118: FUNDAMENTAL PRINCIPLES OF CALCULUS, FALL 2018

STEVEN HEILMAN

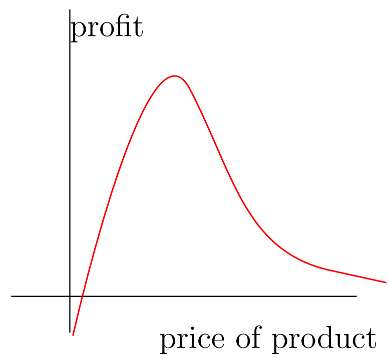
Contents

Date: September 20, 2018.

1. Introduction

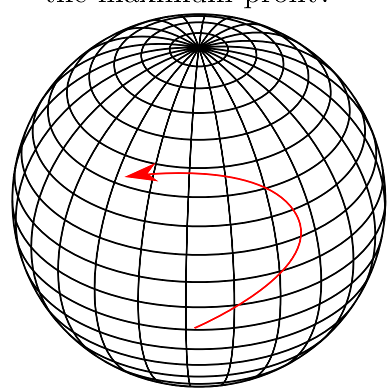
The ability of Calculus to describe the world is one of the great triumphs of mathematics. Below, we will briefly describe some of these successful applications. Some of these applications will be discussed in the second half of this course, and others may be found in your future endeavors.

For example, calculus is very closely related to **probability**, which itself has applications to **statistics** and **algorithms**. For example, the first generation of Google's search technology came from ideas from probability theory. In **economics**, optimization is often used to e.g. maximize profit margins, or to find optimal strategies in **game theory**. Also, the ideas of single variable calculus are developed and generalized within financial mathematics to e.g. **stochastic calculus**. Signal processing and **Fourier analysis** provide some nice applications within many areas of science. For example, our cell phones use Fourier analysis to compress voice signals.



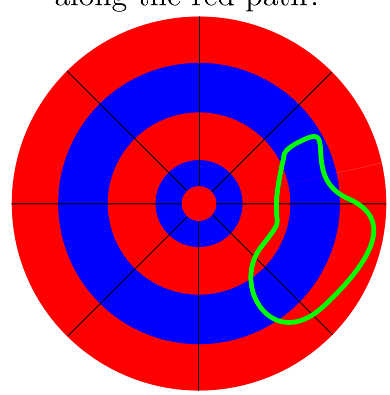
How can we find the maximum profit?

In **physics**, many models of the real world use solutions of **differential equations**. These equations involve the slopes and shape of functions, and their solutions describe the behavior of many physical systems. For example, the famous Navier-Stokes equations of fluid dynamics are expressed in this language. Solutions of the Navier-Stokes equations show us how water behaves, though these equations really just state Newton's second law. Also, **Einstein's Theory of General Relativity** uses a version of Calculus, though geometry is needed here as well.



What forces are felt along the red path?

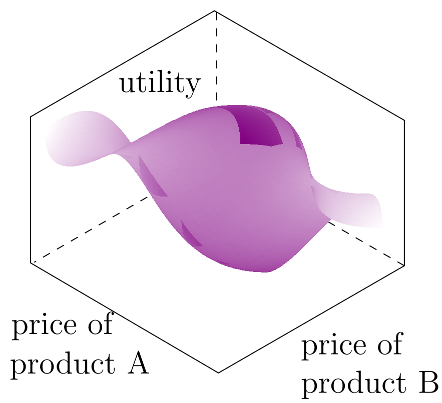
In **mathematics** itself, the fundamental concepts of Calculus reappear in many places, some of which have already been described above. Also, Calculus serves as the foundation of **probability**, which itself serves as the foundation of **statistics**. For example, to prove that a large number of numerical data samples have the distribution of a bell curve, one can use tools from Calculus. As another example, we can repeat Calculus for a single real variable by using a single complex variable, and we get the beautiful subject of **Complex analysis**.



How likely will a dart land inside the green region?

For a complete understanding of **biology**, you need to understand Calculus. For example, suppose someone is given an intravenous drug which is administered at a certain rate. If we know the volume of fluid in the body, the concentration of the administered drug, and the rate of flow of the drug into the body, then we can use Calculus to model the concentration of drug in the person's body. These calculations are done using differential equations. If one wants to fully understand how MRIs and CT-scans work, one needs **multivariable calculus**. Many of the concepts in multivariable calculus originate in single variable calculus, though there are some new things that are needed.

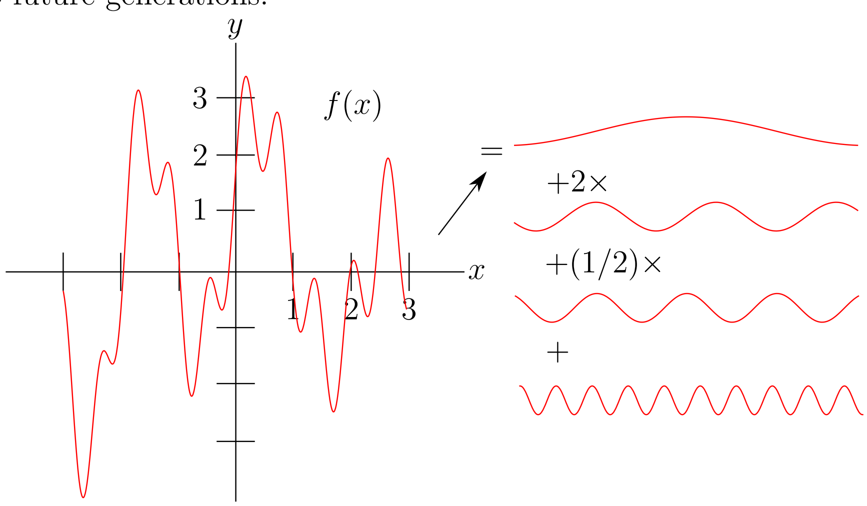
Using differential equations, one can derive some of the equations for chemical concentrations that were used in our high school **chemistry** classes. For example, we could ask: after chemicals interact for along time, what are their concentrations? Also, the ideas of Calculus are used in more advanced chemistry subjects, including **quantum mechanics**.



How to find the maximum utility as a function of two variables?

Even though **computer science** often deals with discrete problems, many of the ideas of Calculus arise in computer science, and sometimes these ideas arise in unexpected ways. For example, the methods used to encode music onto CDs and MP3s use Calculus, with some additional tools from **Fourier analysis**. These same tools compress image and video data for JPEGs and MPEGs, respectively.

The acts of thinking rigorously and using **logic** in our reasoning should become common in this course. We want to transmit one of the great intellectual achievements of humanity, just as we pass down great literature, art and philosophy to future generations.



A sound wave split into its frequency components.

1.0.1. *A Brief History of Calculus.* The rudiments of Calculus can be traced through many ancient cultures including those of Greece, China and India. Calculus in its modern form is generally attributed to Newton and Leibniz in the late 1600s. Newton was mainly motivated by applying Calculus to physics. However, Leibniz invented most of the notation we use today. Even though this topic is now taught in high school and college, it is still over 300 years old. In this course, perhaps we will try to give you glimpses of what happened in the ensuing centuries.

1.0.2. *The Calculus Paradigm.* Mathematics before Calculus usually involves algebra, trigonometry and some planar geometry. The concepts in Calculus are very different from concepts that are learned before, so the following "paradoxes" are meant to help in understanding the new concepts. These paradoxes are known as Zeno's Paradoxes.

Paradox 1.1. In a footrace, suppose a slower runner is in front of a quicker runner. When the quicker runner reaches any point in the race, the slower runner was already at that point in the past. Therefore, the quicker runner can never overtake the slower runner.

Paradox 1.2. Suppose I want to walk through a doorway, and I am standing a meter away from the door. At some point I am at a half meter away from the door, then at another point I am at a quarter of a meter away, then at another point I am at an eighth a meter away, and so on. Therefore, I can never make it through the doorway.

Paradox 1.3. Suppose I shoot at arrow at a target. At any given moment, the arrow occupies a fixed position in space. However, in order for an object to move, it cannot sit in one place. So, the arrow must have no motion at all. The arrow is motionless.

The first two paradoxes are somewhat similar. In the first paradox, we know from empirical observation that a quicker runner can overtake a slower runner. And in order to resolve the paradox, we note that the total time that the quicker runner remains behind the slower runner is finite. Similarly, in the second paradox, I know that it only takes a finite amount of time to pass through a door. Paradox 1.2 seems to occur since I am subdividing the one meter that I travel into an infinite number of smaller steps. Zeno seems to object, saying that an infinite number of subdivisions cannot occur.

In mathematical terms, Zeno objects to the assertion $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots=1$, because there does not appear to be a rigorous way to think about an infinite sum of numbers. By using a limit, we can actually speak rigorously about an infinite sum of numbers, thereby resolving the paradox.

The third paradox is a bit different from the other two. Zeno is really questioning the meaning of an instant of time. How can we rigorously discuss the instantaneous speed of an object? Below, we will use limits to define derivatives, and these derivatives give a rigorous meaning to instantaneous velocity, thereby resolving the paradox.

1.1. **The Notion of a Limit.**

Definition 1.4 (Intuitive Definition of a Limit). Let $f\colon \mathbb{R}\to\mathbb{R}$, and let $x,a,L\in\mathbb{R}$. We say that f has limit L as x approaches a and write

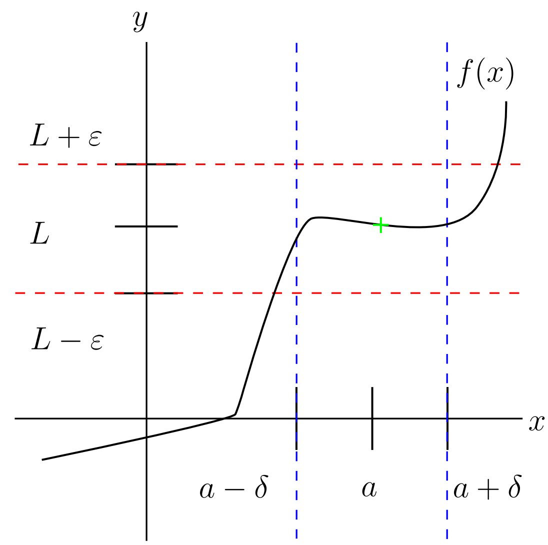
$$\lim_{x\to a} f(x) = L,$$

if $f(x)$ gets closer and closer to L as x gets closer and closer to a, with $x\neq a$.

Definition 1.5 (Formal Definition of a Limit). Fix $a\in\mathbb{R}$ and let x be a variable. We write $\lim_{x\to a} f(x)=L$ if the following occurs.

$$\begin{aligned}&\text{For all } \varepsilon>0, \text{ there exists } \delta=\delta(\varepsilon)>0, \text{ such that:}\\ &\text{if } 0<|x-a|<\delta, \text{ then } |f(x)-L|<\varepsilon.\end{aligned} \tag{1}$$

Put another way, no matter how small I want $|f(x) - L|$ to be, I can always choose a small region around a that is so small such that $f(x)$ is really close to L in this region.



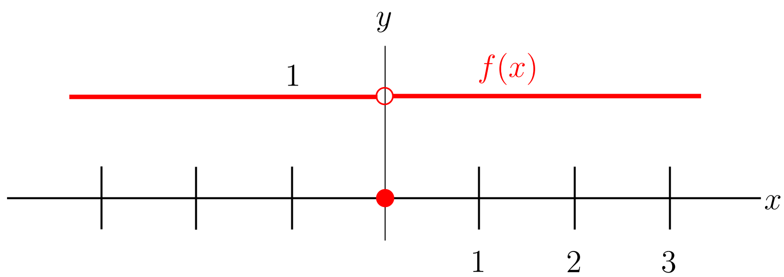
Example 1.6. Let $f(x) = x$ and let $a = 1$. Then $\lim_{x \to a} f(x) = a = 1$.

Example 1.7. Let $f(x) = x^2$ and let $a = 2$. Then $\lim_{x \to a} f(x) = a^2 = 4$.

Remark 1.8. The limit $\lim_{x \to a} f(x)$ does not depend on the value of f at a.

Example 1.9. Let $f(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$

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



Definition 1.10 (Intuitive Definition of One-Sided Limits). Let $f \colon \mathbb{R} \to \mathbb{R}$, and let $x, a, L \in \mathbb{R}$. We say that f has limit L as x approaches a from the left and write

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

if $f(x)$ gets closer and closer to L as x gets closer and closer to a, with $x < a$.

We say that f has limit L as x approaches a from the right and write

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

if $f(x)$ gets closer and closer to L as x gets closer and closer to a, with $x > a$.

Remark 1.11. $\lim_{x \to a} f(x)$ exists if and only if both one-sided limits $\lim_{x \to a^-} f(x)$ and $\lim_{x \to a^-} f(x)$ exist and are equal.

The limit may at first look a bit silly[1], since for a polynomial, we can just plug in the function value and the result agrees with the limit. So why are we defining a limit anyway? First of all, we need limits in order to define tangent lines to functions (derivatives), and derivatives are one of the extremely important concepts in Calculus. Second of all, there are some subtleties to the definition that may not yet be apparent.

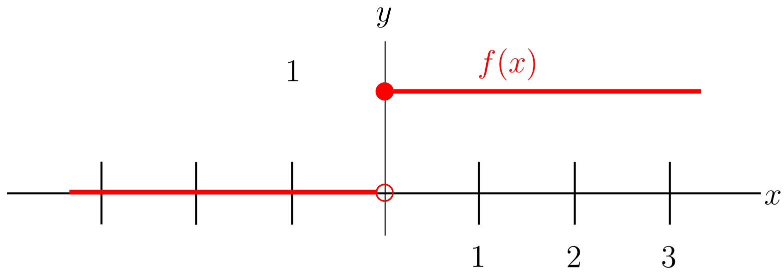
For example, the limit of a function may not always exist. This issue is important because we will see that the derivatives of some functions may not exist. So, there may be no reasonable way to talk about the instantaneous speed of certain trajectories.

[1] Historically, the definition (1) was invented over a century after Newton and Leibniz invented Calculus. The approach of Newton and Leibniz was not quite rigorous by modern standards.

Example 1.12 (**A jump discontinuity**)**.** Define $H\colon \mathbb{R} \to \mathbb{R}$ (the Heaviside function) by the formula

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

We claim that $\lim_{x\to 0^-} H(x) = 0$ and $\lim_{x\to 0^+} H(x) = 1$. Therefore, $\lim_{x\to 0} H(x)$ does not exist.



To see this, note that if $x < 0$, then $H(x) = 0$. That is, $H(x)$ is always close to 0 whenever $x < 0$. We conclude that $\lim_{x\to 0^-} H(x) = 0$.

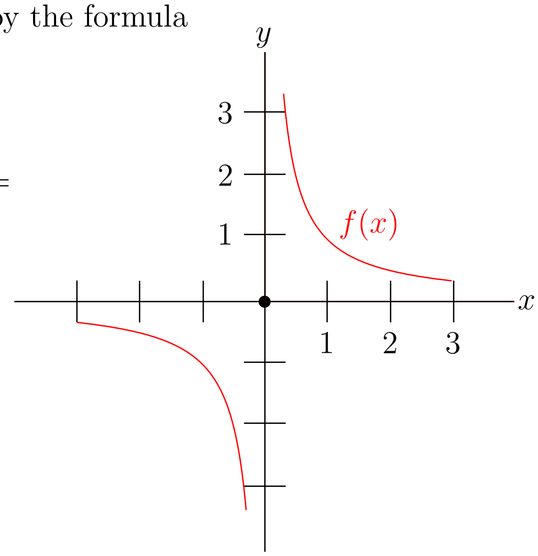
Similarly, if $x > 0$, then $H(x) = 1$. That is, $H(x)$ is always close to 1 whenever $x > 0$. Therefore, $\lim_{x\to 0^+} H(x) = 1$.

Both one-sided limits must be equal in order for $\lim_{x\to 0} H(x)$ to exist. Since we know that $\lim_{x\to 0^+} H(x) = 1 \neq 0 = \lim_{x\to 0^-} H(x)$ we conclude that $\lim_{x\to 0} H(x)$ does not exist.

Remark 1.13. If a function f becomes arbitrarily large as $x \to a$, the limit $\lim_{x\to a} f(x)$ does not exist, but we still write $\lim_{x\to a} f(x) = \infty$. If a function f has $|f(x)|$ arbitrarily large and $f(x) < 0$ as $x \to a$, the limit $\lim_{x\to a} f(x)$ does not exist, but we still write $\lim_{x\to a} f(x) = -\infty$.

Example 1.14 (**A singularity**)**.** Define $f\colon \mathbb{R} \to \mathbb{R}$ by the formula

$$f(x) = \begin{cases} 1/x, & x \neq 0 \\ 0, & x = 0. \end{cases}$$



We claim that $\lim_{x\to 0^+} f(x) = \infty$ and $\lim_{x\to 0^-} f(x) = -\infty$. Therefore, $\lim_{x\to 0} f(x)$ does not exist.

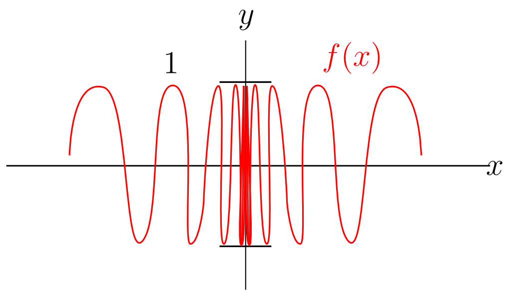
Note that if $x > 0$, then $f(x) > 0$. And, for example, if $x = 1/n$ where n is a positive integer, then $f(x) = n$ becomes arbitrarily large as n becomes large. Therefore, $\lim_{x\to 0^+} f(x) = \infty$. Since $\lim_{x\to 0^+} f(x)$ does not exist, we already know that $\lim_{x\to 0} f(x)$ does not exist.

Similarly, if $x < 0$, then $f(x) < 0$. And if $x = -1/n$ where n is a positive integer, then $f(x) = -n$ becomes negative and large as n becomes large. Therefore, $\lim_{x\to 0^-} f(x) = -\infty$.

Example 1.15 (**Infinite oscillation**)**.** Define $f\colon \mathbb{R} \to \mathbb{R}$ so that

$$f(x) = \begin{cases} \cos(1/x), & x \neq 0 \\ 0, & x = 0. \end{cases}$$

We claim that $\lim_{x\to 0} f(x)$ does not exist. To see this, let n be a positive integer and consider the points $1/(2\pi n)$ and $1/(2\pi n+\pi)$. If n is large, both of these points can become arbitrarily close to zero. However, $f(1/(2\pi n)) = \cos(2\pi n) = 1$, while $f(1/(2\pi n+\pi)) = \cos(2\pi n+\pi) = -1$. So, f never becomes close to any value L as $x \to 0$. So, $\lim_{x\to 0} f(x)$ does not exist.



1.2. **Calculating Limits.** In order to manipulate limits, we need to be able to apply some simple operations to them. The following statements summarize some ways that we can manipulate limits.

Proposition 1.16 (Limit Laws). *Let $f,g\colon \mathbb{R}\to\mathbb{R}$. Let $a,c\in\mathbb{R}$. Assume that $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} g(x)$ exists. Then*

(i) $\lim_{x\to a}(f(x)+g(x)) = (\lim_{x\to a} f(x)) + (\lim_{x\to a} g(x))$.
(i') $\lim_{x\to a}(f(x)-g(x)) = (\lim_{x\to a} f(x)) - (\lim_{x\to a} g(x))$.
(ii) $\lim_{x\to a}[cf(x)] = c(\lim_{x\to a} f(x))$.
(iii) $\lim_{x\to a}[(f(x))(g(x))] = (\lim_{x\to a} f(x))(\lim_{x\to a} g(x))$.
(iv) *If* $\lim_{x\to a} g(x) \neq 0$, *then* $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$.

Example 1.17. We know that $\lim_{x\to 2} x = 2$, so we know that $\lim_{x\to 2} x^2 = (\lim_{x\to 2} x)^2 = 2^2 = 4$, using Limit Law (*iii*).

Example 1.18. Since we know that $\lim_{x\to 2} x^2 = 4$ and $\lim_{x\to 2} x = 2$, we then know that $\lim_{x\to 2} x^3 = (\lim_{x\to 2} x^2)(\lim_{x\to 2} x) = 2^2 2 = 8$, using Limit Law (*iii*).

Example 1.19. Since $\lim_{x\to 0}(x+1) = 1$, and $\lim_{x\to 0}(x^2+3) = 3$, we have

$$\lim_{x\to 0}\frac{x+1}{x^2+3} = \frac{1}{3}.$$

Example 1.20. We know that $\lim_{x\to 0} x = 0$ and $\lim_{x\to 0} x^{-1}$ does not exist. So, the following equality doesn't make any sense:

$$1 = \lim_{x\to 0} 1 = \lim_{x\to 0}(x/x) \stackrel{?}{=} (\lim_{x\to 0} x)(\lim_{x\to 0} x^{-1}).$$

However, the assumptions of the Limit Laws are not satisfied, so we have not found a contradiction within the Limit Laws.

1.3. **Continuity.**

Definition 1.21 (Intuitive Definition of Continuity). Let $f\colon\mathbb{R}\to\mathbb{R}$ and let $a\in\mathbb{R}$. We say that f is **continuous** if we can draw f with a pencil, without lifting the pencil off of the paper.

Alternatively, f is continuous at a if f does not "jump around too much" around a.

Definition 1.22 (Formal Definition of Continuity). Let $f\colon\mathbb{R}\to\mathbb{R}$ let $a\in\mathbb{R}$. We say that f is **continuous at** a if the following three conditions are satisfied

(i) $\lim_{x\to a} f(x)$ exists
(ii) a is in the domain of f
(iii) $\lim_{x\to a} f(x) = f(a)$

We say that a function f is **continuous** if f is continuous on all points of its domain. If f is not continuous at a, we say that f is **discontinuous** at a.

Example 1.23. Let $f(x) = x$. Then f is continuous, since $\lim_{x\to a} f(x) = a = f(a)$ for all points $a \in \mathbb{R}$.

Example 1.24 (Removable Discontinuity). Let

$$f(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $\lim_{x\to 0} f(x) = 0$, but $f(0) = 1$. That is $\lim_{x\to 0} f(x) \neq f(0)$. So f is discontinuous at 0. However, if we redefined f to be equal to 1 at 0, then f would be continuous at 0. For this reason, we say that f has a **removable discontinuity** at 0.

Example 1.25 (Jump Discontinuity)). Recall we defined $H\colon \mathbb{R} \to \mathbb{R}$ where

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases}$$

We showed that $\lim_{x\to 0} H(x)$ does not exist. So, $H(x)$ is discontinuous at $x = 0$.

Example 1.26 (A singularity). Define $f\colon \mathbb{R} \to \mathbb{R}$ by the formula

$$f(x) = \begin{cases} 1/x, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

We showed that $\lim_{x\to 0} f(x)$ does not exist. So, $f(x)$ is discontinuous at $x = 0$. However, f is actually continuous at any nonzero point, by the limit law for quotients. If $a \neq 0$, then

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

Example 1.27 (Infinite oscillation). Define $f\colon \mathbb{R} \to \mathbb{R}$ so that

$$f(x) = \begin{cases} \cos(1/x), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

We showed that $\lim_{x\to 0} f(x)$ does not exist. So, f is discontinuous at $x = 0$. We will show later on that f is continuous at any point $a \neq 0$.

Verifying directly that a given function is continuous can be tedious. Thankfully, we can often verify continuity of a function using the following rules, which are consequence of the Limit Laws, Proposition 1.16.

Proposition 1.28 (Laws of Continuity). *Let $f, g\colon \mathbb{R} \to \mathbb{R}$. Let $a, c \in \mathbb{R}$. Assume that f and g are both continuous at a. Then*

(i) *$f + g$ is continuous at a.*
(i') *$f - g$ is continuous at a*
(ii) *cf is continuous at a.*
(iii) *fg is continuous at a*
(iv) *If $g(a) \neq 0$, then f/g is continuous at a.*

Corollary 1.29. *Let p, q be polynomials. Then p is continuous on the real line, and p/q is continuous at all values a where $q(a) \neq 0$.*

Proof. Note that $p(x)$ is a sum of monomials $c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$. Each monomial is continuous, so p is continuous by the first law of continuity.

So, p, q are both continuous on all of $\mathbb{R}$. Then p/q is continuous at all values a where $q(a) \neq 0$, by the last law of continuity. □

Proposition 1.30.

- *If $b > 0$, then $f(x) = b^x$ is continuous on the real line.*
- *If $b > 0$, then $f(x) = \log_b x$ is continuous whenever $x > 0$.*
- *If n is a positive integer, then $f(x) = x^{1/n}$ is continuous on its domain.*

Example 1.31. $f(x) = 1/x$ is continuous when $x \neq 0$ by the fourth law of continuity.

Example 1.32. The function $f(x) = x^2 e^x$ is continuous on the real line, by the third law of continuity.

Theorem 1.33 (**A Composition of Continuous Functions is Continuous**). *Let f, g be functions. Define $F(x) = f(g(x))$. If g is continuous at $x = a$, and if f is continuous at $g(a)$, then $F(x) = f(g(x))$ is continuous at $x = a$.*

Example 1.34. Consider the function

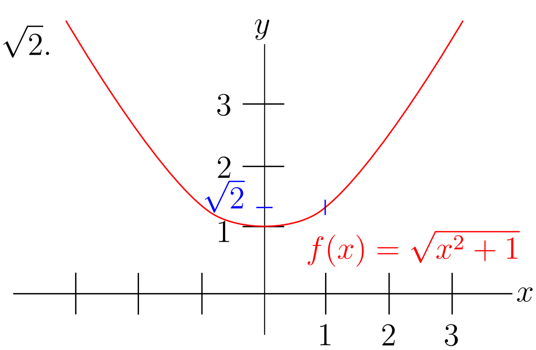
$$F(x) = \begin{cases} e^{1/x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Since $g(x) = 1/x$ is continuous for $x \neq 0$, and $f(y) = e^y$ is continuous for any y, we see that $F(x) = f(g(x))$ is continuous for any $x \neq 0$. Also, F is discontinuous at $x = 0$, since $\lim_{x\to 0} F(x)$ does not exist. More specifically, $\lim_{x\to 0^+} F(x) = \lim_{y\to\infty} e^y = \infty$ and $\lim_{x\to 0^-} F(x) = \lim_{y\to -\infty} e^y = 0$. Since the right and left limits at $x = 0$ do not agree, $\lim_{x\to 0} F(x)$ does not exist, so F is not continuous at $x = 0$.

Remark 1.35. If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, we can move limits inside or outside of f:

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

Example 1.36. $\lim\limits_{x\to 1} \sqrt{x^2+1} = \sqrt{\lim\limits_{x\to 1}(x^2+1)} = \sqrt{2}$.

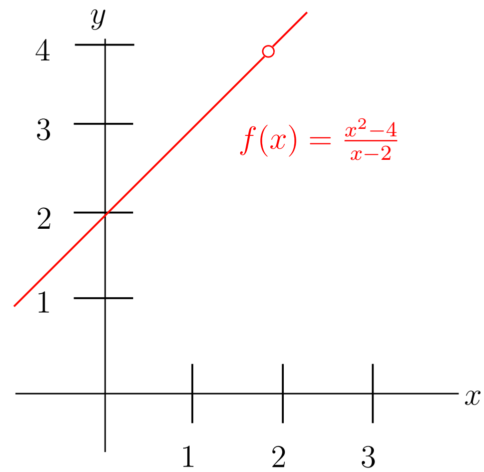


Example 1.37. $\lim_{x\to 2} e^{x^2-3}$
$= e^{\lim_{x\to 2}(x^2-3)} = e^1$.

Sometimes we need to do a bit of algebra before we can move the limits around or apply limits laws.

Example 1.38. $\lim\limits_{x\to 2} \dfrac{x^2-4}{x-2} = \lim\limits_{x\to 2} \dfrac{(x-2)(x+2)}{(x-2)}$
$= \lim\limits_{x\to 2}(x+2) = 4$.

Example 1.39. $\lim_{x\to 1} \frac{\sqrt{x}-1}{x-1} = \lim_{x\to 1} \frac{\sqrt{x}-1}{x-1}\frac{\sqrt{x}+1}{\sqrt{x}+1}$
$= \lim_{x\to 1} \frac{x-1}{(x-1)(\sqrt{x}+1)} = \lim_{x\to 1} \frac{1}{\sqrt{x}+1} = \frac{1}{2}.$

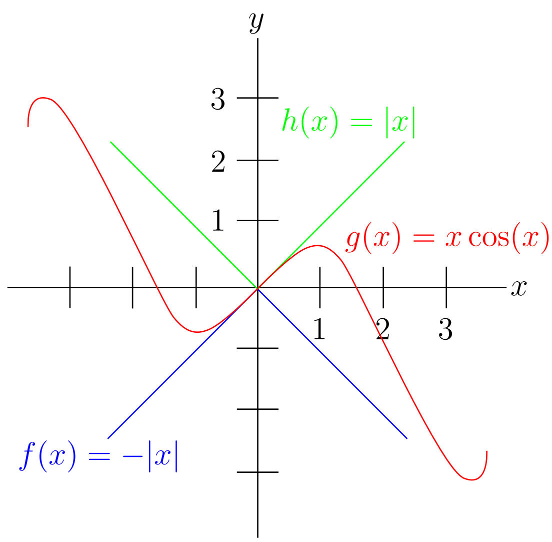


Example 1.40. $\lim_{x\to 0} \left(\frac{1}{4x} - \frac{1}{x(x+4)} \right)$
$= \lim_{x\to 0} \frac{x+4-4}{4x(x+4)} = \lim_{x\to 0} \frac{x}{4x(x+4)}$
$= \lim_{x\to 0} \frac{1}{4(x+4)} = \frac{1}{16}.$

1.4. **Limits at Infinity.**

Theorem 1.41 (**Squeeze Theorem**)**.** *Let $f, g, h\colon \mathbb{R} \to \mathbb{R}$. Suppose $f(x) \le g(x) \le h(x)$ and $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L$. Then $\lim_{x\to a} g(x)$ exists and $\lim_{x\to a} g(x) = L$.*

Example 1.42. We show that $\lim_{x\to 0} x\cos(x) = 0$. Since $-1 \le \cos(x) \le 1$, we have $-|x| \le x\cos x \le |x|$. Since $\lim_{x\to 0}(-|x|) = \lim_{x\to 0} |x| = 0$, we conclude that $\lim_{x\to 0} x\cos x = 0$.



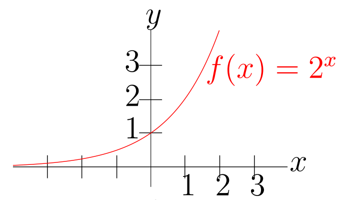
1.4.1. *Limits at Infinity.* So far we have only considered limits of the form $\lim_{x\to a} f(x)$ where a is a real number. We can also consider limits of the form $\lim_{x\to\infty} f(x)$ and $\lim_{x\to -\infty} f(x)$, using essentially the same rules as before. Alternatively, we could make the following definition, if the limit on the right exists:

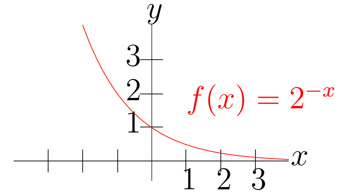
$$\lim_{x\to\infty} f(x) = \lim_{y\to 0^+} f(1/y).$$

Similarly, if the limit on the right exists, we can define

$$\lim_{x\to -\infty} f(x) = \lim_{y\to 0^+} f(-1/y).$$

We can think of $\lim_{t\to\infty} f(t)$ as describing where f eventually ends up, as time goes to infinity.

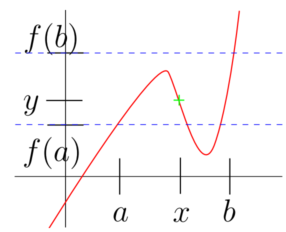




Example 1.43. $\lim_{x\to\infty} 1/x = 0$.

Example 1.44. $\lim_{x\to\infty} 2^x = \infty$, $\lim_{x\to\infty} 2^{-x} = 0$.

Example 1.45. $\displaystyle\lim_{x\to\infty} \frac{4x^2+2}{3x^2+x+3} = \lim_{x\to\infty} \frac{4x^2+2}{3x^2+x+3} \cdot \frac{x^{-2}}{x^{-2}}$
$$= \lim_{x\to\infty} \frac{4+\frac{2}{x^2}}{3+\frac{1}{x}+\frac{3}{x^2}} = \frac{\lim_{x\to\infty} 4+\frac{2}{x^2}}{\lim_{x\to\infty} 3+\frac{1}{x}+\frac{3}{x^2}} = \frac{4}{3}.$$



Example 1.46. More generally, if n is a positive integer and $d_n \neq 0$,
$$\lim_{x\to\infty} \frac{c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0}{d_n x^n + d_{n-1}x^{n-1} + \cdots + d_0}$$
$$= \lim_{x\to\infty} \frac{c_n + \frac{c_{n-1}}{x} + \cdots + \frac{c_0}{x^n}}{d_n + \frac{d_{n-1}}{x} + \cdots + \frac{d_0}{x^n}}$$
$$= \frac{\lim_{x\to\infty} c_n + \frac{c_{n-1}}{x} + \cdots + \frac{c_0}{x^n}}{\lim_{x\to\infty} d_n + \frac{d_{n-1}}{x} + \cdots + \frac{d_0}{x^n}} = \frac{c_n}{d_n}.$$

1.5. **Intermediate Value Theorem.** Continuous functions have many nice properties. One such property is expressed in the following Theorem.

Theorem 1.47 (Intermediate Value Theorem). *Let $a, b \in \mathbb{R}$, $a < b$. Let $f\colon [a,b] \to \mathbb{R}$ be continuous. Then f achieves every value between $f(a)$ and $f(b)$. That is, for every $y \in \mathbb{R}$ in between $f(a)$ and $f(b)$, there exists $x \in [a,b]$ such that $f(x) = y$.*

Example 1.48. For a real-world example, suppose $f\colon [0,10] \to \mathbb{R}$ and $f(t)$ is your position at time t as you walk along a straight path. Suppose $f(0) = 0$ and $f(10) = 2$. Then you must visit every point along the path $[0,2]$ at some time between time 0 and time 10.

Example 1.49. There is some nonzero number x such that $2\sin(x) = x$. To see this, let $g(x) = 2\sin(x) - x$. Then $g(\pi/2) = 2 - \pi/2 > 0$ and $g(\pi) = -\pi < 0$. By the Intermediate Value Theorem, there is some $x \in (\pi/2, \pi)$ such that $g(x) = 0$. At this point, we have $2\sin(x) = x$.

Example 1.50. Continuity is needed for the conclusion of the Intermediate Value Theorem to hold. For example, consider the jump discontinuity
$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$
Then $H(-1) = 0$, $H(1) = 1$, but there does not exist some $x \in (-1,1)$ with $H(x) = 1/2$.

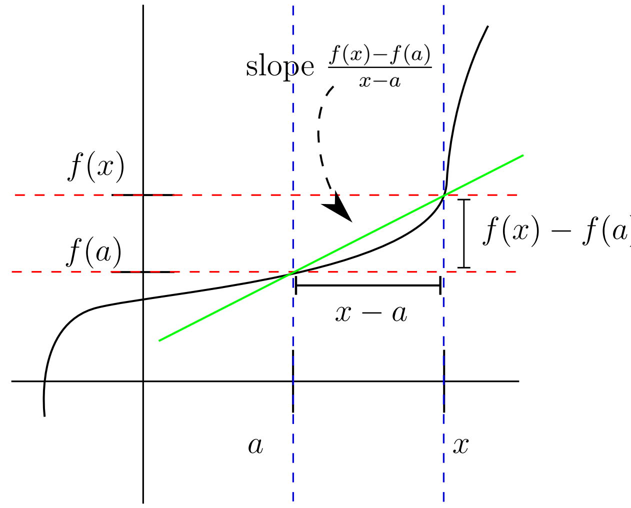
2. The Derivative

2.1. **Definition of the Derivative.** The derivative is one of the central concepts in calculus. It is also one of the most useful. For a function $f\colon \mathbb{R} \to \mathbb{R}$, where $f(t)$ denotes the vertical position of a falling object at time t, the derivative of t is the velocity of the object at time t.

For a general function $f\colon \mathbb{R} \to \mathbb{R}$, the derivative of f is the rate of change of f. We denote the derivative of f at x by $f'(x)$.

We now begin the construction of the derivative. Let $f\colon \mathbb{R} \to \mathbb{R}$. Let $a, x \in \mathbb{R}$ with $a < x$. Consider the points $(a, f(a))$ and $(x, f(x))$. Recall that the line passing through these points has a slope equal to the **rise over run** of the points. That is, the slope of the line passing through these two points is

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



The numerator is the change in the value of f from a to x, and the denominator is the change in the domain of f from a to x. So, $(f(x)-f(a))/(x-a)$ is essentially the rate of change of f at a. The line passing through the points $(a, f(a))$ and $(x, f(x))$ is also known as the **secant line**.

The slope of the secant line is a decent approximation to the slope of the tangent line to f at a. It turns out that as $x \to a$, the slope of the secant line will often go to the slope of the tangent line. That is, suppose as x goes to a, the quantity $\frac{f(x)-f(a)}{x-a}$ goes to some number L. We can then think of L as the infinitesimal rate of change of f. To see this approximation procedure in action, see the following JAVA applet: secant approximation.

This discussion leads the following definition of a derivative

Definition 2.1 (**The Derivative at a Point**)**.** Let $a, b, x \in \mathbb{R}$, $f\colon \mathbb{R} \to \mathbb{R}$. We say that f is **differentiable** at a if the following limit exists.

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

If $f'(a)$ exists, we call $f'(a)$ the **derivative** of f at a. We say that f is **differentiable** if f is differentiable on its domain.

Example 2.2. Let $f(x) = c$ be a constant function. Then $f'(0) = \lim_{h\to 0} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0} \frac{c-c}{h} = \lim_{h\to 0} 0 = 0$.

Let $f(x) = x$. Then $f'(0) = \lim_{h\to 0} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0} \frac{h}{h} = \lim_{h\to 0} 1 = 1$.

Let $f(x) = x^2$. Then $f'(1) = \lim_{h\to 0} \frac{f(1+h)-f(1)}{h} = \lim_{h\to 0} \frac{1+2h+h^2-1}{h} = \lim_{h\to 0}(2 + h) = 2$.

Let $f(x) = x^{-1}$. Then

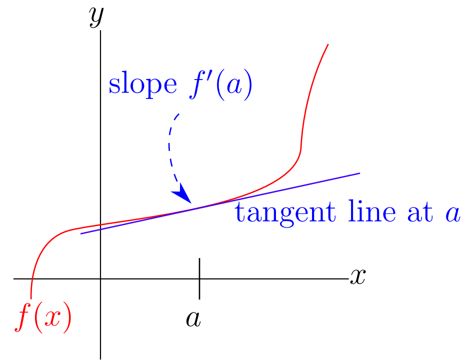
$$f'(3) = \lim_{h\to 0} \frac{f(3 + h) - f(3)}{h} = \lim_{h\to 0} \frac{\frac{1}{3+h} - \frac{1}{3}}{h} = \lim_{h\to 0} \frac{3 - (3 + h)}{3(3 + h)h} = \lim_{h\to 0} \frac{-1}{3(3 + h)} = -\frac{1}{9}.$$

Definition 2.3 (**Tangent Line**)**.** Let $f\colon \mathbb{R} \to \mathbb{R}$. Assume f is differentiable at a. The **tangent line** to the graph $y = f(x)$ at the point $x = a$ is the line with slope $f'(a)$ which passes through the point $(a, f(a))$. So, this line has equation

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

Example 2.4. Let's find the tangent line to the curve $y = x^2$ at $x = 1$. If $f(x) = x^2$, then $f(1) = 1$, and we computed already $f'(1) = 2$. So, the tangent line is $y = 1 + 2(x - 1)$.

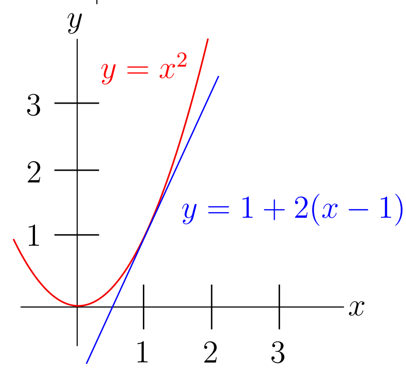
Example 2.5. Let's find the tangent line to the curve $y = x^{-1}$ at $x = 3$. If $f(x) = x^{-1}$, then $f(3) = 1/3$, and we computed already $f'(3) = -1/9$. So, the tangent line is $y = (1/3) - (1/9)(x - 3)$.



Recall that if f is a constant function, we computed $f'(0) = 0$. And if $f(x) = x$, we computed $f'(0) = 1$. Imitating these calculations leads to the following proposition.

Proposition 2.6.

- *Suppose $f(x) = c$ where c is a constant. Then $f'(a) = 0$ for all a.*
- *Suppose $f(x) = mx+b$ where m, b are constants. Then $f'(a) = m$ for all a.*



2.2. **The Derivative as a Function.** Let $x \in \mathbb{R}$, and let $f\colon \mathbb{R} \to \mathbb{R}$. If $f'(x)$ exists, then

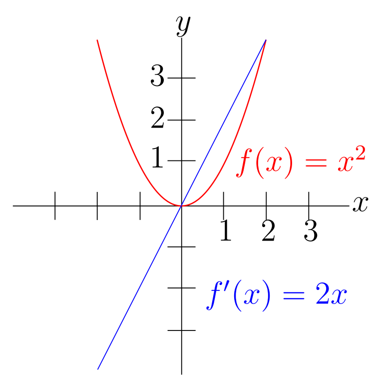
$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = f'(x) = \frac{df}{dx} = \frac{df}{dx}(x) = \frac{d}{dx}f(x).$$

Note that $f'(x)$ is also a function of x.

If $f(t)$ is the position of an object at time t, then $f'(t)$ is called the **velocity** of the object at time t. If $f(x)$ is the utility of a commodity as a function of the price x of the commodity, then $f'(x)$ is called the **marginal utility** of the commodity. In both cases, f' is the instantaneous **rate of change** of f.

Example 2.7. Let $f(x) = x^2$. Then

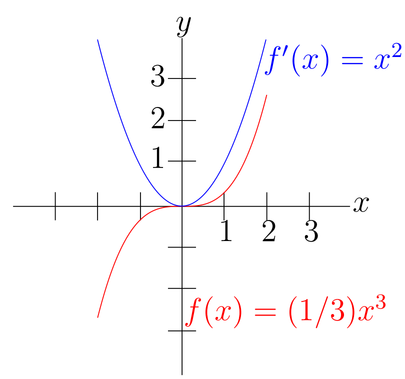
$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} \\
&= \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\
&= \lim_{h\to 0} (2x + h) = 2x.
\end{aligned}$$



So, $f'(x) = 2x$.

We can generalize this computation to arbitrary powers of positive integers.

Proposition 2.8. *Let $f(x) = x^n$, where n is any constant. Then $f'(x) = nx^{n-1}$. (If $n < 1$ and $n \neq 0$, then $f'(0)$ is undefined.)*



Proof. For simplicity, we assume n is a positive integer. Let $a, b > 0$. We need a generalization of the identity $a^2 - b^2 = (a-b)(a+b)$, or $a^3 - b^3 = (a-b)(a^2+ab+b^2)$.

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + ab^{n-2} + b^{n-1}).$$

Using this identity with $a = (x+h)$ and $b = x$, we get

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} \\ &= \lim_{h\to 0} \frac{h}{h}((x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}) \\ &= \lim_{h\to 0}((x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}) \\ &= \underbrace{x^{n-1} + \cdots + x^{n-1}}_{n \text{ times}} = nx^{n-1}. \end{aligned}$$

□

Example 2.9. If $f(x) = x^3$, then $f'(x) = 3x^2$. If $f(x) = x^{2/3}$, then $f'(x) = (2/3)x^{-1/3}$. If $f(x) = x^{-\sqrt{2}}$, then $f'(x) = -\sqrt{2}x^{-\sqrt{2}-1}$.

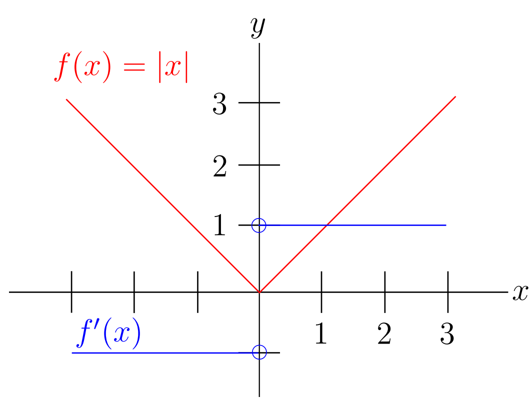
Proposition 2.10. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be differentiable on $\mathbb{R}$. Then f is continuous on $\mathbb{R}$.*

Proof. Let $a \in \mathbb{R}$. Since f is differentiable at a, $\lim_{x\to a} \frac{f(x)-f(a)}{x-a}$ exists. For $x \neq a$, write $f(x) - f(a) = \frac{f(x)-f(a)}{x-a}(x-a)$. Then our limit law for products applies, yielding

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

So, $\lim_{x\to a} f(x) = f(a)$, i.e. f is continuous at a. □

Is there a continuous function that is not differentiable at one point? In other words, does the converse to Proposition 2.10 hold?



Example 2.11 (**A Discontinuous Derivative**)**.** Consider the function $f(x) = |x|$, $x \in \mathbb{R}$. Since $f(x) = x$ for $x > 0$, and $f(x) = -x$ for $x < 0$, we see that $f'(x)$ resembles the Heaviside function

$$f'(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0. \end{cases}$$

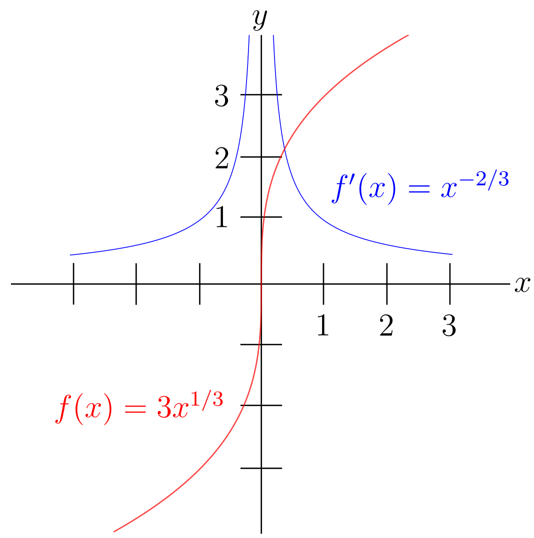
But what happens at zero? Observe,

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

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

So, $\lim_{h\to 0} \frac{f(0+h)-f(0)}{h}$ does not exist, i.e. $f'(0)$ does not exist and the converse of Proposition 2.10 is false. In fact, a stronger statement holds.

From the contrapositive of Proposition 2.10, we know that if f is discontinuous, then f is not differentiable. There are even more ways for f to not be differentiable.



Example 2.12 (**A Derivative Approaching Infinity**)**.** Let $x \in \mathbb{R}$, and let $f(x) = 3x^{1/3}$. For $x \neq 0$, $f'(x) = x^{-2/3}$. So, $\lim_{x\to 0} f'(x) = \infty$, i.e. $\lim_{x\to 0} f'(x)$ does not exist. Also, $f'(0)$ does not exist.

2.3. **Product Rule, Quotient Rule, Chain Rule.**

Proposition 2.13 (**Properties of the Derivative**)**.** *Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be differentiable functions. Let $c \in \mathbb{R}$ be a constant.*

- $\frac{d}{dx}[cf(x)] = c\frac{d}{dx}f(x)$.
- $\frac{d}{dx}(f(x)+g(x)) = \left(\frac{d}{dx}f(x)\right) + \left(\frac{d}{dx}g(x)\right)$
- $\frac{d}{dx}\left(f(x)-g(x)\right) = \left(\frac{d}{dx}f(x)\right) - \left(\frac{d}{dx}g(x)\right)$.
- *(**Product rule**)* $\dfrac{d}{dx}[f(x)g(x)] = f'(x)g(x) + g'(x)f(x)$.
- *(**Quotient rule**) If $g(x) \neq 0$, then*
$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}.$$
- *(**Chain rule**) If $f, g\colon \mathbb{R} \to \mathbb{R}$ are differentiable, then*
$$\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x).$$
Writing $u = g(x)$, and $y = f(g(x))$, we can also write this as
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

Remark 2.14. The quotient rule can be memorized with the mnemonic "low d-hi minus hi d-low over low squared."

Example 2.15.
$$\frac{d}{dx}(x^4+3x^2+2x^{1/2}+1) = (\frac{d}{dx}x^4)+3(\frac{d}{dx}x^2)+2(\frac{d}{dx}x^{1/2})+(\frac{d}{dx}1) = 4x^3+6x+x^{-1/2}.$$

Example 2.16. Let $f(x) = x^2-2x+1 = (x-1)^2$. Note that $f'(x) = 2x-2 = 2(x-1)$. So, when $x > 1$, $f'(x) > 0$, and f is increasing. And when $x < 1$, $f'(x) < 0$, and f is decreasing. Finally, when $x = 1$, $f'(x) = 0$, so f is neither increasing nor decreasing, and the tangent line to f is horizontal.

Example 2.17. Let $f(x) = x^3 + x$ and let $g(x) = x^{1/2} - 3$. Then $(d/dx)(f(x)g(x)) = (d/dx)[(x^3+x)(x^{1/2}-3)] = (x^3+x)(x^{-1/2}/2) + (x^{1/2}-3)(3x^2+1)$.

Example 2.18. Let $f(x) = x^3$ and let $g(x) = x^{1/2}-3$. Then $(d/dx)(f(g(x))) = (d/dx)(x^{1/2}-3)^3 = 3(x^{1/2}-3)^2(1/2)x^{-1/2}$.

Example 2.19. Let $f(x) = \frac{x}{x^2+1}$. Then $f'(x) = \frac{(x^2+1)\frac{d}{dx}(x) - x\frac{d}{dx}(x^2+1)}{(x^2+1)^2} = \frac{(x^2+1)-x(2x)}{(x^2+1)^2}$.

Example 2.20. Let $f(x) = \frac{1}{x^3+2}$. Then $f'(x) = \frac{(x^3+2)\frac{d}{dx}(1) - 1\frac{d}{dx}(x^3+2)}{(x^3+2)^2} = \frac{-3x^2}{(x^3+2)^2}$.

Proof of the Product Rule. Let $x, h \in \mathbb{R}$. Then

$$
\begin{aligned}
&\frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&\quad = \frac{f(x+h)g(x+h) - f(x)g(x+h) + g(x+h)f(x) - f(x)g(x)}{h} \\
&\quad = \frac{f(x+h)g(x+h) - f(x)g(x+h)}{h} + \frac{g(x+h)f(x) - f(x)g(x)}{h} \\
&\quad = g(x+h)\frac{f(x+h)-f(x)}{h} + f(x)\frac{g(x+h)-g(x)}{h}.
\end{aligned}
$$

Since g is differentiable, g is continuous by Proposition 2.10, so that $\lim_{h\to 0} g(x+h) = g(x)$. Applying our limit laws (which ones, and how are they justified?), we get

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

$\square$

Proof of the Quotient Rule. Let x such that $g(x) \neq 0$. We first show that $(1/g(x))$ is differentiable, and

$$\frac{d}{dx}\left[\frac{1}{g(x)}\right] = -\frac{g'(x)}{(g(x))^2}. \qquad (*)$$

Observe,

$$
\begin{aligned}
\frac{1}{h}\left(\frac{1}{g(x+h)} - \frac{1}{g(x)}\right) &= \frac{1}{h}\left(\frac{g(x)}{g(x)g(x+h)} - \frac{g(x+h)}{g(x)g(x+h)}\right) \\
&= \frac{g(x)-g(x+h)}{h}\frac{1}{g(x)g(x+h)}. \qquad (**)
\end{aligned}
$$

Since g is differentiable, g is continuous by Proposition 2.10, so $\lim_{h\to 0} g(x+h) = g(x)$. Using our quotient limit law,

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

So, taking the limit of $(**)$, and using our limit law for products,

$$\lim_{h\to 0}\frac{1}{h}\left(\frac{1}{g(x+h)}-\frac{1}{g(x)}\right)=\left(\lim_{h\to 0}\frac{g(x)-g(x+h)}{h}\right)\left(\lim_{h\to 0}\frac{1}{g(x)g(x+h)}\right)=\frac{g'(x)}{(g(x))^2}.$$

Since $\lim_{h\to 0}\frac{1}{h}\left(\frac{1}{g(x+h)}-\frac{1}{g(x)}\right)$ exists, $1/g(x)$ is differentiable, and formula $(*)$ is proven. We can now conclude by using the product rule. Observe

$$\begin{aligned}\frac{d}{dx}\left(f(x)\frac{1}{g(x)}\right)&=f(x)\frac{d}{dx}\left(\frac{1}{g(x)}\right)+f'(x)\frac{1}{g(x)}\\&=-f(x)\frac{g'(x)}{(g(x))^2}+f'(x)\frac{g(x)}{(g(x))^2}=\frac{g(x)f'(x)-f(x)g'(x)}{(g(x))^2}.\end{aligned}$$

□

Proof sketch of the Chain Rule.

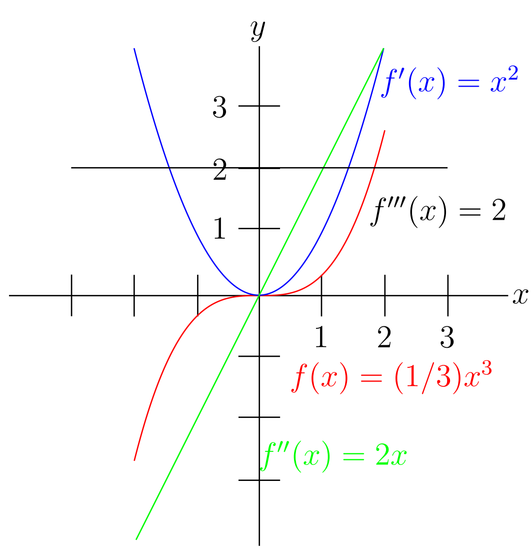
$$\begin{aligned}\frac{d}{dx}f(g(x))&=\lim_{h\to 0}\frac{f(g(x+h))-f(g(x))}{h}\\&=\left(\lim_{h\to 0}\frac{f(g(x+h))-f(g(x))}{g(x+h)-g(x)}\right)\left(\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\right)\\&=\left(\lim_{y\to g(x)}\frac{f(y)-f(g(x))}{y-g(x)}\right)\left(\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\right)=f'(g(x))g'(x).\end{aligned}$$

□

Corollary 2.21. *Let $f\colon\mathbb{R}\to\mathbb{R}$ be differentiable. Let n be a positive integer, let m,b be constants. Then*

- $\frac{d}{dx}(f(x))^n=n(f(x))^{n-1}f'(x)$.
- $\frac{d}{dx}f(x^n)=f'(x^n)nx^{n-1}$.
- $\frac{d}{dx}f(mx+b)=mf'(mx+b)$.

2.4. **Higher Derivatives.** Let $f\colon\mathbb{R}\to\mathbb{R}$. Recall that f' is also a function, and $f'>0$ when f is increasing, while $f'<0$ when f is decreasing. Also, if $f(t)$ represents the position of an object at time t, then $f'(t)$ is the velocity of the object at time t, since $f'(t)$ is the rate of change of the position over time. We can also consider the rate of change of the velocity over time, which is known as acceleration. That is, we can consider the derivative $(d/dt)f'(t)$ of $f'(t)$ to be the acceleration of the object at time t. We denote this second derivative as $f''(t)$.



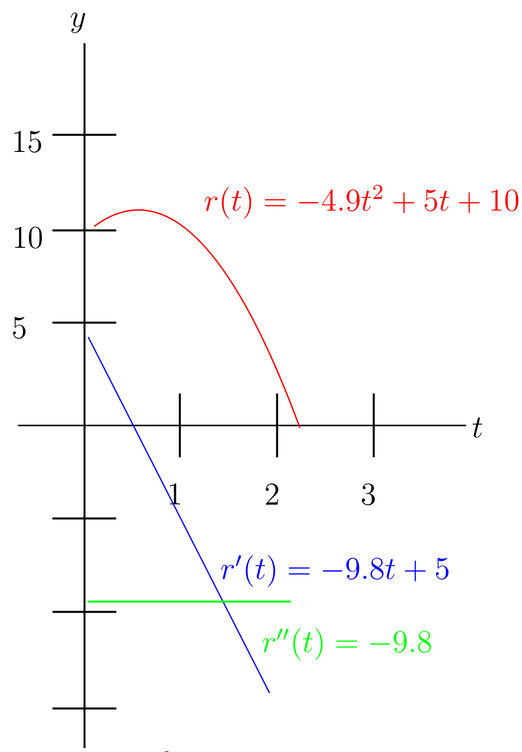
Definition 2.22 (**Higher Derivatives**). Let $f\colon \mathbb{R} \to \mathbb{R}$. We define the **second derivative** of f at the point $x \in \mathbb{R}$ by

$$f''(x) = (d/dx)f'(x) = \frac{d^2}{dx^2}f(x).$$

We similarly define the third derivative of f by

$$f'''(x) = (d/dx)f''(x) = \frac{d^3}{dx^3}f(x),$$

we define the fourth derivative of f by $f^{(4)}(x) = (d/dx)f'''(x) = d^4 f(x)/dx^4$, and so on. In general, for any positive integer n, we define the n^{th} derivative of f by $f^{(n)}(x) = (d/dx)f^{(n-1)}(x) = \frac{d^n}{dx^n}f(x)$.



Example 2.23. Let $f(x) = x^3 - 3x + 1$. Then $f'(x) = 3x^2 - 3$, $f''(x) = 6x$, $f'''(x) = 6$, $f^{(4)}(x) = 0$, $f^{(5)}(x) = 0$, and all higher derivatives of f are zero.

Example 2.24. Suppose you throw a baseball vertically in the air, with initial upward velocity v_0 and initial position r_0 (and we neglect air friction). Then, the position of the baseball (in meters) at time t (in seconds) is

$$r(t) = r_0 + tv_0 - (9.8/2)t^2.$$

Note that $r(0) = r_0$, $r'(0) = v_0$, and $r''(t) = -9.8$. That is, for any time t, the acceleration of the baseball (due to gravity) is constant. Also, all higher derivatives of r are zero: $r'''(t) = 0$, $r^{(4)}(t) = 0$, and so on.

2.5. **Exponential Functions.** From our previous courses (or our calculators), we should have a good understanding of the meaning of the following numbers

$$2^1 = 2, \qquad 3^2 = 9,$$
$$6^{-1} = 1/6, \qquad (1/2)^{1.5} = 2^{-3/2} \approx 0.354.$$

Perhaps it is less obvious what is meant by the numbers

$$2^\pi, \qquad 0^0, \qquad (-1)^{\sqrt{2}}.$$

We will eventually understand the first two numbers, but the last one will not be discussed in this class (though calculators seem to have an idea of what the last number should be). We would like to give a general treatment of numbers of this sort, since they arise so frequently. We therefore will discuss exponential functions below. There are several ways to introduce exponential functions, and we will try to describe these different presentations later on, since they are all useful.

Definition 2.25. Let $b > 0$ with $b \neq 1$. Let x be a real number. An **exponential function** is a function of the form

$$f(x) = b^x.$$

(Since $1^x = 1$ for all real numbers x, the case $b = 1$ is excluded.)

Here are some properties of exponential functions.

Proposition 2.26 (**Properties of Exponential Functions**). *Let $b > 0$, let x, y be real numbers, let n be a positive integer, and let $f(x) = b^x$.*

- *Exponential functions are always positive: $b^x > 0$ for all real numbers x.*
- *If $b > 0$ with $b \neq 1$, the range of $f(x) = b^x$ is the set of all positive real numbers.*
- *If $b > 1$, then $f(x) = b^x$ is increasing. If $0 < b < 1$, then f is decreasing.*
- $b^0 = 1$.
- $b^{x+y} = b^x b^y$.
- $\frac{b^x}{b^y} = b^{x-y}$.
- $b^{-x} = \frac{1}{b^x}$.
- $(b^x)^y = b^{(xy)}$.
- $b^{1/n} = \sqrt[n]{b}$.

2.5.1. *Derivatives of Exponential Functions.* Let $b > 0$, let x be a real number, and let $f(x) = b^x$. Let's try to compute the derivative of $f(x)$.

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

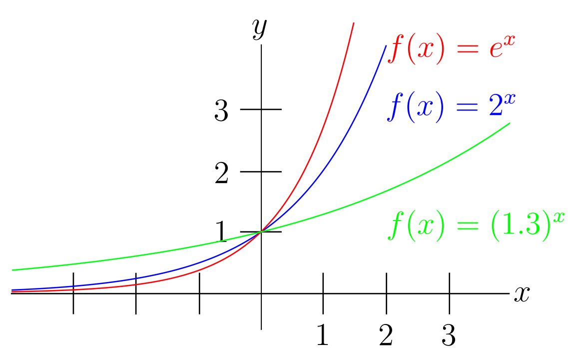
Here we defined $m(b) = \lim_{h \to 0} \frac{b^h - 1}{h}$. In summary,

Proposition 2.27. *Let $b > 0$. The function $f(x) = b^x$ is proportional to its own derivative. That is,*

$$f'(x) = m(b) f(x).$$

We will compute $m(b)$ later on, but for now, let's find a case where $m(b) = 1$.

Note that $m(1) = 0$, $m(3)$ seems larger than 1, and m seems to be continuous. So, by the Intermediate Value Theorem, there seems to be a unique number $0 < e < 3$ such that $m(e) = 1$. In summary,



Proposition 2.28. *There is a unique positive real number e such that*

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

Also, $e \approx 2.718281828\ldots$.

Remark 2.29. Let $b > 0$, $f(x) = b^x$. Then $f(0) = 0$, and $f'(0) = m(b)$ is the slope of f at $x = 0$.

Remark 2.30. The number e was actually discovered in an investigation related to compound interest. We will discuss this later on, but we note for now that

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

n	1	2	3	4	5	$\cdots$	10	$\cdots$	100	$\cdots$	1000	$\cdots$
$\left(1+\frac{1}{n}\right)^n$	2	2.25	2.3704	2.4414	2.4883	$\cdots$	2.5937	$\cdots$	2.7048	$\cdots$	2.7169	$\cdots$

And more generally,

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

The last formula is natural in the following sense. Suppose I have 100 dollars in a bank that computes interest compounded once a year, at a 6 percent interest rate. After a year, I then have

$$100(1+.06).$$

If instead, the bank compounds the interest daily at the 6 percent rate, then after a year I have

$$100(1+.06/(365))^{365}$$

If instead, the bank compounds the interest every second at the 6 percent rate, then after a year I have

$$100(1+.06/(31536000))^{31536000}$$

So, for $x = .06$, $100e^x$ computes the money that I have after a year when the interest rate is x, and when the compounding time goes to zero.

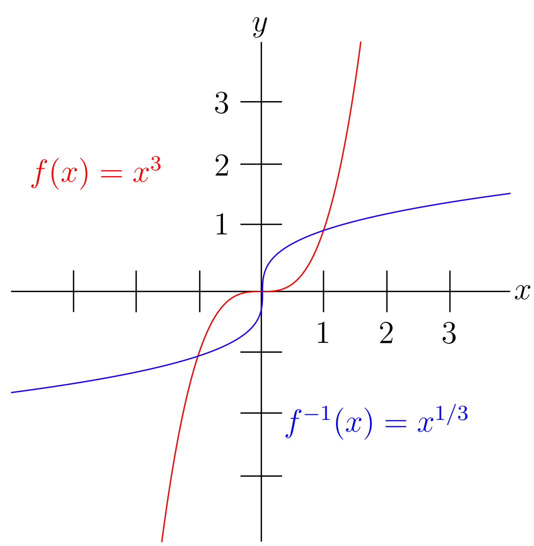
Proposition 2.31. *Let $g\colon \mathbb{R}\to\mathbb{R}$ be a function. Then*

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

In particular, for c, d constants, we have $(d/dx)e^{cx+d} = ce^{cx+d}$.

2.6. **Inverse Functions.** Let f be a real valued function on the real line. An inverse function for f does not always exist. If an inverse function for f exists, it will un-do the effect of f. For example, the function $f(x) = x^3$ has an inverse $f^{-1}(x) = x^{1/3}$. Note that

$$f(f^{-1}(x)) = (x^{1/3})^3 = x,$$
$$f^{-1}(f(x)) = (x^3)^{1/3} = x.$$



The function $f(x) = x^3$ is a bit special, in that it actually has an inverse function. To see what makes this function special, let's consider an example that does not have an inverse. Consider the function $f(x) = x^2$ on the whole real line. Note that $f(2) = f(-2) = 4$. This implies that f does not have an inverse. If we could find a function f^{-1} that un-does the effect of f, then when we apply f^{-1}

to the equality $f(2) = f(-2)$, we would get $2 = -2$, which is clearly false. Let's therefore take this obstruction to creating an inverse, and turn it into a definition.

Definition 2.32. Let f be a function with domain D and range R. We say that f is **one-to-one** if and only if, for every $c \in R$, there exists exactly one $x \in D$ such that $f(x) = c$.

Definition 2.33. Let f be a function with domain D and range R. If there is a function g with domain R and range D such that

$$g(f(x)) = x \quad \text{for all}\, x \in D \qquad \text{and} \qquad f(g(x)) = x \quad \text{for all}\, x \in R,$$

then f is said to be **invertible**. The function g is called the **inverse** of f, and we denote it by $g = f^{-1}$.

Remark 2.34. If f is a one-to-one function with domain D and range R, then f is invertible. If f is not one-to-one, then f is not invertible.

Example 2.35. The function $f(x) = x^3$ with domain $(-\infty, \infty)$ and range $(-\infty, \infty)$ is one-to-one, since any real number c has exactly one cube root x with $x^3 = c$. So, f is invertible. As we showed above, $f^{-1}(x) = x^{1/3}$.

Example 2.36. The function $f(x) = x^2$ with domain $(-\infty, \infty)$ and range $[0, \infty)$ is not one-to-one, since, as we saw above, the number 4 has two distinct numbers 2 and -2 such that $f(2) = f(-2) = 4$. So, f is not invertible when it has the domain $(-\infty, \infty)$. However, if we restrict the domain of f, so that we consider $f(x) = x^2$ with domain $[0, \infty)$ and range $[0, \infty)$, then f is one-to-one. Every nonnegative number c has exactly one nonnegative square root x such that $x^2 = c$. So, f is invertible when it has the domain $[0, \infty)$. In this case, f^{-1} is the square root function, which we write as $f^{-1}(x) = \sqrt{x}$. (Note that the square root function $\sqrt{x}$ is distinct from the concept of "the square roots of x." In particular, $\sqrt{x}$ is only defined when $x \geq 0$, and it holds that $\sqrt{x^2} = |x|$ for any real number x. So, if x is negative, $\sqrt{x^2} \neq x$. Moreover, the answer to "What are the square roots of 9" is 3 and -3.)

Remark 2.37 (Horizontal Line Test). Let f be a function with domain D and range R. Suppose we look at the graph of f over the domain D. Then f is one-to-one if and only if every horizontal line passes through the graph of f in at most one point.

Remark 2.38. The function $f(x) = x^3$ with domain $(-\infty, \infty)$ satisfies the horizontal line test. The function $f(x) = x^2$ with domain $(-\infty, \infty)$ does not satisfy the horizontal line test.

Proposition 2.39. *Let f be a strictly increasing function. That is, if $x < y$, then $f(x) < f(y)$. Then f is one-to-one.*

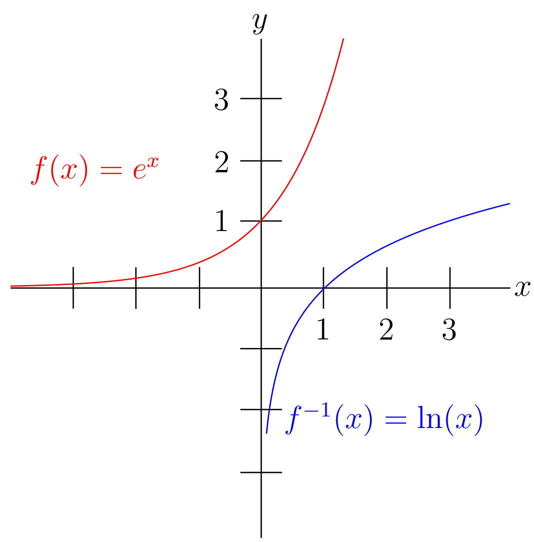
Remark 2.40. Recall that a continuously differentiable function with domain $(-\infty, \infty)$ such that $f'(x) > 0$ is increasing. This follows from the Mean Value Theorem (Theorem 3.18 below). If we had $x < y$ with $f(x) \geq f(y)$, then there would be some c with $x < c < y$ such that $f'(c) = (f(y) - f(x))/(y - x) \leq 0$. But $f'(c) > 0$, so we must have $f(x) < f(y)$, so that f is strictly increasing.

Remark 2.41. Since $(d/dx)e^x = e^x > 0$, the function e^x is increasing on $(-\infty, \infty)$. So, this proposition implies that the exponential function is one-to-one, so that it has an inverse. Similarly, if $b > 1$, then $(d/dx)b^x = m(b)b^x$, and $m(b) > 0$, so b^x is increasing and invertible. And if $b < 1$, then $(d/dx)b^x = m(b)b^x$ and $m(b) < 0$, so $-b^x$ is increasing and invertible.

To better understand the inverse of the exponential function, let's consider the graph of the inverse of the exponential function.

Remark 2.42. Suppose $y = f(x)$ is an invertible function. Then $(x, y) = (x, f(x))$ is a point in the graph of f. Since f is invertible, we can apply f^{-1} to both sides of $y = f(x)$ to get $f^{-1}(y) = x$. So, the point $(y, x) = (y, f^{-1}(y))$ is in the graph of f^{-1}. So, whenever the graph of f contains (x, y), the graph of f^{-1} contains (y, x). Note that (y, x) is the reflection of the point (x, y) across the line $y = x$. Graphically, this means that if we plot the function $y = f(x)$, then the graph of the inverse function f^{-1} is the reflection of the graph of f across the line $y = x$.

Example 2.43. Let's plot the function $f(x) = e^x$. Note that $f(0) = 1$, and $e^x > x$ for all x. This follows since $e^0 > 0$, and since $f'(x) = e^x > 1$ for all $x > 0$, while $(d/dx)x = 1$. To find the graph of the inverse function f^{-1}, we then just reflect the graph of f across the line $y = x$.



A function and its inverse are not only related with respect to their graphs, but also with respect to their derivatives, as we now show.

Theorem 2.44 (Inverse Differentiation). *Let f be a function of a real variable with inverse $g = f^{-1}$. If x is in the domain of g, and if $f'(g(x)) \neq 0$, then $g'(x)$ exists, and*

$$g'(x) = \frac{1}{f'(g(x))}.$$

Remark 2.45. The proof of this identity follows from the Chain Rule. If we additionally assume that g is differentiable at x, then the Chain rule applied to $f(g(x)) = x$ says that

$$f'(g(x)) \cdot g'(x) = \frac{d}{dx} f(g(x)) = \frac{d}{dx} x = 1.$$

That is, $g'(x) = 1/(f'(g(x)))$.

Remark 2.46. This Theorem is another tool that will allow us to understand the inverse of the exponential function. In particular, it will allow us to compute the derivative of the inverse of the exponential function.

Example 2.47. Let's consider again the function $f(x) = x^3$ on $(-\infty, \infty)$, where $f^{-1}(x) = x^{1/3}$. Then $f'(x) = 3x^2$, so

$$(f^{-1})'(x) = [f'(f^{-1}(x))]^{-1} = [3(f^{-1}(x))^2]^{-1} = x^{-2/3}/3.$$

However, we already "knew" that $(f^{-1})'(x) = (1/3)x^{-2/3}$, by our usual differentiation rules.

Remark 2.48. You should not confuse the notation f^{-1}, which denotes the inverse function of f, with a number to the -1 power. That is, it is typically true that $f^{-1}(x) \neq (f(x))^{-1}$.

2.7. **Logarithms.** As discussed in Remark 2.41, if $b > 0$ with $b \neq 1$, then the function $f(x) = b^x$ is invertible. We write $f^{-1}(x) = \log_b(x)$. By the definition of invertibility, we have

Proposition 2.49.

$$b^{\log_b x} = x, \qquad \log_b b^x = x.$$

Remark 2.50. When $b = e$, we write $\log_b(x) = \ln(x)$, and we call $\ln(x)$ the **natural logarithm**.

Example 2.51. Since $2^3 = 8$, we have $\log_2(8) = \log_2(2^3) = 3$, and $2^{\log_2 8} = 8$.
Since $3^{-2} = 1/9$, we have $\log_3(1/9) = \log_3(3^{-2}) = -2$, and $3^{\log_3(1/9)} = 1/9$.

Let $b > 0$ with $b \neq 1$. Let $f(x) = b^x$. If $b > 1$, then

$$\lim_{x\to\infty} b^x = \infty, \qquad \lim_{x\to-\infty} b^x = \lim_{x\to\infty} \frac{1}{b^x} = 0.$$

So, f has domain $(-\infty, \infty)$ and range $(0, \infty)$. So, if $b > 1$, we can see from the graph of $\log_b$ that

$$\lim_{x\to\infty} \log_b(x) = \infty, \qquad \lim_{x\to 0^+} \log_b(x) = -\infty.$$

If $b < 1$, then

$$\lim_{x\to\infty} b^x = 0, \qquad \lim_{x\to-\infty} b^x = \lim_{x\to\infty} \frac{1}{b^x} = \infty.$$

So, f has domain $(-\infty, \infty)$ and range $(0, \infty)$. And, if $b < 1$, we can see from the graph of $\log_b$ that

$$\lim_{x\to\infty} \log_b(x) = -\infty, \qquad \lim_{x\to 0^+} \log_b(x) = +\infty.$$

Remark 2.52. The logarithm of a negative number is undefined. To see why, note that if we could write $x = \log_3(-2)$, then by exponentiation both sides, we would have $3^x = -2$. But this equation has no real solution.

Here is a summary of properties of logarithm functions. These properties typically follow by the corresponding properties of the exponential functions

Proposition 2.53 (**Properties of Logarithm Functions**). *Let $b > 0$, with $b \neq 1$, let x, y be positive real numbers, let n be a positive integer, and let $f(x) = \log_b(x)$.*

- *The domain of $f(x) = \log_b(x)$ is the set of all positive numbers, and the range of f is the set of all real numbers*
- *If $b > 1$, then $f(x) = \log_b(x)$ is increasing. If $0 < b < 1$, then f is decreasing.*
- $\log_b(1) = 0$.
- $\log_b(xy) = \log_b(x) + \log_b(y)$
- $\log_b(x/y) = \log_b(x) - \log_b(y)$.
- $\log_b(1/x) = -\log_b(x)$.
- $\log_b(x^n) = n \log_b x$.

Moreover, if $a > 0$ with $a \neq 1$, then the logarithm functions are proportional in the following sense

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}, \qquad \log_b(x) = \frac{\ln x}{\ln b}$$

To see why some of these properties are true, recall that $b^0 = 1$. So, taking the log of both sides, $\log_b(1) = 0$, proving the third property. To prove the fourth property, note that

$$b^{\log_b(xy)} = xy = (b^{\log_b(x)})(b^{\log_b(y)}) = b^{\log_b(x)+\log_b(y)}.$$

So, taking $\log_b$ of both sides, we get $\log_b(xy) = \log_b(x) + \log_b(y)$. The other properties follow in a similar way.

To see the final property, note that

$$e^{\ln(b)\log_b(x)} = (e^{\ln(b)})^{\log_b(x)} = b^{\log_b(x)} = x.$$

So, taking ln of both sides shows that $\ln(b)\log_b(x) = \ln(x)$, proving the final property. This property then implies that

$$\log_b(x) = \frac{\ln x}{\ln b} = \frac{(\ln x)/(\ln a)}{(\ln b)/(\ln a)} = \frac{\log_a(x)}{\log_a(b)}.$$

Example 2.54.

$$\log_6(9) + \log_6(4) = \log_6(9 \cdot 4) = \log_6(36) = \log_6(6^2) = 2.$$

Using the properties of the logarithm, we can now finally differentiate exponential functions and logarithms.

Theorem 2.55 (**Derivative of Exponential**)**.** *Let $b > 0$. Then*

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

This formula follows readily from the properties of the logarithm, as follows.

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

We can then use the inverse differentiation formula to differentiate the logarithm as follows.

Theorem 2.56 (**Derivative of Logarithm**)**.** *Let $x > 0$. Then*

$$\frac{d}{dx} \ln x = \frac{1}{x}.$$

Consequently, for any $b > 0$ with $b \neq 1$, since $\log_b(x) = \frac{\ln x}{\ln b}$, we have

$$\frac{d}{dx} \log_b(x) = \frac{1}{x \ln b}.$$

To justify this differentiation, we apply the inverse differentiation formula (Theorem 2.44) where $f(x) = e^x$ and $f^{-1}(x) = \ln x$ to get

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

Remark 2.57. Let f be a positive function. From the chain rule, we have

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

This formula can be useful for calculating derivatives.

Example 2.58. Let j be a positive number. Let's calculate the derivative of

$$f(x) = x^j.$$

Strictly speaking, you never learned why the formula $f'(x) = jx^{j-1}$ holds. Let's see why it holds now. Let $x > 0$. Then $\ln f(x) = j \ln x$. So, differentiating both sides and using the above remark,

$$\frac{f'(x)}{f(x)} = \frac{j}{x}.$$

That is, $f'(x) = \frac{f(x)j}{x} = jx^{j-1}$.

We can carry this method further to more exotic functions that are not covered by any rules we have learned before. Consider

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

You might think the Chain Rule applies, but I do not see an obvious method for applying the chain rule. You can use some properties of the logarithm function to write

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

Alternately, we can write

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

So that, once again, $f'(x) = f(x)(1+\ln x) = x^x(1+\ln x)$.

Remark 2.59. Note that for $t < 0$, $(d/dt)\ln(-t) = 1/t$, so for any $t \neq 0$, we have

$$\frac{d}{dt}\ln|t| = \frac{1}{t}$$

So, from the chain rule, if f is a nonzero function, then

$$\frac{d}{dt}\ln|f(t)| = \frac{f'(t)}{f(t)}.$$

2.8. **Application: Exponential Growth.** One of the many reasons to care about the exponential function is that it appears in so many applications. For example, bacteria grow at an exponential rate (if they have unlimited food and space). Money in a bank account grows at an exponential rate. But also, money devalues at an exponential rate (roughly three percent per year in the US.) This last observation (known as inflation) explains why most common material possessions "used to cost a nickel and a dime." Specifically, if money loses roughly three percent of its value per year, then the current value of one dollar will be equal to the value of two dollars in roughly twenty years. Put another way, the amount of goods you can buy with a fixed amount of money gets cut in half, roughly every twenty years. So, one hundred years ago, money was roughly $2^5 = 32$ times more valuable. A dime roughly a hundred years ago could buy the same amount as around three dollars now. A $20,000$ dollar car now would cost roughly 600 dollars one hundred years ago. And indeed, the latter number is comparable to the original price of a Ford Model T. To compare money at different historical periods, we therefore hear monetary amounts that have been inflation adjusted, which means that they take into account this change in the value of money over time.

Let's be more precise about exponential growth. Let t, k and P_0 be real numbers, and define

$$P(t) = P_0 e^{kt}.$$

If $k > 0$, we say that P **grows exponentially**. If $k < 0$, we say that P **decays exponentially**.

For example, if $P(t)$ denotes a population of bacteria at time t, then $P(0) = P_0$ is the population of the bacteria at time zero, and $k > 0$ represents the rate of growth of the bacteria.

Note that $P(t)$ satisfies the differential equation

$$P'(t) = kP(t), \qquad P(0) = P_0.$$

It turns out that this differential equation uniquely characterizes the exponential function.

Theorem 2.60 (**Differential Equation Characterization of Exponential**). *Let $y(t)$ be a continuously differentiable function of t such that*

$$y'(t) = ky(t), \qquad y(0) = P_0.$$

Then we must have $y(t) = P_0 e^{kt}$.

To justify this statement, we differentiate $y(t)e^{-kt}$ to get

$$\frac{d}{dt}(y(t)e^{-kt}) = -ky(t)e^{-kt} + y'(t)e^{-kt} = e^{-kt}(-ky(t) + ky(t)) = 0.$$

Therefore, there exists a constant C such that $y(t)e^{-kt} = C$, so that $y(t) = Ce^{kt}$. Since $P_0 = y(0) = Ce^0 = C$, we conclude that $C = P_0$, so that $y(t) = P_0 e^{kt}$, as desired.

Remark 2.61. We interpret this characterization of the exponential function as follows. Suppose we have a quantity $y(t)$ whose growth is proportional to $y(t)$. Then $y(t)$ must be an exponential function. In this way, the exponential function arises naturally when discussing the growth of bacteria, interest rates, radioactive decay, chemical kinetics, and so on. However, more complicated differential equations have more exotic solutions.

Example 2.62. Certain therapeutic drugs in the body are filtered by the kidneys at a rate which is proportional to their concentration in the blood stream. That is, if $y(t)$ is the concentration of the drug at time t, then there exists a constant $k < 0$ such that $dy/dt = ky$. Therefore, $y(t) = y_0 e^{kt}$.

Example 2.63. In chemistry class, we learned about "first order" reactions. In these reactions, the concentration $y(t)$ of a certain chemical is proportional to its rate of change. So, as before, $dy/dt = ky(t)$ and then $y(t) = y_0 e^{kt}$.

Exercise 2.64. In chemistry, a "second order" reaction satisfies $dy/dt = k(y(t))^2$. If $y_0 = y(0) \neq 0$, verify that we must have

$$y(t) = \frac{1}{y_0^{-1} - kt}.$$

Example 2.65. Radioactive isotopes decay at a rate that is proportional to their concentration.

As we discussed in the introduction, it is helpful to think about the doubling time (if $k > 0$) or half-life (if $k < 0$) of an exponential function $P(t) = P_0 e^{kt}$. Given any real t, the **doubling time** $T > 0$ of $P(t) = P_0 e^{kt}$ is a time such that

$$\begin{cases} P(t+T) = 2P(t) & \text{, if } k > 0 \\ P(t+T) = (1/2)P(t) & \text{, if } k < 0 \end{cases}.$$

It is evident that we have

$$T = \frac{\ln 2}{|k|}.$$

To see this, note that

$$P\left(t + \frac{\ln 2}{|k|}\right) = P(t)e^{\frac{k \ln 2}{|k|}} = P(t)2^{k/|k|} = \begin{cases} 2P(t) & \text{, if } k > 0 \\ (1/2)P(t) & \text{, if } k < 0 \end{cases}.$$

Example 2.66. A radioactive isotope decays exponentially with a half-life of 4 days. Suppose we want to know how long it takes for 80% of the isotope to decay. In this case, we have $P(t) = P_0 e^{-kt}$ where $k > 0$, and the half life satisfies $4 = (\ln 2)/k$, so that $k = (\ln 2)/4$. That is,

$$P(t) = P_0 e^{-t(\ln 2)/4} = \frac{P_0}{2^{t/4}}.$$

We want to find t such that $P(t)/P_0 = .2$. That is, we need to solve $2^{-t/4} = .2$, i.e. $(-t/4)\ln 2 = \ln(1/5)$, so $t = \frac{4 \ln 5}{\ln 2} \approx 9.2877$. Note that this answer is reasonable since the doubling time is 4 days, so it takes 8 days to get 75% decay, so 80% decay should take a bit longer than 8 days (but not more than 12, which would have 87.5% decay).

2.9. **Application: Compound Interest.** As we discussed above, if I have an account that starts with P_0 dollars at time $t = 0$ which earns an annual interest rate r which is compounded M times during the year, then the amount of money in the account after t years is

$$P(t) = P_0 \left(1 + \frac{r}{M}\right)^{Mt}.$$

Also, using the limit formula for e

$$e^r = \lim_{M \to \infty} \left(1 + \frac{r}{M}\right)^M,$$

we see that as the compounding time per year M goes to infinity, we actually reach some finite limiting quantity. That is,

$$\lim_{M \to \infty} P(t) = P_0 \left[\lim_{M \to \infty} \left(1 + \frac{r}{M}\right)^M\right]^t = P_0 (e^r)^t = P_0 e^{rt}.$$

We therefore say that, if an interest rate is **compounded continuously** at the rate r, then

$$P(t) = P_0 e^{rt}.$$

In fact, such a consideration led to the very discovery of the constant e.

We will now return to some of our economic discussions related to interest rates and deflation.

Definition 2.67. Suppose there is a (continuously compounded) interest rate $r > 0$ at which I can lend or borrow money. The **present value** of an amount P of money that is received at time t is defined as Pe^{-rt}.

Example 2.68. If $r = .1$, and if you agree to give me \$1000 twenty years from now, then this would be equivalent to giving me $1000e^{-.1(20)} = 1000e^{-2} \approx \$135$ right now (since I could put this money into an account that earns 10% interest for twenty years.)

Example 2.69. Returning to our discussion of inflation, if $r = .03$ is the rate of inflation of money, then money becomes half as valuable roughly every $(\ln 2)/.03 \approx 23$ years. So, if you agree to buy me a \$20000 car in 115 years, this would be equivalent to buying a car worth $20000e^{-.03(115)} \approx 635$ dollars right now.

To some extent, we can think of an income as a continuous function $R(t)$, in which case we can compute the present value of that income as follows.

2.10. **Application: Terminal Velocity.** Suppose an object of mass m is falling due to the force of gravity $-mg$ in the negative y-direction. The force due to air friction is then $-kv$ where v is the velocity of the object and $k > 0$. That is, the air friction opposes the velocity of the object. The total force on the object is then $-mg - kv$. If a denotes the acceleration of the object, then $a = v'$ and $F = ma$, so $ma = mv' = -mg - kv$. That is, $v'(t) = -g - kv(t)/m$. That is,

$$v'(t) = -g - kv(t)/m = (-k/m)(v(t) + gm/k).$$

If $f(t) = v(t) + gm/k$, then $f'(t) = v'(t) = (-k/m)f(t)$, so $f(t) = Ce^{-kt/m}$. That is,

$$v(t) = Ce^{-kt/m} - gm/k.$$

As $t \to \infty$, we see that

$$\lim_{t\to\infty} v(t) = -gm/k.$$

We therefore refer to $-gm/k$ as the **terminal velocity** of the body. That is, after falling for a while, the forces of gravity and air friction cancel each other out, and the body falls at a constant velocity (the terminal velocity).

A similar differential equation gives Newton's Law of Cooling.

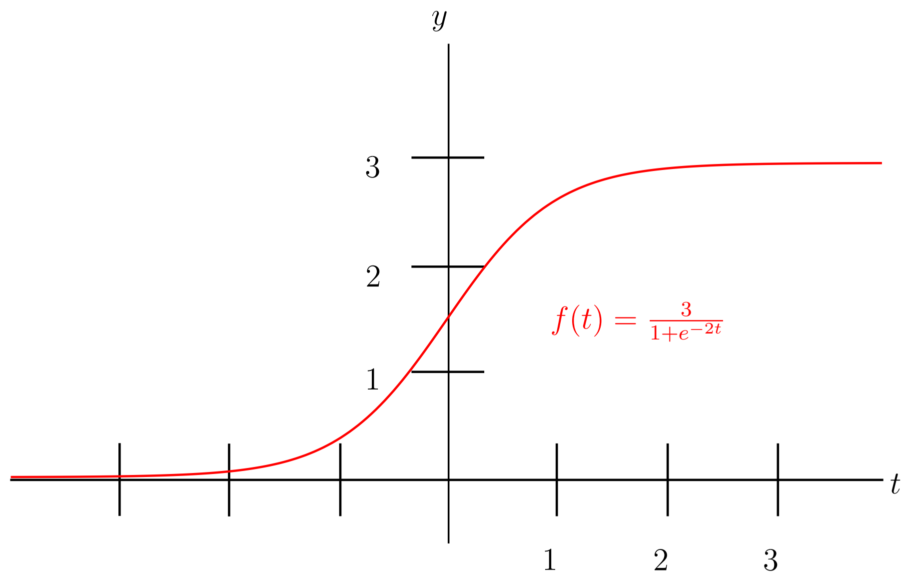
Exercise 2.70 (**Newton's Law of Cooling**)**.** Suppose $y(t)$ is the temperature of an object at time t. If an object is of a different temperature than its surroundings, then the rate of change of the object's temperature is proportional to the difference of the temperature of the object and the temperature of the surroundings. That is, if Y denotes the temperature of the surroundings, and if $y(0) = y_0 \neq Y$, then there exists a constant $k > 0$ such that

$$y'(t) = -k(y(t) - Y).$$

Note that if $y(0) < Y$, then $y'(0) > 0$, so that the temperature of y is increasing to the environment's temperature. And if $y(0) > Y$, then $y'(0) < 0$, so that y is decreasing to the environment's temperature.

Let $f(t) = y(t) - Y$. Verify that $f'(t) = -kf(t)$. Conclude that $f(t) = y(t) - Y = (y_0 - Y)e^{-kt}$. That is, we have Newton's Law of cooling:

$$y(t) = Y + (y_0 - Y)e^{-kt}.$$



Exercise 2.71. The exponential growth model for bacteria is a bit unrealistic, since after a while, the bacteria are limited by their environment and food supply. We therefore consider the **logistic growth** model. Suppose $y(t)$ is the amount of bacteria in a petri dish at time t and $k > 0$ is a constant. Let C be the maximum possible population of the bacteria. We model the growth of the bacteria by the formula

$$y'(t) = ky(t)(C - y(t)), \qquad y(0) = y_0$$

So, when y is small, $y'(t)$ is proportional to y. However, when y becomes close to C, y' becomes very small. That is, the rate of growth of bacteria is constrained by the environment.

- Verify that the following function satisfies the above differential equation.

$$y(t) = \frac{C}{1 + (C^{-1}y_0^{-1} - 1)e^{-kt}}.$$

- Plot the function $y(t)$. (What are the limits of y as t goes to $+\infty$ and $-\infty$?)
- Find out where $y'(t)$ is the largest. (Hint: find the maximum of the function of y: $ky(C - y)$.)

The latter observation explains the "J-curve" scare for human population growth in the 1980s. At this point in time, many people were afraid that the human population would grow too large for the earth to support us. However, it seems that we were simply observing the maximum possible growth rate of the human population at this time (if we believe that logistic growth models the human population reasonably well).

3. Applications of Derivatives

3.1. **Linear Approximation.** When we look at a very small domain of a function f near a point a, the function f looks like a linear function. This heuristic is expressed with the following approximation

$$f(x) \approx f(a) + (x - a)f'(a).$$

That is, when x is near a, the quantity $f(x)$ is approximately equal to the quantity $f(a) + (x - a)f'(a)$.

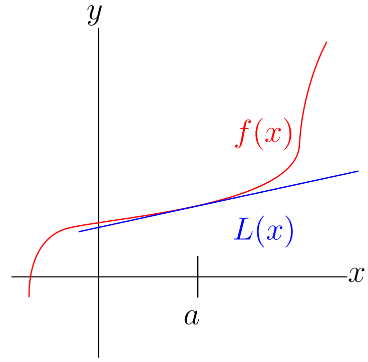
Definition 3.1. Let $f\colon \mathbb{R} \to \mathbb{R}$ and let a be a point in $\mathbb{R}$. The **linearization** of f near a is the following function of x:

$$L(x) = f(a) + (x - a)f'(a).$$

The expression

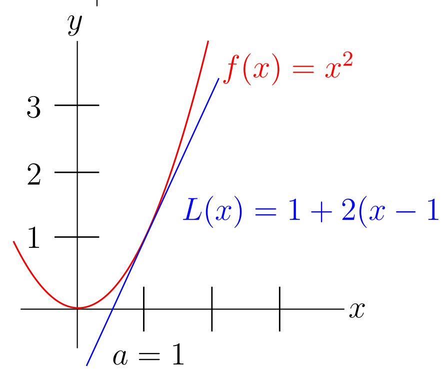
$$f(x) \approx L(x)$$

is referred to as the **linear approximation** of f at a. Note that $y = L(x)$ is exactly the equation for the tangent line to f at a.



Example 3.2. Let's estimate $(1.1)^2$ using linear approximation (even though we know $(1.1)^2 = 1.21$). We approximate $f(x) = x^2$ by $L(x)$ when $a = 1$. We have

$$\begin{aligned} L(x) &= f(1) + (x - 1)f'(1) \\ &= 1 + (x - 1)(2) \\ &= 1 + 2(x - 1). \end{aligned}$$



So, using the linear approximation of $f(x)$ at 1, we have

$$\begin{aligned} f(1 + .1) &\approx L(1 + .1) \\ &= 1 + .2. \end{aligned}$$

Example 3.3. Let's estimate $1/(9.9)$ using linear approximation. We approximate $1/x$ by $L(x)$ when $a = 10$. We have

$$L(x) = 1/10 + (x - 10)(-(10)^{-2}) = 1/10 - (x - 10)/100.$$

So, we have

$$1/(9.9) \approx L(9.9) = 1/10 - (-.1)/100 = 1/10 + 1/1000 = \frac{101}{1000}.$$

3.2. **Extreme Values and Optimization.**

Nothing takes place in the world whose meaning is not that of some maximum or minimum.

Leonhard Euler

Given a function f, it is often desirable to find the maximum or minimum value of f. For example, if $f(x)$ represents the profit obtained from setting the price x of a product, we would like to maximize the profit function f. If f represents the energy or cost of an object moving between two points, we would like to minimize this energy or cost. Many physical principles can be rephrased as statements about minimizing energy or some other quantity. For example, light always travels on the path of shortest length between two points. That is, photons minimize the length over which they travel. And so on. Thankfully there are often very general methods for maximizing or minimizing functions. Unfortunately, these general methods do not always work. For example, suppose the mailman has 1000 houses to visit, and he wants to visit them in the shortest amount of time. We do not yet know an efficient way to find the path that takes the shortest amount of time.

Definition 3.4 (**Extrema**)**.** Let D be a domain in the real line. Let $f \colon D \to \mathbb{R}$. We say that f has an **absolute maximum** on D at the point $c \in D$ if

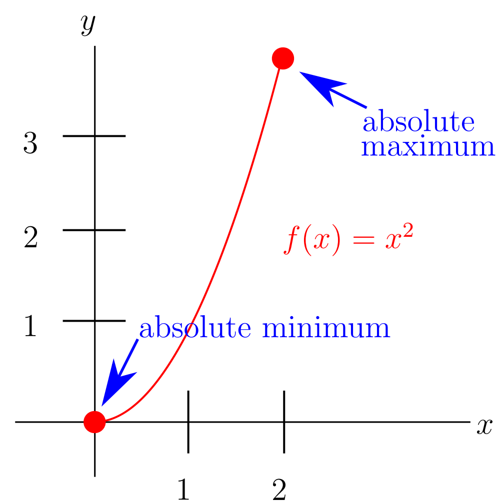
$$f(x) \le f(c) \qquad \text{for all } x \in D.$$

We say that f has an **absolute minimum** on D at the point $c \in D$ if

$$f(x) \ge f(c) \qquad \text{for all } x \in D.$$

We refer to the absolute minimum and absolute maximum values of f as the **extreme values** or **extrema** of f. The process of finding the extrema of f is called **optimization**.

Since we would often like to find the extrema of a function f, let's first look at some examples where the extrema cannot possibly be found (since they don't exist).



Example 3.5. Let $f(x) = 1/x$ where f has domain $(0,1)$. Then f has no absolute maximum since $\lim_{x\to 0^+} f(x) = \infty$. So, a discontinuity of a function can interfere with the existence of extrema.

Example 3.6. Let $f(x) = x^2$ where f has domain $[0,2]$. Then the absolute maximum of f is 4, which occurs at $x = 2$. And the absolutely minimum of f is 0, which occurs at $x = 0$.

Let $f(x) = x^2$ where f has domain $(0,2)$. Then the absolute maximum of f does not exist. And the absolutely minimum of f does not exist either. Given any point $c \in (0,2)$, there are always points $a, b \in (0,2)$ with $a < c < b$, so that $f(a) < f(c) < f(b)$. So, an extreme value cannot occur at any point in $(0,2)$. Even though f is continuous, the open interval is interfering with the existence of extrema.

In summary, if we want extrema to exist, it looks like we need our function to be continuous, and we cannot consider a domain which is an open interval. Fortunately, being continuous on a closed interval guarantees that the extrema exist.

Theorem 3.7 (**Extreme Value Theorem**)**.** *Let $a < b$. Let $f \colon [a,b] \to \mathbb{R}$ be a continuous function. Then f achieves its minimum and maximum values. More specifically, there exist $c, d \in [a,b]$ such that: for all $x \in [a,b]$, $f(c) \le f(x) \le f(d)$.*

The proof of this Theorem is outside the scope of this course.

Sometimes it is also desirable to find extreme values for a function when it is restricted to a small domain. Such points are called **local extrema**.

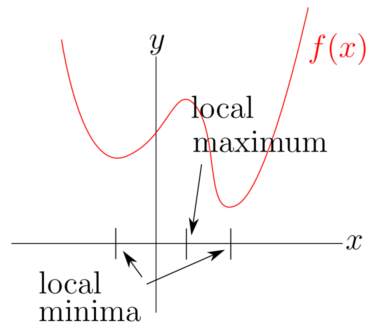
Definition 3.8 (**Local Extrema**)**.** Let $f \colon \mathbb{R} \to \mathbb{R}$. We say that f has an **local maximum** at the point $c \in \mathbb{R}$ if

$$f(x) \le f(c) \qquad \text{for all points } x \text{ in some open interval containing } c.$$

We say that f has an **local minimum** at the point $c \in \mathbb{R}$ if

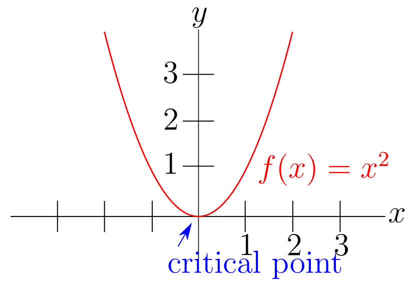
$$f(x) \ge f(c) \qquad \text{for all points } x \text{ in some open interval containing } c.$$

Finding local extrema often reduces to the problem of finding critical points.

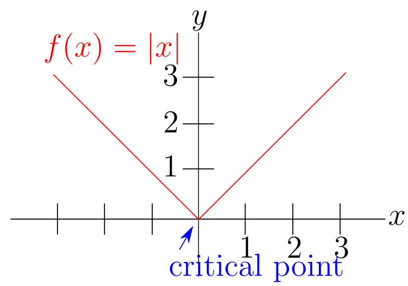


Definition 3.9 (**Critical Point**)**.** Let $f\colon \mathbb{R} \to \mathbb{R}$. We say that f has a **critical point** at the point c if $f'(c) = 0$ or $f'(c)$ does not exist.

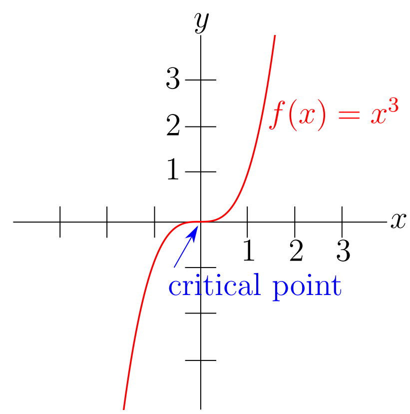
Example 3.10. Let $f(x) = x^2$. Then $f'(x) = 2x$. So, $f'(x) = 0$ only when $x = 0$. That is, $x = 0$ is the only critical point of f. Note that $x = 0$ is also the absolute minimum for f with the domain $\mathbb{R}$.



Example 3.11. Let $f(x) = |x|$. If $x < 0$, then $f'(x) = -1$. If $x > 0$, then $f'(x) = 1$. If $x = 0$, then $f'(x)$ is undefined. So, $x = 0$ is the only critical point of f. Note that $x = 0$ is also the absolute minimum for f with the domain $\mathbb{R}$.



Example 3.12. Let $f(x) = x^3$. Then $f'(x) = 3x^2$. So, $f'(x) = 0$ only when $x = 0$. That is, $x = 0$ is the only critical point of f. Note that $x = 0$ is neither a local maximum nor a local minimum of f. So, a critical point is not necessarily a local extremum.



We have just seen that a critical point is not necessarily a local extremum. However, a local extremum is always a critical point.

Theorem 3.13. *Let $f\colon \mathbb{R} \to \mathbb{R}$. If a local minimum or maximum of f occurs at c, then c is a critical point of f.*

Proof. Suppose c is a local maximum of f and $f'(c)$ exists. Then c is an absolute maximum of f on some open interval $(c-b, c+b)$ where $b > 0$. Let $h \in (-b, b)$. Then $f(x+h) \leq f(c)$, since c is a local maximum of f. If $h > 0$, we have $(f(x+h) - f(c))/h \geq 0$. So,

$$\lim_{h \to 0^+} \frac{f(x+h) - f(c)}{h} \geq 0.$$

If $h < 0$, we have $(f(x+h) - f(c))/h \leq 0$. So,

$$\lim_{h \to 0^-} \frac{f(x+h) - f(c)}{h} \leq 0.$$

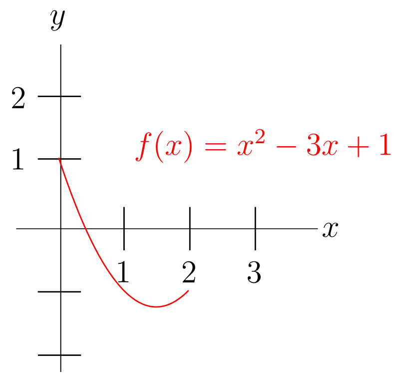
We therefore have $0 \leq f'(c) \leq 0$, so that $f'(c) = 0$ (if $f'(c)$ exists). $\square$

This Theorem allows us to find absolute extrema on closed intervals.

Proposition 3.14 (**Extreme Values on Closed Intervals**)**.** *Let $a < b$. Let $f\colon [a, b] \to \mathbb{R}$ be continuous. Then the extreme values of f on $[a, b]$ occur either at critical points of f, or at the endpoints a, b of the interval $[a, b]$.*

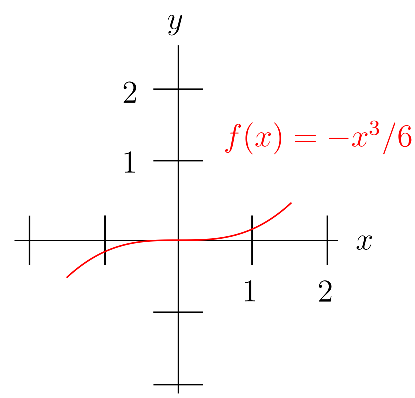
Proof. Let c be a point where an extreme value of f occurs, and assume $c \neq a$ and $c \neq b$. Then c is a local extremum of f on (a, b). So, c is a critical point of f by the Theorem 3.13. $\square$

We now apply Theorem 3.14 to a few examples.



Example 3.15. Let's find the extreme values of $f(x) = x^2 - 3x + 1$ on the interval $[0, 2]$. We have $f'(x) = 2x - 3$, so that $f'(x) = 0$ only when $x = 3/2$. So, the only critical point of f occurs at $x = 3/2$. So, the extreme values of f must occur at elements of the following list: $0, 3/2, 2$. We now check the values of f at these points. We have $f(0) = 1$, $f(3/2) = (9/4) - (9/2) + 1 = -5/4$, and $f(2) = 4 - 6 + 1 = -1$.

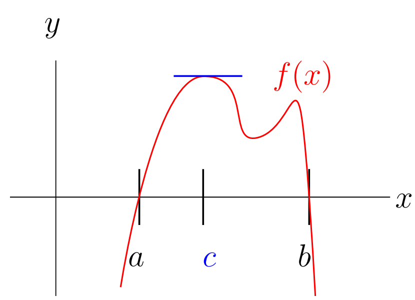
So the absolute maximum of f on $[0, 2]$ is 1, which occurs at $x = 0$. And the absolute minimum of f on $[0, 2]$ is $-5/4$ which occurs at $x = 3/2$.



Example 3.16. Let's find the extreme values of $f(x) = x^3/6$ on the interval $[-\pi/2, \pi/2]$. We have $f'(x) = x^2/2$, so that $f'(x) = 0$ only when $x = 0$. So, $x = 0$ is the only critical point of f. So, the extreme values of f must occur at elements of the following list: $-\pi/2, 0, \pi/2$. We now check the values of f at these points. We have $f(0) = 0$, $f(\pi/2) = \pi^3/48$, and $f(-\pi/2) = -\pi^3/48$.

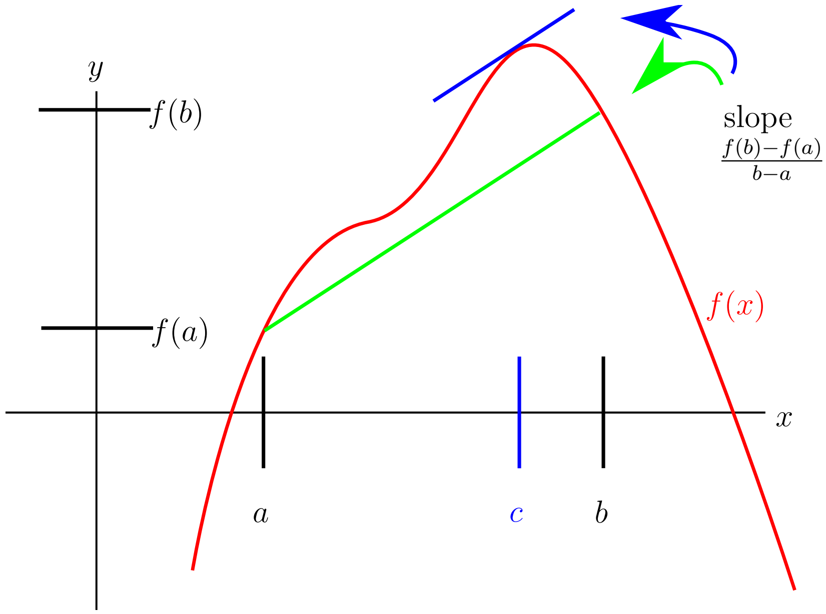
So the absolute maximum of f on $[-\pi/2, \pi/2]$ is $\pi^3/48$, which occurs at $x = \pi/2$. And the absolute minimum of f on $[-\pi/2\pi/1]$ is $-\pi^3/48$ which occurs at $x = -\pi/2$.

3.3. **Mean Value Theorem.** The following theorems allow us to better understand the graphs of functions.



Theorem 3.17 (**Rolle's Theorem**). *Let $f\colon [a, b] \to \mathbb{R}$ be continuous function that is differentiable on (a, b) with $f(a) = f(b) = 0$. Then there exists c with $c \in (a, b)$ and $f'(c) = 0$.*

Proof. From the Extreme Value Theorem, let c be an extreme value of f on $[a, b]$. If an extreme value c occurs in (a, b), then $f'(c) = 0$ by Theorem 3.13. If not, both the max and min occur at the endpoints a, b, which implies that f is a constant function. But if f is a constant, then $f' = 0$ everywhere. So, in any case, there is a $c \in (a, b)$ with $f'(c) = 0$. $\square$



Theorem 3.18. ***(Mean Value Theorem)*** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous function that is differentiable on (a,b). Then there exists c with $c \in (a,b)$ and*

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

Proof. Let $g(x) = f(x) - f(a) - (x-a)((f(b) - f(a))/(b-a))$. Then $g(a) = g(b) = 0$. So, by applying Rolle's Theorem, Theorem 3.17, to g, there exists $c \in (a,b)$ such that

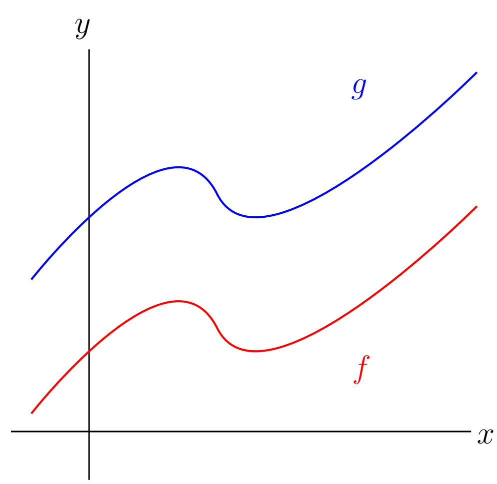
$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}.$$

That is, $f'(c) = (f(b) - f(a))/(b-a)$. □

Remark 3.19. Rolle's Theorem corresponds to the case $f(b) = f(a)$ in the Mean Value Theorem.

Example 3.20. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function such that $f(0) = 0$ and $f(1) = 1$. Then there is some point $x \in (0,1)$ such that $f'(x) = (f(1) - f(0))/(1-0) = 1$.

Recall that a constant function has a derivative that is zero. The converse is also true, though we technically did not know it to be true until now.



Corollary 3.21 (Functions with Zero Derivative are Constant). *Let $a < b$. Let $f\colon (a,b) \to \mathbb{R}$ be differentiable with $f'(x) = 0$ for all $x \in (a,b)$. Then there is a constant C such that $f(x) = C$ for all $x \in (a,b)$.*

Proof. If there is some $d \in (a,b]$ with $f(d) \neq f(a)$, then the Mean Value Theorem says that there is some $c \in (a,d)$ with $f'(c) =$

$(f(d) - f(a))/(d - a) \neq 0$, a contradiction. So, we must have $f(d) = f(a)$ for all $d \in (a, b]$. That is, f is a constant function. $\square$

Corollary 3.22 (Two Functions with the Same Derivative Differ by a Constant). *Let $a < b$. Let $f, g\colon (a, b) \to \mathbb{R}$ be differentiable with $f'(x) = g'(x)$ for all $x \in (a, b)$. Then there is a constant C such that $f(x) = g(x) + C$ for all $x \in (a, b)$.*

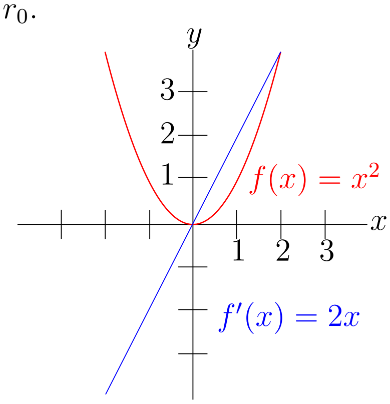
Proof. Apply the previous Corollary to the function $h(x) = f(x) - g(x)$. Note that $h'(x) = 0$, so $h(x) = C$ for some constant C. $\square$

Example 3.23. Let's find the function f such that $f'(x) = e^x$ and such that $f(0) = 2$.

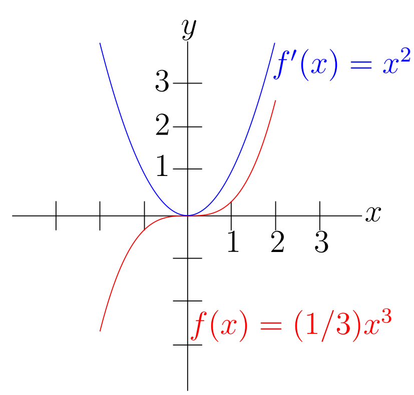
Recall that $(d/dx)e^x = e^x$. So, from the Corollary above, we must have $f(x) = e^x + C$ for some constant C. Since $f(0) = e^0 + C = 1 + C = 2$, we have $C = 1$. So, $f(x) = e^x + 1$.

Example 3.24. Recall our example of an idealized trajectory. Suppose you throw a baseball vertically in the air, with initial upward velocity v_0 and initial position r_0 (and we neglect air friction). Let $r(t)$ denote the position of the baseball (in meters) at time t (in seconds). We know that the acceleration due to gravity is constant, so that $r''(t) = -9.8$. That is, $(d/dt)r'(t) = -9.8$. The function $-9.8t$ also satisfies $(d/dt)(-9.8t) = -9.8$. Therefore, $r'(t) = -9.8t + C$ for some constant C. Since $r'(0) = v_0 = C$, we conclude that $r'(t) = -9.8t + v_0$. Also, the function $(-9.8/2)t^2 + v_0 t$ satisfies $(d/dt)[(-9.8/2)t^2 + v_0 t] = -9.8t + v_0$. So, there is some constant C_1 such that $r(t) = (-9.8/2)t^2 + v_0 t + C_1$. Since $r(0) = C_1 = r_0$, we have found that

$$r(t) = (-9.8/2)t^2 + v_0 t + r_0.$$



By examining the first and second derivatives of a function, we can say a lot about the properties of that function. In the previous section, we saw that the zeros of the first derivative tell us most of the information about the maximum and minimum values of a given function. To see what else the derivatives tell us, see the JAVA Applet, graph features. Below, we will describe several tests on the first and second derivatives that allow us to find several properties of a function, as in this Applet.



Definition 3.25. Let $f\colon (a, b) \to \mathbb{R}$. If $f(x) < f(y)$ whenever $x < y$ with $x, y \in (a, b)$, we say that f is **increasing** on (a, b). If $f(x) > f(y)$ whenever $x < y$ with $x, y \in (a, b)$, we say that f is **decreasing** on (a, b). If f is increasing on (a, b), or if f is decreasing on (a, b), we say that f is **monotonic**.

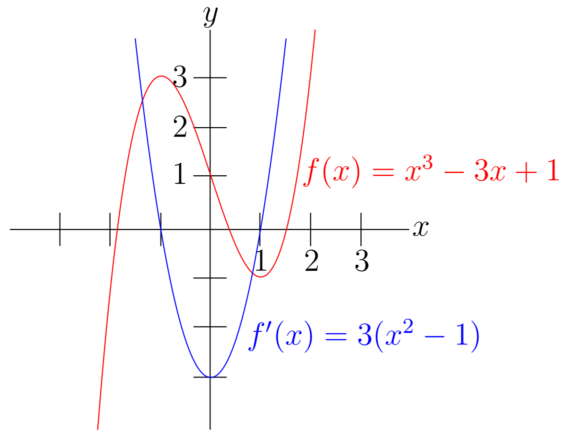
Proposition 3.26 (Increasing/Decreasing Test). *Let $f\colon (a, b) \to \mathbb{R}$ be differentiable.*

- *If $f'(x) > 0$ for all $x \in (a, b)$, then f is increasing on (a, b).*

- *If $f'(x) < 0$ for all $x \in (a,b)$, then f is decreasing on (a,b).*

Proof. We prove the first assertion by contradiction. Suppose f is not increasing but $f' > 0$ on (a,b). Then there are $x, y \in (a,b)$ with $x < y$ but $f(x) \geq f(y)$. That is, $f(y) - f(x) \leq 0$. By the Mean Value Theorem, there is some $c \in (x,y)$ with $f'(c) = (f(y) - f(x))/(y-x) \leq 0$, a contradiction. We conclude that f is increasing. The second assertion is proven similarly. $\square$

The following example applies the increasing/decreasing test.



Example 3.27. Consider $f(x) = x^3 - 3x + 1$. We have $f'(x) = 3x^2 - 3 = 3(x^2 - 1) = 3(x+1)(x-1)$. So, when $x < -1$, $f'(x) > 0$ and f is increasing; when $-1 < x < 1$, $f'(x) < 0$ and f is decreasing; and when $x > 1$, $f'(x) > 0$ and f is increasing.

Proposition 3.28 (First Derivative Test). *Let $c \in (a,b)$ be a critical point for a continuous function $f\colon (a,b) \to \mathbb{R}$. Assume that f is differentiable on (a,c) and on (c,b).*

- *If $f'(x) > 0$ on (a,c), and if $f'(x) < 0$ on (c,b), then f has a local maximum at $x = c$.*
- *If $f'(x) < 0$ on (a,c), and if $f'(x) > 0$ on (c,b), then f has a local minimum at $x = c$.*
- *If $f'(x) > 0$ on $(a,c) \cup (c,b)$, or if $f'(x) < 0$ on $(a,c) \cup (c,b)$, then f does not have a local maximum or a local minimum at $x = c$.*

Proof. Suppose $f'(x) > 0$ on (a,c) and $f'(x) < 0$ on (c,b). Then f is increasing on (a,c) and decreasing on (c,b). So, $f'(c)$ is a local maximum. The other assertions are proven similarly. $\square$

Example 3.29. Consider again $f(x) = x^3 - 3x + 1$. Recall that when $x < -1$, $f'(x) > 0$; when $-1 < x < 1$, $f'(x) < 0$; and when $x > 1$, $f'(x) > 0$. So, $x = -1$ is a local maximum, and $x = 1$ is a local minimum.

Example 3.30. Let $f(x) = x^3$. Then $f'(x) = 3x^2$. So, $f'(0) = 0$, but $f'(x) > 0$ for $x \neq 0$. So, 0 is not a local extremum of f.

3.4. **Graph Sketching.**

Definition 3.31 (Concavity). Let $f\colon (a,b) \to \mathbb{R}$ be a differentiable function.

- If f' is increasing on (a,b), then f is **concave up**. That is, if $y(x) = ax + b$ denotes a tangent line to f, then $f(x) \geq y(x)$ for all $x \in (a,b)$.
- If f' is decreasing on (a,b), then f is **concave down**. That is, if $y(x) = ax + b$ denotes a tangent line to f, then $f(x) \leq y(x)$ for all $x \in (a,b)$.

Example 3.32. Let $f(x) = x^2$. Then $f'(x) = 2x$ is increasing, so f is concave up.
Let $f(x) = -x^2$. Then $f'(x) = -2x$ is decreasing, so f is concave down.

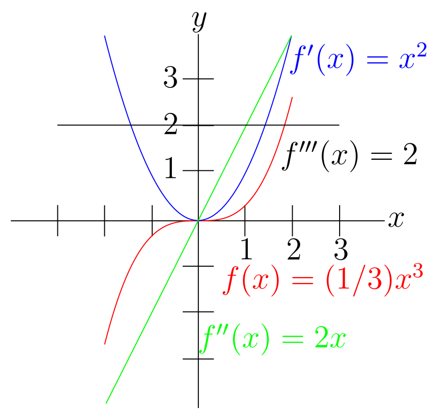
If f'' exists, and if $f'' > 0$ then f' is increasing, and if $f'' < 0$ then f' is decreasing. We therefore get the following test.

Proposition 3.33 (**Concavity Test**)**.** *Let $f\colon (a,b) \to \mathbb{R}$ be a differentiable function such that f'' exists on (a,b).*

- *If $f''(x) > 0$ on (a,b), then f is concave up on (a,b)*
- *If $f''(x) < 0$ on (a,b), then f is concave down on (a,b)*

Definition 3.34 (**Inflection Point**)**.** Let $f\colon (a,b) \to \mathbb{R}$ be a differentiable function and let $c \in (a,b)$. If f is concave up on one side of c, and if f is concave down on the other side of c, then c is called an **inflection point**. If f'' exists on (a,b), then an inflection point occurs when f'' changes sign.

The prototypical example of an inflection point is $x = 0$ for the function $f(x) = x^3$.



Example 3.35. Let $f(x) = x^3$. Then $f''(x) = 6x$, so f'' changes sign at $x = 0$. That is, $x = 0$ is an inflection point of f. Also, $f''(x) < 0$ when $x < 0$, so f is concave down when $x < 0$. Similarly, $f''(x) > 0$ when $x > 0$ so f is concave up when $x > 0$.

Proposition 3.36 (**Second Derivative Test**)**.** *Let $f\colon (a,b) \to \mathbb{R}$. Let $c \in (a,b)$. Assume that $f'(c)$ and $f''(c)$ exist. Assume also that $f'(c) = 0$, and that f'' is continuous near c.*

(1) *If $f''(c) > 0$, then f has a local minimum at c.*
(2) *If $f''(c) < 0$, then f has a local maximum at c.*
(3) *If $f''(c) = 0$, then f may or may not have a local extremum at c.*

Example 3.37. Let $f(x) = x^2$. Then $f'(0) = 0$ and $f''(0) = 2 > 0$. So, f has a local minimum at $x = 0$.

Let $f(x) = -x^2$. Then $f'(0) = 0$ and $f''(0) = -2 < 0$. So, f has a local maximum at $x = 0$.

The functions $f(x) = x^4$, $f(x) = -x^4$ and $f(x) = x^3$ all satisfy $f'(0) = 0$ and $f''(0) = 0$, though they have a local minimum, maximum and neither, respectively.

Proof. We only prove (1), since (2) is proven similarly. Since f'' is continuous near c, the definition of continuity says that there is a small interval containing c where $f''(x) > 0$. Then the Concavity Test, Proposition 3.33, says that $f(c+h) \geq f'(c)h + f(c)$ for all sufficiently small h. Since $f'(c) = 0$, we have $f(c+h) \geq f(c)$, so c is a local minimum of f. $\square$

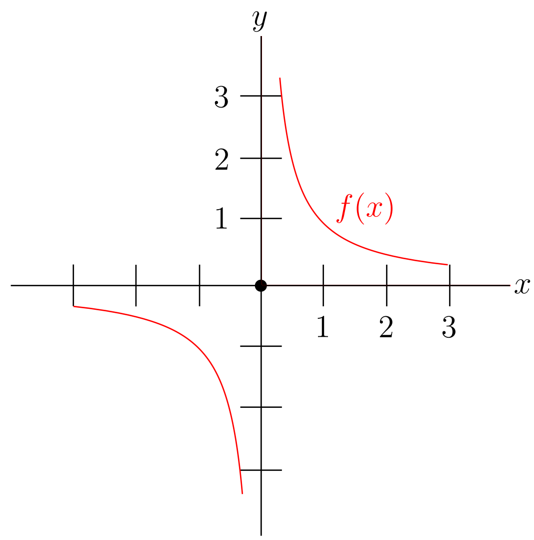
Example 3.38. Consider the function $f(x) = x^3/3 - x - 1$. Identify all local maxima, minima and inflection points. Identify where f is increasing and decreasing. Identify where f is concave up and concave down. Then, sketch the function f.

We have $f'(x) = x^2 - 1 = (x+1)(x-1)$ and $f''(x) = 2x$. So, $x = 0$ is an inflection point and $x = 1, -1$ are critical points. The function f is increasing when $x < -1$, decreasing when $-1 < x < 1$ and increasing when $x > 1$. So, $x = -1$ is a local maximum and $x = 1$ is a local minimum. Alternatively, $f''(-1) < 0$ and $f''(1) > 0$, which again implies that $x = -1$ is a local max and $x = 1$ is a local min. Lastly, $f''(x) < 0$ when $x < 0$ and $f''(x) > 0$ when $x > 0$. So, f is concave down when $x < 0$, and f is concave up when $x > 0$.

Definition 3.39 (**Asymptotes**)**.** Let f be a function and let L be a constant. A horizontal line $y = L$ is called a **horizontal asymptote** of f if $\lim_{x\to\infty} f(x) = L$ or if $\lim_{x\to-\infty} f(x) = L$. A vertical line $x = L$ is called a **vertical asymptote** if $\lim_{x\to L^+} f(x) = \pm\infty$ or if $\lim_{x\to L^-} f(x) = \pm\infty$.

Example 3.40. Let's identify all asymptotes of the function $f(x) = 1/x$.

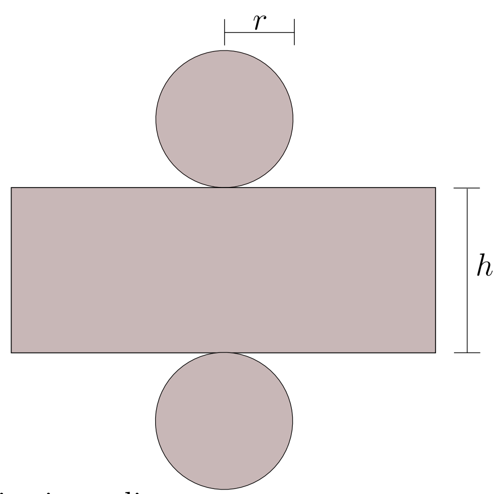
We have $\lim_{x\to\pm\infty} f(x) = 0$. Therefore, $y = 0$ is a horizontal asymptote of f. Also, $\lim_{x\to 0^+} f(x) = \infty$ and $\lim_{x\to 0^-} f(x) = -\infty$. So, $x = 0$ is a vertical asymptote. At all other points, f is a bounded function, so these are the only asymptotes of f.



To sketch the graph of the function f, note that $f'(x) = -x^{-2}$ and $f''(x) = 2x^{-3}$. So, f is concave up when $x > 0$ and concave down when $x < 0$.

3.5. **Applied Optimization.** Optimization problems occur in all applications of mathematics. Here is a simplified example.

Example 3.41. Suppose we want to design a cylindrical soda can with a minimal amount of material. The can's volume is 1 liter (1000 cm^3), and it will be made from aluminum of a fixed thickness. What dimensions should the can have?



Suppose the can has radius r and height h where $r, h > 0$. Then the volume of the can is $\pi r^2 h$. And the surface area of the can is $2\pi r^2 + 2\pi r h$. So, we have $\pi r^2 h = 1000$, or $h = 1000/(\pi r^2)$. And we want to minimize the surface area

$$f(r) = 2\pi r^2 + 2000/r, \qquad r > 0.$$

We look for critical points of f. We have $f'(r) = 4\pi r - 2000r^{-2}$. So, $f'(r) = 0$ when $4\pi r = 2000r^{-2}$, i.e. when $r^3 = 500/\pi$, so that $r = (500/\pi)^{1/3}$. So, a critical point occurs with radius $r = (500/\pi)^{1/3}$ and height $h = 1000/(\pi r^2) = 1000/(\pi^{1/3}(500)^{2/3}) = 2(500/\pi)^{1/3}$. That is, the critical soda can has a height which is twice its radius.

Note that $f'(r) < 0$ when $0 < r < (500/\pi)^{1/3}$ and $f'(r) > 0$ when $r > (500/\pi)^{1/3}$. So, the value $r = (500/\pi)^{1/3}$ is an absolute minimum of f.

Note also that this optimization problem has no absolute maximum, since $\lim_{r\to 0} f(r) = \infty$ and $\lim_{r\to\infty} f(r) = \infty$.

An algorithm for solving optimization problems can be described as follows.

Algorithm 1.

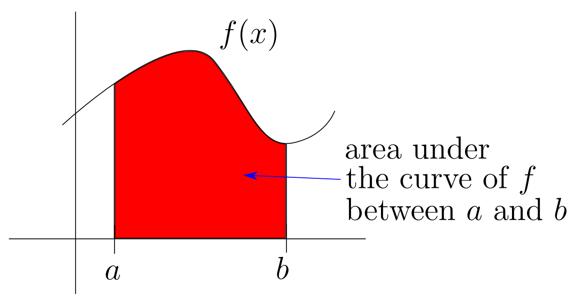
- Introduce variables, and introduce a function f to be optimized.
- Identify the domain of the optimization. Then, apply our usual optimization procedure:

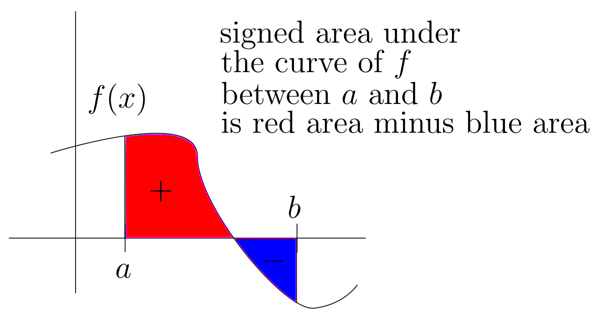
- Find critical points of f in the domain of f.
- Test the critical points of f, and check the endpoints of the domain of f.
- Choose the largest and smallest values of f from these points.

Example 3.42. Find two numbers which sum to 50 and whose product is a maximum.

Given two numbers x, y such that $x + y = 50$, we want to maximize the product xy. Since $y = 50 - x$, we want to maximize $f(x) = xy = x(50 - x)$ over all of $x \in \mathbb{R}$. We check for critical points. We have $f(x) = -x^2 + 50x$, so $f'(x) = -2x + 50$. And $f'(x) = 0$ when $x = 25$. So, the only critical points occurs when $x = 25$. At this point, we have $y = 50 - x = 25$. Note that $f'(x) > 0$ when $x < 25$ and $f'(x) < 0$ when $x > 25$. So, f has an absolute maximum at $x = 25$, and it is unnecessary to check the endpoints of the domain. Note however that $\lim_{x\to\pm\infty} f(x) = -\infty$, so f has no absolute minimum.

4. The Integral





Along with the derivative, the integral is one of the two most fundamental concepts that we find in Calculus. Unfortunately, the formal definition of the integral is more complicated than that of the derivative. However, we should still try to understand these formal definitions, since the ideas that go into the construction of the derivative and the integral are pervasive throughout mathematics and the sciences. In the case of the integral, the quantity $\int_a^b f(x)dx$ intuitively represents the area under the curve $y = f(x)$ on the interval $[a, b]$ (if f is positive on the interval $[a, b]$).

Ultimately, we want to find the area under any given curve. The strategy is similar in spirit to our construction of the derivative. We would like to perform some process that requires infinitely many steps, and as noted by Zeno, doing so does not make any sense. To resolve this issue, we approximate some infinite thing by a finite number of steps. And we hope that, as our approximation gets "finer," some number will approach some limit.

Using this paradigm, we first approximate the area under a given curve by a finite number of rectangles. We know the area of a rectangle, so we therefore know the area of several non-overlapping rectangles. We then want to make our approximation of rectangles finer and finer, and then take some limit. If we complete this process in the right way, and if our curve is nice enough, then this limit will exist. Unfortunately, the limit of the sum of the areas of these rectangles may not always exist, so we have to be careful in our construction of the integral. So, although the notation below and the details may appear pedantic or unnecessary, these things really are necessary in order to get a sensible answer in the end.

4.1. **Summation Notation.** Below, we will be using summation notation often.

Definition 4.1 (Summation Notation). Let $y_1, y_2, \ldots, y_n$ be a set of numbers. We define

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + \cdots + y_{n-1} + y_n.$$

Example 4.2. $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + (n-1) + n = n(n+1)/2.$

Example 4.3. $\sum_{i=1}^{n} i^2 = 1 + 4 + 9 + \cdots + (n-1)^2 + n^2 = n(n+1)(2n+1)/6.$

Example 4.4. If $y_i = (-1)^i$, then

$$\sum_{i=1}^{n} y_i = -1 + 1 - 1 + 1 - 1 + \cdots + (-1)^{n-1} + (-1)^n = \begin{cases} -1 & \text{, if } n \text{ is odd} \\ 0 & \text{, if } n \text{ is even.} \end{cases}$$

Proposition 4.5 (Properties of Finite Sums). *Let $y_1, \ldots, y_n$ be a set of numbers, and let $z_1, \ldots, z_n$ be a set of numbers. Let c be a constant. Then*

- $\sum_{i=1}^{n}(y_i + z_i) = (\sum_{i=1}^{n} y_i) + (\sum_{i=1}^{n} z_i)$.
- $\sum_{i=1}^{n}(y_i - z_i) = (\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} z_i)$.
- $\sum_{i=1}^{n} cy_i = c\sum_{i=1}^{n} y_i$.
- $\sum_{i=1}^{n} c = cn$.

Example 4.6. $\sum_{i=1}^{n}(3i - i^2) = 3\left(\sum_{i=1}^{n} i\right) - \left(\sum_{i=1}^{n} i^2\right)$.

4.2. **The Definite Integral.** Let $a < b$. Let $f\colon [a,b] \to \mathbb{R}$ be a positive function. We would like to compute the area under the curve of f on the interval $[a,b]$. We will eventually do this, but for now we will settle for an approximation. We will first approximate this area by a set of rectangles.

Definition 4.7 (Riemann Sums). Let $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$. The **Riemann sum** of f on $[a,b]$ evaluated at the right endpoints of the rectangles is the quantity

$$\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i).$$

The **Riemann sum** of f on $[a,b]$ evaluated at the left endpoints is the quantity

$$\sum_{i=1}^{n}(x_i - x_{i-1})f(x_{i-1}).$$

The **Riemann sum** of f on $[a,b]$ evaluated at the midpoints is the quantity

$$\sum_{i=1}^{n}(x_i - x_{i-1})f\left(\frac{x_{i-1} + x_i}{2}\right).$$

In each case, we are approximating the area under the curve f by a set of rectangles. For the Riemann sum evaluated at the right endpoints, the quantity $(x_i - x_{i-1})$ is the width of the i^{th} rectangle, and the quantity $f(x_i)$ is the height of the i^{th} rectangle, where $1 \leq i \leq n$.

Each of these Riemann sums provides a good approximation to the area under the curve of f when n is large. In fact, some of these Riemann sums even have a limiting value as $n \to \infty$, if we make the right choices for the points $x_0, \ldots, x_n$.

Example 4.8. Let $f(x) = x$ and consider the interval $[0,1]$. We know that the area under the curve of f on $[0,1]$ is a triangle of area $1/2$. Let's show that the Riemann sums defined above converge to $1/2$ as $n \to \infty$, if we choose $x_0 = 0$, $x_1 = 1/n$, $x_2 = 2/n$, $x_3 = 3/n$, $\ldots$, $x_{n-1} = (n-1)/n$, $x_n = 1$. That is, we have $x_i = i/n$ for all $i \in \{1, \ldots, n\}$. With these choices, we always have $x_i - x_{i-1} = i/n - (i-1)/n = 1/n$, for any $i \in \{1, \ldots, n\}$. So, the Riemann sum evaluated at the right endpoints is equal to

$$\sum_{i=1}^{n} (x_i - x_{i-1}) f(x_i) = \sum_{i=1}^{n} \frac{1}{n} f(i/n) = \frac{1}{n} \sum_{i=1}^{n} f(i/n) = \frac{1}{n} \sum_{i=1}^{n} i/n = \frac{1}{n} \frac{n(n+1)}{2n} = \frac{n+1}{2n}.$$

So,

$$\lim_{n\to\infty} \sum_{i=1}^{n} (x_i - x_{i-1}) f(x_i) = \lim_{n\to\infty} \frac{n+1}{2n} = \frac{1}{2}.$$

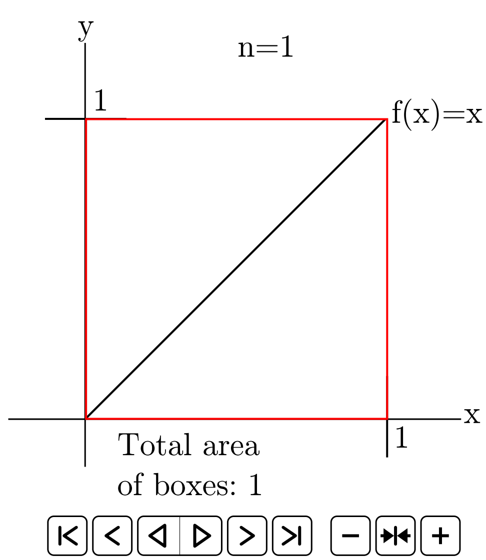
Similarly,

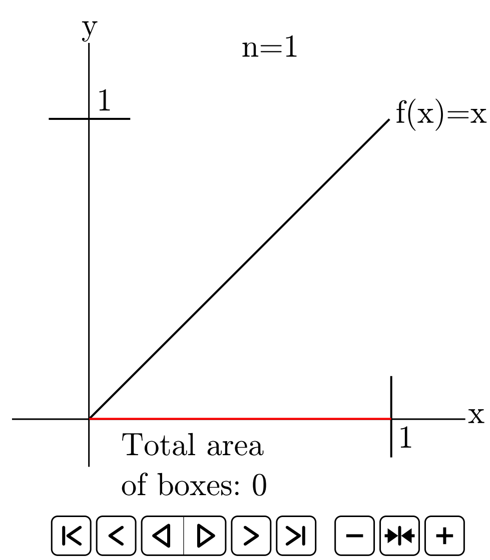
$$\lim_{n\to\infty} \sum_{i=1}^{n} (x_i - x_{i-1}) f(x_{i-1}) = \frac{1}{2}$$

and

$$\lim_{n\to\infty} \sum_{i=1}^{n} (x_i - x_{i-1}) f\left(\frac{x_{i-1} + x_i}{2}\right) = \frac{1}{2}.$$

That is, each Riemann sum approaches the area under the curve, as $n \to \infty$. More precisely, the Riemann sum approaches the area under the curve, as the maximum spacing between the points goes to zero. That is, as $\max_{i=1,\ldots,n}(x_i - x_{i-1}) \to 0$, the Riemann sum approaches the area under the curve.





Example 4.9. Let $f(x) = x^2$ and consider the interval $[0,1]$. Let's show that the Riemann sums defined above converge as $n \to \infty$, if we choose $x_0 = 0$, $x_1 = 1/n$, $x_2 = 2/n$, $x_3 = 3/n$, $\ldots$, $x_{n-1} = (n-1)/n$, $x_n = 1$. That is, we have $x_i = i/n$ for all $i \in \{1, \ldots, n\}$. With these choices, we always have $x_i - x_{i-1} = i/n - (i-1)/n = 1/n$, for any $i \in \{1, \ldots, n\}$. So, the

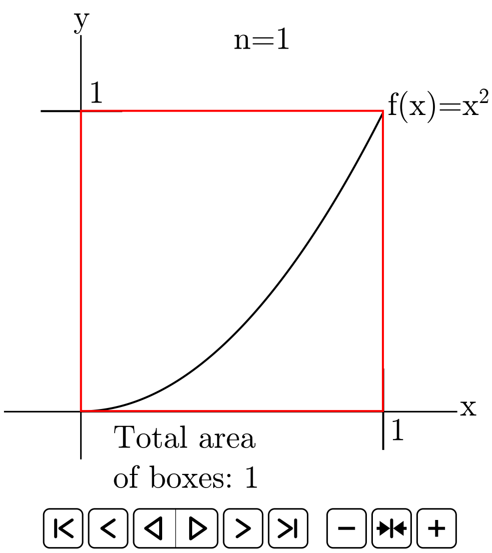
Riemann sum evaluated at the right endpoints is equal to

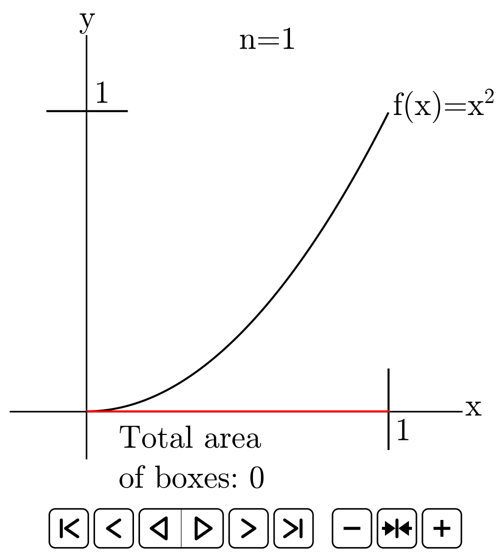
$$\begin{aligned}\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i) &= \sum_{i=1}^{n}\frac{1}{n}f(i/n) = \frac{1}{n}\sum_{i=1}^{n}f(i/n) = \frac{1}{n}\sum_{i=1}^{n}i^2/n^2 \\ &= \frac{1}{n}\frac{n(n+1)(2n+1)}{6n^2} = \frac{(n+1)(2n+1)}{6n^2}.\end{aligned}$$

So,

$$\lim_{n\to\infty}\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i) = \lim_{n\to\infty}\frac{(n+1)(2n+1)}{6n^2} = \lim_{n\to\infty}\frac{2n^2+3n+1}{6n^2} = \frac{2}{6} = \frac{1}{3}.$$

That is, as $\max_{i=1,\ldots,n}(x_i - x_{i-1}) \to 0$, the Riemann sum approaches the the value $1/3$, which is presumably the area under the curve of f on the interval $[0,1]$.





We recommend seeing this picture in action with the help of the JAVA applet, Riemann sums.

Definition 4.10 (**Riemann Sums**)**.** A general **Riemann sum** of f on $[a,b]$ is defined as follows. Let $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$. Let $c_1 \in [x_0, x_1]$, $c_2 \in [x_1, x_2]$, $\ldots$, $c_n \in [x_{n-1}, x_n]$. A general Riemann sum is any sum of the form

$$\sum_{i=1}^{n}(x_i - x_{i-1})f(c_i).$$

Example 4.11. Choosing $c_i = x_i$, or $c_i = x_{i-1}$ or $c_i = (x_{i-1}+x_i)/2$ yields the right endpoint, left endpoint, and midpoint Riemann sums, respectively.

Definition 4.12. Let $a < b$, and let $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. The maximum width of the rectangles of the Riemann sum is denoted by

$$\max_{i=1,\ldots,n}(x_i - x_{i-1}).$$

This number is the maximum of the numbers $(x_1 - x_0), (x_2 - x_1), (x_3 - x_2), \ldots, (x_n - x_{n-1})$. If $\max_{i=1,\ldots,n}(x_i - x_{i-1})$ is small, then the partition is very fine. More specifically, all of our approximating rectangles will have small width. In order to construct the integral, we will let $\max_{i=1,\ldots,n}(x_i - x_{i-1})$ approach zero.

We can finally define the Definite Integral.

Definition 4.13 (The Definite Integral). Let $f\colon [a,b] \to \mathbb{R}$. If the following limit exists, we say that f is integrable on $[a,b]$.

$$\int_a^b f(x)dx = \lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1}) f(c_i).$$

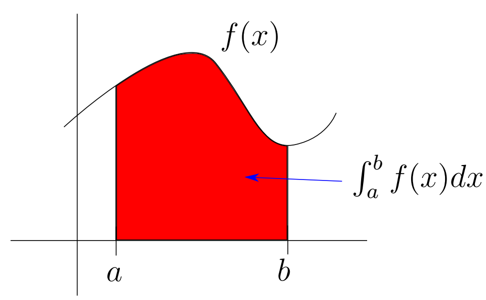
Remark 4.14. We now describe the limit appearing in the Definite Integral more explicitly.

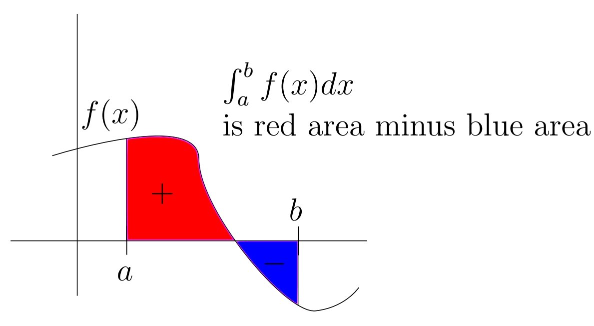
For $\lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1}) f(c_i)$ to exist and be equal to I, we mean the following. For every $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon) > 0$ such that: for any choices of $a = x_0 < x_1 < \cdots < x_n = b$, and for any choices of $c_i \in [x_{i-1}, x_i]$, as long as $\max_{i=1,\ldots,n}(x_i - x_{i-1}) < \delta$, we have

$$\left| \sum_{i=1}^n (x_i - x_{i-1}) f(c_i) - I \right| < \varepsilon.$$

That is, for the limit $\lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1}) f(c_i)$ to exist, we require that any sufficiently fine partition has a Riemann sum that is close to the value I.

The following terminology will be used frequently below.



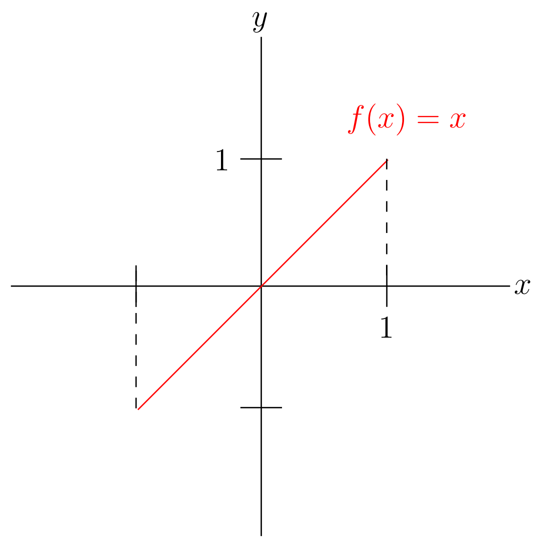


Remark 4.15.

- We refer to $\int_a^b f(x)dx$ as the **integral** of f on $[a,b]$.
- The function f inside the integral is called the **integrand**.
- The numbers a, b representing the interval $[a,b]$ are called the **limits of integration**. In the integral $\int_a^b f(x)dx$, the variable x is just a placeholder, which has no intrinsic meaning. For example, we could just as easily write the integral of f on $[a,b]$ as $\int_a^b f(z)dz$ or $\int_a^b f(s)ds$.

Remark 4.16 (Geometric Interpretation of the Integral). If $f\colon [a,b] \to \mathbb{R}$ has $f(x) \geq 0$ for all $x \in [a,b]$, then $\int_a^b f(x)dx$ represents the area under the curve of f. If $f\colon [a,b] \to \mathbb{R}$ has some negative values, then $\int_a^b f(x)dx$ represents the **signed area** under the curve of f. That is, $\int_a^b f(x)dx$ is the area enclosed by f lying above the x-axis, minus the area enclosed by f lying below the x-axis.

Example 4.17. Let $f(x) = x$. Then $\int_{-1}^{1} f(x)dx = 0$, since the area of f above the x-axis is a triangle of area $1/2$, and the area of f below the x-axis is also a triangle of area $1/2$, so $\int_{-1}^{1} f(x)dx = 1/2 - 1/2 = 0$.



For a function $f: [a, b] \to \mathbb{R}$, we do not yet have a way to determine whether or not $\int_a^b f(x)dx$ exists. Thankfully, if $f: [a, b] \to \mathbb{R}$ is continuous, then $\int_a^b f(x)dx$ exists, as the following very important theorem shows. However, there are situations where the integral of a function does not exist. We will investigate these situations more below. If we understand when integrals do not exist, then our understanding of the integral is improved, just as an understanding of nonexistence of derivatives improves our understanding of derivatives.

Theorem 4.18 (**Continuous Functions on Closed Intervals are Integrable**)**.** *Let $a < b$. Let $f: [a, b] \to \mathbb{R}$ be continuous. Then $\int_a^b f(x)dx$ exists.*

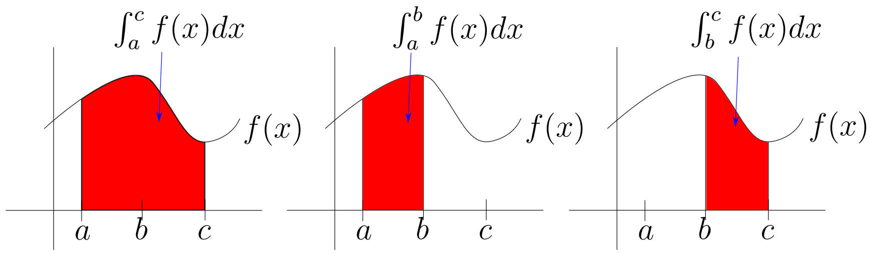
Remark 4.19. We should also mention that the closed interval condition is crucial in Theorem 4.18. For example, $f(x) = 1/x$ is continuous on $(0, 1)$, but $\int_0^1 f(x)dx$ does not exist.

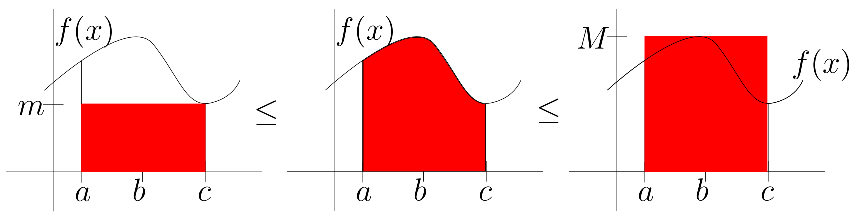
Since a composition of continuous functions is continuous, and a product of continuous functions is continuous, we have the following Corollary of Theorem 4.18.

Corollary 4.20. *Let $f, g: \mathbb{R} \to \mathbb{R}$ be continuous. Then $\int_a^b f(g(x))dx$ exists, and also $\int_a^b f(x)g(x)dx$ exists.*

Proposition 4.21 (**Properties of the Definite Integral**)**.** *Let $a, b, c, k \in \mathbb{R}$, $a < b < c$. Let $f, g: \mathbb{R} \to \mathbb{R}$ be integrable on any closed interval.*

(1) $\int_a^b f(x)dx = -\int_b^a f(x)dx$.
(2) $\int_a^a f(x)dx = 0$.
(3) $\int_a^b k\,dx = k(b - a)$.
(4) $\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$.
(5) $\int_a^b [f(x) - g(x)]dx = \int_a^b f(x)dx - \int_a^b g(x)dx$.
(6) $\int_a^b kf(x)dx = k\int_a^b f(x)dx$.
(7) $\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$.
(8) *If $f \geq 0$, then $\int_a^b f(x)dx \geq 0$.*
(9) *If $f \geq g$, then $\int_a^b f(x)dx \geq \int_a^b g(x)dx$.*
(10) *If $m \leq f \leq M$, then $m(b - a) \leq \int_a^b f(x)dx \leq M(b - a)$.*
(11) $|\int_a^b f(x)dx| \leq \int_a^b |f(x)|\,dx$.





(12) $\int_0^1 f(x)dx = \lim_{n\to\infty} \sum_{i=1}^n \frac{1}{n} f\left(\frac{i}{n}\right)$

(13) $\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^n \left(\frac{b-a}{n}\right) f\left(a + \frac{i(b-a)}{n}\right)$.

Remark 4.22. Property (12) can be used to evaluate certain infinite sums.

All of these properties can be proven in similar ways, by going back to the definition of the integral. Let's just prove property (4) for the sake of illustration.

Proof sketch of (4). Let $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ and let $c_i \in [x_{i-1}, x_i]$ for each $i \in \{1, \ldots, n\}$. Then

$$\sum_{i=1}^n (x_i - x_{i-1})(f(c_i) + g(c_i)) = \sum_{i=1}^n (x_i - x_{i-1}) f(c_i) + \sum_{i=1}^n (x_i - x_{i-1}) g(c_i).$$

So, letting $\max_{i=1,\ldots,n}(x_i - x_{i-1}) \to 0$, and using the limit law for sums,

$$\int_a^b (f(x) + g(x))dx = \lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1})(f(c_i) + g(c_i))$$

$$= \lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1}) f(c_i) + \lim_{\left(\max\limits_{i=1,\ldots,n}(x_i - x_{i-1})\right) \to 0} \sum_{i=1}^n (x_i - x_{i-1}) g(c_i)$$

$$= \int_a^b f(x)dx + \int_a^b g(x)dx.$$

□

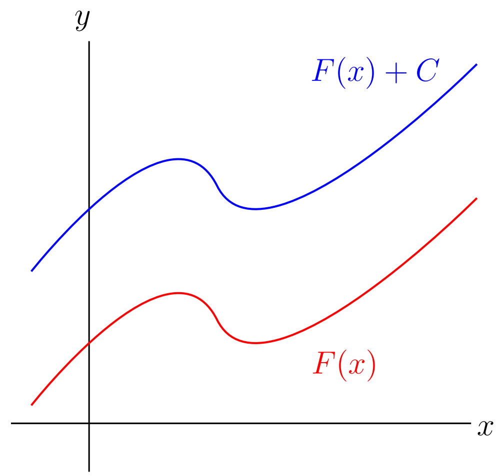
Example 4.23. Suppose $\int_0^2 f(x)dx = 3$ and $\int_0^2 g(x)dx = -2$. Then $\int_0^2 (f(x) + g(x))dx = 3 - 2 = 1$, $\int_0^2 (2f(x) - g(x))dx = 2 \cdot 3 - (-2) = 8$ and $\int_2^0 f(x)dx = -3$.

Example 4.24. Let's use property (10) to estimate the integral $\int_0^2 \sqrt{1 + x^2}dx$. Since $0 \leq x \leq 2$, we have $0 \leq x^2 \leq 4$ and $1 \leq 1 + x^2 \leq 5$. So, $0 \leq \sqrt{1 + x^2} \leq \sqrt{5}$, and by property

(10), we have

$$0 \leq \int_0^2 \sqrt{1+x^2}\, dx \leq \int_0^2 \sqrt{5} dx = 2\sqrt{5}.$$

4.3. **The Indefinite Integral.** So far, we know that a continuous function can be integrated on a closed interval. But we cannot yet compute very many integrals. We now head towards our goal of computing many integrals.



Recall Corollary 3.22: if $f'(x) = g'(x)$ for all $x \in \mathbb{R}$, then there is a constant $C \in \mathbb{R}$ such that $f(x) = g(x) + C$.

Definition 4.25 (**Antiderivative, Indefinite Integral**). Let $f : \mathbb{R} \to \mathbb{R}$. We say that F is an **antiderivative** of f if F is differentiable, and for all $x \in \mathbb{R}$ we have $F'(x) = f(x)$. We then use the notation

$$\int f(x)dx = F(x) + C$$

where C is any constant. We refer to $\int f(x)dx$ as the **indefinite integral** of f.

Remark 4.26. The indefinite integral is not associated to any interval. Also, while the definite integral is a number, the indefinite integral is a function. The definite and indefinite integral are related to each other, as we will see below, but they are not quite the same.

Remark 4.27. Let F, G be antiderivatives of f, so that $F'(x) = G'(x) = f$. From Corollary 3.22, there must be a constant C such that $F(x) = G(x)+C$. So, if we have one antiderivative F of f, then the set of all antiderivatives of f is given by the set of all $F(x) + C$, $C \in \mathbb{R}$.

Example 4.28. Let $f(x) = x^2$. Then the set of all antiderivatives of f is given by $F(x) = (1/3)x^3 + C$, $C \in \mathbb{R}$.

Example 4.29 (**Idealized Trajectories**). Recall our example of idealized trajectories. Suppose I throw a ball straight up in the air at a velocity v_0 m/s, with initial vertical position s_0 meters, ignoring air friction. Suppose the ball has mass m kg. The only acceleration that acts on the ball is a constant acceleration due to gravity, of roughly $a(t) = -9.8$ m/s^2, where t is the time after the ball is thrown, measured in seconds. Taking the antiderivative and using Remark 4.27, the ball must have velocity $v(t) = -9.8t + C$. Since $v(0) = v_0$, we conclude that $v(t) = -9.8t + v_0$. Taking the antiderivative again and using Remark 4.27, the ball must have position $s(t) = -(9.8/2)t^2 + v_0 t + C$. My initial vertical position is s_0, so we conclude that the ball has position

$$s(t) = -4.9t^2 + v_0 t + s_0.$$

For now, antiderivatives may seem a bit strange. Also, if we have a function f, how can we know whether or not an antiderivative exists? It turns out that, if f is continuous, then you can create an antiderivative of f by measuring the areas under the curve f. So, calculating areas under a curve has the "opposite" effect of taking a derivative of a curve. This statement will be made more precise when we state the Fundamental Theorem of Calculus. For now, we give a precise description of how to find the area under a curve, via the Riemann integral.

Theorem 4.30 (**Indefinite Integral of Powers of x**). *Let $n \neq -1$. Then*

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C.$$

Proof. Recall that $\frac{d}{dx}\left(\frac{x^{n+1}}{n+1} + C\right) = \frac{n+1}{n+1}x^n = x^n$. $\square$

Example 4.31. $\int x^3 dx = x^4/4 + C$. $\quad \int x^{3/2} dx = (2/5)x^{5/2} + C$

Proposition 4.32 (**Properties of the Indefinite Integral**). *Let c be a constant and let $f, g\colon \mathbb{R} \to \mathbb{R}$.*

- $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$.
- $\int [f(x) - g(x)]dx = \int f(x)dx - \int g(x)dx$.
- $\int cf(x)dx = c \int f(x)dx$.

Example 4.33.

- $\int e^x dx = e^x + C$.
- $\int e^{cx+d} dx = \frac{1}{c} e^{cx+d} + C$, $c \neq 0$.

4.4. **The Fundamental Theorem of Calculus.** We can finally describe the precise manner in which differentiation and integration "cancel each other out." The following Theorem will also allow us to compute many integrals.

Theorem 4.34 (**Fundamental Theorem of Calculus**). *Let $a < b$.*

(i) *Let $f\colon [a, b] \to \mathbb{R}$ be differentiable. Assume also that $f'\colon [a, b] \to \mathbb{R}$ is continuous. Then*

$$\int_a^b f'(x)dx = f(b) - f(a).$$

(ii) *Let $f\colon [a, b] \to \mathbb{R}$ be continuous. For $x \in (a, b)$ define $g(x) = \int_a^x f(t)dt$. Then g is an antiderivative of f, i.e.*

$$g'(x) = \frac{d}{dx} \int_a^x f(t)dt = f(x).$$

Remark 4.35. Part (i) of Theorem 4.34 can be used to evaluate many different integrals. For example, we have the following two corollaries.

Corollary 4.36 (**Integrating Powers of x**). *Let $n \in \mathbb{R}$, $n \neq -1$, $0 < a < b$. Then*

$$\int_a^b x^n dx = \left[\frac{1}{n+1} x^{n+1}\right]_{x=a}^{x=b} = \frac{1}{n+1}(b^{n+1} - a^{n+1}).$$

Proof. Let $f(x) = (1/(n+1))x^{n+1}$. Note that $f'(x) = x^n$, and then apply the Fundamental Theorem, Theorem 4.34(i) to get

$$\int_a^b x^n dx = \int_a^b f'(x) = f(b) - f(a) = \frac{1}{n+1}(b^{n+1} - a^{n+1}).$$

$\square$

Remark 4.37. What happens if we allow $a = -1$, $b = 1$, $n < 0$?

Proposition 4.38. *Let $a < b$ and let c, d be real numbers with $c \neq 0$. Then*

$$\int e^x dx = e^x + C, \qquad \int_a^b e^x dx = \int_a^b (d/dx)e^x dx = e^b - e^a.$$

$$\int e^{cx+d} dx = c^{-1}e^{cx+d} + C, \qquad \int_a^b e^{cx+d} dx = \int_a^b (d/dx)c^{-1}e^{cx+d} dx = c^{-1}(e^{bc+d} - e^{ac+d}).$$

Example 4.39.

$$\int_0^1 x\,dx = [x^2/2]_{x=0}^{x=1} = 1/2 - 0 = 1/2.$$

$$\int_0^1 x^2\,dx = [x^3/3]_{x=0}^{x=1} = 1/3 - 0 = 1/3.$$

$$\int_1^3 (x^4 - x^{-2})dx = [x^5/5 + x^{-1}]_{x=1}^{x=3} = 3^5/5 + 3^{-1} - 1/5 - 1.$$

$$\int_{-1}^2 e^{2x} dx = [(1/2)e^{2x}]_{x=-1}^{x=2} = (1/2)(e^4 - e^{-2}).$$

Example 4.40.

$$\frac{d}{dx}\int_1^x e^t dt = e^x.$$

$$\frac{d}{dx}\int_2^x \frac{1}{1+t^2}dt = \frac{1}{1+x^2}.$$

$$\frac{d}{dx}\int_1^{x^2} \ln(t)dt = 2x\ln(x^2).$$

For the last example, write $\int_1^{x^2}\ln(t)dt = f(g(x))$, where $f(y) = \int_1^y \ln(t)dt$ and $g(x) = x^2$. Then the Chain Rule says $(d/dx)f(g(x)) = f'(g(x))g'(x)$.

Remark 4.41. Since $(d/dt)\ln t = 1/t$ for $t > 0$ and $\ln(1) = 0$, we have the following formula for the natural logarithm. Let $x > 0$. Then from the fundamental theorem of calculus,

$$\ln(x) = \ln(x) - \ln(1) = \int_1^x \frac{d}{dt}\ln(t)dt = \int_1^x \frac{1}{t}dt.$$

Example 4.42.

$$\int_3^9 \frac{dt}{t} = \ln(9) - \ln(3) = \ln(9/3) = \ln 3.$$

Exercise 4.43. Evaluate

$$\int_{-9}^{-3} \frac{1}{t}dt.$$

Definition 4.44. Let $r > 0$ be an interest rate, let $R(t)$ be an income stream that pays $R(t)$ per year continuously for T years. Then the **present value** of this income stream is

$$\int_0^T R(t)e^{-rt}dt.$$

Example 4.45. Suppose I am earning \$10000 per year continuously for 5 years, and once I get the money, I put it into an interest bearing account with interest rate $r = .03$. We therefore have $R(t) = 10000$, and the present value of this income is

$$\int_0^5 10000e^{-.03t}dt = 10000\frac{1}{-.03}[e^{-.03t}]_{t=0}^{t=5} = 10000\frac{1}{-.03}[e^{-.15} - 1] \approx 46431.$$

That is, it would be equivalent for me to get \$46431 right now and put it into the interest bearing account, rather than continuously putting the money into the account.

Proof of of Theorem 4.34(i). Suppose we have a partition of $[a, b]$. That is, we have

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Then

$$\begin{aligned} f(b) - f(a) &= f(x_n) - f(x_0) \\ &= f(x_n) + [-f(x_{n-1}) + f(x_{n-1})] + \cdots + [-f(x_1) + f(x_1)] - f(x_0) \\ &= [f(x_n) - f(x_{n-1})] + [f(x_{n-1}) - f(x_{n-2})] + \cdots + [f(x_2) - f(x_1)] + [f(x_1) - f(x_0)] \\ &= \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] \end{aligned}$$

By the Mean Value Theorem, there exists $c_i \in [x_i, x_{i-1}]$ such that, for $i = 1, \ldots, n$,

$$(x_i - x_{i-1})f'(c_i) = f(x_i) - f(x_{i-1}).$$

Therefore,

$$f(b) - f(a) = \sum_{i=1}^{n} (x_i - x_{i-1})f'(c_i). \qquad (*)$$

The right side of $(*)$ is a Riemann sum for f'. Since f' is continuous, we know that f' is integrable. So, letting $\max_{i=1,\ldots,n}(x_i - x_{i-1}) \to 0$ and applying the definition of the definite integral, $(*)$ becomes our desired equality:

$$f(b) - f(a) = \int_a^b f'(t)dt.$$

□

Proof sketch of Theorem 4.34(ii). We treat the difference quotient directly. Let $h \in \mathbb{R}$. Then

$$\frac{g(x+h) - g(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)dt.$$

For simplicity, assume that f is differentiable. Then, using the linear approximation of f for values of t near the point x, we have $f(t) \approx f(x) + f'(x)(t - x)$, so

$$\begin{aligned} \frac{g(x+h) - g(x)}{h} &\approx \frac{1}{h}\int_x^{x+h} [f(x) + f'(x)(t - x)]dt = \frac{h}{h}f(x) + f'(x)\frac{1}{h}\int_x^{x+h}(t - x)dt \\ &= f(x) + f'(x)\frac{1}{h}[(t - x)^2/2]_{t=x}^{t=x+h} = f(x) + f'(x)\frac{h^2/2}{h} = f(x) + hf'(x)/2. \end{aligned}$$

So, letting $h \to 0$ we get $\lim_{h\to 0} \frac{g(x+h)-g(x)}{h} = f(x)$. □

4.5. **Integration by Substitution.**

Theorem 4.46 (Change of Variables/ Substitution). *Let $a < b$, $c < d$. Let $g\colon [a,b] \to [c,d]$ be differentiable. Also, suppose that $f\colon [c,d] \to \mathbb{R}$ is continuous. Then*

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(t)dt.$$

Or, in indefinite form, with $u = g(x)$, we have

$$\int f(g(x))g'(x)dx = \int f(u)du.$$

Example 4.47. The following integral may appear difficult if not impossible to evaluate, but Theorem 4.46 allows us to evaluate it. For $x > 0$, let $f(t) = e^t$, and let $g(x) = 1/x$. Applying Theorem 4.46 and then Theorem 4.34(i),

$$\int_1^2 \frac{e^{1/x}}{x^2}dx = -\int_1^2 f(g(x))g'(x)dx = -\int_{g(1)}^{g(2)} e^t dt = -\int_1^{1/2} e^t dt = \int_{1/2}^1 \frac{d}{dt}e^t dt = e - \sqrt{e}.$$

Example 4.48. Let $f(t) = e^t$, and let $g(x) = x^2$. Applying Theorem 4.46 and then Theorem 4.34(i),

$$\begin{aligned}\int_2^3 xe^{x^2}dx &= \frac{1}{2}\int_2^3 f(g(x))g'(x)dx = \frac{1}{2}\int_{g(2)}^{g(3)} f(t)dt = \frac{1}{2}\int_4^9 e^t dt\\ &= \frac{1}{2}\int_4^9 \frac{d}{dt}e^t dt = \frac{1}{2}(e^9 - e^4).\end{aligned}$$

Example 4.49. Let $u = x^2 + 1$ so that $du = 2xdx$, i.e. $xdx = (1/2)du$. Then

$$\int \frac{x}{(x^2+1)^2}dx = \frac{1}{2}\int u^{-2}du = -\frac{1}{2}u^{-1} = -\frac{1}{2(x^2+1)}.$$

And indeed, we can verify that $-\frac{d}{dx}\frac{1}{2(x^2+1)} = \frac{4x}{(2(x^2+1))^2} = \frac{x}{(x^2+1)^2}$.

Example 4.50. Let $u = x + 1$ so that $du = dx$. Then

$$\begin{aligned}\int x\sqrt{x+1}dx &= \int (u-1)u^{1/2}du = \int u^{3/2} - u^{1/2}du\\ &= (2/5)u^{5/2} - (2/3)u^{3/2} = (2/5)(x+1)^{5/2} - (2/3)(x+1)^{3/2}.\end{aligned}$$

And we can verify that $\frac{d}{dx}[(2/5)(x+1)^{5/2} - (2/3)(x+1)^{3/2}] = (x+1)^{3/2} - (x+1)^{1/2} = (x+1-1)\sqrt{x+1} = x\sqrt{x+1}$.

Example 4.51. Consider the function $f(x) = \frac{x}{x^2+1}$. This function is of the form $(1/2)g'(x)/g(x)$ where $g(x) = x^2 + 1$. So, an antiderivative for f is $(1/2)\ln|x^2+1|$.

Consider also the function $f(x) = \tan(x) = \sin(x)/\cos(x)$. This function is of the form $-g'(x)/g(x)$ where $g(x) = \cos(x)$. So, an antiderivative for f is $-\ln|\cos(x)|$.

Proof of Theorem 4.46. For $a < x < b$, define $F(x) = \int_a^x f(t)dt$. From the Fundamental Theorem of Calculus, Theorem 4.34(ii), $F'(x) = f(x)$. Also, by the Chain Rule,

$(d/dx)[F(g(x))] = F'(g(x))g'(x) = f(g(x))g'(x)$. Note also that $f(g(x))g'(x)$ is integrable by Corollary 4.20. Putting everything together, we have

$$\begin{aligned}\int_a^b f(g(x))g'(x)dx &= \int_a^b \frac{d}{dx}[F(g(x))]dx = F(g(b)) - F(g(a)), \qquad \text{by Theorem 4.34(i)}\\ &= \int_{g(b)}^{g(a)} \frac{d}{dx}F(x)dx, \qquad \text{by Theorem 4.34(i)}\\ &= \int_{g(b)}^{g(a)} f(x)dx.\end{aligned}$$

□

4.6. **Average Value.**

Definition 4.52 (Average Value). Let $a < b$. Let $f\colon [a,b] \to \mathbb{R}$. The **average value** of f on the interval $[a,b]$ is defined to be

$$\frac{1}{b-a}\int_a^b f(x)dx.$$

Example 4.53. The average value of $f(x) = x$ on the interval $[0,20]$ is

$$\frac{1}{20}\int_0^{20} xdx = \frac{1}{20}[x^2/2]_0^{20} = \frac{1}{40}(400) = 10.$$

Example 4.54. The average value of $f(x) = x/(x^2+1)^2$ on the interval $[3,6]$ is (using $u = x^2+1$ so $du = 2xdx$, i.e. $xdx = du/2$)

$$\frac{1}{6-3}\int_3^6 \frac{x}{(x^2+1)^2}dx = \frac{1}{3}\int_{10}^{37} \frac{u^{-2}}{2}du = \frac{1}{6}[-u^{-1}]_{u=10}^{u=37} = \frac{1}{6}\left(-\frac{1}{37}+\frac{1}{10}\right)$$

5. Methods of Integration

We now turn our attention to various methods that help to compute integrals. By the end of the chapter, you should feel like you can integrate almost anything.

5.1. **Integration by Parts.**

This first method of integration allows us to essentially move a derivative from one term to another while we are inside an integral.

Theorem 5.1 (Integration by Parts). *Let u, v be continuously differentiable functions on the real line. Let $a < b$. Then*

$$\int_a^b u(x)v'(x)dx = u(b)v(b) - u(a)v(a) - \int_a^b v(x)u'(x)dx.$$

This rule can be memorized by the mnemonic "integral of udv equals uv minus integral of vdu."

The integration by parts formula follows almost immediately from the product rule.

$$(uv)' = u'v + v'u.$$

Integrating both sides on $[a,b]$ and applying the Fundamental Theorem of Calculus to the left side,

$$u(b)v(b) - u(a)v(a) = \int_a^b u'(x)v(x)dx + \int_a^b u(x)v'(x)dx.$$

Example 5.2.

$$\int_a^b xe^x dx = \int_a^b x(d/dx)e^x dx = be^b - ae^a - \int_a^b e^x dx$$
$$= be^b - ae^a - e^b + e^a.$$

Written another way,

$$\int xe^x dx = (x-1)e^x + C.$$

Remark 5.3. If we instead wrote $\int_a^b xe^x dx = \int_a^b \cos e^x (d/dx)(x^2/2)dx$, then things would have just become more complicated. So, we have to choose carefully how to apply integration by parts.

Example 5.4.

$$\int_1^4 \ln x\, dx = \int_1^4 \ln x(d/dx)xdx = 4\cdot\ln 4 - 1\cdot\ln 1 - \int_1^4 x(1/x)dx = 4\ln 4 - 1\ln 1 - 4.$$

Written another way,

$$\int \ln x\, dx = x\ln x - x + C$$

5.2. **Improper Integrals.** It is sometimes nice to integrate a function at infinity. We can do this via the following definition.

Definition 5.5. Fix a real number a. Suppose f is integrable on $[a,b]$ for all $b > a$. The **improper integral** of f on $[a,\infty)$ is defined as the following limit (if the limit exists):

$$\int_a^\infty f(x)dx = \lim_{R\to\infty}\int_a^R f(x)dx.$$

If this limit exists and is finite, we say that the improper integral **converges**. If the limit does not exist, we say that the improper integral **diverges**.

Remark 5.6. We similarly define

$$\int_{-\infty}^a f(x)dx = \lim_{R\to\infty}\int_{-R}^a f(x)dx.$$

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^\infty f(x)dx.$$

Example 5.7. Let $a > 0$. Then

$$\int_0^\infty e^{-at}dt = \lim_{R\to\infty}\int_0^R e^{-at}dt = \lim_{R\to\infty}(-1/a)[e^{-aR} - 1] = 1/a.$$

Theorem 5.8. *Let $a > 0$. If $p > 1$, then*

$$\int_a^\infty \frac{1}{x^p} dx = (p-1)^{-1} a^{-p+1}.$$

If $p \leq 1$, then $\int_a^\infty \frac{1}{x^p} dx$ diverges.

To see this, note that for $p \neq 1$ we have

$$\int_a^\infty \frac{1}{x^p} dx \lim_{R\to\infty} \int_a^R x^{-p} dx = \lim_{R\to\infty} (-p+1)^{-1}[R^{-p+1} - a^{-p+1}]$$

If $p > 1$, then $\lim_{R\to\infty} R^{-p+1} = 0$, so

$$\int_a^\infty \frac{1}{x^p} dx = (p-1)^{-1} a^{-p+1}.$$

If $p < 1$, then $\lim_{R\to\infty} R^{-p+1} = \infty$, so $\int_a^\infty \frac{1}{x^p} dx$ diverges.

In the remaining case $p = 1$, we have

$$\int_a^\infty \frac{1}{x} dx \lim_{R\to\infty} \int_a^R x^{-1} dx = \lim_{R\to\infty} (\ln|R| - \ln|a|).$$

Since $\lim_{R\to\infty} \ln|R| = \infty$, we conclude that $\int_a^\infty \frac{1}{x} dx$ diverges.

It is possible to similarly integrate the discontinuities of integrals, by approaching them in a limiting sense.

Definition 5.9. Let $a < b$. Suppose f is continuous on $[a, b)$ but discontinuous at b. We define the integral of f on $[a, b]$ as the following limit (if the limit exists):

$$\int_a^b f(x) dx = \lim_{R\to b^-} \int_a^R f(x) dx.$$

If this limit exists and is finite, we say that the improper integral **converges**. If the limit does not exist, we say that the improper integral **diverges**.

Remark 5.10. Similarly, if f is continuous on $(a, b]$ but discontinuous at a. We define the integral of f on $[a, b]$ as the following limit (if the limit exists):

$$\int_a^b f(x) dx = \lim_{R\to a^+} \int_R^b f(x) dx.$$

Remark 5.11 (**Integrating over a Discontinuity**)**.** If f is discontinuous only at c, if $a < c < b$, and if $\int_a^c f(x) dx$ and $\int_c^b f(x) dx$ are finite, then we define

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Otherwise, we say that $\int_a^b f(x) dx$ diverges. For example, the integral

$$\int_{-1}^1 \frac{dx}{x}$$

diverges, since $\int_0^1 dx/x$ diverges.

Example 5.12. Let $a > 0$. If $p < 1$, then
$$\int_0^a \frac{dx}{x^p} = (1-p)^{-1}a^{1-p}.$$
If $p \geq 1$, then $\int_0^a dx/x^p$ diverges. To see this, let $p \neq 1$ and observe
$$\int_0^a \frac{dx}{x^p} = \lim_{R\to 0^+} \int_R^a \frac{dx}{x^p} = \lim_{R\to 0^+} (-p+1)^{-1}[a^{-p+1} - R^{-p+1}].$$
If $p < 1$, then $\lim_{R\to 0^+} R^{-p+1} = 0$, so $\int_0^a \frac{dx}{x^p} = (1-p)^{-1}a^{1-p}$. If $p > 1$, then $\lim_{R\to 0^+} R^{-p+1} = \infty$, so $\int_0^a \frac{dx}{x^p}$ diverges. In the remaining case $p = 1$, we have
$$\int_0^a \frac{dx}{x} = \lim_{R\to 0^+} \int_R^a \frac{dx}{x} = \lim_{R\to 0^+} [\ln|a| - \ln|R|] = -\infty.$$
So, the integral $\int_0^a \frac{dx}{x}$ diverges.

5.2.1. *Comparing Integrals.* We have reached somewhat of an end to our investigation of methods of integration. Sometimes, it is actually impossible to get an explicit formula of an integral. Instead, we need to know how to estimate integrals in various ways. In the next section, we will see a few ways that a computer can estimate an integral. For now, we will simply get some rough estimates on integrals. For example, we will just look at a way to check whether or not an improper integral converges.

Proposition 5.13. *Fix a real number a. Suppose $f(x) \geq g(x) \geq 0$ for all $x \geq a$.*

- *If $\int_a^\infty f(x)dx$ converges, then $\int_a^\infty g(x)dx$ converges.*
- *If $\int_a^\infty g(x)dx$ diverges, then $\int_a^\infty f(x)dx$ diverges.*

Example 5.14. We will demonstrate that $\int_1^\infty \frac{dx}{\sqrt{x^3+1}}$ converges.

Let $x \geq 1$. Then $\sqrt{x^3+1} \geq \sqrt{x^3} = x^{3/2} \geq 0$. So, $0 \leq (x^3+1)^{-1/2} \leq x^{-3/2}$. And
$$0 \leq \int_1^\infty x^{-3/2}dx = \lim_{R\to\infty} \int_1^R x^{-3/2}dx = \lim_{R\to\infty} (-2)(R^{-1/2} - 1) = 2.$$
Therefore, $\int_1^\infty \frac{dx}{\sqrt{x^3+1}}$ converges.

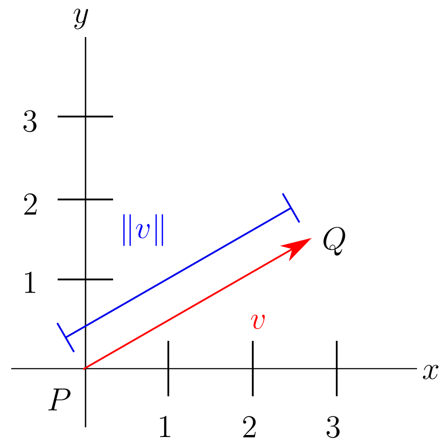
We can do the same test for an endpoint discontinuity as follows.

Example 5.15. We will demonstrate that $\int_0^1 \frac{dx}{x^2+x^8}$ diverges. Let $0 < x \leq 1$. Note that $x^2 + x^8 \leq 2x^2$, so $(x^2+x^8)^{-1} \geq (2x^2)^{-1} \geq 0$. However, we know that $\int_0^1 x^{-2}dx$ diverges. We therefore conclude that $\int_0^1 dx/(x^2+x^8)$ diverges.

6. Vectors

When we change from calculus in one dimension to calculus in two and three dimensions, we naturally need to deal with the extra dimensions. The most convenient way to deal with these extra dimensions is to introduce vectors. We will see that the operations of addition and subtraction on real numbers will extend to vectors. However, we cannot multiply two vectors like they are real numbers. And there are new things to consider for vectors that have no analogue for real numbers. For example, two vectors have an angle between them.

We begin by defining vectors in the plane.



6.1. **Vectors in the Plane.** We denote the plane as the set of all ordered pairs (x, y) where x and y are both real numbers. We use the notation $\mathbb{R}^2$ to denote the plane.

Definition 6.1. A **two-dimensional vector** v is a directed line segment. This line segment has a beginning point P and a terminal point Q, where $P, Q \in \mathbb{R}^2$. Suppose $P = (x_1, y_1)$ and if $Q = (x_2, y_2)$, where x_1, x_2, y_1, y_2 are real numbers. We define the **length** or **magnitude** of the vector v by

$$\|v\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

That is, by the Pythagorean Theorem, $\|v\|$ is the length of the hypotenuse of the right triangle, whose side lengths are $|x_1 - x_2|$ and $|y_1 - y_2|$.

Remark 6.2. Unless otherwise stated, all vectors from now on will have their beginning point P at the origin, so that $P = (0, 0)$. We then will write any two-dimensional vector v in the form $v = (x, y)$. Then

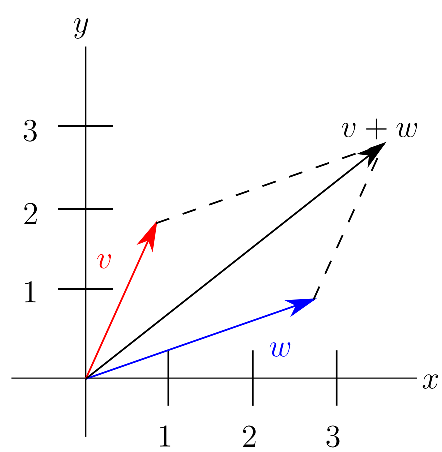
$$\|v\| = \sqrt{x^2 + y^2}.$$

Let λ be a real number. (We often refer to real numbers as **scalars**.) Define $\lambda v = (\lambda x, \lambda y)$. Note that

$$\|\lambda v\| = \sqrt{\lambda^2 x^2 + \lambda^2 y^2} = \sqrt{\lambda^2 (x^2 + y^2)} = |\lambda| \, \|v\| \, .$$

Definition 6.3. Two vectors $v, w \in \mathbb{R}^2$ are said to be **parallel** if there is some scalar $\lambda \in \mathbb{R}$ such that $v = \lambda w$.

We can add vectors together as follows.



Definition 6.4. Suppose $v = (x_1, y_1)$ and $w = (x_2, y_2)$. Define

$$v + w = (x_1 + x_2, y_1 + y_2).$$

Visually, the vector $v + w$ is obtained by taking the vector v, placing the beginning point of w at the endpoint of v, so that the resulting endpoint is that of $v + w$. We also define

$$v - w = (x_1 - x_2, y_1 - y_2).$$

Note that

$$v + (0, 0) = (0, 0) + v = v, \qquad v - v = (0, 0).$$

Example 6.5.

$$(1, 2) + 2(3, 0) = (1, 2) + (6, 0) = (7, 2).$$

Proposition 6.6. *For any vectors $u, v, w \in \mathbb{R}^2$ and for any scalar $\lambda \in \mathbb{R}$,*

$$v + w = w + v.$$

$$u + (v + w) = (u + v) + w.$$

$$\lambda(u + v) = (\lambda u) + (\lambda v).$$

Definition 6.7. A vector $v \in \mathbb{R}^2$ such that $\|v\| = 1$ is called a **unit vector**. If v is any nonzero vector, then the vector

$$\frac{v}{\|v\|}$$

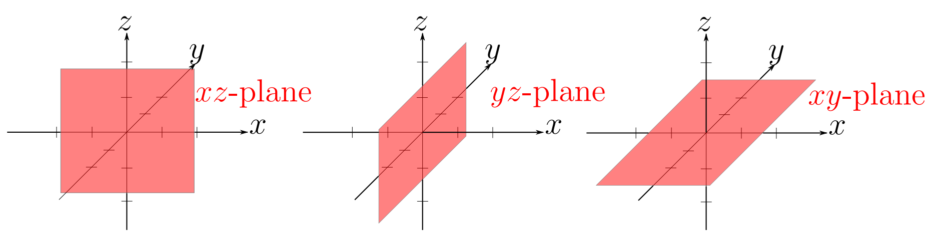
is a unit vector which is parallel to v. To see that $\frac{v}{\|v\|}$ is a unit vector, we use Remark 6.2 to get $\left\|\frac{v}{\|v\|}\right\| = \frac{\|v\|}{\|v\|} = 1$.

Example 6.8. Consider the vector $v = (1, 2)$. Then $\frac{v}{\|v\|} = \frac{(1,2)}{\sqrt{5}} = (1/\sqrt{5}, 2/\sqrt{5})$ is a unit vector, which is parallel to v.

Remark 6.9. Some textbooks use the notation $\vec{i} = (1, 0)$ and $\vec{j} = (0, 1)$, so that a vector $v = (x, y) \in \mathbb{R}^2$ can be written as $v = x\vec{i} + y\vec{j}$. However, we will not use this notation.

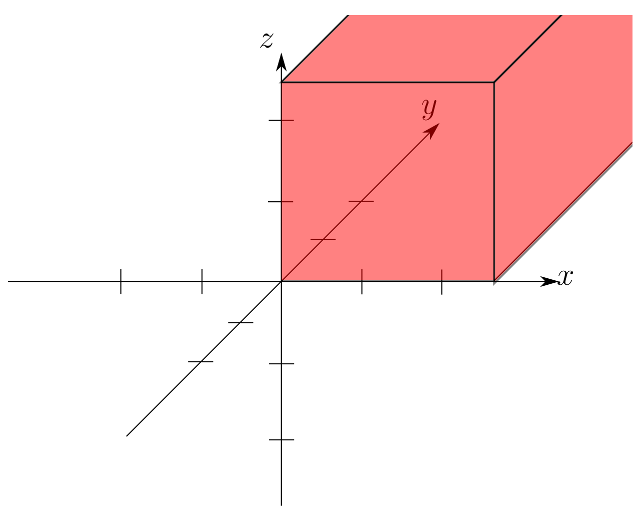
Theorem 6.10 (**Triangle Inequality**). *For any vectors $v, w \in \mathbb{R}^2$, we have*

$$\|v + w\| \leq \|v\| + \|w\|.$$

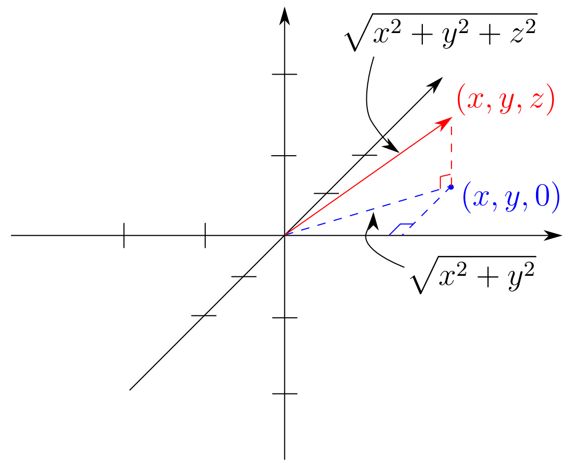


6.2. **Vectors in Three Dimensions.** We denote three-dimensional space as the set of all ordered pairs (x, y, z) where x, y and z are all real numbers. We use the notation $\mathbb{R}^3$ to denote three-dimensional space.

When drawing vectors in the plane, we always use the convention that the x-axis is the horizontal axis, and the y-axis is the vertical axis. We will use a similar convention when plotting vectors in three-dimensional space $\mathbb{R}^3$. However, we now need to orient the axes using the right hand rule. Using your right hand, if the fingers curl from the x-axis to the y-axis, then the z-axis will point in the direction of your extended thumb.



In addition to coordinate axes, we also now have coordinate planes. The xy**-plane** is the set of all $(x, y, z) \in \mathbb{R}^3$ such that $z = 0$. The xz**-plane** is the set of all $(x, y, z) \in \mathbb{R}^3$ such that $y = 0$. The yz**-plane** is the set of all $(x, y, z) \in \mathbb{R}^3$ such that $x = 0$. These are the three coordinate planes. These planes divide $\mathbb{R}^3$ into eight regions, which we call **octants**. The octant where $x \geq 0, y \geq 0$ and $z \geq 0$ is called the **first octant**.



We can now derive a distance formula between points, using the Pythagorean Theorem, just as we did in the plane $\mathbb{R}^2$.

Theorem 6.11. *The distance from the point* $(x, y, z) \in \mathbb{R}^3$ *to the origin* $(0, 0, 0)$ *is*

$$\sqrt{x^2 + y^2 + z^2}.$$

Proof. Consider the point $(x, y, 0)$ in the xy-plane. From our distance formula in the plane, the distance from this point to the origin is $\sqrt{x^2 + y^2}$. Consider the right triangle with vertices $(0, 0, 0)$, $(x, y, 0)$ and (x, y, z). By the Pythagorean Theorem, the length of the hypotenuse (which is also the distance from (x, y, z) to the origin) is

$$\sqrt{(\sqrt{x^2 + y^2})^2 + z^2} = \sqrt{x^2 + y^2 + z^2}.$$

$\square$

Remark 6.12. Consequently, the distance between any two points $(x_1, y_1, z_1) \in \mathbb{R}^3$ and $(x_2, y_2, z_2) \in \mathbb{R}^3$ is

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

We now give a few examples of surfaces in $\mathbb{R}^3$.

Definition 6.13. A **sphere** of radius $r > 0$ and center $(a, b, c) \in \mathbb{R}^3$ is defined as the set of points of distance exactly r from the point (a, b, c). So, the sphere is the set of all $(x, y, z) \in \mathbb{R}^3$ such that

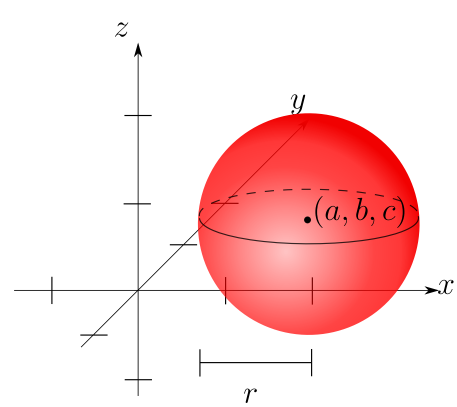
$$\sqrt{(x - a)^2 + (y - b)^2 + (z - c)^2} = r.$$

Equivalently, this sphere is described as the set of all $(x, y, z) \in \mathbb{R}^3$ such that

$$(x - a)^2 + (y - b)^2 + (z - c)^2 = r^2.$$

Example 6.14. The set of all (x, y, z) such that $x^2 + y^2 + z^2 \leq r^2$ is the set of all points of distance at most r from the origin. This set is known as the **solid ball** of radius r.

Example 6.15. The set of all (x, y, z) such that $x^2 + y^2 + z^2 = r^2$ and $z \geq 0$ is the **upper hemisphere** of the sphere of radius r.



Remark 6.16. Note that when $z = 0$ and $c = 0$, the equation $(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$ reduces to the equation for a circle of radius R and center (a, b):

$$(x-a)^2 + (y-b)^2 = R^2.$$

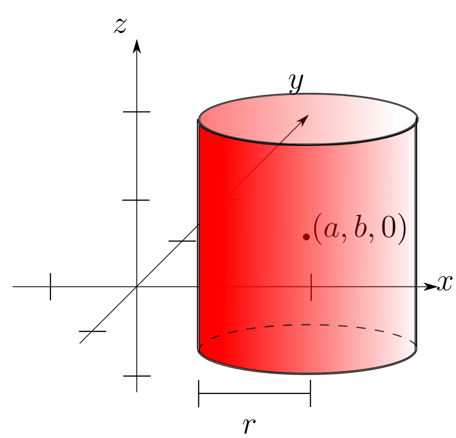
That is, when $c = 0$, the sphere of radius R is cut by the plane $z = 0$, producing a circle.

Definition 6.17. A **cylinder** of radius $r > 0$ with displacement $(a, b, 0) \in \mathbb{R}^3$ is defined as the set of points $(x, y, z) \in \mathbb{R}^3$ such that

$$(x-a)^2 + (y-b)^2 = r^2.$$

Note that the variable z is unrestricted.

We can readily extend addition, length, etc. to the case of three-dimensional vectors.



Definition 6.18. A **three-dimensional vector** $v = (x, y, z)$ is a directed line segment from the origin $(0, 0, 0)$ to the point (x, y, z). We define the **length** or **magnitude** of the vector v by

$$\|v\| = \sqrt{x^2 + y^2 + z^2}.$$

Let $\lambda \in \mathbb{R}$. Define $\lambda v = (\lambda x, \lambda y, \lambda z)$. Note that

$$\|\lambda v\| = |\lambda| \, \|v\| \, .$$

Two vectors $v, w \in \mathbb{R}^3$ are said to be **parallel** if there is some scalar $\lambda \in \mathbb{R}$ such that $v = \lambda w$.

We can add vectors together as follows.

Definition 6.19. Suppose $v = (x_1, y_1, z_1)$ and $w = (x_2, y_2, z_2)$. Define

$$v + w = (x_1 + x_2, y_1 + y_2, z_1 + z_2).$$

$$v - w = (x_1 - x_2, y_1 - y_2, z_1 - z_2).$$

Note that

$$v + (0, 0, 0) = (0, 0, 0) + v = v, \qquad v - v = (0, 0, 0).$$

For any vectors $u, v, w \in \mathbb{R}^3$ and for any scalar $\lambda \in \mathbb{R}$,

$$v + w = w + v.$$

$$u + (v + w) = (u + v) + w.$$

$$\lambda(u + v) = (\lambda u) + (\lambda v).$$

Definition 6.20. A vector $v \in \mathbb{R}^3$ such that $\|v\| = 1$ is called a **unit vector**. If v is any nonzero vector, then the vector

$$\frac{v}{\|v\|}$$

is a unit vector which is parallel to v. To see that $\frac{v}{\|v\|}$ is a unit vector, note that $\left\| \frac{v}{\|v\|} \right\| = \frac{\|v\|}{\|v\|} = 1$.

Example 6.21. Consider the vector $v = (1, 2)$. Then $\frac{v}{\|v\|} = \frac{(1,2)}{\sqrt{5}} = (1/\sqrt{5}, 2/\sqrt{5})$ is a unit vector, which is parallel to v.

Remark 6.22. Some textbooks use the notation $\vec{i} = (1, 0, 0)$, $\vec{j} = (0, 1, 0)$ and $\vec{k} = (0, 0, 1)$ so that a vector $v = (x, y, z) \in \mathbb{R}^3$ can be written as $v = x\vec{i} + y\vec{j} + z\vec{k}$. However, we will not use this notation.

Theorem 6.23 (**Triangle Inequality**). *For any vectors $v, w \in \mathbb{R}^3$, we have*

$$\|v + w\| \leq \|v\| + \|w\| .$$

6.3. **Dot Product.** The dot product operation allows us to combine two vectors, producing a scalar. In some sense, the dot product measures how close two vectors are. In another sense, the dot product of two vectors measures both their length and their angle simultaneously. For now we will just define the dot product, but later on we will give more geometric meaning to it.

Definition 6.24 (**Dot Product**). Let $v = (x_1, y_1)$ and let $w = (x_2, y_2)$ be vectors in $\mathbb{R}^2$. We define the **dot product** of v and w by

$$v \cdot w = x_1 x_2 + y_1 y_2.$$

Definition 6.25 (**Dot Product**). Let $v = (x_1, y_1, z_1)$ and let $w = (x_2, y_2, z_2)$ be vectors in $\mathbb{R}^3$. We define the **dot product** of v and w by

$$v \cdot w = x_1 x_2 + y_1 y_2 + z_1 z_2.$$

Note that $\sqrt{v \cdot v} = \sqrt{x_1^2 + y_1^2 + z_1^2} = \|v\|$.

Example 6.26.

$$(1, 2, 3) \cdot (0, 5, 7) = 0 \cdot 1 + 2 \cdot 5 + 3 \cdot 7 = 10 + 21 = 31.$$

Proposition 6.27 (**Properties of the dot product**). *Let $u, v, w \in \mathbb{R}^3$ and let $\lambda \in \mathbb{R}$. Then*

- $(0, 0, 0) \cdot v = v \cdot (0, 0, 0) = 0.$
- $v \cdot w = w \cdot v.$
- $(\lambda v) \cdot w = v \cdot (\lambda w) = \lambda(v \cdot w).$
- $u \cdot (v + w) = (u \cdot v) + (u \cdot w)$, *and* $(v + w) \cdot u = (v \cdot u) + (w \cdot u).$

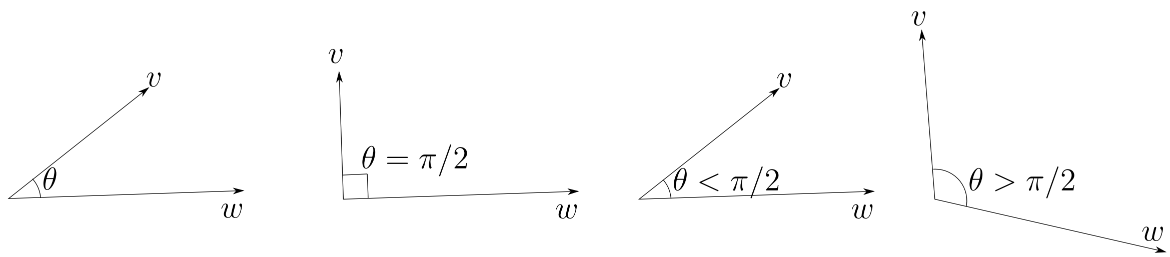
Recall that if a right triangle has an angle θ with edge lengths a and h, where a is the length of the edge adjacent to θ and h is the length of the hypotenuse, then $\cos\theta = a/h$. That is, $\cos\theta$ is the ratio of the length of the adjacent edge and the length of the hypotenuse. And in order to define the inverse of cos, we restrict the domain of cos to $[0, \pi]$. So, $\cos\colon [0, \pi] \to [-1, 1]$ and $\cos^{-1}\colon [-1, 1] \to [0, \pi]$. If $\pi/2 < \theta < \pi$, we define $\cos\theta$ to be $\cos(\pi - \theta)$.

Definition 6.28. The **angle** $0 \leq \theta \leq \pi$ between two nonzero vectors v, w is defined to be

$$\theta = \cos^{-1}\left(\frac{v}{\|v\|} \cdot \frac{w}{\|w\|}\right).$$

That is, the dot product of the unit vector in the direction v with the unit vector in the direction w is the cosine of θ. Written another way,

$$v \cdot w = \|v\| \, \|w\| \cos\theta, \qquad \cos\theta = \frac{v}{\|v\|} \cdot \frac{w}{\|w\|}.$$



In this sense, the dot product of v and w contains information about the lengths of v and w, as well as the angle between these vectors.

Remark 6.29. Recall that the range of $\cos^{-1}$ is $[0, \pi]$, so our definition of θ automatically means that $\theta \in [0, \pi]$.

Definition 6.30. Two nonzero vectors v, w are said to be **perpendicular** or **orthogonal** if the angle θ between them is $\pi/2$. In this case, we write $v \perp w$. Since $\cos(\pi/2) = 0$, and since $v \cdot w = \|v\| \, \|w\| \cos\theta = 0$, we see that the vectors v, w are perpendicular if and only if $v \cdot w = 0$. (If $v \cdot w = 0$, then $\theta = \cos^{-1}(0)$, so $\theta = \pi/2$.)

Remark 6.31. If $v \cdot w > 0$, then $\theta < \pi/2$, that is, the angle between v and w is less than $\pi/2$. If $v \cdot w < 0$, then $\theta > \pi/2$, that is, the angle between v and w is more than $\pi/2$. Both assertions follow from the definition of θ.

6.4. **Planes in Three Dimensions.**

Definition 6.32. Let $n = (a, b) \in \mathbb{R}^2$ be fixed and nonzero, let d be a constant, and let $(x, y) \in \mathbb{R}^2$ be a variable point. Recall that a **line** in the plane can be specified as the set of all points $(x, y) \in \mathbb{R}^2$ such that

$$ax + by = d.$$

Written another, way, a line is the set of all points $(x, y) \in \mathbb{R}^2$ such that

$$n \cdot (x, y) = d.$$

That is, a line is the set of all vectors (x, y) which have a constant dot product with the vector n. In the case $d = 0$, the line is the set of all vectors that are orthogonal to n. For this reason, we call n a **normal vector** to the line. (In the case $d \neq 0$, the line $n \cdot (x, y) = d$ is a translation of the line $n \cdot (x, y) = 0$.)

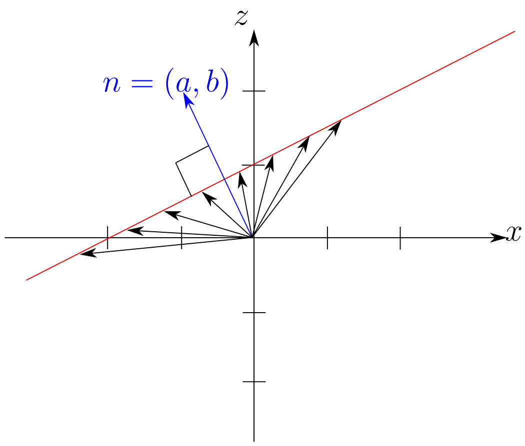
Definition 6.33. Let $n = (a, b, c) \in \mathbb{R}^3$ be fixed and nonzero, let d be a constant and let $v = (x, y, z) \in \mathbb{R}^3$ be a variable point. Then a **plane** is defined as the set of all points $(x, y, z) \in \mathbb{R}^3$ such that

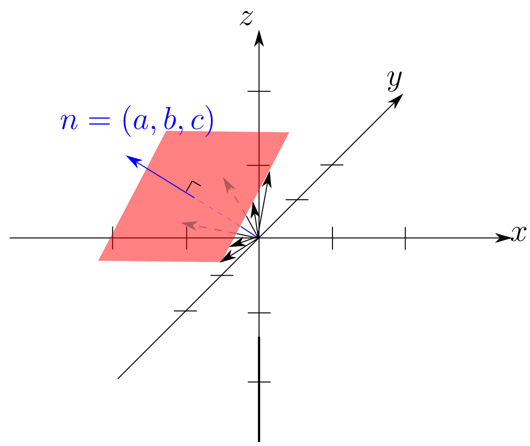
$$ax + by + cz = d.$$

Written another, way, a plane is the set of all points $(x, y, z) \in \mathbb{R}^3$ such that

$$n \cdot (x, y, z) = d.$$

That is, a plane is the set of all vectors (x, y, z) which have a constant dot product with the vector n. In the case $d = 0$, the plane is the set of all vectors that are orthogonal to n. For this reason, we call n a **normal vector** to the plane. (In the case $d \neq 0$, the plane $n \cdot (x, y, z) = d$ is a translation of the plane $n \cdot (x, y, z) = 0$.)





Remark 6.34. A plane which passes through the point (x_0, y_0, z_0) with normal vector $n = (a, b, c)$ can be written as the set of all (x, y, z) such that

$$n \cdot (x - x_0, y - y_0, z - z_0) = 0.$$

Example 6.35. The plane which passes through $(0, 1, 2)$ with normal vector $(3, 0, 2)$ is described as the set of all (x, y, z) such that $(3, 0, 2) \cdot (x, y - 1, z - 2) = 0$.

Example 6.36 (**Distance from a Point to a Plane**)**.** What is the distance of the point $w \in \mathbb{R}^3$ from the plane $ax + by + cz = 0$?

The distance from w to the plane $ax + by + cz = 0$ is the smallest value of $\|w - (x, y, z)\|$ where (x, y, z) ranges over all points in the plane $ax + by + cz = 0$. As in the case of a line, the point (x, y, z) where $\|w - (x, y, z)\|$ is smallest occurs when $w - (x, y, z)$ is perpendicular to the plane. We claim that $\|w - (x, y, z)\|$ is smallest when $w - (x, y, z) = \mathrm{proj}_{(a,b,c)}(w)$. To see this, consider the right triangle with edges $w - (x, y, z)$ and $\mathrm{proj}_{(a,b,c)}(w)$. Since this is a right triangle, the Pythagorean Theorem implies that the length of the hypotenuse is greater than the lengths of the other edges. That is, $\|w - (x, y, z)\| \geq \|\mathrm{proj}_{(a,b,c)}(w)\|$. In conclusion, the distance from w to the plane $ax + by + cz = 0$ is

$$\left\|\mathrm{proj}_{(a,b,c)}(w)\right\|.$$

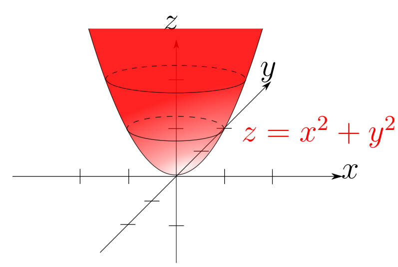
For example, let's find the distance of $w = (1, 0, 2)$ to the plane $x + 2y + 3z = 0$. Since $(a, b, c) = (1, 2, 3)$, the distance is given by

$$|w \cdot (a, b, c) / \|a, b, c\|| = |(1, 0, 2) \cdot (1, 2, 3)| / \sqrt{14} = 7/\sqrt{14}.$$

Definition 6.37. The **angle** between two planes is the angle between their normal vectors. So, if the planes are $ax + by + cz = d$ and $px + qy + rz = s$, then their angle is

$$\cos^{-1}\left(\frac{|(a, b, c) \cdot (p, q, r)|}{\|(a, b, c)\| \, \|(p, q, r)\|}\right).$$

7. Functions of Two or Three Variables



In a previous course, you discussed functions $f : \mathbb{R} \to \mathbb{R}$. We could then graph these functions in the xy-plane. In this class, we have been discussing functions $r : \mathbb{R} \to \mathbb{R}^2$ and $r : \mathbb{R} \to \mathbb{R}^3$. We can think of these functions as curves. We now consider functions with larger domains as well. For example, we will first consider functions $f\colon \mathbb{R}^2 \to \mathbb{R}$.

Example 7.1. Let $(x, y) \in \mathbb{R}^2$. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2$. We can plot this function in three-dimensions by considering the value of f to be the z-coordinate. That is, we identify the function $f(x, y) = x^2 + y^2$ with the surface $z = f(x, y) = x^2 + y^2$. We recall that surface $z = x^2 + y^2$ is a paraboloid.

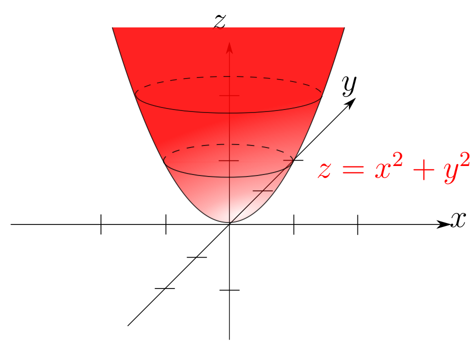
Example 7.2. Suppose we have a domain D in the plane, which we think of as a map of California. Given $(x, y) \in D$, we can think of $f(x, y)$ as the current temperature at the point (x, y). Then $f\colon D \to \mathbb{R}$.

Example 7.3. Suppose we have a domain D in three-dimensional space $\mathbb{R}^3$. For example, we can identify the unit ball $x^2 + y^2 + z^2 \leq 1$ with the earth. Given $(x, y, z) \in D$, we can think of $f(x, y, z)$ as the current temperature at the point (x, y, z) on the earth. Then $f\colon D \to \mathbb{R}$.

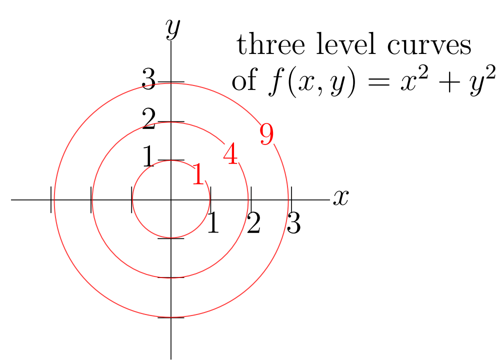
More generally, given any positive integer n, we can consider functions $f\colon \mathbb{R}^n \to \mathbb{R}$. We write a general point in $\mathbb{R}^n$ as $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. And at each such point $(x_1, \ldots, x_n) \in \mathbb{R}^n$, we let $f(x_1, \ldots, x_n)$ be some number.

Remark 7.4. Given a function $f\colon \mathbb{R}^2 \to \mathbb{R}$, we said we could identify the function f with the surface $z = f(x, y)$ in $\mathbb{R}^3$. However, visualizing functions $f\colon \mathbb{R}^3 \to \mathbb{R}$ is a bit more difficult. One way to visualize these functions is to think of $f(x, y, z)$ as a color intensity assigned to the point (x, y, z). For example, if $f(x, y, z)$ is a large number, we can think of (x, y, z) as having a light grey color. And if $f(x, y, z)$ is a very small number, we can think of (x, y, z) as having a dark grey color. Drawing a picture like this is perhaps impossible, but the visualization is perhaps more tractable. For example, if $f(x, y, z)$ is the temperature of the earth at the point (x, y, z) with $x^2 + y^2 + z^2 \leq 1$, we could think of $f(x, y, z)$ as a bright red color when the point (x, y, z) is very hot, and we could think of $f(x, y, z)$ as a dark blue color when the point (x, y, z) is very cold.

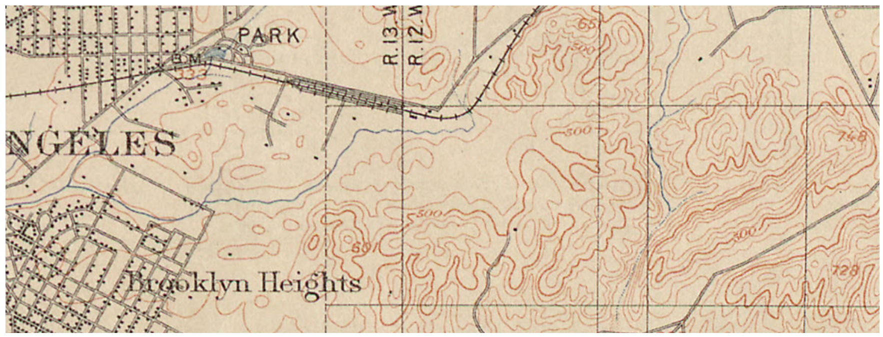
Example 7.5 (**Drawing Functions of Two Variables**)**.** Suppose we want to draw the function $f(x, y) = x^2 + y^2$.



So, we want to draw the surface $z = x^2 + y^2$. It is sometimes convenient to draw the **level curves** of the function f. That is, we can maybe draw the curve $x^2 + y^2 = 1$ in the plane $z = 1$, and then draw the curve $x^2 + y^2 = 4$ in the plane $z = 4$. Doing so results in two circles. On each individual circle, the function f is constant (i.e. the surface has constant height). We can also draw the intersection of $z = x^2 + y^2$ with other hyperplanes. For example, in the plane $x = 0$, we could draw the hyperbola $z = y^2$; in the plane $x = 1$ we could draw the hyperbola $z = 1 + y^2$; in the plane $y = 1$, we could

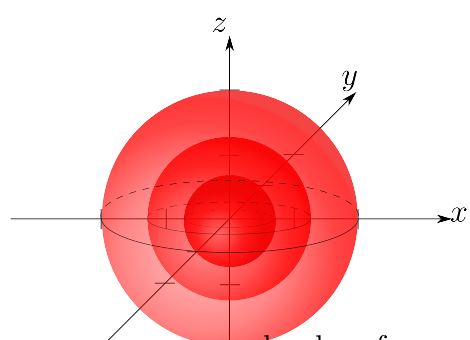


draw the parabola $z = x^2 + 1$, and so on.
Drawing the intersection of $z = x^2 + y^2$ with several hyperplanes in this way gives a good sketch of the function.



Example 7.6 (Level Surfaces in Three Variables). Consider the function $f(x, y, z) = x^2 + y^2 + z^2$. As we discussed above, this function is perhaps difficult to draw. However, we can still draw the **level surfaces** of the function.

For example, the level surface $x^2 + y^2 + z^2 = 1$ is the unit sphere, and we can then draw this surface. And the level surface $x^2 + y^2 + z^2 = 4$ is the sphere of radius 2 centered at the origin, which can also be drawn, and so on. The function f is constant on any individual level surface.



level surfaces of
$f(x, y, z) = x^2 + y^2 + z^2$

Example 7.7 (Domains). Multivariable domains can be a bit more complicated than single-variable domains. Suppose we have a function $f : D \to \mathbb{R}$ where D is a subset of $\mathbb{R}^2$ defined as the set of (x, y) such that $0 \leq x \leq 1$ and $0 \leq y \leq 1$. We can draw this domain by intersecting the domain corresponding to each condition. For example, the set $0 \leq x \leq 1$ is an infinite rectangle lying between the lines $x = 0$ and $x = 1$. And the set $0 \leq y \leq 1$ is an infinite rectangle lying between the lines $y = 0$ and $y = 1$. The intersection of these two rectangles is the unit square.

For another example, consider the domain D in $\mathbb{R}^3$ defined as the set of all (x, y, z) such that $x^2 + y^2 \leq 4$ and such that $y^2 + z^2 \geq 1$. The first condition gives an infinite solid cylinder, bounded by the cylinder $x^2 + y^2 = 4$. From this solid cylinder, we remove the cylinder $y^2 + z^2 < 1$. In summary, the domain D is an infinite solid cylinder with a hole drilled through it along the x-axis.

7.1. **Limits and Continuity.** In single-variable calculus, we learned a lot about functions by using the notions of continuity and derivatives. We can play a similar game now for multivariable functions. However, the continuity and derivative are now a bit different.

In this section, we only consider functions $f : \mathbb{R}^2 \to \mathbb{R}$. Treating functions with domain $\mathbb{R}^3$ can be done analogously.

Definition 7.8 (**Limit, Informal Definition**)**.** Let $(a,b) \in \mathbb{R}^2$ and let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let $L \in \mathbb{R}$. We say that $\lim_{(x,y)\to(a,b)} f(x,y) = L$ if, whenever (x,y) is close to (a,b), we know that $f(x,y)$ is close to L.

Definition 7.9 (**Limit, Formal Definition**)**.** Let $(a,b) \in \mathbb{R}^2$ and let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let $L \in \mathbb{R}$. We say that $\lim_{(x,y)\to(a,b)} f(x,y) = L$ if, given any $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, if (x,y) satisfies $0 < \|(x,y) - (a,b)\| < \delta$, then $|f(x,y) - L| < \varepsilon$.

Example 7.10. Let $f(x,y) = x + y$. Let $(a,b) = (1,2)$. Let $L = 3$. Let's verify from the formal definition that $\lim_{(x,y)\to(a,b)} f(x,y) = 3$. Let $\varepsilon > 0$. Then choose $\delta = \varepsilon/2$. Assume $0 < \|(x,y) - (1,2)\| < \delta$. Then $|x-1| < \delta$ and $|y-2| < \delta$. So, $|f(x,y) - L| = |x+y-3| = |(x-1)+(y-2)| \leq |x-1| + |y-2| < \delta + \delta = \varepsilon$. So, $\lim_{(x,y)\to(a,b)} f(x,y) = 3$.

Example 7.11. Let $f(x,y) = x^2/(x^2+y^2)$ We show that the following limit does not exist: $\lim_{(x,y)\to(0,0)} f(x,y)$. Consider $(x_n,y_n) = (0,1/n)$, $n \geq 1$. Then $(x_n,y_n) \to (0,0)$ as $n \to \infty$, and $f(x_n,y_n) = 0$. However, if $(c_n,d_n) = (1/n,0)$, $n \geq 1$, then $(c_{,n}, d_n) \to (0,0)$, but $f(c_n,d_n) = 1$. If $\lim_{(x,y)\to(0,0)} f(x,y)$ exists, then $\lim_{n\to\infty} f(x_n,y_n)$ would have to be equal to $\lim_{n\to\infty} f(c_n,d_n)$. So, $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

Proposition 7.12 (**Limit Laws**)**.** *Let $f,g\colon \mathbb{R}^2 \to \mathbb{R}$. Assume that $\lim_{(x,y)\to(a,b)} f(x,y)$ exists and that $\lim_{(x,y)\to(a,b)} g(x,y)$ exists. Then*

- $\lim_{(x,y)\to(a,b)}[f(x,y) + g(x,y)] = (\lim_{(x,y)\to(a,b)} f(x,y)) + (\lim_{(x,y)\to(a,b)} g(x,y))$.
- *For any constant c, $\lim_{(x,y)\to(a,b)}[cf(x,y)] = c \cdot \lim_{(x,y)\to(a,b)} f(x,y)$.*
- $\lim_{(x,y)\to(a,b)}[f(x,y)g(x,y)] = (\lim_{(x,y)\to(a,b)} f(x,y))(\lim_{(x,y)\to(a,b)} g(x,y))$.
- *If $\lim_{(x,y)\to(a,b)} g(x,y) \neq 0$, then*

$$\lim_{(x,y)\to(a,b)}(f(x,y)/g(x,y)) = (\lim_{(x,y)\to(a,b)} f(x,y))/(\lim_{(x,y)\to(a,b)} g(x,y)).$$

Definition 7.13 (**Continuity**)**.** Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. We say that f is **continuous** at the point (a,b) if

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

We say that f is continuous on a domain D if f is continuous at (a,b) for every (a,b) in D.

Example 7.14. Let $f\colon \mathbb{R}^3 \to \mathbb{R}$, so that $f(x,y,z) = x^3 + xy^2 + zxy$. Then f is a polynomial, so f is continuous.

Let $g(x,y,z) = x\sin(z) + y$. Then g is continuous, and so is the product of the continuous functions $f(x,y,z) \cdot g(x,y,z)$.

Let $h(x,y,z) = x^4y^4z^2$. Let $F(x,y,z) = f(x,y,z)/g(x,y,z)$. Then F is a quotient of continuous functions, so F is continuous at any point (x,y,z) where $g(x,y,z) \neq 0$.

Theorem 7.15 (**A Composition of Continuous Functions is Continuous**)**.** *Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous and let $g\colon \mathbb{R} \to \mathbb{R}$ be continuous. Then the composition $g(f(x,y))$ is continuous.*

Example 7.16. The function $\ln(x^2+y^2)$ is continuous for all (x,y) with $(x,y) \neq (0,0)$. However, this function is discontinuous at $(x,y) = (0,0)$.

7.2. **Partial Derivatives.** Now that we have discussed continuity for multivariable functions, we now continue extending calculus to multivariable functions. Our next topic is now differentiability. For a single-variable function $f\colon \mathbb{R}\to\mathbb{R}$, we defined

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

And we interpreted $f'(x)$ as the rate of change of f, as x increases.

Now, consider a function $f\colon \mathbb{R}^2\to\mathbb{R}$. Now that there are two variables in the domain, there are many more directions in which we can measure the rate of change of the function. For example, how much does the function change as x increases (but y is fixed)? How much does the function change as y increases (but x is fixed)? How much does the function change as y increases at a rate 3.5 times larger than the rate that x increases? And so on. We will begin by answering the first two questions, and we will save the other one for later.

Definition 7.17 (**Partial Derivatives**)**.** Let $f\colon \mathbb{R}^2\to\mathbb{R}$. Let $(a,b)\in\mathbb{R}$. We define the partial derivative of f at the point (a,b) in the x-direction by the following limit (if it exists):

$$\frac{\partial f}{\partial x}(a,b)=f_x(a,b)=\lim_{h\to 0}\frac{f(a+h,b)-f(a,b)}{h}.$$

That is, the partial derivative of f in the x-direction is just the derivative of $f(x,y)$ with respect to x, if we consider y to be fixed.

We similarly define the partial derivative of f at the point (a,b) in the y-direction by

$$\frac{\partial f}{\partial y}(a,b)=f_y(a,b)=\lim_{h\to 0}\frac{f(a,b+h)-f(a,b)}{h}.$$

Example 7.18. Let $f(x,y)=x^2y^3+y^2+x$. Then

$$\frac{\partial f}{\partial x}=2xy^3+1,\qquad \frac{\partial f}{\partial y}=3x^2y^2+2y.$$

Example 7.19. Let $f(x,y,z)=x^2+y^2+z^2$. Then

$$\frac{\partial f}{\partial x}=2x,\qquad \frac{\partial f}{\partial y}=2y\qquad \frac{\partial f}{\partial z}=2z.$$

Remark 7.20. We can interpret a partial derivative in the same way we always interpret derivatives. For example, $\partial f/\partial x$ is the rate of change of f, as x increases (while other variables are held fixed).

Example 7.21. Since a partial derivative is essentially a one-variable derivative, it follows all of the usual rules such as the chain rule, product rule, quotient rule, etc.

Let $f(x,y)=\frac{x(x^2+1)^2}{y^2+1}$. Then

$$\frac{\partial f}{\partial x}=\frac{2x(x^2+1)(2x)+(x^2+1)^2}{y^2+1},\qquad \frac{\partial f}{\partial y}=\frac{-x(x^2+1)^2(2y)}{(y^2+1)^2}.$$

For single variable functions, after defining the first derivative $f'(x)$ of a function $f\colon\mathbb{R}\to\mathbb{R}$, we then defined $f''(x)=(d/dx)f'(x)$, $f'''(x)=(d/dx)f''(x)$, and so on. We will do something similar now for multivariable functions. However, now that we have more than one variable, there are many more derivatives to consider.

Definition 7.22 (**Iterated Partial Derivatives**)**.** Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. We define the second order partial derivatives of f in the following way.

$$\frac{\partial^2 f}{\partial x^2} = f_{xx} = \frac{\partial}{\partial x}\frac{\partial f}{\partial x}, \qquad \frac{\partial^2 f}{\partial y^2} = f_{yy} = \frac{\partial}{\partial y}\frac{\partial f}{\partial y}.$$

$$\frac{\partial^2 f}{\partial x \partial y} = f_{xy} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}, \qquad \frac{\partial^2 f}{\partial y \partial x} = f_{yx} = \frac{\partial}{\partial y}\frac{\partial f}{\partial x}.$$

Example 7.23. Let's compute some iterated partial derivatives of $f(x, y, z) = xyz + x^2$. We have

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(yz + 2x) = 2.$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(xz) = 0.$$

$$\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial}{\partial y}\frac{\partial f}{\partial z} = \frac{\partial}{\partial y}(xy) = x, \qquad \frac{\partial^2 f}{\partial z \partial y} = \frac{\partial}{\partial z}\frac{\partial f}{\partial y} = \frac{\partial}{\partial z}(xz) = x$$

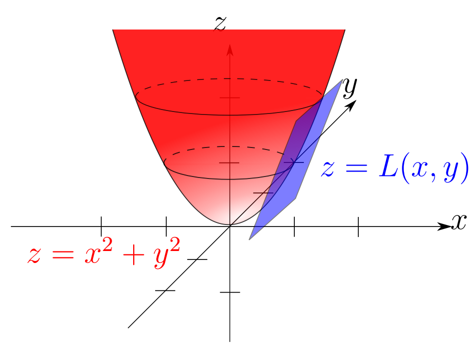
In the last example, we found that $f_{yz} = f_{zy}$. That is, the order of the iterated partial derivative does not matter. This observation is in fact generic.

Theorem 7.24 (**Clairaut's Theorem**)**.** *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Suppose f_{xy} and f_{yx} both exist and are continuous. Then $f_{xy} = f_{yx}$.*

7.3. **Differentiability and Tangent Planes.** Suppose $f\colon \mathbb{R} \to \mathbb{R}$. Fix $a \in \mathbb{R}$. Recall that the **linearization** of f at the point a is given by the following function of x.

$$L(x) = f(a) + (x - a)f'(a).$$

This function is the linearization of f since L is a linear function of x, $L(a) = f(a)$ and $L'(a) = f'(a)$. We also refer to $L(x)$ as the **tangent line** to f at a. We can do something similar for function several variables.



Definition 7.25. Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$. Fix $(a, b) \in \mathbb{R}$. We define the **linearization** of f at the point (a, b) to be the following function of $(x, y) \in \mathbb{R}^2$.

$$L(x, y) = f(a, b) + (x - a)f_x(a, b) + (y - b)f_y(a, b) = f(a, b) + ((x, y) - (a, b)) \cdot \nabla f(a, b).$$

Here we defined the **gradient** of f at the point (a, b) to be the following vector in $\mathbb{R}^2$.

$$\nabla f(a, b) = (f_x(a, b), f_y(a, b)).$$

If $f\colon \mathbb{R}^3 \to \mathbb{R}$ and if $(a, b, c) \in \mathbb{R}$, we similarly define the **gradient** of f at (a, b, c) to be the vector in $\mathbb{R}^3$.

$$\nabla f(a, b, c) = (f_x(a, b, c), f_y(a, b, c), f_z(a, b, c)).$$

And we similarly define the **linearization** of f at the point (a, b, c) to be the following function of $(x, y, z) \in \mathbb{R}^3$.

$$L(x, y, z) = f(a, b, c) + ((x, y, z) - (a, b, c)) \cdot \nabla f(a, b, c).$$

Example 7.26. We find the linearization of the function $f(x,y) = x^2 + y^2$ at the point $(1,3)$. The linearization is given by

$$\begin{aligned} L(x,y) &= f(1,3) + ((x,y) - (1,3)) \cdot \nabla f(1,3) \\ &= 10 + ((x,y) - (1,3)) \cdot (2,6) = 10 + 2(x-1) + 6(y-3). \end{aligned}$$

Remark 7.27. Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$. The linearization L gives an equation for a plane $z = L(x,y)$. We say that the plane $z = L(x,y)$ is then the **tangent plane** to the surface $z = f(x,y)$ at the point (a,b). The tangent plane at (a,b) is a linear approximation to the surface $z = f(x,y)$ in the same way that a single-variable function $f\colon \mathbb{R} \to \mathbb{R}$ can have a tangent line.

Example 7.28. Continuing the previous example, the tangent plane to the paraboloid $z = x^2 + y^2$ at $(1,3)$ is given by $z = 10 + 2(x-1) + 6(y-3)$.

Example 7.29. Let's find the tangent plane to the cone $z^2 = x^2 + y^2$, $z \geq 0$ at $(3,4)$. Solving for z, we have $z = \sqrt{x^2+y^2}$. Setting $f(x,y) = \sqrt{x^2+y^2}$, we have $\nabla f(x,y) = (x(x^2+y^2)^{-1/2}, y(x^2+y^2)^{-1/2})$. So, $\nabla f(3,4) = (3/5, 4/5)$. So, the tangent plane is given by $z = f(3,4) + ((x,y) - (3,4)) \cdot \nabla f(3,4) = 5 + (x-3)(3/5) + (y-4)(4/5)$.

Note that the gradient of f is discontinuous at $(x,y) = (0,0)$, and there is similarly no clear geometric meaning to the tangent plane at the tip of the cone.

Remark 7.30 (**Linear Approximation**)**.** Recall that for a single-variable function $f\colon \mathbb{R} \to \mathbb{R}$, we have the heuristic $f(x) \approx L(x)$ when x is near a. We similarly have a heuristic for functions of multiple variables. For example, if $f\colon \mathbb{R}^2 \to \mathbb{R}$, then $f(x,y) \approx L(x,y)$ when (x,y) is near the point (a,b). This approximation holds since f and L agree to first order, just as in the single-variable case. Indeed, we have

$$L(a,b) = f(a,b), \qquad \nabla L(a,b) = (L_x(a,b), L_y(a,b)) = (f_x(a,b), f_y(a,b)) = \nabla f(a,b).$$

7.4. **Gradient and Directional Derivative.** Let $f\colon \mathbb{R} \to \mathbb{R}$. Recall that if the derivative of f is positive, then the function is increasing; if the derivative is negative, then the function is decreasing, and if the derivative is zero, then the function has a critical point. So, the derivative of f tells us how the function f is changing. There is an analogous thing to say for functions of two or more variables. However, we now have many variables, so a single derivative may not necessarily tell us what is going on. That is, we will need to look at a few different derivatives, but then some notion of positivity or negativity is then less straightforward.

Example 7.31. Consider the function $f(x,y) = x^2 - y^2$. If $y = 0$, then $f(x,y) = x^2$. That is, when we restrict the function to the line $y = 0$, we see that $f(x,y)$ has a local minimum at $x = 0$. However, if $x = 0$, then $f(x,y) = -y^2$. That is, when we restrict the function to the line $x = 0$, we see that $f(x,y)$ has a local *maximum* at $y = 0$. This behavior can be understood by looking at both derivatives of f. We have $f_x = 2x$, so that moving x towards zero will always decrease the value of f. And $f_y = -2y$, so moving y towards zero will always increase the value of f. This increasing/decreasing nature of f cannot be understood just by looking at a single derivative of f.

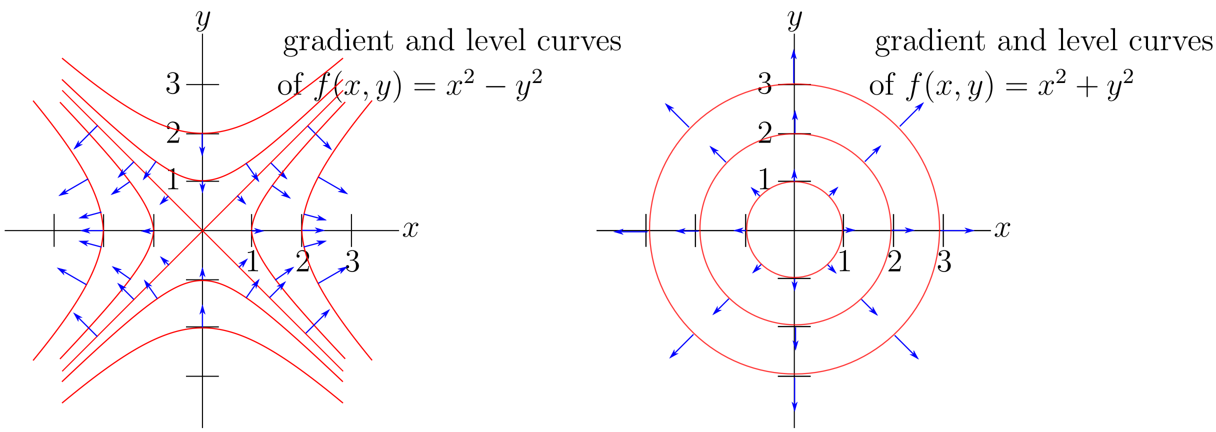
Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ Recall that we defined the gradient of f at (a,b) by

$$\nabla f(a,b) = (f_x(a,b), f_y(a,b)).$$

Remark 7.32. The gradient vector $\nabla f(a,b)$ points in the direction in which the function f is increasing, at the point (a,b). If $\nabla f(a,b) = (0,0)$, or if f_x or f_y is undefined, we say that (a,b) is a **critical point** of the function f.

Example 7.33. Let $f(x,y) = x^2 - y^2$ as in the previous example. Then $\nabla f(a,b) = (2a,-2b) = 2(a,-b)$. The vector $(a,-b)$ points in the direction in which f increases, at the point (a,b). If $(a,b) = (0,0)$, then $\nabla f(a,b) = (0,0)$, so $(0,0)$ is a critical point of f. However, note that this point is neither a local maximum nor a local minimum of f. We will discuss critical points later on in the course.

Example 7.34. Let $f(x,y) = x^2 + y^2$. Then $\nabla f(a,b) = (2a,2b)$. The vector (a,b) points in the direction in which f increases, at the point (a,b). That is, moving away from the origin always increases the value of f. If $(a,b) = (0,0)$, then $\nabla f(a,b) = (0,0)$, so $(0,0)$ is a critical point of f. In this case, the origin $(0,0)$ is the global minimum of the function f.



Proposition 7.35 (Properties of the Gradient). *Let $f,g\colon \mathbb{R}^2 \to \mathbb{R}$ and let $c \in \mathbb{R}$.*

- $\nabla(f+g) = (\nabla f) + (\nabla g)$
- $\nabla(cf) = c\cdot(\nabla f)$.
- $\nabla(fg) = f\cdot(\nabla g) + g\cdot(\nabla f)$. ***(Product Rule)***
- *Let $F\colon \mathbb{R}\to\mathbb{R}$. Then $\nabla F(f(x,y)) = F'(f(x,y))\cdot\nabla f(x,y)$.* ***(Chain Rule)***

Example 7.36. Let $f(x,y) = x^2+y^2$, let $g(x,y) = 2x$. Then at the point (a,b), we have $\nabla(f+g) = \nabla f + \nabla g = (2a,2b) + (2,0) = (2a+2,2b)$.

Let $h(x,y) = (x^2+y^2)$ and let $F(t) = t^3$. Then at the point (a,b), we have $\nabla F(h(a,b)) = 3(h(a,b))^2 \nabla h(a,b) = 3(a^2+b^2)^2(2a,2b)$.

Proposition 7.37 (Chain Rule for Vector-Valued Functions). *Let $f\colon \mathbb{R}^3\to\mathbb{R}$. Let $r\colon \mathbb{R}\to\mathbb{R}^3$. Then*

$$\frac{d}{dt}(f(r(t))) = (\nabla f(r(t)))\cdot r'(t).$$

Proof Sketch. We use the linear approximation, $r(t+h) \approx r(t)+hr'(t)$. Also, $f((x,y,z)+v) \approx f(x,y,z) + v \cdot \nabla f(x,y,z)$. Substituting one approximation into the other,

$$\frac{d}{dt}f(r(t)) = \lim_{h\to 0}\frac{f(r(t+h)) - f(r(t))}{h} \approx \lim_{h\to 0}\frac{f(r(t)+hr'(t)) - f(r(t))}{h}$$
$$\approx \lim_{h\to 0}\frac{[f(r(t)) + hr'(t)\cdot \nabla f(r(t))] - f(r(t))}{h} = \lim_{h\to 0}\frac{hr'(t)\cdot \nabla f(r(t))}{h} = r'(t)\cdot \nabla f(r(t)).$$

□

Example 7.38. Let $f(x,y) = x^2 + y^2$. Let $s(t) = (t,t^3)$. Then $f(s(t)) = t^2 + t^6$. From the Chain Rule, we have $(d/dt)f(s(t)) = \nabla f(s(t)) \cdot s'(t) = (2t, 2t^3) \cdot (1, 3t^2) = 2t + 6t^5$.

7.5. **Directional Derivatives.** From the Chain Rule for vector-valued functions, we saw that, if $f\colon \mathbb{R}^3 \to \mathbb{R}$ and if $r\colon \mathbb{R} \to \mathbb{R}^3$, then

$$\frac{d}{dt}(f(r(t))) = \lim_{h\to 0}\frac{f(r(t)+hr'(t)) - f(r(t))}{h} = (\nabla f(r(t)))\cdot r'(t).$$

We can similarly allow $r'(t)$ to be any fixed vector $v \in \mathbb{R}^3$, and turn this limit into a definition.

Definition 7.39 (**Directional Derivative**)**.** Let $f\colon \mathbb{R}^3 \to \mathbb{R}$. Let $(a,b,c) \in \mathbb{R}^3$ and let $v \in \mathbb{R}^3$. We define the derivative $D_v f(a,b,c)$ of f with respect to v at the point (a,b,c) by

$$D_v f(a,b,c) = \lim_{h\to 0}\frac{f((a,b,c)+hv) - f(a,b,c)}{h} = (\nabla f(a,b,c))\cdot v.$$

If v is a unit vector, we call $D_v f(a,b,c)$ the **directional derivative** of f in the direction v at the point (a,b,c).

Remark 7.40. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ and let $v \in \mathbb{R}^2$. The directional derivative $D_v f(a,b)$ measures the rate of growth of f at the point (a,b) in the direction v.

Example 7.41. Let $f(x,y) = x^2 + y^2$. Let $v = (1,2)$. Then $D_v f(0,-1) = (\nabla f(0,-1)) \cdot (1,2) = (0,-2)\cdot(1,2) = -4$.

Directional derivatives allow us to better understand the information contained in the gradient itself. For example, suppose $v \in \mathbb{R}^2$ is a unit vector and let $(a,b) \in \mathbb{R}^2$. Then, if θ denotes the angle between $\nabla f(a,b)$ and v, we have

$$D_v f = \nabla f(a,b)\cdot v = \|\nabla f(a,b)\| \cos\theta.$$

That is, $D_v f$ is the most positive when $\theta = 0$, and $D_v f$ is most negative when $\theta = \pi$. In the case $\theta = 0$, v points in the same direction as $\nabla f(a,b)$. And in the case that $\theta = \pi$, v points in the opposite direction of $\nabla f(a,b)$.

Finally, in the case $\theta = \pi/2$, $D_v f = 0$. So, when v is perpendicular to $\nabla f(a,b)$, f is neither increasing nor decreasing in the direction v. We summarize these observations below.

Definition 7.42 (**Optimization Interpretation of Gradient and Directional Derivatives**)**.** Let $f\colon \mathbb{R}^2 \to \mathbb{R}$, let $v \in \mathbb{R}^2$ be a unit vector, and let $(a,b) \in \mathbb{R}^2$. Assume $\nabla f(a,b) \neq (0,0)$. Then

(i) $\nabla f(a,b)$ points in the direction of greatest increase of f. And $-\nabla f(a,b)$ points in the direction of greatest decrease of f.
(ii) $\nabla f(a,b)$ is orthogonal to the level curve of f at (a,b).

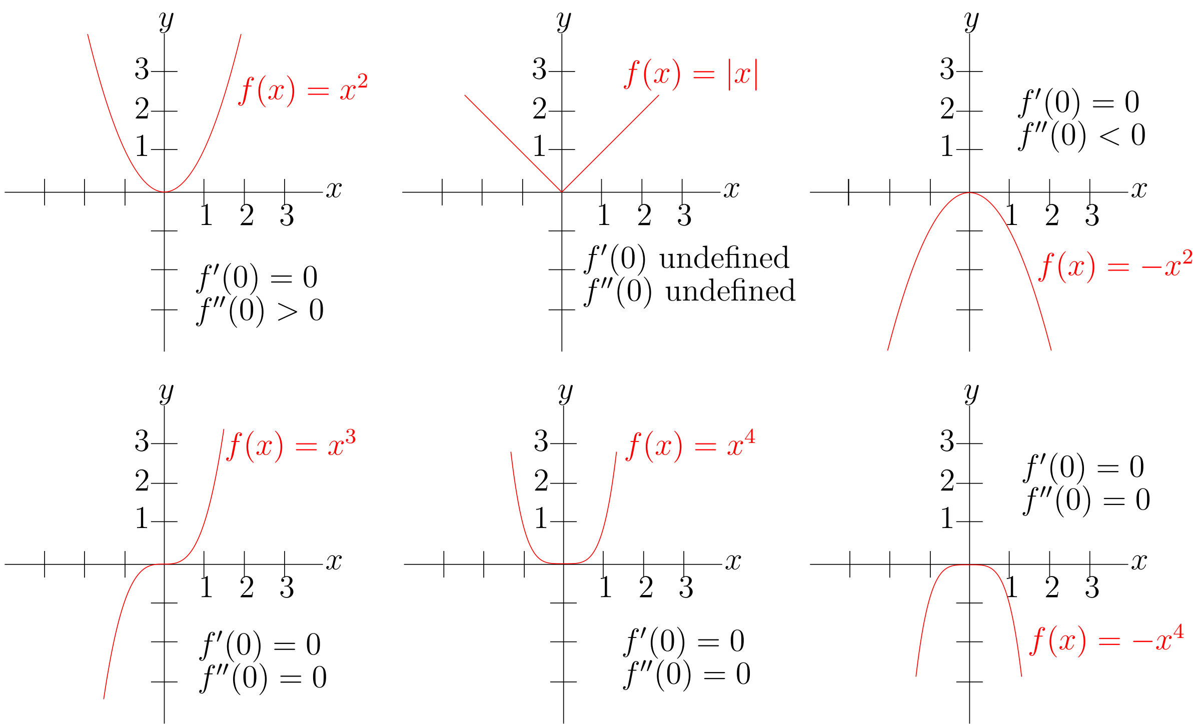
In the case $f\colon \mathbb{R}^3 \to \mathbb{R}$, then $\nabla f(a,b,c)$ is normal to the level surface of f at (a,b,c).

Proof of (ii). Let $r\colon \mathbb{R} \to \mathbb{R}^2$ be a level curve of f. That is, assume that $f(r(t))$ is constant as t varies, and assume that $r(0) = (a,b,c)$. Then using the Chain Rule,

$$0 = \frac{d}{dt} f(r(t)) = \nabla f(r(t)) \cdot r'(t).$$

That is, $r'(t)$ is orthogonal to $\nabla f(r(t))$ for any t (in particular for $t = 0$). And $r'(t)$ is tangent to the level curve $r(t)$. So, $\nabla f(r(t))$ is orthogonal to the level curve. □

8. Optimization



We have now built up enough machinery of differential calculus to begin optimizing functions of multiple variables. We have seen that the gradient of a function points in the direction of greatest increase, and the negative of the gradient of a function points in the direction of greatest decrease. We will exploit this and other properties of the gradient to optimize functions. We first recall optimization for a single variable.

Suppose $f\colon \mathbb{R} \to \mathbb{R}$. Recall that a critical point occurs at x when $f'(x) = 0$ or $f'(x)$ is undefined. If x is a critical point for f, then x may or may not be a local extremum of f. For example, if $f(x) = x^2$, then $f'(x) = 0$ occurs only at $x = 0$. And we know that $x = 0$ is a global minimum for f. That is, $f(x) \geq f(0)$ for all $x \in \mathbb{R}$. However, if $f(x) = x^3$, then $f'(x) = 0$ occurs only at $x = 0$. But $x = 0$ is not a local maximum or a local minimum. That is, no matter how close we look near $x = 0$, there are always points x_1, x_2 near zero such that $f(x_1) > f(0) > f(x_2)$. The issue here is that $f''(x)$ changes sign at $x = 0$, so that $x = 0$ is an inflection point of f. So, the first derivative can be helpful sometimes for optimizing functions, but not always. Lastly, recall that if $f''(x) > 0$, then f has a shape similar to $g(x) = x^2$ (f is concave up), and if $f''(x) < 0$, then f has a shape similar to $g(x) = -x^2$ (f is concave down). So, the second derivative of f can also tell us some things about local extrema of a function. Note that $f(x) = |x|$ has a global minimum at $x = 0$ even though $f'(0)$ is undefined. Note also that $f(x) = x^4$ has a global minimum at $x = 0$, while $f'(0) = f''(0) = 0$, so the value of the second derivative at a single point cannot always identify a local extremum.

Let's recall the second derivative test in one variable.

Example 8.1. Let $f : \mathbb{R} \to \mathbb{R}$. We say that c is a local maximum for f if all points x near c satisfy $f(x) \leq f(c)$. We say that c is a local minimum for f if all points x near c satisfy $f(x) \geq f(c)$. We say that c is a local extremum for f if c is either a local maximum or a local minimum of f.

(**Second Derivative Test on $\mathbb{R}$**)

- If $f'(c) = 0$ and $f''(c) > 0$ then c is a local minimum of f.
- If $f'(c) = 0$ and if $f''(c) < 0$, then c is a local maximum of f.
- If $f'(c) = 0$ and if $f''(c) = 0$, then this test is inconclusive. That is, c may or may not be a local extremum of f.

We will come up with an analogous test for functions of more variables. However, the test will be a bit more complicated than before. We begin with some definitions

Definition 8.2 (**Local Extremum**)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$.

- We say that f has a **local maximum** at (a, b) if $f(x, y) \leq f(a, b)$ for all (x, y) near (a, b). (Formally: there is some $t > 0$ such that, if $\|(a, b) - (x, y)\| < t$, then $f(x, y) \leq f(a, b)$.)
- We say that f has a **local minimum** at (a, b) if $f(x, y) \geq f(a, b)$ for all (x, y) near (a, b). (Formally: there is some $t > 0$ such that, if $\|(a, b) - (x, y)\| < t$, then $f(x, y) \leq f(a, b)$.)

We say that f has a **local extremum** at (a, b) if (a, b) is either a local maximum or a local minimum.

Example 8.3. $f(x, y) = x^2 + y^2$ has a local minimum at $(0, 0)$, since $f(x, y) \geq 0 = f(0, 0)$ for all $(x, y) \in \mathbb{R}^2$. Note that $\nabla f(0, 0) = (0, 0)$.

$g(x, y) = -x^2 - y^2$ has a local maximum at $(0, 0)$, since $g(x, y) \leq 0 = g(0, 0)$ for all $(x, y) \in \mathbb{R}^2$. Note that $\nabla g(0, 0) = (0, 0)$.

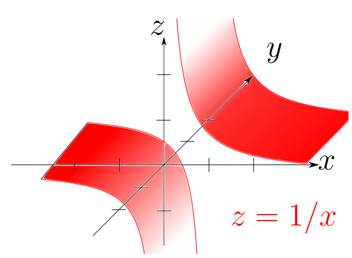
$h(x, y) = x^2 - y^2$ does not have a local extremum at $(0, 0)$. Still, we have $\nabla h(0, 0) = (0, 0)$.

Definition 8.4 (**Critical Point**)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ and let $(a, b) \in \mathbb{R}^2$. We say that (a, b) is a **critical point** of f if either:

- $\nabla f(a, b) = (0, 0)$, or
- One component of $\nabla f(a, b)$ is undefined.

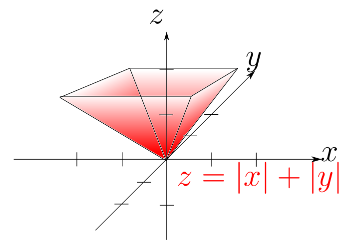
Here are some examples of critical points.

Example 8.5. The point $(0, 0)$ is the only critical point of $f(x, y) = x^2 + y^2$, of $g(x, y) = -x^2 - y^2$, and of $h(x, y) = x^2 - y^2$. The point $(0, 0)$ is a critical point of $f(x, y) = 1/x$.

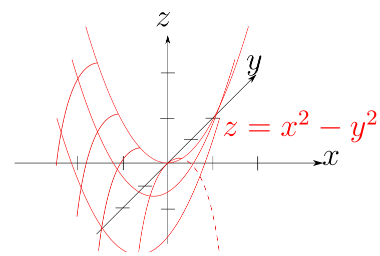


Example 8.6. Let $f(x, y) = |x| + |y|$. Then (a, b) is a critical point of f whenever $a = 0$ or $b = 0$. The global minimum of f occurs at $(0, 0)$.

Note that $\nabla f(0, 0)$ is undefined.



Definition 8.7 (Saddle Point). Let $f: \mathbb{R}^2 \to \mathbb{R}$ and let $(a, b) \in \mathbb{R}^2$. We say that (a, b) is a **saddle point** of f if (a, b) is a critical point of f and if (a, b) is not a local extremum of f. Formally: (a, b) is a saddle point of f if (a, b) is a critical point of f, and for any $t > 0$, there exist points (x_1, y_1) and (x_2, y_2) such that $\|(x_1, y_1) - (a, b)\| < t$, $\|(x_2, y_2) - (a, b)\| < t$, and such that $f(x_1, y_1) > f(a, b) > f(x_2, y_2)$.



Example 8.8. The point $(0, 0)$ is a saddle point for $h(x, y) = x^2 - y^2$. We have $\nabla h(0, 0) = (0, 0)$, and for any $n > 1$, we have $h(1/n, 0) > h(0, 0) > h(0, 1/n)$.

Theorem 8.9. *Let $f: \mathbb{R}^2 \to \mathbb{R}$. If f has a local extremum at (a, b), then (a, b) is a critical point of f.*

Proof Sketch. Assume that $\nabla f(a, b)$ exists. Using the linear approximation of f for points (x, y) near (a, b), we have

$$f(x, y) \approx f(a, b) + ((x, y) - (a, b)) \cdot \nabla f(a, b).$$

Let ε be a small number and choose $(x, y) = (a, b) + \varepsilon \nabla f(a, b)$, so that $f(x, y) \approx f(a, b) + \varepsilon \|\nabla f(a, b)\|$. If $\nabla f(a, b) \neq (0, 0)$, then we can choose $\varepsilon > 0$ to get $f(x, y) > f(a, b)$. And we can choose $\varepsilon < 0$ so that $f(x, y) < f(a, b)$. Since (a, b) is a local extremum of f, we cannot find points near (a, b) that lie above and below $f(a, b)$. We conclude that $\nabla f(a, b) = (0, 0)$, as desired. $\square$

From the previous Theorem, if we want to find the local extrema of a function, it suffices to find the critical points of f. Once we have found a critical point, we would like to identify it as a local maximum, local minimum, or saddle point. For functions of two variables, the following quantity allows such an identification.

Definition 8.10. Let $f: \mathbb{R}^2 \to \mathbb{R}$. Let $(a, b) \in \mathbb{R}^2$. Define the **discriminant** $D(a, b)$ of f at (a, b) by

$$D(a, b) = \det \begin{pmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{pmatrix} = f_{xx}(a, b) f_{yy}(a, b) - (f_{xy}(a, b))^2.$$

$D(a, b)$ is sometimes called the **Hessian** of f at (a, b).

Theorem 8.11 (Second Derivative Test on $\mathbb{R}^2$). *Let $f: \mathbb{R}^2 \to \mathbb{R}$. Let $(a, b) \in \mathbb{R}^2$. Assume that (a, b) is a critical point of f, and that the second derivatives f_{xx}, f_{xy}, f_{yy} exist and are continuous near (a, b). Then*

- *If $D > 0$ and if $f_{xx}(a, b) > 0$, then (a, b) is a local minimum of f.*
- *If $D > 0$ and if $f_{xx}(a, b) < 0$, then (a, b) is a local maximum of f.*
- *If $D < 0$, then (a, b) is a saddle point of f.*
- *If $D = 0$, then no conclusion can be drawn in general.*

Example 8.12. If $f(x,y) = x^2 + y^2$, then $D(0,0) = 4 > 0$. Since $f_{xx}(0,0) = 2 > 0$, we know that $(0,0)$ is a local minimum of f.

If $g(x,y) = -x^2 - y^2$, then $D(0,0) = 4 > 0$. Since $f_{xx}(0,0) = -2 < 0$, we know that $(0,0)$ is a local maximum of g.

If $h(x,y) = x^2 - y^2$, then $D(0,0) = -4 < 0$, so $(0,0)$ is a saddle point of h.

If $p(x,y) = x^4 + y^4$, then $D(0,0) = 0$, but $(0,0)$ is a global minimum of p. If $p(x,y) = -x^4 - y^4$, then $D(0,0) = 0$, but $(0,0)$ is a global maximum of p. If $p(x,y) = x^4 - y^4$, then $D(0,0) = 0$, but $(0,0)$ is saddle point of p.

Proof sketch of Theorem 8.11. For simplicity, assume $(a,b) = (0,0)$. It turns out that there are multivariable versions of Taylor's Theorem. In the present case, we have the following Taylor expansion near $(0,0)$. (Note the resemblance to the linear approximation of f.)

$$f(x,y) \approx f(0,0) + (x,y) \cdot \nabla f(0,0) + x^2 f_{xx}(0,0)/2 + y^2 f_{yy}(0,0)/2 + xy f_{xy}(0,0) + \cdots$$

If $\nabla f(0,0) = 0$, then the first-order terms go away, so that we have

$$f(x,y) \approx f(0,0) + x^2 f_{xx}(0,0)/2 + y^2 f_{yy}(0,0)/2 + xy f_{xy}(0,0) + \cdots$$

That is, the most significant part of the function is the quadratic term

$$q(x,y) = x^2 f_{xx}(0,0)/2 + y^2 f_{yy}(0,0)/2 + xy f_{xy}(0,0).$$

(Note that $q_{xx}(0,0) = f_{xx}(0,0)$, $q_{yy}(0,0) = f_{yy}(0,0)$ and $q_{xy}(0,0) = f_{xy}(0,0)$). To make things easier, we complete the square to get rid of the xy term.

$$q(x,y) = \frac{1}{2} f_{xx}(0,0) \left(x + y \frac{f_{xy}(0,0)}{f_{xx}(0,0)} \right)^2 + \frac{1}{2 f_{xx}(0,0)} y^2 \left(f_{yy}(0,0) f_{xx}(0,0) - (f_{xy}(0,0))^2 \right).$$

If $D > 0$ and $f_{xx}(0,0) > 0$, then $q(x,y) > 0$, so $(0,0)$ is a local minimum. The other cases are handled similarly. □

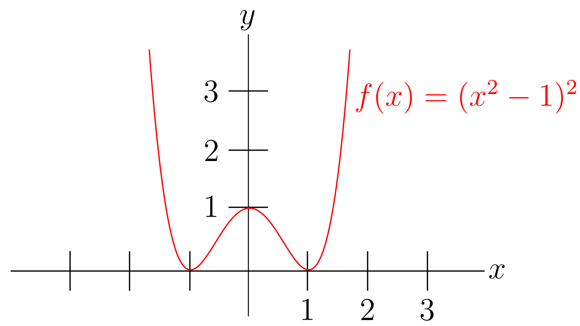
Finding local extrema is nice, but it would be better to find the maximum or minimum values of a function on a given domain. Doing so requires more than just looking at the critical points of the function.

Definition 8.13. Let D be a domain in $\mathbb{R}^2$, and let $f\colon D \to \mathbb{R}$. The **global maximum** of f on D is the largest value of f on the domain D (if such a value exists). The **global minimum** of f on D is the smallest value of f on the domain D (if such a value exists).

Example 8.14. Let $f(x) = x$ on the domain $0 \leq x \leq 1$, so that $f\colon [0,1] \to \mathbb{R}$. The global maximum of f occurs at $x = 1$, and the global minimum of f occurs at $x = 0$. Note that $f'(x)$ is never zero, so we cannot identify the global maximum and minimum just by examining critical points. We also need to examine the boundary of the domain $0 \leq x \leq 1$, which in this case is exactly $x = 0$ and $x = 1$. Note also that if we consider f to have the domain $-\infty < x < \infty$, then f has no maximum or minimum value.

Example 8.15. Suppose we want to maximize $f(x) = (x^2 - 1)^2 = x^4 - 2x^2 + 1$, where $-\infty < x < \infty$. We see that $f'(x) = 4x(x^2 - 1)$, so that $f'(x) = 0$ when $x = 0, 1, -1$.

Also, $f''(x) = 12x^2 - 4$, so $f''(0) < 0$. That is, $x = 0$ is a local maximum. However, f has no global maximum on $-\infty < x < \infty$, since $\lim_{x\to\infty} f(x) = \infty$. Also, the other two critical points are local minima. So, even though there is only one local maximum, this does not necessarily imply that a global maximum exists. On the other hand, we can come up with a condition on the domain of the function f such that, if this condition is satisfied, then we can find the global maximum or minimum of f with an algorithm.



Definition 8.16. Let D be a domain in $\mathbb{R}^2$. Let $(a, b) \in \mathbb{R}^2$. We say that (a, b) is a **boundary point** if, given any $t > 0$, the open disk $\{(x, y) \in \mathbb{R}^2 \colon \|(x, y) - (a, b)\| < t\}$ contains at least one point in D, and it contains at least one point not in D. A domain D is **closed** if it contains all of its boundary points. A domain D is called **bounded** if there exists some $R > 0$ such that D is contained in the disk $\{(x, y) \in \mathbb{R}^2 \colon \|(x, y) - (a, b)\| < R\}$.

Remark 8.17. These definitions can be applied to Euclidean space of any dimension.

Example 8.18. The closed interval $[0, 1]$ is closed. Its boundary points are 0 and 1, and both points are contained in $[0, 1]$.

The open interval $(0, 1)$ is not closed. Its boundary points are 0 and 1, and both points are not contained in $(0, 1)$.

Both of the intervals $[0, 1]$ and $(0, 1)$ are bounded. However, the interval $(0, \infty)$ is not bounded.

As we saw from the above examples, if the domain of a function is not closed, then the global maximum may not exist. Similarly, if the domain of a function is not bounded, then the global maximum may not exist. Fortunately, if the domain of a continuous function is both closed and bounded, then the global maximum does exist. The following theorem is proven is a proof-based calculus class, otherwise known as analysis. The one-dimensional analogue of the theorem below was stated in Theorem 3.7.

Theorem 8.19 (Extreme Value Theorem). *Let D be a domain in $\mathbb{R}^2$. Assume that D is closed and bounded. Let $f\colon D \to \mathbb{R}$ be continuous. Then*

- *The function f achieves both its maximum and minimum values in D.*
- *The extreme values of f occur either at critical points of f in D, or on the boundary of D.*

We can then turn this Theorem into an algorithm for finding the extreme values of a function

Algorithm 2 (Finding Extreme Values). Let D be a domain in $\mathbb{R}^2$. Assume that D is closed and bounded. Let $f\colon D \to \mathbb{R}$ be continuous. To find the maximum and minimum values of f, we perform the following procedure:

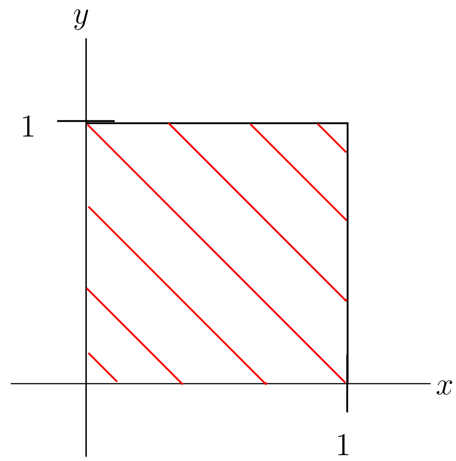
(i) Find all critical points of f in D.
(ii) Find the maximum and minimum of f on the boundary points of D.

(iii) Among the points found in parts (i) and (ii), choose those points with the largest and smallest values of f.

The maximum and minimum values of f on D must occur at the points found in Step (iii).

Remark 8.20. Step (ii) requires care when D is a domain in $\mathbb{R}^3$, as we demonstrate below.

Example 8.21. Consider $f(x, y) = x + y$ on the domain D where $0 \leq x \leq 1$ and $0 \leq y \leq 1$.



level curves of $f(x, y) = x + y$

We have $\nabla f(x, y) = (1, 1)$, so f has no critical points in D, and Step (i) is then complete. We now check the boundary of the domain. This boundary consists of four line segments, and we need to check each one separately. The first line segment is when $y = 0$ and $0 \leq x \leq 1$. In this case $f(x, y) = x$. We know $\partial f / \partial x = 1 \neq 0$ on this line, so the extrema on this line occur on its boundary, which is $x = 0$ and $x = 1$. So, we have added the points $(0, 0)$ and $(1, 0)$ to our list of points in Step (ii). The next line segment is when $y = 1$ and $0 \leq x \leq 1$. In this case $f(x, y) = x + 1$. We know $\partial f / \partial x = 1 \neq 0$ on this line, so the extrema on this line occur on its boundary, which is $x = 0$ and $x = 1$. So, we have added the points $(0, 1)$ and $(1, 1)$ to our list of points in Step (ii). Similarly, we can check the third and fourth line segments ($x = 0$, $0 \leq y \leq 1$ and $x = 1$, $0 \leq y \leq 1$), which add the points $(0, 0), (0, 1)$ and $(1, 0), (1, 1)$ respectively to our list of points in Step (ii). However, we already added these points to our list, so we did not add any new points.

In conclusion, from Steps (i) and (ii), we found a list of all possible candidate extrema, and this list is: $(0, 0), (0, 1), (1, 0)$ and $(1, 1)$. So, it remains to check the values of f at these points. We have $f(0, 0) = 0$, $f(1, 0) = 1$, $f(0, 1) = 1$ and $f(1, 1) = 2$. The minimum and maximum of these values are the minimum and maximum of f on D. That is, the minimum value of f on D is 0, and it occurs only at $(0, 0)$. And the maximum value of f on D is 2, and it occurs only at $(1, 1)$.

Example 8.22. Consider $f(x, y) = x^2 + y^2 + z^2$ on the domain D where $-1 \leq x \leq 1$, $-1 \leq y \leq 1$ and $-1 \leq z \leq 1$. We have $\nabla f(x, y) = (2x, 2y, 2z)$, so f only has a critical point at $(0, 0, 0)$, and Step (i) is then complete. We now check the boundary of the domain. This boundary consists of six squares, and we need to check each one separately. The first square is $x = -1$, $-1 \leq y \leq 1$ and $-1 \leq z \leq 1$. In this case $f(x, y, z) = 1 + y^2 + z^2$. That is, we are confronted with a two-variable optimization of the function $g(y, z) = 1 + y^2 + z^2$ where $-1 \leq y \leq 1$ and $-1 \leq z \leq 1$. We therefore perform Algorithm 2 on the function g. We see that $\nabla g = (2y, 2z)$, so a critical point only occurs at $(y, z) = (0, 0)$. We then need to check the boundary of the square, which consists of four line segments. The first line segment occurs when $y = 1$ and $-1 \leq z \leq 1$. On this line, we have $g(y, z) = 2 + z^2$. We have $\partial g / \partial z = 2z$, so that $z = 0$ is the only critical point of g on the line. We finally add the boundary points of the line segment to our list of points, where $z = 1, -1$.

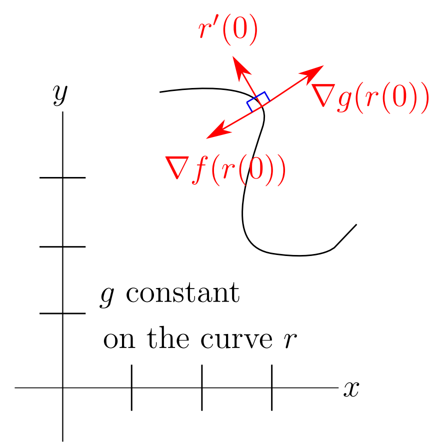
In summary, so far we have added the following points to our list of potential extreme values: $(0,0,0)$, $(-1,0,0)$, $(-1,1,0)$, $(-1,1,1)$, $(-1,1,-1)$. We still have to examine the three other sides of the square $x=-1$, $-1 \le y \le 1$ and $-1 \le z \le 1$. And we have to examine the five other squares of the domain D. Once we do this procedure, we will get a list of points consisting of all possible triples of points from the set $\{-1,0,1\}$.

It remains to check the values of f at these points. We have $f(0,0,0)=0$, $f(\pm 1,\pm 1,\pm 1)=3$, and all other candidate extrema have value 1 or 2. So, the minimum value of f on D is 0, and it occurs only at $(0,0,0)$. And the maximum value of f on D is 3, and it occurs only at the set of eight points $(\pm 1,\pm 1,\pm 1)$.

8.1. **Constrained Optimization: Lagrange Multipliers.** We have now reached the culmination of this chapter and of the course. As we have seen from the previous section, many optimization problems naturally require us to consider constraints. For example, when we tried to maximize the function $f(x,y,z)=x^2+y^2+z^2$ over the box $-1 \le x \le 1, -1 \le y \le 1, -1 \le z \le 1$, we were forced to optimize the function f on the boundary of the box. For example, we had to optimize the function f subject to the constraint $x=-1$, $-1 \le y \le 1$ and $-1 \le z \le 1$. This constraint was relatively easy to deal with, since we could just substitute the value $x=-1$ into the function f, and then optimize over y and z. However, for more general constraints, this substitution procedure **may not work**, or it could make things unnecessarily complicated.

Example 8.23. Consider the function $f(x,y)=x^2$ subject to the constraint $x+y=1$. There may not be a way to substitute the constraint $x+y=1$ into the function $f(x,y)=x^2$ to get a single variable function, which can be optimized. However, in this case, this procedure works: you could try to parametrize the line $x+y=1$ as $r(t)=(t,1-t)$, $t \in \mathbb{R}$, and then consider $f(x,y)=x^2=t^2$ which has a critical point only at $t=0$, which corresponds to the point $(0,1)$. And $(0,1)$ is the global minimum of f on $x+y=1$.

However, there is an automatic procedure which works in any dimension, and it does not require any parametrization.



Theorem 8.24 (**Lagrange Multipliers**). *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ and let $g\colon \mathbb{R}^2 \to \mathbb{R}$ be functions such that ∇f and ∇g exist and are continuous. Consider the problem of maximizing or minimizing f subject to the constraint $g(x,y)=0$. If f has a local maximum or local minimum at the point (a,b) subject to the constraint $g(x,y)=0$, and if $\nabla g(x,y) \neq (0,0)$, then there exists a real number λ such that*

$$\nabla f(a,b) = \lambda \nabla g(a,b).$$

Remark 8.25. The condition $\nabla f(a,b)=\lambda\nabla g(a,b)$ says that the vectors ∇f and ∇g are parallel.

Proof Sketch. Suppose $r(t)$ is a parametrization of the constraint curve $g(x,y)=0$ where $r(0)=(a,b)$. Then $f(r(t))$ is a function of the real variable t which has a local maximum or minimum at $t = 0$. That is, $(d/dt)f(r(t)) = 0$ when $t = 0$. From the Chain Rule, this equation says

$$\nabla f(r(0)) \cdot r'(0) = 0.$$

That is, $r'(0)$ is perpendicular to $\nabla f(a, b) = \nabla f(r(0))$. Since $r'(0)$ is parallel to the curve $g(x, y) = 0$ and $\nabla g(a, b)$ is perpendicular to the curve $g(x, y) = 0$, we see that $r'(0)$ is perpendicular to $\nabla g(a, b)$. And since $r'(0)$ is also perpendicular to $\nabla f(a, b)$, we see that $\nabla f(a, b)$ is parallel to $\nabla g(a, b)$. $\square$

Remark 8.26. This Theorem also holds for functions $f : \mathbb{R}^3 \to \mathbb{R}$ and constraints $g : \mathbb{R}^3 \to \mathbb{R}$.

Example 8.27. Maximize $f(x, y) = x^2 + y^2$ subject to the constraint $x^2 + 2y^2 = 1$.

We use the method of Lagrange Multipliers, with $f(x, y) = x^2 + y^2$ and $g(x, y) = x^2 + 2y^2 - 1$. We need to solve for x, y and λ in the equation

$$\nabla f(x, y) = \lambda \nabla g(x, y).$$

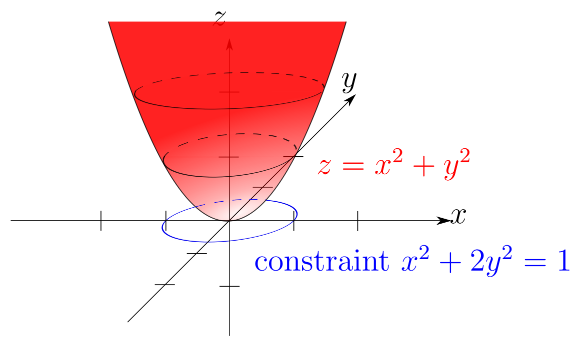
That is, we need to solve for

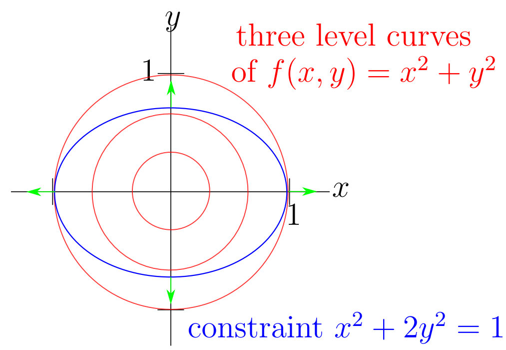
$$(2x, 2y) = \lambda(2x, 4y).$$

We have the system of three equations in three unknowns

$$\begin{cases} 2x = \lambda(2x) \\ 2y = \lambda(4y) \\ x^2 + 2y^2 = 1. \end{cases}$$

If $x \neq 0$, we have $\lambda = 1$, so that the second equation implies that $y = 0$. The third equation then implies that $x = 1$ or $x = -1$. So, the points $(1, 0)$ and $(-1, 0)$ are candidate maxima for f, subject to the constraint $g(x, y) = 0$.





In the remaining case that $x = 0$, the last equation implies that $y = 1/\sqrt{2}$ or $y = -1/\sqrt{2}$. So, the points $(0, 1/\sqrt{2})$ and $(0, -1/\sqrt{2})$ are candidate maxima for f.

In summary, the only possible maxima for f subject to the constraint $g(x, y) = 0$ are the four points: $(1, 0), (-1, 0), (0, 1/\sqrt{2})$ and $(0, -1/\sqrt{2})$. By inspecting the values of f at these points, we see that $f(1, 0) = f(-1, 0) = 1$ and $f(0, 1/\sqrt{2}) = (0, -1/\sqrt{2}) = 1/2$. The largest of these values is 1.

In conclusion, subject to the constraint $x^2+2y^2 = 1$, the function f has its global maximum of 1 at the points $(1, 0)$ and $(-1, 0)$. We did not ask for it, but the global minimum of f is $1/2$ occurring at the points $(0, 1/\sqrt{2})$ and $(0, -1/\sqrt{2})$. (Note that $\nabla g(x, y) \neq (0, 0)$ when $x^2 + 2y^2 = 1$, since if $x = 0$, then $y \neq 0$. So, the Lagrange Multiplier Theorem applies.)

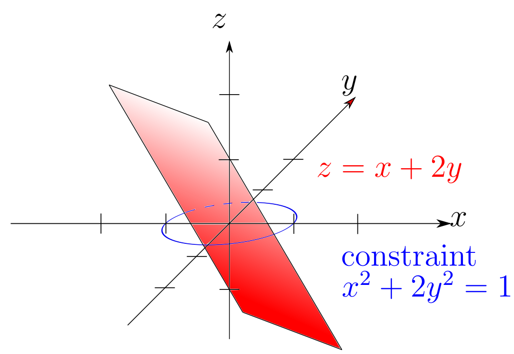
Example 8.28. Let's find the extreme values of the function $f(x, y) = x + 2y$ on the ellipse $x^2 + 2y^2 = 1$.

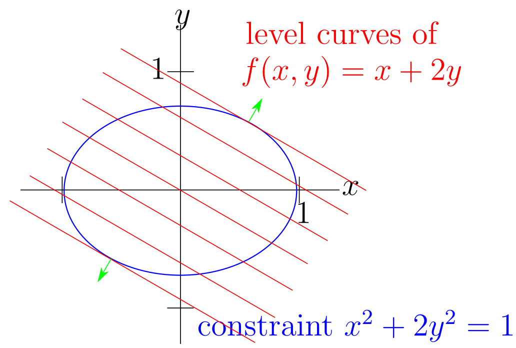
We use the method of Lagrange Multipliers, with $f(x, y) = x + 2y$ and $g(x, y) = x^2 + 2y^2 - 1$. We need to solve for x, y and λ in the equation

$$\nabla f(x, y) = \lambda \nabla g(x, y).$$

That is, we need to solve for

$$(1, 2) = \lambda(2x, 4y).$$





We have the system of three equations in three unknowns

$$\begin{cases} 1 = \lambda(2x) \\ 2 = \lambda(4y) \\ x^2 + 2y^2 = 1. \end{cases}$$

We have $2\lambda(2x) = 2 = \lambda(4y)$. So, if $\lambda \neq 0$, we have $4x = 4y$. That is, $x = y$. Substituting this into the last equation, we have

$$1 = 3x^2.$$

Solving for x, we get $x = \pm 1/\sqrt{3}$, so that $y = \pm 1/\sqrt{3}$ as well. So, two candidate points for extreme values are $(1/\sqrt{3}, 1/\sqrt{3})$ and $(-1/\sqrt{3}, -1/\sqrt{3})$.

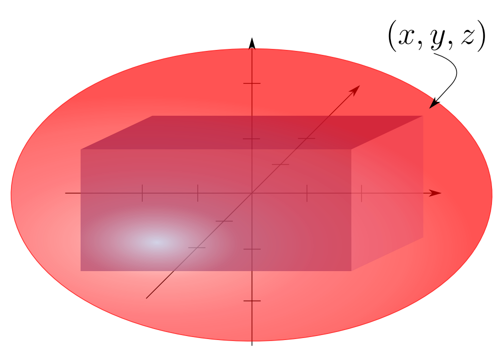
Note that the case $\lambda = 0$ cannot occur, since it would say that $(1, 2) = (0, 0)$, which is not true. So, the extreme values can only occur at $(1/\sqrt{3}, 1/\sqrt{3})$ and $(-1/\sqrt{3}, -1/\sqrt{3})$. By inspection, $f(1/\sqrt{3}, 1/\sqrt{3}) = 3/\sqrt{3} = \sqrt{3}$, and $f(-1/\sqrt{3}, -1/\sqrt{3}) = -3/\sqrt{3} = -\sqrt{3}$.

In conclusion, subject to the constraint $x^2 + 2y^2 = 1$, the function f has its global maximum of $\sqrt{3}$ at the point $(1/\sqrt{3}, 1/\sqrt{3})$ and its global minimum of $-\sqrt{3}$ at the point $(-1/\sqrt{3}, -1/\sqrt{3})$. (Note that $\nabla g(x, y) \neq (0, 0)$ when $x^2 + 2y^2 = 1$, since if $x = 0$, then $y \neq 0$. So, the Lagrange Multiplier Theorem applies.)

Example 8.29. Let $a, b, c > 0$. Find the dimensions of the box of maximal volume, whose edges are parallel to the coordinate axes, which can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Suppose the box B has vertices at the coordinates $(\pm x, \pm y, \pm z) \in \mathbb{R}^3$, where $x, y, z \geq 0$. Then the area of the box is $8xyz$. Also, since the vertices intersect the ellipsoid, we have $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$. To find the maximum area box, we therefore maximize the function $f(x, y, z) = xyz$ with the constraint $g(x, y, z) = x^2/a^2 + y^2/b^2 + z^2/c^2 - 1 = 0$, with $x, y, z, \geq 0$. Let



$$D = \{(x, y, z) \in \mathbb{R}^3 \colon x > 0, y > 0, z > 0\}.$$

Then

$$\nabla f = \begin{pmatrix} yz \\ xz \\ xy \end{pmatrix}, \qquad \nabla g = \begin{pmatrix} 2x/a^2 \\ 2y/b^2 \\ 2z/c^2 \end{pmatrix}.$$

(Since $\nabla g \neq (0,0,0)$ on D, the Lagrange Multiplier Theorem applies.)

We need to solve the equation $\nabla f = \lambda \nabla g$. This system of equations says

$$yz = \lambda 2x/a^2, \qquad xz = \lambda 2y/b^2, \qquad xy = \lambda 2z/c^2.$$

Since $\nabla f \neq (0,0,0)$ on D, we may assume that $\lambda \neq 0$. Substituting the first equation into the second gives $(yza^2/(2\lambda))z = 2\lambda y/b^2$, so $z^2 = 4\lambda^2/(a^2b^2)$. So, $z = 2\left|\lambda\right|/(ab)$. Substituting the second equation into the third gives $(xzb^2/(2\lambda))x = 2\lambda z/c^2$, so $x = 2\left|\lambda\right|/(bc)$. Similarly, $y = 2\left|\lambda\right|/(ac)$. Plugging these equalities for x, y, z into the condition $g(x,y,z) = 0$ shows that $12\lambda^2 = a^2b^2c^2$, so $2\sqrt{3}\left|\lambda\right| = abc$. So, the only critical point we have found in D is

$$(x,y,z) = (a/\sqrt{3}, b/\sqrt{3}, c/\sqrt{3}).$$

On the boundary of D, $f = 0$. Also, $f(a/\sqrt{3}, b/\sqrt{3}, c/\sqrt{3}) = abc3^{-3/2} > 0$. So, we have found the unique maximum of f. The box of maximal volume with edges parallel to the coordinate axes therefore has dimensions $2a/\sqrt{3}, 2b/\sqrt{3}, 2c/\sqrt{3}$.

Example 8.30. Suppose that we have a probability distribution on the set $\{1,\ldots,n\}$, i.e. we have a set of number $p_1,\ldots,p_n \in [0,1]$ such that $\sum_{i=1}^n p_i = 1$. A fundamental quantity for a probability distribution is its *entropy*

$$f(p_1,\ldots,p_n) = -\sum_{i=1}^n p_i \log p_i\,.$$

(We extend the function $x\log x$ to 0 by continuity, so that $0\log 0 = 0$.) The entropy of $p_1,\ldots,p_n$ measures the disorder or lack of information in p. This quantity is important in statistical physics.

We will try to maximize the entropy f subject to the constraint $\sum_{i=1}^n p_i = 1$. That is, we optimize f subject to the constraint $g(p_1,\ldots,p_n) = (\sum_{i=1}^n p_i) - 1 = 0$. Note that

$$\nabla f(p_1,\ldots,p_n) = \begin{pmatrix} -1-\log p_1 \\ \vdots \\ -1-\log p_n \end{pmatrix}, \qquad \nabla g(p_1,\ldots,p_n) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

(Since $\nabla g \neq (0,\ldots,0)$, the Lagrange Multiplier Theorem applies.)

We need to solve the equation $\nabla f(p_1,\ldots,p_n) = \lambda \nabla g(p_1,\ldots,p_n)$. This system of equations says

$$-1-\log p_1 = \lambda, \ldots, -1-\log p_n = \lambda.$$

That is, $p_1 = p_2 = \cdots = p_n$. Since $g(p_1,\ldots,p_n) = 0 = (\sum_{i=1}^n p_i) - 1 = np_1 - 1$, we conclude that $p_1 = 1/n$, so that $p_1 = \cdots = p_n = 1/n$. That is, we have found only one candidate extreme value of f subject to the constraint $g(p_1,\ldots,p_n) = 0$.

In order to optimize f, we now need to check the boundary of the domain. The boundary consists of $p_1,\ldots,p_n$ such that $\sum_{i=1}^n p_i = 1$, and there is some p_j which is equal to 1 or 0, so that $p_j \log p_j = 0$. So, when we are on the boundary of the domain, we have $n-1$ nonnegative numbers $p_1,\ldots,p_{j-1},p_{j+1},\ldots,p_n$ that add to 1. That is, we can exactly repeat the above calculation with $n-1$ replaced by n. We then see that the only critical point we find occurs when the $n-1$ numbers are equal to $1/(n-1)$. So, our next candidate extreme value of f occurs at such a point. We now need to check the boundary of this region, and so on. Iterating, we see that the only candidate extreme values occur when some of the numbers are zero, and the rest are equal to each other.

So, our candidate extreme values occur when $(p_1,\ldots,p_n) = (1/n,\ldots,1/n)$, $(p_1,\ldots,p_n) = (0, 1/(n-1),\ldots,1/(n-1))$, $(p_1,\ldots,p_n) = (0,0,1/(n-2),\ldots,1/(n-2))$, and so on, along with any permutation of these points. By inspection, we have

$$f(1/n,\ldots,1/n) = \sum_{i=1}^{n} -(1/n)\log(1/n) = (n/n)\log n = \log n.$$

$$f(0,1/(n-1),\ldots,1/(n-1)) = \sum_{i=1}^{n-1} -\frac{1}{n-1}\log\left(\frac{1}{n-1}\right) = \frac{n-1}{n-1}\log(n-1) = \log(n-1).$$

And so on. We therefore see that the maximum entropy occurs at the point $(1/n,\ldots,1/n)$. (And the minimum entropy occurs at the point $(0,\ldots,0,1)$.)

9. Double Integrals

Let $a<b$ be real numbers, and let $f\colon [a,b]\to\mathbb{R}$ be a continuous function. Recall that the integral $\int_a^b f(x)dx$ is approximated by **Riemann sums**. That is, if

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b,$$

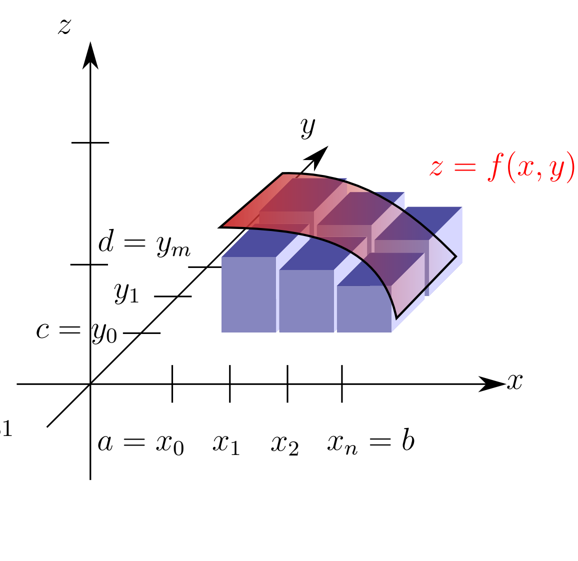
then $\int_a^b f(x)dx$, i.e. the area under the curve f, is approximated by the areas under boxes that approximate the function f. For example, if we evaluate f at the left endpoints of the boxes, we have

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n}(x_i - x_{i-1})f(x_{i-1}).$$

The term on the right is the sum of areas of n boxes, where the i^{th} box has width $(x_i - x_{i-1})$ and height $f(x_{i-1})$. More specifically, we have the following limiting expression for the integral

$$\int_a^b f(x)dx = \lim_{(\max_{i=1}^n (x_i - x_{i-1}))\to 0} \sum_{i=1}^{n}(x_i - x_{i-1})f(x_{i-1}). \qquad (*)$$

That is, no matter how we choose the points $a = x_0 < x_1 < \cdots < x_n = b$, as long as the quantity $(\max_{i=1}^n (x_i - x_{i-1}))$ goes to zero as $n\to\infty$, then the right side of $(*)$ approaches some real number. And that real number is called the **integral of** f from a to b, and it is denoted by the left side of $(*)$.



For a two-variable function, it turns out that we can adapt this procedure. Let $a<b$ and let $c<d$ be real numbers. Let $f\colon [a,b]\times[c,d]\to\mathbb{R}$ be a continuous function. We then split up the domain $[a,b]\times[c,d]$ into boxes. That is, let

$$a = x_0 < x_1 < \cdots < x_n = b,$$

$$c = y_0 < y_1 < \cdots < y_m = d.$$

We will then approximate the area under the function f by boxes. If $1 \leq i \leq n$ and if $1 \leq j \leq m$, then the box labelled (i,j) will have area $(x_i - x_{i-1})(y_j - y_{j-1})$. If we choose to evaluate f at the "lower-left" corner of each box, then this box will have height $f(x_{i-1}, y_{j-1})$. The volumes of these boxes will then form a **Riemann sum** as follows.

$$\int_{y=c}^{y=d} \int_{x=a}^{x=b} f(x,y)dxdy \approx \sum_{j=1}^{m} \sum_{i=1}^{n} (x_i - x_{i-1})(y_j - y_{j-1}) f(x_{i-1}, y_{j-1}).$$

More specifically, we have the following limiting expression for the integral

$$\int_{y=c}^{y=d} \int_{x=a}^{x=b} f(x,y)dxdy = \lim_{\substack{(\max_{i=1}^{n}(x_i - x_{i-1}))\to 0 \\ (\max_{j=1}^{m}(y_j - y_{j-1}))\to 0}} \sum_{j=1}^{m} \sum_{i=1}^{n} (x_i - x_{i-1})(y_j - y_{j-1}) f(x_{i-1}, y_{j-1}). \qquad (*)$$

That is, no matter how we choose the points $a = x_0 < x_1 < \cdots < x_n = b$ and $c = y_0 < y_1 < \cdots < y_m = d$, as long as the quantities $(\max_{i=1}^{n}(x_i - x_{i-1}))$ and $(\max_{j=1}^{m}(y_j - y_{j-1}))$ go to zero as $n, m \to \infty$, then the right side of $(*)$ approaches some real number. And that real number is called the **integral of** f on $[a,b] \times [c,d]$, and it is denoted by the left side of $(*)$.

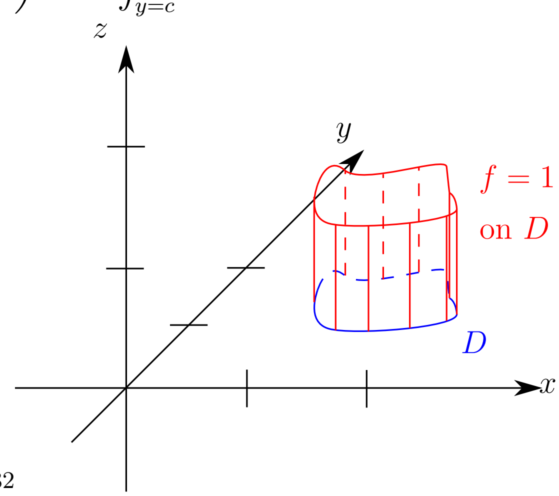
Remark 9.1. The same idea holds for an integral of $f(x,y)$ over a more general region D of the plane, but the Riemann sum is just more difficult to write. In the case that we integrate over a region D of the plane, we denote this integral by

$$\iint_D f dA.$$

Remark 9.2. In practice, computing a double integral is often done by computing two separate single-variable integrals.

Example 9.3. Let $a < b$ and let $c < d$ be real numbers. For any $a \leq x \leq b$ and $c \leq y \leq d$, let $f(x,y) = 1$. We evaluate the integral of f over $[a,b] \times [c,d]$ by first integrating in the x variable (while considering y fixed), and then by integrating in the y variable.

$$\int_{y=c}^{y=d} \int_{x=a}^{x=b} f(x,y)dxdy = \int_{y=c}^{y=d} \left(\int_{x=a}^{x=b} dx \right) dy = \int_{y=c}^{y=d} (b-a)\, dy = (d-c)(b-a).$$



So, $\int_{y=c}^{y=d} \int_{x=a}^{x=b} dxdy$ gives the **area** of the rectangle $[a,b] \times [c,d]$. More generally, if D is a region in the plane, then the area of D is given by $\iint_D dA$.

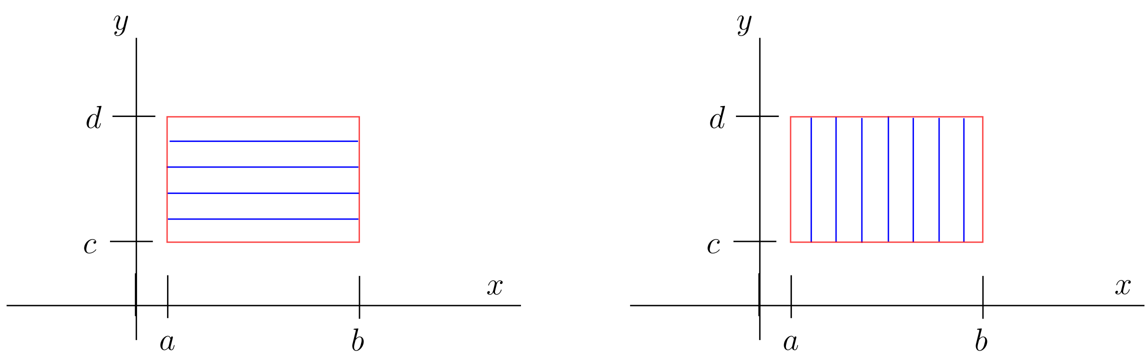
Example 9.4. For any $0 \leq x \leq 1$ and $0 \leq y \leq 1$ let $f(x,y) = x + y$. We evaluate the integral of f by first integrating in the x variable (while considering y fixed), and then by integrating in the y variable. That is,

$$\int_{y=0}^{y=1}\int_{x=0}^{x=1} f(x,y)dxdy = \int_{y=0}^{y=1}\left(\int_{x=0}^{x=1}(x+y)dx\right)dy = \int_{y=0}^{y=1}\left[((1/2)x^2+yx)|_{x=0}^{x=1}\right]dy$$
$$= \int_{y=0}^{y=1}((1/2)+y)\,dy = ((1/2)y+(1/2)y^2)|_{y=0}^{y=1}$$
$$= (1/2)+(1/2) = 1.$$

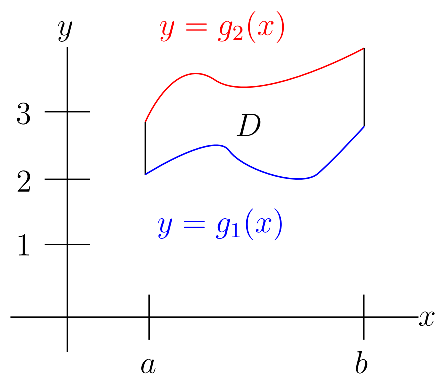
Remark 9.5. When we integrate on a rectangle $[a,b]\times[c,d]$, then we can change the order of integration. We can also change the order of integration when we integrate over general regions D, but we then need to be careful about describing the region D correctly.

Theorem 9.6 (**Change of Order of Integration**)**.** *Let $a<b$ and let $c<d$ be real numbers. Let $f\colon[a,b]\times[c,d]\to\mathbb{R}$ be a continuous function. Let D denote the domain $D=[a,b]\times[c,d]$. Then*

$$\iint_D f dA = \int_{y=c}^{y=d}\int_{x=a}^{x=b} f(x,y)dxdy = \int_{x=a}^{x=b}\int_{y=c}^{y=d} f(x,y)dydx.$$



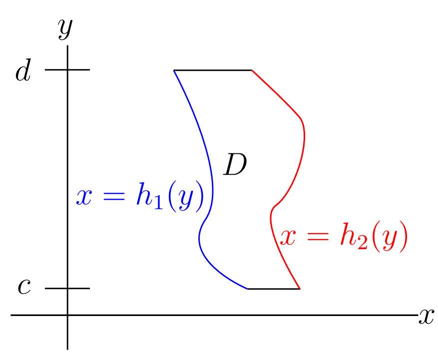
In order to change the order of integration of a more general region D, we need to make sure that the limits of the double integral define the region correctly.



Theorem 9.7 (**Integration over General Regions D**)**.** *Let D be a region in the plane $\mathbb{R}^2$. Let $a<b$ and let $c<d$ be real numbers.*

- *Let $g_1\colon[a,b]\to\mathbb{R}$ and let $g_2\colon[a,b]\to\mathbb{R}$ be continuous functions. Suppose D is defined as the set of all (x,y) in the plane such that $a\leq x\leq b$ and such that $g_1(x)\leq y\leq g_2(x)$. Then we compute the integral $\iint_D f dA$ as follows.*

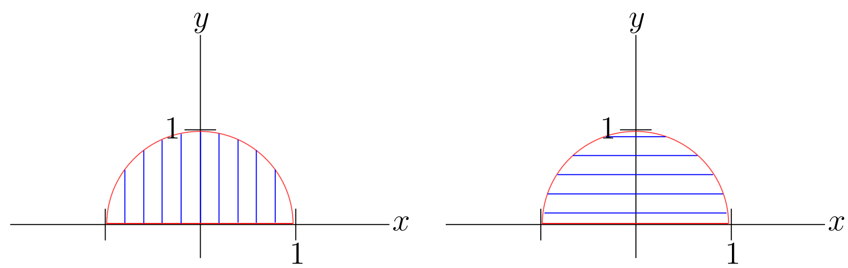
$$\iint_D f dA = \int_{x=a}^{x=b}\int_{y=g_1(x)}^{y=g_2(x)} f(x,y)dydx.$$



- *Let $h_1\colon [c,d] \to \mathbb{R}$ and let $h_2\colon [c,d] \to \mathbb{R}$ be continuous functions. Suppose D is defined as the set of all (x,y) in the plane such that $c \leq y \leq d$ and such that $h_1(y) \leq x \leq h_2(x)$. Then we compute the integral $\iint_D f dA$ as follows.*

$$\iint_D f dA = \int_{y=c}^{y=d} \int_{x=h_1(y)}^{x=h_2(y)} f(x,y) dx dy.$$

Remark 9.8. Note that both items in this Theorem are the same, if we reverse the roles of x and y.



Example 9.9. Consider the region D defined as the set of all (x,y) in the plane such that $-1 \leq x \leq 1$ and such that $0 \leq y \leq \sqrt{1-x^2}$. Then D is a half of a disc. We will integrate D in two different ways, and we will get the same answer. Consider $f(x,y) = 2y$. Then

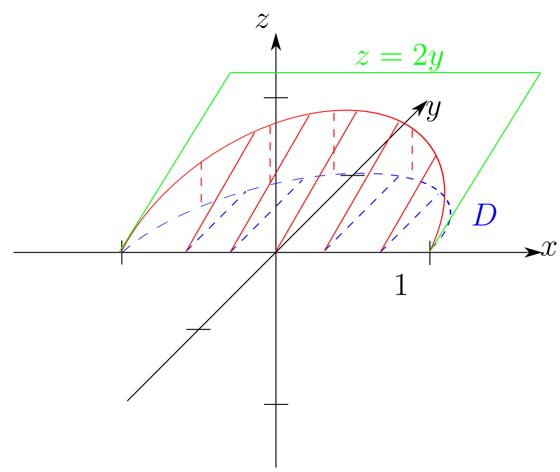
$$\begin{aligned}\iint_D f dA &= \int_{x=-1}^{x=1} \int_{y=0}^{y=\sqrt{1-x^2}} 2y dy dx = \int_{x=-1}^{x=1} [(y^2)|_{y=0}^{y=\sqrt{1-x^2}}] dx \\ &= \int_{x=-1}^{x=1} (1-x^2) dx = [x - (1/3)x^3]|_{x=-1}^{x=1} \\ &= 1 - (1/3) + 1 - (1/3) = 4/3.\end{aligned}$$

We now compute this integral by first integrating with respect to x, and then with respect to y. Note that D can be defined as the set of all (x,y) in the plane such that $0 \leq y \leq 1$ and such that $-\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2}$. So,

$$\begin{aligned}\iint_D f dA &= \int_{y=0}^{y=1} \int_{x=-\sqrt{1-y^2}}^{x=\sqrt{1-y^2}} 2y dx dy = \int_{y=0}^{y=1} (2y)(2\sqrt{1-y^2}) dy \\ &= \int_{u=1}^{u=0} (-2)\sqrt{u} du \qquad \text{, substituting } u = 1-y^2 \\ &= \int_{u=0}^{u=1} 2\sqrt{u} du = (4/3)u^{3/2}|_{u=0}^{u=1} = 4/3.\end{aligned}$$

Remark 9.10. In the above example, note that since $f(x,y) \geq 0$ on D,

$\iint_D f(x,y)dxdy$ gives the **volume** between the surfaces $z = f(x,y)$ and $z = 0$, over the region D.



Remark 9.11 (**An Important Remark concerning the Square Root Function**). In high school, you learned that the number 9 has two square roots 3 and -3. When we write the symbol $\sqrt{9}$, we mean that the symbol $\sqrt{9}$ stands for the nonnegative number that, when squared, gives 9. So, $\sqrt{9} = 3$. So, do not confuse the **square root function** $\sqrt{x}$, which is a function from $[0,\infty)$ to $[0,\infty)$, with the **square roots of a number**. The square root function only has one value, and not two values.

Since the square root function applied to a positive number always gives a positive number, the following identity holds for any real number x

$$\sqrt{x^2} = |x| \,.$$

We used this identity in the above example. Note that, if x is negative, we have $\sqrt{x^2} \neq x$.

Example 9.12. Sometimes we are given a double integral, and we need to change the order of integration, since such a change may simplify the calculation of the integral. In such a case, drawing a picture of the region is helpful. For example, suppose we have the integral

$$\int_{x=0}^{x=1} \int_{y=x^2}^{y=x} f(x,y)dydx.$$

By drawing a picture, we can see that the set of (x,y) in the plane with $0 \leq x \leq 1$ and with $x^2 \leq y \leq x$ is equal to the set of (x,y) in the plane with $0 \leq y \leq 1$ and with $x = y$ to $x = \sqrt{y}$. That is, we have

$$\int_{x=0}^{x=1} \int_{y=x^2}^{y=x} f(x,y)dydx = \int_{y=0}^{y=1} \int_{x=y}^{x=\sqrt{y}} f(x,y)dxdy.$$

The following properties are useful for computing various integrals. These properties should be familiar from the setting of single-variable integrals.

Proposition 9.13 (**Properties of Integrals**). *Let f and g be continuous functions in the plane, and let D be a region in the plane.*

- *If c is a real number, then $\iint_D cf dA = c\iint_D f dA$.*
- *$\iint_D (f+g)dA = \iint_D f dA + \iint_D g dA$.*
- *$\iint_D (f-g)dA = \iint_D f dA - \iint_D g dA$.*
- *If $f(x,y) \geq 0$ for all (x,y) in D, then $\iint_D f dA \geq 0$.*
- *If $f(x,y) \geq g(x,y)$ for all (x,y) in D, then $\iint_D f dA \geq \iint_D g dA$.*
- *If D is the union of two nonoverlapping regions D_1 and D_2, then*

$$\iint_D f dA = \iint_{D_1} f dA + \iint_{D_2} f dA.$$

10. Triple Integrals

Triple integrals are constructed in an analogous way to double integrals. We now sketch the idea.

Let $a < b$ and let $c < d$ and let $e < f$ be real numbers. Let $F\colon [a,b]\times[c,d]\times[e,f]\to\mathbb{R}$ be a continuous function. We then split up the domain $[a,b]\times[c,d]\times[e,f]$ into boxes. That is, let

$$\begin{aligned} a &= x_0 < x_1 < \cdots < x_n = b \\ c &= y_0 < y_1 < \cdots < y_m = d \\ e &= z_0 < z_1 < \cdots < z_\ell = f. \end{aligned}$$

We will then approximate the integral of f by boxes. If $1 \leq i \leq n$, if $1 \leq j \leq m$, and if $1 \leq k \leq \ell$, then the box labelled (i,j,k) will have volume $(x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1})$. If we choose to evaluate f at the "lower-left" corner of each box, then this box will have height $f(x_{i-1}, y_{j-1}, z_{k-1})$. The volumes of these boxes multiplied by the value of f on the corner of the boxes will then form a **Riemann sum** as follows.

$$\begin{aligned} &\int_{z=e}^{z=f}\int_{y=c}^{y=d}\int_{x=a}^{x=b} F(x,y,z)dxdydz \\ &\approx \sum_{k=1}^{\ell}\sum_{j=1}^{m}\sum_{i=1}^{n}(x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1})F(x_{i-1}, y_{j-1}, z_{k-1}). \end{aligned}$$

More specifically, we have the following limiting expression for the integral

$$\begin{aligned} &\int_{z=e}^{z=f}\int_{y=c}^{y=d}\int_{x=a}^{x=b} F(x,y,z)dxdydz \\ &= \lim_{\substack{(\max_{i=1}^{n}(x_i - x_{i-1}))\to 0 \\ (\max_{j=1}^{m}(y_j - y_{j-1}))\to 0 \\ (\max_{k=1}^{\ell}(z_k - z_{k-1}))\to 0}} \sum_{k=1}^{\ell}\sum_{j=1}^{m}\sum_{i=1}^{n}(x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1})F(x_{i-1}, y_{j-1}, z_{k-1}). \qquad (*) \end{aligned}$$

That is, no matter how we choose the points $a = x_0 < x_1 < \cdots < x_n = b$, $c = y_0 < y_1 < \cdots < y_m = d$, and $e = z_0 < z_1 < \cdots < z_\ell = f$, as long as the quantities $(\max_{i=1}^{n}(x_i - x_{i-1}))$, $(\max_{j=1}^{m}(y_j - y_{j-1}))$ and $(\max_{k=1}^{m}(z_k - z_{k-1}))$ go to zero as $n, m, \ell \to \infty$, then the right side of $(*)$ approaches some real number. And that real number is called the **integral of** F on $[a,b]\times[c,d]\times[e,f]$, and it is denoted by the left side of $(*)$.

Remark 10.1. The same idea holds for an integral of $F(x,y,z)$ over a more general region D of Euclidean space, but the Riemann sum is just more difficult to write. In the case that we integrate over a region D of Euclidean space, we denote this integral by

$$\iiint_D F dV.$$

Remark 10.2. In practice, computing a triple integral is often done by computing three separate single-variable integrals.

Example 10.3. Let $a < b$ and let $c < d$ and let $e < f$ be real numbers. For any $a \leq x \leq b$, $c \leq y \leq d$, and $e \leq z \leq f$, let $F(x,y,z) = 1$. We evaluate the integral of F over $[a,b]\times[c,d]\times[e,f]$ by first integrating in the x variable (while considering y and z fixed), and then by integrating in the y variable (while considering z fixed), and then by integrating in the z variable. That is,

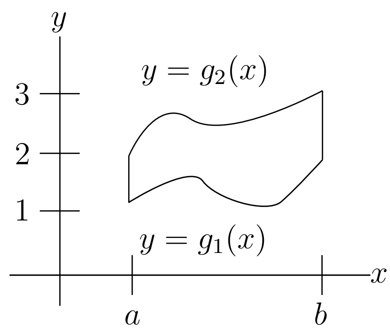
$$\begin{aligned}\int_{z=e}^{z=f}\int_{y=c}^{y=d}\int_{x=a}^{x=b} F(x,y,z)dxdydz \\ &= \int_{z=e}^{z=f}\int_{y=c}^{y=d}\left(\int_{x=a}^{x=b} dx\right)dydz = \int_{z=e}^{z=f}\left(\int_{y=c}^{y=d}(b-a)\,dy\right)dz \\ &= \int_{z=e}^{z=f}(d-c)(b-a)dz = (f-e)(d-c)(b-a).\end{aligned}$$

So, $\int_{z=e}^{z=f}\int_{y=c}^{y=d}\int_{x=a}^{x=b} dxdydz$ gives the **volume** of the box $[a,b]\times[c,d]\times[e,f]$. More generally, if D is a region in Euclidean space $\mathbb{R}^3$, then the volume of D is given by $\iiint_D dV$.

Remark 10.4 (**Visualizing Higher Dimensions**)**.** For a function $F(x,y)$ of two variables x,y, we often think of $F(x,y)$ as the height of the function F above the point (x,y). Using this same visualization in three-dimensions doesn't work so well, since a function value $F(x,y,z)$ of three variables (x,y,z) would be interpreted as a fourth-dimension. Instead of thinking of $F(x,y,z)$ as a height, you can think of $F(x,y,z)$ being the color of a point (x,y,z). For example, if $F(x,y,z)$ is large and positive, you can think of the point (x,y,z) as colored dark gray, and if $F(x,y,z)$ is very negative, you can think of the point (x,y,z) as colored light gray. So, a function on three-dimensional Euclidean space can be thought of as a grayscale coloring of Euclidean space.

Theorem 10.5 (**Integration over General Regions D**)**.** *Let D be a region in the Euclidean space $\mathbb{R}^3$. Let $a<b$, $c<d$ and let $e<f$ be real numbers. Let F be a continuous function on $\mathbb{R}^3$.*

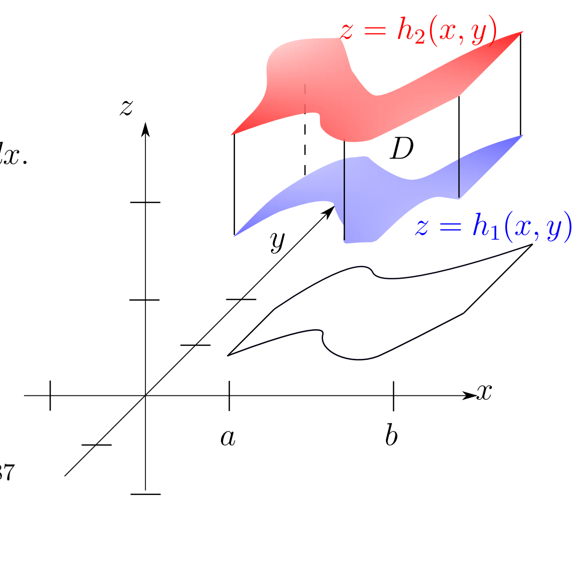
Let $g_1\colon[a,b]\to\mathbb{R}$ and let $g_2\colon[a,b]\to\mathbb{R}$ be continuous functions. Let $h_1\colon[a,b]\times[c,d]\to\mathbb{R}$ and let $h_2\colon[a,b]\times[c,d]\to\mathbb{R}$ be continuous functions. Suppose D is defined as the set of all (x,y,z) in $\mathbb{R}^3$ such that:



- $a\le x\le b$
- $g_1(x)\le y\le g_2(x)$
- $h_1(x,y)\le z\le h_2(x,y)$

Then we compute the integral $\iiint_D FdV$ as follows.

$$\iiint_D FdV = \int_{x=a}^{x=b}\int_{y=g_1(x)}^{y=g_2(x)}\int_{z=h_1(x,y)}^{z=h_2(x,y)} F(x,y,z)dzdydx.$$



Remark 10.6. We can permute the roles of x,y,z in the above theorem as needed in applications.

Example 10.7. Find the volume V of the region above the rectangle where $0\le x\le 2$,

$0 \leq y \leq 1$ and $z = 0$, bounded by the plane
$y + z = 1$. Let's find the limits of integration and apply Theorem 10.5. C is defined as the set of points where $0 \leq x \leq 2$, where $0 \leq y \leq 1$, and where $0 \leq z \leq 1 - y$. So, the volume V of C satisfies

$$\begin{aligned} V &= \int_{x=0}^{x=2} \int_{y=0}^{y=1} \int_{z=0}^{z=1-y} dz dy dx = \int_{x=0}^{x=2} \int_{y=0}^{y=1} (1-y) dy dx \\ &= \int_{x=0}^{x=2} (y - (1/2)y^2)|_{y=0}^{y=1} dx = \int_{x=0}^{x=2} (1/2) dx = 1. \end{aligned}$$

Triple integrals satisfy the same properties as double integrals.

Proposition 10.8 (**Properties of Integrals**)**.** *Let f and g be continuous functions on $\mathbb{R}^3$, and let D be a region in $\mathbb{R}^3$*

- *If c is a real number, then $\iiint_D cf dV = c \iiint_D f dV$.*
- *$\iiint_D (f+g) dV = \iiint_D f dV + \iiint_D g dV$.*
- *$\iiint_D (f-g) dV = \iiint_D f dV - \iiint_D g dV$.*
- *If $f(x,y,z) \geq 0$ for all (x,y,z) in D, then $\iiint_D f dV \geq 0$.*
- *If $f(x,y,z) \geq g(x,y,z)$ for all (x,y,z) in D, then*

$$\iiint_D f dV \geq \iiint_D g dV.$$

- *If D is the union of two nonoverlapping regions D_1 and D_2, then*

$$\iiint_D f dV = \iiint_{D_1} f dV + \iiint_{D_2} f dV.$$

10.1. **Applications of Triple Integrals.**

Definition 10.9 (**Average Value**)**.** Let D be a region in Euclidean space $\mathbb{R}^3$. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a function. Then the **average value** of f on D is defined as

$$\frac{\iiint_D f dV}{\iiint_D dV}.$$

Definition 10.10 (**Mass**)**.** Let D be a region in Euclidean space $\mathbb{R}^3$. For each point (x,y,z) in D, let $f(x,y,z)$ denote the density of the object D. Then the **mass** M of D is defined as

$$M = \iiint_D f(x,y,z) dx dy dz.$$

Definition 10.11 (**Center of Mass**)**.** Let D be a region in Euclidean space $\mathbb{R}^3$. For each point (x,y,z) in D, let $f(x,y,z)$ denote the density of the object D. Then the **center of mass** of D is defined as the following point in Euclidean space $\mathbb{R}^3$.

$$\left(\frac{1}{M} \iiint_D x f(x,y,z) dx dy dz, \frac{1}{M} \iiint_D y f(x,y,z) dx dy dz, \frac{1}{M} \iiint_D z f(x,y,z) dx dy dz \right).$$

Definition 10.12 (**Moment of inertia**)**.** Let D be a region in Euclidean space $\mathbb{R}^3$. For each point (x,y,z) in D, let $f(x,y,z)$ denote the density of the object D. Then the **moment of inertia** of D about the z axis is defined as

$$\iiint_D (x^2 + y^2) f(x,y,z) dx dy dz.$$

The **moment of inertia** of D about the y axis is defined as

$$\iiint_D (x^2+z^2)f(x,y,z)dxdydz.$$

The **moment of inertia** of D about the x axis is defined as

$$\iiint_D (y^2+z^2)f(x,y,z)dxdydz.$$

The moment of inertia measures how difficult it is to rotate an object about an axis.

Exercise 10.13. Suppose I want to design a structure of bounded height and with minimal moment of inertia. Specifically, suppose I have a region D in Euclidean space $\mathbb{R}^3$, and D lies between the planes $\{(x,y,z)\in\mathbb{R}^3 \colon z=0\}$ and $\{(x,y,z)\in\mathbb{R}^3 \colon z=1\}$. Suppose also that D has uniform density, and the mass of D is equal to 1. I then want to find the D with the smallest moment of inertia around the z axis. Which D should I use?

11. Review of Course

11.0.1. *Limits.*

- A limit may or may not exist. $\lim_{x\to 0} x = 0$, $\lim_{x\to 0} 1/x$ DNE.
- If f is continuous, then $\lim_{x\to a} f(x) = f(a)$.
- If f is a polynomial, then $\lim_{x\to a} f(x) = f(a)$.
- If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist, then

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

- (**Intermediate Value Theorem**) If $f\colon [a,b]\to\mathbb{R}$ is continuous, then f takes every value between $f(a)$ and $f(b)$.

11.0.2. *Derivatives.*

- $f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$. If f is position, then f' is velocity, f'' is acceleration.
- (**Product Rule**) $\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + g'(x)f(x)$.
- (**Quotient Rule**) $\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{g(x)f'(x)-f(x)g'(x)}{(g(x))^2}$.
- (**Chain Rule**) $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$.
- **Extreme Value Theorem** (allows us to do optimization)
- **Linear Approximation**.

Example. Approximate $(16.32)^{1/4}$ using a linear approximation of $f(x)=x^{1/4}$ near $a=16$. $f'(x)=(1/4)x^{-3/4}$, $f'(16)=(1/4)2^{-3}=1/32$. So $f(x)\approx f(16)+(x-16)f'(16) = 2+(x-16)(1/32)$. So $f(16.32)\approx 2+(.32)/32 = 2.01$.

Is this number larger or smaller than $(16.32)^{1/4}$? $f''(x) = -(3/16)x^{-7/4}$, so f is concave down, so the tangent line at $x=16$ lies above the function f. So, $(16.32)^{1/4} < 2.01$.

11.0.3. *Graph Sketching.*

- (**Critical Point**) x where $f'(x)=0$, or $f'(x)$ does not exist.
- (**First Derivative Test**) $f'(x)>0$ means f is increasing, $f'(x)<0$ means f is decreasing.
- (**Second Derivative Test**) $f''(x)>0$ means f is concave up, $f''(x)<0$ means f is concave down.

- (**Mean Value Theorem**) $f'(c) = \frac{f(b)-f(a)}{b-a}$ for some c between a and b, if f' exists and is continuous.
- (**Rolle's Theorem**) If $f(a) = f(b)$, MVT reduces to: $f'(c) = 0$ for some c between a and b, if f' exists and is continuous.
- (**Extreme Value Theorem**) If $f\colon [a,b] \to \mathbb{R}$ is continuous, then the max and min of f exist on $[a,b]$.
- Exponential growth, compound interest

Example. Fix $a > 1$. Find the minimum distance from the point $(0,a)$ in the plane to the parabola $y = x^2$.

We are asked to minimize $f(x) = \sqrt{x^2 + (y-a)}^2$ on the parabola $y = x^2$. That is, we are asked to minimize $f(x) = \sqrt{x^2 + (x^2-a)}^2$ over all $x \in \mathbb{R}$. Equivalently, we need to minimize $F(x) = x^2 + (x^2-a)^2$ over all $x \in \mathbb{R}$. We solve $F'(x) = 2x + 4x(x^2-a) = 0$ so that critical points occur when $x = 0$ or $x^2 - a = -1/2$. Note that $F(0) = a^2$ and $F(\sqrt{a-1/2}) = F(-\sqrt{a-1/2}) = (a-1/2) + (1/2)^2 = a - 1/4$. Since $a > 1$, $a^2 > a$, so $a^2 > a - 1/4$. So, the minimum distance occurs when $x^2 = a - 1/2$. That is, the minimum distance is $\sqrt{a - 1/2 + (-1/2)^2} = \sqrt{a - 1/4}$.

11.0.4. *Integration.*

- (**Fundamental Theorem of Calculus**) $\int_a^b \frac{d}{dx} f(x) = f(b) - f(a)$, $\frac{d}{dx}\int_a^x f(t)dt = f(x)$. (Exercise. Compute $\frac{d}{dx}\int_{x^3}^{x^4}\cos(t^5)dt$)
- Integrating exponential and logarithm functions.
- Inverse derivative formula. $f(f^{-1}(x)) = x$, so $f'(f^{-1}(x))(d/dx)f^{-1}(x) = 1$.
- Integration by substitution, trigonometric substitution
- Integration by parts. $\int u dv = uv - \int v du$
- Improper integrals, integrating $1/\sqrt{x}$ near $x = 0$, integrating e^{-x} on $[0,\infty]$.

Some examples of integration, using e.g. substitution.

$\int_2^5 e^x e^{e^x} dx$. Substitute $u = e^x$, so $du = e^x dx$, get $\int_2^5 e^x e^{e^x} dx = \int_{u(2)}^{u(5)} e^u du = \int_{e^2}^{e^5} e^u du = e^{e^5} - e^{e^2}$.

$\int_{-3}^3 \frac{x}{10-x^2} dx$. Substitute $u = 10 - x^2$, so $du = -2x dx$, get $\int_{-3}^3 \frac{x}{10-x^2} dx = \int_{u(-3)}^{u(3)} (-1/2)u^{-1/2} du = \int_1^1 u^{-1/2} du = 0$.

$\int_{-1}^1 x\sqrt{1+x} dx$. Substitute $u = 1 + x$, so $du = dx$, get $\int_{-1}^1 x\sqrt{1+x} dx = \int_0^2 (u-1)u^{1/2} du = \int_0^2 u^{3/2} - u^{1/2} du = \cdots$.

$\int_0^1 x e^{x^2} dx$. Substitute $u = x^2$, so $du = 2x dx$, get $\int_0^1 x e^{x^2} dx = \int_{u(0)}^{u(1)} (1/2) e^u du = (1/2)\int_0^1 e^u du = (1/2)[e^u]_{u=0}^{u=1} = (1/2)[e^1 - e^0] = (1/2)(e-1)$.

$\int_{-1}^1 \frac{x^3}{e^x + e^{-x}} dx = 0$ since the integrand is an odd function integrated over an interval that is symmetric about $x = 0$.

$\int_0^1 \frac{d}{dx}(e^{4x}\ln((x+2)/(x+1)))dx = e \cdot \ln(3/2) - \ln(2)$, by the Fundamental Theorem of Calculus

$\int_{-1}^1 |x|\, dx = \int_0^1 x dx + \int_{-1}^0 (-x) dx = \int_0^1 x dx - \int_{-1}^0 (x) dx = 1/2 - (-1/2) = 1$.

11.0.5. *Multivariable Calculus.*

- Vectors, dot product, angle. $(a,b,c)\cdot(d,e,f) = ad + be + cf$. $\|(a,b,c)\| = \sqrt{a^2+b^2+c^2}$.

- $(a,b,c) \cdot (d,e,f) = \|(a,b,c)\| \, \|(d,e,f)\| \cos\theta$, where θ is the angle between the two vectors.
- Limits and continuity (nearly the same as for one variable).
- Partial derivatives. $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\frac{\partial}{\partial x} f(a,b) = \lim_{h\to 0} \frac{f(a+h,b)-f(a,b)}{h}$, $\frac{\partial}{\partial y} f(a,b) = \lim_{h\to 0} \frac{f(a,b+h)-f(a,b)}{h}$.
- Gradient. $\nabla f(a,b) = (\frac{\partial}{\partial x} f(a,b), \frac{\partial}{\partial y} f(a,b))$. Points in the direction of greatest increase of f. The gradient is orthogonal to the level curves of f (curves where f is constant).
- Directional derivative. $D_v f(a,b) = \nabla f(a,b) \cdot v$, $v \in \mathbb{R}^2$. The rate of change of f at (a,b) in the direction v.

Example. $f(x,y) = x^2 + 2xy + 3y^2$. $\nabla f(x,y) = (2x+y, 8y+2x)$. $D_{(1,1)} f(2,0) = \nabla f(2,0) \cdot (1,1) = (4,4) \cdot (1,1) = 8$. $D_{(1,0)} f(0,2) = \nabla f(0,2) \cdot (1,0) = (2,12) \cdot (1,0) = 2$.

Example. $\lim_{(x,y)\to(0,0)} \frac{x^2+xy+y^2}{x^2+y^2}$ DNE. (Along the line $x = 0$, get 1, and along the line $x = y$, get 3/2.)

- Tangent planes (similar to tangent lines, just with more variables). $f\colon \mathbb{R}^2 \to \mathbb{R}$, has an associated surface $z = f(x,y)$. The tangent plane at (a,b) is the plane $z = L(x,y) = f(a,b) + ((x,y) - (a,b)) \cdot \nabla f(a,b)$.
- (**Extreme Value Theorem**) If $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a continuous function on a closed and bounded set D. Then the max and min of f exist on D.
- Critical point. x such that $\nabla f(x) = 0$ or one component of $\nabla f(x)$ does not exist.
- (**Second Derivative Test** for critical points x) Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ with $\nabla f(x) = 0$. Let
$$D(a,b) = \det \begin{pmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{pmatrix} = f_{xx}(a,b) f_{yy}(a,b) - (f_{xy}(a,b))^2.$$
  - If $D > 0$ and if $f_{xx}(a,b) > 0$, then (a,b) is a local minimum of f.
  - If $D > 0$ and if $f_{xx}(a,b) < 0$, then (a,b) is a local maximum of f.
  - If $D < 0$, then (a,b) is a saddle point of f.
  - If $D = 0$, then no conclusion can be drawn in general.
- (**Lagrange Multipliers** for constraint optimization) Maximize $f\colon \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint $g = 0$. Solve for x, y, λ: $\nabla f = \lambda \nabla g$ (and $g = 0$).

Example. Minimize $f(x,y,z) = x^2 + y^2 + z^2$ subject to $g(x,y,z) = x + y + z - 3/2 = 0$. Solve $\nabla f(x,y,z) = \lambda \nabla g(x,y,z)$ and $g(x,y,z) = 0$. We have the system of four equations in four unknowns
$$\begin{cases} 2x = \lambda \\ 2y = \lambda \\ 2z = \lambda \\ x + y + z = 3/2. \end{cases}$$
The first three equations imply $x = y = z$. So, the fourth equation implies that $3x = 3/2$, so $x = 1/2 = y = z$. So $(x,y,z) = (1/2, 1/2, 1/2)$ is the only candidate critical point. This critical point must be the global minimum since f becomes arbitrarily large for the point $(x, -x, 3/2)$ as $x \to \infty$.

12. Appendix: Notation

Let x be a real number. Let $a > 0$

$\mathbb{R}$ denotes the set of real numbers

$\in$ means "is an element of." For example, $2 \in \mathbb{R}$ is read as "2 is an element of $\mathbb{R}$."

$f\colon A \to B$ means f is a function with domain A taking values in B. For example,

$f\colon [0,1] \to \mathbb{R}$ means that f is a function with domain $[0,1]$ taking values in $\mathbb{R}$

e^x denotes the exponential function

$\ln(x)$ denotes the natural logarithm of $x > 0$, i.e. the inverse of the exponenial function

$\log_a(x) = \ln(x)/\ln(a)$

$\cos^{-1}(x)$ denotes the inverse of $\cos\colon [0,\pi] \to [-1,1]$

Let $f\colon \mathbb{R} \to \mathbb{R}$, and let $a, b, x \in \mathbb{R}$ with $a < b$. Let n be a positive integer.

$\lim_{x\to a} f(x)$ denotes the limit of $f(x)$ as x approaches a

$f'(x) = \dfrac{df(x)}{dx} = \dfrac{d}{dx}f(x) = \lim_{h\to 0}\dfrac{f(x+h)-f(x)}{h}$ denotes the derivative of f with respect to x

$f''(x) = \dfrac{d}{dx}f'(x)$ denotes the second derivative of f with respect to x

$f^{(n)}(x)$ denotes the n^{th} derivative of f with respect to x

$\int_a^b f(x)\,dx$ denotes the definite integral of f on the interval $[a,b]$

$\int f(x)\,dx$ denotes the indefinite integral of f

Remark 12.1. Whenever a fraction has a radical of a number in the denominator, we prefer to move the radical to the numerator, as in the following example:

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

If a variable occurs in the radical, then we usually leave the radical in the denominator.

$$\mathbb{R}^2 = \{(x_1, x_2)\colon x_1 \in \mathbb{R} \text{ and } x_2 \in \mathbb{R}\}$$
$$\mathbb{R}^3 = \{(x_1, x_2, x_3)\colon x_1 \in \mathbb{R} \text{ and } x_2 \in \mathbb{R} \text{ and } x_3 \in \mathbb{R}\}$$

Let $v = (x_1, y_1, z_1)$ and let $w = (x_2, y_2, z_2)$ be vectors in Euclidean space $\mathbb{R}^3$.

$$v \cdot w = x_1x_2 + y_1y_2 + z_1z_2 \text{ , the dot product of } v \text{ and } w$$
$$\|v\| = \sqrt{x_1^2 + y_1^2 + z_1^2} \text{ , the length of the vector } v$$

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let $(a,b) \in \mathbb{R}^2$. Let $v \in \mathbb{R}^2$ be a vector.

$$\frac{\partial f}{\partial x}(a,b) = f_x(a,b), \text{ denotes the partial derivative of } f \text{ in the } x\text{-direction}$$
$$\frac{\partial f}{\partial y}(a,b) = f_y(a,b), \text{ denotes the partial derivative of } f \text{ in the } y\text{-direction}$$
$$\nabla f(a,b) = (f_x(a,b), f_y(a,b)), \text{ denotes the gradient vector of } f \text{ at } (a,b)$$
$$D_v f(a,b) = \nabla f(a,b) \cdot v, \text{ denotes the derivative of } f \text{ with respect to the direction } v$$
$$D(a,b) = \det \begin{pmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{pmatrix} = f_{xx}(a,b)f_{yy}(a,b) - (f_{xy}(a,b))^2$$
$$\text{denotes the discriminant, or Hessian, of } f \text{ at } (a,b)$$

Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be a real valued function.

$$\nabla f = \left(\frac{\partial}{\partial x} f, \frac{\partial}{\partial y} f, \frac{\partial}{\partial z} f \right)$$

USC Department of Mathematics, Los Angeles, CA 90089-2532
E-mail address: stevenmheilman@gmail.com