \, dx \):

  \[
  \int \frac{\cos(\ln |x|)}{x} \, dx = \int \cos(\ln |x|) \cdot \left( \frac{1}{x} \, dx \right) = \int \cos(u) \, du = \sin(u) + C = \sin(\ln |u|) + C.
  \]

• Manipulating first

  You will see many integrals where there doesn’t seem to be a possible substitution right away. Often you will have to do some kind of manipulation of the integrand, by algebra, trigonometry or exponential/logarithmic formulas, to get it into a form where you can do a substitution (or some other method that we’ll learn later). In the last section of this chapter I will remind you of some of the trig or exponential identities that you might have to use.

  ◦ For example, consider

  \[
  \int \frac{(x + 1)^2}{x} \, dx.
  \]

  You can see this as a product in several ways, but none will lead to a useful substitution. However, a little algebra will make the integral easy (and we won’t even need substitution):

  \[
  \int \frac{(x + 1)^2}{x} \, dx = \int \frac{x^2 + 2x + 1}{x} \, dx = \int \left( \frac{x^2}{x} + \frac{2x}{x} + \frac{1}{x} \right) \, dx = \int x \, dx + \int 2 \, dx + \int \frac{1}{x} \, dx = \frac{1}{2} x^2 + 2x + \ln |x| + C.
  \]

  ◦ A different kind of example is

  \[
  \int \frac{1}{x^2 + 9} \, dx.
  \]
This should make you think of \( \int \frac{1}{x^2 + 1} \, dx = \arctan(x) + C \), except that the 9 is in the way. But you can fix that with some algebra:

\[
\frac{1}{x^2 + 9} = \frac{1}{9} \cdot \frac{1}{x^2 + 1} = \frac{1}{9} \cdot \frac{1}{\left(\frac{x}{3}\right)^2 + 1}.
\]

But now there's a 3 – why is that better? Because you can substitute it away, using \( u = x/3 \), \( du = \frac{1}{3} \, dx \):

\[
\int \frac{1}{x^2 + 9} \, dx = \frac{1}{9} \int \frac{1}{u^2 + 1} \cdot 3 \, du = \frac{3}{9} \arctan(u) + C = \frac{1}{3} \arctan \left( \frac{x}{3} \right) + C.
\]

○ Another one:

\[
\int e^x + e^{-x} \, dx.
\]

It's hard to guess what to do. But maybe (maybe) it reminds you of one we did earlier, namely \( \int \frac{e^x}{(e^x)^2 + 1} \, dx \). And we can turn it into that one:

\[
\frac{1}{e^x + e^{-x}} = \frac{e^x}{e^x} \cdot \frac{1}{e^x + e^{-x}} = \frac{e^x}{(e^x)^2 + e^x \cdot e^{-x}} = \frac{e^x}{(e^x)^2 + 1}.
\]

So we can do it with the substitution \( u = e^x \), like before:

\[
\int \frac{1}{e^x + e^{-x}} \, dx = \int \frac{e^x}{(e^x)^2 + 1} \, dx = \int \frac{1}{u^2 + 1} \, du = \arctan(u) + C = \arctan(e^x) + C.
\]

○ Now consider

\[
\int \frac{x + 1}{x - 1} \, dx.
\]

One way to do this is with the substitution \( u = x - 1 \): then \( du = dx \), and to get rid of the \( x \) in the numerator we can plug in \( x = u + 1 \) (which we get by rewriting \( u = x - 1 \)):

\[
\int \frac{x + 1}{x - 1} \, dx = \int \frac{u + 1}{u} \, du = \int \frac{(u + 1)}{u} \, du = \int \frac{u + 2}{u} \, du
\]

\[
= \int \left( \frac{u}{u} + \frac{2}{u} \right) \, du = \int 1 \, du + \int \frac{2}{u} \, du = u + 2 \ln |u| + C_1 = x - 1 + 2 \ln |x - 1| + C_1.
\]

But we can also do this one without a substitution, if we use the following algebra trick:

\[
\int \frac{x + 1}{x - 1} \, dx = \int \frac{(x - 1 + 1)}{x - 1} \, dx = \int \frac{x - 1}{x - 1} + \frac{2}{x - 1} \, dx
\]

\[
= \int 1 \, dx + \int \frac{2}{x - 1} \, dx = x + 2 \ln |x - 1| + C_2.
\]

That was a little easier.

But wait! They’re different: the first has \( x - 1 \), the second \( x \). But this is okay, it just means that the \( C \)'s are different, which is why I’ve given them different names. If for \( C_2 \) we plug in \( C_2 = C_1 - 1 \), the second result becomes the same as the first.

○ Here's a harder example that has two fairly different choices for \( u \), neither of which is easily recognizable:

\[
\int \frac{x}{\sqrt{x - 2}} \, dx.
\]

The 'smaller' function would be \( x \), but if \( u' = x \) then \( u = \frac{1}{2} x^2 \), and we can't substitute that. So we'll have to try any of the smaller functions that occurs inside. There are 2 and it turns out they both work:
First we’ll use \( u = x - 2 \), so \( du = dx \). Then we’ll have to use \( x = u + 2 \) to get rid of the \( x \):

\[
\int \frac{x}{\sqrt{x - 2}} \, dx = \int \frac{u + 2}{\sqrt{u}} \, du
\]

This is an integral we can do after a little algebra: \( \frac{u+2}{\sqrt{u}} = \frac{u}{\sqrt{u}} + \frac{2}{\sqrt{u}} = u^{1/2} + 2u^{-1/2} \). Using that we get:

\[
= \int u^{1/2} \, du + 2 \int u^{-1/2} \, du = \frac{1}{3/2} u^{3/2} + 2 \cdot \frac{1}{1/2} u^{1/2} + C
\]

\[
= \frac{2}{3} (x - 2)^{3/2} + 4(x - 2)^{1/2} + C.
\]

The other option is to take \( u = \sqrt{x - 2} \), so \( 2 \, du = \frac{1}{\sqrt{x - 2}} \, dx \), and to get rid of \( x \) we can solve for \( x = u^2 + 2 \). Then

\[
\int x \cdot \left( \frac{1}{\sqrt{x - 2}} \right) \, dx = 2 \int (u^2 + 2) \, du = \frac{2}{3} u^3 + 4u + C
\]

\[
= \frac{2}{3} (x - 2)^{3/2} + 4(x - 2)^{1/2} + C.
\]

**Using trig formulas**

Another class of integrals is those where you have to first manipulate the integrand with some trig identities, like:

\[
\int \tan(x) \, dx = \int \frac{\sin(x)}{\cos(x)} \, dx = \int \frac{1}{u} (\sin(x) \, dx) = \int \frac{1}{u} \cdot \frac{du}{\sin(x)} \, dx
\]

\[
- \int \frac{1}{u} \, du = - \ln |u| + C = \ln |\cos(x)| + C.
\]

That was just using the definition of tan, but here’s a common one where you need the identity \( \sin^2(x) = \frac{1}{2} - \frac{1}{2} \cos(2x) \) to lower the exponent of sin:

\[
\int \sin^2(x) \, dx = \int \left( \frac{1}{2} - \frac{1}{2} \cos(2x) \right) \, dx = \int \frac{1}{2} \, dx + \frac{1}{2} \int \cos(2x) \, dx = \frac{1}{2} x + \frac{1}{4} \sin(2x) + C.
\]

We’ll see many more of these later.

**Exponentials with other bases**

We also need to be able to deal with functions that involve exponentials to bases other than \( e \), like \( 10^x \). Fortunately, with the identity \( a^x = e^{\ln(a)x} \) you can rewrite those in terms of an exponential to base \( e \), and use the (easy) antiderivative that you know for \( e^x \).

For instance,

\[
\int 10^x \, dx = \int \left( e^{\ln(10)} \right)^x \, dx = \int e^{\ln(10)x} \, dx = \frac{1}{\ln(10)} e^{\ln(10)x} + C = \frac{1}{\ln(10)} 10^x + C.
\]

Note that \( \ln(10) \) is just a number, so \( \int e^{\ln(10)x} \, dx \) works just like \( \int e^{3x} \, dx = \frac{1}{3} e^{3x} + C \). Of course we can make them a little more complicated:

\[
\int \frac{2^n}{\sqrt{x}} \, dx = \int \left( \frac{1}{\sqrt{x}} \right)^{2n} \, dx = 2^{n} \int \left( \frac{1}{\sqrt{x}} \right)^{2} \, dx = \frac{2}{2n} \int 2^n \, du
\]

75
\[ \int e^{\ln(2)u} du = 2 \cdot \frac{1}{\ln(2)}e^{\ln(2)u} + C = \frac{2}{\ln(2)}2^{\sqrt{2}} + C. \]

**Trig & exponential identities**

Here I will remind you of a few formulas that you will often need in integration. You should also be familiar with the graphs of all these functions; I won’t reproduce those here, but you can find them in the book in sections 1.3, 1.4 and 6.7.

First the fundamental identities for sin and cos:

\[ \sin^2(x) + \cos^2(x) = 1, \quad \sin(2x) = 2\sin(x)\cos(x), \quad \cos(2x) = \cos^2(x) - \sin^2(x). \]

I call these ‘fundamental’ because you can derive all the other ones from them. For instance, plugging the first one into the third one in the form \( \sin^2(x) = 1 - \cos^2(x) \) gives

\[ \cos(2x) = \cos^2(x) - (1 - \cos^2(x)) = 2\cos^2(x) - 1, \]

which gives us the following very useful formula for lowering the exponent of a cos; and I’ll also give the analogous one for sin, which is derived similarly:

\[ \cos^2(x) = \frac{1}{2} + \frac{1}{2}\cos(2x), \quad \sin^2(x) = \frac{1}{2} - \frac{1}{2}\cos(2x). \]

Then there are the other trig functions, defined as follows:

\[ \sec(x) = \frac{1}{\cos(x)}, \quad \csc(x) = \frac{1}{\sin(x)}, \quad \tan(x) = \frac{\sin(x)}{\cos(x)}, \quad \cot(x) = \frac{1}{\tan(x)} = \frac{\cos(x)}{\sin(x)}. \]

The only two identities for these that you might need are the ones that you get by dividing \( \sin^2(x) + \cos^2(x) = 1 \) by respectively \( \cos^2(x) \) and \( \sin^2(x) \):

\[ \tan^2(x) + 1 = \sec^2(x), \quad 1 + \cot^2(x) = \csc^2(x). \]

Finally, there’s inverse functions for all of these, but there aren’t many important identities for those. The ones you’ll see most are \( \arctan(x) = \tan^{-1}(x) \) and \( \arcsin(x) = \sin^{-1}(x) \); occasionally you might see \( \sec^{-1}(x) \).

There are also some identities that you should know for \( e^x \) and \( \ln(x) \):

\[ e^0 = 1, \quad e^{a+b} = e^a \cdot e^b, \quad e^{a-b} = (e^a)^b = (e^b)^a, \quad e^{-a} = \frac{1}{e^a}. \]

\[ \ln(1) = 0, \quad \ln(a \cdot b) = \ln(a) + \ln(b), \quad \ln(a^b) = b \cdot \ln(a), \quad \ln \left( \frac{1}{a} \right) = -\ln(a). \]

The following two express that \( e^x \) and \( \ln(x) \) are inverse functions of one another:

\[ e^{\ln(x)} = x, \quad \ln(e^x) = x. \]

These are particularly useful in dealing with exponentials to different bases:

\[ T^x = \left( e^{\ln(T)} \right)^x = e^{\ln(T) \cdot x}, \]

as well as with those pesky tower functions (that have an \( x \) in both the base and the exponent):

\[ x^x = \left( e^{\ln(x)} \right)^x = e^{x \ln(x)}, \quad (\sin(x))^{\cos(x)} = \left( e^{\ln(\sin(x))} \right)^{\cos(x)} = e^{\cos(x) \ln(\sin(x))}. \]
Chapter 12
Definite Integrals

12.1 Definite integrals

Temporary definition of definite integral
We define the new concept of a definite integral:

\[ \int_{a}^{b} f(t) \, dt \] is the net area under the graph of \( f(x) \) between \( a \) and \( b \).

Several specifications should be made:

- 'under' means between the graph and the \( x \)-axis.
- 'net' area means that we count area above the \( x \)-axis positively, and we count area below the \( x \)-axis negatively.
- Normally \( a < b \) and we look at the area from left to right. If \( b < a \), then the area should be counted negatively, as if it’s going from right to left.
- Although we’re using the same symbol \( \int \) as in the last chapter, right now these symbols have nothing to do with each other: the indefinite integral \( \int f(t) \, dt \) stands for the general antiderivative of \( f \), while \( \int_{a}^{b} f(t) \, dt \) is defined as an area.
- This definition is 'temporary' because we’re not defining what area is; it’s an intuitive notion, but in mathematics that’s not quite good enough. Later in this chapter we’ll see how to define area and definite integrals precisely, using Riemann sums.
- The function \( f(t) \) inside is referred to as the integrand. The numbers \( a \) and \( b \) are the limits of integration; \( a \) is the lower limit and \( b \) the upper limit.

Properties of definite integrals
We can deduce several properties of definite integrals that we’ll need later.
To begin with:

\[ \int_{a}^{a} f(t) \, dt = 0, \quad \int_{b}^{a} f(t) \, dt = - \int_{a}^{b} f(t) \, dt. \]
The first follows from the word 'between' – there is no area between $a$ and $a$. The second captures the remark above that if you go from right to left instead of from left to right, the area counts negatively.

The next two properties describe what happens to a definite integral if you modify the integrand:

\[
\int_a^b c \cdot f(t)\,dt = c \cdot \int_a^b f(t)\,dt,
\]
\[
\int_a^b (f(t) + g(t)) \,dt = \int_a^b f(t)\,dt + \int_a^b g(t)\,dt.
\]

The first says that the area under a scaled function equals the area under that function, scaled by the same factor. The second says that to calculate the area under the sum of two functions you can simply take the areas under the two functions separately, and add the two results together. But note that this only makes sense if the two integrals have the same limits of integration.

Both of these make perfect sense if you think of how areas work. Note that these happen to exactly match the properties of the indefinite integral $\int f(t)\,dt$.

Finally, we can combine two integrals that have the same integrand but different limits of integration, but only in the special case where the upper limit of one integral is the lower limit of the other:

\[
\int_a^b f(t)\,dt + \int_b^c f(t)\,dt = \int_a^c f(t)\,dt.
\]

Again this makes perfect sense if you look at the picture of such areas. But note that it only works if the integrands really are the same.

- **Examples using geometry**

Soon we’ll see a strong technique for evaluating definite integrals, but right now we can only do a few using area formulas that we know from geometry. For instance, when the integrand is a linear function, an area under the graph (a line) is a combination of rectangles and triangles. Another type that we can do is where the area is a half or quarter circle.

Here are some examples of these.

- Let’s evaluate $\int_0^3 2x\,dx$. If we draw the picture, we see that the area between 0 and 3 under the graph is a triangle, with base 3 and height $2 \cdot 3 = 6$. Hence its area is $\frac{1}{2} \cdot 3 \cdot 6 = 9$, so we have $\int_0^3 2x\,dx = 9$.

- Now we do $\int_{-2}^2 (5 - 2x)\,dx$. Again we draw the picture, and we see a triangle on top of a rectangle. The rectangle is 4 by 1, so has area 4. The triangle is 4 by 8, so has area 16. Hence we have $\int_{-2}^2 (2x + 5)\,dx = 20$.

- A different one is $\int_0^4 (1 - x)\,dx$. Now the area consists of 2 triangles: one is 1 by 1 and above the $x$-axis, the other is 3 by 3, but below the $x$-axis. As defined above, the second triangle should be counted negatively, giving us $\int_0^4 (1 - x)\,dx = \frac{1}{2} \cdot 1 \cdot 1 - \frac{1}{2} \cdot 3 \cdot 3 = -4$.

- We evaluate $\int_{-3}^3 \sqrt{9 - x^2}\,dx$. The area is a half circle, with center at the origin, and radius 3. Hence the area of the whole circle would be $\pi \cdot 3^2$, and the area of a half circle is $\frac{\pi \cdot 3^2}{2}$. So $\int_{-3}^3 \sqrt{9 - x^2}\,dx = \frac{\pi \cdot 3^2}{2}$.

### 12.2 The Fundamental Theorem of Calculus

- **Area Functions**

By an area function of $f$ we shall mean a function of the form

\[
A(x) = \int_a^x f(t)\,dt.
\]
This type of function will be important in the Fundamental Theorem that’s coming up.

- As an example, consider \( A(x) = \int_0^x (2t + 1) dt \). Then when we draw the picture we can see that \( A(0) = 0 \), \( A(1) = 2 \) (1 by 1 rectangle and a 1 by 2 triangle), and \( A(3) = 12 \) (3 by 1 rectangle, 3 by 6 triangle).

In fact, we can find a general formula: \( A(x) \) is the area of an \( x \) by 1 rectangle plus an \( x \) by \((2x + 1) - 1 = 2x \) triangle, which gives us \( A(x) = x + \frac{1}{2} \cdot x \cdot 2x = x^2 + x \).

You may notice that this is an antiderivative of \( A(x) \). The next theorem tells us that this always happens.

- **Fundamental Theorem of Calculus (FTC)**
  
  We have defined indefinite integrals (antiderivatives) and definite integrals (like area functions) in very different ways; we even used the same symbol. Of course, this was for a reason, which is given by the Fundamental Theorem:

  \[ \text{FTC: An area function of } f \text{ is an antiderivative of } f. \]

  This is the short version; what we will most often make use of are the following two consequences (and especially of the first one), which together make up the long version:

  - \( \int_a^b f(x) dx = F(b) - F(a) \), if \( F \) is an antiderivative of \( f \), and \( f \) is continuous on \([a, b]\)
  - \( \frac{d}{dx} \int_a^x f(t) dt = f(x) \)

  There are different approaches as to what to call FTC: the book calls the two consequences the FTC, I prefer the short version above. Either way, these are just different versions of the same idea, and it’s okay to refer to any of it as the FTC.

  In the first one, the restriction that \( f \) must be continuous on the interval \([a, b]\) is important: if there is a discontinuity of \( f \) between \( a \) and \( b \), then the formula might be wrong. We’ll see more about this later.

- **Why?**
  
  You can find a detailed proof of the FTC in the book; here I will give an outline of the idea, but first I’ll say why the consequences are consequences.

  - The second is just a restatement (if \( A(x) \) is an antiderivative of \( f(x) \), then \( f(x) \) is a derivative of \( A(x) \)).

  - For the first consequence, take the area function \( A(x) = \int_a^x f(t) dt \) and let \( F(x) \) be any antiderivative of \( f(x) \). Then we know that they only differ by some constant, so \( F(x) = A(x) + C \). But then, since \( A(a) = \int_a^a f(t) dt = 0 \), we already have

    \[ F(b) - F(a) = (A(b) + C) - (A(a) + C) = A(b) - A(a) = A(b) = \int_a^b f(t) dt. \]

  - To show that the FTC is true, we show that the derivative of \( A(x) = \int_a^x f(t) dt \) is in fact \( f(x) \). We’ll have to use the definition of derivative, which tells us

    \[ A'(x) = \lim_{h \to 0} \frac{A(x + h) - A(x)}{h}. \]

    With the sum property for definite integrals above we get

    \[ A(x + h) - A(x) = \int_a^{x+h} f(t) dt - \int_a^x f(t) dt = \int_x^{x+h} f(t) dt \]
By definition of the definite integral, this is the thin area under the graph of \( f(x) \) between \( x \) and \( x+h \). Since in the limit \( h \) is going to zero, this strip gets thinner and thinner. Because it’s so thin, it’s pretty much a rectangle with width \( (x+h) - x = h \) and height \( f(x) \), so the area of the strip will approach the area of the rectangle, which is width times height, \( h \cdot f(x) \). Then

\[
A'(x) = \lim_{h \to 0} \frac{A(x+h) - A(x)}{h} = \lim_{h \to 0} \frac{h \cdot f(x)}{h} = \lim_{h \to 0} f(x) = f(x).
\]

• Examples using \( \int_b^a f(x)dx = F(b) - F(a) \)
This is one of the most important tools in this course: to evaluate a definite integral, find an antiderivative \( F(x) \) of the integrand, and plug in the limits. Since we do not yet have that many techniques for finding antiderivatives (substitution is in the next section), I will only do a few examples here.

One piece of notation: since the form \( F(b) - F(a) \) will occur a lot, and it might be annoying to write out the \( F \) twice if it’s complicated, we will use a vertical bar like this:

\[
F(b) - F(a) = \{F(x)\} \bigg|_a^b.
\]

The examples don’t require much explanation, I’ll just write down the solutions:

\[
\int_1^2 x^3 dx = \left. \frac{1}{3} x^3 \right|_1^2 = \left( \frac{1}{3} \cdot 2^3 \right) - \left( \frac{1}{3} \cdot 1^3 \right) = \frac{8}{3} - \frac{1}{3} = \frac{7}{3},
\]

\[
\int_\pi^{\pi/2} \sin(x) dx = \left. -\cos(x) \right|_\pi^{\pi/2} = \left( -\cos(\pi) \right) - \left( -\cos(\pi/2) \right) = \left( -1 \right) - \left( 0 \right) = 1,
\]

\[
\int_{-1}^2 (x^2 - 3x + 2) dx = \left. \frac{1}{3} x^3 - \frac{3}{2} x^2 + 2x \right|_{-1}^2 = \left( \frac{1}{3} \cdot 8 - \frac{3}{2} \cdot 4 + 2 \cdot 2 \right) - \left( \frac{1}{3} \cdot (-1)^3 - \frac{3}{2} \cdot (-1)^2 + 2(-1) \right)
\]

\[
= \frac{8}{3} - 6 + 4 + \frac{1}{3} + \frac{3}{2} + 2 = \frac{9}{2}.
\]

• Examples using \( \frac{d}{dx} \int_0^x f(t) dt = f(x) \)
This formula is usually pretty straightforward to use. Here are two examples:

\[
\frac{d}{dx} \int_2^x t^5 dt = x^5,
\]

\[
\frac{d}{dx} \int_x^3 e^{-t^2} dt = \frac{d}{dx} \left( -\int_3^x e^{-t^2} dt \right) = -\frac{d}{dx} \int_3^x e^{-t^2} dt = -e^{-x^2}.
\]

The second one used the property \( \int_a^b = -\int_b^a \) of definite integrals that we saw before.

There is one tricky variation to this type of question, for instance

\[
\frac{d}{dx} \int_0^{x^3} \sin(t) dt.
\]

Here instead of an \( x \) in the limit of integration, we have a function of \( x \). So this is a composition \( F(G(x)) \) with the outer function \( F(x) = \int_0^x \sin(t) dt \) and the inner function \( G(x) = x^3 \). To
differentiate a composition like that, we use the chain rule, \( \frac{d}{dx} F(G(x)) = F'(G(x)) \cdot G'(x) \). We know that \( G'(x) = 3x^2 \), and we can also do

\[
F'(x) = \frac{d}{dx} \int_0^x \sin(t) dt = \sin(x).
\]

Hence we have

\[
\frac{d}{dx} \int_0^x \sin(t) dt = F'(G(x)) \cdot G'(x) = \sin(x^3) \cdot 3x^2.
\]

### 12.3 Substitution

We’ve seen how to use substitution for indefinite integrals, but for definite integrals there is a twist to how the limits are plugged in. There are two ways to do this:

(A) Work out the integral with substitution, return to \( x \), and then plug in the limits.

(B) Adjust the limits to the substitution, and plug them in without having to return to \( x \).

I’ll explain this in a few examples.

- I’ll do the integral \( \int_2^3 2x \sin(x^2) dx \) both ways, with the substitution \( u = x^2 \), \( du = 2xdx \):

\[
(A) : \int_2^3 2x \sin(x^2) dx = \int_2^3 \sin(u) du = -\cos(u) \bigg|_{x^2=2}^{x^2=3} = -\cos(9) + \cos(4).
\]

The important thing here is to make clear that the limits 2 and 3 should not be plugged in for \( u \), but for \( x \). So to write \( \int_2^3 \sin(u) du \) would have been very wrong, because it would mean that you’re going to plug in \( u = 2 \) and \( u = 3 \).

For method (B), we write the limits in terms of \( u = x^2 \):

\[
x = 2 \Rightarrow u = 2^2 = 4, \quad x = 3 \Rightarrow u = 3^2 = 9,
\]

and use that to adjust the limits:

\[
(B) : \int_2^3 2x \sin(x^2) dx = \int_4^9 \sin(u) du = -\cos(u) \bigg|_{4}^{9} = -\cos(9) + \cos(4).
\]

In general method (B) is slightly more efficient, but it’s important to understand both ways (and even more important not to confuse them).

- **Question:** Find the net area under the graph of \( \frac{\ln(x)}{x} \) on the interval \([e, e^2]\).

The net area is given by the definite integral

\[
\int_e^{e^2} \frac{\ln(x)}{x} dx.
\]

I’ll use method (B). The substitution to use is

\[
u = \ln(x), \quad du = \frac{1}{x} dx,
\]

so the limits become

\[
x = e \Rightarrow u = \ln(e) = 1, \quad x = e^2 \Rightarrow u = \ln(e^2) = 2.
\]

Then we get

\[
\int_e^{e^2} \frac{\ln(x)}{x} dx = \int_1^2 u du = \frac{1}{2} u^2 \bigg|_1^2 = \frac{1}{2} (2^2 - 1^2) = \frac{3}{4}.
\]
12.4 Riemann sums

• Real definition of the definite integral
At the beginning of this chapter I gave a 'temporary' definition of the definite integral, which depended on an intuitive sense of what area is. Now I will introduce the real definition, and then I will explain what the components of the formula mean.

\[ \int_{a}^{b} f(x) \, dx = \lim_{n \to \infty} \sum_{k=1}^{n} f(x_k) \Delta x, \]

where \( x_k = a + (k-1) \cdot \frac{b-a}{n}, \quad \Delta x = \frac{b-a}{n}. \)

Note: The book uses slightly different notation in this definition. It writes \( \lim_{\Delta \to 0} \) instead of \( \lim_{n \to \infty} \), \( x_k \) instead of \( x_k \), and \( \Delta x_k \) instead of \( \Delta x \). You can use either notation.

• Approximating areas with rectangles

Note: you need to look at the relevant picture here – either the ones I showed you in class, or for instance Figure 5.9 in the book.

The definition above basically gives a formula for the area of a shape where one side can be 'curvy'. The idea behind it is to fill up such an area with smaller shapes whose area we know (rectangles), and then adding the areas of those rectangles to get an approximation for the larger area. By using more and more smaller and smaller rectangles, the approximation should get better and better, at least if we do it right.

This can be done in general, but here we use a simple and systematic approach: we use rectangles that all have the same width, and we match their heights to the graph. Let’s work out formulas for the resulting approximation, for a function \( f(x) \) on the interval \([a, b] \).

If we use \( n \) rectangles of equal width, we should divide the interval up into \( n \) equal parts, so the width of a rectangle must be

\[ \Delta x = \frac{b-a}{n}. \]

Then the \( k \)-th rectangle will have its left side above

\[ x_k = a + (k-1) \cdot \Delta x = a + (k-1) \cdot \frac{b-a}{n}. \]

To see that there should be a \( k-1 \) there, note that the first rectangle has its left side above \( a = a + 0 \cdot \Delta x \), the second above \( a + 1 \cdot \Delta x \), etc.

Then the height of each rectangle should match the height of the graph (at the left side of the rectangle), so the height of the \( k \)-th rectangle should be \( f(x_k) \).

Finally we can write down the sum of the areas (width \( \times \) height) of the rectangles, which is the approximation for the area under the graph:

\[ (f(x_1) \cdot \Delta x) + (f(x_2) \cdot \Delta x) + (f(x_3) \cdot \Delta x) + \cdots + (f(x_n) \cdot \Delta x). \]

Such a sum is called a Riemann sum.

• Sum notation

For convenience, we will write out Riemann sums using the summation symbol \( \sum \), which is defined by

\[ \sum_{k=1}^{n} f(k) = f(1) + f(2) + f(3) + \cdots + f(n). \]
For example:

\[ \sum_{k=1}^{5} k^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2, \]

\[ \sum_{k=1}^{3} \frac{1}{k+1} = \frac{1}{1+1} + \frac{1}{2+1} + \frac{1}{3+1}. \]

- **Refining the rectangles**

So the Riemann sum for \( f(x) \) on \([a, b]\) with \( n \) rectangles is given by

\[ \sum_{k=1}^{n} f(x_k) \Delta x. \]

The only step left is to make the rectangles smaller and smaller, which you’ll have to believe makes the area approximation better and better. This is done by letting \( n \) get larger and larger, which makes \( \Delta x \) smaller and smaller, hence also the rectangles. This process is exactly described by the limit \( \lim_{n \to \infty} \), so we have our whole definition:

\[ \int_{a}^{b} f(x) \, dx = \lim_{n \to \infty} \sum_{k=1}^{n} f(x_k) \Delta x; \]

where \( x_k = a + (k - 1) \cdot \frac{b-a}{n}, \quad \Delta x = \frac{b-a}{n}. \)

- **Different kinds of Riemann sums**

Above we chose to let the rectangle touch the graph with its top left corner. This is just a choice, we could also have chosen the top right corner, or even another point on the top edge, like the midpoint. These choices give different Riemann sums, and we name them as follows.

The sum \( \sum_{k=1}^{n} f(x_k) \Delta x \) is called a

- **left** Riemann sum if \( x_k = a + (k - 1) \cdot \frac{b-a}{n}; \)
- **right** Riemann sum if \( x_k = a + k \cdot \frac{b-a}{n}; \)
- **midpoint** Riemann sum if \( x_k = a + (k - \frac{1}{2}) \cdot \frac{b-a}{n}. \)

I used left Riemann sums to define the definite integral, but I could just as well have used any of the others, since the resulting limit has the same value. This is because as the rectangles get smaller and smaller, it matters less and less which point of the top edge is on the graph.

- **Questions you might be asked about Riemann sums**

Here are the two main types of questions that you should be able to answer about Riemann sums.

\( \circ \) **Write the right Riemann sum with \( n = 3 \) for \( \int_{2}^{4} \sin(x) \, dx. \)**

We first compute the width and the \( x_k \):

\[ \Delta x = \frac{4 - 2}{3} = \frac{2}{3}, \quad x_k = 2 + k \cdot \frac{2}{3}. \]

Then we write out the formula:

\[ \sum_{k=1}^{n} f(x_k) \Delta x = \sum_{k=1}^{3} \sin \left( 2 + k \cdot \frac{2}{3} \right) \cdot \frac{2}{3} = \sin \left( 2 + \frac{2}{3} \right) \cdot \frac{2}{3} + \sin \left( 2 + \frac{4}{3} \right) \cdot \frac{2}{3} + \sin (4) \cdot \frac{2}{3}. \]
Identify the definite integral that this is a Riemann sum for:

$$\lim_{n \to \infty} \sum_{k=1}^{n} \sqrt{1 + \frac{3k-3}{n} \cdot \frac{3}{n}}.$$ 

This type of question is a bit like puzzling, you have to recognize the parts of the Riemann sum in it, and extract the $f$, the $a$ and the $b$.

In this case, it’s easy to guess that $\frac{3}{n}$ is the $\Delta x$, so $\frac{b-a}{n} = \frac{3}{n}$ and $b - a = 3$.

Then $\sqrt{1 + \frac{3k-3}{n}}$ should be $f(x_k)$, so a good guess seems to be $x_k = 1 + \frac{3k-3}{n} = 1 + (k - 1) \cdot \frac{3}{n}$, which makes sense because it has $\Delta x = \frac{3}{n}$ in it. But now we can read off that $a = 1$, which together with $b - a = 3$ gives $b = 4$. Finally, since we know what $x_k$ is, we also see that $f(x) = \sqrt{x}$.

Together this gives us the definite integral that the sum above is a left (because of the $(k - 1)$) Riemann sum for:

$$\int_{1}^{4} \sqrt{x} \, dx.$$ 

Actually, we could also have chosen $x_k = \frac{3k-3}{n}$ and $f(x) = \sqrt{1 + x}$. Then $a = 0$, so $b = 3$, and the integral we get as answer is

$$\int_{0}^{3} \sqrt{1 + x} \, dx.$$ 

This is also correct, and equivalent to the first integral, because we can obtain it by substituting $u = x + 1$ and adjusting the limits of integration accordingly.

As you see, this type of question can have more than one correct answer.
Chapter 13

Integration by parts and other topics

13.1 Integration by parts

Turning the product rule around

Earlier we saw that there is no chain rule for integration, but that we can still turn the chain rule for differentiation around, which gave us the substitution method. Similarly, there is no product rule for integration, but we can still derive an integration rule from the product rule for differentiation, as follows:

\[(uv)’ = u’v + uv’ \implies uv’ = (uv)’ - u’v\]

\[\implies \int uv’ \, dx = \int (uv)’ \, dx - \int u’v \, dx\]

\[\implies \int u \, dv = uv - \int v \, du.\]

To get this last formula we used that \(\frac{dv}{dx} = v’\), so \(v’ \, dx = dv\), and similarly \(u’ \, dx = du\); and we used that the integral of the derivative of \(uv\) is just \(uv\), so \(\int (uv)’ \, dx = uv + C\) (and it’s okay to leave out the \(C\) here, because there will be a \(C\) from the second integral anyway).

This is the indefinite integral version; let’s right away give the definite integral version:

\[\int_{a}^{b} u(x)v'(x) \, dx = u(x)v(x) \bigg|_{a}^{b} - \int_{a}^{b} v(x)u'(x) \, dx.\]

Note that here we can’t use the \(du\) and \(dv\) notation, because it has to be clear that we’re plugging in \(x = a\) and \(x = b\).

A common mistake in using integration by parts for definite integrals is to forget the \(\bigg|_{a}^{b}\), i.e. to just leave the \(u(x)v(x)\) in the answer. That’s clearly wrong, because the result of a definite integral cannot have an \(x\) in it.

I will give several examples of applying this method, and then I will try to give a heuristic for choosing the \(u\) and \(v\), which is probably the hardest thing about this method.

• Examples
  o Let’s do the integral
\[ \int x \cos(x) dx. \]

The choice for for \( u \) and \( dv \) to make here is:

\[
\begin{array}{c|c}
   u = x & dv = \cos(x) dx \\
   du = dx & v = \sin(x)
\end{array}
\]

So we have \( uv = x \sin(x) \) and \( vdu = \sin(x) dx \), and we get

\[
\Rightarrow \int x \cos(x) dx = x \sin(x) - \int \sin(x) dx = x \sin(x) + \cos(x) + C.
\]

With integration by parts, it’s usually worth it to check your answer, so let’s differentiate it (with the product rule):

\[
(x \sin(x) + \cos(x))' = (x \sin(x))' + (\cos(x))' = (\sin(x) + x \cos(x)) - \sin(x) = x \cos(x).
\]

Here’s what the result will look like for a definite integral:

\[
\int_0^\pi x \cos(x) dx = (x \sin(x) + \cos(x)) \bigg|_0^\pi = (\pi \sin(\pi) + \cos(\pi)) - (0 + \cos(0)) = (-1) - (1) = -2.
\]

We see that doing integration by parts for definite integrals is not very different, since we can simply do it for the indefinite integral first, and the plug in the limits.

Let’s do

\[ \int \frac{\ln(x)}{x^2} dx. \]

It’s harder to choose \( u \) and \( dv \) here, so let me write out several choices:

\[
\begin{array}{c|c}
   u = \frac{1}{x^2} & dv = \ln(x) dx \\
   du = -\frac{1}{x^3} dx & v = ?
\end{array}
\]

\[
\begin{array}{c|c}
   u = \frac{\ln(x)}{x^2} & dv = \frac{1}{x} dx \\
   du = \frac{-1 - \ln(x)}{x^3} dx & v = \ln |x|
\end{array}
\]

\[
\begin{array}{c|c}
   u = \ln(x) & dv = \frac{1}{x^2} dx \\
   du = \frac{-1}{x} dx & v = -\frac{1}{x}
\end{array}
\]

We see right away that the first one is useless, since we can’t integrate \( \ln(x) \) (yet).

In the second one, \( dv \) does integrate nicely, and we can differentiate \( u \), but then \( du \) is pretty complicated. That’s not good, since we’d be getting an integrand \( vdu \) that’s uglier than the integrand we started with.

The third one looks better: we get \( uv = -\frac{\ln(x)}{x} \) and \( vdu = \frac{1}{x} \cdot -\frac{1}{x^2} dx = -\frac{1}{x^3} dx \), and we get

\[
\int \frac{\ln(x)}{x^2} dx = -\frac{\ln(x)}{x} - \int -\frac{1}{x^2} dx = -\frac{\ln(x)}{x} - \frac{1}{x} + C = -\frac{\ln(x) + 1}{x} + C.
\]

Again, let’s check by differentiating (with the quotient rule):

\[
\left(-\frac{\ln(x) + 1}{x}\right)' = -\frac{(\ln(x) + 1)' \cdot x - (\ln(x) + 1) \cdot (x)'}{x^2} = -\frac{\frac{1}{x} \cdot x - (\ln(x) + 1)}{x^2} = \frac{\ln(x)}{x^2}.
\]

• **Choosing \( u \) and \( dv \)**

There is no perfect strategy for choosing \( u \) and \( dv \), but two guidelines are:

- \( u \) should become simpler when you differentiate it;
- \( dv \) should be easy to integrate.
This was clear for $\frac{\ln(x)}{x^2}$: we couldn’t integrate $\ln(x)$, so $dv = \ln(x)dx$ is not a good choice; and $\frac{\ln(x)}{x}$ actually becomes more complicated when you differentiate, so $u = \frac{\ln(x)}{x}$ is also no good. But be aware that even when these conditions hold, it sometimes still doesn’t work...

**Examples**

○ Let’s apply this to $\int x \sec^2(x) dx$.

If we choose $u = x$, then $du = dx$ is very simple, and we can integrate $dv = \sec^2(x) dx$, since $(\tan(x))' = \sec^2(x)$ so $v = \tan(x)$.

No other choices would work: we can’t integrate $\sec(x)$ (yet), and $\sec^2(x)$ does not have a simple derivative. In general, if the integrand is $x$ times something more complicated, you should always try $u = x$, since $du = dx$ is one of the best simplifications that you can get: the $x$ just goes away!

To continue with the integral, with $uv = x \tan(x)$ and $vdu = \tan(x) dx$ we have

$$\int x \sec^2(x) dx = x \tan(x) - \int \tan(x) dx.$$

As often happens, the integral we’re left with is simpler, but we’re not done yet. Luckily we’ve seen this one before (in the substitution section), we have to use the definition of $\tan(x)$ first, and then substitute $u = \cos(x)$, $du = -\sin(x) dx$:

$$\int \tan(x) dx = \int \frac{\sin(x)}{\cos(x)} dx = -\int \frac{1}{u} du = -\ln|u| + C = -\ln|\cos(x)| + C.$$

Then the integral is

$$\int x \sec^2(x) dx = x \tan(x) + \ln|\cos(x)| + C.$$

○ Now consider $\int \ln(x) dx$.

The integrand doesn’t look like a product, but we could see it as the product of $\ln(x)$ and 1. We still don’t know how to integrate $\ln(x)$, but its derivative is nice and simple; and we do know how to integrate 1.

So let’s choose

| $u = \ln(x)$ | $dv = 1 \cdot dx$ |
| $du = \frac{1}{x} dx$ | $v = x$ |

Then $uv = x \ln(x)$ and $vdu = x \cdot \frac{1}{x} dx = dx$, so

$$\int \ln(x) dx = x \ln(x) - \int dx = x \ln(x) - x + C.$$

Let’s check that (it’s an important one – you might want to memorize it):

$$(x \ln(x) - x)' = (x \ln(x))' - (x)' = \left(1 \cdot \ln(x) + x \cdot \frac{1}{x}\right) - 1 = \ln(x).$$

**Repeated integration by parts**

Now consider $\int x^2 e^x dx$. 

87
Since the derivative of $x^2$ is a bit simpler and $e^x$ is easy to integrate, let’s try

<table>
<thead>
<tr>
<th>$u = x^2$</th>
<th>$dv = e^x , dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$du = 2x , dx$</td>
<td>$v = e^x$</td>
</tr>
</tbody>
</table>

Then $uv = x^2e^x$ and $vdu = 2xe^x \, dx$, so

$$\int x^2e^x \, dx = x^2e^x - 2 \int xe^x \, dx.$$ 

This is not an overwhelming success, since $\int xe^x \, dx$ is not very easy, but it is an improvement. In fact, if we now apply integration by parts to $\int xe^x \, dx$, with the easy choice

<table>
<thead>
<tr>
<th>$u = x$</th>
<th>$dv = e^x , dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$du = dx$</td>
<td>$v = e^x$</td>
</tr>
</tbody>
</table>

then $uv = xe^x$ and $vdu = e^x \, dx$, and we can work out

$$\int xe^x \, dx = xe^x - \int e^x \, dx = xe^x - e^x + C.$$ 

Putting these together we get

$$\int x^2e^x \, dx = x^2e^x - 2(xe^x - e^x) + C = (x^2 - 2x - 2)e^x + C.$$

**• Repeat and solve for $I$**

A different type is

$$\int e^{2x} \cos(x) \, dx.$$ 

A reasonable choice is

<table>
<thead>
<tr>
<th>$u = e^{2x}$</th>
<th>$dv = \cos(x) , dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$du = 2e^{2x} , dx$</td>
<td>$v = \sin(x)$</td>
</tr>
</tbody>
</table>

Then $uv = e^{2x} \sin(x)$ and $vdu = 2e^{2x} \sin(x) \, dx$, so

$$\int e^{2x} \cos(x) \, dx = e^{2x} \sin(x) - 2 \int e^{2x} \sin(x) \, dx.$$ 

This doesn’t look like an improvement – it’s just like the one we started with! However, if we try the new integral anyway, with the choice

<table>
<thead>
<tr>
<th>$u = e^{2x}$</th>
<th>$dv = \sin(x) , dx$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$du = 2e^{2x} , dx$</td>
<td>$v = -\cos(x)$</td>
</tr>
</tbody>
</table>

then we get

$$\int e^{2x} \sin(x) \, dx = -e^{2x} \cos(x) + 2 \int e^{2x} \cos(x) \, dx.$$ 

We end up with the integral we started with! That may sound bad, but actually it’s a good thing, because if we put them together, we get

$$\int e^{2x} \cos(x) \, dx = e^{2x} \sin(x) - 2 \left(-e^{2x} \cos(x) + 2 \int e^{2x} \cos(x) \, dx\right)$$

$$= e^{2x} \sin(x) + 2e^{2x} \cos(x) - 4 \int e^{2x} \cos(x) \, dx.$$ 

88
So we have our original on both sides of the equation, which means we can solve for it. It’s usually easiest to call the integral $I$, and solve for $I$:

$$I = \int e^{2x} \cos(x) \, dx = e^{2x} \sin(x) + 2e^{2x} \cos(x) - 4I$$

$$\Rightarrow 5I = e^{2x} \sin(x) + 2e^{2x} \cos(x) + C \Rightarrow I = \frac{e^{2x}}{5} (\sin(x) + 2 \cos(x)) + C.$$

### 13.2 Trigonometric integrals

In this section we will see how to deal with integrals of the forms

$$\int \sin^m(x) \cos^n(x) \, dx, \quad \int \tan^m(x) \sec^n(x) \, dx.$$  

**Strategy for $\int \sin^m(x) \cos^n(x) \, dx$**

Here are two special cases of this kind of integral where we can use a simple substitution:

$$\int \sin^m(x) \cos(x) \, dx \quad \frac{u = \sin(x)}{du = \cos(x)} \quad \int u^m \cos(x) \, du = \frac{1}{m+1} u^{m+1} + C = \frac{1}{m+1} \sin^{m+1}(x) + C,$$

$$\int \cos^n(x) \sin(x) \, dx \quad \frac{u = \cos(x)}{du = -\sin(x)} \quad \int u^n \sin(x) \, du = -\frac{1}{n+1} u^{n+1} + C = -\frac{1}{n+1} \cos^{n+1}(x) + C.$$

So when either $\cos(x)$ or $\sin(x)$ occurs to the power 1, the integral is easy. In the general case, our strategy will be to reduce it to (a combination of) these simpler integrals.

We will do this reducing by using one of the following familiar trigonometric formulas:

$$\sin^2(x) = 1 - \cos^2(x), \quad \sin^2(x) = \frac{1}{2} - \frac{1}{2} \cos(2x),$$

$$\cos^2(x) = 1 - \sin^2(x), \quad \cos^2(x) = \frac{1}{2} + \frac{1}{2} \cos(2x).$$

Which one to use will depend on whether one of $m$ and $n$ is odd, or if both are even:

- **If $m$ is odd**, we split off one $\sin(x)$, and in the remaining even power of $\sin(x)$ we substitute $\sin^2(x)$ to get a combination of integrals of the form $\int \cos^n(x) \sin(x) \, dx$.
- **If $n$ is odd**, we can similarly split off one $\cos(x)$ and substitute $\cos^2(x)$ to get integrals of the form $\int \sin^m(x) \cos(x) \, dx$.
- **If both $m$ and $n$ are even**, then we substitute $\sin^2(x) = \frac{1}{2} - \frac{1}{2} \cos(2x)$ and $\cos^2(x) = \frac{1}{2} + \frac{1}{2} \cos(2x)$ whenever possible, and see what new integrals we get.

**Examples of $\int \sin^m(x) \cos^n(x) \, dx$**

- **$\int \cos^3(x) \, dx$**

Since there is an odd power of $\cos(x)$, we split off one $\cos(x)$ and substitute $\cos^2(x) = 1 - \sin^2(x)$:

$$\int \cos^3(x) \, dx = \int \cos^2(x) \cdot \cos(x) \, dx = \int (1 - \sin^2(x)) \cos(x) \, dx$$

$$\frac{u = \sin(x)}{du = \cos(x)} \int (1 - u^2) \, du = u - \frac{1}{3} u^3 + C = \sin(x) - \frac{1}{3} \sin^3(x) + C.$$
\[ \int \sin^2(x) \, dx \]

There are no odd powers, so we use \( \sin^2(x) = \frac{1}{2} - \frac{1}{2} \cos(2x) \):

\[
\int \sin^2(x) \, dx = \int \left( \frac{1}{2} - \frac{1}{2} \cos(2x) \right) \, dx = \frac{1}{2} x - \frac{1}{4} \sin(2x) + C.
\]

\[ \int \sin^4(x) \, dx \]

Again we use \( \sin^2(x) = \frac{1}{2} - \frac{1}{2} \cos(2x) \), but now we’re not done right away:

\[
\int \sin^4(x) \, dx = \int (\sin^2(x))^2 \, dx = \int \left( \frac{1}{4} - \frac{1}{2} \cos(2x) + \frac{1}{4} \cos^2(2x) \right) \, dx.
\]

To integrate the third term, we need to separately use \( \cos^2(2x) = \frac{1}{2} + \frac{1}{2} \cos(4x) \) (which is the formula \( \cos^2(x) = \frac{1}{2} + \frac{1}{2} \cos(2x) \) with \( 2x \) plugged in for \( x \)):

\[
\int \frac{1}{4} \cos^2(2x) \, dx = \int \left( \frac{1}{8} + \frac{1}{8} \cos(4x) \right) \, dx = \frac{1}{8} x + \frac{1}{32} \sin(4x) + C.
\]

With this we can finish the main integral:

\[
\int \sin^4(x) \, dx = \int \left( \frac{1}{4} - \frac{1}{2} \cos(2x) \right) \, dx + \int \frac{1}{4} \cos^2(2x) \, dx
= \left( \frac{1}{4} x - \frac{1}{4} \sin(2x) \right) + \left( \frac{1}{8} x + \frac{1}{32} \sin(4x) + C \right) = \frac{3}{8} x - \frac{1}{4} \sin(2x) + \frac{1}{32} \sin(4x) + C.
\]

\[ \int \sin^5(x) \cos^2(x) \, dx \]

Since we have an odd power of \( \sin(x) \), we split off one \( \sin(x) \) and substitute \( \sin^2(x) = 1 - \cos^2(x) \):

\[
\int \sin^5(x) \cos^2(x) \, dx = \int (\sin^2(x))^2 \cdot \cos^2(x) \cdot \sin(x) \, dx = \int (1 - \cos^2(x))^2 \cdot \cos^2(x) \cdot \sin(x) \, dx
= \int \left( \cos^2(x) - 2 \cos^4(x) + \cos^6(x) \right) \cdot \sin(x) \, dx_{u=\cos(x)} = \int \left( u^2 - 2u^4 + u^6 \right) \, du
= -\frac{1}{3} \cos^3(x) + \frac{2}{5} \cos^5(x) - \frac{1}{7} \cos^7(x) + C.
\]

- **The integrals** \( \int \tan(x) \, dx \), \( \int \sec(x) \, dx \), and \( \int \sec^3(x) \, dx \)

Before attacking integrals like \( \int \tan^n(x) \sec^n(x) \, dx \), we will first do some of the basic ones, because these are not done with the general strategy.

- We’ve seen the integral of \( \tan(x) \) before, obtained by substituting \( u = \cos(x) \):

\[ \int \tan(x) \, dx = \ln |\cos(x)| + C. \]

- \( \int \sec(x) \, dx \) is notoriously tricky, it’s very hard to come up with its solution. I’ll try to explain where it comes from, but if you find this hard to follow, this is one integral that’s worth
memorizing.
The trick comes down to observing (out of the blue) that
\[(\sec(x) + \tan(x))' = \sec(x)\tan(x) + \sec^2(x) = (\sec(x) + \tan(x)) \cdot \sec(x),\]
so the derivative of \(f(x) = \sec(x) + \tan(x)\) equals itself times \(\sec(x)\), in other words
\[f'(x) = f(x) \cdot \sec(x) \implies \sec(x) = \frac{f'(x)}{f(x)} = \frac{d}{dx} \ln |f(x)|.\]

What we have here is \(\sec(x)\) as the derivative of something else, so we know an antiderivative:
\[
\int \sec(x) \, dx = \ln |\sec(x) + \tan(x)| + C.
\]

- A related integral of this type that I should mention is \(\int \sec^3(x) \, dx\) (note that \(\int \sec^2(x) \, dx = \tan(x) + C\) is easy). It requires integration by parts, with the following choice for \(u\) and \(dv\):

<table>
<thead>
<tr>
<th>(u = \sec(x))</th>
<th>(dv = \sec^2(x) , dx)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(du = \sec(x) \tan(x) , dx)</td>
<td>(v = \tan(x))</td>
</tr>
</tbody>
</table>

so we get \(uv = \sec(x) \tan(x)\) and \(vdu = \sec(x) \tan^2(x)\), hence (using \(\tan^2(x) = \sec^2(x) - 1\), see below)
\[
I = \int \sec^3(x) \, dx = \sec(x) \tan(x) - \int \sec(x) \tan^2(x) \, dx
= \sec(x) \tan(x) - \int \sec(x) (\sec^2(x) - 1) \, dx
= \sec(x) \tan(x) - \int (\sec^3(x) - \sec(x)) \, dx
= \sec(x) \tan(x) - I + \ln |\sec(x) + \tan(x)|
\]
\[
\implies I = \frac{1}{2} \sec(x) \tan(x) + \frac{1}{2} \ln |\sec(x) + \tan(x)| + C.
\]

Here we used the "solve for \(I'\)" trick that we encountered in the section on integration by parts.

- **Strategy for** \(\int \tan^n(x) \sec^n(x) \, dx\)

For integrals of the form \(\int \tan^n(x) \sec^n(x) \, dx\) there is a similar strategy to the one we used for \(\int \sin^m(x) \cos^n(x) \, dx\), although it differs in that it doesn’t work in all cases.

Because the derivatives of \(\tan(x)\) and \(\sec(x)\) are
\[(\tan(x))' = \sec^2(x), \quad (\sec(x))' = \sec(x) \tan(x),\]
we again have certain cases that can be done by a simple substitution:
\[
\int \tan^m(x) \cdot \sec^2(x) \, dx \quad u = \tan(x) \quad \int u^m \, du = \frac{1}{m+1} \tan^{m+1}(x) + C,
\]
\[
\int \sec^n(x) \cdot (\sec(x) \tan(x)) \, dx \quad u = \sec(x) \quad \int u^n \, du = \frac{1}{n+1} \sec^{n+1}(x) + C.
\]
And again we try to do the other ones by reducing to these. For that we need the basic relation between \(\tan(x)\) and \(\sec(x)\), which we derive as follows:
\[
\sin^2(x) + \cos^2(x) = 1 \implies \frac{\sin^2(x)}{\cos^2(x)} + \frac{\cos^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} \implies \tan^2(x) + 1 = \sec^2(x).
\]

There are now two cases where this approach works, but also one where it doesn’t:
• If $n$ is even, we split off one $\sec^2(x)$, substitute $\sec^2(x) = \tan^2(x) + 1$ in the remaining power of $\sec(x)$, to get a combination of integrals of the form $\int \tan^n(x) \cdot \sec^2(x) \, dx$;
• If $m$ is odd and $n \geq 1$, we split off $\sec(x) \tan(x)$, substitute $\tan^2(x) = 1 - \sec^2(x)$ in the power of $\tan(x)$ that remains, to get a combination of integrals of the form $\int \sec^n(x) \cdot (\sec(x) \tan(x)) \, dx$;
• When neither of these conditions applies, we have to try something else.

**Examples for $\int \tan^m(x) \sec^n(x) \, dx$**

- $\int \tan^6(x) \sec^4(x) \, dx$

  We have an even power of $\sec(x)$, so we split off one $\sec^2(x)$ and substitute $\sec^2(x) = \tan^2(x) + 1$ for the other:

  $$\int \tan^6(x) \sec^4(x) \, dx = \int \tan^6(x)(\tan^2(x) + 1) \cdot \sec^2(x) \, dx = \int (\tan^8(x) + \tan^6(x)) \cdot \sec^2(x) \, dx$$

  $$u = \tan(x)$$

  $$\int (u^8 + u^6) \, du = \frac{1}{9} \tan^9(x) + \frac{1}{7} \tan^7(x) + C.$$  

- $\int \tan^5(x) \sec^7(x) \, dx$

  Now we have an odd power of $\tan(x)$, so we split off $\sec(x) \tan(x)$ and then substitute $\tan^2(x) = \sec^2(x) - 1$ in the rest:

  $$\int \tan^5(x) \sec^7(x) \, dx = \int (\sec^2(x) - 1)^2 \sec^6(x) \cdot (\sec(x) \tan(x)) \, dx$$

  $$= \int (\sec^{10}(x) - 2 \sec^8(x) + \sec^6(x)) \cdot (\sec(x) \tan(x)) \, dx$$

  $$u = \sec(x)$$

  $$\int (u^{10} - 2u^8 + u^6) \, du = \frac{1}{11} \sec^{11}(x) - \frac{2}{9} \sec^9(x) + \frac{1}{7} \sec^7(x) + C.$$  

### 13.3 Odd and even functions

**Definition**

A function $f(x)$ is called:

- **Even** if $f(-x) = f(x)$ for all $x$ in the domain of $f$;
- **Odd** if $f(-x) = -f(x)$ for all $x$ in the domain of $f$.

Some examples:

- $f(x) = x^2$ is even because $f(-x) = (-x)^2 = x^2 = f(x)$,
- $f(x) = x^3$ is odd because $f(-x) = (-x)^3 = -x^3 = -f(x)$,
- $x^6 - \sqrt{1 - x^2}$ is even because $(-x)^6 - \sqrt{1 - (-x)^2} = x^6 - \sqrt{1 - x^2}$,
- $\cos(x)$ is even because $\cos(-x) = \cos(x)$,
- $\sin(x)$ is odd because $\sin(-x) = -\sin(x)$,
- $x \sin(x)$ is even because $(-x) \sin(-x) = x \sin(x)$,
- $\tan(x)$ is odd because $\tan(-x) = \frac{\sin(-x)}{\cos(-x)} = -\frac{\sin(x)}{\cos(x)} = -\tan(x)$,

Note: most function are neither odd nor even.

**Symmetry**

These two properties have clear meanings in terms of symmetry of the graph of the function:
• If \( f(x) \) is even, then its graph \( y = f(x) \) is symmetric in the \( y \)-axis. This means that if you reflect it in the \( y \)-axis, in other words you look at the graph \( y = f(-x) \), then that’s the same graph as \( y = f(x) \).

• If \( f(x) \) is odd, then its graph \( y = f(x) \) is symmetric in the origin. This means that if you reflect it in the \( y \)-axis and then in the \( x \)-axis, and get the graph \( y = -f(-x) \), then that’s the same graph as \( y = f(x) \).

• Integrals of odd and even Functions
If you know that a function is odd or even, then that can help you do some integrals of it, namely the ones that have symmetric limits like \( \int_{-a}^{a} \). For even functions, the left part of the graph will have the same area as the right part, so we only have to compute one and multiply it by two. For odd functions, the left part of the graph will have the same area as the right part, but counted negatively, so the areas of the left and right part will cancel each other out, and the integral equals 0.

• If \( f(x) \) is odd and continuous on \([-a, a]\), then \( \int_{-a}^{a} f(x) \, dx = 0 \).

• If \( f(x) \) is even and continuous on \([-a, a]\), then \( \int_{-a}^{a} f(x) \, dx = 2 \int_{0}^{a} f(x) \, dx \).
The second one will merely make the calculation a bit more efficient (plugging in 0 is usually easy), but the first one can make the entire calculation redundant.

A few examples of integrals of odd and even functions (check for yourself if the function is odd or even):
\[
\int_{-7}^{7} x^4 \, dx = 2 \int_{0}^{7} x^4 \, dx = 2 \cdot \frac{1}{5} x^5 \bigg|_{0}^{7} = \frac{2}{5} \cdot (7^5 - 0) = \frac{2}{5} \cdot 7^5,
\]
\[
\int_{-1}^{1} \tan(x) \, dx = 0,
\int_{-2}^{2} x^5 \cos(x) \, dx = 0.
\]

13.4 Averages and the Mean Value Theorem

• Averages
You already know what a discrete average is: for instance, if \( f(k) \) gives your score on the \( k \)-th homework, and there are \( n \) homeworks in total, then your average score is
\[
\overline{f} = \frac{1}{n} \sum_{k=1}^{n} f(k).
\]
This is called a discrete average because this is a discrete function, which means it only takes values on isolated numbers.
In calculus however, we deal with functions that take values on all numbers in an interval \([a, b]\), so we couldn’t take a sum like above. What we can do instead is take the integral over the interval, and we divide by the length of the interval (similarly to dividing by \( n \) above). This gives the average value of a function:
\[
\overline{f} = \frac{1}{b - a} \int_{a}^{b} f(x) \, dx.
\]

• Example
◦ Suppose that the elevation (in km) of a hiking trail is given by
\[
h(x) = \frac{1}{10}(2 + 6x - x^2),
\]
where $x$ is the distance along the trail in km, running from $x = 0$ to $x = 6$. (This is actually pretty much going up and down the Grouse Grind, if you know what that is...) Then the average elevation of the trail is

$$
\bar{h} = \frac{1}{6-0} \int_0^6 \frac{1}{10}(2 + 6x - x^2)dx = \frac{1}{60}(2x + 3x^2 - \frac{1}{3}x^3)\bigg|_0^6 = \frac{1}{60}(12 + 108 - 72) = \frac{4}{5}.
$$

So the average height of this trail is 800m. Why would this be useful? It’s in a way the best summary that you can give of the trail in a single number. And that’s useful because if you have two completely different trails, you couldn’t really compare their elevation functions, but you can compare their average elevations. This is of course also why we compute the average of your grades at the end of a course!

• **The Mean Value Theorem for integrals**

The following fact about averages is called the *Mean Value Theorem* for integrals:

\[\text{If } f \text{ is a continuous function on } [a,b], \text{ then there is a point } c \text{ in } [a,b] \text{ for which } f(c) = \bar{f}.\]

In other words, on an interval, there is always a point where the value of the function equals its average value over that interval.

• **Example**

○ For the hiking trail example above, this theorem says that there should be a $c$ in $[0,6]$ such that $h(c) = \bar{h} = \frac{4}{5}$. Because $h(x)$ is simple enough, we can find this $c$ algebraically:

$$
\frac{4}{5} = h(c) = \frac{1}{10}(2 + 6c - c^2) \Rightarrow 40 = 10 + 30c - 5c^2 \Rightarrow 5c^2 - 30c + 30 = 0 \Rightarrow c^2 - 6c + 6 = 0
$$

$$
\Rightarrow c = \frac{1}{2} \left(6 \pm \sqrt{36 - 24}\right) = 3 \pm \sqrt{3},
$$

using the quadratic formula. So in fact there are two $c$, $c = 3 - \sqrt{3} \approx 1.3$ and $c = 3 + \sqrt{3} \approx 4.7$, as we would expect, since we go up and down the same trail.
Chapter 14

Trigonometric Substitution, Partial Fractions, and Numerical Integration

14.1 Trigonometric substitution

• Introduction

We do not yet know how to handle integrals involving variations of the functions

\[ \sqrt{1 - x^2}, \quad \sqrt{1 + x^2}, \quad \sqrt{x^2 - 1}. \]

For instance, we cannot yet do integrals like

\[ \int \sqrt{1 - x^2} \, dx, \quad \int \frac{1}{\sqrt{4 + x^2}}, \quad \int \frac{x^2}{(2x^2 - 3)^{3/2}}. \]

The basic observation that we need for solving these is that

\[ \sqrt{1 - \sin^2(\theta)} \overset{\ast}{=} \cos(\theta), \quad \sqrt{1 + \tan^2(\theta)} \overset{\ast}{=} \sec(\theta), \quad \sqrt{\sec^2(\theta) - 1} \overset{\ast}{=} \tan(\theta). \]

I’ve put a * over these equalities because they’re not quite true: the right hand sides should be \(|\cos(\theta)|, |\sec(\theta)|, \text{ and } |\tan(\theta)|, \) since \(\sqrt{x^2} = |x|, \) not \(x.\) But I’m going to ignore this for now, and go into the finer details later.

Here is why these equations help: to do for instance \(\int \sqrt{1 - x^2} \, dx,\) we can use the substitution

\[ x = \sin(\theta) \implies dx = \cos(\theta) \, d\theta \]

to get

\[ \int \sqrt{1 - x^2} \, dx = \int \sqrt{1 - \sin^2(\theta)} \cdot (\cos(\theta) \, d\theta) = \int \cos^2(\theta) \, d\theta, \]

and this is a trigonometric integral that we can do.

Note that this substitution is a bit different from the ones we’re used to: instead of something like \(u = x^2\) when we see \(x^2\) in the integrand, we’re writing \(x = \sin(\theta),\) even though we don’t see a \(\sin(\theta)\) in the integrand.

Similarly, we can use the substitutions \(x = \tan(\theta), \ dx = \sec^2(\theta) \, d\theta\) to do

\[ \int \sqrt{1 + x^2} \, dx = \int \sqrt{1 + \tan^2(\theta)} \cdot (\sec^2(\theta) \, d\theta) = \int \sec^3(\theta) \, d\theta, \]
an integral we saw in the last chapter. And we can use \( x = \sec(\theta), \, dx = \sec(\theta) \tan(\theta) d\theta \) for
\[
\int \sqrt{x^2 - 1} \, dx = \int \sqrt{\sec^2(\theta) - 1} \cdot (\sec(\theta) \tan(\theta)) d\theta = \int \tan^2(\theta) \sec(\theta) d\theta,
\]
which we can finish using \( \tan^2(\theta) = 1 - \sec^2(\theta) \).

**Different forms of** \( \sqrt{1 - x^2}, \sqrt{1 + x^2}, \sqrt{x^2 - 1} \).

There are a number of different forms of these root functions for which we can use these substitutions. I will illustrate these here without doing the integrals yet.

For instance, for \( \sqrt{9 - x^2} \) we couldn’t use \( x = \sin(\theta) \) itself, but we can modify it to \( x = 3 \sin(\theta) \):
\[
\sqrt{9 - x^2} \quad x = 3 \sin(\theta) \quad \sqrt{3^2 - 3^2 \sin^2(\theta)} = 3 \sqrt{1 - \sin^2(\theta)} = 3 \cos(\theta).
\]

Here are some examples of modified substitutions that work, with increasing difficulty:
\[
\begin{align*}
\sqrt{1 + 4x^2} & \quad x = \frac{1}{2} \tan(\theta) \quad \sqrt{1 + 2^2 \cdot (1/2)^2 \cdot \tan^2(\theta)} = \sqrt{1 + \tan^2(\theta)} = \sec(\theta), \\
\sqrt{(x + 1)^2 - 25} & \quad x = 5 \sec(\theta) - 1 \quad \sqrt{(5 \sec(\theta))^2 - 5^2} = 5 \sqrt{\sec^2(\theta) - 1} = 5 \tan(\theta), \\
\sqrt{2 - 3x^2} & \quad x = \frac{\sqrt{2}}{3} \sin(\theta) \quad \sqrt{2 - 3 \cdot (\sqrt{2}/\sqrt{3})^2 \sin^2(\theta)} = \sqrt{2 - 2 \sin^2(\theta)} = \sqrt{2} \cos(\theta).
\end{align*}
\]

**Examples of integrals**

\( \int \frac{1}{\sqrt{1 - x^2}} \, dx \)

We actually already know an antiderivative for the integrand, but it’s a good example, so let’s pretend we don’t know that antiderivative. Since we see a \( \sqrt{1 - x^2} \), we should use the substitution \( x = \sin(\theta), \, dx = \cos(\theta) d\theta \), so
\[
\int \frac{1}{\sqrt{1 - x^2}} \, dx = \int \frac{1}{\sqrt{1 - \sin^2(\theta)}} \cdot \cos(\theta) d\theta = \int \frac{\cos(\theta)}{\cos(\theta)} d\theta = \int d\theta = \theta + C,
\]
which is not the answer yet, because we have to return to \( x \). To do that, we have to invert the relation \( x = \sin(\theta) \), which gives \( \theta = \sin^{-1}(x) \). So the integral is what we thought it would be:
\[
\int \frac{1}{\sqrt{1 - x^2}} \, dx = \theta + C = \sin^{-1}(x) + C.
\]

\( \int \frac{1}{(4x^2 + 9)^{3/2}} \, dx \)

We use the substitution \( x = \frac{3}{2} \tan(\theta), \, dx = \frac{3}{2} \sec^2(\theta) \), so
\[
\int \frac{1}{(4x^2 + 9)^{3/2}} \, dx = \int \frac{1}{(9 \tan^2(\theta) + 9)^{3/2}} \left( \frac{3}{2} \sec^2(\theta) d\theta \right) = \frac{3}{2} \int \frac{\sec^2(\theta)}{9^{3/2}(\sec^2(\theta))^{3/2}} d\theta \\
= \frac{3}{2} \cdot \frac{3}{3} \int \frac{\sec^2(\theta)}{\sec^3(\theta)} d\theta = \frac{1}{18} \int \cos(\theta) d\theta = -\frac{1}{18} \sin(\theta) + C.
\]

To return to \( x \), we invert \( x = \frac{3}{2} \tan(\theta) \) to get \( \theta = \tan^{-1}(2x/3) \), which we plug in:
\[
= -\frac{1}{18} \sin(\tan^{-1}(2x/3)) + C = -\frac{1}{18} \frac{2x/3}{\sqrt{(2x/3)^2 + 1}} + C = -\frac{1}{9} \frac{x}{\sqrt{2x^2/9 + 1}} + C
\]
Finally, the ones with \( \sin \) outside we can do as follows:

\[
\begin{align*}
\sin(\tan^{-1}(x)) &= \frac{x}{\sqrt{x^2 + 1}}.
\end{align*}
\]

Here we used the identity \( \sin(\tan^{-1}(x)) = \frac{x}{\sqrt{x^2 + 1}} \), which I will explain after the next integral.

\[\int_3^6 \frac{\sqrt{x^2 - 9}}{x} \, dx\]

We do the indefinite integral first, using the substitution \( x = 3 \sec(\theta) \), \( dx = 3 \sec(\theta) \tan(\theta) \, d\theta \):

\[
\int \frac{\sqrt{x^2 - 9}}{x} \, dx = \int \frac{\sqrt{3^2 \sec^2(\theta) - 9^2}}{3 \sec(\theta)} (3 \sec(\theta) \tan(\theta) \, d\theta) = \int \frac{3 \tan(\theta)}{3 \sec(\theta)} (3 \sec(\theta) \tan(\theta) \, d\theta)
\]

\[= 3 \int \tan^2(\theta) \, d\theta = 3 \int (\sec^2(\theta) - 1) \, d\theta = 3 \tan(\theta) - 3\theta + C.\]

To return to \( x \), we invert \( x = 3 \sec(\theta) \) to get \( \theta = \sec^{-1}(x/3) \), which we plug in:

\[
= 3 \tan(\sec^{-1}(x/3)) - 3 \sec^{-1}(x/3) + C = 3\sqrt{(x/3)^2 - 1} - 3 \sec^{-1}(x/3) + C
\]

\[= \frac{x}{\sqrt{x^2 - 1}},\]

which we plug in:

\[
= \int_3^6 \frac{\sqrt{x^2 - 9}}{x} \, dx
\]

Here we used \( \tan(\sec^{-1}(x)) = \sqrt{x^2 - 1} \), which I will explain now.

\[\cdot \text{ Identities like } \sin(\tan^{-1}(x)) = \frac{x}{\sqrt{x^2 + 1}}.\]

When doing trigonometric substitution integrals, you very often end up with formulas like \( \sin(\sin^{-1}(x)) \), \( \tan(\sec^{-1}(x)) \), or \( \sin(\tan^{-1}(x)) \). You have to simplify these using identities, but there are too many different ones to memorize them, so you will have to know how to derive them. There are two ways to do this: one is by ‘reference triangle’, which the book does in 7.3; I’ll do the other method here because it doesn’t require a drawing.

The easiest ones are

\[
\sin(\sin^{-1}(x)) = x, \quad \tan(\tan^{-1}(x)) = x, \quad \sec(\sec^{-1}(x)) = x.
\]

The ones involving \( \sec \) and \( \tan \) we can reduce to the above using \( \tan = \sqrt{\sec^2 - 1} \) and \( \sec = \sqrt{\tan^2 + 1} \):

\[
\tan(\sec^{-1}(x)) = \sqrt{(\sec(\sec^{-1}(x)))^2 - 1} = \sqrt{x^2 - 1},
\]

\[
\sec(\tan^{-1}(x)) = \sqrt{(\tan(\tan^{-1}(x)))^2 + 1} = \sqrt{x^2 + 1}.
\]

The ones with \( \sin^{-1} \) inside we can do using \( \tan = \frac{\sin}{\cos}, \sec = \frac{1}{\cos}, \) and \( \cos = \sqrt{1 - \sin^2} \):

\[
\tan(\sin^{-1}(x)) = \frac{\sin(\sin^{-1}(x))}{\cos(\sin^{-1}(x))} = \frac{x}{\sqrt{1 - (\sin(\sin^{-1}(x)))^2}} = \frac{x}{\sqrt{1 - x^2}},
\]

\[
\sec(\sin^{-1}(x)) = \frac{1}{\cos(\sin^{-1}(x))} = \frac{1}{\sqrt{1 - (\sin(\sin^{-1}(x)))^2}} = \frac{1}{\sqrt{1 - x^2}}.
\]

Finally, the ones with \( \sin \) outside we can do as follows:

\[
\sin(\tan^{-1}(x)) = \frac{\sin(\tan^{-1}(x))}{\cos(\tan^{-1}(x))} \cdot \frac{\cos(\tan^{-1}(x))}{1} = \tan(\tan^{-1}(x)) \cdot \frac{1}{\sec(\tan^{-1}(x))} = \frac{x}{\sqrt{x^2 + 1}},
\]

\[
\sin(\sec^{-1}(x)) = \frac{\sin(\sec^{-1}(x))}{\cos(\sec^{-1}(x))} \cdot \frac{\cos(\sec^{-1}(x))}{1} = \tan(\sec^{-1}(x)) \cdot \frac{1}{\sec(\sec^{-1}(x))} = \frac{\sqrt{x^2 - 1}}{x}.
\]
• The area of a circle!
Let’s see an important application of this technique, namely the area of a circle. You already
know what that is, of course, but you’ve probably never seen why.
A circle with center at the origin and radius \( r \) is given by the equation \( x^2 + y^2 = r^2 \), so the
top half of the circle is given by the function \( f(x) = \sqrt{r^2 - x^2} \). Then the area \( A \) of the circle
is given by 4 times the area of a quarter circle, which is given by the definite integral
\[
A = 4 \int_{0}^{r} \sqrt{r^2 - x^2} \, dx.
\]
To compute this integral, we use the substitution \( x = r \sin(\theta) \), \( dx = r \cos(\theta) \, d\theta \). Here it’s easier
to change the limits right away, using \( \theta = \sin^{-1}(x/r) \): \( x = 0 \) gives \( \theta = \sin^{-1}(0) = 0 \), \( x = r \) gives
\( \theta = \sin^{-1}(1) = \frac{\pi}{2} \). Then
\[
A = 4 \int_{0}^{\pi/2} \sqrt{r^2 - r^2 \sin^2(\theta)} \cdot (r \cos(\theta)) \, d\theta = 4r \int_{0}^{\pi/2} \sqrt{r^2 \cos^2(\theta) \cos(\theta)} \, d\theta = 4r^2 \int_{0}^{\pi/2} \cos^2(\theta) \, d\theta
\]
\[
= 4r^2 \left[ \frac{1}{2} \sin(2\theta) \right]_{0}^{\pi/2} = 4r^2 \left( \frac{1}{2} \cdot 1 + \frac{1}{4} \sin(\pi) \right) - (0 + 0) = 4r^2 \cdot \frac{\pi}{4} = \pi r^2.
\]

• Completing the square
In the examples above, underneath the square root we always had two terms, one with an \( x^2 \)
and one with a constant. But we can also deal with cases like \( \sqrt{x^2 + 2x + 10} \), by completing
the square:
\[
\sqrt{x^2 + 2x + 2} = \sqrt{x^2 + 2x + 1 + 1} = \sqrt{(x + 1)^2 + 1},
\]
so we can use the substitution \( x = \tan(\theta) - 1 \) to simplify the integral
\[
\int \sqrt{x^2 + 2x + 2} \, dx = \int \sqrt{(x + 1)^2 + 1} \, dx = \int \sqrt{\tan^2(\theta) + 1} \sec^2(\theta) \, d\theta = \int \sec^3(\theta) \, d\theta.
\]
In general, completing the square goes like this:
\[
x^2 + ax + b = \left( x^2 + 2 \cdot \frac{a}{2} x + \frac{a^2}{4} \right) + \left( b - \frac{a^2}{4} \right) = \left( x + \frac{a}{2} \right)^2 + \left( b - \frac{a^2}{4} \right)
\]
If this looks a bit mysterious, here’s the idea: we want \( x^2 + ax \) to look like \( (x + c)^2 = x^2 + 2cx + c^2 \),
so we take \( 2c = a \), hence \( c = \frac{a}{2} \), and then we compensate for the extra \( \frac{a^2}{4} \) by subtracting it
from \( b \).
Here’s a harder example of an integral with completing the square:
\[
\int \frac{1}{(5 + 4x - x^2)^{3/2}} \, dx = \int \frac{1}{(9 - (x - 2)^2)^{3/2}} \, dx = \int \frac{1}{9^{3/2}(1 - \sin^2(x))^{3/2}} \cos(\theta) \, d\theta
\]
\[
= \frac{1}{9} \int \frac{\cos(\theta)}{\cos^3(\theta)} \, d\theta = \frac{1}{9} \int \sec^2(\theta) \, d\theta = \frac{1}{9} \tan(\theta) + C = \frac{1}{9} \tan(\sin^{-1}((x - 2)/3)) + C
\]
\[
= \frac{1}{9} \frac{x - 2}{\sqrt{1 - \left( \frac{x - 2}{3} \right)^2}} + C = \frac{1}{9} \sqrt{9 - (x - 2)^2} + C = \frac{1}{9} \sqrt{5 + 4x - x^2} + C.
\]
Here are some examples that illustrate the sec-subtlety (we have respectively the cases you use depends on what values $x$ actually dealing with and those $x$). As it turns out, this is no problem for the sin and tan substitutions, but for the sec things get a little tricky. To see how this works, we have to take a closer look at what values $\theta$ can take (its range); I’m assuming you’re familiar with the graphs of sin$^{-1}$, tan$^{-1}$, sec$^{-1}$ (if not, look them up; see p.470 of the book).

When we’re substituting $x = \sin(\theta)$, we’re actually defining the new variable $\theta = \sin^{-1}(x)$, and the range of $\sin^{-1}(x)$ is $-\pi/2 \leq \theta \leq \pi/2$. So in our manipulations, we’re are only dealing with those $\theta$. And luck has it that $-\pi/2 \leq \theta \leq \pi/2$ happens to be an interval on which $\cos(\theta)$ is positive, which implies $|\cos(\theta)| = \cos(\theta)$. So for the sin substitution, the formula $\sqrt{1 - \sin^2(\theta)} = \sqrt{\cos^2(\theta)} = \cos(\theta)$ is correct after all.

Let me summarize it like this:

$$\theta = \sin^{-1}(x) \Rightarrow -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \Rightarrow \cos(\theta) \geq 0 \Rightarrow \sqrt{1 - \sin^2(\theta)} = \cos(\theta).$$

For the tan substitution, the same thing happens (note that $-\pi/2$ and $\pi/2$ are not in the range of tan$^{-1}$, because it has horizontal asymptotes there):

$$\theta = \tan^{-1}(x) \Rightarrow -\frac{\pi}{2} < \theta < \frac{\pi}{2} \Rightarrow \sec(\theta) \geq 0 \Rightarrow \sqrt{\tan^2(\theta) + 1} = \sec(\theta).$$

However, for the sec substitution, something different happens. We’re defining $\theta = \sec^{-1}(x)$, but as you can see from the graph of that function, it has a weird range: it consists of two intervals, $0 \leq \theta < \pi/2$ and $\pi/2 < \theta \leq \pi$. Similarly, its domain consists of two pieces, $x \leq -1$ and $x \geq 1$. Because of this, there are actually two different sec substitutions, and which one you use depends on what values $x$ takes (this is why integrals for which you need a secant substitution are usually given as definite integrals). The two cases look like this:

$$x \leq -1 \Rightarrow \frac{\pi}{2} < \theta \leq \pi \Rightarrow \tan(\theta) < 0 \Rightarrow \sqrt{\sec^2(\theta) - 1} = |\tan(\theta)| = -\tan(\theta)$$

$$x \geq 1 \Rightarrow 0 \leq \theta < \frac{\pi}{2} \Rightarrow \tan(\theta) > 0 \Rightarrow \sqrt{\sec^2(\theta) - 1} = |\tan(\theta)| = \tan(\theta)$$

As if this isn’t hard enough, we usually substitute something like $x = 3\sec(\theta)$, so we’re actually dealing with $\theta = \sec^{-1}(x/3)$. That means we have to replace $x \geq 1$ by $x/3 \geq 1$ and $x \leq -1$ by $x/3 \leq -1$.

Here are some examples that illustrate the sec-subtlety (we have respectively the cases $x \geq 1$, $x \leq -1$ and $x/3 \leq -1$):

$$\int_{-1}^{-2} \sqrt{x^2 - 1} dx = \int_{x=1}^{x=-1} \sqrt{\sec^2(\theta) - 1} \cdot \sec(\theta) \tan(\theta) d\theta = \int_{x=1}^{x=-1} \tan^2(\theta) \sec(\theta) d\theta,$$

$$\int_{-3}^{-4} \sqrt{x^2 - 9} dx = \int_{x=-4}^{x=-3} \sqrt{3\sec(\theta)^2 - 3^2} \cdot 3\sec(\theta) \tan(\theta) d\theta = \int_{x=-4}^{x=-3} \tan^2(\theta) \sec(\theta) d\theta.$$
14.2 Partial fractions

- Introduction
A partial fraction decomposition of a rational function like \( \frac{1}{(x+1)(x+2)} \) is a decomposition as a sum of simpler fractions, like

\[
\frac{1}{(x+1)(x+2)} = \frac{1}{x+1} - \frac{1}{x+2}.
\]

If you read this from right to left, it is nothing but "making common denominators". But in this direction it turns out to be very useful for integration, as in

\[
\int \frac{1}{(x+1)(x+2)} \, dx = \int \left( \frac{1}{x+1} - \frac{1}{x+2} \right) \, dx
\]

\[
= \int \frac{1}{x+1} \, dx - \int \frac{1}{x+2} \, dx = \ln |x+1| - \ln |x+2| + C.
\]

We could also have done this integral with completing the square and a trigonometric substitution, but that would have been a lot more work. So by manipulating the integrand with a little algebra first, we can make some such integrals a lot easier. We will also see integrals like this that we can’t do any other way. But first we’ll need a systematic way of getting such a partial fraction decomposition.

- Partial fractions by matching coefficients
  - Here’s how we could obtain the decomposition above:

\[
\frac{1}{(x+1)(x+2)} = \frac{A}{x+1} + \frac{B}{x+2} = \frac{A(x+2) + B(x+1)}{(x+1)(x+2)};
\]

since the denominators are equal, the numerators should also be equal:

\[
\Rightarrow 1 = A(x+2) + B(x+1) = (A+B)x + (2A + B).
\]

This means that we have two polynomials, \( 1 = 0 \cdot x^1 + 1 \cdot x^0 \) and \( (A+B)x + (2A + B) = (A+B)x^1 + (2A+B)x^0 \), that are equal; this means that the coefficient of corresponding powers of \( x \) must be equal:

\[
\Rightarrow 0 = A + B, \quad 1 = 2A + B.
\]

These two equations in two unknowns are easy to solve: the first gives \( B = -A \), and plugging that into the second gives

\[
1 = 2A + (-A) = A \Rightarrow A = 1, \quad B = -A = -1.
\]

Plugging those into the equation defining \( A \) and \( B \) gives us the decomposition.

  - Let’s find the partial fraction decomposition for \( \frac{3}{(x-2)(3x+5)} \).

We write

\[
\frac{3}{(x-2)(3x+5)} = \frac{A}{x-2} + \frac{B}{3x+5} = \frac{A(3x+5) + B(x-2)}{(x-2)(3x+5)}
\]

Numerators equal: \( 3 = A(3x+5) + B(x-2) = (3A + B)x + (5A - 2B) \)

Match coefficients: \( 0 = 3A + B, \quad 3 = 5A - 2B \)

Solve equations: \( B = -3A \Rightarrow 3 = 5A - 2(-3A) = 11A \Rightarrow A = 3/11, \quad B = -3A = -9/11. \)
So
\[
\frac{3}{(x-2)(3x+5)} = \frac{3}{11} \frac{1}{x-2} - \frac{9}{11} \frac{1}{3x+5},
\]
and we can compute the integral
\[
\int \frac{3}{(x-2)(3x+5)} \, dx = \int \left( \frac{3}{11} \frac{1}{x-2} - \frac{9}{11} \frac{1}{3x+5} \right) \, dx
\]
\[
= \frac{3}{11} \int \frac{1}{x-2} \, dx - \frac{9}{11} \int \frac{1}{3x+5} \, dx = \frac{3}{11} \ln |x-2| - \frac{9}{11} \frac{1}{3} \cdot \ln |3x+5| + C
\]
\[
= \frac{3}{11} \ln |x-2| - \frac{3}{11} \ln |3x+5| + C.
\]
Notice the \( \frac{1}{3} \) in \( \int \frac{1}{3x+5} \, dx = \frac{1}{3} \ln |3x+5| \); it’s a common mistake to forget about that in a calculation like this.

Let’s find the partial fraction decomposition for \( \frac{2x+1}{(x-2)(x-1)} \).

\[
\frac{2x+1}{(x-2)(x-1)} = \frac{A}{x-2} + \frac{B}{x-1} = \frac{A(x-1)+B(x-2)}{(x-2)(x-1)}
\]

**Numerators equal:** \( 2x + 1 = A(x-1) + B(x-2) = (A + B)x + (-A - 2B) \)

**Match coefficients:** \( 2 = A + B, \quad 1 = -A - 2B \)

**Solve equations:** \( B = 2 - A \Rightarrow 1 = -A - 2(2 - A) = A - 4 \Rightarrow A = 5, \quad B = 2 - 5 = -3. \)

So
\[
\frac{2x+1}{(x-2)(x-1)} = \frac{5}{x-2} - \frac{3}{x-1},
\]
and the integral is
\[
\int \frac{2x+1}{(x-2)(x-1)} \, dx = \int \frac{5}{x-2} \, dx - \int \frac{3}{x-1} \, dx = 5 \ln |x-2| - 3 \ln |x-2| + C.
\]

**Partial fractions by plugging in roots**
The method above is the most natural, but now we will see an alternative method that is a little faster.

Let’s find the partial fraction decomposition for \( \frac{1}{(x+1)(x+2)} \), we again set
\[
\frac{1}{(x+1)(x+2)} = \frac{A}{x+1} + \frac{B}{x+2} = \frac{A(x+2) + B(x+1)}{(x+1)(x+2)},
\]
and we set the numerators equal:
\[
1 = A(x+2) + B(x+1).
\]
But then we plug in the roots of the linear factors (\( x = -1 \) is the root of \( x+1 \), \( x = -2 \) that of \( x+2 \)) to simply pick out the \( A \) and \( B \):

\[\text{Plug in } x = -1: \quad 1 = A \cdot (-1 + 2) + B \cdot 0 = A \Rightarrow A = 1,\]

\[\text{Plug in } x = -2: \quad 1 = A \cdot 0 + B \cdot (-2 + 1) = -B \Rightarrow B = -1.\]
We get the same $A$ and $B$ as before. The trick here is that by plugging in a root, the corresponding linear factor becomes 0, taking the $A$ or the $B$ out of the equation, and leaving the other easy to read off.

- Let’s do this for $\frac{x + 1}{x(2x + 3)}$.

$$\frac{x + 1}{x(2x + 3)} = A \frac{x}{x} + B \frac{2x + 3}{x(2x + 3)} = A(2x + 3) + Bx$$

**Numerators equal:** $x + 1 = A(2x + 3) + Bx$

- **Plug in $x = 0$:** $0 + 1 = A \cdot (0 + 3) + B \cdot 0 = 3A \Rightarrow A = \frac{1}{3}$

- **Plug in $x = \frac{-3}{2}$:** $\frac{-3}{2} + 1 = A \cdot 0 + B \cdot \frac{-3}{2} \Rightarrow B = \left(\frac{-2}{3}\right) \cdot \left(\frac{-1}{2}\right) = \frac{1}{3}$

So

$$\int \frac{x + 1}{x(2x + 3)} \, dx = \frac{1}{3} \int \frac{1}{x} \, dx + \frac{1}{3} \int \frac{1}{2x + 3} \, dx = \frac{1}{3} \ln |x| + \frac{1}{3} \cdot \frac{1}{2} \ln |2x + 3| + C$$

$$= \frac{1}{3} \ln |x| + \frac{1}{6} \ln |2x + 3| + C.$$

**Three linear factors**

- We can also see integrals of rational functions like $\int \frac{1}{(x + 1)(x + 2)(x + 3)} \, dx$.

For these ones, the plugging-in-roots method is considerably more convenient, because in the matching-coefficients method we would have to solve three equations in three unknowns. The procedure is the same:

$$\frac{1}{(x + 1)(x + 2)(x + 3)} = \frac{A}{x + 1} + \frac{B}{x + 2} + \frac{C}{x + 3}$$

$$= \frac{A(x + 2)(x + 3) + B(x + 1)(x + 3) + C(x + 1)(x + 2)}{(x + 1)(x + 2)(x + 3)}$$

**Numerators equal:** $1 = A(x + 2)(x + 3) + B(x + 1)(x + 3) + C(x + 1)(x + 2)$

- **Plug in $x = -1$:** $1 = A \cdot 1 \cdot 2 + B \cdot 0 + C \cdot 0 = 2A \Rightarrow A = 1/2$

- **Plug in $x = -2$:** $1 = A \cdot 0 + B \cdot (-1) \cdot (1) + C \cdot 0 = -B \Rightarrow B = -1$

- **Plug in $x = -3$:** $1 = A \cdot 0 + B \cdot 0 + C \cdot (-2) \cdot (-1) = C \Rightarrow C = 1/2$

So

$$\int \frac{1}{(x + 1)(x + 2)(x + 3)} \, dx = \int \left(\frac{1}{x + 1} - \frac{1}{x + 2} + \frac{1}{2x + 3}\right) \, dx$$

$$= \frac{1}{2} \ln |x + 1| - \ln |x + 2| + \frac{1}{2} \ln |x + 3| + C.$$

- Another example: $\int \frac{x^2 + 1}{x(1 - 3x)(x - 3)} \, dx$

$$\frac{x^2 + 1}{x(1 - 3x)(x - 3)} = \frac{A}{x} + \frac{B}{1 - 3x} + \frac{C}{x - 3} = \frac{A(1 - 3x)(x - 3) + Bx(x - 3) + Cx(1 - 3x)}{x(1 - 3x)(x - 3)}$$
Numerator equal: \(x^2 + 1 = A(1 - 3x)(x - 3) + Bx(x - 3) + Cx(1 - 3x)\)

Plug in \(x = 0\): \(0 + 1 = A \cdot 1 \cdot (-3) + B \cdot 0 + C \cdot 0 = -3A \Rightarrow A = -1/3\)

Plug in \(x = \frac{1}{3}\): \(\frac{1}{9} + 1 = A \cdot 0 + B \cdot \frac{1}{3} \cdot \left(-\frac{8}{3}\right) + C \cdot 0 = -\frac{8}{9}B \Rightarrow B = -\frac{5}{4}\)

Plug in \(x = 3\): \(9 + 1 = A \cdot 0 + B \cdot 0 + C \cdot (-8) = -8C \Rightarrow C = -\frac{5}{4}\)

So
\[
\int \frac{x^2 + 1}{x(1-3x)(x-3)} \, dx = \int \left( -\frac{11}{3x} - \frac{5}{1 - 3x} - \frac{5}{4x - 3} \right) \, dx
\]
\[
= -\frac{1}{3} \ln |x| - \frac{5}{4} \cdot \frac{1}{-3} \ln |1 - 3x| - \frac{5}{4} \ln |x - 3| + C
\]
\[
= -\frac{1}{3} \ln |x| + \frac{5}{12} \ln |1 - 3x| - \frac{5}{4} \ln |x - 3| + C.
\]

• For which functions can we use this?
We can use the method above whenever we have a rational function whose denominator factors completely into linear factors, without repeated factors.

For example, it does not work for the rational functions
\[
\frac{1}{x^2 + 1}, \quad \frac{x - 3}{x^2 + x + 1}, \quad \frac{1}{(x - 2)^2}, \quad \frac{x + 1}{x^3(x + 3)}.
\]
the first two have a denominator that does not factor into linear factors, as you can check using the quadratic formula \((D = b^2 - 4ac\) is negative, so there are no real roots); the third has the factor \((x - 2)\) twice; and the fourth has the factor \(x\) three times.

• Manipulating algebraically first
For a rational function whose denominator factors into linear factors, without repeated factors, there are two steps that we might have to undertake to get it into the form that we can work with:

○ Factoring: The examples above were all provided in factored form, but you may also have to factor them yourself. For instance, to evaluate \(\int \frac{1}{x^2 + 3x + 2} \, dx\) you first have to factor the denominator, using the quadratic formula or any other method that you like. You’d get
\[
\int \frac{1}{x^2 + 3x + 2} \, dx = \int \frac{1}{(x + 1)(x + 2)} \, dx,
\]
an example we saw above.

○ Long division: The examples above also all had the property that the degree of the numerator is strictly less than the degree of the denominator. If this isn’t the case, you will have to first change that by doing long division.

For example, take \(\int \frac{x^3 - 2x}{x^2 + 3x + 2} \, dx = \int \frac{x^3 - 2x}{(x + 1)(x + 2)} \, dx\).
The degree of the numerator is 3, and the degree of the denominator is 2. If we tried to do the method above right away, it would fail, since when we set the numerators equal, we’d get

\[x^3 - 3x = A(x + 2) + B(x + 1),\]
which is not possible (setting the coefficients of \(x^3\) equal gives \(1 = 0\)). What’s risky is that plugging-in-roots might still look possible, but don’t do it because it will give wrong answers. What we should do instead is use long division to make the degree of the numerator less:

\[
\frac{x^3 - 2x}{x^2 + 3x + 2} = x - 3 + \frac{5x + 6}{x^2 + 3x + 2},
\]

and then we can integrate the right hand side:

\[
\int \frac{x^3 - 2x}{x^2 + 3x + 2} \, dx = \int (x - 3) \, dx + \int \frac{5x + 6}{(x + 1)(x + 2)} \, dx.
\]

I won’t explain polynomial long division here; see the bottom of p.482 of the book for a worked out example.

14.3 Numerical integration

Numerical integration is the approximation of integrals by sums, which can be useful when you can’t evaluate a definite integral, but you would still like to have an approximate value for it. One way to do this is using Riemann sums: although we introduced these to be able to define definite integrals, we could just as well use them to approximate integrals.

• The Midpoint Rule

Of the three types of Riemann sums (left, right, and midpoint), the midpoint sums are the most suitable for approximation. You can see this if you draw the picture: a rectangle in the midpoint sum will always be partly above the graph and partly below, and the extra area from above the graph will partly compensate for the missed area below the graph.

We get the midpoint sum by splitting up the interval \([a, b]\) into \(n\) equal smaller intervals, using \(x_k = a + k \cdot \frac{b - a}{n}\). Then the \(k\)-th midpoint rectangle will have its bottom corners at \(x_{k-1}\) and \(x_k\), so it has width \(\Delta x = \frac{b - a}{n}\). Next we choose its height so that the graph of \(f(x)\) hits the top of the rectangle in the middle, which happens if the height is \(f\left(\frac{x_{k-1} + x_k}{2}\right)\). Then the area of the \(k\)-th rectangle is

\[
f\left(\frac{x_{k-1} + x_k}{2}\right) \cdot \Delta x.
\]

Adding up all these areas we get the resulting formula, that approximates \(\int_a^b f(x) \, dx\) with \(n\) rectangles:

\[
\text{Midpoint Rule: } M(n) = \sum_{k=1}^{n} f\left(\frac{x_{k-1} + x_k}{2}\right) \cdot \Delta x
\]

where \(\Delta x = \frac{b - a}{n}\), \(x_k = a + k \cdot \Delta x\)

• The Trapezoid Rule

The other approximation rule that we will use is obtained by using trapezoids instead of rectangles. As you will see when you draw the picture, a trapezoid can fit much more snugly underneath a graph than a rectangle can.

For this we will need the area of a trapezoid with a horizontal base of width \(w\) and two different heights (and so a slanted side on top) of height \(a\) and \(b\), which is \(\frac{a+b}{2} \cdot w\). When we fit such a trapezoid underneath a graph, with its bottom corners at \(x_{k-1}\) and \(x_k\) like for the rectangles
above (and so with width \( \Delta x \)), then the two heights will be \( f(x_{k-1}) \) and \( f(x_k) \). Hence one such trapezoid will have area
\[
\frac{f(x_{k-1}) + f(x_k)}{2} \cdot \Delta x.
\]
To approximate an integral \( \int_a^b f(x) \, dx \), we again divide up the interval using \( x_k = a + k \cdot \Delta x \), and add up the areas of all the trapezoids:
\[
\sum_{k=1}^{n} \frac{f(x_{k-1}) + f(x_k)}{2} \cdot \Delta x.
\]
This is not quite what we will call the trapezoid rule, because there is a slight improvement that we can make. I’ll illustrate this for \( n = 3 \):
\[
\sum_{k=1}^{3} \frac{f(x_{k-1}) + f(x_k)}{2} \cdot \Delta x = \left( \frac{f(x_0) + f(x_1)}{2} + \frac{f(x_1) + f(x_2)}{2} + \frac{f(x_2) + f(x_3)}{2} \right) \cdot \Delta x
\]
\[
= \left( \frac{f(x_0)}{2} + \frac{f(x_1)}{2} + \frac{f(x_1)}{2} + \frac{f(x_2)}{2} + \frac{f(x_2)}{2} + \frac{f(x_3)}{2} \right) \cdot \Delta x
\]
\[
= \left( \frac{f(x_0)}{2} + f(x_1) + f(x_2) + \frac{f(x_3)}{2} \right) \cdot \Delta x
\]
As you see, except for the first and last term, each two consecutive terms \( \frac{f(x_k)}{2} + \frac{f(x_{k+1})}{2} \) add up to \( f(x_k) \). This happens in general, and it leads to the final form of our trapezoid rule.
The resulting formula, that approximates \( \int_a^b f(x) \, dx \) with \( n \) trapezoids, is:

<table>
<thead>
<tr>
<th>Trapezoid Rule:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T(n) = \left( \frac{f(x_0)}{2} + \sum_{k=1}^{n-1} f(x_k) + \frac{f(x_n)}{2} \right) \cdot \Delta x )</td>
</tr>
<tr>
<td>with ( \Delta x = \frac{b-a}{n} ), ( x_k = a + k \cdot \Delta x )</td>
</tr>
</tbody>
</table>

- **Example**

I’ll give one easy example here; for more examples please see the book, which gives pictures and calculates the actual values of the approximations.

I’ll approximate the following integral with \( n = 4 \):
\[
\int_2^4 x^2 \, dx.
\]
To use the midpoint rule, we first need \( \Delta x = \frac{4-2}{4} = \frac{1}{2} \) and \( x_k = 2 + k \cdot \frac{1}{2} \). Then let’s write out all the midpoints:
\[
\frac{x_0 + x_1}{2} = \frac{2 + 2.5}{2} = 2.25, \quad \frac{x_1 + x_2}{2} = 2.75, \quad \frac{x_2 + x_3}{2} = 3.25, \quad \frac{x_3 + x_4}{2} = 3.75.
\]
Then we get the approximation
\[
M(4) = f\left( \frac{x_0 + x_1}{2} \right) \cdot \Delta x + f\left( \frac{x_1 + x_2}{2} \right) \cdot \Delta x + f\left( \frac{x_2 + x_3}{2} \right) \cdot \Delta x + f\left( \frac{x_3 + x_4}{2} \right) \cdot \Delta x
\]
\[
= 2.25^2 \cdot \frac{1}{2} + 2.75^2 \cdot \frac{1}{2} + 3.25^2 \cdot \frac{1}{2} + 3.75^2 \cdot \frac{1}{2} = (18.625).
\]
For the trapezoid rule, we use the same $\Delta x$ and $x_k$, and we get

$$T(4) = \frac{f(x_0)}{2} + f(x_1) + f(x_2) + f(x_3) + \frac{f(x_4)}{2} = \frac{2^2}{2} + 2.5^2 + 3^2 + 3.5^2 + \frac{4^2}{2} \ (= 18.75).$$

Just for comparison, in this case the exact value of the definite integral is

$$\int_2^4 x^2 \, dx = \frac{1}{3} x^3 \bigg|_2^4 = \frac{1}{3} (4^3 - 2^3) = \frac{56}{3} \approx 18.666 \cdots.$$
Chapter 15

Improper Integrals

15.1 Summary

By our definitions, what we will call improper integrals are not really integrals. But we will see below that we can handle them as limits of normal (proper) integrals, and they are not much harder to compute.

There are 2 types, which are improper integrals with:

- **Infinite intervals**: For example
  \[ \int_{0}^{\infty} e^{-x} \, dx; \]
  these have an infinite interval of integration, which means that they represent the area of an infinitely long region. But we will see that such a region can still have a finite area, if it is 'thin enough'.
  
  There are three types of infinite intervals: those that extend infinitely far to the right \((\int_{a}^{\infty})\), to the left \((\int_{-\infty}^{a})\), or to both sides \((\int_{-\infty}^{\infty})\).

- **Unbounded functions**: For example
  \[ \int_{0}^{1} \frac{1}{\sqrt{x}} \, dx; \]
  these have an integrand that is unbounded somewhere on the interval of integration (in this example at \(x = 0\)). So they represent the area of a region that is infinitely tall, which again can have a finite area.

Calculating improper integrals

Both types can be calculated as a limit of a proper integral. I will illustrate that for both by example:

\[
\int_{0}^{\infty} e^{-x} \, dx = \lim_{c \to \infty} \int_{0}^{c} e^{-x} \, dx = \lim_{c \to \infty} \left( -e^{-x} \right) \bigg|_{0}^{c} = \lim_{c \to \infty} (-e^{-c} + 1) = 0 + 1 = 1,
\]

\[
\int_{0}^{1} \frac{1}{\sqrt{x}} \, dx = \lim_{c \to 0^+} \int_{c}^{1} \frac{1}{\sqrt{x}} \, dx = \lim_{c \to 0^+} \left( 2\sqrt{x} \right) \bigg|_{c}^{1} = \lim_{c \to 0^+} (2 - 2\sqrt{c}) = 2 - 0 = 2.
\]

So for both types, we circumvent the 'problem' in the interval using a limit.
• for infinite intervals we use (leaving out the integrand for brevity):

\[ \int_a^\infty = \lim_{c \to \infty} \int_a^c, \quad \int_\infty^a = \lim_{c \to \infty} \int_c^a, \quad \int_\infty^\infty = \int_0^0 + \lim_{c \to \infty} \int_0^c + \lim_{c \to -\infty} \int_c^0; \]

• for a point \( p \) where the integrand is unbounded we use

\[ \int_p^b = \lim_{c \to p^+} \int_c^b, \quad \int_p^a = \lim_{c \to p^-} \int_c^a. \]

If the resulting limit exists, we say the improper integral converges. If the limit does not exist, we say the improper integral diverges. This happens for instance in

\[ \int_1^\infty \frac{1}{x} \, dx = \lim_{c \to \infty} \int_1^c \frac{1}{x} \, dx = \lim_{c \to \infty} \ln|x| \bigg|_1^c = \lim_{c \to \infty} \ln|c| = \infty \Rightarrow \text{integral diverges}. \]

Subtleties of improper integrals with unbounded functions

• Recognizing improper integrals with unbounded functions

While improper integrals with infinite intervals are easy to spot (they have \( \pm \infty \) in one of the limits of integration), the ones with unbounded functions can be harder to spot. Here are some common examples:

\[ \int_0^1 \frac{1}{x^p} \, dx, \quad \int_a^b \frac{1}{(x-b)^p} \, dx, \quad \int_a^b \frac{1}{(x-c)^p} \, dx, \]

for any \( p > 0 \). In the first two you clearly have division by 0 when you plug in one of the limits, but the third is trickier: here we have an unbounded function if \( a \leq c \leq b \), but not otherwise. To compute this, we would have to split it up as \( \int_a^b = \int_a^c + \int_c^b \).

In other examples it is not so clear that there is division by 0:

\[ \int_0^1 \frac{1}{\sqrt{1-x^2}} \, dx, \quad \int_0^{\pi/2} \tan(x) \, dx, \quad \int_0^1 \ln(x) \, dx. \]

For the last two we have to know that \( \tan(x) \) has a vertical asymptote at \( x = \pi/2 \), and that \( \ln(x) \) has a vertical asymptote at \( x = 0 \).

• One-sided limits

As we saw in the example

\[ \int_0^1 \frac{1}{\sqrt{x}} \, dx = \lim_{c \to 0^+} \int_c^1 \frac{1}{\sqrt{x}} \, dx \]

above, we have to use one-sided limits for these improper integrals. Another example, where the limit is from the other side, is

\[ \int_0^1 \frac{1}{\sqrt{1-x^2}} \, dx = \lim_{c \to 1^-} \int_0^c \frac{1}{\sqrt{1-x^2}} \, dx. \]

We have to do this because the function has an asymptote at 0, and we have to approach this from inside the integration interval (in fact, in the examples above the function doesn’t exist on the other side), otherwise the integral inside the limit would still include the asymptote. Fortunately, in most cases this does not make the limit more difficult to evaluate.

• Be careful with the FTC

Note that if a function is unbounded on \([a,b]\), then it is not continuous on \([a,b]\). This means
that the Fundamental Theorem of Calculus does not hold for this function (since continuity is a condition for that). For example, for the following integral we cannot use the FTC right away:

\[ \int_{-1}^{1} \frac{1}{x} \, dx \neq \ln |x| \bigg|_{-1}^{1}, \]

since \(1/x\) is unbounded at 0. And the result would be wrong, since the integral is the sum of the two diverging improper integrals \( \int_{-1}^{0} \frac{1}{x} \, dx \) and \( \int_{0}^{1} \frac{1}{x} \, dx \), so it must diverge as well.

### 15.2 Examples

**Examples of improper integrals with infinite intervals**

- **Evaluate** \( \int_{0}^{\infty} \frac{1}{(x+2)^3} \, dx \).

  \[
  \int_{0}^{\infty} \frac{1}{(x+2)^3} \, dx = \lim_{c \to \infty} \int_{0}^{c} \frac{1}{(x+2)^3} \, dx = \lim_{c \to \infty} \left( -\frac{1}{2} \frac{1}{(c+2)^2} - \frac{1}{4} \right)_{0}^{c} = -\frac{1}{2} \lim_{c \to \infty} \left( \frac{1}{(c+2)^2} - \frac{1}{4} \right) = -\frac{1}{2} \cdot \left( 0 - \frac{1}{4} \right) = \frac{1}{8}.
  \]

- **Evaluate** \( \int_{-\infty}^{1} \frac{1}{x^2} \, dx \).

  \[
  \int_{-\infty}^{1} \frac{1}{x^2} \, dx = \lim_{c \to -\infty} \int_{c}^{1} \frac{1}{x^2} \, dx = \lim_{c \to -\infty} \left( -\frac{1}{x} \right)_{c}^{1} = \lim_{c \to -\infty} \left( 1 + \frac{1}{c} \right) = 1 + 0 = 1.
  \]

- **Evaluate** \( \int_{-\infty}^{\infty} \frac{1}{x^2 + 1} \, dx \).

  \[
  \int_{-\infty}^{\infty} \frac{1}{x^2 + 1} \, dx = \int_{-\infty}^{0} \frac{1}{x^2 + 1} \, dx + \int_{0}^{\infty} \frac{1}{x^2 + 1} \, dx = \left( \lim_{c \to -\infty} \int_{c}^{0} \frac{1}{x^2 + 1} \, dx \right) + \left( \lim_{c \to \infty} \int_{0}^{c} \frac{1}{x^2 + 1} \, dx \right)
  = \left( \lim_{c \to -\infty} \arctan(x) \bigg|_{c}^{0} \right) + \left( \lim_{c \to \infty} \arctan(x) \bigg|_{0}^{c} \right)
  = \left( \lim_{c \to -\infty} \left( \arctan(0) - \arctan(c) \right) \right) + \left( \lim_{c \to \infty} \left( \arctan(c) - \arctan(0) \right) \right)
  = \left( \lim_{c \to -\infty} \left( -\arctan(c) \right) \right) + \left( \lim_{c \to \infty} \arctan(c) \right) = -\left( -\frac{\pi}{2} \right) + \frac{\pi}{2} = \pi.
  \]

Here we need to know the graph of \( \arctan(x) \), and in particular that it has the horizontal asymptote \( y = \pi/2 \) to the right, and the horizontal asymptote \( y = -\pi/2 \) to the left. These correspond to the vertical asymptotes of \( \tan(x) \).

- **Evaluate** \( \int_{-\infty}^{1} xe^{2x} \, dx \).

  This integral requires integration by parts, and it’s better to do it as an indefinite integral first:

  \[
  \int xe^{2x} \, dx = \frac{1}{2}xe^{2x} - \frac{1}{2} \int e^{2x} \, dx = \frac{1}{2}xe^{2x} - \frac{1}{4} e^{2x} + C.
  \]

  Then we can work out the improper integral in one go:

  \[
  \int_{-\infty}^{1} xe^{2x} \, dx = \lim_{c \to -\infty} \int_{c}^{1} xe^{2x} \, dx = \lim_{c \to -\infty} \left( \left( \frac{1}{2}xe^{2x} - \frac{1}{4} e^{2x} \right) \bigg|_{c}^{1} \right)
  \]

Here we need to know the graph of \( \arctan(x) \), and in particular that it has the horizontal asymptote \( y = \pi/2 \) to the right, and the horizontal asymptote \( y = -\pi/2 \) to the left. These correspond to the vertical asymptotes of \( \tan(x) \).
\[
\lim_{c \to -\infty} \left( \frac{1}{2} e^{2c} - \frac{1}{4} e^2 \right) - \left( \frac{1}{2} e^{2c} - \frac{1}{4} e^2 \right) = \frac{e^2}{4} - 0 = \frac{e^2}{4}.
\]

Note that we used \( \lim_{c \to -\infty} c e^{2c} = 0 \), which follows by l'Hopital's rule, or by observing that after replacing \( c \) by \(-u\), it is equivalent to \( \lim_{-u \to -\infty} (-u)e^{-2u} = -\lim_{-u \to -\infty} \frac{u}{e^{2u}} \), which is 0 since the denominator \( e^{2u} \) grows much faster than the numerator \( u \).

**Examples of improper integrals with unbounded functions**

- Evaluate \( \int_2^3 \frac{1}{(x - 2)^{3/4}} \, dx \).

  \[
  \int_2^3 \frac{1}{(x - 2)^{3/4}} \, dx = \lim_{c \to 2^+} \int_c^3 \frac{1}{(x - 2)^{3/4}} \, dx = \lim_{c \to 2^+} 4(x - 2)^{1/4} \bigg|_c^3 = 4 \lim_{c \to 2^+} (1 - (c - 2)^{1/4}) = 4 \cdot (1 - 0) = 4.
  \]

- Evaluate \( \int_0^1 \frac{1}{\sqrt{1 - x^2}} \, dx \).

  \[
  \int_0^1 \frac{1}{\sqrt{1 - x^2}} \, dx = \lim_{c \to 1^-} \int_0^c \frac{1}{\sqrt{1 - x^2}} \, dx = \lim_{c \to 1^-} \arcsin(x) \bigg|_0^c = \lim_{c \to 1^-} (\arcsin(c) - \arcsin(0))
  \]

  \[
  = \lim_{c \to 1^-} \arcsin(c) = \arcsin(1) = \frac{\pi}{2}.
  \]

- Evaluate \( \int_0^3 \frac{1}{x} \, dx \).

  \[
  \int_0^3 \frac{1}{x} \, dx = \lim_{c \to 0^+} \int_c^3 \frac{1}{x} \, dx = \lim_{c \to 0^+} (\ln |3| - \ln |c|) = -\infty,
  \]

  so this integral diverges.

- Evaluate \( \int_0^2 \frac{1}{\sqrt{x} - 1} \, dx \). This is one where the asymptote \( x = 1 \) of the integrand lies inside the interval of integration, so we have to split up the integral:

  \[
  \int_0^2 \frac{1}{\sqrt{x} - 1} \, dx = \int_0^1 \frac{1}{\sqrt{x} - 1} \, dx + \int_1^2 \frac{1}{\sqrt{x} - 1} \, dx
  \]

  \[
  = \left( \lim_{c \to 1^-} \int_0^c \frac{1}{\sqrt{x} - 1} \, dx \right) + \left( \lim_{d \to 1^+} \int_d^2 \frac{1}{\sqrt{x} - 1} \, dx \right)
  \]

  \[
  = \left( \lim_{c \to 1^-} \frac{3}{2} (x - 1)^{2/3} \bigg|_0^c \right) + \left( \lim_{d \to 1^+} \frac{3}{2} (x - 1)^{2/3} \bigg|_d^2 \right)
  \]

  \[
  = \frac{3}{2} \left( \lim_{c \to 1^-} ((c - 1)^{2/3} - 1) \right) + \frac{3}{2} \left( \lim_{d \to 1^+} (1 - (c - 1)^{2/3}) \right)
  \]

  \[
  = \frac{3}{2} (0 - 1) + \frac{3}{2} (1 - 0) = 0.
  \]
16.1 Introduction

- **Differential equation:** A differential equation (DE) is an equation involving a variable $t$, a function $y(t)$, and its derivative $y'(t)$ (or some higher derivative like $y''(t)$). For instance:

  \[
  y'(t) = 2y(t) + 3, \quad y' = t^2y^2, \quad y' = \frac{e^t}{1 + y^2}.
  \]

  The study of DEs is a large and important part of mathematics, and in this course we will only scratch its surface. We will see examples like the ones above, but there are many other forms that we will not see, like $\cos(y') = y^2 + t$ or $y'' = 2y + 3$. All DEs in this course will be first-order and separable, which we will define below.

- **First-order:** A DE is called first-order if the only derivative it involves is $y'$ (so no higher derivatives like $y''$).

- **Linear:** A first-order DE is called linear if we can put it in the form $y' = ay + b$.

- **Solution:** A solution to a DE is a function $y(t)$ that satisfies the equation. For example, $y(t) = e^{2t}$ is a solution of $y' = 2y$, because $y'(t) = (e^{2t})' = 2e^{2t} = 2y(t)$.

- **General solution:** The general solution of a first-order DE is a function that depends on a constant $C$ and gives all possible solutions. For example, $y(t) = C \cdot e^{2t}$ is the general solution of $y' = 2y$.

  *Note:* An antiderivative $\int f(t)dt$ is the general solution to the DE $y'(t) = f(t)$.

- **Initial Value Problem:** An initial value problem (IVP) is a DE together with an initial condition, which gives one value of the function. For instance:

  \[
  y'(t) = 2y(t), \quad y(0) = 3.
  \]

  An IVP has a unique solution, which you get from the general solution by determining the constant using the initial condition. For example, the IVP above has the unique solution $y(t) = 3e^{2t}$, which we obtain from the general solution as follows:

  \[
  y(t) = Ce^{2t} \Rightarrow 3 = y(0) = Ce^0 = C \Rightarrow C = 3.
  \]
Separable: A first-order DE is called separable if we can put it in the form
\[ y'(t) = \frac{g(t)}{h(y)}. \]

An example is \( y'(t) = \frac{t^2+1}{y^2} \). But often a separable DE is not given in this form, and we have to put it in that form ourselves:
\[ y' = t^3 y^2 \Rightarrow y' = \frac{t^3}{1/y^2}; \quad y' = e^{t+y} \Rightarrow y' = \frac{e^t}{e^{-y}}. \]

Separable equations are the only type of DE that we will learn to solve in this course.

Method for solving (some) separable equations

\[
\begin{align*}
\frac{dy}{dt} &= \frac{g(t)}{h(y)} \\
\Rightarrow h(y)dy &= g(t)dt \\
\Rightarrow \int h(y)dy &= \int g(t)dt \\
\Rightarrow H(y) &= G(t) + C, \quad \text{where } H' = H, \; G' = g \\
\Rightarrow \text{Solve this equation for } y \text{ to get the general solution.}
\end{align*}
\]

This method works for many separable equations, though not all, since the last two steps might not be possible:

- It might not be possible to find an antiderivative for \( h(y) \) or \( g(t) \);
- It might not be possible to solve the equation \( H(y) = G(t) + C \) for \( y \).

In this course, you will usually only see DEs for which this works out, but you should be aware of these limitations.

16.2 Examples

- Find the general solution to \( y' = \frac{t^2}{y^2} \).

\[
\begin{align*}
\frac{dy}{dt} &= \frac{t^2}{y^2} \Rightarrow y^2dy &= t^2dt \Rightarrow \int y^2dy &= \int t^2dt \\
\Rightarrow \frac{1}{3}y^3 &= \frac{1}{3}t^3 + C_1 \Rightarrow y^3 = t^3 + 3C_1 = t^3 + C \\
\Rightarrow [y(t)] &= \sqrt[3]{t^3 + C}.
\end{align*}
\]

- Solve the initial value problem \( y' = y, \; y(0) = 2 \).

\[
\begin{align*}
\frac{dy}{dt} &= y \Rightarrow \int \frac{1}{y}dy &= \int dt \Rightarrow \ln |y| &= t + C \\
\Rightarrow |y| &= e^{t+C} = e^C \cdot e^t \\
\Rightarrow y &= \pm e^C \cdot e^t.
\end{align*}
\]
⇒ \( y = Ae^t \).

This often happens with DEs like this: we get \(|y| = f(t)\), where \( f(t) \) is always positive. This means that either \( y = f(t) \) or \( y = -f(t) \), both of which are possible. In this case we have \( e^C e^t \) or \( -e^C e^t \), and we can glue these together as \( Ae^t \) with arbitrary \( A \) (which also includes \( y = 0 \), a solution that we missed before because we divided by \( y \)).

\[
y(0) = 2 \Rightarrow 2 = y(0) = Ae^0 = A \Rightarrow A = 2
\]

⇒ \( y(t) = 2e^t \).

• Solve the initial value problem \( y' = \frac{t+1}{ty} \), \( y(1) = -1 \).

\[
\frac{dy}{dt} = \frac{t+1}{ty} \Rightarrow \int y dy = \int \frac{t+1}{t} dt = \int \left( 1 + \frac{1}{t} \right) dt
\]

\[
\Rightarrow \frac{1}{2} y^2 = t + \ln |t| + C_1 \Rightarrow y^2 = 2t + 2 \ln |t| + 2C_1
\]

⇒ \( y = \pm \sqrt{2t + 2 \ln |t| + C} \).

When taking square roots we have to be careful: the general solution consists of two families of functions, those with + and those with −. Then from the initial condition \( y(1) = -1 \) we have to see which family we’re in; in this case the one with −.

\[
y(1) = -1 \Rightarrow -1 = y(1) = -\sqrt{2 + 0 + C} \Rightarrow C + 2 = 1 \Rightarrow C = -1
\]

⇒ \( y(t) = -\sqrt{2t + 2 \ln |t| - 1} \).

• Solve the initial value problem \( y' = t^2 y^3 \), \( y(1) = 3 \).

\[
\frac{dy}{dt} = t^2 y^3 \Rightarrow \int \frac{1}{y^3} dy = \int t^2 dt \Rightarrow -\frac{1}{2} y^{-2} = \frac{1}{3} t^3 + C_1
\]

\[
\Rightarrow \frac{1}{y^2} = -\frac{2}{3} t^3 - 2C_1 = -\frac{2}{3} t^3 + C \Rightarrow y^2 = \frac{1}{C - \frac{2}{3} t^3}
\]

⇒ \( y = \pm \frac{1}{\sqrt{C - \frac{2}{3} t^3}} \).

Again we have two options, and now \( y(1) = 3 \) tells us we should take the +. Then

\[
3 = y(1) = \frac{1}{\sqrt{C - \frac{2}{3} t^3}} \Rightarrow C - \frac{2}{3} = \frac{1}{9} \Rightarrow C = \frac{7}{9}
\]

⇒ \( y(t) = \frac{1}{\sqrt{\frac{7}{9} - \frac{2}{3} t^3}} = \frac{3}{\sqrt{7 - 6t^3}} \).
16.3 Word problems

• Population growth

Let \( P(t) \) be the number of fruit flies in thousands that we have in a box after \( t \) days, starting with 100 fruit flies. A reasonable model for this population is given by the DE

\[
P'(t) = kP(t) \left(1 - \frac{P(t)}{K}\right).
\]

This is called the logistic equation. The factor \( kP(t) \) reflects that the rate at which the population reproduces is proportional to the population size; the factor \( 1 - \frac{P(t)}{K} \) reflects that because of limited space and food, the growth will slow down as the population approaches some maximum \( K \).

Solve this initial value problem under the assumption that \( k = 1, \ K = 1 \) (this is not so realistic, but it makes the calculation easier).

So the DE is \( P' = P(1 - P) \), which is separable, and we can solve it as before:

\[
\int \frac{1}{P(1-P)} dP = \int dt,
\]

where we need partial fractions to get \( \frac{1}{P(1-P)} = \frac{1}{P} + \frac{1}{1-P} \), so we get (note the – in front of \( \ln |1 - P| \)):

\[
t + C_1 = \int \frac{1}{P} dP + \int \frac{1}{1-P} dP = \ln |P| - \ln |1-P| = \ln \left| \frac{P}{1-P} \right|
\]

\[
\Rightarrow \left| \frac{P}{1-P} \right| = e^{t+C_1} = e^{C_1} e^t = Ce^t.
\]

To be able to solve for \( P \), we’ll have to get rid of the absolute value. We know that \( P(t) \geq 0 \), so \( |P| = P \); and \( P(t) < 1 \) since it can’t grow past its maximum of 1000, so \( |1-P| = 1-P \); therefore we have \( \left| \frac{P}{1-P} \right| = \frac{P}{1-P} \). Then:

\[
\frac{P}{1-P} = Ce^t \Rightarrow P = Ce^t - Ce^t \cdot P \Rightarrow P(1 + Ce^t) = Ce^t
\]

\[
\Rightarrow P(t) = C e^t \cdot \frac{e^{-t}}{e^{-t} + C} = \frac{C}{e^{-t} + C}.
\]

Now we plug in the initial condition \( P(0) = 0.1: \)

\[
\frac{1}{10} = \frac{C}{1+C} \Rightarrow 1 + C = 10C \Rightarrow C = \frac{1}{9},
\]

\[
\Rightarrow P(t) = \frac{1/9}{e^{-t} + 1/9} = \frac{1}{9e^{-t} + 1}.
\]

• A savings account

Suppose you have a savings account with an initial balance of $1000. The interest rate is 5%, annually and compounded continuously. You make deposits at a continuous annual rate of $500.

Set up an initial value problem for the balance \( B(t) \) (with \( t \) in years) of this account and solve it. How long will it take until you have $10,000 in this account?

The rate of change of the balance, \( B'(t) \), equals the rate of change due to the deposits plus the rate of change due to interest. The first is simply 500, and the second 5% of the current amount, so 0.05\( B(t) \). This gives

\[
B'(t) = 0.05B(t) + 500.
\]
The initial condition is \(B(0) = 1000\).

To solve this initial value problem, we first find the general solution to this separable DE:

\[
\frac{dB}{dt} = 0.05B + 500 \implies \int \frac{1}{0.05B + 500} dB = \int dt
\]

\[
\implies \frac{1}{0.05} \ln |0.05B + 500| = t + C_1 \implies \ln |0.05B + 500| = 0.05t + C_2
\]

\[
\implies |0.05B + 500| = e^{0.05t+C_2} = e^{C_2} \cdot e^{0.05t} = C_3 e^{0.05t}.
\]

We could argue away the absolute value like in the examples above, but for word problems it’s often easier: we know that \(B > 0\) (it starts positive and only grows), hence also \(0.05B + 500 > 0\), so we get

\[
\implies 0.05B + 500 = C_3 \cdot e^{0.05t} \implies B = \frac{1}{0.05} (C_3 e^{0.05t} - 500) = 20C_3 e^{0.05t} - 10000
\]

\[
\implies B(t) = C e^{0.05t} - 10000.
\]

Now we plug in the initial condition:

\[
1000 = B(0) = C - 10000 \implies C = 11000 \implies B(t) = 11000 e^{0.05t} - 10000.
\]

To answer the last question, we solve:

\[
11000 e^{0.05t} - 10000 = 10000 \implies e^{0.05t} = \frac{20000}{11000} = \frac{20}{11}
\]

\[
\implies t = \frac{1}{0.05} \ln \left(\frac{20}{11}\right) \approx 12 \text{ years}.
\]
Chapter 17
Two-variable Limits and Continuity

17.1 Introduction

Here is the definition of a 2-variable limit:

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

means that as \((x,y)\) approaches \((a,b)\) along any path, the value \(f(x,y)\) approaches \(L\). If different paths give different \(L\), we say that the limit does not exist (DNE).

This is similar to the definition of a 1-variable limit \(\lim_{x \to a} f(x)\), with the distinction that in 2 dimensions there are many different ways for \((x,y)\) to approach \((a,b)\) (ie different 'paths'), whereas in 1 dimension \(x\) can only approach \(a\) from the left or the right. This will matter when we want to prove non-existence (see the 2-path test below), though not really when we’re computing the value of a limit that does exist.

We say that a function \(f(x,y)\) is continuous at \((a,b)\) if

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

In other words, a function is continuous at \((a,b)\) if we can compute \(\lim_{(x,y) \to (a,b)} f(x,y)\) by just plugging in. So we can show that a function is continuous by computing a limit, but once we know that most functions are continuous, that also lets us compute many limits by plugging in.

Fact: All combinations (by addition, multiplication, division, composition) of basic functions (rational functions, roots, exponentials, trigonometric, etc.) are continuous, unless there somehow is: division by zero, \(\ln(0)\), or a piecewise defined function.

Of these 3 exceptions, we will probably only see division by zero; the others I just have to mention for correctness, I will ignore them below.

Computing limits
To compute a limit \(\lim_{(x,y) \to (a,b)} f(x,y)\), here’s what you do:

- Try to plug in \((a,b)\) (in your head or on the side);
• If there is no division by zero, then \( f(x,y) \) is continuous at \((a,b)\), so you can compute the limit by plugging in.

• If there is division by zero, and it’s of the form \( \frac{\text{zero}}{\text{zero}} \), then it’s possible that the limit does exist; you can try one of the following tricks:
  - factoring
  - multiplying by the conjugate
  - substituting to 1 variable

If none of these applies, there might be a different trick you have to use, or the limit might not exist.

**Showing a limit does not exist**
If plugging in gives \( \frac{\text{zero}}{\text{zero}} \), but no trick seems to work, you might have to use the

2-path test for non-existence:
If \( f(x,y) \) approaches different values for 2 different paths approaching \((a,b)\),
then the limit does not exist.

In most examples that we will see, we will have \((a,b) = (0,0)\), and the different paths are lines \((y = mx)\), parabolas \((y = mx^2)\), or something similar \((y = mx^k)\).

## 17.2 Examples

For each of the following limits, compute the limit, or show that it doesn’t exist.

• \( \lim_{(x,y) \to (1,2)} \frac{x + y}{9x - y^2} \)
  When we try to plug in, we don’t get division by zero, hence the function is continuous at \((1,2)\),
  hence we can compute the limit by plugging in:
  \[
  \lim_{(x,y) \to (1,2)} \frac{x + y}{9x - y^2} = \frac{1 + 2}{9 \cdot 1 - 2^2} = \frac{3}{5}.
  \]

• \( \lim_{(x,y) \to (1,3)} \frac{x^2 - 1}{xy + x - y - 1} \)
  Here plugging in would give division by zero, and we have \( \frac{\text{zero}}{\text{zero}} \), so we need to do something else first. The form \( x^2 - 1 \) is begging to be factored, which suggests we have to factor the bottom as well:
  \[
  \lim_{(x,y) \to (1,3)} \frac{x - 1}{y + 1} = \frac{x - 1}{y + 1} = \frac{2}{4} = \frac{1}{2}.
  \]

• \( \lim_{(x,y) \to (1,1)} \frac{\sqrt{x} - \sqrt{y}}{x - y} \)
  Again we have \( \frac{\text{zero}}{\text{zero}} \), and in this case the square roots suggest that we should multiply by a conjugate on top and on bottom:
  \[
  \lim_{(x,y) \to (1,1)} \frac{\sqrt{x} - \sqrt{y}}{x - y} = \frac{\sqrt{x} + \sqrt{y}}{\sqrt{x} + \sqrt{y}} = \lim_{(x,y) \to (1,1)} \frac{x - y}{(\sqrt{x} + \sqrt{y})^2} = \lim_{(x,y) \to (1,1)} \frac{1}{\sqrt{x} + \sqrt{y}} = \frac{1}{2}.
  \]
Note that another way to do the same thing is to factor \( x - y = (\sqrt{x} - \sqrt{y})(\sqrt{x} + \sqrt{y}) \) on the bottom.

- \( \lim_{(x,y) \to (1,1)} \frac{x^2 y^2 - 1}{xy - 1} \)

Here we could note that \( x \) and \( y \) always occur in the same form \( xy \). That allows us to make the substitution \( u = xy \), reducing the limit to a 1-variable one. Note that as \( (x, y) \to (1, 1) \) we have \( u \to 1 \cdot 1 = 1 \).

\[
\lim_{(x,y) \to (1,1)} \frac{x^2 y^2 - 1}{xy - 1} = \lim_{u \to 1} \frac{u^2 - 1}{u - 1} = \lim_{u \to 1} \frac{(u - 1)(u + 1)}{u - 1} = \lim_{u \to 1} u + 1 = 2.
\]

- \( \lim_{(x,y) \to (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2} \)

Just like for the one above, \( x \) and \( y \) only occur in one form, as \( x^2 + y^2 \), so we can substitute \( u = x^2 + y^2 \). Then \( u \to 0^2 + 0^2 = 0 \), so

\[
\lim_{(x,y) \to (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2} = \lim_{u \to 0} \frac{\sin(u)}{u} = 1.
\]

Note that here we used the standard trigonometric limit \( \lim_{x \to 0} \frac{\sin(x)}{x} = 1 \); if this isn’t familiar, see p.155 of the book or ask me. The other one that you should know is \( \lim_{x \to 0} \frac{\cos(x) - 1}{x} = 0 \), and you should be able to do variations of these.

- \( \lim_{(x,y) \to (0,0)} \frac{x + y}{x - y} \)

Here plugging in would give \( \frac{\text{zero}}{\text{zero}} \), and none of the tricks above seems to work. That suggests the limit might not exist, which we might be able to prove with the 2-path test.

The easiest type of path to try is a line, so let’s try an arbitrary line \( y = mx \), and then see if for two different \( m \) we get different values:

\[
\lim_{(x,y) \to (0,0)} \frac{x + y}{x - y} = \lim_{x \to 0} \frac{x + mx}{x - mx} = \lim_{x \to 0} \frac{x(1 + m)}{x(1 - m)} = \lim_{x \to 0} \frac{1 + m}{1 - m} = \frac{1 + m}{1 - m}.
\]

So for two different \( m \), for instance \( m = 2 \) and \( m = 3 \), the results are \( \frac{1+2}{1-2} = -3 \) and \( \frac{1+3}{1-3} = -2 \) (in fact pretty much any two \( m \) give different results). Therefore if \( (x, y) \) approaches \((0,0)\) along the paths \( y = 2x \) and \( y = 3x \), the function approaches different values, so by the 2-path test this limit does not exist.

- \( \lim_{(x,y) \to (0,0)} \frac{x^2 + 2y}{2x^2 + y} \)

This is similar to the one above, but now lines \( y = mx \) do not help:

\[
\lim_{(x,y) \to (0,0)} \frac{x^2 + 2y}{2x^2 + y} = \lim_{x \to 0} \frac{x^2 + 2mx}{2x^2 + mx} = \lim_{x \to 0} \frac{x + 2m}{2x + m} = \frac{0 + 2m}{0 + m} = 2.
\]

So whatever line we approach along, the function value approaches 2. This does not mean that the test failed, just that with lines it does not work.

We can try something else, like parabolas \( y = mx^2 \):

\[
\lim_{(x,y) \to (0,0)} \frac{x^2 + 2y}{2x^2 + y} = \lim_{x \to 0} \frac{x^2 + 2mx^2}{2x^2 + mx^2} = \lim_{x \to 0} \frac{1 + 2m}{2 + m} = \frac{1 + 2m}{2 + m}.
\]
This is what we want, since now for instance \( m = 1 \) and \( m = 2 \) give the values 1 and \( 5/4 \). That means that if we approach along \( y = x^2 \) and \( y = 2x^2 \), we get different values, so by the 2-path test this limit does not exist.

Note that you do not have to guess these paths out of thin air; they’re usually of the form \( y = mx^k \), and you choose the exponent \( k \) so that all the \( x \)'s will factor out. For instance, if the denominator is \( x^3 + y^4 \), plugging in \( y = mx^{3/4} \) gives \( x^3 + m^4(x^{3/4})^4 = x^3 + m^4x^3 = x^3(1 + m^4) \).
Chapter 18

Partial Derivatives and the Chain Rule

18.1 Introduction

Definitions
The partial derivatives of \( f(x,y) \) are defined as follows:

- \( \frac{\partial f}{\partial x} \): the partial derivative of \( f(x,y) \) with respect to \( x \) is the derivative of \( f(x,y) \) with \( y \) considered constant;

- \( \frac{\partial f}{\partial y} \): the partial derivative of \( f(x,y) \) with respect to \( y \) is the derivative of \( f(x,y) \) with \( x \) considered constant.

The symbol \( \partial \) is just pronounced ‘d’, and it signifies that we’re not dealing with a normal 1-variable derivative, but with a partial derivative.

Some other notation for partial derivatives that we will use:

\[
\frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x,y) = f_x(x,y), \quad \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x,y) = f_y(x,y).
\]

Examples
There will be more examples below, but here are a few to illustrate:

\( f(x,y) = x^2 + y \Rightarrow \frac{\partial f}{\partial x} = 2x. \)

After a bit of practice this will come naturally, but at first you can think of really putting in a constant \( c \) for \( y \) to see what the derivative looks like: \( \frac{\partial f}{\partial x} = \frac{d}{dx} (x^2 + c) = 2x + 0 = 2x. \)

\( f(x,y) = xy \Rightarrow \frac{\partial f}{\partial x} = y; \)

here you can think of \( \frac{\partial f}{\partial x} = \frac{d}{dx} (cx) = c. \) Of course, in the real answer you have to make sure to put a \( y \) and not a \( c. \)

It can get a bit trickier when you have to use the chain rule:

\[
\frac{\partial}{\partial y} \sqrt{xy} = \frac{1}{2\sqrt{xy}} \cdot \frac{\partial}{\partial y} (xy) = \frac{1}{2\sqrt{xy}} \cdot x = \frac{1}{2} \frac{x}{\sqrt{xy}} = \frac{1}{2} \frac{\sqrt{x}}{\sqrt{y}}.
\]
here you can think of \( \frac{\partial f}{\partial y} = \frac{d}{dy}\sqrt{cy} = \frac{1}{\sqrt{2c}} \cdot c. \)

**Partial derivative at a point** \((a,b)\)

A partial derivative of a 2-variable function is a 2-variable function itself, so we can evaluate it at a point \((a,b)\), which we will write as \( \frac{\partial f}{\partial x}(a,b) \). For example, if we want the partial derivative with respect to \( x \) of \( f(x,y) = 3x^2 + 2xy + 5y \) at the point \((1,4)\), then we write

\[
\frac{\partial f}{\partial x} = 6x + 2y \quad \Rightarrow \quad \frac{\partial f}{\partial x}(1,4) = 6 \cdot 1 + 2 \cdot 4 = 14.
\]

Different notation for the same thing would be

\[
\frac{\partial f}{\partial x}(1,4) = (6x + 2y)(x,y) = (1,4) = 14,
\]

\[
\frac{\partial}{\partial x}(3x^2 + 2xy + 5y)(x,y) = (6x + 2y)(x,y) = (1,4) = 14.
\]

These forms can come in handy if you don’t want to write out the partial derivative by itself, or if you don’t want to separately define \( f \).

Here’s a trick that can save a lot of work, and also tests if you understand this notation; on the other hand, please do not use this if you do not fully understand it.

\[
\frac{\partial}{\partial x} \left( x^4 \right)(x,y) = \frac{d}{dx} \left( x^4 \right) \bigg|_{x=1} = \frac{d}{dx} x^2 \bigg|_{x=1} = 2x \bigg|_{x=1} = 2.
\]

The trick is that instead of differentiating with respect to \( x \) and plugging in \((a,b)\), you can 'slip in' \( y = b \) first, which gives you a 1-variable derivative that is often easier (in this case you would have had to do an elaborate quotient rule).

**Warning**

It is important to note that \( \frac{\partial f}{\partial x}(a,b) \) means ‘differentiate \( f(x,y) \) with respect to \( x \), then plug in \((a,b)\), and not the other way around. So:

\[
\frac{\partial f}{\partial x}(a,b) \neq \frac{\partial}{\partial x} (f(a,b));
\]

the second is a derivative of a constant, hence always zero. The same distinction exists for functions of 1 variable \((f'(a) \neq \frac{d}{dx} f(a) = 0)\), but with 2 variables it’s easier to get confused.

Similarly, the vertical bar notation \( \bigg|_{(x,y)=(a,b)} \) means that you plug in \((a,b)\) after you’ve differentiated.

**Geometric Interpretation**

I can’t give the pictures here, but I’ll describe the graphical interpretation of partial derivatives. Please draw the picture yourself.

- \( \frac{\partial f}{\partial x}(a,b) \) = slope of the curve \( z = f(x,b) \) in the plane \( y = b \);
- \( \frac{\partial f}{\partial y}(a,b) \) = slope of the curve \( z = f(a,y) \) in the plane \( x = a \).

In other words: to find \( \frac{\partial f}{\partial x}(a,b) \), you hold \( y \) constant at \( b \); this is the same as intersecting the surface \( z = f(x,y) \) with the plane \( y = b \), the result of which is the curve \( z = f(x,b) \); then you differentiate \( f(x,b) \) and plug in \( a \), which gives the slope of the tangent line above \( x = a \).
18.2 The Chain Rule for $z = f(x(t), y(t))$

In two variables, there are different kinds of Chain Rules; the one we will see is for situations like $z = f(x(t), y(t))$, where $z$ is a function of $x$ and $y$, and $x$ and $y$ are both functions of $t$. In such a situation, $t$ is called the independent variable, because it doesn’t depend on any other variable, while $x$, $y$ and $z$ are called dependent variables, since they depend on other variables.

Recall that one way to write the 1-variable Chain Rule for $y = f(x(t))$ is $\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}$. The same notation works best for the 2-variable Chain Rule with 1 independent variable:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt}$$

Note that since $z = f(x,y)$ is a 2-variable function, $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ are partial derivatives, so are written with $\partial$’s, while the other derivatives are of 1-variable functions, so are written with $d$’s. This notation is nice and short, but can be a bit misleading; we should also write it out more explicitly, with $z = g(t)$ denoting $z$ directly as a function of $t$:

$$z = g(t) = f(x(t), y(t)) \Rightarrow \frac{dz}{dt} = f_x(x(t), y(t)) \cdot x'(t) + f_y(x(t), y(t)) \cdot y'(t).$$

Then if we plug in a value $t = a$, we get:

$$\left. \frac{dz}{dt} \right|_{t=a} = g'(a) = f_x(x(a), y(a)) \cdot x'(a) + f_y(x(a), y(a)) \cdot y'(a).$$

Note that in theory we could always write out $z = g(t)$ explicitly, and differentiate that as a 1-variable function; but in practice this might be too much work, since this will be a more complicated function, hence more work to differentiate.

**Examples**

Let’s apply this to

$$z = f(x, y) = x^2 + 5y^2, \quad x(t) = t^2, \quad y(t) = t^3.$$

Then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt} = \left( \frac{\partial}{\partial x} (x^2 + 5y^2) \right) \cdot \left( \frac{dt}{dt} \right) + \left( \frac{\partial}{\partial y} (x^2 + 5y^2) \right) \cdot \left( \frac{dt}{dt} \right) = 2x \cdot 2t + 10y \cdot 3t^2 = 2t^2 \cdot 2t + 10t^4 \cdot 3t^2 = 4t^3 + 30t^5.$$

In this case it’s not hard to write out $z = g(t)$ explicitly:

$$z = f(t^2, t^3) = (t^2)^2 + 5(t^3)^2 = t^4 + 5t^6,$$

and then differentiate this as a 1-variable function, which gives the same result:

$$\frac{dz}{dt} = \frac{dt}{dt} (t^4 + 5t^6) = 4t^3 + 30t^5.$$

Here is another example:

$$z = f(x, y) = xy, \quad x(t) = \cos(t), \quad y(t) = \sin(t).$$
Then
\[
\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt} = \left( \frac{\partial}{\partial x} \cos(t) \right) \cdot \left( \frac{d}{dt} \cos(t) \right) + \left( \frac{\partial}{\partial y} \sin(t) \right) \cdot \left( \frac{d}{dt} \sin(t) \right)
\]

\[
= y \cdot (-\sin(t)) + x \cdot \cos(t)
\]

\[
= \sin(t) \cdot (-\sin(t)) + \cos(t) \cdot \cos(t)
\]

\[
= \cos^2(t) - \sin^2(t) = \cos(2t).
\]

Again, we can check this:

\[
z = f(\cos(t), \sin(t)) = \cos(t) \sin(t) \Rightarrow \frac{dz}{dt} = (\sin(t)) \cdot \cos(t) + \cos(t) \cdot \sin(t) = \cos^2(t) - \sin^2(t).
\]

Implicit Differentiation

With this Chain Rule, we can revisit implicit differentiation, a topic you have probably seen in your differential calculus course. The goal is to find \( \frac{dy}{dx} \) if \( y \) and \( x \) are given implicitly by an equation

\[F(x, y) = 0.
\]

For instance, for the circle \( x^2 + y^2 - 1 = 0 \), we would like to find \( \frac{dy}{dx} \), the slope of the tangent line, without having to write \( y \) as a function of \( x \).

Assuming \( y = y(x) \) is a function of \( x \), we can do this as follows, using the Chain Rule:

\[0 = F(x, y(x)) \Rightarrow 0 = \frac{\partial F}{\partial x} \frac{dx}{dx} + \frac{\partial F}{\partial y} \frac{dy}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \frac{dy}{dx}.
\]

For the example of the circle, \( F(x, y) = x^2 + y^2 - 1 \), this gives

\[0 = 2x + 2y \cdot \frac{dy}{dx} \Rightarrow \frac{dy}{dx} = -\frac{x}{y}.
\]

In general, we get the following formula for implicit differentiation:

\[0 = F(x, y(x)) \Rightarrow 0 = F_x + F_y \cdot \frac{dy}{dx} \Rightarrow \frac{dy}{dx} = -\frac{F_x}{F_y}.
\]

Geometric Interpretation: Curves on Surfaces

The situation where

\[z = f(x, y), \quad x = x(t), \quad y = y(t),
\]

can be interpreted as a curve on a surface. The surface is \( z = f(x, y) \) (actually a graph), and the curve on it is

\[(x(t), y(t), f(x(t), y(t))).
\]

This is called a parametrized curve: for each value of \( t \) (the parameter), you get a 3D point, and all those points together make up a curve in 3D. Since by construction each of those points \( (x, y, z) \) satisfies \( z = f(x, y) \), this curve lies on the surface.

Another way to see this is that \( (x(t), y(t)) \) gives a parametrized curve in the \( xy \)-plane, and you create the 3D curve by drawing a point above \( (x(t), y(t)) \) at height \( f(x(t), y(t)) \).

In applications, \( t \) might represent time, so that the curve represents a path on the surface that is traversed over time. Then a typical thing that you might want to know is the rate at which at your height is changing, which is

\[\frac{dz}{dt} = f_x \cdot x'(t) + f_y \cdot y'(t).
\]
For example, suppose you’re walking on the plane
\[ x + y + z = 1 \quad \Rightarrow \quad z = f(x, y) = 1 - x - y, \]
above the unit circle (so you’re walking along an ellipse on the plane), which is parametrized by
\[ x(t) = \cos(t), \quad y(t) = \sin(t). \]
Then you’re height is changing at the rate
\[
\frac{dz}{dt} = f_x \cdot x'(t) + f_y \cdot y'(t) = -1 \cdot (-\sin(t)) + -1 \cdot \cos(t) = \sin(t) - \cos(t).
\]
For instance, at time \( t = 0 \) you’re at \( (x, y, z) = (\cos(0), \sin(0), f(\cos(0), \sin(0))) = (1, 0, 0), \) and your height is changing at the rate
\[
\left. \frac{dz}{dt} \right|_{t=0} = \sin(0) - \cos(0) = -1,
\]
which means you’re going downhill.
Chapter 19

Two-variable Optimization

19.1 Optimization

Local extrema

Our main goal in optimization of 2-variable functions will be to find their local extrema:

- $f(x, y)$ has a local maximum at $(a, b)$ if $f(x, y) \leq f(a, b)$ for all $(x, y)$ around $(a, b)$.
- $f(x, y)$ has a local minimum at $(a, b)$ if $f(x, y) \geq f(a, b)$ for all $(x, y)$ around $(a, b)$.

'Local' means it is the largest/smallest value 'in a neighborhood' around $(a, b)$; there could be larger/smaller values further away. The largest/smallest value overall is called the absolute maximum/minimum, but we won’t cover those in this course.

To find local extrema, we won’t really use these definitions directly, we will use critical points.

Critical points

The point $(a, b)$ is called a critical point of $f(x, y)$ if $f_x(a, b) = 0$ and $f_y(a, b) = 0$, or if $f_x$ or $f_y$ does not exist at $(a, b)$.

Actually, we won’t see the second type (non-existing partial derivatives) much in this course, we’ll focus on the type where the partial derivatives are zero.

Compare this definition to that for 1-variable functions $f(x)$: they have a critical point at $x = a$ if $f'(a) = 0$ or if $f'(a)$ does not exist. Just like for 1-variable functions, the reason for considering critical points of 2-variable functions is this:

Fact: If $f(x, y)$ has a local maximum or minimum at $(a, b)$, then $(a, b)$ is a critical point.

$\Rightarrow$ Critical points are the candidates to be local maxima or minima.

Saddle points

A saddle point is a critical point $(a, b)$ of $f(x, y)$ around which both $f(x, y) > f(a, b)$ and $f(x, y) < f(a, b)$ occur.

So a saddle point is a critical point that is not a local maximum or minimum. The corresponding type of point for 1-variable functions is an inflection point. For instance $x = 0$ for $f(x) = x^3$: it is a critical point since $f'(0) = 3 \cdot 0^2 = 0$, but a quick look at the graph shows that there are lower and higher points around it.

The typical example of a saddle point is $(0, 0)$ on the hyperbolic paraboloid $z = x^2 - y^2$; in fact, it is called a saddle point because this surface looks like a saddle (if you draw it right). But keep in mind that not every saddle point has to look like a saddle.
Second Derivative Test

This is our main method for classifying a critical point as a local maximum, local minimum or saddle point. Given \((a, b)\) with \(f_x(a, b) = 0\) and \(f_y(a, b) = 0\), compute the discriminant \(D(x, y) = f_{xx}f_{yy} - f_{xy}^2\). Then:

- \(D(a, b) > 0\) and \(f_{xx}(a, b) < 0 \implies (a, b)\) is a local maximum;
- \(D(a, b) > 0\) and \(f_{xx}(a, b) > 0 \implies (a, b)\) is a local minimum;
- \(D(a, b) < 0 \implies (a, b)\) is a saddle point;
- \(D(a, b) = 0 \implies the test is inconclusive.

Compare this to the Second Derivative Test for critical points \((x = a\) with \(f'(a) = 0\)\) of 1-variable functions: if \(f''(a) < 0\), then \(x = a\) is a local maximum; if \(f''(a) > 0\), then \(x = a\) is a local minimum; and if \(f''(a) = 0\), the test is inconclusive.

Examples

- **Find the critical points of** \(x^2 + y^2 - 2x - 6y + 14\), and classify them as local maximum, local minimum or saddle point.

To find the critical points, we compute the partial derivatives and set them equal to 0:

\[
f_x = 2x - 2 = 0, \quad f_y = 2y - 6.
\]

In this case, solving the equations is easy: the first gives \(x = 1\), the second gives \(y = 3\), so the only critical point is \((1, 3)\).

To classify this critical point, we compute the second partial derivatives and the discriminant:

\[
f_{xx} = 2, \quad f_{yy} = 2, \quad f_{xy} = 0
\]

\[
\implies D(x, y) = f_{xx}f_{yy} - f_{xy}^2 = 2 \cdot 2 - 0^2 = 4.
\]

Then \(D(1, 3) = 4 > 0\), so we have to look at \(f_{xx}(1, 3) = 2 > 0\), from which we can tell that \((1, 3)\) is a local minimum.

- **Find and classify the critical points of** \(x^3 - 12xy + 8y^3\).

\[
\implies f_x = 3x^2 - 12y = 0, \quad f_y = -12x + 24y^2 = 0.
\]

To solve these equations, we can isolate \(y = \frac{1}{4}x^2\) from the first equation, and plug that into the second:

\[
0 = -12x + 24 \left( \frac{1}{4}x^2 \right)^2 = -12x + \frac{24}{16}x^4 = x \left( \frac{3}{2}x^3 - 12 \right) \implies x = 0, \quad x = \sqrt[3]{\frac{2}{3}} \cdot 12 = 2.
\]

The corresponding \(y\)-values are \(\frac{1}{4}0^2 = 0\) and \(\frac{1}{4}2^2 = 1\), so the critical points are \((0, 0)\) and \((2, 1)\).

\[
f_{xx} = 6x, \quad f_{yy} = 48y, \quad f_{xy} = -12
\]

\[
\implies D(x, y) = 6 \cdot 48y - (-12)^2 = 288xy - 144 = 144(2xy - 1).
\]

Then \(D(0, 0) = -144 < 0\), so \((0, 0)\) is a saddle point. Also \(D(2, 1) = 144 \cdot 3 > 0\), and \(f_{xx}(2, 1) = 12 > 0\), so \((2, 1)\) is a local minimum.

- **Find and classify the critical points of** \(f(x, y) = 3y^2 - 2y^3 - 3x^2 + 6xy\).

\[
\implies f_x = -6x + 6y = 0, \quad f_y = 6y - 6y^2 + 6x = 0
\]
To solve these equations, we can get $y = x$ from the first equation, and plug that into the second:

$$\Rightarrow \quad 6y - 6y^2 + 6y = 0 \quad \Rightarrow \quad 0 = 12y - 6y^2 = 6y(2 - y) \quad \Rightarrow \quad y = 0, y = 2.$$ 

So the critical points are $(0, 0)$ and $(2, 2)$.

$$f_{xx} = -6, \quad f_{yy} = 6 - 12y, \quad f_{xy} = 6$$

$$\Rightarrow \quad D(x, y) = -6(6 - 12y) - 6^2 = -72 + 72y = 72(y - 1).$$

Then for the critical point $(0, 0)$ we have $D(0, 0) = -72 < 0$, so $(0, 0)$ is a saddle point.

For $(2, 2)$ we have $D(2, 2) = 72 > 0$, so we look at $f_{xx}(2, 2) = -6 < 0$, from which we can tell that $(2, 2)$ is a local maximum.

**Find and classify the critical points of $f(x, y) = e^x(x^2 - y^2)$.

$$\Rightarrow \quad f_x = e^x(x^2 + 2x - y^2) = 0, \quad f_y = -2ye^x = 0$$

From the second equation we immediately get $y = 0$. Plugging that into the second gives $0 = x^2 + 2x = x(x + 2)$, so $x = 0$ or $x = -2$. Hence our critical points are $(0, 0)$ and $(-2, 0)$.

$$f_{xx} = e^x(x^2 + 4x - y^2 + 2), \quad f_{yy} = -2e^x, \quad f_{xy} = -2ye^x$$

$$\Rightarrow \quad D(x, y) = -2e^{2x}(x^2 + 4x - y^2 + 2) - 4y^2e^{2x} = -2e^{2x}(x^2 + 4x + y^2 + 2).$$

Then for the critical point $(0, 0)$ we have $D(0, 0) = -4 < 0$, so $(0, 0)$ is a saddle point.

For $(-2, 0)$ we have $D(-2, 0) = -2e^{-4}(4 - 8 + 0 + 2) = 4e^{-4} > 0$, so we look at $f_{xx}(-2, 0) = e^{-2} \cdot (-2) > 0$, which tells us that $(2, 2)$ is a local maximum.

**Find and classify the critical points of $f(x, y) = \frac{1}{x} + xy + \frac{1}{y}$.

$$\Rightarrow \quad f_x = -\frac{1}{x^2} + y = 0, \quad f_y = x - \frac{1}{y^2} = 0$$

$$\Rightarrow \quad y = \frac{1}{x^2}, \quad x = \frac{1}{y^2}.$$ 

Plugging the first equation into the second gives

$$x = \frac{1}{(1/x^2)^2} = x^4 \quad \Rightarrow \quad 1 = x^3 \quad \Rightarrow \quad x = 1 \quad \Rightarrow \quad y = \frac{1}{1^2} = 1.$$ 

Note that we divided by $x$ when we went from $x = x^4$ to $1 = x^3$, which would normally mean that $x = 0$ can be a solution, but here $x = 0$ is not in the domain of $f$ because it has $\frac{1}{x}$ in it. So the only critical point is $(1, 1)$.

$$f_{xx} = \frac{2}{x^3}, \quad f_{yy} = \frac{2}{y^3}, \quad f_{xy} = 1$$

$$\Rightarrow \quad D(x, y) = \frac{4}{x^3y^3} - 1.$$ 

Then $D(1, 1) = 3 > 0$, and $f_x(1, 1) = 2 > 0$, so $(1, 1)$ is a local minimum.
A constrained optimization problem (in 2 variables) consists of an objective function \( f(x,y) \) and a constraint \( g(x,y) = 0 \).

The goal is to optimize the objective function subject to the constraint; or in other words, among the points \((a,b)\) satisfying \( g(a,b) = 0 \), find the ones for which \( f(a,b) \) is the largest or smallest.

This kind of situation is quite common. Here are two different kinds of examples.

- For instance, suppose you want to know what rectangle with perimeter 4 has the largest area. If you call the sides of the rectangle \( x \) and \( y \), then the area is \( xy \) and the perimeter is \( 2x + 2y = 4 \), so you have to solve the constrained optimization problem

  \[
  f(x,y) = xy, \quad g(x,y) = 2x + 2y - 4 = 0.
  \]

  In this case, you can probably guess that the answer is a square with sides \( x = y = 1 \).

- An example from economics is optimizing a production function subject to a budget: if \( x^{1/2}y^{1/2} \) is the number of units produced using \( x \) units of labor and \( y \) units of capital, and \( 3x + 2y = 100 \), for what choice of labor and capital will production be highest? Here the constraint can be seen as a budget of 100 dollars, with labor costing 3 dollars per unit and capital 2 dollars per unit. Then you have to solve the constrained optimization problem:

  \[
  f(x,y) = x^{1/2}y^{1/2}, \quad g(x,y) = 3x + 2y - 100 = 0.
  \]

**Turning constrained optimization into just optimization**

Before looking at the general method for solving constrained optimization problems, first we’ll look at a trick that works for easy examples.

- Let’s look at the rectangle example above. From the constraint \( 2x + 2y = 4 \), we can isolate \( y = 2 - x \), and plug that into the objective function \( xy \) to get a new 1-variable function

  \[
  h(x) = 2x - x^2.
  \]

  That one is easy to optimize: set \( h'(x) = 2 - 2x = 0 \), so \( x = 1 \). Then from the constraint we get \( y = 1 \), and that tells us that the square with sides \( x = y = 1 \) has the largest area.

- For the production function example, doing the same thing gives

  \[
  3x + 2y = 100 \quad \Rightarrow \quad y = 50 - \frac{3}{2}x \quad \Rightarrow \quad h(x) = f(x, 50 - \frac{3}{2}x) = \sqrt{50x - \frac{3}{2}x^2}
  \]

  \[
  \Rightarrow \quad h'(x) = \frac{50 - 3x}{2\sqrt{50x - \frac{3}{2}x^2}} = 0 \quad \Rightarrow \quad 50 - 3x = 0 \quad \Rightarrow \quad x = \frac{50}{3}, \quad y = 50 - \frac{3}{2}\cdot \frac{50}{3} = 25.
  \]

The first example was quite easy to do this way, the second was more work. For more complicated functions, this method can get too complicated. What’s more, for many constraints, it is impossible: take for example \( g(x, y) = y^3 + xy + x^5 = 0 \); then we wouldn’t be able to isolate \( y \) (or \( x \)) at all, and we simply can’t use this approach.
The method of Lagrange Multipliers

To optimize \( f(x, y) \) with the constraint \( g(x, y) = 0 \), solve the 3 equations in 3 unknowns

\[
\nabla f = \lambda \nabla g, \quad g = 0.
\]

Plug all the solutions \((a, b)\) into \( f \), and then compare to find the largest and smallest.
If there is only one solution \((a, b)\), find any other point \((c, d)\) that satisfies \( g = 0 \), and compare \( f(a, b) \) to \( f(c, d) \) to see if \((a, b)\) is a maximum or a minimum.

More specifically, the 3 equations in 3 unknowns are

\[
\begin{align*}
 f_x(x, y) &= \lambda \cdot g_x(x, y), \\
 f_y(x, y) &= \lambda \cdot g_y(x, y), \\
 g(x, y) &= 0.
\end{align*}
\]

To solve these equations, first isolate \( \lambda \) from one equation and then plug it into the other. That leaves you with 2 equations in the 2 unknowns \( x \) and \( y \). Then try to isolate \( x \) or \( y \) and plug it into the other equation.

Note that \( \lambda \) (the multiplier) is what is called a dummy variable: it’s there to make the equations possible, but we do not care what its value is; we only need the \( x \) and \( y \) values of a solution.

Explanation

- **Explanation by pictures:** This won’t come off very well in these typed notes, but I can still describe the graphical explanation. See Figures 12.95 and 12.98 in the book for appropriate pictures.

Think of the constraint \( g(x, y) = 0 \) as a curve in the \( xy \)-plane above which you walk on the surface \( z = f(x, y) \). The main observation behind Lagrange multipliers is this: at a maximum \( f(a, b) = M \), the curve \( g = 0 \) is tangent to the level curve \( f(x, y) = M \), since it reaches that height, but no higher level curve – that’s what it means to be a maximum.

Then note that \( \nabla f(a, b) \) is a vector orthogonal to the level curve \( f(x, y) = M \), and similarly \( \nabla g(a, b) \) is orthogonal to \( g(x, y) = 0 \) (because you can also think of it as a level curve of \( g \), at height 0).

So at the maximum \((a, b)\) the two gradients are orthogonal to two curves that are tangent, which means that these vectors must be parallel, ie one is a multiple of the other: \( \nabla f = \lambda \nabla g \).

- **Explanation by equations:** Now let’s write the curve \( g(x, y) = 0 \) as a parametrized curve \((x(t), y(t))\), and suppose we’re walking over the surface \( z = f(x, y) \) above this parametrized curve. As we pass over a maximum, we must be switching from walking uphill \((dz \, dt > 0)\) to walking downhill \((dz \, dt < 0)\), so at the maximum itself, we should have \( \frac{dz}{dt} = 0 \). But as we’ve learned, we can get a formula for \( \frac{dz}{dt} \) using the Chain Rule:

\[
0 = \frac{dz}{dt} = f_x \cdot x' + f_y \cdot y' \quad \Rightarrow \quad \frac{f_x}{f_y} = \frac{-y'}{x'}.
\]

We can get a something similar by differentiating \( g(x(t), y(t)) = 0 \) on both sides with respect to \( t \):

\[
0 = \frac{d}{dt} g(x(t), y(t)) = g_x \cdot x' + g_y \cdot y' \quad \Rightarrow \quad \frac{g_x}{g_y} = \frac{-y'}{x'}.
\]

Comparing these, we get

\[
\frac{f_x}{f_y} = \frac{-y'}{x'} = \frac{g_x}{g_y} \quad \Rightarrow \quad \frac{f_x}{g_x} = \frac{f_y}{g_y}.
\]
So at a maximum (or minimum), $f_x$ and $f_y$ have the same value, which we can call $\lambda$. But then

$$f_x = \lambda g_x, \quad f_y = \lambda g_y \quad \Rightarrow \quad \nabla f = \lambda \nabla g.$$ 

**Examples**

- Let’s reconsider the problem of optimizing the area of a rectangle with perimeter 4:

  $$f(x, y) = xy, \quad g(x, y) = 2x + 2y - 4 = 0.$$ 

  We have $\nabla f = \langle y, x \rangle$ and $\nabla g = \langle 2, 2 \rangle$, so our 3 equations are

  $$y = 2\lambda, \quad x = 2\lambda, \quad 2x + 2y = 4.$$ 

  The first equation gives $\lambda = y/2$, and plugging that into the second equation gives $x = y$. Plugging that into the third equation gives $2x + 2y = 4$, so $x = 1$ and then also $y = 1$. To see whether it is a maximum or a minimum, we need another point satisfying $2x + 2y = 4$, like $x = 0$, $y = 2$. Then we compare $f(1, 1) = 1$ and $f(0, 2) = 0$, which tells us that $x = 1$, $y = 1$ is a maximum.

- Now let’s do this for the production function with budget from above:

  $$f(x, y) = x^{1/2}y^{1/2}, \quad g(x, y) = 3x + 2y - 100 = 0.$$ 

  We have $\nabla f = \langle \frac{\sqrt{y}}{2\sqrt{x}}, \frac{\sqrt{x}}{2\sqrt{y}} \rangle$ and $\nabla g = \langle 3, 2 \rangle$, so our 3 equations are

  $$\frac{\sqrt{y}}{2\sqrt{x}} = 3\lambda, \quad \frac{\sqrt{x}}{2\sqrt{y}} = 2\lambda, \quad 3x + 2y = 100.$$ 

  Then isolating $\lambda$ from the first equation and plugging that into the second gives:

  $$\lambda = \frac{\sqrt{y}}{6\sqrt{x}} \quad \Rightarrow \quad \frac{\sqrt{x}}{2\sqrt{y}} = 2 \cdot \frac{\sqrt{y}}{6\sqrt{x}} \quad \Rightarrow \quad 3x = 2y.$$ 

  Plugging that into the constraint gives $100 = 3x + 3x = 6x$, so $x = 100/6 = 50/3$, and $y = 50/2 = 25$. Comparing $f(50/3, 25) = \sqrt{50/3}\sqrt{25} = 25\sqrt{2}/3$ and for instance $f(0, 50) = 0$, we see that we have a maximum.

- Find the extreme values of the function $f(x, y) = x^2 + 2y^2$ on the circle $x^2 + y^2 = 1$.

  We have $\nabla f = \langle 2x, 4y \rangle$, and with $g(x, y) = x^2 + y^2 - 1$ we have $\nabla g = \langle 2x, 2y \rangle$, so our 3 equations are

  $$2x = \lambda \cdot 2x, \quad 4y = \lambda \cdot 2y, \quad x^2 + y^2 = 1.$$ 

  From the first equation we have $\lambda = 1$ or $x = 0$. If $\lambda = 1$, then the second equation gives that $y = 0$, and the third gives $x = \pm 1$. If we have $x = 0$, then the third equation gives $y = \pm 1$. So the possible extreme values are at the points

  $$(1, 0), \quad (-1, 0), \quad (0, 1), \quad (0, -1).$$ 

  We evaluate $f$ at each of these:

  $$f(1, 0) = 1, \quad f(-1, 0) = 1, \quad f(0, 1) = 2, \quad f(0, -1) = 2.$$
So we have two minima at \((1, 0)\) and \((-1, 0)\), and two maxima at \((0, 1)\) and \((0, -1)\).

- **Find the extreme values of the function** \(f(x, y) = x^2 + xy + y^2 - 2x - 5y\) **subject to the constraint** \(x - y = 1\).

  We have \(\nabla f = \langle 2x + y - 2, x + 2y - 5 \rangle\), and with \(g(x, y) = x - y - 1\) we have \(\nabla g = \langle 1, -1 \rangle\), so our 3 equations are

  \[
  2x + y - 2 = \lambda, \quad x + 2y - 5 = -\lambda, \quad x - y = 1.
  \]

  Eliminating \(\lambda\) from the first two equations, we get

  \[
  2x + y - 2 = \lambda = -x - 2y + 5 \quad \Rightarrow \quad 3x + 3y = 7 \quad \Rightarrow \quad y = \frac{7}{3} - x,
  \]

  and plugging this into \(x - y = 1\) gives

  \[
  x - \frac{7}{3} + x = 1 \quad \Rightarrow \quad 2x = 1 + \frac{7}{3} = \frac{10}{3} \quad \Rightarrow \quad x = \frac{5}{3}, \quad y = \frac{7}{3} - \frac{5}{3} = \frac{2}{3}.
  \]

  Evaluating at this point gives

  \[
  f\left(\frac{5}{3}, \frac{2}{3}\right) = \frac{25}{9} + \frac{10}{9} + \frac{4}{9} - \frac{30}{9} - \frac{30}{9} = -\frac{21}{9} = -\frac{7}{3},
  \]

  and evaluating at some other point that satisfies \(x - y = 1\), like \((1, 0)\), gives \(f(1, 0) = -1\), so we conclude that \((5/3, 2/3)\) is a minimum for \(f\) subject to \(x - y = 1\), and this is its only such extreme value.

- **Suppose that** \(x\) **units of labor and** \(y\) **units of capital can produce** \(f(x, y) = 60x^{3/4}y^{1/4}\) **units of a certain product, and that a unit of labor costs** $100, **while a unit of capital costs** $200. **Assume that** $30,000 **is available to spend on production. How many units of labor and how many units of capital should be utilized to maximize production?**

  The constraint equation is

  \[g(x, y) = 100x + 200y - 30000 = 0.\]

  Then

  \[
  \nabla f = \left\langle 45 \frac{y^{1/4}}{x^{1/4}}, 15 \frac{x^{3/4}}{y^{3/4}} \right\rangle, \quad \nabla g = \langle 100, 200 \rangle,
  \]

  and the 3 equations are

  \[
  45 \frac{y^{1/4}}{x^{1/4}} = 100\lambda, \quad 15 \frac{x^{3/4}}{y^{3/4}} = 200\lambda, \quad 100x + 200y = 30000.
  \]

  Eliminating \(\lambda\) gives

  \[
  \frac{45}{100} \frac{y^{1/4}}{x^{1/4}} = \lambda = \frac{15}{200} \frac{x^{3/4}}{y^{3/4}} \quad \Rightarrow \quad y = \frac{100}{45} \cdot \frac{15}{200} \cdot x = \frac{1}{6} x.
  \]

  Plugging that into the third equation gives

  \[
  30000 = 100x + \frac{200}{6} x = \frac{800}{6} x \quad \Rightarrow \quad x = \frac{6 \cdot 30000}{800} = 225 \quad \Rightarrow \quad y = \frac{1}{6} \cdot 225 = \frac{75}{2}.
  \]

  So there is an extreme value at \((225, 75/2)\). However, evaluating \(f(225, 75/2)\) would be hard without a calculator. Fortunately, we don’t have to, because another point satisfying the constraint is \((300, 0)\), and there \(f(300, 0) = 0\), while clearly \(f(225, 75/2) > 0\). So we can conclude that production is maximized by 225 units of labor and 37.5 units of capital.

  The last trick usually works in these production function problems: since for 0 capital or 0 labor, production must be 0, and otherwise production is positive, we know that a unique extreme value must be a maximum.