Math Prefresher Lecture Notes

Rochester Department of Political Science

August 19, 2014

Contents

List of Named Theorems

List of Examples

Chapter 1

Set Theory

1.1 Defining Sets

We begin by defining sets.

Definition 1.1. A *set* is a grouping of distinct objects considered as a whole; these objects are called the *elements* of the set. □

To denote that an object x is an element of the set A, we write $x \in A$. Conversely, to denote that it is not an element of A, we write $x \notin A$.

Example 1.1 (Set membership)

1. Let X be the set of items you can order at a typical coffee shop.
   - latte $\in X$
   - muffin $\in X$
   - pizza $\notin X$
2. Let $\mathbb{N}$ be the set of natural numbers: $1, 2, 3, 4, 5, \ldots$
   - $2 \in \mathbb{N}$.
   - $-13 \notin \mathbb{N}$.
   - $2.4 \notin \mathbb{N}$.

3. Let $\mathbb{R}$ be the set of real numbers: every number that can be placed on the "number line".
   - $2 \in \mathbb{R}$
   - $\sqrt{2} \in \mathbb{R}$
   - $-\sqrt{2} \in \mathbb{R}$
   - $\sqrt{-2} \notin \mathbb{R}$
   - pizza $\notin \mathbb{R}$

There are two ways to write sets, either (1) by listing the elements of the set between brackets, or (2) using set builder notation.

Example 1.2 (Set builder notation)

1. A is the set of all integers between 1 and 20.
   - $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$
   - $A = \{x \,|\, 1 \leq x \leq 20, x \in \mathbb{N}\}$
   - or more succinctly: $A = \{x \in \mathbb{N} \,|\, 1 \leq x \leq 20\}$
2. B is the set of all even natural numbers:
   - $B = \{2, 4, 6, 8, 10, \ldots\}$
   - $B = \{x \,|\, x = 2n, n \in \mathbb{N}\}$
   - or more succinctly: $B = \{2n \,|\, n \in \mathbb{N}\}$
3. C is the set of all real numbers
   - Not possible
   - $C = \{x \,|\, x \in \mathbb{R}\}$
4. D is the sequence of numbers described by n^2+1, s.t., n is a natural number: $2, 5, 10, 17, 26, \ldots$

- Not possible
- $D = \{n^2 + 1 \mid n \in \mathbb{N}\}$

5. $E = \{a, e, i, o, u\}$
6. $F = \{3, 3.1, 3.14, 3.141, 3.1415, \ldots\}$

1.2 Set Properties

Definition 1.2. The *cardinality* of a set S, written as $|S|$, is the number of elements in S. □

For example, $A = \{a, b, c, \ldots, x, y, z\}$ has 26 elements. The set of primary colors has three elements (red, yellow, blue).

Sets can have infinitely many elements, such as the set of integers $\mathbb{Z}$ or the set of real numbers $\mathbb{R}$. Sets can also have 0 cardinality. We call such a set the empty, or null, set, denoted $\emptyset$. For example, the set of all 2-sided triangles is an empty set.

We also need to talk about subsets.

Definition 1.3. A set S is a *subset* of T is every element of S is an element of T. □

We write this as $S \subseteq T$ and equivalently $T \supseteq S$. The symbol $\subseteq$ is an inclusion or containment. So in the first case, we'll say that S is included or contained in T. In the second case, we say that T is a superset of S.

Equality of sets is a closely related concept. The sets S and T are equal if they have the same members, in which case we write $S = T$. For the sets to be equal, all elements of S must be elements of T, and vice versa. Therefore, $S = T$ if and only if $S \subseteq T$ and $T \subseteq S$.

Example 1.3 (Subsets)

1. Suppose $T = \{1, 2, 3, 4, 5\}$

- $\{1\}$ is a subset of T.
- $\{1, 2, 3, 4, 5\}$ is a subset of T.
- $\{1, 2, 3, 7\}$ is *not* a subset of T.

2. Suppose $T = \{\frac{m}{n} \mid m, n \in \mathbb{N}\}$.
   - $\{\frac{3}{n} \mid n \in \mathbb{N}\} \subseteq T$.

You may have intuitively thought that S cannot be a subset of T if they are equal. This is true when we're talking about proper subsets.

Definition 1.4. A set S is a *proper subset* of a set T if $S \subseteq T$, and there is $t \in T$ such that $t \notin S$. We write this as $S \subset T$ or $S \subsetneq T$. □

The number of subsets of a set is $2^{|S|}$, where $|S|$ is the cardinality of the set. For example, what are the subsets of $S = \{1, 2, 3, 4\}$?

$$\begin{array}{l}
\emptyset \\
\{1\}, \{2\}, \{3\}, \{4\}, \\
\{1,2\}, \{1,3\}, \{1,4\}, \\
\{2,3\}, \{2,4\}, \{3,4\}, \\
\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \\
\{1,2,3,4\}
\end{array}$$

How would you prove that subsets are transitive—i.e., that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$?

Proof. Take any element in A, call it t. Then by definition, since $t \in A$ and $A \subseteq B$, we know that $t \in B$. Since $B \subseteq C$, we have that $t \in C$. Thus, if an element $t \in A$, it must be the case that $t \in C$. In other words, every element of A must also be an element of C. So $A \subseteq C$. ■

1.3 Set Operations

Definition 1.5. Let A and B be sets. The *intersection* of A and B, written $A \cap B$, is the set of all elements that are in both A and B. □

Definition 1.6. The sets A and B are said to be *disjoint* if they have no elements in common: $A \cap B = \emptyset$. $\square$

Properties of intersections:

1. $A \cap B = B \cap A$.

2. $A \cap B \subseteq A$.
   Proof: take any $x \in A \cap B$. Then it must be the case, by definition, that $x \in A$. Thus $A \cap B \subseteq A$.

3. $A \cap A = A$
   Proof: take any $x \in A \cap A$. Then $x \in A$ and $x \in A$. Thus, it must be the case that $A \cap A \subseteq A$. Now take any $x \in A$. Then $x \in A$ and $x \in A$. Thus $x \in A \cap A$. So $A \cap A = A$.

4. $A \cap \emptyset = \emptyset$. [You get to show this yourselves in the homework.]

Example 1.4 (Intersections)
Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{8, 9, 10\}$

1. $A \cap B = \{3\}$

2. $A \cap C = \emptyset$

3. $B \cap C = \emptyset$

Definition 1.7. Let A and B be two sets. Then the *union* of A and B, denoted $A \cup B$, is the set of all elements that are members of A, or members of B, or members of both. $\square$

Properties of unions:

1. $A \cup B = B \cup A$.

2. A is a subset of $A \cup B$. $A \subseteq A \cup B$.

3. $A \cup A = A$.

4. $A \cup \emptyset = A$.

Example 1.5 (Unions)
Same sets as in Example 1.4.

1. $A \cup B = \{1, 2, 3, 4, 5\}$
2. $A \cup C = \{1, 2, 3, 8, 9, 10\}$
3. $B \cup C = \{3, 4, 5, 8, 9, 10\}$
4. $\{1, 2\} \cup \{R, W\} = \{1, 2, R, W\}$
5. $\{1, 2\} \cup \{1, 2\} = \{1, 2\}$

Definition 1.8. Let A and B be sets. The *set difference* of A and B, written $A \setminus B$, is the set of elements in A that are not in B: $A \setminus B = \{a \in A \,|\, a \notin B\}$. This is also known as the relative complement of B in A. □

Example 1.6 (Set difference)
Same sets as in the last two examples.

1. $A \setminus B = \{1, 2\}$.
2. $A \setminus A = \emptyset$

Definition 1.9. If we define a universal set U and a set A such that $A \subseteq U$, then the absolute complement of A, written $\bar{A}$ or A^C, is $U \setminus A$. □

Properties of complements:

- $A \cup \bar{A} = U$
- $A \cap \bar{A} = \emptyset$

- $\overline{\overline{A}} = A$
- $A \setminus \bar{A} = A$
- $A \setminus B = A \cap \bar{B}$
  Proof: Take any $x \in A \setminus B$. Then, by definition, $x \in A$ but $x \notin B$. Then this implies that $x \in \bar{B}$. So $x \in A \cap \bar{B}$. Thus $x \in A$ and $x \in \bar{B}$. So $x \in A \cap \bar{B}$, and $A \setminus B \subseteq A \cap \bar{B}$. Now, take any $x \in A \cap \bar{B}$. Then $x \in A$ and $x \in \bar{B}$. So $x \in A$ and $x \notin B$. Then by definition $x \in A \setminus B$. Thus, $A \setminus B \supseteq A \cap \bar{B}$. Thus $A \setminus B = A \cap \bar{B}$

Example 1.7 (Complements)

1. Suppose that $U = \mathbb{R}$, and $A = \{x \,|\, 0 \leq x \leq 1\}$. Then $\bar{A} = \{x \,|\, x < 0 \text{ or } x > 1\}$.
2. Suppose that U is the set of all integers. Let $A = \{2k \,|\, k \in \mathbb{Z}\}$ and $B = \{2k+1 \,|\, k \in \mathbb{Z}\}$. Then $\bar{A} = B$ and $\bar{B} = A$.

1.4 Numbers

We've referred to the natural numbers, the integers, and the real numbers without being very rigorous about them. Let's look at them in a bit more detail now.

- **Natural Numbers**: the positive integers $1, 2, 3, 4, \ldots$. We write the set of natural numbers as $\mathbb{N}$.
- **Integers**: these are the positive natural numbers $(1, 2, 3, 4, \ldots)$, their negatives (additive inverses) $(-1, -2, -3, -4, \ldots)$ and 0. We write the set of integers as $\mathbb{Z}$.
- **Rational Numbers**: the set of rational numbers is the set of numbers that can be expressed as a fraction of two integers: an integer numerator and a non-zero integer denominator. Examples: $\frac{3}{4}$, $\frac{-7}{22}$, $\frac{21}{5}$, etc... We write in set notation that
$$\mathbb{Q} = \left\{ \frac{m}{n} \,\middle|\, m, n \in \mathbb{Z}; n \neq 0 \right\}.$$

- **Irrational Numbers**: the set of all numbers with non-terminating, non-repeating decimals. Note: these cannot be written as ratios of the integers.

  Let's show that $\sqrt{2}$ is irrational.

  Proof. Suppose not. Then $\sqrt{2}$ is rational, and we can write it as $\sqrt{2} = \frac{m}{n}$, where $\frac{m}{n}$ is some irreducible ratio—i.e., m and n have no factors in common—with $m, n \in \mathbb{Z}$, $n \neq 0$. This implies that $\frac{m^2}{n^2} = 2$ and thus $m^2 = 2n^2$. Since n^2 is an integer, it follows that m^2 is even. Furthermore, as even squares have even roots, this implies that m is even, and hence there exists $k \in \mathbb{Z}$ such that $m = 2k$. This gives

  $$n^2 = \frac{1}{2}m^2 = \frac{1}{2}(2k)^2 = 2k^2.$$

  Since k^2 is an integer, this implies that n is even. However, because m and n are both even, this means 2 is a common factor of m and n. This means $\frac{m}{n}$ is not irreducible, a contradiction. Therefore, $\sqrt{2}$ is irrational. ■

  Some other examples of irrational numbers (not going to prove them) are π and e.

- **Real Numbers**: The set of all rational and irrational numbers. We write this set as $\mathbb{R}$.

  Informally, we can think of the real numbers as labels for points along a horizontal line of infinite length: exact measures of the distance from a fixed point called zero and labeled the origin.

Note that if we go by our set descriptions earlier, there exists a specific relationship between $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$. It is $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. In addition, the set of irrational numbers is equivalent to $\mathbb{R} \setminus \mathbb{Q}$.

Now we'll move on to describing various properties of the real numbers and some common operations.

Field properties (you may or may not study fields in 404)

1. Commutation: $x + y = y + x$ and $xy = yx$
2. Association: $x + (y + z) = (x + y) + z$ and $x(yz) = (xy)z$

3. Distribution: $x(y + z) = xy + xz$
4. Existence of identity elements: there are two distinct numbers 0 and 1 that function as identities for $+$ and $\cdot$ respectively: $x+0 = x$ and $x \cdot 1 = x$.
5. Existence of inverse elements: the additive inverse of x is $-x$ and the multiplicative inverse of x is $\frac{1}{x}$.

   So for all x, we have that $x + (-x) = 0$ (the additive inverse) and, for $x \neq 0$, $x \cdot \frac{1}{x} = 1$ (the multiplicative inverse).

Order properties (you will definitely study orders in 407)

1. Trichotomy: if x and y are numbers, exactly one of the following holds: $x = y$, $x > y$, or $x < y$
2. Transitivity: $x < y$ and $y < z$ implies $x < z$. $x = y$ and $y = z$ implies $x = z$
3. Addition: $x < y$ implies $x + z < y + z$
4. Multiplication: For $z > 0$, $x < y$ implies $zx < zy$; for $z < 0$, $x < y$ implies $zx > zy$

Properties of absolute values

1. $|ab| = |a| \cdot |b|$
2. $\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$
3. $|a + b| \leq |a| + |b|$ (the oft-used triangle inequality)
4. $|a - b| \geq ||a| - |b||$

Properties of exponents First of all, what is an exponent? Consider an iterated multiplication: $m \times m \times m \times ... \times m$ (n times). Then we write m^n, where n is the exponent.

For all that follows, $a, b \in \mathbb{R}$:

1. $a^1 = a$
2. $a^{-1} = \frac{1}{a}$
3. $a^{-m} = \frac{1}{a^m}$
4. $a^{m+n} = (a^m)(a^n)$

5. $a^{m-n} = \frac{a^m}{a^n}$
6. $a^0 = 1$ (Proof: $a^0 = a^{2-2} = \frac{a^2}{a^2} = 1$)
7. $(a^m)^n = a^{mn}$
8. $a^{\frac{m}{n}} = (a^{\frac{1}{n}})^m = (a^m)^{\frac{1}{n}}$
9. $(ab)^m = a^m b^m$
10. $(\frac{a}{b})^m = \frac{a^m}{b^m}$

Properties of Logarithms We define logarithms in the following way: if $x = b^n$, then $\log_b(x) = n$. That is, the log of a given base for a number is the exponent to which you would have to raise the base so that it equaled the number.

1. $\log_b(xy) = \log_b(x) + \log_b(y)$
2. $\log_b(\frac{x}{y}) = \log_b(x) - \log_b(y)$
3. $\log_b(x^y) = y \log_b(x)$
4. $\log_b(\sqrt{x}) = \frac{1}{2} \log_b(x)$
5. $b^{\log_b(x)} = x$

Examples:

$$\begin{array}{cc}
9 = 3^2 & \log_3(9) = 2 \\
x^{\sqrt{2}} = \pi & \log_x(\pi) = \sqrt{2} \\
\log_{\sqrt{2}}(\pi) = x & \pi = \sqrt{2}^x \\
\ln(x) = e & x = e^e \\
\ln(\frac{1}{x}) - 2\ln(x) & \ln(\frac{1}{x} \cdot \frac{1}{x^2}) = \ln(\frac{1}{x^3})
\end{array}$$

Example: Solve y as a function of $x > 0$ and $C > 0$:

$$\begin{aligned}
\ln(y-5) &= ln(3x^2) + \ln(C) \\
e^{\ln(y-5)} &= e^{ln(3x^2)+\ln(C)} \\
y - 5 &= e^{ln(3x^2)} \cdot e^{\ln(C)} \\
y - 5 &= 3x^2 \cdot C \\
y &= 3Cx^2 + 5
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{2}\ln y &= \ln(4x^3) + C \\
\ln y &= 2\ln(4x^3) + 2C \\
e^{\ln y} &= e^{2\ln(4x^3)+2C} \\
y &= e^{\ln(4x^3)^2} \cdot e^{2C} \\
y &= (4x^3)^2 e^{2C} \\
y &= 16x^6 e^{2C}
\end{aligned}
$$

1.5 Real Analysis

Now let's consider some more abstract structure; this will hopefully be useful to have seen just a bit of when you hit it in 407.

Consider a set A of real numbers.

Definition 1.10. A set $A \subseteq \mathbb{R}$ is *bounded above* if there is a number $x \in \mathbb{R}$ such that $a \leq x$ for all $a \in A$. We call an element x for which this holds an *upper bound* of A. □

Definition 1.11. A set $A \subseteq \mathbb{R}$ is *bounded below* if there is a number $y \in \mathbb{R}$ such that $a \geq y$ for all $a \in A$. We call an element y for which this holds a *lower bound* of A. □

Example 1.8 (Upper and lower bounds)

1. Let $A = \{a \,|\, 2 \leq a \leq 5\}$.
    - A is bounded above. 6 is an upper bound of A, as is 5.5, as is 5.25, and so on.
    - A is bounded below. 1 is a lower bound of A, as is 1.5, as is 1.75, and so on.
2. Let B be the set of even natural numbers: 2, 4, 6, 8, ...
    - B is not bounded above. No number can be an upper bound for B.
    - B is bounded below. 0 is a lower bound of B, as is 1, as is 1.5, and so on.

Note there are often many possible upper and lower bounds. But we can narrow it down a bit by talking about the least upper bound and the greatest lower bound.

Definition 1.12. The least upper bound of a non-empty set A that is bounded above is called the *supremum* of A, denoted $\sup A$. □

Definition 1.13. The greatest lower bound of a non-empty set A that is bounded below is called the *infimum* of A, denoted $\inf A$. □

In the example above, $\sup A = 5$ and $\inf A = 2$.

We say that a set A has a a maximum when $\sup A \in A$, and we say it has a minimum when $\inf A \in A$.

Example 1.9 (Supremum, infimum, maximum, minimum)

1. $A = \{1, 2, 3, 4, 5\}$. $\sup A = 5$ and $\inf A = 1$. Similarly, the maximum and minimum of A are 5 and 1 respectively.

2. $A = \{x \,|\, 2 \leq x < 3\}$. $\sup A = 3$ and $\inf A = 2$. The set has no maximum, but its minimum is 2.

3. $A = \{x \,|\, x < 0\}$. $\sup A = 0$. A is not bounded below, so it does not have an infimum. The set has neither a maximum nor a minimum.

Assertion 1.1 (Completeness Axiom) *Every non-empty set of real numbers which is bounded above has a smallest upper bound. Every non-empty set of real numbers which is bounded below has a largest lower bound.*

All this talk of sets of real numbers leads us naturally to begin talking about intervals. We can talk about intervals formally in the following way.

Definition 1.14. A non-empty subset X of $\mathbb{R}$ is an *interval* if and only if for all $a, b \in X$ and $c \in \mathbb{R}$, $a \leq c \leq b$ implies that $c \in X$. □

So intuitively, an interval is a connected portion of the real line. If every number between any two members of the set also belongs to the set, then the set is an interval.

Bounded intervals come in the following flavors:

- $(a, b) \equiv \{x \mid a < x < b\}$. We call this interval open. $\inf A = a$ and $\sup A = b$, but it has no max or min.
- $[a, b] \equiv \{x \mid a \leq x \leq b\}$. We call this interval closed. $\inf A = a = \min A$ and $\sup A = b = \max A$.
- $(a, b] \equiv \{x \mid a < x \leq b\}$. This is half open. The set has no min, but $\inf A = a$ and $\sup A = b = \max A$.
- $[a, b) \equiv \{x \mid a \leq x < b\}$. This is also half open. The set has no max, but $\inf A = a = \min A$ and $\sup A = b$.

We also have the following unbounded intervals:

- $[a, \infty) = \{x \mid x \geq a\}$. $\inf A = a = \min A$, no max or sup.
- $(a, \infty) = \{x \mid x > a\}$. $\inf A = a$, no sup, min, or max.
- $(-\infty, b] = \{x \mid x \leq b\}$. No inf or min, $\sup A = b = \max A$.
- $(-\infty, b) = \{x \mid x < b\}$. No inf or min, $\sup A = b$, no max.

Finally, before we move to bigger and better things, we can formalize our notions of open and closed sets (note: moving beyond intervals) by considering just a little bit of real analysis.

Definition 1.15. Given a set $A \subseteq \mathbb{R}$, we say that $x \in \mathbb{R}$ is a *boundary point* of A if, for all $\epsilon > 0$,

1. $(x - \epsilon, x + \epsilon) \cap A \neq \emptyset$,
2. $(x - \epsilon, x + \epsilon) \cap \bar{A} \neq \emptyset$.

The set of all boundary points of A is written as $\operatorname{bd} A$. □

So what does this mean in words? x is a boundary point of set A if every $\epsilon > 0$ creates an interval around x that includes both points inside A, and outside A. Think about an open interval on the real line.

Example 1.10 (Boundary points)

1. $\operatorname{bd} \emptyset = \emptyset$

2. $\operatorname{bd}\mathbb{R} = \emptyset$

3. $\operatorname{bd}[a, b] = \operatorname{bd}(a, b) = \operatorname{bd}[a, b) = \operatorname{bd}(a, b] = \{a, b\}$

4. $\operatorname{bd}\{1, 2, 3, 4\} = \{1, 2, 3, 4\}$

5. $\operatorname{bd}\mathbb{N} = \mathbb{N}$

6. $\operatorname{bd}(\{1, 5\} \cup [2, 3]) = \{1, 5, 2, 3\}$

Using the concept of boundary points, we can define open and closed sets (not just for intervals) more generally.

Definition 1.16. A set A is *open* if it does not contain any of its boundary points; i.e., $A \cap \operatorname{bd} A = \emptyset$. □

Definition 1.17. A set A is *closed* if it contains all of its boundary points; i.e., $\operatorname{bd} A \subseteq A$. □

Important note: A set may be neither open nor closed! (In fact, "most" sets are neither.) So if you find that a set is not closed — e.g., it fails to contain one of its boundary points — that does not necessarily mean it is open.

Example 1.11 (Open and closed sets)

- $\{a, b, c, d\}$ is closed.
- $[a, b] \cup [c, d]$ is closed.
- $(a, b) \cup (c, d)$ is open.
- $[5, 6)$ and $(0, 1) \cup \{2\}$ are neither open nor closed.

There are even two special cases of sets that are both open and closed:

- The empty set is both open and closed.

  Proof. The empty set is open because $\emptyset \cap \operatorname{bd} \emptyset = \emptyset$. The empty set is closed because $\operatorname{bd} \emptyset = \emptyset \subseteq \emptyset$. ■

- $\mathbb{R}$ is both open and closed.

  Proof. If $\mathbb{R}$ is the universal domain, then $\bar{\mathbb{R}} = \emptyset$. It is never possible to pick a $\epsilon > 0$ so that $(x - \epsilon, x + \epsilon) \cap \bar{\mathbb{R}} \neq \emptyset$, meaning that $\mathbb{R}$ has no boundary points: $\operatorname{bd} \mathbb{R} = \emptyset$. Therefore, $\mathbb{R} \cap \operatorname{bd} \mathbb{R} = \mathbb{R} \cap \emptyset = \emptyset$, so $\mathbb{R}$ is open. At the same time, since $\operatorname{bd} \mathbb{R} = \emptyset$ and $\emptyset$ is a subset of every set, $\operatorname{bd} \mathbb{R} \subseteq \mathbb{R}$, so $\mathbb{R}$ is closed. ■

To finish off, we're going to talk about two more types of points related to sets: interior points and closure points.

Definition 1.18. Given a set A, we say that x is an interior point of A if there exists $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subseteq A$. The set of all interior points of A is written as $\operatorname{int} A$. □

Example 1.12 (Interior points)

1. $\operatorname{int} \emptyset = \emptyset$
2. $\operatorname{int} \mathbb{R} = \mathbb{R}$
3. $\operatorname{int}(a, b) = \operatorname{int}[a, b] = \operatorname{int}(a, b] = \operatorname{int}[a, b) = (a, b)$
4. $\operatorname{int}\{x_1, x_2, x_3, x_4, ..., x_n\} = \emptyset$
5. $\operatorname{int} \overline{\{x_1, x_2, x_3, x_4, ..., x_n\}} = \overline{\{x_1, x_2, x_3, x_4, ..., x_n\}}$

Definition 1.19. Given a set A, we say that x is a closure point of A if, for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap A \neq \emptyset$. The set of all closure points of A is written as $\operatorname{clos} A$. □

Example 1.13 (Closure points)

1. $\operatorname{clos} \emptyset = \emptyset$

2. $\operatorname{clos} \mathbb{R} = \mathbb{R}$

3. $\operatorname{clos}[a,b] = \operatorname{clos}(a,b) = \operatorname{clos}(a,b] = \operatorname{clos}[a,b) = [a,b]$

Now let's see an example of how we would go about using all of this stuff. We will prove that $\operatorname{clos} A = A \cup \operatorname{bd} A$.

Proof. Take any $x \in \operatorname{clos} A$. Then, by definition, for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap A \neq \emptyset$. Either $x \in A$ or $x \notin A$. If $x \in A$ then $x \in A \cup \operatorname{bd} A$, as needed. If $x \notin A$ then $x \in \bar{A}$. In addition, for all $\epsilon > 0, x \in (x - \epsilon, x + \epsilon)$, therefore $x \in (x - \epsilon, x + \epsilon) \cap \bar{A}$, and thefore $(x - \epsilon, x + \epsilon) \cap \bar{A} \neq \emptyset$. Since also, as noted, for all $\epsilon > 0, (x - \epsilon, x + \epsilon) \cap A \neq \emptyset$, $x \in \operatorname{bd} A$, and $x \in A \cup \operatorname{bd} A$. Implying $\operatorname{clos} A \subseteq A \cup \operatorname{bd} A$.

Now take $x \in A \cup \operatorname{bd} A$. Then $x \in A$ or $x \in \operatorname{bd} A$ or both. If $x \in A$, then for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap A \neq \emptyset$, since $x \in (x - \epsilon, x + \epsilon)$. Thus $x \in \operatorname{clos} A$. Now, suppose $x \in \operatorname{bd} A$. Then, by definition, for all $\epsilon > 0$, $(x - \epsilon, x + \epsilon) \cap A \neq \emptyset$. So $x \in \operatorname{clos} A$. Thus $A \cup \operatorname{bd} A \subseteq \operatorname{clos} A$. Since $\operatorname{clos} A \subseteq A \cup \operatorname{bd} A$ and $A \cup \operatorname{bd} A \subseteq \operatorname{clos} A$, we have $\operatorname{clos} A = A \cup \operatorname{bd} A$. ∎

Chapter 2

Sequences and Series

2.1 Sequences and Convergence

Informally, we can think of a sequence as simply an ordered set of real numbers.

Definition 2.1. A *sequence* is a rule which assigns to each natural number n a unique real number $x_n \in \mathbb{R}$. This is written as $\{x_n\}$. □

If you already know the definition of a function, you may see that this means that a sequence is a type of function. In particular, it is a function that maps from the natural numbers $\mathbb{N}$ into the real line $\mathbb{R}$.

Example 2.1 (Sequences)

1. $\{x_n\} = \{n\} = \{1, 2, 3, 4, \ldots\}$
2. $\{x_n\} = \{\frac{1}{2^n}\} = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$
3. $\{x_n\} = \{(-1)^{n+1}\} = \{1, -1, 1, -1, \ldots\}$
4. $\{x_n\} = \{3.1, 3.14, 3.141, 3.1415, \ldots\}$
5. $\{x_n\} = \{\frac{n^2+1}{n}\} = \{2, \frac{5}{2}, \frac{10}{3}, \frac{17}{4}, \ldots\}$
6. $\{x_n\} = \{2^{n(-1)^n}\} = \{\frac{1}{2}, 4, \frac{1}{8}, 16, \ldots\}$
7. $\{x_n\} = \{\frac{n+1}{n}\} = \{\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \ldots\}$

8. The Fibonacci sequence, defined as follows:

$$F_1 = 0$$
$$F_2 = 1$$
$$\{F_n\} = \{F_{n-2} + F_{n-1}\} = \{0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots\}$$

Definition 2.2. A sequence $\{x_n\}$ is *weakly increasing* if, for all $n \in \mathbb{N}$, $x_{n+1} \geq x_n$. The sequence is *strictly increasing* if, for all $n \in \mathbb{N}$, $x_{n+1} > x_n$.□

That is, the next term in the series is always at least as large as the term immediately preceding it. Or larger than the previous term in the series. In our examples on the previous page, which are strictly increasing, and weakly increasing? [Sequences 1 and 5 are strictly increasing; sequences 4 and 8 are weakly increasing.]

Definition 2.3. A sequence $\{x_n\}$ is *weakly decreasing* if, for all $n \in \mathbb{N}$, $x_{n+1} \leq x_n$. The sequence is *strictly decreasing* if for all $n \in \mathbb{N}$, $x_{n+1} < x_n$. □

From the examples above, sequences 2 and 7 are strictly decreasing. The remainder of these sequences (3 and 6) are neither increasing nor decreasing. Now, for some added structure, we can apply some terminology from earlier to these sequences.

Definition 2.4. A sequence $\{x_n\}$ is *bounded above* if the set $\{x_n \,|\, n \in \mathbb{N}\}$ is bounded above. That is, if there exists $k \in \mathbb{R}$ such that $x_n \leq k$ for all n, then $\{x_n\}$ is bounded above. Similarly, we say that a sequence is *bounded below* if the set $\{x_n \,|\, n \in \mathbb{N}\}$ is bounded below. □

Definition 2.5. Let $\{x_n\}$ be a sequence and x be a real number. We say that $\{x_n\}$ *converges* to x if, for each $\epsilon > 0$, there exists N such that $|x_n - x| \leq \epsilon$ for all $n \geq N$. In this case, we write $x_n \to x$ or $\lim_{n\to\infty} x_n = x$, and we call x the *limit* of the sequence. □

Definition 2.6. A sequence $\{x_n\}$ *diverges to infinity* if, for all $c \in \mathbb{R}$, there exists N such that $x_n \geq c$ for all $n \geq N$. □

Although it is often easy to tell intuitively what the limit of a sequence is, proving it formally may be a different matter.

Example 2.2 (Limit of a sequence)
Show that $\frac{n^2-1}{n^2+1} \to 1$.

Pick any $\epsilon > 0$. We want to find an N such that

$$\left|\frac{n^2-1}{n^2+1} - 1\right| \leq \epsilon$$

for all $n \geq N$.

By our rules of absolute value, we have

$$\begin{aligned} -\epsilon \leq \frac{n^2-1}{n^2+1} - 1 \leq \epsilon \\ 1 - \epsilon \leq \frac{n^2-1}{n^2+1} \leq 1 + \epsilon \end{aligned}$$

We want to find that n is greater than or equal to something, so we will focus on the left side of the inequality.

$$\begin{aligned} 1 - \epsilon &\leq \frac{n^2-1}{n^2+1} \\ (1-\epsilon)(n^2+1) &\leq n^2 - 1 \\ n^2 + 1 - \epsilon(n^2+1) &\leq n^2 - 1 \\ -\epsilon(n^2+1) &\leq -2 \\ \epsilon(n^2+1) &\geq 2 \\ n^2 + 1 &\geq \frac{2}{\epsilon} \\ n &\geq \sqrt{\frac{2}{\epsilon} - 1} \end{aligned}$$

By choosing any $N \geq \sqrt{\frac{2}{\epsilon} - 1}$, we have that $|\frac{n^2-1}{n^2+1} - 1| \leq \epsilon$ for all $n \geq N$. Therefore, $\frac{n^2-1}{n^2+1} \to 1$.

There are three types of sequences that we might see:

1. Sequences that converge to something (such as in examples 2, 4, and 7)
2. Sequences that increase or decrease without bound, hence diverging to ∞ (examples 1, 5, and 8)
3. Sequences that neither increase/decrease without bound nor converge (examples 3 and 6)

Properties of convergent sequences:
Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences so that $x_n \to x$ and $y_n \to y$.

1. $\alpha x_n + \beta y_n \to \alpha x + \beta y$.
2. $x_n y_n \to xy$.
3. If $y \neq 0$, $\frac{x_n}{y_n} \to \frac{x}{y}$.
4. If there is some M such that $x_n \geq y_n$ for all $n \geq M$, then $x \geq y$.
5. If $\{x_n\}$ is convergent, then it is bounded above and below.
6. If $\{x_n\}$ is increasing and bounded above, then $x_n \to \sup\{x_n \mid n \in \mathbb{N}\}$.
7. A sequence has at most one limit to which it converges.

Now we have a number of results that can help us out with determining if sequences converge.

Theorem 2.1 (Squeeze Theorem) *Suppose that $\{a_n\}$ and $\{c_n\}$ both converge to some value L and that $a_n \leq b_n \leq c_n$ for all $n \geq K$, a fixed integer. Then $\{b_n\}$ also converges to L.*

Example 2.3 (Squeeze Theorem)
Show that $\{x_n\} = \left\{\frac{\sin(n)}{n}\right\}$ converges to 0.

[Plot $\sin(n)$.] Note that $\sin(n)$ oscillates between -1 and 1. Therefore, for $n \geq 1$, we know that

$$\frac{-1}{n} \leq \frac{\sin(n)}{n} \leq \frac{1}{n}.$$

Since both $\frac{1}{n}$ and $\frac{-1}{n}$ converge to 0, so does $\frac{\sin(n)}{n}$.

Theorem 2.2 *If $|a_n| \to 0$, then $a_n \to 0$.*

Proof. Since $-|a_n| \leq a_n \leq |a_n|$, this result follows from the Squeeze Theorem. ■

Theorem 2.3 *If U is an upper bound for a weakly increasing sequence $\{a_n\}$ then the sequence converges to a limit $A \leq U$. Similarly, if L is a lower bound for a weakly decreasing sequence $\{b_n\}$, then the sequence $\{b_n\}$ converges to a limit $B \geq L$.*

Example 2.4 (Convergence of a decreasing sequence)
Show that the sequence $\{b_n\} = \frac{n^2}{2^n}$ converges.

Let's write out the first few terms of the sequence:

$$\frac{1}{2}, 1, \frac{9}{8}, 1, \frac{25}{32}, \frac{36}{64}, \frac{49}{128}, \frac{64}{256}, \frac{81}{512}, \ldots$$

So it appears as though the sequence is decreasing (after the first few terms), but we want to check to be sure. We want to see that

$$\begin{aligned}
\frac{n^2}{2^n} &> \frac{(n+1)^2}{2^{n+1}} \\
\frac{n^2}{1} &> \frac{(n+1)^2}{2} \\
2n^2 &> n^2 + 2n + 1 \\
n^2 - 2n &> 1 \\
n(n-2) &> 1
\end{aligned}$$

This holds for all $n \geq 3$. Since $n \geq 1$, the numerator is positive, and the denominator is positive, and so every term in the sequence is non-negative. The sequence thus is bounded below by 0. Thus, the conditions of Theorem 2.3 hold, and we know that the sequence converges.

2.2 Series

Series are closely related to sequences.

Definition 2.7. Given an infinite sequence of real numbers $\{x_n\}$, a *series* is the sum of the terms in the sequence:

$$\sum_{k=1}^{\infty} x_k = x_1 + x_2 + x_3 + \dots.$$

The *partial sum* S_n is the sum of the first n terms:

$$S_n = \sum_{k=1}^{n} x_k = x_1 + x_2 + x_3 + \dots + x_n.$$

□

We are able to talk about the convergence of a series in much the same way as we talk about the convergence of sequences. It is a similar idea: we want to know if S_n approaches some fixed value as $n \to \infty$. By characterizing the partial sum as a sequence we are able to do this. In particular, we call the sum of an infinite series the limit of the sequence of partial sums, $\{S_n\}$. If this limit exists, then we say that the sequence converges.

Definition 2.8. Let $\{x_n\}$ be an infinite sequence and let $\{S_n\}$ be the associated sequence of partial sums. The infinite series $\sum_{k=1}^{\infty} x_k$ *converges* and has sum S if $S_n \to S$. If $\{S_n\}$ does not converge, then the series *diverges* and has no sum. □

Example 2.5 (Convergence of a series)
Consider the series $\sum_{n=1}^{\infty} (\frac{1}{2})^n$, whose partial sum can be written

$$S_n = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots + \frac{1}{2^n}.$$

Multiplying each side by $\frac{1}{2}$, we have

$$\frac{1}{2} S_n = \frac{1}{2^2} + \frac{1}{2^3} + \dots + \frac{1}{2^{n+1}}.$$

Now subtract the two series to yield

$$\begin{aligned}\frac{1}{2}S_n = S_n - \frac{1}{2}S_n &= \left(\frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^n}\right) - \left(\frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^{n+1}}\right)\\ &= \frac{1}{2} + \left(\frac{1}{2^2} - \frac{1}{2^2}\right) + \ldots + \left(\frac{1}{2^n} - \frac{1}{2^n}\right) - \frac{1}{2^{n+1}}\\ &= \frac{1}{2} - \frac{1}{2^{n+1}}.\end{aligned}$$

This implies

$$S_n = 2\left(\frac{1}{2} - \frac{1}{2^{n+1}}\right) = 1 - \frac{1}{2^n}.$$

What happens as $n \to \infty$? We have $\frac{1}{2^n} \to 0$, so we end up with

$$\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n = \lim_{n\to\infty} S_n = 1.$$

Example 2.6 (Geometric series)
Consider the *geometric series*

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \ldots.$$

We want to show that this series converges if $|x| < 1$ but diverges if $|x| \geq 1$. We can use the same tricks as in the last example. Start with the partial sum:

$$\begin{aligned}
S_n &= \sum_{k=0}^{n} x^k = 1 + x + x^2 + x^3 + \ldots + x^n. \\
xS_n &= x\sum_{k=0}^{n} x^k = x + x^2 + x^3 + \ldots + x^n + x^{n+1}. \\
S_n - xS_n &= 1 + (x - x) + (x^2 - x^2) + \ldots + (x^n - x^n) - x^{n+1}. \\
(1 - x)S_n &= 1 - x^{n+1} \\
S_n &= \frac{1 - x^{n+1}}{1 - x}.
\end{aligned}$$

If $|x| < 1$, then $x^{n+1} \to 0$ as $n \to \infty$, which means $S_n \to \frac{1}{1-x}$ as $n \to \infty$. Therefore, the geometric series converges to $\frac{1}{1-x}$ when $|x| < 1$. However, if $|x| \geq 1$, then the sequence $\{x^{n+1}\}$ diverges, as does $\{S_n\}$, so the series diverges.[1]

The geometric series is just one of a few common types of series:

1. Geometric series: $\sum_{n=0}^{\infty} ax^n = \frac{a}{1-x}$ for $|x| < 1$.

2. Harmonic series: $\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$

   Do you think this series converges or diverges? Despite looking like one that might converge, this series actually diverges, which we'll show now. Consider the partial sum:

$$\begin{aligned}
S_n &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \ldots \\
&= 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \ldots \\
&\geq 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \ldots \\
&= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \ldots
\end{aligned}$$

So by choosing large enough n, we can introduce enough groups of value at least $\frac{1}{2}$ to make S_n as large as we please. Therefore, the sequence of partial sums $\{S_n\}$ diverges, as does the series.

3. Telescoping series: $\sum_{n=0}^{\infty}(b_n - b_{n+1})$, so that the partial sums can be reduced to a fixed number of terms.

4. Alternating series: $\sum_{n=0}^{\infty}(-1)^n a_n$, where $a_n \geq 0$ for all n, so that the signs of the terms alternate. A famous example is the alternating harmonic series, $\sum_{n=1}^{\infty} \frac{1}{n}(-1)^{n-1}$, which does converge.

So how can we tell if these things converge? Just as with sequences, there are a number of theorems that can help us out.

Theorem 2.4 (n'th term test) *If the series $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. Equivalently, if $\lim_{n\to\infty} a_n \neq 0$ or if $\lim_{n\to\infty} a_n$ does not exist, then the series diverges.*

Note that $\lim_{n\to\infty} a_n = 0$ is necessary, but not sufficient, for the series to converge. Recall the harmonic series.

Theorem 2.5 *If the series $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ both converge, and c is some constant, then:*

1. *$\sum_{k=1}^{\infty} ca_k$ and $\sum_{k=1}^{\infty}(a_k + b_k)$ also converge*

2. *$\sum_{k=1}^{\infty} ca_k = c\sum_{k=1}^{\infty} a_k$*

3. *$\sum_{k=1}^{\infty}(a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k$*

Theorem 2.6 *If $\sum_{k=1}^{\infty} a_k$ diverges and $c \neq 0$, then $\sum_{k=1}^{\infty} ca_k$ also diverges.*

To derive a couple of additional results, we need to refine our concept of convergence a bit.

Definition 2.9. A series $\sum_{n=1}^{\infty} a_n$ *converges absolutely* if the series $\sum_{n=1}^{\infty} |a_n|$ converges. A series that converges but is not absolutely convergent is *conditionally convergent.* □

For example, the alternating harmonic series $\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$ is conditionally convergent, because $\sum_{n=1}^{\infty} |(-1)^{n-1}\frac{1}{n}| = \sum_{n=1}^{\infty} \frac{1}{n}$ is the ordinary harmonic series, which diverges.

Theorem 2.7 *If the series $\sum_{n=1}^{\infty} a_n$ converges absolutely, then it converges.*

Theorem 2.8 *If the series $\sum_{n=1}^{\infty} a_n = A$ and converges absolutely, then any series formed from a rearrangement of its terms also converges to A.*

Theorem 2.9 (Riemann Series Theorem) *If the series $\sum_{n=1}^{\infty} a_n$ converges conditionally, then for all $A \in \mathbb{R}$ there exists a series formed from a rearrangement of its terms that converges to A. Moreover, there also exists a series formed from a rearrangement of its terms that diverges.*

Example 2.7 (Riemann Series Theorem)
The standard alternating harmonic series is

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots = \ln(2).$$

But by the Reimann Series Theorem, we can rearrange this conditionally convergent series to converge to any given value. One such rearrangement is

$$\left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) + \ldots,$$

which is equivalent to

$$\frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} \ldots\right),$$

half of the original sum. This is an illustration of a principle you will run into often in 407: weird things can happen when you're dealing with infinite amounts of numbers.

Chapter 3

Functions

Heading toward calculus, we're going to start talking about functions, their basic properties, and what we can do with them.

3.1 Definitions

Definition 3.1. Let S and T be sets. A *function* or *mapping* f from S to T, written $f : S \to T$, is a rule that assigns to each element $s \in S$ a unique element $t \in T$. □

Consider the following examples of functions:

1. Let S be the set of all people in the world and T be the set of all countries. Let f be the rule that assigns to every person his or her country of citizenship.

   Is this a function? No: some people have no citizenship, and some have dual citizenship.

2. $f : \mathbb{R} \to \mathbb{R}$, where $f(a) = a^2$, is a function.

3. $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(a, b) = a^2 - b^2$, is a function.

The intuition for our purposes is simple—a function is something that takes inputs from one set of elements and associates them with outputs in another set of elements. But now let's get a little more formal.

Definition 3.2. An *ordered pair* (x, y) is a set with two elements whose order is important and cannot be changed. □

With ordinary sets, $\{x, y\} = \{y, x\}$. However, with ordered pairs, $(x, y) \neq (y, x)$ in general.

Definition 3.3. The *Cartesian product* of two sets S and T, denoted $S \times T$, is the set of all ordered pairs (x, y) where $x \in S$ and $y \in T$. □

Definition 3.4. A *relation* B between two sets S and T is any subset of their Cartesian product, $B \subseteq S \times T$. If $(x, y) \in B$, we say that the relation B holds between x and y and write xBy or $x \to y$. □

Definition 3.5. A *function* from S to T is a relation B such that:

1. For all $x \in S$, there exists $y \in T$ such that xBy.
2. If xBy for some $y \in T$, then it is not the case that xBz for any $z \in T \setminus \{y\}$. □

For functions on the real numbers, these restrictions give us the familiar "vertical line rule": that each vertical line at a point in S would intersect with the graph of the function in one and only one place.

Example 3.1 (Relations and functions)
Let $A = \{1, 2, 3, 4\}$, $B = \{14, 7, 234\}$, and $C = \{a, b, c\}$. These have the following Cartesian products:

$$\begin{aligned} A \times B = \{&(1,14),(1,7),(1,234),(2,14),(2,7),(2,234),\\ &(3,14),(3,7),(3,234),(4,14),(4,7),(4,234)\}\\ A \times C = \{&(1,a),(1,b),(1,c),(2,a),(2,b),(2,c),\\ &(3,a),(3,b),(3,c),(4,a),(4,b),(4,c)\}\\ B \times C = \{&(14,a),(14,b),(14,c),(7,a),(7,b),(7,c),(234,a),(234,b),(234,c)\} \end{aligned}$$

Now consider the following relations. Which of them is a function?

1. Let e be the relation between A and B that associates $1 \to 234$, $2 \to 7$, $3 \to 14$, $4 \to 234$, and $2 \to 234$. This is a subset of the Cartesian product of A and B, and therefore is a relation. But it is not a function, because 2 is associated with both 7 and 234.

2. Let f be the association between A and C that associates $1 \to c$, $2 \to b$, $3 \to a$, $4 \to b$. Again, a relation. But this time, it is also a function.

3. Let g be the association between A and C that associates $1 \to a$, $2 \to a$, $3 \to a$. It is a relation. But it is not a function, as 4 has nothing associated with it.

3.2 Properties

We need to establish some additional definitions before we can start talking about graphing functions, roots of functions, and things of that nature.

In all of the following definitions, let A and B be sets and $f : A \to B$ be a function.

Definition 3.6. The set A is called the *domain* and B is called the *codomain*. □

Definition 3.7. Given a function $f : A \to B$ which assigns to each $a \in A$ a unique $b \in B$, the element b is called the *image of the element* a and written $b = f(a)$. □

Definition 3.8. Given a subset of the domain $S \subseteq A$, the *image of the subset* is

$$f(S) = \{b \in B \mid f(a) = b \text{ for some } a \in S\} = \{f(a) \mid a \in S\}.$$

The image of A, written $f(A)$, is also called the *range* of the function. Note that $f(A)$ need not equal the whole codomain B. □

Definition 3.9. Given a subset of the codomain $T \subseteq B$, the *preimage* or *inverse image* of T is

$$f^{-1}(T) = \{a \in A \mid f(a) = b \text{ for some } b \in T\}.$$ □

Example 3.2 (Images and preimages)

1. Draw a picture where $a \to 4$, $b \to 1$, $c \to 4$, $d \to 5$, and $e \to 2$.
   (a) What are the domain and range of f?
   (b) What is the image of $C = \{c, d, e\}$?
   (c) What is the preimage of $D = \{1, 4\}$?
   (d) What is the image of A?
   (e) What is the preimage of B?
2. Consider $f(x) = x^2$. (Graph.)
   (a) What are the domain ($\mathbb{R}$) and range ($\mathbb{R}_+$)?
   (b) What is the image of $[0, 2]$?
   (c) What is the preimage of $[1, 4]$?
3. Consider $f(x) = \frac{1}{x}$. (Graph.) What are the domain and range? (Both are $\mathbb{R} \setminus \{0\}$.)

We sometimes want to take the inverses of functions, but we need to establish some additional properties first to ensure that a function inverse is well-defined.

Definition 3.10. A function $f : A \to B$ is *one-to-one*, or *injective*, if for all $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ implies that $a_1 = a_2$. $\square$

Intuitively, this means that two input values cannot result in the same output value.

Definition 3.11. A function $f : A \to B$ is *onto*, or *surjective*, if for all $b \in B$, there exists $a \in A$ such that $f(a) = b$. $\square$

This means that every element in the codomain of f is associated by the function with something in the domain.

Definition 3.12. A function that is one-to-one and onto is called *bijective*. If and only if $f : A \to B$ is bijective, then there exists the *inverse function* $f^{-1} : B \to A$, such that $f^{-1}(b) = a$ if and only if $f(a) = b$. $\square$

Example 3.3 (One-to-one and onto functions)

1. $f(x) = 2x + 24$. [Graph.]

   This is one-to-one and onto. Note that, in fact, all linear functions are one-to-one and onto, and thus an inverse will always exist for a linear function.

2. $f(x) = \cos(x)$. [Graph.]

   This is periodic, so it is not one-to-one.

3. $f : \mathbb{N} \to \mathbb{N}$, $f(n) = 4n + 3$.

   This is not onto. There will exist natural numbers in $\mathbb{N}$ that cannot be written as $4n + 3$. (For example, 1 or 2 or 3.) It is one-to-one, since $f(n_1) = f(n_2)$ implies $4n_1 - 3 = 4n_2 - 3$, which means $n_1 = n_2$.

Operations on functions:

1. Addition: if $S \subseteq \mathbb{R}$ and f and g are two functions from S to $\mathbb{R}$, then we define the function $f + g$ to be the function from S to $\mathbb{R}$ that satisfies $(f + g)(x) = f(x) + g(x)$, $x \in S$.

2. Multiplication by a constant: if $\lambda \in \mathbb{R}$ and $f : S \to \mathbb{R}$, then $(\lambda f)(x) = \lambda f(x)$, $x \in S$.

3. Multiplication by a function: $(fg)(x) = f(x)g(x)$.

4. Division: $(f/g)(x) = f(x)/g(x)$ where $g(x) \neq 0$ for all $x \in S$.

5. Composition: Let S and T be subsets of $\mathbb{R}$ so that $g : S \to T$ and $f : T \to \mathbb{R}$. Then the composite function $(f \circ g) : S \to \mathbb{R}$ is $(f \circ g)(x) = f(g(x))$, $x \in S$.

Example 3.4 (Operations on functions)
Take $f : \mathbb{R} \to \mathbb{R}$, such that $f(x) = \frac{3}{1+x^2}$ and $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = x^3$.

- $(f + g)(x) = \frac{3}{1+x^2} + x^3$
- $(2f)(x) = 2\frac{3}{1+x^2} = \frac{6}{1+x^2}$
- $(fg)(x) = \frac{3x^3}{1+x^2}$
- $(f \circ g)(x) = f(g(x)) = \frac{3}{1+(g(x))^2} = \frac{3}{1+(x^3)^2} = \frac{3}{1+x^6}$
- $(g \circ f)(x) = g(f(x)) = (f(x))^3 = \left(\frac{3}{1+x^2}\right)^3 = \frac{27}{(1+x^2)^3}$

Now let's go through a couple of proofs that use these properties.

Assertion 3.1 *Let $f : A \to B$ and $g : B \to C$. Consider $g \circ f$. Prove that if f and g are one-to-one then $g \circ f$ is one-to-one.*

Proof. Recall that for a one-to-one function h, $h(x_1) = h(x_2)$ implies that $x_1 = x_2$. Take any $x_1, x_2 \in A$ such that $(g \circ f)(x_1) = (g \circ f)(x_2)$. Then, by definition, we have $g(f(x_1)) = g(f(x_2))$. Since g is one-to-one, it must be the case that $f(x_1) = f(x_2)$. Since f is one-to-one, $x_1 = x_2$. So $(g \circ f)(x_1) = (g \circ f)(x_2)$ implies that $x_1 = x_2$, and therefore $g \circ f$ is one-to-one. ■

Assertion 3.2 *Let $f : A \to B$ and $g : B \to C$. If $g \circ f$ is onto, then g is onto.*

Proof. Suppose that $g \circ f$ is onto. Then we know that for $g \circ f : A \to C$, for every $c \in C$, there exists $a \in A$ such that $(g \circ f)(a) = c$. Thus, for all $c \in C$, there exists $a \in A$ such that $g(f(a)) = c$. So for every $c \in C$, there exists $b = f(a)$ such that $g(b) = c$. Thus, g is onto. ■

Types of functions:

1. Linear: $f(x) = mx + b$, where m is the slope and b is the intercept. (Graph example.)

2. Monomial: $f(x) = cx^k$, c is the coefficient and k is the exponent. (Graph c, cx, cx^2.)

3. Polynomial: A sum of monomials of different degrees and with possibly different coefficients,

$$f(x) = \sum_{k=0}^{n} c_k x^k = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n.$$

The value of the largest exponent is referred to as the degree of the polynomial, provided the associated coefficient is not zero.

4. Rational functions: A ratio of two polynomials. For example,

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

5. Exponential functions: $f(x) = ak^x$, where a is the coefficient, k is the base, and x is the exponent. Often, the term exponential function will refer to a function of the form $f(x) = e^x$.

6. Trigonometric functions: These are functions of an angle. Thus, they tend to be used mostly in the natural sciences.

3.3 Limits

Calculus is in many ways the study of limits. The two major branches of calculus, differential calculus and integral calculus, both deal with this. So first we want an intuitive understanding of exactly what a limit is. We already had some acquaintance with the notion of limits when we were talking about infinite sequences and series. In those cases, we were concerned with the behavior of the sequence as the number of terms grew infinitely. Now we're going to see limits as the behavior of a function infinitesimally close to a particular point.

Definition 3.13 (intuitive). To say that $\lim_{x\to c} f(x) = L$ means that when x is near but not equal to c, then $f(x)$ is near L. □

We don't necessarily care what happens *at* c. The notion of a limit is concerned only with what happens *near* c.

What about when our function looks like this?

To deal with this type of case, we need some additional concepts.

Definition 3.14 (Right- and left-hand limits, intuitive). To say that $\lim_{x\to c^+} f(x) = L$ means that when x is near but greater than c, then $f(x)$ is near L. Similarly, to say that $\lim_{x\to c^-} f(x) = L$ means that when x is near but less than c, then $f(x)$ is near L. □

Now that we've explained limits intuitively, it's time to be more formal.

Definition 3.15. To say that the *limit* of f at c is L, written $\lim_{x\to c} f(x) = L$, means that for each $\epsilon > 0$, there exists a corresponding $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all x with $0 < |x - c| < \delta$. □

So what this means is: pick any ϵ. Then we can find a range of width δ around x so that the difference between $f(x)$ and L is always less than ϵ within this range. [Use the picture of the function with the point discontinuity.]

Example 3.5 (Formal proofs of limits)

1. Consider $\lim_{x\to 3} 2x + 1$. Take $\epsilon = 0.01$. We guess that the limit equals 7. Can we find a δ corresponding to $\epsilon = 0.01$? That is, can we find a δ such that $|(2x+1)-7| < 0.01$ whenever $0 < |x-3| < \delta$?

$$\begin{aligned} |(2x+1) - 7| < 0.01 &\Leftrightarrow |2x - 6| < 0.01 \\ &\Leftrightarrow 2|x - 3| < 0.01 \\ &\Leftrightarrow |x - 3| < 0.005. \end{aligned}$$

   So $\delta = 0.005$ will work. That is, we can make $2x + 1$ within 0.01 of 7 given that x is within $\frac{0.01}{2}$ of 3.

2. Prove that $\lim_{x\to 2} \frac{2x^2-3x-2}{x-2} = 5$.

We are looking for δ such that

$$0 < |x-2| < \delta \Rightarrow \left|\frac{2x^2-3x-2}{x-2} - 5\right| < \epsilon$$

$$\left|\frac{2x^2-3x-2}{x-2} - 5\right| < \epsilon \Leftrightarrow \left|\frac{(2x+1)(x-2)}{x-2} - 5\right| < \epsilon$$

$$|(2x+1)-5| < \epsilon \Leftrightarrow |2x-4| < \epsilon$$

$$2|x-2| < \epsilon \Leftrightarrow |x-2| < \frac{\epsilon}{2}$$

So it seems that $\delta = \frac{\epsilon}{2}$ works. To write this formally, we would say:

Proof. Take any $\epsilon > 0$ and choose $\delta = \frac{\epsilon}{2}$. Then $0 < |x-2| < \delta$ implies

$$\left|\tfrac{2x^2-3x-2}{x-2} - 5\right| = \left|\tfrac{(2x+1)(x-2)}{x-2} - 5\right| = |2x+1-5| = 2|x-2| < 2\delta = \epsilon.$$

■

3. Prove that $\lim_{x\to c}(mx+b) = mc+b$.

   We are looking for δ such that $0 < |x-c| < \delta \Rightarrow |(mx+b)-(mc+b)| < \epsilon$.

   First play around with what is inside the absolute values: $|mx+b-mc-b| = |mx-mc| = |m(x-c)| = |m|\cdot|x-c| < \epsilon$. Therefore, we can use $|x-c| < \frac{\epsilon}{|m|}$. So $\delta = \frac{\epsilon}{|m|}$ works as long as $m \neq 0$.

   Proof. Let $\epsilon > 0$ be given, and let $\delta = \frac{\epsilon}{|m|}$. Then $0 < |x-c| < \delta$ implies that $|(mx+b)-(mc+b)| = |mx-mc| = |m|\cdot|x-c| < |m|\delta = \epsilon$. ■

Using this formal conception of a limit, we can restate our intuitive definition of right- and left-hand limits as follows.

Definition 3.16. To say $\lim_{x\to c^+} f(x) = L$ means that for each $\epsilon > 0$, there exists a corresponding $\delta > 0$ such that $|f(x)-L| < \epsilon$ for all x with $0 <$

$x - c < \delta$. To say $\lim_{x \to c-} f(x) = L$ means that for each $\epsilon > 0$, there exists a corresponding $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all x with $0 < -(x-c) < \delta$.□

The following theorem formalizes this intuition:

Theorem 3.1 $\lim_{x \to c} f(x) = L$ *if and only if* $\lim_{x \to c-} f(x) = L$ *and* $\lim_{x \to c+} f(x) = L$.

Properties of limits: Let n be a positive integer, k be a constant, and f and g be functions that have limits at c.

1. $\lim\limits_{x \to c} k = k$
2. $\lim\limits_{x \to c} x = c$
3. $\lim\limits_{x \to c} kf(x) = k \lim\limits_{x \to c} f(x)$
4. $\lim\limits_{x \to c}(f(x) + g(x)) = \lim\limits_{x \to c} f(x) + \lim\limits_{x \to c} g(x)$
5. $\lim\limits_{x \to c}(f(x) - g(x)) = \lim\limits_{x \to c} f(x) - \lim\limits_{x \to c} g(x)$
6. $\lim\limits_{x \to c}(f(x) \cdot g(x)) = \lim\limits_{x \to c} f(x) \cdot \lim\limits_{x \to c} g(x)$
7. $\lim\limits_{x \to c} \left(\dfrac{f(x)}{g(x)} \right) = \dfrac{\lim\limits_{x \to c} f(x)}{\lim\limits_{x \to c} g(x)}$, provided that $\lim\limits_{x \to c} g(x) \neq 0$.
8. $\lim\limits_{x \to c}(f(x)^n) = (\lim\limits_{x \to c} f(x))^n$
9. $\lim\limits_{x \to c} \sqrt[n]{f(x)} = \sqrt[n]{\lim\limits_{x \to c} f(x)}$, provided that the limit is positive when n is even.

We can use these to make our lives easier when dealing with limits. In many simple cases, these properties let us make statements about limits without going through lengthy proofs of the type we saw earlier.

Example 3.6 (Properties of limits)

1. Find $\lim_{x\to 3} 2x^4$. By statement 3,

$$\lim_{x\to 3} 2x^4 = 2\lim_{x\to 3} x^4$$

By statement 8,

$$2\lim_{x\to 3} x^4 = 2(\lim_{x\to 3} x)^4$$

By statement 2,

$$2(\lim_{x\to 3} x)^4 = 2(3)^4 = 162$$

2. Find $\lim_{x\to 4}(3x^2 - 2x)$. By statement 5,

$$\lim_{x\to 4}(3x^2 - 2x) = \lim_{x\to 4} 3x^2 - \lim_{x\to 4} 2x$$

By statements 2 and 3,

$$\lim_{x\to 4} 3x^2 - \lim_{x\to 4} 2x = 3\lim_{x\to 4} x^2 - 8$$

By statement 8,

$$3\lim_{x\to 4} x^2 - 8 = 3(\lim_{x\to 4} x)^2 - 8$$

By statement 2,

$$3(\lim_{x\to 4} x)^2 - 8 = 3\cdot 16 - 8 = 40$$

We can do some of this even more directly by application of the next theorem.

Theorem 3.2 *If f is a polynomial function or rational function, then $\lim_{x\to c} f(x) = f(c)$, provided that $f(c)$ is defined, and, in the case of a rational function, the denominator does not equal zero at c.*

Example 3.7 (Limits of polynomial and rational functions)

1. We have

$$\lim_{x\to 2}\frac{7x^5-10x^4-13x+6}{3x^2-6x-8}=\frac{7(2)^5-10(2)^4-13(2)+6}{3(2)^2-6(2)-8}=\frac{-11}{2}.$$

2. What is $\lim_{x\to 1}\frac{x^3+3x+7}{x^2-2x+1}$?

   The denominator is 0, so we cannot apply the theorem here. We'll have to use more advanced methods, which we'll see later on in the prefresher.

3. Find $\lim_{t\to 2}\frac{t^2+3t-10}{t^2+t-6}$.

   Again, if we just substitute in, we see that the denominator is zero — but so is the numerator. When that happens, it is wise to try some algebraic simplification (factoring, long division) and then try to apply the theorem again.

$$\lim_{t\to 2}\frac{t^2+3t-10}{t^2+t-6}=\lim_{t\to 2}\frac{(t+5)(t-2)}{(t+3)(t-2)}=\lim_{t\to 2}\frac{t+5}{t+3}=\frac{7}{5}$$

From our discussion of limits of sequences, you may recall the result we called the "Squeeze Theorem." As it happens, a very similar result also holds for limits of functions.

Theorem 3.3 (Squeeze Theorem) *Let f, g, and h be functions and let c be a constant. Suppose there exists $\nu > 0$ such that $f(x) \le g(x) \le h(x)$ for all x such that $0 < |x-c| < \nu$. [In other words, $f(x) \le g(x) \le h(x)$ for all x near c.] If $\lim_{x\to c} f(x) = \lim_{x\to c} h(x) = L$, then $\lim_{x\to c} g(x) = L$.*

Proof. Take any $\epsilon > 0$. By definition of limit, there exists δ_1 such that $L-\epsilon < f(x) < L+\epsilon$ for all x with $0 < |x-c| < \delta_1$. Similarly, there exists δ_2 such that $L-\epsilon < h(x) < L+\epsilon$ for all x with $0 < |x-c| < \delta_2$. Now let $\delta = \min\{\delta_1, \delta_2, \nu\}$. We have that for all x such that $0 < |x-c| < \delta$,

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

We then conclude from the definition of a limit that $\lim_{x\to c} g(x) = L$. ■

We can also talk about the limiting behavior of a function when its argument approaches infinity.

Definition 3.17. Consider a function f defined on $[c, \infty)$ for some number c. We say that $\lim_{x\to\infty} f(x) = L$ if for each $\epsilon > 0$, there exists a corresponding m such that $|f(x) - L| < \epsilon$ for all $x > m$. □

Definition 3.18. Consider a function f defined on $(-\infty, c]$ for some number c. We say that $\lim_{x\to-\infty} f(x) = L$ if for all $\epsilon > 0$, there exists a corresponding m such that $|f(x) - L| < \epsilon$ for all $x < m$. □

Example 3.8 (Limits as $x \to \infty$)

1. Show that if k is a positive integer, then $\lim_{x\to\infty} \frac{1}{x^k} = 0$.

   As before, we require some preliminary work. We want to find an $m > 0$ such that $x > m$ implies $|\frac{1}{x^k} - 0| < \epsilon$. Let $m = \sqrt[k]{1/\epsilon}$ and suppose $x > m$. We have

   $$\frac{1}{x^k} < \frac{1}{m^k} = \epsilon,$$

   as required.

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

   Instead of bothering with ϵ and δ, this time let's just use some of our helpful tricks from earlier. Here, we end up with $\frac{\infty}{\infty}$ if we just substitute in for x, and thus this is ambiguous. So let's use another trick, and divide everything by the highest term:

   $$\lim_{x\to\infty} \left(\frac{x}{1+x^2} \cdot \frac{1/x^2}{1/x^2} \right) = \lim_{x\to\infty} \frac{1/x}{1/x^2+1} = \frac{0}{0+1} = 0.$$

We'll wrap up with some definitions and examples for infinite-valued limits.

Definition 3.19. We say that $\lim_{x\to c} f(x) = \infty$ if for each number M, there exists $\delta > 0$ such that $f(x) > M$ for all x with $0 < |x-c| < \delta$. The definitions for infinite-valued left- and right-hand limits, as well as limits valued $-\infty$, are analogous. $\square$

This will allow us to do a bit better than simply saying the limit does not exist in situations where we cannot do away with a denominator of 0.

Example 3.9 (Infinite-valued limits)

1. Find $\lim_{x\to 1} \frac{1}{(x-1)^2}$.

   Intuitively, it should be clear that the limit is infinity. To prove this, take any real number $m > 0$. We want to show that there exists δ such that $\frac{1}{(x-1)^2} > m$ for all x such that $0 < |x-1| < \delta$. In particular, choose any $\delta \le \sqrt{\frac{1}{m}}$. Then $|x-1| < \delta$ implies

   $$(x-1)^2 < \delta^2 \le \frac{1}{m},$$

   and thus $\frac{1}{(x-1)^2} > m$, as required.

3.4 Continuity

Our last building block before moving on to calculus is continuity of functions. You may recall the intuitive definition of continuity from earlier algebra or calculus classes: a real-valued function is continuous if you can trace the graph of the function without lifting your pen from your paper. This notion is related to limits, which we use in the formal definition of continuity.

Definition 3.20. Let f be defined on an open interval containing c. We say that the function f is *continuous at the point* c if $\lim_{x\to c} f(x) = f(c)$. $\square$

Thus, this definition really requires three things:

1. $\lim_{x\to c} f(x)$ exists

2. $f(c)$ exists (c is in the domain of f)
3. $\lim_{x\to c} f(x) = f(c)$

Note that if any of these properties are violated, then it cannot be the case that f is continuous at c.

Example 3.10 (Continuity of a function)
Let $f(x) = \frac{x^2-4}{x-2}$ for all $x \neq 2$. How might we define $f(x)$ at $x = 2$ to make this a continuous function?

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

So if we define $f(x) = 4$ when $x = 2$, $f(x)$ will be continuous at 2.

Now let's go through some results that will be useful in applications.

Theorem 3.4 *A polynomial function is continuous at every real number. A rational function is continuous at every real number for which it is defined; that is, a rational function is continuous at every real number where the denominator is not zero.*

Theorem 3.5 *The absolute value function is continuous at every real number. If n is odd, then the nth root function is continuous at every real number. If n is even, then the nth root function is continuous for all positive real numbers.*

How would we prove that the absolute value function $f(x) = |x|$ is continuous at 0? Note that for $x < 0$, $f(x) = -x$, and for $x > 0$, $f(x) = x$. These are both polynomials, so by the previous theorem, we know that

$$\lim_{x\to 0-} f(x) = \lim_{x\to 0-} -x = 0 = \lim_{x\to 0+} x = \lim_{x\to 0+} f(x),$$

which implies $\lim_{x\to 0} f(x) = 0$. Since $f(0) = |0| = 0$, this proves continuity of f at 0. In fact, since f is a polynomial everywhere else, this proves that f is continuous everywhere.

Theorem 3.6 *Let f and g be functions defined on the real numbers, c and k be real numbers, and n be a positive integer. If f and g are continuous at c, then so are kf, $f+g$, $f-g$, fg, $\frac{f}{g}$ (provided that $g(c) \neq 0$), f^n and $\sqrt[n]{f}$ (provided that $f(c) > 0$ if n is even).*

Theorem 3.7 (Composite Limit Theorem) *If $\lim_{x\to c} g(x) = L$ and f is continuous at L, then $\lim_{x\to c} f(g(x)) = f(\lim_{x\to c} g(x)) = f(L)$. Therefore, if g is continuous at a point c and f is continuous at the point $g(c)$, then the composition $f \circ g$ is continuous at c.*

Example 3.11 (Composite Limit Theorem)

1. Show that $h(x) = |x^2 - 3x + 6|$ is continuous at all $c \in \mathbb{R}$. Consider this as a composite: $f(x) = |x|$, $g(x) = x^2 - 3x + 6$. Note that $g(x)$ is a polynomial, and so, by the results above, continuous everywhere. Note that $f(x) = |x|$ is continuous everywhere, as we just saw. Then, also by the previous result, $(f \circ g)(x) = f(g(x)) = |x^2 - 3x + 6|$ is continuous.

We also have the notion of continuity on an interval: this we can divide into speaking about continuity on an open interval and continuity on a closed interval. In the case of an open interval, this means what we would intuitively expect it to mean: at each point in the interval, the function is continuous.

What about the case of a closed interval? There is a problem here ... It could be the case that f is not defined at all to the left or right of the endpoints of an interval. For example, $f(x) = \sqrt{x}$ is not defined (on the real line, at least) for anything less than 0. This violates one of our properties, since $\lim_{x\to 0} f(x)$ will not exist in this case.

Definition 3.21. The function $f(x)$ is *right-continuous at a* if $\lim_{x\to a+} f(x) = f(a)$ and *left-continuous at b* if $\lim_{x\to b-} f(x) = f(b)$. We say that f is continuous on the open interval (a, b) if it is continuous at each point of that interval. It is continuous on the closed interval $[a, b]$ if it is continuous on (a, b), right-continuous at a and left-continuous at b. □

Example 3.12 (Continuity on a closed interval)
What is the largest interval on which the function $h(x) = \sqrt{4 - x^2}$ is continuous? The domain on which the function is well-defined is $[-2, 2]$. Let $f(x) = \sqrt{x}$ and $g(x) = 4 - x^2$, so that $h = f \circ g$. As a polynomial, g is continuous everywhere; f is continuous at all x such that $f(x) > 0$. We thus have that h is continuous on $(-2, 2)$, and now we just need to consider whether it is continuous at the endpoints. Using the Composite Limit Theorem, we have

$$\lim_{x\to 2-} \sqrt{4 - x^2} = \sqrt{4 - \lim_{x\to 2-} x^2} = \sqrt{4 - 4} = 0 = h(2)$$
$$\lim_{x\to -2+} \sqrt{4 - x^2} = \sqrt{4 - \lim_{x\to -2+} x^2} = \sqrt{4 - 4} = 0 = h(-2)$$

So $h(x)$ is right-continuous at -2 and left-continuous at 2. Thus, $h(x)$ is continuous on its entire domain $[-2, 2]$.

Chapter 4

Univariate Calculus

4.1 Differential Calculus

4.1.1 Definitions

This is one of the two primary areas within calculus. This has applications everywhere, and is ubiquitous in the social sciences.

Definition 4.1. The *derivative* of a function f is another function f', whose value at any number c is

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

provided that this limit exists and is not ∞ or $-\infty$. $\square$

In the intuitive sense, as was alluded to above, the derivative of a function f at c is simply the rate of change of the function at c. How much $f(x)$ changes with a change in x at c. For a (straight) line, the derivative of the line at a particular point x is simply the slope of the line. For a curve, the derivative is simply the slope of the line tangent to the curve at x.

A note on notation: the derivative $f'(x)$ can also be written $\frac{df}{dx}$ (sometimes $\frac{dy}{dx}$ when the function is written as $y = f(x)$) and as $D_x f(x)$.

Example 4.1 (Derivative, first definition)

1. Let $f(x) = 13x - 6$. Find $f'(4)$.

$$\begin{aligned} f'(4) &= \lim_{h\to 0} \frac{f(4+h) - f(4)}{h} = \lim_{h\to 0} \frac{13(4+h) - 6 - (13(4) - 6)}{h} \\ &= \lim_{h\to 0} \frac{52 + 13h - 6 - 52 + 6}{h} = \lim_{h\to 0} \frac{13h}{h} = 13. \end{aligned}$$

2. Let $f(x) = x^3 + 7x$, and let c be given. Find $f'(c)$.

$$\begin{aligned} f'(c) &= \lim_{h\to 0} \frac{f(c+h) - f(h)}{h} \\ &= \lim_{h\to 0} \frac{(c+h)^3 + 7(c+h) - c^3 - 7c}{h} \\ &= \lim_{h\to 0} \frac{(c+h)(c+h)(c+h) + 7(c+h) - c^3 - 7c}{h} \\ &= \lim_{h\to 0} \frac{c^3 + 3c^2h + 3ch^2 + h^3 + 7h - c^3}{h} \\ &= \lim_{h\to 0} \frac{3c^2h + 3ch^2 + h^3 + 7h}{h} \\ &= \lim_{h\to 0} 3c^2 + 3ch + h^2 + 7 \\ &= 3c^2 + 7. \end{aligned}$$

3. Let $f(x) = \frac{1}{x}$, and let $c \neq 0$ be given. Find $f'(x)$.

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

Note: we derive the form of the derivative from the following. This implies that we can write down an equivalent definition of the derivative,

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

The two definitions are, of course, identical in their results: you can use whichever of the two forms you prefer.

Example 4.2 (Derivative, second definition)
Find $g'(c)$ where $g(x) = \frac{2}{x+3}$.

$$\begin{aligned}
g'(c) &= \lim_{x\to c} \frac{g(x)-g(c)}{x-c} \\
&= \lim_{x\to c} \frac{\frac{2}{x+3} - \frac{2}{c+3}}{x-c} \\
&= \lim_{x\to c} \frac{1}{x-c}\left(\frac{2(c+3)}{(c+3)(x+3)} - \frac{2(x+3)}{(c+3)(x+3)}\right) \\
&= \lim_{x\to c} \frac{1}{x-c}\left(\frac{2(c+3)-2(x+3)}{(c+3)(x+3)}\right) \\
&= \lim_{x\to c} \frac{1}{x-c}\left(\frac{-2(x-c)}{(c+3)(x+3)}\right) \\
&= \lim_{x\to c} \frac{-2}{(c+3)(x+3)} = \frac{-2}{(c+3)^2}
\end{aligned}$$

So now we have established what the derivative is and what it means to be differentiable at a point. The derivative is also closely related to the continuity of functions.

Theorem 4.1 *If $f'(c)$ exists, then f is continuous at c.*

Proof. We need to show that $\lim_{x\to c} f(x) = f(c)$. We can write the following:

$$f(x) = f(c) + f(x) - f(c) = f(c) + \frac{f(x)-f(c)}{x-c}(x-c),$$

where $x \neq c$. Thus, taking the limit of this expression, we see that

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

By the rules of limits previously described, we can write this as:

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

So $\lim_{x\to c} f(x) = f(c)$. Therefore, $f(x)$ is continuous at c. ■

Note that the converse of this statement is not generally true; if a function is continuous at a point, the derivative does not have to exist there. To see this, consider the function $f(x) = |x|$ at the point $c = 0$. Let's ask if the derivative exists at the point 0. For any h, we have

$$\frac{f(0+h) - f(0)}{h} = \frac{|0+h| + |0|}{h} = \frac{|h|}{h}.$$

This implies

$$\lim_{h \to 0+} \frac{|h|}{h} = 1 \neq \lim_{h \to 0-} \frac{|h|}{h} = -1.$$

Since the left- limit and right-hand limits are not the same, the limit and hence the derivative do not exist at $x = 0$. In general, the function is not going to be differentiable at missing points, corners or kinks, and vertical asymptotes.

4.1.2 Rules of Differentiation

So this process works for us to set up derivatives and evaluate functions for the value of a derivative at a particular point ... but it is tedious and annoying. Let's come up with some helpful tricks.

Theorem 4.2 (Constant Function Rule) *If $f(x) = k$, where k is a constant, then for all x, $f'(x) = 0$.*

Proof. Note that $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h \to 0} \frac{k-k}{h} = \lim_{h \to 0} \frac{0}{h} = 0$. ■

Theorem 4.3 *If $f(x) = x$, then $f'(x) = 1$.*

Proof. Note that $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h \to 0} \frac{x+h-x}{h} = \lim_{h \to 0} \frac{h}{h} = 1$. ■

Theorem 4.4 (Power Rule) *If $f(x) = x^n$, where n is a positive integer, then $f'(x) = nx^{n-1}$.*

Proof. We have

$$\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{(x+h)^n - x^n}{h} \\
&= \lim_{h\to 0}\frac{x^n + nx^{n-1}h + n(n-1)\frac{1}{2}x^{n-2}h^2 + \ldots + nxh^{n-1} - x^n}{h} \\
&= \lim_{h\to 0}\frac{h}{h}\left(nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \ldots + nxh^{n-2} + h^{n-1}\right) \\
&= nx^{n-1},
\end{aligned}$$

as claimed. ■

The Power Rule actually works for any real number n when x is positive; the proof is just more involved in that case.

Theorem 4.5 (Constant Multiple Rule) *If k is a constant and f is a differentiable function, then $(kf)'(x) = k \cdot f'(x)$.*

Proof. Let $g(x) = k \cdot f(x)$. Then $g'(x) = \lim_{h\to 0}\frac{g(x+h)-g(x)}{h} = \lim_{h\to 0}\frac{kf(x+h)-kf(x)}{h} = k \cdot \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = k \cdot f'(x)$. ■

Theorem 4.6 (Sum Rule) *If f and g are differentiable functions, then $(f+g)'(x) = f'(x) + g'(x)$.*

Proof. Let $F(x) = f(x) + g(x)$. We then have

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

as claimed. ■

Theorem 4.7 (Difference Rule) *If f and g are differentiable functions, then $(f-g)'(x) = f'(x) - g'(x)$.*

Proof. Let $F(x) = f(x) - g(x) = f(x) + (-1)g(x)$. Then the result is immediate from the Constant Multiple Rule and the Sum Rule. ∎

Theorem 4.8 (Product Rule) *If f and g are differentiable functions, then $(fg)'(x) = f(x)g'(x) + g(x)f'(x)$.*

Proof. Let $F(x) = f(x)g(x)$. We have

$$\begin{aligned}
F'(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)[g(x+h) - g(x)] + g(x)[f(x+h) - f(x)]}{h} \\
&= \lim_{h\to 0} f(x+h) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} + g(x) \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= f(x)g'(x) + g(x)f'(x),
\end{aligned}$$

as claimed. ∎

Theorem 4.9 *Let f and g be differentiable functions with $g(x) \neq 0$. Then*

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

Proof. Proof: Let $F(x) = \frac{f(x)}{g(x)}$. We have

$$\begin{aligned}
F'(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\
&= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - g(x+h)f(x)}{g(x)g(x+h)} \frac{1}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - g(x+h)f(x)}{g(x)g(x+h)} \frac{1}{h} \\
&= \lim_{h\to 0} \frac{[f(x+h) - f(x)]g(x) - f(x)[g(x+h) - g(x)]}{h} \frac{1}{g(x)g(x+h)} \\
&= \lim_{h\to 0} \left[g(x) \frac{f(x+h) - f(x)}{h} - f(x) \frac{g(x+h) - g(x)}{h} \right] \frac{1}{g(x)g(x+h)} \\
&= \frac{f'(x)g(x) - g'(x)f(x)}{(g(x))^2},
\end{aligned}$$

as claimed. ■

There is a mnemonic for the quotient rule that goes "ho de-hi minus hi de-ho all over ho ho." If this is useful to you, great. If not, just Google quotient rule when you need to use it.

Example 4.3 (Rules of differentiation)

1. Find the derivatives of:
   (a) $y = \frac{2}{x^4+1} + \frac{3}{x}$
   (b) $y = 3x^2 + 2x^{\frac{1}{3}}$
   (c) $y = x^3(2x^4)$
   (d) $f(x) = \frac{x^2+1}{x^2-1}$
2. Find all points of $y = x^3 - x^2$ where the line tangent to the curve is horizontal.

We could of course do this graphically, but that isn't possible in general, which is why we can use derivatives. We have

$$\frac{dy}{dx} = 3x^2 - 2x$$

When the tangent line to the curve is horizontal, what is the slope? It's 0. Why? Recall that the slope of a line between (x_1, y_1) and (x_2, y_2) is $\frac{y_2 - y_1}{x_2 - x_1}$ (rise over run). If the line is horizontal, then $y_1 = y_2$ and therefore the slope is zero. So we must solve the following:

$$x(3x - 2) = 0$$

Which holds when $x = 0$ and when $x = \frac{2}{3}$.

There is another rule that we haven't talked about yet: how do we deal with composite functions? Consider trying to take the derivative of something like $y = (x^2 + 3x + 6)^{100}$. We could either generalize the multiplication rule, which would be unpleasant, or we could multiply out the entire polynomial, which would be painful. So instead we use the Chain Rule. In fact, the Chain Rule lets us do these things in a single step (or perhaps two). It is so important in calculus, that we rarely deal with things without using the Chain Rule.

Theorem 4.10 (Chain Rule) *Let f and g be differentiable functions, and consider their composition $(f \circ g)(x) = f(g(x))$. The derivative of the composition is $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$.*

In other words, the derivative of a composite function is the derivative of the outer function evaluated at the inner function times the derivative of the inner function.

This can give rise to some confusion when we have to apply multiple rules all at the same time. Consider

$$y = (x^2 - 3x)^2(x - 1)^2.$$

The derivative is gotten to via:

1. Apply the multiplication rule
2. Apply the chain rule to each term

So we have:

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

The basic lesson is that order matters. We have an informal rule in how to apply the rules of calculus:

The Last First Rule: the last step in calculation corresponds to the first step in differentiation.

So in our previous example, the last step in computing would be the multiplication of our two terms, hence this is the first step we will take in differentiating.

Example 4.4 (Chain Rule)

1. $y=(x^2+3x+6)^{100}$.

   Consider the above in terms of $F(x)=f(g(x))$, where $f(x)=x^{100}$ and $g(x)=x^2+3x+6$.

   $$\begin{aligned}F'(x) &= f'(g(x))\cdot g'(x)\\ &= 100(x^2+3x+6)^{99}\cdot(2x+3).\end{aligned}$$

2. $y=\frac{1}{(2x^5-7)^3}$.

   We can do this two ways:

   (a) Use the Quotient Rule and the Chain Rule

   (b) Think of the function using the rules of exponents. Namely, $F(x)=f(g(x))=(2x^5-7)^{-3}$, where $f(x)=x^{-3}$ and $g(x)=2x^5-7$.

   $$\begin{aligned}F'(x) &= f'(g(x))\cdot g'(x)\\ &= -3(2x^5-7)^{-4}\cdot 10x^4\\ &= \frac{-30x^4}{(2x^5-7)^4}.\end{aligned}$$

3. $y = \left(\frac{t^3-2t+1}{t^4+3}\right)^{13}$.

   Let $F(t) = f(g(t))$, where $f(t) = t^{13}$ and $g(t) = \frac{t^3-2t+1}{t^4+3}$.

$$\begin{aligned} F'(x) &= f'(g(x)) \cdot g'(x) \\ &= 13\left(\frac{t^3-2t+1}{t^4+3}\right)^{12} \cdot \frac{(3t^2-2)(t^4+3)-(4t^3)(t^3-2t+1)}{(t^4+3)^2}. \end{aligned}$$

Note that we can also apply the Chain Rule more than once; in principle the process can be iterated as many times as is necessary to get the job done. For example, consider $y = 10((x^2-3)^4 - x)^3$. We have

$$\frac{dy}{dx} = 30((x^2-3)^4 - x)^2 \cdot (4(x^2-3)^3(2x) - 1)$$

By now we've effectively covered the Chain Rule, though we'll revisit it when we talk about multivariate calculus. Now we have a few more foundational things to cover before we talk about some of the applications of differential calculus.

4.1.3 Higher-Order Derivatives

The operation of differentiation takes a function f and produces a new function f' which, evaluated at a point, is the slope of the original function at a line tangent to that point. There is no reason why we can't simply do this again, to produce f'', which is the change in the slope of the function at a particular point. The nth-order derivative can be written in any of the following ways:

- $f'^{\cdots\prime}(x)$ or $f^{(n)}(x)$
- $\frac{d^n f}{dx^n}$ or $\frac{d^n y}{dx^n}$
- $D_{x\cdots x} f(x)$

Example 4.5 (Higher-order derivatives)

1. $y = x^2$.

   Here, $\frac{dy}{dx} = 2x$ and $\frac{d^2y}{dx^2} = 2$. (Graph the function.) So in our little graph, as required, we have the original function, the derivative, a measure of the change in the original function at any particular point x and the 2nd derivative, a measure of the change in the change of the function at any particular point.

2. $y = 2x^3 - 4x^2 + 7x - 8$.

$$\frac{dy}{dx} = 6x^2 - 8x + 7$$
$$\frac{d^2y}{dx^2} = 12x - 8$$
$$\frac{d^3y}{dx^3} = 12$$
$$\frac{d^4y}{dx^4} = 0$$

We'll also return to higher-order differentiation when we talk about calculus in multiple dimensions. Until then, there's one more topic to cover before we look at applications of derivates.

4.1.4 Implicit Differentiation

What do we do when we cannot separate out the variables in a function? Consider, for example, $y^3 + 7y = x^3$. We call this an implicit function: there is no convenient way to give y as a function of x. So what do we do when we want $\frac{dy}{dx}$? We differentiate *both sides* of the equation with respect to x.

Example 4.6 (Implicit differentiation)

1. $y^3 + 7y = x^3$.

$$\begin{aligned}\frac{d(y^3)}{dx} + \frac{d(7y)}{dx} &= \frac{d(x^3)}{dx}\\ 3y^2\frac{dy}{dx} + 7\frac{dy}{dx} &= 3x^2\\ \frac{dy}{dx}(3y^2+7) &= 3x^2\\ \frac{dy}{dx} &= \frac{3x^2}{3y^2+7}\end{aligned}$$

2. $x^2 + 5y^3 = x + 9$.

$$\begin{aligned}\frac{d(x^2)}{dx} + \frac{d(5y^3)}{dx} &= \frac{d(x)}{dx} + \frac{d(9)}{dx}\\ 2x + 15y^2\frac{dy}{dx} &= 1\\ \frac{dy}{dx} &= \frac{1-2x}{15y^2}\end{aligned}$$

3. $x^4 + 2y^2 = 8$.

$$\begin{aligned}\frac{d(x^4)}{dx} + \frac{d(2y^2)}{dx} &= \frac{d(8)}{dx}\\ 4x^3 + 4y\frac{dy}{dx} &= 0\\ \frac{dy}{dx} &= -\frac{4x^3}{4y} = -\frac{x^3}{y}\end{aligned}$$

This is great for when we have a point in the cartesian plane (x, y) and we want to know what the derivative is at that point. It maybe isn't so great when we want to find a functional form for the derivative explicitly as a function of one variable. Still, it's useful to see, and you may end up manipulating things in this way, especially when deriving equations.

4.1.5 Applications of Differentiation

Often, we are interested, in political science, in the best way to do something — for example, what might a strategic politician want to maximize? Votes, as a function of, say, direct mailings and TV commercials. Such problems can be formulated as maximizing or minimizing a function over a specified set. When we want to do this, we can use the tools that we've just discussed for differential calculus to solve the problem. Of course, the first thing to do is decide whether f even has a maximum or minimum value on the specified set.

Definition 4.2. Let S be the domain of a function f, and suppose that it contains some point c.

- We say that $f(c)$ is the *maximum value* of f on S if $f(c) \geq f(x)$ for all $x \in S$.
- We say that $f(x)$ is the *minimum value* of f on S if $f(c) \leq f(x)$ for all $x \in S$.
- $f(c)$ is an *extreme value* of f on S if it is either the maximum value or the minimum value.
- We call the function we want to maximize or minimize the *objective function.* □

Now that we have a definition of these types of extreme points, we want to ask ourselves when they exist. For example, on the set $S = (0, \infty)$, the function $f(x) = \frac{1}{x}$ has no extreme value.

We have a partial answer at our disposal.

Theorem 4.11 *If f is continuous and its domain is a closed interval $[a, b]$, then f attains both a maximum value and a minimum value.*

So where do these extreme points occur? They will occur at one of these places:

1. Stationary points: these are points where the derivative of the objective function is 0. That is, stationary points are where the slope of the line tangent to the curve is 0, or the line is horizontal.

   These typically occur where the function takes on a maximum or minimum value. (Draw picture of $y = x^2$.) However, stationary points are not always extreme points: consider $y = x^3$ at $x = 0$. (Draw picture.)

2. Singular points: these are points in $\operatorname{int} S$ where the derivative of the objective function does not exist. These are places where the graph of f will have a sharp corner, vertical asymptote, or discontinuity. This kind of point is usually quite rare in practical problems. (But consider $y = |x|$.)

3. The end points of the interval. (Draw picture of a linear function.)

We call one of these types of points a *critical point*. The following theorem summarizes the relationship between critical points and extreme points.

Theorem 4.12 (Critical Point Theorem) *Let a function f be defined on an interval I containing the point c. If $f(c)$ is an extreme value, then c must be a critical point:*

1. *an endpoint of I*

2. *a stationary point of f on I*

3. *a singular point of f on I*

Keep in mind that the relationship does not go in reverse: a critical point is not necessarily an extreme point.

This suggests a really simple process for finding the maximum and minimum values of a function on a closed interval I.

1. Find all the critical points of the function f on I.

2. Evaluate the function at those critical points; the largest of these will be the maximum value, the smallest the minimum value.

Example 4.7 (Optimization)

1. Let's find the maximum and minimum values of $f(x) = -2x^3 + 3x^2$ on $[-\frac{1}{2}, 2]$. First, we must identify the critical points:

   (a) The endpoints, $-\frac{1}{2}$ and 2

(b) Stationary points

$$\begin{aligned} f'(x) = -6x^2 + 6x = 0 &\Leftrightarrow 6x = 6x^2 \\ &\Leftrightarrow x = x^2 \\ &\Leftrightarrow x \in \{0, 1\} \end{aligned}$$

(c) There are no singular points, since this is a polynomial.

So we have critical points at $-\frac{1}{2}$, 0, 1, 2.

Now we simply evaluate the function at each point.

- $f(-\frac{1}{2}) = -2(-\frac{1}{2})^3 + 3(-\frac{1}{2})^2 = \frac{2}{8} + \frac{3}{4} = 1$
- $f(0) = -2(0)^3 + 3(0)^2 = 0$
- $f(1) = -2(1)^3 + 3(1)^2 = 1$
- $f(2) = -2(2)^3 + 3(2)^2 = -4$

Therefore, we have a maximum value at $(-\frac{1}{2}, 1)$ and at $(1, 1)$ and a minimum value at $(2, -4)$.

2. Find the maximum values of $x^{\frac{2}{3}} = y$ on $[-1, 2]$.

First we find the critical values:

(a) the endpoints -1 and 2

(b) $\frac{dy}{dx} = \frac{2}{3}x^{\frac{-1}{3}}$. Note that this will never equal zero and this yields no critical points.

(c) Note also that the derivative does not exist at $x = 0$: we have a singular point at $x = 0$.

So we have critical points at $x = -1, 0, 2$. Now, we simply evaluate the function at each of these points.

- $(-1)^{\frac{2}{3}} = 1$
- $(0)^{\frac{2}{3}} = 0$
- $(2)^{\frac{2}{3}} = \sqrt[3]{4} \simeq 1.59$

So we have a maximum at $x = 2$ and a minimum at $x = 0$.

3. Suppose we are making a cardboard box by cutting out identical squares from each corner and turning up the edges. What is the maximum volume that can be attained if the cardboard piece is 24×9?

   Let x be the length of one side of the square to be cut out, and recall that $V = l \cdot w \cdot h$. The formula for the volume of the box is

   $$\begin{aligned} V &= x(24 - 2x)(9 - 2x) \\ &= (24x - 2x^2)(9 - 2x) \\ &= 216x - 48x^2 - 18x^2 + 4x^3 \\ &= 216x - 66x^2 + 4x^3 \end{aligned}$$

   Note also that $0 \leq x \leq 4.5$. So we are trying to maximize $V = 216x - 66x^2 + 4x^3$ on $[0, 4.5]$.

   Find the critical points:

   (a) The end points 0 and 4.5

   (b) The stationary points, obtained from

   $$\frac{dV}{dx} = 216 - 132x + 12x^2 = 12(18 - 11x + x^2) = 0$$

   This is an opportune time to remind ourselves of the quadratic formula: the roots of a degree 2 polynomial of the form $f(x) = ax^2 + bx + c$ are given by

   $$f(x) = 0 \quad \Leftrightarrow \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

   Here we have $a = 1$, $b = -11$, and $c = 18$, so we get

   $$x = \frac{-(-11) \pm \sqrt{121 - 4 \cdot 1 \cdot 18}}{2} = \frac{11 \pm \sqrt{49}}{2} = \frac{11 \pm 7}{2},$$

   and $x = 2$ or $x = 9$.

   Note that 9 is outside the interval of interest, so we have a critical point at 2.

   (c) There are no singular points, since V is a polynomial.

   Now we just evaluate the function at 0, 2, 4.5. At 0 and 4.5 the volume of the box is 0. Thus, we have a maximum at 2.

So we have some nice results that help us find and sort out global maxima and minima on the interval of interest. But we can introduce more nuance into the whole thing: *local* maxima and minima. To do this, we are going to have to introduce some new concepts.

Definition 4.3. Let the function f be defined as an interval I (open, closed, neither). We say that:

- f is *strictly increasing* on I if, for every pair of numbers x_1 and x_2 in I, $x_1 < x_2$ implies that $f(x_1) < f(x_2)$
- f is *strictly decreasing* on I if, for every pair of numbers x_1 and x_2 in I, $x_1 < x_2$ implies that $f(x_1) > f(x_2)$
- f is *strictly monotonic* on I if it is either increasing on I or decreasing on I. □

Now, if we remember that the derivative simply is a measure of the slope of the function, the next thing should be relatively straightforward:

Theorem 4.13 (Monotonicity Theorem) *Let f be continuous on an interval I and differentiable at every interior point of I.*

1. *If $f'(x) > 0$ for all $x \in \operatorname{int} I$, then f is increasing on I.*
2. *If $f'(x) < 0$ for all $x \in \operatorname{int} I$, then f is decreasing on I.*

Example 4.8 (Derivatives and monotonicity)
Let $f(x) = 2x^3 - 3x^2 - 12x + 7$ and find where f is increasing and where f is decreasing. We have

$$f'(x) = 6x^2 - 6x - 12 = 0,$$

which is equivalent to

$$x^2 - x - 2 = (x + 1)(x - 2) = 0.$$

So we want to know where $x^2 - x - 2 > 0$ and where $x^2 - x - 2 < 0$. Note that $f'(x) = 0$ at $x = -1$ and $x = 2$. These two points split the interval into three intervals: $(-\infty, -1)$, $(-1, 2)$, and $(2, \infty)$. Now we simply evaluate the function within each interval. Let's use these points: -2, 0, 3.

- $f'(-2) = (-2)^2 - (-2) - 2 = 6$
- $f'(0) = (0)^2 - (0) - 2 = -2$
- $f'(3) = (3)^2 - 3 - 2 = 4$

So the function is increasing on $(-\infty, -1)$ and $(2, \infty)$ and decreasing on $(-1, 2)$.

Now we'll move from increasing and decreasing to talk about something particularly important when defining preferences.

Definition 4.4. Let f be differentiable on an open interval I. We say that f is *concave up*, or *convex*, on I if f' is increasing on I. We say that f is *concave down*, or simply *concave*, on I if f' is decreasing on I. □

It should be clear how we can define concavity in terms of derivatives.

Theorem 4.14 (Concavity Theorem) *Let a function f be twice differentiable on the open interval I.*

- *If $f''(x) > 0$ for all x in I, then f is concave up on I.*
- *If $f''(x) < 0$ for all x in I, then f is concave down on I.*

Points where a function switches from concave up to concave down, or vice versa, are called inflection points.

Definition 4.5. Let f be a function that is continuous at c. The point $(c, f(c))$ is an *inflection point* if f is concave up on one side of c and down on the other side of c. □

Just as there is a relationship between extreme points and the first derivative, there is one between inflection points and the second derivative. In specific, the candidates for inflection points are those where $f''(x) = 0$ or where $f''(x)$ does not exist. Therefore, the process for finding inflection points is:

1. Find all the critical points of the first derivative (not counting the endpoints).

2. Check concavity on either side of each critical point.

Example 4.9 (Concavity and inflection points)

1. Find all the inflection points of $y = x^{\frac{1}{3}} + 2$.

   The first two derivatives are:

   $$\frac{dy}{dx} = \frac{1}{3}x^{\frac{-2}{3}} = \frac{1}{3x^{\frac{2}{3}}}$$
   $$\frac{d^2y}{dx^2} = \frac{-2}{9}x^{\frac{-5}{3}} = \frac{-2}{9x^{\frac{5}{3}}}$$

   Note that the second derivative is never zero, but it does fail to exist at $x = 0$. So now, we simply make use of our Concavity Theorem, and examine the intervals $(-\infty, 0)$ and $(0, \infty)$. Since $x = 0$ is the only critical point, if the function turns out to be concave up on one side and concave down on the other, we can say that $x = 0$ is an inflection point of f.

   So we want to know where $\frac{d^2y}{dx^2} > 0$ and where $\frac{d^2y}{dx^2} < 0$. For all $x < 0$, the denominator will be negative and so $\frac{d^2y}{dx^2} > 0$. For all $x > 0$, the denominator will be positive and so $\frac{d^2y}{dx^2} < 0$. Hence y is concave up on $(-\infty, 0)$ and concave down $(0, \infty)$, so $x = 0$ is an inflection point.

2. Where is $f(x) = \frac{1}{3}x^3 - x^2 - 3x + 4$ increasing, decreasing, concave up, and concave down? We have

   $$f'(x) = x^2 - 2x - 3 = (x - 3)(x + 1)$$
   $$f''(x) = 2x - 2$$

   Let's start by finding where it is increasing and decreasing. We need to determine where $f'(x) > 0$ and where $f'(x) < 0$, then we apply the Monotonicity Theorem. The intervals of interest are $(-\infty, -1)$, $(-1, 3)$, and $(3, \infty)$. Taking as test points -2, 0, 4:

   - $f'(-2) = (-2)^2 - 2(-2) - 3 = 4 + 4 - 3 = 5 > 0$
   - $f'(0) = 0^2 - 2(0) - 3 = -3 < 0$

- $f'(4) = 4^2 - 2(4) - 3 = 5 > 0$

So $f(x)$ is increasing on $(-\infty, -1)$ and $(3, \infty)$ and decreasing on $(-1, 3)$. Now let's look for inflection points. Since $f''(x)$ is a polynomial, it exists everywhere, and it is clear to see that $f''(x) < 0$ if $x < 1$ and $f''(x) > 0$ if $x > 1$. Therefore, f is concave down on $(-\infty, 1)$ and concave up on $(1, \infty)$, which means $x = 1$ is an inflection point.

With all this in place, we can finally talk about local maxima and minima. We'll begin by defining these terms precisely.

Definition 4.6. Let S be the domain of the function f and pick a point $c \in S$. We say:

1. $f(c)$ is a local maximum value of f if there exists $\epsilon > 0$ such that for all $x \in (c - \epsilon, c + \epsilon) \cap S$, $f(c) \geq f(x)$.
2. $f(c)$ is a local minimum value of f if there exists $\epsilon > 0$ such that for all $x \in (c - \epsilon, c + \epsilon) \cap S$, $f(c) \leq f(x)$.
3. $f(c)$ is a local extreme value if it is either a local maximum or a local minimum. □

So you can think of a local maximum as a point that is a global maximum with respect to a sufficiently small interval around it. This means that the criteria from the Critical Point Theorem also apply to local maxima and minima—and that not every critical point is necessarily a local extreme point.

Theorem 4.15 (First Derivative Test) *Let f be continuous on an open interval (a, b), and let $c \in (a, b)$ be a critical point of f.*

- *If $f'(x) > 0$ for all x in (a, c) and $f'(x) < 0$ for all x in (c, b), then $f(c)$ is a local maximum of f.*
- *If $f'(x) < 0$ for all x in (a, c) and $f'(x) > 0$ for all x in (c, b), then $f(c)$ is a local minimum of f.*

- *If $f'(x)$ has the same sign on both sides of c then $f(c)$ is not a local extreme value of f.*

Example 4.10 (First Derivative Test)
We want to find the local extreme values of $f(x) = \frac{1}{3}x^3 - x^2 - 3x + 4$ on $(-\infty, \infty)$. Recall that

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

and there are no endpoints or singular points, so we are only concerned with critical points. $f'(x) = x^2 - 2x - 3 = (x-3)(x+1) = 0$. This will be zero at $x = 3$ and $x = -1$. So we have three intervals: we need to evaluate $f(x)$ on $(-\infty, -1)$, $(-1, 3)$ and $(3, \infty)$. As before, let's simply pick some test points: $x = -2$, $x = 0$, and $x = 4$.

- $f'(-2) = (-2)^2 - 2(-2) - 3 = 4 + 4 - 3 = 5 > 0$
- $f'(0) = 0^2 - 2(0) - 3 = -3 < 0$
- $f'(4) = 4^2 - 2(4) - 3 = 5 > 0$

So we have a local maximum at $x = -1$ and a local minimum at $x = 3$.

It's usually even easier than this to determine whether a stationary point is a local extreme point, because there is another test that relies on the concept of concavity.

Theorem 4.16 (Second Derivative Test) *Let f' and f'' exist at every point in an open interval (a, b) containing c and suppose that $f'(c) = 0$.*

1. *If $f''(c) < 0$, $f(c)$ is a local maximum value of f.*
2. *If $f''(c) > 0$, $f(c)$ is a local minimum value of f.*

Example 4.11 (Second Derivative Test)

1. First we'll go back to the example we used to illustrate the First Derivative Test:

$$\begin{aligned} f(x) &= \frac{1}{3}x^3 - x^2 - 3x + 4 \\ f'(x) &= x^2 - 2x - 3 & &= (x-3)(x+1) \\ f''(x) &= 2x - 2 \end{aligned}$$

   As before, the stationary points are $x = 3$ and $x = -1$. We have $f''(3) = 4 > 0$, so $x = 3$ is a local minimum; and $f''(-1) = -3$, so $x = -1$ is a local maximum.

2. Find the local extreme values of the function $y = x^2 - 6x + 5$ on $(-\infty, \infty)$. So we need to find the critical points. There are obviously no endpoints. Moreover, y is a polynomial, so there are no singular points. The stationary points are given by

$$\frac{dy}{dx} = 2x - 6 = 0,$$

   which is true only for $x = 3$. We have

$$\frac{d^2y}{dx^2} = 2,$$

   meaning the function is always concave up, so $x = 3$ must be a local minimum.

4.1.6 L'Hopital's Rule

Using derivatives, we are ready to go back and deal with a problem encountered back in the discussion of limits: how can we deal with indeterminate forms, where the numerator and denominator are both zero, or both infinite? For example, what is

$$\lim_{x\to 3} \frac{x^2 - 9}{x^2 - x - 6}?$$

In a situation like this, you'll find that the easiest way to proceed is by the rule given in the next two theorems. We first consider forms of the type $\frac{0}{0}$.

Theorem 4.17 (L'Hopital's Rule for $\frac{\mathbf{0}}{\mathbf{0}}$) *Suppose that* $\lim_{x\to u} f(x) = \lim_{x\to u} g(x) = 0$*, where* u *may be finite or infinite. If* $\lim_{x\to u} \frac{f'(x)}{g'(x)}$ *exists, then:*

$$\lim_{x\to u} \frac{f(x)}{g(x)} = \lim_{x\to u} \frac{f'(x)}{g'(x)}.$$

Theorem 4.18 (L'Hopital's Rule for $\pm\frac{\infty}{\infty}$) *Suppose that* $\lim_{x\to u} |f(x)| = \lim_{x\to u} |g(x)| = \infty$*, where* u *may be finite or infinite. If* $\lim_{x\to u} \frac{f'(x)}{g'(x)}$ *exists, then:*

$$\lim_{x\to u} \frac{f(x)}{g(x)} = \lim_{x\to u} \frac{f'(x)}{g'(x)}.$$

Example 4.12 (L'Hopital's Rule)

1. Find $\lim_{x\to 3} \frac{x^2-9}{x^2-x-6}$.

   Here we have a $\frac{0}{0}$ form. Applying L'Hopital's Rule, we get

   $$\lim_{x\to 3} \frac{x^2-9}{x^2-x-6} = \lim_{x\to 3} \frac{2x}{2x-1} = \frac{6}{5}.$$

2. Find $\lim_{x\to\infty} \frac{x}{x^2+1}$.

   This is a case where the limit has the form $\frac{\infty}{\infty}$. Applying L'Hopital's Rule, we get

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

4.2 Integral Calculus

4.2.1 Indefinite Integrals

So far our focus in calculus has been on the derivative. Now we're going to consider its inverse, known as integration or antidifferentiation.

Definition 4.7. We say that the function F is an *antiderivative* of the function f on the interval I if $F'(x) = f(x)$ for all $x \in I$. If F is an antiderivative for f on its entire domain, we write

$$\int f(x)\,dx = F(x) + c,$$

which is also known as an *indefinite integral*, where the function $f(x)$ is the *integrand*. □

The c in the definition is taken to be an arbitrary constant, which we will explain in a second.

Example 4.13 (Antiderivatives)

1. Find an antiderivative of $4x^3$ on $(-\infty, \infty)$.

   We must find y so that $\frac{dy}{dx} = 4x^3$. One such function that will satisfy this would be $y = x^4$.

   However, this is not the *only* antiderivative of $4x^3$. Another would be $y = x^4 + 1$, and yet another would be $y = x^4 + 42$. In fact, a function with one antiderivative has infinitely many, which is why we write $\int f(x)\,dx = F(x) + c$. The c stands for the fact that you can add any constant to an antiderivative and it will still be an antiderivative of the original function.

2. Find an antiderivative of $\frac{1}{3}x^3 - x^2 - 3x + 4$.

   As before, we want y such that $\frac{dy}{dx} = \frac{1}{3}x^3 - x^2 - 3x + 4$. We can satisfy this with

$$\begin{aligned} y &= \frac{1}{3}\left(\frac{1}{4}x^4\right) - \frac{1}{3}x^3 - 3\left(\frac{1}{2}x^2\right) + 4x \\ &= \frac{x^4}{12} - \frac{x^3}{3} - \frac{3x^2}{2} + 4x. \end{aligned}$$

You may notice a pattern that we used to obtain these antiderivatives, which is generalized in the following theorem.

Theorem 4.19 (Power Rule) *If r is any rational number other than -1, then*

$$\int x^r dx = \frac{x^{r+1}}{r+1} + c.$$

Proof. Pick any rational number $r \neq -1$ and let $F(x) = \frac{x^{r+1}}{r+1} + c$ for some $c \in \mathbb{R}$. By the Power Rule for differentiation, we have

$$F'(x) = \left(\frac{1}{r+1}\right)(r+1)x^r = x^r,$$

as claimed. ∎

One thing to note about the theorem we've just presented: the conclusion can only be valid on an interval where x^r is defined. So, for example, for any f in which $r < 0$, the interval cannot include 0, since f would not be defined at 0.

Theorem 4.20 *Let f and g be functions that have antiderivatives and let k be a constant. We have:*

1. $\int kf(x)\,dx = k\int f(x)\,dx$
2. $\int (f(x)+g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$
3. $\int (f(x)-g(x))\,dx = \int f(x)\,dx - \int g(x)\,dx$

That is, indefinite integration is a linear operator.

Proof. For part 1, we have

$$\frac{d(k\int f(x)\,dx)}{dx} = \frac{k(d(\int f(x)\,dx)}{dx} = kf(x).$$

For part 2, we have

$$\begin{aligned}\frac{d}{dx}\left(\int f(x)\,dx + \int g(x)\,dx\right) &= \frac{d}{dx}\int f(x)\,dx + \frac{d}{dx}\int g(x)\,dx \\ &= f(x) + g(x)\end{aligned}$$

Part 3 follows directly from parts 1 and 2. ∎

Example 4.14 (Integration as a linear operator)

1. $\int (3x^2 + 4x)\,dx = \int 3x^2\,dx + \int 4x\,dx = x^3 + 2x^2 + c.$

2. What is $\int (u^{\frac{3}{2}} - 3u + 14)\,du$?

$$\begin{aligned} \int (u^{\frac{3}{2}} - 3u + 14)\,du &= \int u^{\frac{3}{2}}\,du - \int 3u\,du + \int 14\,du \\ &= \frac{u^{\frac{3}{2}+\frac{3}{3}}}{\frac{3}{2}+\frac{3}{3}} - \frac{3u^{1+1}}{1+1} + 14u + c \\ &= \frac{u^{\frac{5}{2}}}{\frac{5}{2}} - \frac{3}{2}u^2 + 14u + c. \end{aligned}$$

3. What is $\int (\frac{1}{t^2} + \sqrt{t})\,dt$?

$$\begin{aligned} \int \left(\frac{1}{t^2} + \sqrt{t}\right)\,dt &= \int \frac{1}{t^2}\,dt + \int \sqrt{t}\,dt \\ &= \int t^{-2}\,dt + \int t^{\frac{1}{2}}\,dt \\ &= -t^{-1} + \frac{t^{\frac{1}{2}+\frac{2}{2}}}{\frac{1}{2}+\frac{2}{2}} + c \\ &= \frac{-1}{t} + \frac{2t^{\frac{3}{2}}}{3} + c. \end{aligned}$$

We can actually generalize the Power Rule we're using here even further.

Theorem 4.21 (Generalized Power Rule) *Let g be a differentiable function and take any $r \neq -1$. Then*

$$\int (g(x))^r g'(x)\,dx = \frac{g(x)^{r+1}}{r+1} + c.$$

This is the integral answer to the technique of differentiation we called the Chain Rule. Basically, we are taking the integral of something we got from applying the Chain Rule. The most important thing to do here is to recognize the different parts of the integrand. One part will correspond to $g(x)^r$ and another to $g'(x)$. This can be tough: we may see some tricks later that help with problems like this.

Example 4.15 (Generalized Power Rule)

1. Evaluate $\int (x^4 + 3x)^{30}(4x^3 + 3)\,dx$.

   First, we ought to recognize this as the output of something with the Chain Rule applied to it. What are the respective parts?

   $$g(x)^r = (x^4 + 3x)^{30}$$
   $$g'(x) = 4x^3 + 3$$

   So the integral above can be evaluated by applying the theorem:

   $$\int (x^4 + 3x)^{30}(4x^3 + 3)\,dx = \frac{(x^4 + 3x)^{31}}{31} + c.$$

2. Evaluate $\int (x^3 + 6x)^5(6x^2 + 12)\,dx$.

   We'll use this as a simple example of the method of substitution: we define a new variable u from x, rewrite the integral in terms of u (hopefully in a form we find more friendly, such as one where we can easily apply the Generalized Power Rule), and then substitute x back in at the end.

   Let $u = x^3 + 6x$, so $du = 3x^2 + 6\,dx$ and $6x^2 + 12\,dx = 2(3x^2 + 6)\,dx = 2\,du$. So our problem becomes the following:

   $$\int (x^3 + 6x)^5(6x^2 + 12)\,dx = \int u^5 \cdot 2\,du = \frac{2}{6}u^6 + c = \frac{1}{3}(x^3 + 6x)^6 + c.$$

3. Evaluate $\int (x^2 + 4)^{10} x\,dx$.

   This time we'll make the substitution $u = x^2 + 4$, which means $du = 2x\,dx$. So have have

   $$\int (x^2 + 4)^{10} x\,dx = \int u^{10} \cdot \frac{1}{2}\,du = \frac{1}{22}u^{11} + c = \frac{1}{22}(x^2 + 4)^{11} + c.$$

4.2.2 Definite Integrals

So what is our interpretation of the integral? Recall that a derivative gives the rate of change of a function at each point in its domain. It was, in essence, the slope of the line tangent to the curve at each point. There is a similar interpretation for the integral: the output from integrating a function over an interval is the (signed) area under the curve of the function on that interval.

It is easy to compute the area of a rectangle: it's the product of the width and the length. But what about the area under a smooth curve? Archimedes gave us an answer: we inscribe polygons within the curve. So consider $y = x^2$. Suppose we are interested in calculating the region under the curve along the interval from 0 to 2. Then what we can do is partition that interval into n subintervals, each of equal length $\frac{2}{n} = \Delta x$ by picking $n+1$ points on the interval. Each of these inscribed rectangles will have width $\Delta x = x_i - x_{i-1}$ and height $f(x_{i-1}) = (x_{i-1})^2$. Then we can approximate the area under the curve by simply summing up the rectangles.

$$\begin{aligned}
Area &= f(x_0)(x_1 - x_0) + f(x_1)(x_2 - x_1) + \ldots + f(x_{n-1})(x_n - x_{n-1}) \\
&= f(x_0)\Delta x + f(x_1)\Delta x + \ldots + f(x_{n-1})\Delta x \\
&= \frac{2}{n}((x_0)^2 + (x_1)^2 + \ldots + (x_{n-1})^2)
\end{aligned}$$

Now note that $x_0 = 0$, $x_1 = \frac{2}{n}$, $x_2 = \frac{2}{n} \cdot 2$, $\ldots$, $x_{n-1} = \frac{2}{n} \cdot (n-1)$, which implies $x_i = i\frac{2}{n}$. Thus:

$$\begin{aligned}
Area &= \frac{2}{n}\left(\left(\frac{2}{n} \cdot 0\right)^2 + \left(\frac{2}{n} \cdot 1\right)^2 + \ldots + \left(\frac{2}{n} \cdot (n-1)\right)^2\right) \\
&= \frac{2}{n} \cdot \frac{4}{n^2} \sum_{i=0}^{n-1} i^2 \\
&= \frac{8}{n^3} \sum_{i=0}^{n-1} i^2
\end{aligned}$$

As it turns out, there is a special formula for this sum, which is

$$\sum_{i=0}^{n-1} i^2 = \frac{(n-1)(n)(2n-1)}{6}.$$

So we have

$$\begin{aligned} Area &= \frac{8}{n^3}\frac{(n-1)(n)(2n-1)}{6} \\ &= \frac{8}{6n^3}(2n^3 - n^2 - 2n^2 + n) \\ &= \frac{8}{6n^2}(2n^2 - 3n + 1) \\ &= \frac{8}{3} - \frac{4}{n} + \frac{4}{3n^2} \end{aligned}$$

We can make the approximation more accurate by adding more and more points. So what happens as we take the limit of this as $n \to \infty$?

$$\lim_{n\to\infty} = \frac{8}{3} - \frac{4}{n} + \frac{4}{3n^2} = \frac{8}{3}$$

So what this is actually saying is that as we let the bases of the rectangles get smaller and smaller, we end up converging on the true area under the curve.

Having defined the relationship between the area of a curve and how we can approximate it, we are now ready to begin to discuss the definite integral. The modern definition of the definite integral we owe to the Reimann sum, and so we'll need to talk about that (quickly) first.

Definition 4.8. Consider a function f defined on a closed interval $[a, b]$ and suppose that we partition the interval into n subintervals (not necessarily equal) by means of points $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$, and let $\Delta x_i = x_i - x_{i-1}$. On each subinterval, pick an arbitrary point $\bar{x}_i$, which we will call a sample point for the interval. Then we call the sum

$$R_p = \sum_{i=1}^{n} f(\bar{x}_i)\Delta x_i$$

a *Reimann sum* for the function f corresponding to the partition p. □

Example 4.16 (Riemann sums)
Evaluate the Reimann sum R_p for $f(x) = (x+1)(x-2)(x-4)$ on the interval $[0, 5]$ using partition p with points $0 < 1.1 < 2 < 3.2 < 4 < 5$

and sample points $\bar{x}_1 = 0.5$, $\bar{x}_2 = 1.5$, $\bar{x}_3 = 2.5$, $\bar{x}_4 = 3.6$, and $\bar{x}_5 = 4.5$. We have

$$\begin{aligned}
R_p &= \sum_{i=1}^{n} f(\bar{x}_i)\Delta x_i \\
&= f(\bar{x}_1)\Delta x_1 + f(\bar{x}_2)\Delta x_2 + f(\bar{x}_3)\Delta x_3 + f(\bar{x}_4)\Delta x_4 + f(\bar{x}_5)\Delta x_5 \\
&= f(\bar{x}_1)(1.1 - 0) + f(\bar{x}_2)(2 - 1.1) + f(\bar{x}_3)(3.2 - 2) \\
&\quad + f(\bar{x}_4)(4 - 3.2) + f(\bar{x}_5)(5 - 4) \\
&\simeq 8.66 + 2.81 - 3.15 - 2.36 + 6.88 \\
&\simeq 12.84.
\end{aligned}$$

We use the Reimann sum to define the definite integral as follows.

Definition 4.9. Let f be a function on the closed interval $[a, b]$. Denote a partition of $[a, b]$ by p, and let $|p|$ denote the length of the longest subinterval of the partition. If

$$\lim_{|p|\to 0} \sum_{i=1}^{n} f(\bar{x}_i)\Delta x_i$$

exists, we say f is *integrable* on $[a, b]$. Moreover $\int_a^b f(x)\,dx$, called the *definite integral* from a to b, is given by

$$\int_a^b f(x)\,dx = \lim_{|p|\to 0} \sum_{i=1}^{n} f(\bar{x}_i)\Delta x_i.$$

□

So how does the definite integral relate to area? It is simply the generalized version of the inscribed polygons that we just discussed. In particular, $\int_a^b f(x)\,dx$ gives us the signed area under the curve. Take note of the term *signed* area. The integral attaches a positive sign to areas under the curve yet above the x axis and a negative sign to areas under the x axis and above the curve. That is,

$$\int_a^b f(x)\,dx = \text{area above } x \text{ axis} - \text{area below } x \text{ axis}$$

Now let's consider some simple and hopefully helpful properties of integrals.

Theorem 4.22 (Integrability Theorem) *If a function f is bounded on $[a, b]$ and is continuous on $[a, b]$ except for a finite number of points at most, then f is integrable on $[a, b]$. In particular, if f is continuous on the whole interval $[a, b]$ it is integrable on $[a, b]$.*

Some implications of this theorem and the definition of integration:

1. A polynomial function is integrable on every closed interval $[a, b]$
2. A rational function is integrable on any closed interval $[a, b]$ that contains no points where the function's denominator equals 0
3. $\int_a^a f(x)\,dx = 0$
4. $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$

Theorem 4.23 *If f is integrable on an interval containing the points a, b, and c, then*

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

We could spend time on evaluating Reimann sums ... but I'm not convinced that this is a particularly useful topic (since you will never do it in practice, nor is it instructive toward another topic). So let's move on to some fundamental (literally) results in calculus.

Theorem 4.24 (First Fundamental Theorem of Calculus) *Let f be continuous on the closed interval $[a, b]$, and let x be a variable point in (a, b). Then:*

$$\frac{d}{dx}\left(\int_a^x f(t)\right) = f(x)$$

In words, we might say that the rate of accumulation at $t = x$ is equal to the value of the function being accumulated at $t = x$. What this theorem establishes is that integration is indeed the inverse of differentiation—taking the derivative of an integral gives you the original function.

Example 4.17 (First Fundamental Theorem of Calculus)

1. Find $\frac{d}{dx}(\int_1^x t^3\,dt)$.

   By the First Fundamental Theorem of Calculus,

$$\frac{d}{dx}\left(\int_1^x t^3\,dt\right) = x^3.$$

2. Find $\frac{d}{dx}(\int_1^{x^2}(3t-1)\,dt$.

   What do we do when x^2 is one of the limits of integration? Consider this as an application of the Chain Rule, where $f(x) = \int_1^x (3t-1)\,dt$, $g(x) = x^2$, and $h(x) = f(g(x)) = \int_1^{x^2}(3t-1)\,dt$. We then have, by the Chain Rule and the First Fundamental Theorem,

$$\begin{aligned} h'(x) &= f'(g(x)) \cdot g'(x) \\ &= (3g(x) - 1) \cdot (2x) \\ &= 2x(3x^2 - 1). \end{aligned}$$

Theorem 4.25 (Second Fundamental Theorem of Calculus) *Let f be continuous (and thus integrable) on $[a, b]$ and let F be any antiderivative of f on $[a, b]$. Then*

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

This is the result that you will make use of almost all the time when you are evaluating functions and integrating in your classes. It is also why we really never need to bother with actually computing the limit for the Reimann Sum. As such, this is one of the most important theorems you will learn in this course.

Finally, we can also establish various nice properties about definite integrals that make them easier to compute in practice.

Theorem 4.26 *If f and g are integrable on $[a, b]$ and $f(x) \leq g(x)$ for all x in $[a, b]$, then*

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

Theorem 4.27 *If f is integrable on $[a,b]$ and $m \leq f(x) \leq M$ for all $x \in [a,b]$, then $m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a)$.*

Theorem 4.28 *Suppose that f and g are integrable on $[a,b]$ and that k is a constant. Then kf and $f+g$ are integrable and*

- $\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$
- $\int_a^b (f(x) \pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$

Example 4.18 (Second Fundamental Theorem of Calculus)

1. Show that if k is a constant, $\int_a^b k\,dx = k(b-a)$.

   $F(x) = kx$ is an antiderivative of $f(x) = k$. So we now apply the Second Fundamental Theorem of Calculus:

   $$\int_a^b k\,dx = F(b) - F(a) = kb - ka = k(b-a).$$

2. Show that $\int_a^b x\,dx = \frac{b^2}{2} - \frac{a^2}{2}$.

   As before, $F(x) = \frac{x^2}{2}$ is an antiderivative of $f(x) = x$. Thus, by the Second Fundamental Theorem of Calculus:

   $$\int_a^b x\,dx = F(b) - F(a) = \frac{b^2}{2} - \frac{a^2}{2}.$$

3. Evaluate $\int_{-1}^2 (4x - 6x^2)\,dx$.

   Applying the various results from above, we have

   $$\begin{aligned}\int_{-1}^2 (4x-6x^2)\,dx &= \int_{-1}^2 4x\,dx - \int_{-1}^2 6x^2\,dx \\ &= 2x^2\Big|_{x=-1}^2 - 2x^3\Big|_{x=-1}^2 \\ &= 2(2)^2 - 2(-1)^2 - (2(2)^3 - 2(-1)^3) \\ &= 8 - 2 - 16 - 2 \\ &= -12.\end{aligned}$$

4. Evaluate $\int_0^4 \sqrt{x^2+x}(2x+1)\,dx$.

   What do we do to find the antiderivative here? We use the same substitution method as in our previous examples of the Generalized Power Rule. Let $u = x^2 + x$ then $du = 2x + 1\,dx$. Now making the relevant substitutions, we can readily evaluate the integral:

$$\begin{aligned}\int_0^4 \sqrt{x^2+x}(2x+1)\,dx &= \int_0^{20} u^{\frac{1}{2}}\,du \\ &= \frac{2}{3}u^{\frac{3}{2}}\bigg|_{u=0}^{20} \\ &= \frac{2}{3}20^{\frac{3}{2}}.\end{aligned}$$

   Notice that the limits of integration change in the first line when we go from x to u. In particular, the upper limit changes to $u = 4^2 + 4 = 20$ and the lower limit changes to $u = 0^2 + 0 = 0$.

5. Evaluate $\int_0^1 x^2 + (x^2+1)^4 x\,dx$.

   By linearity,

$$\int_0^1 x^2 + (x^2+1)^4 x\,dx = \int_0^1 x^2\,dx + \int_0^1 (x^2+1)^4 x\,dx$$

   We can easily do the first integral:

$$\int_0^1 x^2\,dx = \frac{x^3}{3}\bigg|_{x=0}^{1} = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}.$$

   For the second part, let $u = x^2+1$. Then $du = 2x\,dx$, and $\frac{du}{2} = x\,dx$. Then, making the relevant substitutions, we have

$$\begin{aligned}\int_0^1 (x^2+1)^4 x\,dx &= \int_1^2 \frac{u^4}{2}\,du \\ &= \frac{u^5}{10}\bigg|_{u=1}^{2} \\ &= \frac{32}{10} - \frac{1}{10} = \frac{31}{10}.\end{aligned}$$

(Note again how we changed the limits of integration.) So the overall integral is $\frac{1}{3} + \frac{31}{10}$.

The next couple of theorems formalize the technique of substitution that we've been using.

Theorem 4.29 *Let g be a differentiable function and suppose that F is an antiderivative of f. Then, if we let $u = g(x)$,*

$$\int f(g(x))g'(x)\,dx = \int f(u)\,du = F(u) + c = F(g(x)) + c$$

This just tells us that our substitution method works and can be applied broadly. The next theorem formalizes our way of dealing with the limits of integration when evaluating definite integrals.

Theorem 4.30 *Let g have a continuous derivative on $[a, b]$ and let f be continuous on the range of g. Then, letting $u = g(x)$, we have*

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du.$$

The next theorem allows us to simplify our calculations when the integrand (the function being integrated) takes a particular form. First we need to define even and odd functions.

Definition 4.10. An *even function* is a function for which $f(x) = f(-x)$ for all x in the domain of f. An *odd function* is a function for which $f(-x) = -f(x)$ for all x in the domain of f. □

Now we can give the result.

Theorem 4.31 *If f is an integrable even function, then*

$$\int_{-a}^a f(x)\,dx = 2\int_0^a f(x)\,dx.$$

If f is an integrable odd function, then

$$\int_{-a}^a f(x)\,dx = 0.$$

The two assertions in the above theorem correspond to cases where area is the same on both sides of the origin, and where area on one side of the origin cancels area on the other side of the origin.

Example 4.19 (Even and odd functions)

1. Evaluate $\int_{-3}^{3} |x|\, dx$. This is an even function. Thus, we can do the following:

$$\int_{-3}^{3} |x|\, dx = 2\int_{0}^{3} x\, dx = 2 \cdot \left.\frac{x^2}{2}\right|_{x=0}^{3} = 9$$

2. Evaluate $\int_{-5}^{5} \frac{x^5}{x^2+4}\, dx$. Is this function even or odd? Substitute $-x$ in for x:

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

This function is odd. Thus,

$$\int_{-5}^{5} \frac{x^5}{x^2+4}\, dx = 0.$$

4.2.3 Natural Logarithms

Notice that we haven't really dealt with any kind of exponential functions yet. And in fact, we have a gap in our knowledge of derivatives.

$$\begin{aligned}
\frac{d(\frac{x^3}{3})}{dx} &= x^2 \\
\frac{d(\frac{x^2}{2})}{dx} &= x \\
\frac{d(x)}{dx} &= x^0 \\
\frac{d(???)}{dx} &= x^{-1} \\
\frac{d(x^{-1})}{dx} &= x^{-2}
\end{aligned}$$

Recall that the first fundamental theorem of calculus asserted that we can find the anti-derivative of any continuous function. However, this does not mean that we can describe that anti-derivative in terms of what we already know—so we'll need some new machinery to describe the antiderivative of x^{-1}.

Definition 4.11. The natural logarithm function, denoted $\ln(x)$, is defined by

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

for all $x > 0$. $\square$

The natural logarithm function measures the area under the curve $y = \frac{1}{t}$ between 1 and x if $x > 1$, and the negative of the area for $0 < x < 1$. Now we note that the derivative of this new function is exactly what we were looking for: by the First Fundamental Theorem of Calculus,

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

Let's show that $\frac{d \ln|x|}{dx} = \frac{1}{x}$ for all $x \neq 0$. We have two cases to consider: either $x > 0$ or $x < 0$. If $x > 0$, $|x| = x$ and $\frac{d \ln|x|}{dx} = \frac{d(\ln(x))}{dx} = \frac{1}{x}$. If $x < 0$, $|x| = -x$ and so: $\frac{d \ln|x|}{dx} = \frac{d \ln(-x)}{dx} = \frac{1}{-x}(-1) = \frac{1}{x}$. Thus, what happens when we integrate $\frac{1}{x}$ is the following:

$$\int \frac{1}{x}\, dx = \ln|x| + c.$$

Example 4.20 (Natural logarithm)

1. Evaluate $\int \frac{5}{2x+7}\, dx$. Let $u = 2x + 7$. Then $du = 2\, dx$. So we can rewrite the problem as

$$\frac{5}{2} \int \frac{1}{2x+7} \cdot 2\, dx = \frac{5}{2} \int \frac{1}{u}\, du = \frac{5}{2} ln|u| + c = \frac{5}{2} \ln|2x + 7| + c.$$

2. Evaluate $\int_{-1}^{3} \frac{x}{10-x^2}\, dx$. Let $u = 10 - x^2$. Then $du = -2x\, dx$ and thus $x\, dx = \frac{-du}{2}$.

Making the proper substitutions, we see that:

$$\begin{aligned}\int_{-1}^{3} \frac{x}{10-x^2}\,dx &= \frac{-1}{2}\int_{9}^{1} \frac{1}{u}\,du \\ &= \frac{-1}{2}\ln|u|\bigg|_{u=9}^{1} \\ &= \frac{-1}{2}(\ln 1 - \ln 9) \\ &= \frac{1}{2}\ln 9.\end{aligned}$$

Since we've defined ln as the natural logarithm, by now you should be able to guess what its inverse is: an exponent.

Definition 4.12. The inverse of ln is called the *natural exponential function* and is denoted by exp. Thus $x = e^y$ if and only if $y = \ln x$. Note that the symbol exp(1) or e denotes the unique positive number such that $\ln e = 1$.□

By defining e in this way, it follows immediately that the following two things hold:

- $e^{\ln(x)} = x$ for all $x > 0$
- $\ln(e^y) = y$ for all y

Note that the number e, which is itself irrational ($e \simeq 2.718$) is one of the most important numbers in math; you'll find yourselves making quite a bit of use of e in PSC 404 with various probability distributions.

The natural exponent has the usual properties of an exponent, such as $e^a \cdot e^b = e^{a+b}$ and $e^a/e^b = e^{a-b}$. But it also has a special property in the context of differentiation: if $y = e^x$, then

$$\frac{dy}{dx} = e^x = y.$$

That is, the natural exponent is its own derivative. This is important enough that we'll go ahead and prove it.

Proof. Suppose $y = e^x$. We then have

$$\begin{aligned} x &= \ln y. \\ \frac{d}{dx} x &= \frac{d}{dx} \ln y. \\ 1 &= \frac{1}{y} \cdot \frac{dy}{dx}. \\ y &= \frac{dy}{dx}. \end{aligned}$$

Therefore, $\frac{dy}{dx} = e^x$, as claimed. ■

Example 4.21 (Natural logarithm and exponent)

1. Find $\frac{dy}{dx}$ if $y = e^{x^2 \ln(x)}$. Applying the Chain Rule and the Product Rule, we get

$$\frac{dy}{dx} = e^{x^2 \ln x} \cdot \left(2x \cdot \ln x + x^2 \cdot \frac{1}{x}\right) = e^{x^2 \ln x}(x + 2x \ln x).$$

2. Let $f(x) = xe^{\frac{x}{2}}$. Find where f is increasing, where it is decreasing, where it is concave up and concave down.

   We'll start by taking the first two derivatives of f and finding their stationary points.

$$\begin{aligned} f'(x) &= e^{\frac{x}{2}} + \frac{x}{2} e^{\frac{x}{2}} = e^{\frac{x}{2}} \left(\frac{x+2}{2}\right) \\ f''(x) &= \frac{e^{\frac{x}{2}}}{2} + \frac{e^{\frac{x}{2}}}{2} + \frac{x}{4} e^{\frac{x}{2}} = e^{\frac{x}{2}} \left(\frac{1}{2} + \frac{1}{2} + \frac{x}{4}\right) \end{aligned}$$

   Since $e^{\frac{x}{2}}$ is always positive, $f'(x)$ must be negative when $x < -2$, zero at $x = -2$, and positive when $x > -2$. Therefore, the function is decreasing on $(-\infty, -2)$ and increasing on $(-2, \infty)$, with a local minimum at $x = -2$.

   Similarly, the sign of $f''(x)$ only depends on that of $1 + \frac{x}{4}$. We have $f''(x) < 0$ for $x < -4$ and $f''(x) > 0$ for $x > -4$, so the function is concave down on $(-\infty, -4)$ and concave up on $(-4, \infty)$.

3. Evaluate $\int e^{-4x}\,dx$.

   Recall that $\frac{de^x}{dx} = e^x$. So we want a function such that $\frac{df(x)}{dx} = e^{-4x}$. Let $u = -4x$. Then $du = -4\,dx$ and $dx = \frac{-1}{4}\,du$. Making the substitutions, we have

$$\int e^{-4x}\,dx = -\frac{1}{4}\int e^u du = -\frac{1}{4}e^u + c = -\frac{1}{4}e^{-4x} + c.$$

4. Evaluate $\int x^2 e^{-x^3}\,dx$. Let $u = -x^3$. Then $du = -3x^2\,dx$ and $x^2\,dx = -\frac{1}{3}\,du$. Making the relevant substitutions, we see that:

$$\int x^2 e^{-x^3}\,dx = -\frac{1}{3}\int e^u du = -\frac{1}{3}e^u + c = -\frac{1}{3}e^{-x^3} + c.$$

We've talked about the natural log function and the exponential function e^x. But now what about a^x and its derivatives?

Theorem 4.32 *For $a > 0$ and any real number x, $a^x = e^{x \ln a}$.*

Note that as we've seen before with exponents, the usual laws apply when there is a variable in the exponent. The rules are summarized in the next theorem.

Theorem 4.33 *For any $a > 0$,*

1. $\frac{d(a^x)}{dx} = a^x \ln(a)$
2. $\int a^x\,dx = \frac{1}{\ln(a)}a^x + c$ *for* $a \neq 1$.

Proof. For statement 1, we use the definition above and apply the Chain Rule:

$$\begin{aligned} \frac{d(a^x)}{dx} &= \frac{d(e^{x\ln a})}{dx} \\ &= e^{x\ln a} \cdot \frac{d(x \ln a)}{dx} \\ &= e^{x\ln a} \cdot \ln a \\ &= a^x \ln a. \end{aligned}$$

We then have
$$\frac{d}{dx}\left(\frac{1}{\ln a}a^x + c\right) = \frac{1}{\ln a}(a^x \ln a) = a^x,$$
which proves statement 2. ■

Example 4.22 (Derivatives of exponential functions)

1. Find $\frac{d(3^{\sqrt{x}})}{dx}$.

   By the Chain Rule,
   $$\frac{d(3^{\sqrt{x}})}{dx} = 3^{\sqrt{x}}\ln 3 \cdot \frac{d(\sqrt{x})}{dx} = 3^{\sqrt{x}}\ln 3 \cdot \frac{1}{2}x^{-\frac{1}{2}} = \frac{3^{\sqrt{x}}\ln 3}{2\sqrt{x}}.$$

2. Find $\frac{dy}{dx}$ if $y = (x^4+2)^5 + 5^{x^4+2}$.
   $$\frac{dy}{dx} = 5(x^4+2)^4 \cdot 4x^3 + 5^{x^4+2}\ln(5) \cdot (4x^3).$$

3. Find $\int 2^{x^3} x^2\, dx$.

   We can do a simple substitution here: let $u = x^3$. Then $du = 3x^2\, dx$ and so $\frac{du}{3} = x^2\, dx$. Making the substitutions:
   $$\int 2^{x^3} x^2\, dx = \int \frac{2^u}{3}\, du = \frac{1}{3\ln 2}2^u = \frac{1}{3\ln 2}2^{x^3} + c$$

4.2.4 Techniques of Integration

Generally, differentiation of the sort of functions we run into is straightforward and usually quite simple. Anti-differentiation is not. There are three main techniques that people use to do this stuff. The first is trigonometric functions, which we'll disregard because they are rarely useful in political science applications. The second is substitutions—which we've already been doing.

Substitutions

Recall: let g be a differentiable function and let F be an antiderivative of f. Then if $u = g(x)$,

$$f(g(x))g'(x)\,dx = \int f(u)du = F(u) + c = F(g(x)) + c.$$

So basically, all you really want to do when you want to integrate a function by substitutions is the following:

1. Identify the part of the function to assign to u. Generally, you'll want this to differentiate so that you can replace the remaining parts of the function.
2. Differentiate u.
3. Substitute u in for x.
4. Integrate.
5. Substitute x back in for u.
6. Evaluate (if it is a definite integral).

Note that, as a check, you can always differentiate the result to see if you get back to the original integral, and also to practice your skills! Also note that for integrals that involve some sort of radical like $\sqrt[n]{ax+b}$, for example, making the substitution $u = \sqrt[n]{ax+b}$ will eliminate that radical and ease your pain.

Integration by Parts

When integration by substitution fails, it may be possible to use a kind of double substitution called integration by parts.

Consider first the simple indefinite integral. Let $u = u(x)$ and $v = v(x)$. Then

$$\frac{d(u(x)v(x))}{dx} = u(x)v'(x) + u'(x)v(x),$$

which implies

$$u(x)v'(x) = \frac{d(u(x)v(x))}{dx} - u'(x)v(x).$$

By integrating both sides of the equation we get

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int u'(x)v(x)\,dx.$$

Since $dv = v'(x)\,dx$ and $u'(x)\,dx = du$ we can write this in the following manner (this is how you might encounter it if you looked this up online or in a text):

$$\int u\,dv = uv - \int v\,du.$$

From this, it's quite a short hop to definite integrals:

$$\int_a^b u(x)v'(x)\,dx = u(x)v(x)|_{x=a}^b - \int_a^b v(x)u'(x)\,dx.$$

Example 4.23 (Integration by parts)

1. Find $\int_1^2 \ln(x)\,dx$. Let $u = \ln x$ and $dv = dx$. Then $du = \frac{1}{x}\,dx$ and $v = x$.

$$\int_1^2 \ln(x)\,dx = x\ln(x)|_{x=1}^2 - \int_1^2 x \cdot \frac{1}{x}\,dx = 2\ln(2) - 1$$

2. Find $\int xe^x\,dx$. Let $u = x$ and $dv = e^x\,dx$. Then $du = dx$ and $v = e^x$.

$$\int xe^x\,dx = xe^x - \int e^x\,dx = xe^x - e^x + c = (x-1)e^x + c$$

How do we choose which function is to be u and which function is to be dv? There is a simple rule of thumb that we can use: choose u to be whatever in your expression comes first in this list:

- I: inverse trigonometric functions $\arctan(x)$, $\arcsin(x)$, etc
- L: logarithmic functions $\ln(x)$
- A: algebraic functions x^2, $3x^{50}$, etc

- T: trigonometric functions $\sin(x)$, $\cos(x)$, etc
- E: Exponential functions e^x

Note that, in general, the functions lower on the list have easier antiderivatives than do functions higher on the list. Also, note that we will pretty much never see inverse trig functions and trig functions; they simply aren't that useful for what we do in political science.

Methods for Rational Functions

Recall that a rational function is simply the quotient of two polynomials. Let's begin by making an additional distinction.

Definition 4.13. A rational function is a *proper rational function* if the degree of the numerator is less than the degree of the denominator. □

An example is the rational function

$$f(x) = \frac{2x+2}{x^2-4x+8}.$$

Note that we can always write an improper rational function as a polynomial plus a proper rational function by simply doing long division.

Long divison. As an example, suppose that $f(x) = \frac{x^2-3x-10}{x+2}$.

$$\begin{array}{r|r} & x \quad -5 \\ \hline x+2 & x^2 - 3x - 10 \\ & -\ x^2 - 2x \quad\quad \\ \hline & -\ 5x - 10 \\ & 5x + 10 \\ \hline & 0 \end{array}$$

Therefore another way to write $f(x) = \frac{x^2-3x-10}{x+2}$ is $f(x) = x - 5$

As another example, suppose that $f(x) = \frac{x^5+2x^3-x+1}{x^3+5x}$.

$$
\begin{array}{rrrrr}
 & & x^2 & & -3 \\
\cline{2-5}
x^3+5x \big) & x^5 & +2x^3 & -x & +1 \\
 & -x^5 & -5x^3 & & \\
\cline{2-3}
 & & -3x^3 & -x & \\
 & & 3x^3 & +15x & \\
\cline{3-4}
 & & & 14x &
\end{array}
$$

Now we can't take this any further, since the degree of the polynomial is greater than our remainder, so we get

$$
f(x) = x^2 - 3 + \frac{14x+1}{x^3+5x}.
$$

This long division can often make integrals easier to work with.

Completing the square. This is a simple operation where we can replace an expression of the form $x^2 + bx$ with $(x + c)^2 + d$. This reduces any problem involving a quadratic polynomial to one involving a square quadratic polynomial and a constant.

For example, we have

$$
x^2 - 4x + 8 = x^2 - 4x + 4 + 4 = (x-2)^2 + 4.
$$

In general, we won't use this too much, since it often results in having to integrate and get trigonometric functions.

Partial fraction decomposition. Adding fractions is a standard part of our arsenal. This is a process whereby we go the opposite way — we undo fraction addition. There are obviously an infinite variety of possible partial fraction decompositions; we'll go through a few examples here.

1. **Distinct linear factors:** these will be things that we can break down into polynomials of degree 1. Consider the following example:

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

So we require A and B so that the following relationship holds:

$$\begin{aligned}\frac{A}{x-3}+\frac{B}{x+2} &= \frac{3x-1}{x^2-x-6}\\ 3x-1 &= A(x+2)+B(x-3)\\ 3x-1 &= Ax+2A+Bx-3B\\ 3x-1 &= (A+B)x+2A-3B\end{aligned}$$

So we see that it must be the case that $A+B=3$ and then $2A-3B=-1$. We now need only solve this little system of equations. The first implies $A=3-B$, so we have

$$2(3-B)-3B=-1 \quad \Leftrightarrow \quad 6-5B=-1 \quad \Rightarrow \quad B=\frac{7}{5},$$

which in turn implies $A=\frac{8}{5}$.

Thus our partial fraction decomposition yields

$$\frac{3x-1}{x^2-x-6}=\frac{\frac{8}{5}}{x-3}+\frac{\frac{7}{5}}{x+2}.$$

Now consider the problem of trying to integrate this function,

$$\int \frac{3x-1}{x^2-x-6}\,dx.$$

We might try substitution with $u=x^2-x-6$, but then $du=(2x-1)\,dx$. This won't work — we cannot make du equal to the numerator in our expression. But, as we have just seen, we can rewrite the problem as

$$\begin{aligned}\int \frac{\frac{8}{5}}{x-3}+\frac{\frac{7}{5}}{x+2}\,dx &= \int \frac{\frac{8}{5}}{x-3}\,dx+\int \frac{\frac{7}{5}}{x+2}\,dx\\ &= \frac{8}{5}\ln|x-3|+\frac{7}{5}\ln|x+2|+c.\end{aligned}$$

The actual integration step is easy.

As another example, find $\int \frac{5x+3}{x^3-2x^2-3x}\,dx$. Obviously we'll be using partial fraction decomposition. The denominator factors into

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

Then we require A, B, and C so that

$$\frac{5x+3}{x^3-2x^2-3x} = \frac{A}{x} + \frac{B}{x-3} + \frac{C}{x+1}.$$

This gives us the condition

$$5x+3 = A(x-3)(x+1) + B(x+1)x + C(x-3)x.$$

This time, instead of multiplying out and combining the coefficients, let's simply substitute in some values for x. In particular, we'll pick values so that two of the three terms drop out, and we can solve for the last.

- Setting $x = 0$, we get $3 = -3A$, so $A = -1$.
- Setting $x = -1$, we get $-2 = 4C$, so $C = \frac{-1}{2}$.
- Setting $x = 3$, we get $18 = 12B$, so $B = \frac{3}{2}$.

Going back to the integral, we have

$$\begin{aligned}\int \frac{5x+3}{x^3-2x^2-3x}\,dx &= \int \frac{-1}{x}\,dx + \int \frac{-\frac{1}{2}}{x-3}\,dx + \int \frac{\frac{3}{2}}{x+1}\,dx \\ &= -\ln|x| - \frac{1}{2}\ln|x-3| + \frac{3}{2}\ln|x+1| + c.\end{aligned}$$

2. **Repeated linear factors:** Consider the problem of finding $\int \frac{x}{(x-3)^2}\,dx$. Again, we'll use a partial fraction decomposition to solve this problem. This time, the decomposition shall take the form:

$$\frac{x}{(x-3)^2} = \frac{A}{x-3} + \frac{B}{(x-3)^2}$$

Why the square in the second term? Its easier to deal with in the following step:

$$\begin{aligned}x &= A(x-3) + B.\\ x &= Ax - 3A + B.\end{aligned}$$

Substituting in $x = 3$, we see that $B = 3$. Then, substituting in $x = 0$, we get $3A = B$, meaning $A = 1$. Therefore, our integral becomes

$$\begin{aligned}\int \frac{x}{(x-3)^2}\,dx &= \int \frac{1}{x-3}\,dx + \int \frac{3}{(x-3)^2}\,dx \\ &= \ln|x-3| - \frac{3}{x-3} + c.\end{aligned}$$

Note that the general rule for decomposing functions with repeated linear factors in the denominator is that for each factor $(ax+b)^k$ in the denominator there are k terms in the partial fraction decomposition:

$$\frac{A_1}{ax+b} + \frac{A_2}{(ax+b)^2} + \frac{A_3}{(ax+b)^3} + \ldots + \frac{A_k}{(ax+b)^k}.$$

Here's an example to do yourselves: find $\int \frac{3x^2-8x+13}{(x+3)(x-1)^2}\,dx$. The decomposition is

$$\begin{aligned}\frac{3x^2-8x+13}{(x+3)(x-1)^2} &= \frac{A}{x+3} + \frac{B}{x-1} + \frac{C}{(x-1)^2} \\ 3x^2-8x+13 &= A(x-1)^2 + B(x+3)(x-1) + C(x+3).\end{aligned}$$

Solve for A, B, and C by substitutions:

- Setting $x = 1$, we get $4C = 8$, so $C = 2$.
- Setting $x = -3$, we get $16A = 64$, so $A = 4$.
- Setting $x = 0$, we get $A - 3B + 3C = 4 - 3B + 6 = 13$, so $B = -1$.

Finally, the integral is

$$\int \frac{4}{x+3}\,dx + \int \frac{2}{(x-1)^2}\,dx - \int \frac{dx}{x-1} = 4\ln|x+3| - \ln|x-1| + 2\frac{1}{x-1} + c.$$

3. **Single quadratic factor:** we use this when we find it impossible to factorize the denominator into a linear factor without introducing complex numbers. For example, consider

$$\int \frac{6x^2-3x+1}{(4x+1)(x^2+1)}\,dx.$$

In this case, the appropriate form of the decomposition is

$$\begin{aligned}\frac{6x^2-3x+1}{(4x+1)(x^2+1)} &= \frac{A}{4x+1}+\frac{Bx+C}{x^2+1}\\ 6x^2-3x+1 &= A(x^2+1)+(Bx+C)(4x+1)\\ &= Ax^2+A+4Bx^2+Bx+4Cx+C\\ &= (A+4B)x^2+(B+4C)x+A+C\end{aligned}$$

We thus have the system

$$\begin{aligned}6 &= A+4B\\ -3 &= B+4C\\ 1 &= A+C\end{aligned}$$

We can subtract the third from the first, and multiply the second by -4, to get:

$$\begin{aligned}5 &= 4B-C\\ 12 &= -4B-16C\end{aligned}$$

We thus have $17=-17C$, so $C=-1$. We then have $A=1-C=2$ and $B=-3-4C=1$. The original integral becomes

$$\int \frac{2}{4x+1}+\frac{x-1}{x^2+1}\,dx.$$

You can integrate this, though you'll find one of the integrals results in a trigonometric function.

4.2.5 Improper Integrals

This is our last topic before we begin with multivariate calculus. When we defined definite integrals,

$$\int_a^b f(x)\,dx,$$

we implicitly assumed that both a and b were finite. But how do we deal with things like

$$\int_{-\infty}^{\infty} f(x)\,dx?$$

We'll start by defining such an integral more formally.

Definition 4.14. An *improper integral* takes one of the following forms:

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

If the limit exists and has a finite value, we say that the corresponding improper integral converges and has this value. Otherwise, the integral is said to diverge. □

Example 4.24 (Infinite limits of integration)
Find if possible $\int_{-\infty}^{-1} xe^{-x^2}\,dx$.

$$\begin{aligned}
\int_{-\infty}^{-1} xe^{-x^2}\,dx &= -\frac{1}{2}\int_{-\infty}^{-1} -2xe^{-x^2}\,dx \\
&= \lim_{a\to-\infty}\frac{-1}{2}\int_{a}^{-1} -2xe^{-x^2} \\
&= \lim_{a\to-\infty}\frac{-1}{2}e^{-x^2}\Big|_{x=a}^{-1} \\
&= \lim_{a\to-\infty}\left(\frac{-1}{2e}+\frac{1}{2e^{a^2}}\right) = \frac{-1}{2e}.
\end{aligned}$$

Definition 4.15. If there is a finite c such that $\int_{-\infty}^{c} f(x)\,dx$ and $\int_{c}^{\infty} f(x)\,dx$ converge, then $\int_{-\infty}^{\infty} f(x)\,dx$ converges and has value

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{c} f(x)\,dx + \int_{c}^{\infty} f(x)\,dx.$$ □

Finally, what will happen when we have an infinite integrand? We might end up with things that look really simple, but in fact turn out not to be so. Consider the following example:

$$\int_{-2}^{1}\frac{1}{x^2}\,dx = \frac{-1}{x}|_{-2}^{1} = \frac{-1}{1} - \frac{-1}{-2} = -\frac{3}{2}$$

This doesn't seem right if you look at the graph of the function, which has a vertical asymptote at $x = 0$. For a function to be integrable in the standard sense—that is, for us to treat it as we just did—the function must be bounded.

There are two possibilities that we shall have to deal with.

Infinite at an endpoint.

Definition 4.16. Let f be continuous on the half-open interval $[a, b)$ and suppose that $\lim_{x\to b-} |f(x)| = \infty$. Then

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

provided that this limit exists and is finite, in which case we say that the integral converges. Otherwise, the integral diverges. (The definition when f approaches infinity at a is analogous.) □

Example 4.25 (Integral when f is infinite at an endpoint)

1. Evaluate, if possible, $\int_0^{16} \frac{1}{\sqrt[4]{x}}\,dx$.

$$\begin{aligned}\int_0^{16} \frac{1}{\sqrt[4]{x}}\,dx &= \lim_{t\to 0+}\int_t^{16} x^{\frac{-1}{4}}\,dx \\ &= \lim_{t\to 0+} \frac{4}{3}x^{\frac{3}{4}}\bigg|_{x=t}^{16} \\ &= \lim_{t\to 0+} \frac{4}{3}\left((16)^{\frac{3}{4}} - t^{\frac{3}{4}}\right) \\ &= \lim_{to\to 0+}\left(\frac{32}{3} - \frac{4}{3}t^{\frac{3}{4}}\right) = \frac{32}{3}.\end{aligned}$$

2. Evaluate, if possible, $\int_0^1 \frac{1}{x}\,dx$.

$$\int_0^1 \frac{1}{x}\,dx = \lim_{t\to 0+}\int_t^1 \frac{1}{x}\,dx = \lim_{t\to 0+} \ln|x|\big|_t^1 = \lim_{t\to 0+} \ln 1 - \ln t = \lim_{t\to 0+} -\ln t = \infty.$$

The integral, in this case, diverges.

Infinite at an interior point. This was the case we originally considered.

Definition 4.17. Let f be continuous on $[a, b]$ except at a number c, where $a < c < b$, and suppose that $\lim_{x \to c} |f(x)| = \infty$. Then

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

provided that both integrals on the right hand side converge. Otherwise we say that $\int_a^b f(x)\,dx$ diverges. □

Example 4.26 (Integral when f is infinite an at interior point) Evaluate if possible the improper interval $\int_0^3 \frac{1}{(x-1)^{\frac{2}{3}}}\,dx$. The first thing to observe is that this integral tends toward infinity at $x = 1$. Thus, we break up the interval over which we are integrating into two parts, as suggested by our definition:

$$\begin{aligned}
\int_0^3 \frac{1}{(x-1)^{\frac{2}{3}}}\,dx &= \int_0^1 \frac{1}{(x-1)^{\frac{2}{3}}}\,dx + \int_1^3 \frac{1}{(x-1)^{\frac{2}{3}}}\,dx \\
&= \lim_{t \to 1-} \int_0^t \frac{1}{(x-1)^{\frac{2}{3}}}\,dx + \lim_{s \to 1+} \int_s^3 \frac{1}{(x-1)^{\frac{2}{3}}}\,dx \\
&= \lim_{t \to 1-} 3(x-1)^{\frac{1}{3}}\Big|_{x=0}^{t} + \lim_{s \to 1+} 3(x-1)^{\frac{1}{3}}\Big|_{x=s}^{3} \\
&= 3 \cdot \lim_{t \to 1-} \left[(t-1)^{\frac{1}{3}} - (0-1)^{\frac{1}{3}}\right] + 3 \cdot \lim_{s \to 1+} \left[(2)^{\frac{1}{3}} - (s-1)^{\frac{1}{3}}\right] \\
&= 3 + 3 \cdot 2^{\frac{1}{3}}.
\end{aligned}$$

Chapter 5

Multivariate Calculus

Up to this point we've talked entirely about calculus as applied to functions in the Cartesian plane, with just an x and y axis. This is just a special case of a more general n-space, of which we've restricted our attention to only two dimensions. Our task now is to generalize our notions of derivatives and integrals to n-space. Before we can do that, we need to talk about vectors.

5.1 Vectors

There are some things we talk about that are simply scalars. Speed, length, mass, etc., are easily represented as numbers (magnitude). But there are other things—like velocity and force—that require both a magnitude and a direction. We call these vectors.

Generally, we might represent these as arrows emanating from the origin: the length of the arrow corresponds to the magnitude of the vector, and the direction of the arrow is its direction. The vector is set to originate at the origin, its magnitude and direction are determined uniquely by choosing the coordinates in the Cartesian plane associated with the end of the vector. We write this as $\vec{u} = (u_1, u_2)$. (For simplicity, we will be focusing on two-dimensional vectors, but all of the concepts introduced here generalize straightforwardly to n-dimensional vectors.)

Operations on vectors: Let $\vec{u} = (u_1, u_2)$ and $\vec{v} = (v_1, v_2)$ be vectors.

1. Vector equality: $\vec{u} = \vec{v}$ if $u_1 = v_1$ and $u_2 = v_2$.

2. Vector addition: $\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2)$

3. Scalar multiplication: for a scalar c, $c\vec{u} = (cu_1, cu_2)$.
4. Magnitude of vectors: The length of $\vec{u}$ is given by $\|\vec{u}\| = \sqrt{u_1^2 + u_2^2}$.
5. Dot product: Given two vectors, $\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2$.

   There is a second formula for the dot product: $\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos(\theta)$, where θ is the angle between the two vectors. Note that this implies that two vectors will be perpendicular to one another if and only if $\vec{u} \cdot \vec{v} = 0$.

Theorem 5.1 *For any vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$ and any scalars a and b the following hold:*

1. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
2. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
3. $\vec{u} + \vec{0} = \vec{0} + \vec{u} = \vec{u}$
4. $\vec{u} + (-\vec{u}) = \vec{0}$
5. $a(b(\vec{u})) = (ab)\vec{u} = \vec{u}(ab)$
6. $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$
7. $(a + b)\vec{u} = a\vec{u} + b\vec{u}$
8. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
9. $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$
10. $\|a\vec{u}\| = |a| \|\vec{u}\|$
11. $a(\vec{u} \cdot \vec{v}) = (a\vec{u}) \cdot \vec{v} = \vec{u} \cdot (a\vec{v})$
12. $\vec{0} \cdot \vec{u} = 0$
13. $\vec{u} \cdot \vec{u} = \|\vec{u}\|^2$

Example 5.1 (Vectors)

1. Suppose $\vec{u} = (4, -3)$. Compute $\|\vec{u}\|$ and $\|-2\vec{u}\|$.

- $\|\vec{u}\| = \sqrt{4^2 + (-3)^2} = \sqrt{16+9} = \sqrt{25} = 5$
- $\|-2\vec{u}\| = |-2|\|\vec{u}\| = 2\sqrt{25} = 10$

2. Find b so that $(8,6)$ and $(3,b)$ are orthogonal.

$$\begin{aligned}(8,6)\cdot(3,b) &= 0\\ 24 + 6b &= 0\\ b &= -4\end{aligned}$$

5.2 Differential Calculus

In this section we'll be considering the derivatives of functions in n-space. The basic idea is to take the derivative of the function with respect to a single variable, using our typical methods, while holding all other variables fixed (treating them as if they were constants).

5.2.1 Partial Derivatives

Definition 5.1. The *partial derivative* of a function of n variables at $(x_1^0, x_2^0, \ldots, x_n^0)$ with respect to x_i is

$$\frac{\partial f(x_1^0, x_2^0, \ldots, x_n^0)}{\partial x_i} = \lim_{h\to 0} \frac{f(x_1^0, x_2^0, \ldots, x_i^0 - h, \ldots, x_n^0) - f(x_1^0, x_2^0, \ldots, x_n^0)}{h}. \qquad \square$$

So the partial derivative is just defined in terms of a one-dimensional limit, which we already know how to deal with. Let's do a couple of examples to illustrate.

Example 5.2 (Partial derivatives)

1. Suppose $f(x, y) = x^2y + 3y^3$. Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. To do this, simply differentiate with respect to one variable while treating the other as though it were simply a constant:

$$\frac{\partial f(x, y)}{\partial x} = 2xy.$$
$$\frac{\partial f(x, y)}{\partial y} = x^2 + 9y^2.$$

2. Consider the volume of a cone given by $V = \frac{r^2 h\pi}{3}$ where r is the radius and h is the height. Find the rate of change of the cone's volume as its radius varies and its height remains constant.

   To do this, simply take the partial derivative with respect to r:

$$\frac{\partial V}{\partial r} = \frac{2rh\pi}{3}.$$

Since a partial derivative of a function of x and y is another function of those same two variables, it can be differentiated yet again in the same way, either with respect to x or with respect to y. Thus, in the case of two variables, each first derivative has two possible second derivatives:

$$\frac{\partial}{\partial x}\left(\frac{\partial f(x, y)}{\partial x}\right) = \frac{\partial^2 f(x, y)}{\partial x^2}$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f(x, y)}{\partial x}\right) = \frac{\partial^2 f(x, y)}{\partial y \partial x}$$

When we've differentiated with respect to multiple different variables, we call that a mixed partial derivative.

Example 5.3 (Second partial derivatives)
Find all the second partial derivatives of $f(x, y) = xe^y + x^3y^2$.

$$\frac{\partial f}{\partial x} = e^y + 3x^2y^2$$
$$\frac{\partial^2 f}{\partial x^2} = 6xy^2$$
$$\frac{\partial^2 f}{\partial y \partial x} = e^y + 6x^2y$$

$$\frac{\partial f}{\partial y} = xe^y + 2x^3y$$
$$\frac{\partial^2 f}{\partial y^2} = xe^y + 2x^3$$
$$\frac{\partial^2 f}{\partial x \partial y} = e^y + 6x^2y$$

Now let's extend our notions of limits and continuity to n-dimensional space.

Definition 5.2. Let $\vec{x} = (x_1, \ldots, x_n)$ be an n-dimensional variable and $\vec{a} = (a_1, \ldots, a_n)$ be a constant vector. To say that $\lim_{\vec{x} \to \vec{a}} f(x_1, \ldots, x_n) = L$ means that for each $\epsilon > 0$, there is a corresponding $\delta > 0$ such that $\|f(x_1, \ldots, x_n) - L\| < \epsilon$ when $0 < \|\vec{x} - \vec{a}\| < \delta$. □

There are 3 things to note about this definition of limits in n-space:

1. The path by which we approach $\vec{a}$ is totally irrelevant. Thus, if different paths of approach lead to different values of L, the limit does not exist.

2. The actual behavior of the function at $\vec{a}$ is irrelevant; the function does not even have to be defined at $\vec{a}$. It only matters what happens as we get arbitrarily close to $\vec{a}$.

3. All of our previous limit results hold. So if we run into problems with new n-space limits, we can refer back to our previous results.

Example 5.4 (Limits in n-space)
Find $\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$ or show that it doesn't exist. We have

$$\lim_{(x,0)\to(0,0)} \frac{x^2-0}{x^2+0} = \lim_{(x,0)\to(0,0)} 1 = 1$$
$$\lim_{(0,y)\to(0,0)} \frac{0-y^2}{0+y^2} = \lim_{(0,y)\to(0,0)} -1 = -1$$

And so the limit does not exist.

Having defined limits in n-space, we can now consider the continuity of functions.

Definition 5.3. We say that $f(x_1, \ldots, x_n)$ is continuous at a point $\vec{a} = (a_1, \ldots, a_n)$ if

1. f has a value at $\vec{a}$
2. f has a limit at $\vec{a}$
3. $\lim_{\vec{x}\to\vec{a}} f(x_1, \ldots, x_n) = f(a_1, \ldots, a_n)$ □

Intuitively, this is essentially the same as before: no jumps, no wild fluctuations in the value of the function, and no unbounded behavior.

As in the case of functions of just one variable, we still have that polynomials are continuous everywhere, rational functions are continuous everywhere that the denominator is not zero, and the composition of two continuous functions is continuous.

The next result gives a condition for symmetry of the mixed partial derivatives.

Theorem 5.2 (Young's Theorem) *If $\frac{\partial^2 f}{\partial x\,\partial y}$ and $\frac{\partial^2 f}{\partial y\,\partial x}$ exist and are continuous on an open set S, then at each point of that set S,*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

The theorem is stated just for the two-dimensional case, but its generalization to higher-order partial derivatives is straightforward.

5.2.2 The Gradient

Recall that when we were discussing functions of a single variable, differentiability of f at x meant the existence of the derivative $f'(x)$, which was equivalent to the graph having a tangent line at x. How can we generalize the notion of differentiability? What plays the role of the derivative for a function of two or more variables? It can't be the derivative itself, because there is more than one of them.

Another way that we can think of a 1-variable derivative is that if f is differentiable at a point c, there exists a tangent line that approximates the function for the values of x near c. We can write this formally for the 2-dimensional case as follows.

Definition 5.4. Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. We say that f is *locally linear* at (c_1, c_2) if $f(c_1 + h_1, c_2 + h_2) = f(c_1, c_2) + h_1 \frac{\partial f(c_1,c_2)}{\partial x_1} + h_2 \frac{\partial f(c_1,c_2)}{\partial x_2} + \epsilon(h_1, h_2)$ where $\epsilon(h_1, h_2) \to 0$ as $(h_1, h_2) \to 0$. □

We can rewrite this in vector notation: let $\vec{c} = (c_1, c_2)$, $\vec{h} = (h_1, h_2)$ and $\epsilon(\vec{h}) = \epsilon(h_1, h_2)$. Then this simplifies to:

$$f(\vec{c} + \vec{h}) = f(\vec{c}) + \left(\frac{\partial f(\vec{c})}{\partial x_1}, \frac{\partial f(\vec{c})}{\partial x_2} \right) \cdot \vec{h} + \epsilon(\vec{h}).$$

We have a special name for the vector $(\frac{\partial f(\vec{c})}{\partial x_1}, \frac{\partial f(\vec{c})}{\partial x_2})$ given above. We write this as $\nabla f(\vec{c})$ and call it the gradient of f at $\vec{c}$.

Definition 5.5. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\vec{c}$. The *gradient* of f at $\vec{c}$, written $\nabla f(\vec{c})$, is the vector of partial deriviatives of f evaluated at $\vec{c}$,

$$\nabla f(\vec{c}) = \left(\frac{\partial f(\vec{c})}{\partial x_1}, \frac{\partial f(\vec{c})}{\partial x_2}, \dots, \frac{\partial f(\vec{c})}{\partial x_n} \right).$$

□

Now we can define differentiability in terms of local linearity for a function of n variables.

Definition 5.6. A function f is differentiable at $\vec{c}$ if and only if it is locally linear at $\vec{c}$; i.e., if

$$f(\vec{c} + \vec{h}) = f(\vec{c}) + \nabla f(\vec{c}) \cdot \vec{h} + \epsilon(\vec{h}),$$

where $\epsilon(\vec{h}) \to 0$ as $\vec{h} \to 0$. □

In practice we do not typically need to check this condition, thanks to the following result.

Theorem 5.3 *If $f(\vec{c})$ has continuous partial derivatives $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}$ on a set $S \subseteq \mathbb{R}^n$ whose interior contains $\vec{c}$, then f is differentiable at $\vec{c}$.*

In other words, in order to have differentiability at a point, all we need are continuous partial derivatives at that point.

Example 5.5 (Gradients and differentiability)
Calculate the gradient of $f(x, y) = xe^y + x^2y$ and show that it is differentiable everywhere. We have

$$\frac{\partial f}{\partial x} = e^y + 2xy$$
$$\frac{\partial f}{\partial y} = xe^y + x^2,$$

Note that these are both continuous everywhere and so by the theorem given above, $f(x, y)$ is differentiable everywhere. The gradient is simply the vector formed by these partial derivatives:

$$\nabla f(x, y) = (e^y + 2xy, xe^y + x^2).$$

Properties of the gradient:

1. $\nabla(f(\vec{c}) + g(\vec{c})) = \nabla f(\vec{c}) + \nabla g(\vec{c})$
2. $\nabla(\alpha f(\vec{c})) = \alpha \nabla f(\vec{c})$
3. $\nabla(f(\vec{c})g(\vec{c})) = f(\vec{c})\nabla g(\vec{c}) + g(\vec{c})\nabla f(\vec{c})$

We also can establish that the relationship we saw between differentiability and continuity in the one-dimensional case is also true in n dimensions.

Theorem 5.4 *If f is differentiable at $\vec{c}$, then f is continuous at $\vec{c}$.*

The same caveat applies as before: all differentiable functions are continuous, but not all continuous functions are differentiable.

5.2.3 Directional Derivatives

The partial derivatives tell us the rate of change in $f(x, y)$ as we move along the x or y axis. What if we wanted to consider the rate of change in some other direction?

Definition 5.7. A *unit vector* is a vector $\vec{u}$ such that $\|\vec{u}\| = 1$. □

Note that if $\vec{p}$ is non-zero, then $\vec{u} = \frac{1}{\|\vec{p}\|}\vec{p}$ is a unit vector in the same direction as $\vec{p}$.

Definition 5.8. For any unit vector $\vec{u}$, consider the limit

$$D_{\vec{u}} f(\vec{c}) = \lim_{h \to 0} \frac{f(\vec{c} + h\vec{u}) - f(\vec{c})}{\vec{h}}.$$

If this limit exists and is finite, we call it the *directional derivative* of f at $\vec{c}$ in direction $\vec{u}$. □

The directional derivative has a very useful connection with the gradient—one that lets us avoid computing this limit explicitly.

Theorem 5.5 *Let f be differentiable at $\vec{c}$. Then f has a directional derivative at $\vec{c}$ in the direction of the unit vector $\vec{u}$ and*

$$D_{\vec{u}} f(\vec{c}) = \vec{u} \cdot \nabla f(\vec{c}).$$

Example 5.6 (Directional derivatives)

1. If $f(x, y) = 4x^2 - xy + 4y^2$, find the directional derivative of f at $(-2, 1)$ in the direction of $(4, 3)$.

   First, we compute the gradient:

   $$\nabla f(x, y) = (8x - y, 8y - x) = (-17, 10).$$

   Before computing the directional derivative, we need to convert the direction into a unit vector,

   $$\vec{u} = \frac{1}{\|\vec{c}\|}\vec{c} = \frac{1}{\sqrt{4^2 + 3^2}}(4, 3) = \left(\frac{4}{5}, \frac{3}{5}\right).$$

Therefore, we have

$$D_{\vec{u}}f(-2,1) = \left(\frac{4}{5}, \frac{3}{5}\right) \cdot (-17, 10) = \frac{-68}{5} + \frac{30}{5} = \frac{-38}{5}.$$

2. Find the directional derivative of $f(x,y,z,t) = xyzt$ at $(1,1,-3,5)$ in the direction $(1,2,2,4)$.

   The gradient is

$$\nabla f(x,y,z,t) = (yzt, xzt, xyt, xyz) = (-15, -15, 5, -3).$$

   The unit vector for the direction is

$$\vec{u} = \frac{1}{\|\vec{c}\|}\vec{c} = \frac{1}{\sqrt{1^2+2^2+2^2+4^2}}(1,2,2,4) = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, \frac{4}{5}\right).$$

   We then compute the directional derivative as

$$\begin{aligned} D_{\vec{u}}f(1,1,-3,5) &= \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, \frac{4}{5}\right) \cdot (-15,-15,5,-3) \\ &= -3 - 6 + 2 - \frac{12}{5} = \frac{-47}{5}. \end{aligned}$$

Theorem 5.6 (Direction of Steepest Ascent) *A function increases most rapidly at $\vec{c}$ in the direction of the gradient with rate $\|\nabla f(\vec{c})\|$ and decreases most rapidly in the opposite direction with rate $-\|\nabla f(\vec{c})\|$.*

Example 5.7 (Direction of steepest ascent)
Suppose we are at a point $(1,1,0)$ on a paraboliod defined by $f(x,y) = y^2 - x^2$. What direction should we move if we are looking for the steepest climb and what will be the slope as we begin? The gradient is

$$\nabla f(x,y) = (-2x, 2y) = (-2, 2).$$

So we should move in direction $(-2,2)$ for the steepest climb. This will be of slope $\|\nabla f(\vec{c})\| = \sqrt{(-2)^2+2^2} = \sqrt{8}$.

5.2.4 The Chain Rule Revisited

Remember that the chain rule was a means of differentiating composite functions. This too can be generalized to functions of multiple variables.

Theorem 5.7 (Chain Rule, Part 1) *Let $x = x(t)$ and $y = y(t)$ both be differentiable at t, and let $z = f(x, y)$ be differentiable at $(x(t), y(t))$. Then $z = f(x(t), y(t))$ is differentiable at t and*

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

In this case, $\frac{dz}{dt}$ is also sometimes called the *total derivative*.

Example 5.8 (Total derivative)

1. Suppose $z = x^3 y$ where $x = 2t$ and $y = t^2$, and find $\frac{dz}{dt}$. We have

$$\begin{aligned}\frac{dz}{dt} &= 3x^2 y \cdot 2 + x^3 \cdot 2t \\ &= 3(2t)^2 \cdot 2t^2 + (2t)^3 \cdot 2t \\ &= 24t^4 + 16t^4 = 40t^4.\end{aligned}$$

   Now you might be thinking this is lame — we didn't really need the chain rule for that, since a simple substitution would have turned this into a one-variable problem. But the next example is not trivial.

2. As a right solid cylinder is heated, its radius r and height h increase; thus, so does its surface area:

$$S = 2\pi rh + 2\pi r^2.$$

   Suppose that at the instant when the radius $r = 10$ and $h = 100$, r is increasing at 0.2 units per hour, while h is increasing at 0.5 units per hour. How fast is surface area increasing at this instant?

   We want to find $\frac{dS}{dt}$. The partial deriatives of S are

$$\begin{aligned}\frac{\partial S}{\partial r} &= 2\pi h + 4\pi r. \\ \frac{\partial S}{\partial h} &= 2\pi r.\end{aligned}$$

Therefore, using the Chain Rule, we have

$$\begin{aligned}\frac{dS}{dt} &= \frac{\partial S}{\partial r}\frac{dr}{dt} + \frac{\partial S}{\partial h}\frac{dh}{dt} \\ &= (2\pi h + 4\pi r)0.2 + (2\pi r)0.5 \\ &= (200\pi + 40\pi)0.2 + (20\pi)0.5 \\ &= 58\pi.\end{aligned}$$

That is, surface area is increasing at 58π squared units per hour.

We can also state the Chain Rule for the case where the component functions are themselves multivariate.

Theorem 5.8 (Chain Rule, Part 2) *Let $x = x(s,t)$ and $y = y(s,t)$ have first partial derivatives at (s,t) and let $z = f(x,y)$ be differentiable at $(x(s,t), y(s,t))$. Then $z = f(x(s,t), y(s,t))$ has first derivatives given by*

$$\begin{aligned}\frac{dz}{ds} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}. \\ \frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.\end{aligned}$$

Example 5.9 (Total derivative with multivariate components) Let $z = 3x^2 - y^2$ where $x = 2s + 7t$ and $y = 5st$, and find $\frac{dz}{dt}$ and $\frac{dz}{ds}$ expressed in terms of s and t. The gradient of each function is as follows:

$$\begin{aligned}\nabla f(x,y) &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (6x, -2y) \\ \nabla x(s,t) &= \left(\frac{\partial x}{\partial s}, \frac{\partial x}{\partial t}\right) = (2, 7) \\ \nabla y(s,t) &= \left(\frac{\partial y}{\partial s}, \frac{\partial y}{\partial t}\right) = (5t, 5s).\end{aligned}$$

Therefore, using the Chain Rule, the first derivatives of $z = f(x(s,t), y(s,t))$ are

$$\begin{aligned}\frac{dz}{ds} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}\\ &= 6x(2) - 2y(5t)\\ &= 12(2s+7t) - 2(5st)(5t)\\ &= 24s + 84t - 50st^2.\end{aligned}$$

and

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}\\ &= 6x(7) - 2y(5s)\\ &= 42(2s+7t) - 2(5st)(5s)\\ &= 84s + 294t - 50s^2t.\end{aligned}$$

5.2.5 Optimization

We'll now look at using the derivative in n dimensions for the same purposes as in one dimension: finding the maxima and minima of functions. In all that follows, let $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{x}_0 = (x_1^0, x_2^0, \ldots, x_n^0)$.

Definition 5.9. A *neighborhood* of a point $\vec{c} \in \mathbb{R}^n$ is a set of the form $\{\vec{x} \in \mathbb{R}^n \mid \|\vec{x} - \vec{c}\| < \epsilon\}$ for some $\epsilon > 0$. □

Note that in the one-dimensional case, a neighborhood is simply an interval of the form $(c - \epsilon, c + \epsilon)$.

Definition 5.10. Consider a function $f : S \to \mathbb{R}$ where $S \subseteq \mathbb{R}^n$, and let $\vec{x}_0$ be any point in S.

- $f(\vec{x}_0)$ is a *global maximum value* of f on S if $f(\vec{x}_0) \geq f(\vec{x})$ for all $\vec{x} \in S$.
- $f(\vec{x}_0)$ is a *local maximum value* of f on S if there exists a set $U \subseteq S$, where U is a neighborhood of $\vec{x}_0$, such that $f(\vec{x}_0)$ is a global maximum value of f on U.

The definitions of global and local minimum values are analogous. □

Our result from the one-dimensional case on the existence of maxima and minima also holds up in this context.

Theorem 5.9 *If f is continuous on a closed, bounded set S, then f attains both a global maximum value and a global minimum value on S.*

So where do our extreme values occur in multidimensional space? It ends up being analogous to what we saw before. The critical points of a function f on S are again of three types:

1. Boundary points of S
2. Stationary points: we call $\vec{x}_0$ a stationary point if $\vec{x}_0$ is an interior point of S where f is differentiable and $\nabla f(\vec{x}_0) = \vec{0}$.
3. Singular points: we call $\vec{x}_0$ a singular point if $\vec{x}_0$ is an interior point of S where f is not differentiable.

Theorem 5.10 (Critical Point Theorem) *Let f be defined on a set S containing $\vec{x}_0$. If $f(\vec{x}_0)$ is an extreme value, then $\vec{x}_0$ must be a critical point:*

1. *a boundary point of S*
2. *a stationary point of S*
3. *a singular point of S*

Example 5.10 (Critical points in $\mathbb{R}^n$)

1. Find the extreme points of $f(x, y) = x^2 - 2x + \frac{y^2}{4}$. This is a polynomial, so there are no boundary points or singular points. Thus, we only have to look for stationary points. The gradient is

$$\nabla f(x, y) = (2x - 2, \tfrac{1}{2}y),$$

so the sole stationary point is $(x, y) = (1, 0)$. Now we have to determine if it is a maximum or a minimum. We have $f(1, 0) = -1$. Note that $x^2 - 2x \geq -1$ for any x and $y^2/4 \geq 0$ for any y. Therefore, we have found a minimum.

2. Find the extreme values of $f(x,y) = \frac{-x^2}{a^2} + \frac{y^2}{b^2}$. Again, no boundary points or singular points, so we only need to look for stationary points. The gradient is

$$\nabla f(x,y) = \left(\frac{-2x}{a^2}, \frac{2y}{b^2}\right),$$

so obviously the only critical point is $(x,y) = (0,0)$.

Now for the tricky part: is this a maximum or a minimum? It is neither: instead it is a saddle point, which is a maximum along one dimension and a minimum along another. So when we speak about the overall function, we get neither a maximum nor a minimum.

The second example illustrates something a little troubling about our extreme points test, namely that for the gradient of a function to be zero at a given point it is necessary for extrema but not sufficient.

Theorem 5.11 (Second Partials Test) *Suppose that $f(x,y)$ has continuous second partial derivatives in a neighborhood of (x_0, y_0) and that $\nabla f(x_0, y_0) = \vec{0}$. Let*

$$D(x_0,y_0) = \frac{\partial^2 f(x_0,y_0)}{\partial x^2}\frac{\partial^2 f(x_0,y_0)}{\partial y^2} - \left(\frac{\partial^2 f(x_0,y_0)}{\partial y \partial x}\right)^2.$$

Then:

1. *If $D(x_0,y_0) > 0$ and $\frac{\partial^2 f(x_0,y_0)}{\partial x^2} < 0$, $f(x_0,y_0)$ is a local maximum value.*

2. *If $D(x_0,y_0) > 0$ and $\frac{\partial^2 f(x_0,y_0)}{\partial x^2} > 0$, $f(x_0,y_0)$ is a local minimum value.*

3. *If $D(x_0,y_0) < 0$, $f(x_0,y_0)$ is a saddle point, and thus not an extreme value.*

4. *If $D(x_0,y_0) = 0$, the test is inconclusive.*

Example 5.11 (Second Partials Test)

1. Find the extrema of $f(x,y) = 3x^3 + y^2 - 9x + 4y$

   The gradient is

   $$\nabla f(x,y) = (9x^2 - 9, 2y + 4),$$

   so the critical points are $(x,y) = (1,-2)$ and $(x,y) = (-1,-2)$. We'll now compute the second partial derivatives in order to apply the above theorem:

   $$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= 18x \\ \frac{\partial^2 f}{\partial y^2} &= 2 \\ \frac{\partial^2 f}{\partial x \partial y} &= 0.\end{aligned}$$

   Using the formula from the theorem, we have

   $$D(x,y) = \frac{\partial^2 f(x_0,y_0)}{\partial x^2}\frac{\partial^2 f(x_0,y_0)}{\partial y^2} - \left(\frac{\partial^2 f(x_0,y_0)}{\partial y \partial x}\right)^2 = 36x.$$

   We thus have $D(1,-2) = 36 > 0$ and $\frac{\partial^2 f(1,-2)}{\partial x^2} = 18 > 0$, so $(1,-2)$ is a local minimum. For the other critical point, we have $D(-1,-2) = -36 < 0$, so this is a saddle point.

2. Find the minimum distance between the origin and the surface $z^2 = x^2y + 4$. First we need to come up with an expression for this distance, which we will call d:

   $$\begin{aligned}d &= \sqrt{x^2 + y^2 + z^2} \\ d^2 &= x^2 + y^2 + z^2 \\ &= x^2 + y^2 + x^2y + 4.\end{aligned}$$

   Any extreme value of d is also an extreme value of d^2, so we'll use the latter since it's easier to work with. The gradient is

   $$\nabla d^2(x,y) = (2x + 2xy, 2y + x^2).$$

We need to solve the following nonlinear system to find critical points:

$$\begin{aligned} 2x + 2xy &= 0 \\ 2y + x^2 &= 0 \end{aligned}$$

The second equation implies $2y = -x^2$, so substituting that into the first yields

$$2x - x^3 = 0.$$

We thus have either $x = 0$ or $x = \pm\sqrt{2}$. If $x = 0$, then the second equation implies $y = 0$. If $x = \pm\sqrt{2}$, it implies $y = -1$. So the three critical points we must consider are $(0, 0)$, $(\sqrt{2}, -1)$, and $(-\sqrt{2}, -1)$.

The second partials are:

$$\begin{aligned} \frac{\partial^2(d^2)}{\partial x^2} &= 2 + 2y \\ \frac{\partial^2(d^2)}{\partial y^2} &= 2 \\ \frac{\partial^2(d^2)}{\partial x \partial y} &= 2x, \end{aligned}$$

so we have

$$D = (2 + 2y)(2) - (2x)^2 = 4 + 4y - 4x^2.$$

Applying this to our three critical points, we see that $(0, 0)$ is a minimum, while the other two are both saddle points.

We've focused on two-dimensional examples throughout this section. Although most of the other results generalize to higher dimensions, it is considerably more complicated to apply the Second Partials Test with more than two dimensions. We'll come back to that when we do matrix algebra.

5.3 Integral Calculus

5.3.1 Double Integration over a Rectangle

Remember we defined the Riemann integral for functions of one variable in the following way: we formed an arbitrary partition P on the interval of interest, call it $[a, b]$, where we created subintervals of length Δx_k. We then picked a sample point $\bar{x}_k$ from the kth subinterval and wrote

$$\int_a^b f(x)\,dx = \lim_{|P|\to 0} \sum_{k=1}^{n} f(\bar{x_k})\Delta x_k.$$

We will define the integral in multiple dimensions similarly. Suppose we are interested in integrating over a rectangular region, call it R, where

$$R = \{(x, y) \,|\, a \le x \le b, c \le y \le d\}.$$

Next, we form a partition P of the rectangular region R by means of lines parallel to the x and y axes. We divide R into subrectangles, say n of them, denoted R_k, $k = 1, 2, \ldots, n$. Let Δx_k and Δy_k be lengths of the sides of the rectangles, so that $A_k = \Delta x_k \Delta y_k$ is the area of the kth rectangle. Then, as before, we pick a sample point for each of the n rectangles $(\bar{x}_k, \bar{y}_k)$.

Now consider the Reimann sum,

$$\sum_{k=1}^{n} f(\bar{x}_k, \bar{y}_k)A_k.$$

This is the sum of the volumes of these n boxes, where the height of each represents the value of f at the respective sample point. Just as in the univariate case, we will define the integrability of a function in terms of the existence of a limit of such sums.

Definition 5.11. Let f be a function of two variables that is defined on a closed rectangle R. If

$$\lim_{|P|\to 0} \sum_{k=1}^{n} f(\bar{x}_k, \bar{y}_k)A_k$$

exists and is finite, we say that f is *integrable* on R. We write the value of the limit as $\iint\limits_R f(x, y)\,dA$, also called the *double integral* of f over R. □

The double integral given above represents the volume of the region under the curve given by the function of two variables $f(x, y)$.

There is an analogous sufficient condition for integrability to the one we saw in the univariate case.

Theorem 5.12 *If f is bounded on the closed rectangle R and is continuous there except on at most a finite number of smooth curves, then f is integrable on R.*

Properties of the double integral:

1. $\iint\limits_R kf(x,y)\,dA = k\iint\limits_R f(x,y)\,dA$.
2. $\iint\limits_R f(x,y) + g(x,y)\,dA = \iint\limits_R f(x,y)\,dA + \iint\limits_R g(x,y)\,dA$.
3. The double integral is additive on rectangles that overlap only on a line segment: $\iint\limits_R f(x,y)\,dA = \iint\limits_{R_1} f(x,y)\,dA + \iint\limits_{R_2} f(x,y)\,dA$. Where $R = R_1 \cup R_2$.
4. If $f(x,y) \leq g(x,y)$ for all $(x,y) \in R$, then $\iint\limits_R f(x,y) \leq \iint\limits_R g(x,y)\,dA$.

Example 5.12 (Riemann sum in two dimensions)
Let f be a staircase function where:

$$f(x,y) = \begin{cases} 1 & 0 \leq x \leq 3,\ 0 \leq y \leq 1 \\ 2 & 0 \leq x \leq 3,\ 1 < y \leq 2 \\ 3 & 0 \leq x \leq 3,\ 2 < y \leq 3 \end{cases}$$

We introduce the following partition,

$$\begin{aligned} R_1 &= \{(x,y)\,|\,0 \leq x \leq 3, 0 \leq y \leq 1\} = [0,3] \times [0,1], \\ R_2 &= \{(x,y)\,|\,0 \leq x \leq 3, 1 \leq y \leq 2\} = [0,3] \times [1,2], \\ R_3 &= \{(x,y)\,|\,0 \leq x \leq 3, 2 \leq y \leq 3\} = [0,3] \times [2,3], \end{aligned}$$

where the area of each rectangular segment is $A_{R_i} = 3 \cdot 1 = 3$. We can then evaluate the integral as

$$\begin{aligned}\iint\limits_R f(x,y)\,dA &= \iint\limits_{R_1} f(x,y)\,dA + \iint\limits_{R_2} f(x,y)\,dA + \iint\limits_{R_3} f(x,y)\,dA \\ &= 1 \cdot A_{R_1} + 2 \cdot A_{R_2} + 3 \cdot A_{R_3} \\ &= (1+2+3) \cdot 3 = 18.\end{aligned}$$

Now let's never use Reimann sums again. There is another, more precise and much less putzy way to compute the volume of solids under curves: the iterated integral.

Consider the problem of ascertaining the volume of some solid by cutting it into "slices" along the y axis. We can approximate the volume of the slice at y_i by multiplying its width by the area of its face, which is

$$A(y_i) = \int_a^b f(x, y_i)\,dx.$$

Our approximation to the overall volume would then be

$$V \simeq \sum_{i=1}^{n} \Delta y_i A(y_i) = \sum_{i=1}^{n} \Delta y_i \cdot \int_a^b f(x, y_i)\,dx.$$

If we kept making the slices smaller and smaller, until the width of each became infinitesimal, we would end up with an *iterated integral*,

$$V = \int_c^d A(y)\,dy = \int_c^d \left(\int_a^b f(x,y)\,dx \right) dy.$$

But recall that $V = \iint\limits_R f(x,y)\,dA$. So we get the result:

$$\iint\limits_R f(x,y)\,dA = \int_c^d \left(\int_a^b f(x,y)\,dx \right) dy = \int_a^b \left(\int_c^d f(x,y)\,dy \right) dx.$$

This is the technique we will almost always use to evaluate integrals in practice.

Example 5.13 (Iterated integral)

1. Evaluate $\iint 2x + 3y\, dA$ on $R = [1,2] \times [0,3]$.

$$\begin{aligned}
\iint_R 2x + 3y\, dA &= \int_0^3 \int_1^2 2x + 3y\, dx\, dy \\
&= \int_0^3 x^2 + 3yx\Big|_{x=1}^{2}\ dy \\
&= \int_0^3 4 + 6y - 1 - 3y\, dy \\
&= 3y + \frac{3}{2}y^2\Big|_{y=0}^{3} \\
&= 9 + \frac{3}{2} \cdot 9 = \frac{45}{2}.
\end{aligned}$$

2. Evaluate $\iint \frac{1}{16}(64 - 8x + y^2)\, dA$ on $R = [0,4] \times [0,8]$.

$$\begin{aligned}
\iint_R \frac{1}{16}(64 - 8x + y^2)\, dA &= \frac{1}{16}\int_0^8 \int_0^4 64 - 8x + y^2\, dx\, dy \\
&= \frac{1}{16}\int_0^8 64x - 4x^2 + y^2x\Big|_{x=0}^{4}\ dy \\
&= \frac{1}{16}\int_0^8 256 - 64 + 4y^2\, dy \\
&= \int_0^8 12 + \frac{1}{4}y^2\, dy \\
&= 12y + \frac{1}{12}y^3\Big|_{y=0}^{8} \\
&= 96 + \frac{512}{12} = \frac{416}{3}.
\end{aligned}$$

5.3.2 Double Integration over Non-Rectangular Spaces

This is all well and good for rectangular regions, but what about more arbitrary spaces?

Obviously, this can get to be quite complicated. For our purposes here, we will consider two types of areas.

Definition 5.12. We call a set S y-simple if there are functions ϕ_1 and ϕ_2 on $[a, b]$ such that

$$S = \{(x, y) \mid \phi_1(x) \leq y \leq \phi_2(x),\ a \leq x \leq b\}.$$

We call a set S x-simple if there are functions ψ_1 and ψ_2 on $[c, d]$ such that

$$S = \{(x, y) \mid \psi_1(y) \leq x \leq \psi_2(y),\ c \leq y \leq d\}. \qquad \square$$

We can integrate over a y-simple region via

$$\iint\limits_S f(x, y)\, dA = \int_a^b \int_{\phi_1(x)}^{\phi_2(x)} f(x, y)\, dy\, dx.$$

Similarly, we can integrate over an x-simple region via

$$\iint\limits_S f(x, y)\, dA = \int_c^d \int_{\psi_1(y)}^{\psi_2(y)} f(x, y)\, dx\, dy.$$

Example 5.14 (Integrals over non-rectangular spaces)

1. Evaluate the iterated integral $\int_3^5 \int_{-x}^{x^2} 4x + 10y\, dy\, dx$.

We have

$$\begin{aligned}
\int_3^5 \int_{-x}^{x^2} 4x + 10y \, dy \, dx &= \int_3^5 4xy + 5y^2 \Big|_{y=-x}^{x^2} \, dx \\
&= \int_3^5 4x^3 + 5x^4 - (-4x^2 + 5x^2) \, dx \\
&= \int_3^5 4x^3 + 5x^4 - x^2 \, dx \\
&= x^4 + x^5 - \frac{x^3}{3} \Big|_{x=3}^5 \\
&= 5^4 + 5^5 - \frac{5^3}{3} - 3^4 - 3^5 + \frac{3^3}{3}.
\end{aligned}$$

2. Use double integration to find the volume of the tetrahedron bounded by the coordinate planes and the plane $3x+6y+4z-12 = 0$.

   It's always helpful in these sorts of problems to draw the region. First, solve for z: $4z = 12 - 3x - 6y$ and so $z = 3 - \frac{3}{4}x - \frac{3}{2}y$. (Setting $x = z = 0$, we get a y-intercept at 2; setting $z = y = 0$, we get an x-intercept at 4.)

   Now let's call the area of interest in the xy plane the region S. This corresponds to the following triangle. The line is $y = 2 - \frac{1}{2}x$, so this corresponds to either the y-simple region

   $$S = \{(x, y) \,|\, 0 \le y \le 2 - \tfrac{1}{2}x,\ 0 \le x \le 4\}$$

   or the x-simple region

   $$S = \{(x, y) \,|\, 0 \le y \le 2,\ 0 \le x \le 4 - 2y\}.$$

   We'll get the same answer either way, so let's just use the y-simple one.

$$
\begin{aligned}
\iint_S f(x,y)\,dA &= \int_0^4 \int_0^{2-\frac{x}{2}} 3 - \frac{3}{4}x - \frac{3}{2}y\,dy\,dx \\
&= \frac{3}{4}\int_0^4 4y - xy - y^2\Big|_{y=0}^{2-\frac{x}{2}}\,dx \\
&= \frac{3}{4}\int_0^4 8 - 2x - x(2-\frac{x}{2}) - (2-\frac{x}{2})^2\,dx \\
&= \frac{3}{4}\int_0^4 8 - 2x - 2x + \frac{x^2}{2} - (4 - 2x + \frac{x^2}{4})\,dx \\
&= \frac{3}{4}\int_0^4 8 - 4x + \frac{x^2}{2} - 4 + 2x - \frac{x^2}{4}\,dx \\
&= \frac{3}{4}\int_0^4 8 - 2x + \frac{x^2}{4}\,dx \\
&= \frac{3}{4}\left(4x - x^2 + \frac{x^3}{12}\right)\Big|_{x=0}^4 \\
&= \frac{3}{4}\cdot\frac{64}{12} = 4.
\end{aligned}
$$

3. Evaluate $\int_0^4 \int_{\frac{x}{2}}^2 e^{y^2}\,dy\,dx$.

The first thing to recognize here is that e^{y^2} has no simple antiderivative: it is impossible to do this as written. But we can convert the given y-simple region into an x-simple one.

$$\{(x,y)\,|\,0\le x\le 4,\ \tfrac{x}{2}\le y\le 2\} = \{(x,y)\,|\,0\le x\le 2y,\ 0\le y\le 2\}.$$

We now have

$$\begin{aligned}
\int_0^4 \int_{\frac{x}{2}}^2 e^{y^2}\,dy\,dx &= \int_0^2 \int_0^{2y} e^{y^2}\,dx\,dy \\
&= \int_0^2 e^{y^2} x\Big|_{x=0}^{2y}\,dy \\
&= \int_0^2 2ye^{y^2}\,dy \\
&= e^{y^2}\Big|_{y=0}^{2} = e^4 - 1.
\end{aligned}$$

Note that when the region of interest is not x- or y-simple, we can typically break it up into multiple simple regions and then integrate each individually.

5.3.3 Application: Center of Mass

Consider a surface of variable density $d(x, y)$ covering a region S in the xy-plane. Then the center of mass of this surface is given by the following formulas:

$$\bar{x} = \frac{\iint\limits_S xd(x, y)\,dA}{\iint\limits_S d(x, y)\,dA}$$

$$\bar{y} = \frac{\iint\limits_S yd(x, y)\,dA}{\iint\limits_S d(x, y)\,dA}$$

When d is a probability density functions, the denominator integrates to 1 and we call this the *expected value.*

Example 5.15 (Center of mass)
Find the center of mass of a surface with density $d(x, y) = xy$ bounded by the x-axis, the line $x = 8$ and the curve $y = x^{\frac{2}{3}}$. The denominator for

both $\bar{x}$ and $\bar{y}$ is

$$\int_0^8 \int_0^{x^{\frac{2}{3}}} xy\,dy\,dx = \int_0^8 \left.\frac{xy^2}{2}\right|_{y=0}^{x^{\frac{2}{3}}} dx = \int_0^8 \frac{x^{\frac{7}{3}}}{2}\,dx = \left.\frac{3}{20}x^{\frac{10}{3}}\right|_{x=0}^{8} = \frac{3}{20}2^{10}.$$

The numerator for $\bar{x}$ is

$$\int_0^8 \int_0^{x^{\frac{2}{3}}} x(xy)\,dy\,dx = \int_0^8 \left.\frac{x^2y^2}{2}\right|_{y=0}^{x^{\frac{2}{3}}} dx = \int_0^8 \frac{x^{\frac{10}{3}}}{2}\,dx = \left.\frac{3}{26}x^{\frac{13}{3}}\right|_{x=0}^{8} = \frac{3}{26}2^{13},$$

so we have

$$\bar{x} = \frac{\frac{3}{26}2^{13}}{\frac{3}{20}2^{10}} = \frac{10}{13}\cdot 8 = \frac{80}{13}.$$

The numerator for $\bar{y}$ is

$$\int_0^8 \int_0^{x^{\frac{2}{3}}} y(xy)\,dy\,dx = \int_0^8 \left.\frac{xy^3}{3}\right|_{y=0}^{x^{\frac{2}{3}}} dx = \int_0^8 \frac{x^3}{3}\,dx = \left.\frac{x^4}{12}\right|_{x=0}^{8} = \frac{1}{12}2^{12},$$

so we have

$$\bar{y} = \frac{\frac{1}{12}2^{12}}{\frac{3}{20}2^{10}} = \frac{4}{12}\cdot\frac{20}{3} = \frac{20}{9}.$$

Therefore, the center of mass is $(\bar{x},\bar{y}) = (\frac{80}{13},\frac{20}{9})$.

Chapter 6

Matrix Algebra

6.1 Operations on Matrices

6.1.1 Definitions

Now, we are going to completely switch gears and talk a bit about matrices and the operations we perform on them, an area of mathematics known as linear algebra. We'll come back to calculus later and use the things we talk about here to come up with conditions for maxima in multiple dimensions.

Definition 6.1. A *matrix* is a rectangular array of numbers denoted by

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

If a matrix has m rows and n columns, we call this an $m \times n$ matrix. We say that the ith row of A is $[a_{i1}, a_{i2}, \dots, a_{in}]$ and the jth row is

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Note that in the special case where $m = n$, we say that A is a square matrix of order n, and moreover that elements a_{11}, a_{22}, $\dots$, a_{nn} are on the main diagonal of A. □

Definition 6.2. Two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are *equal* if they agree entry by entry, that is, if $a_{ij} = b_{ij}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. □

Example 6.1 (Matrix equality)
The following matrices,

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & -3 & 4 \\ 0 & -4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & w \\ 2 & x & 4 \\ y & -4 & z \end{bmatrix},$$

are equal if and only if $x = -3$, $y = 0$, $w = -1$, and $z = 5$.

Definition 6.3. If $A = [a_{ij}]$ and $B = [b_{ij}]$ are both $m \times n$ matrices, then the sum $A + B$ is a $m \times n$ matrix $C = [c_{ij}]$ defined by $c_{ij} = a_{ij} + b_{ij}$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. □

Example 6.2 (Matrix addition)
With

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 2 & -1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 3 & -4 \end{bmatrix},$$

we have

$$A + B = \begin{bmatrix} 1 & 0 & 4 \\ 3 & 2 & 0 \end{bmatrix}.$$

Definition 6.4. If $A = [a_{ij}]$ is an $m \times n$ matrix and r is a real number, then the scalar multiple of A by r, denoted rA, is the $m \times n$ matrix $C = [c_{ij}]$ where $c_{ij} = ra_{ij}$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. □

Example 6.3 (Scalar multiplication of a matrix)
Let $r = -2$ and

$$A = \begin{bmatrix} 4 & -2 & -3 \\ 7 & -3 & 2 \end{bmatrix}.$$

Then we have

$$rA = \begin{bmatrix} -8 & 4 & 6 \\ -14 & 6 & -4 \end{bmatrix}.$$

Note that we can define the difference between two matrices using the previous two properties; that is, for matrices A and B, both $m \times n$, $A - B = A + (-1)B$.

Definition 6.5. If $A = [a_{ij}]$ is an $m \times n$ matrix and $B = [b_{ij}]$ is an $n \times p$ matrix, then the product of A and B, written $AB = C = [c_{ij}]$, is an $m \times p$ matrix defined by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \ldots + a_{in} b_{nj}$$

for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, p$. $\square$

Example 6.4 (Matrix multiplication)
With

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 5 \\ 4 & -3 \\ 2 & 1 \end{bmatrix},$$

we have

$$AB = \begin{bmatrix} 1 \cdot -2 + 2 \cdot 4 + (-1) \cdot 2 & 1 \cdot 5 + 2 \cdot (-3) + -1 \cdot 1 \\ 3 \cdot (-2) + 1(4) + 4(2) & 3(5) + 1(-3) + 4(1) \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 6 & 16 \end{bmatrix}$$

What about BA? We know from the definition that it will be 3×3, and hence not the same thing. This is our first hint that matrix multiplication does *not* satisfy commutation, meaning it is possible (and in fact is almost always the case) that $AB \neq BA$.

Note that AB does not exist if the number of columns in A is not equal to the number of rows in B.

Definition 6.6. If $A = [a_{ij}]$ is an $m \times n$ matrix, then the transpose of A, $A^T = [a^T_{ij}]$ is the $n \times m$ matrix defined by $a^T_{ij} = a_{ji}$. The transpose is also sometimes written A'. □

Example 6.5 (Transpose)
With

$$A = \begin{bmatrix} 1 & 2 & -1 \\ -3 & 2 & 7 \end{bmatrix},$$

we have

$$A^T = \begin{bmatrix} 1 & -3 \\ 2 & 2 \\ -1 & 7 \end{bmatrix}.$$

6.1.2 Properties

With these definitions in hand, we can consider some of their implications. For the sake of time, I won't offer proofs of most of the following facts, but you can look at any introductory linear algebra text to find them.

Theorem 6.1 (Properties of Matrix Addition) *Let A, B, and C be $m \times n$ matrices.*

1. $A + B = B + A$.
2. $A + (B + C) = (A + B) + C$.
3. *There is a unique additive identity, an $m \times n$ matrix 0_{mn} such that $A + 0_{mn} = A$.*
4. *There is a unique additive inverse, an $m \times n$ matrix D such that $A + D = 0_{mn}$. We write D as $-A$ so that $A - A = 0_{mn}$.*

Theorem 6.2 (Properties of Scalar Multiplication) *Let r and s be real numbers, and let A and B be matrices.*

1. $r(sA) = (rs)A$

2. $(r + s)A = rA + sA$

3. $r(A + B) = rA + rB$

4. $A(rB) = r(AB) = (rA)B$

Theorem 6.3 (Properties of Matrix Multiplication) *Let A, B and C be matrices.*

1. $A(BC) = (AB)C$

2. $(A + B)C = AC + BC$

3. $C(A + B) = CA + CB$

Theorem 6.4 (Properties of Transpose) *Let r be a scalar and let A and B be matrices.*

1. $(A^T)^T = A$

2. $(A + B)^T = A^T + B^T$

3. $(AB)^T = B^T A^T$

4. $(rA)^T = rA^T$

Now let me quickly note some common confusions that can arise when we deal with matrices. They look like scalars at times, but they are not scalars and do not operate in the same way real numbers do:

1. If a and b are real numbers, then $ab = 0$ can hold only if $a = 0$ or $b = 0$. This is not true for matrices.

2. If a, b, and c are real numbers and $ab = ac$ with $a \neq 0$, then $b = c$. However, we cannot simply cancel out A when dealing with matrices, and hence C may not equal B even if $AB = AC$.

3. As already observed, AB need not equal BA.

6.2 Special Types of Matrices

There are a few special types of matrices that we might care about.

Definition 6.7. An $n \times n$ matrix $A = [a_{ij}]$ is called a *diagonal matrix* if $a_{ij} = 0$ for all $i, j \in \{1, \dots, n\}$ where $i \neq j$. That is, all the terms off the main diagonal are equal to zero. □

Definition 6.8. A diagonal matrix where all the elements on the main diagonal are the same is called a *scalar matrix*. □

Definition 6.9. The scalar matrix $I_n = [a_{ij}]$ where $a_{ii} = 1$ for all i and $a_{ij} = 0$ for all $i \neq j$ is called the $n \times n$ *identity matrix*. □

Example 6.6 (Special types of matrices)
Consider the following matrices:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Of these A is diagonal, B is scalar (hence also diagonal), and C is the 3×3 identity matrix (hence also scalar and diagonal).

Theorem 6.5 *If A is an $m \times n$ matrix, $AI_n = I_m A = A$.*

Proof. Let $AI_n = C = [c_{ij}]$. Take any $i = 1, \dots, m$ and $j = 1, \dots, n$, and consider the matrix entry c_{ij}. By definition of matrix multiplication, we have

$$\begin{aligned} c_{ij} &= \sum_{k=1}^{n} a_{ik} I_{kj} \\ &= a_{i1} I_{1j} + \ldots + a_{ij} I_{jj} + \ldots + a_{in} I_{nj} \\ &= a_{i1} \cdot 0 + \ldots + a_{ij} \cdot 1 + \ldots + a_{in} \cdot 0 \\ &= a_{ij}. \end{aligned}$$

The proof for $I_m A$ is analogous. ■

Definition 6.10. An $n \times n$ matrix $A = [a_{ij}]$ is *upper triangular* if $a_{ij} = 0$ for all $i > j$. Similarly, we say that A is *lower triangular* if $a_{ij} = 0$ for all $i < j$. □

Example 6.7 (Triangular matrices)
Consider

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 0 & 2 & 5 \\ 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 3 & 5 & 3 \end{bmatrix},$$

A is upper triangular and B is lower triangular.

Definition 6.11. An $n \times n$ matrix $A = [a_{ij}]$ is *symmetric* if $A^T = A$. It is *skew symmetric* if $A^T = -A$. □

Definition 6.12. An $n \times n$ matrix $A = [a_{ij}]$ is *non-singular*, or *invertible* if there exists an $n \times n$ matrix B such that $AB = BA = I_n$. B is called the inverse of A. If no such matrix B exists, A is called singular, or non-invertible. □

Example 6.8 (Inverse of a matrix)
The matrices A and B are inverses of each other:

$$A = \begin{bmatrix} 4 & 1 \\ 7 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & -1 \\ -7 & 4 \end{bmatrix}$$

Theorem 6.6 *The inverse of a matrix, if it exists, is unique.*

Proof. Let B and C be inverses of A. Then we have

$$B = BI_n = B(AC) = (BA)C = I_nC = C,$$

so $B = C$. ■

Theorem 6.7 *If A and B are both non-singular $n \times n$ matrices, then AB is non-singular and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. We know that A^{-1} and B^{-1} exist and are both $n \times n$. We then have

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n.$$

Since the inverse of a matrix is unique, we conclude $(AB)^{-1} = B^{-1}A^{-1}$. ■

Theorem 6.8 *If A is a non-singular matrix, then A^{-1} is non-singular and $(A^{-1})^{-1} = A$.*

Proof. Immediate from the fact that $AA^{-1} = A^{-1}A = I_n$ for any non-singular A. ■

Theorem 6.9 *If A is a non-singular matrix, then A^T is non-singular and $(A^T)^{-1} = (A^{-1})^T$.*

Proof. Note that $AA^{-1} = I_n$, which implies

$$(AA^{-1})^T = (A^{-1})^TA^T = I_n^T = I_n.$$ ■

6.3 Gauss-Jordan Reduction

6.3.1 Solving Linear Systems

Now it's time to start applying matrix algebra. To do this systematically, we need to develop some basic tools, and then we'll be able to solve for inverses by hand, as well as linear systems of equations.

Suppose that we have the following system of equations:

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ x_2 + x_3 &= 2 \\ x_2 &= 1 \end{aligned}$$

We can represent this as something called an augmented matrix where we take the coefficients from the unknowns and the solutions to the equations and combine them in the following way:

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

There is a systematic way to manipulate this system, represented by an augmented matrix, into a form from which the solution can be found quite easily. To do this, we'll need some definitions.

Definition 6.13. An $m \times n$ matrix $A = [a_{ij}]$ is said to be in *reduced row echelon form* if it satisfies the following:

1. All zero rows, if there are any, appear at the bottom of the matrix.

2. The first non-zero entry of a non-zero row is a 1. This entry is called a leading 1 of its row.

3. For each non-zero row, the leading 1 appears to the right and below any leading ones in preceding rows.

4. If a column contains a leading 1, then all other entries in that column are zero. □

Example 6.9 (Reduced row echelon form)
An example of a matrix in reduced row echelon form is

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & -2 & 4 \\ 0 & 1 & 0 & 0 & 4 & 8 \\ 0 & 0 & 0 & 1 & 7 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Definition 6.14. The *elementary row operations* are:

1. Interchange any two rows

2. Multiply a row by a non-zero number

3. Add a multiple of one row to another □

Example 6.10 (Elementary row operations)

1. Find a matrix in reduced row echelon form starting with the matrix

$$A = \begin{bmatrix} 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 2 & 2 & -5 & 2 & 4 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}.$$

   (a) Interchange rows 1 and 3

$$A_1 = \begin{bmatrix} 2 & 2 & -5 & 2 & 4 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

   (b) Multiply R_1 by $\frac{1}{2}$

$$A_2 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 2 & 0 & -6 & 9 & 7 \end{bmatrix}$$

   (c) $-2R_1 + R_4$

$$A_3 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & -2 & -1 & 7 & 3 \end{bmatrix}$$

   (d) R_2 swapped with R_3

$$A_4 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & -2 & -1 & 7 & 3 \end{bmatrix}$$

   (e) $R_2 + R_4$

$$A_5 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 2 & 3 & 4 \end{bmatrix}$$

(f) $-R_3 + R_4$

$$A_6 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 2 & 3 & -4 & 1 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(g) Multiply both R_2 and R_3 by $\frac{1}{2}$

$$A_7 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(h) $-\frac{3}{2}R_3 + R_2$

$$A_8 = \begin{bmatrix} 1 & 1 & -\frac{5}{2} & 1 & 2 \\ 0 & 1 & 0 & -\frac{17}{4} & \frac{-5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(i) $\frac{5}{2}R_3 + R_1$

$$A_9 = \begin{bmatrix} 1 & 1 & 0 & \frac{19}{4} & 7 \\ 0 & 1 & 0 & -\frac{17}{4} & \frac{-5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(j) $R_1 - R_2$

$$A_{10} = \begin{bmatrix} 1 & 0 & 0 & 9 & \frac{19}{2} \\ 0 & 1 & 0 & -\frac{17}{4} & \frac{-5}{2} \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that A_{10} is in reduced row echeleon form. The process we used here is called Gauss-Jordan reduction.

2. Consider the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 9 \\ 2x_1 - x_2 + x_3 &= 8 \\ 3x_1 - x_3 &= 3 \end{aligned}$$

This corresponds to the following augmented matrix:

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 3 & 0 & -1 & 3 \end{bmatrix}$$

Let's apply Gauss-Jordan reduction to solve the system:

(a) $R_3 - 3R_1 \rightarrow R_3$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 2 & -1 & 1 & 8 \\ 0 & -6 & -10 & -24 \end{bmatrix}$$

(b) $R_2 - 2R_1$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 0 & -5 & -5 & -10 \\ 0 & -6 & -10 & -24 \end{bmatrix}$$

(c) $\frac{1}{5}R_2 \rightarrow R_2$ and $\frac{1}{6}R_3 \rightarrow R_3$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 0 & -1 & -1 & -2 \\ 0 & -1 & -\frac{10}{6} & -4 \end{bmatrix}$$

(d) $-R_2 + R_3 \rightarrow R_3$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & -\frac{4}{6} & -2 \end{bmatrix}$$

(e) $-\frac{6}{4}R_3 \rightarrow R_3$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

(f) $R_3 + R_2 \rightarrow R_2$

$$\begin{bmatrix} 1 & 2 & 3 & 9 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

(g) $2R_3 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 0 & 3 & 11 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

(h) $-1R_2 \to R_2$ and $-3R_3 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

This corresponds to the solution $x_1 = 2$, $x_2 = -1$, $x_3 = 3$.

3. Consider the linear system given by the following augmented matrix:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 & 6 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

This system will have no solution; the last equation suggests that $0x_1 + 0x_2 + 0x_3 + 0x_4 = 1$, which is not possible.

4. Consider the linear system given by the following augmented matrix in reduced row echelon form:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & -1 & 7 \\ 0 & 0 & 1 & 2 & 3 & 7 \\ 0 & 0 & 0 & 1 & 2 & 9 \end{bmatrix}$$

This corresponds to:

$$\begin{aligned} x_4 + 2x_5 &= 9 \\ x_3 + 2x_4 + 3x_5 &= 7 \\ x_2 + 2x_3 + 3x_4 - x_5 &= 7 \\ x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 &= 6 \end{aligned}$$

Thus:

$$\begin{aligned}
x_4 &= 9 - 2x_5 \\
x_3 &= 7 - 2x_4 + 3x_5 = 7 - 2(9 - 2x_5) - 3x_5 = -11 + x_5 \\
x_2 &= 7 + x_5 - 2x_3 - 3x_4 = 7 + x_5 - 2(-11 + x_5) - 3(9 - 2x_5) = 2 + 5x_5 \\
x_1 &= 6 - 2x_2 - 3x_3 - 4x_4 - 5x_5 = -1 - 10x_5
\end{aligned}$$

This system therefore has infinitely many solutions!

$$\begin{aligned}
x_1 &= -1 - 10x_5 \\
x_2 &= 2 + 5x_5 \\
x_3 &= -11 + x_5 \\
x_4 &= 9 - 2x_5 \\
x_5 &= \text{any real number!}
\end{aligned}$$

6.3.2 Inverting Matrices

We can use the same techniques to find the inverses of non-singular matrices.

Example 6.11 (Matrix inversion)
Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{bmatrix}.$$

We form the following augmented matrix:

$$\begin{bmatrix} A & \vdots & I_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & 2 & 3 & \vdots & 0 & 1 & 0 \\ 5 & 5 & 1 & \vdots & 0 & 0 & 1 \end{bmatrix}.$$

We now begin Gauss-Jordan reduction:

1. $-5R_1 + R_3 \to R_3$

$$\begin{bmatrix} 1 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & 2 & 3 & \vdots & 0 & 1 & 0 \\ 0 & 0 & -4 & \vdots & -5 & 0 & 1 \end{bmatrix}$$

2. $-\frac{1}{4}R_3 \to R_3$ and $\frac{1}{2}R_2 \to R_2$

$$\begin{bmatrix} 1 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & 1 & \frac{3}{2} & \vdots & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & \vdots & \frac{5}{4} & 0 & \frac{-1}{4} \end{bmatrix}$$

3. $-\frac{3}{2}R_3 + R_2 \to R_2$

$$\begin{bmatrix} 1 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & 1 & 0 & \vdots & -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & 0 & 1 & \vdots & \frac{5}{4} & 0 & \frac{-1}{4} \end{bmatrix}$$

4. $-R_3 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 1 & 0 & \vdots & -\frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 & \vdots & -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & 0 & 1 & \vdots & \frac{5}{4} & 0 & \frac{-1}{4} \end{bmatrix}$$

5. $R_1 - R_2 \to R_1$

$$\begin{bmatrix} 1 & 0 & 0 & \vdots & \frac{13}{8} & \frac{-1}{2} & \frac{-1}{8} \\ 0 & 1 & 0 & \vdots & -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & 0 & 1 & \vdots & \frac{5}{4} & 0 & \frac{-1}{4} \end{bmatrix}$$

So our inverse is

$$\begin{bmatrix} \frac{13}{8} & \frac{-1}{2} & \frac{-1}{8} \\ -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{5}{4} & 0 & \frac{-1}{4} \end{bmatrix}.$$

6.4 The Determinant

A determinant is a function that associates a scalar value $\det(A)$, to every $n \times n$ square matrix A. To talk about determinants we first need to talk quickly about permutations.

Definition 6.15. Let $S = \{1, 2, \ldots, n\}$ be the set of integers from 1 to n, arranged in ascending order. A rearrangement of these integers is called a *permutation* of S. That is, a permutation is a one-to-one mapping of S onto itself. □

Example 6.12 (Permutations)
Let $S = \{1, 2, 3, 4\}$. Then 4321 is a 4-permutation of S corresponding to the function:

$$\begin{aligned} f(1) &= 4 \\ f(2) &= 3 \\ f(3) &= 2 \\ f(4) &= 1 \end{aligned}$$

Note that the general form for the number of permutations of n objects can be constructed by simply considering how many possibilities are available for the first position, how many are available for the second position, how many for the third, etc. This results in $n! = n(n-1)(n-2)(n-3)\ldots(3)(2)(1)$ possibilities.

Definition 6.16. A permutation $j_1, j_2, \ldots, j_n$ is said to have an *inversion* if a larger integer, call it j_r precedes a smaller one, call it j_s. A permutation is called *even* if the total number of inversions in it is even and *odd* if the total number of inversions in it is odd. □

Example 6.13 (Even and odd permutations)

1. In the 4-permutation 4312:

- 4 precedes 1, 2, and 3
- 3 precedes 1 and 2

Hence, there are 5 total inversions, so this is an odd permutation.

2. Consider $S_3 = \{1, 2, 3\}$. There are 3! permutations:
   - 123 with 0 inversions
   - 213 with 1 inversion
   - 231 with 2 inversions
   - 132 with 1 inversion
   - 312 with 2 inversions
   - 321 with 3 inversions

We use the notion of even and odd permutations to define the determinant.

Definition 6.17. Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then we define the *determinant* as

$$\det(A) = \sum \pm a_{1j_1} a_{2j_2} \dots a_{nj_n},$$

where the summation is over all permutations $j_1 j_2 \dots j_n$ of the set $S = \{1, 2, \dots, n\}$ and the sign is $+$ if the permutation $j_1 j_2 \dots j_n$ is even and $-$ if it is odd. $\square$

In each term $\pm a_{1j_1} a_{nj_n} \dots a_{nj_n}$ of $\det(A)$ the row subscripts are in their natural numerical order, while the column subscripts are in a permuted order.

Thankfully, we never use this. It's a pain. We have algorithms that we use instead.

Determinant of a 2×2. Consider the 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

To obtain the determinant of A, we need to consider the two permutations: $j_1j_2 = 12$ and $j_1'j_2' = 21$. The first has no inversions and hence is even; the second has one inversion and hence is odd. So, using the definition of a determinant, we have

$$\det(A) = \pm a_{1j_1}a_{2j_2} \pm a_{1j_1'}a_{2j_2'} = a_{11}a_{22} - a_{12}a_{21}.$$

Example 6.14 (Determinant of a 2 × 2)
If

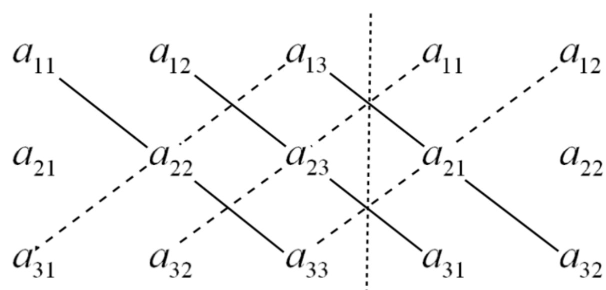
$$A = \begin{bmatrix} 2 & -3 \\ 4 & 5 \end{bmatrix},$$

then $\det(A) = 2(5) - (4(-3)) = 22$.

Determinant of a 3×3. Consider the 3×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

In this case, we can use the following rule:[1]



You take the product along each of the diagonal lines, multiplying said product by 1 if it goes down to the right, and by -1 if it runs down to the left. This produces the following:

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}.$$

[1]Graphic from Wikipedia.

Example 6.15 (Determinant of a 3 × 3)
Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

The determinant is

$$\begin{aligned} \det(A) &= 1(1)(2) + 2(3)(3) + 3(2)(1) - 2(2)(2) - 1(3)(1) - 3(1)(3) \\ &= 2 + 18 + 6 - 8 - 3 - 9 \\ &= 6. \end{aligned}$$

We are now unfortunately out of clever tricks, and can't apply these methods to larger matrices. You could probably do the algebra for a 4×4 matrix, but anything beyond this would be quite difficult.

If you ever had to take the determinant of a larger matrix by hand, you would probably want to use cofactor expansion. This is a method whereby we find the determinant of an n-order matrix by reducing it to the evaluation of several determinants of order $n - 1$.

Before we move on to definiteness, let's go over some properties of the determinant and its use for solving linear systems.

Theorem 6.10 (Properties of Determinants) *Let A and B be $n \times n$ matrices, and let c be a real number.*

1. $\det(A) = \det(A^T)$.
2. *If a row or column of A consists entirely of zeros, then* $\det(A) = 0$.
3. *If A is upper or lower triangular, then* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.
4. *A is non-singular if and only if* $\det(A) \neq 0$. *In this case,* $\det(A^{-1}) = \frac{1}{\det(A)}$.
5. $\det(AB) = \det(A)\det(B)$.

6. *If B results from A by interchanging two rows or columns of A, then* $\det(B) = -\det(A)$.

7. *If B is obtained from A by multiplying a row or column of A by c, then* $\det(B) = c\det(A)$.

8. $\det(cA) = c^n \det(A)$.

We can also use determinants to solve linear systems of equations. The method can be a bit ... awkward ... but if you greatly prefer computing determinants to row reducing, then by all means go for it!

Theorem 6.11 (Cramer's Rule) *Consider a linear system of n equations in n unknowns,*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n, \end{aligned}$$

and let $A = [a_{ij}]$ be the coefficient matrix so that we can write the given system as $Ax = b$, where

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

If $\det(A) \neq 0$, then the system has the unique solution:

$$\begin{aligned} x_1 &= \frac{\det(A_1)}{\det(A)} \\ x_2 &= \frac{\det(A_2)}{\det(A)} \\ &\vdots \\ x_n &= \frac{\det(A_n)}{\det(A)}, \end{aligned}$$

where each A_i is the matrix obtained from A by replacing the ith column of A by b.

Example 6.16 (Cramer's Rule)
Consider the linear system

$$\begin{aligned} 3x_1 + 2x_2 &= 7, \\ -2x_1 + 2x_2 &= 2. \end{aligned}$$

We have

$$\begin{aligned} A &= \begin{bmatrix} 3 & 2 \\ -2 & 2 \end{bmatrix}, & \det(A) &= 6 - (-4) = 10, \\ A_1 &= \begin{bmatrix} 7 & 2 \\ 2 & 2 \end{bmatrix}, & \det(A_1) &= 14 - 4 = 10, \\ A_2 &= \begin{bmatrix} 3 & 7 \\ -2 & 2 \end{bmatrix}, & \det(A_2) &= 6 - (-14) = 20. \end{aligned}$$

The solution to the system is thus

$$(x_1, x_2) = \left(\frac{\det(A_1)}{\det(A)}, \frac{\det(A_2)}{\det(A)} \right) = (1, 2).$$

6.5 Definiteness

Finally, we'll talk about definiteness at a strictly mechanical level—its implications will be left for PSC 404 and PSC 407.

Definition 6.18. An $n \times n$ matrix A is:

- *positive definite* if $x^T Ax > 0$ for all non-zero length-n vectors x,
- *negative definite* if $x^T Ax < 0$ for all non-zero length-n vectors x,
- *indefinite* if there exists x for which $x^T Ax < 0$ and y for which $y^T Ay > 0$. □

Definition 6.19. Let A be a $n \times n$ matrix. Any $k \times k$ matrix formed by deleting $n-k$ columns $i_1, i_2, \ldots, i_{n-k}$ and the same $n-k$ rows is a *kth-order principal submatrix* of A. The determinant of a $k \times k$ principal submatrix is called a *kth-order principal minor* of A. The principal submatrix formed by deleting the last $n-k$ rows is the *kth-order* leading *principal submatrix*, and its determinant is the *kth-order* leading *principal minor*. □

Example 6.17 (Principal minors)
Consider the general 3×3 matrix,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

1. There are three first-order principal minors: $\det[a_{11}]$ (leading), $\det[a_{22}]$, and $\det[a_{33}]$.
2. There are three second-order principal minors:
   (a) delete row and column 3 (leading):
   $$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$
   (b) delete row and column 2:
   $$\det \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}$$
   (c) delete row and column 1:
   $$\det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$
3. There is one third-order principal minor, $\det(A)$ (leading).

Theorem 6.12 *Let A be an $n \times n$ symmetric matrix. Then:*

1. *A is positive definite if and only if all of its n leading principal minors are strictly positive.*

2. *A is negative definite if and only if its odd-order leading principal minors are negative and its even-order leading principal minors are positive.*

3. *If some leading principal minor is not zero, but does not fit this pattern, then the matrix is indefinite.*

And that's it for the prefresher!