Math 140: Foundations of Real Analysis

Todd Kemp
# Contents

## Part 1. Math 140A

### Chapter 1. Ordered Sets, Ordered Fields, and Completeness
1. Lecture 1: January 5, 2016
2. Lecture 2: January 7, 2016
4. Lecture 4: January 14, 2014

### Chapter 2. Sequences and Limits
1. Lecture 5: January 19, 2016
2. Lecture 6: January 21, 2016
3. Lecture 7: January 26, 2016
4. Lecture 8: January 28, 2016

### Chapter 3. Extensions of $\mathbb{R}$: the Extended Real Numbers $\mathbb{R}$ and the Complex Numbers $\mathbb{C}$

### Chapter 4. Series
1. Lecture 11: February 9, 2016

### Chapter 5. Metric Spaces
1. Lecture 14: February 18, 2016
2. Lecture 15: February 23, 2016
4. Lecture 17: February 29, 2016

### Chapter 6. Limits and Continuity
2. Lecture 19: March 8, 2016
3. Lecture 20: March 10, 2016

## Part 2. Math 140B

### Chapter 7. More on Continuity
1. Lecture 1: March 29, 2016
2. Lecture 2: March 31, 2016
3. Lecture 3: April 5, 2016

### Chapter 8. Differentiation of Functions of a Real Variable
1. Lecture 4: April 7, 2016  
2. Lecture 5: April 12, 2016  
4. Lecture 7: April 19, 2016

Chapter 9. Integration

1. Lecture 8: April 21, 2016  
2. Lecture 9: April 26, 2016  
Part 1

Math 140A
CHAPTER 1

Ordered Sets, Ordered Fields, and Completeness

1. Lecture 1: January 5, 2016

- \( \mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C} \).
- \( \mathbb{R} \) is the “Réal numbers”. There is nothing real about them! That is the first, most important lesson to learn in this class. We will encounter many “obvious” statements that are, in fact, false. We will also see some counterintuitive statements that turn out to be true.
- Mathematicians roughly split into two groups: analysts and algebraists. (There’s lots of overlap, though.) Roughly speaking, algebraists are largely concerned about equalities, while analysts are largely concerned about inequalities.

**Definition 1.1.** A total order is a binary relation \( < \) on a set \( S \) which satisfies:

1. transitive: if \( x, y, z \in S, x < y \), and \( y < z \), then \( x < z \).
2. ordered: given any \( x, y \in S \), exactly one of the following is true: \( x < y \), \( x = y \), or \( y < x \).

The usual order relation on \( \mathbb{Q} \) (and its subsets \( \mathbb{Z} \) and \( \mathbb{N} \)) is a total order. As usual, we write \( x > y \) to mean \( y < x \), and \( x \leq y \) to mean “\( x < y \) or \( x = y \)”.

**Definition 1.2.** Let \( (S, <) \) be a totally ordered set. Let \( E \subseteq S \). A lower bound for \( E \) is an element \( \alpha \in S \) with the property that \( \alpha \leq x \) for each \( x \in E \). A upper bound for \( E \) is an element \( \beta \in S \) with the property that \( x \leq \beta \) for each \( x \in E \). If \( E \) possesses an upper bound, we say \( E \) is bounded above; if it possesses a lower bound, it is bounded below.

For example, the set \( \mathbb{N} \) is bounded below in \( \mathbb{Z} \), but it is not bounded above. Any set that has a maximal element is bounded above by its maximum; similarly, any set with a minimal element is bounded below by its minimum.

**Definition 1.3.** Let \( (S, <) \) be a totally ordered set, and let \( E \subseteq S \) be bounded above. The least upper bound or supremum of \( E \), should it exist, is

\[
\sup E \equiv \min \{ \beta \in S : \beta \text{ is an upper bound of } E \}.
\]

Similarly, if \( F \) is bounded below, the greatest lower bound or infimum of \( F \), should it exist, is

\[
\inf F \subseteq S \equiv \max \{ \alpha \in S : \alpha \text{ is a lower bound of } F \}.
\]

To work with the definition (of \( \sup \), say), we rewrite it slightly. A number \( \sigma \in S \) is the supremum of \( E \) if the following two properties hold:

1. \( \sigma \) is an upper bound of \( E \).
2. Given any \( s \in E \) with \( s < \sigma \), \( s \) is not an upper bound of \( E \); i.e. there exists some \( x \in E \) with \( s < x \leq \sigma \).

**Example 1.4.** Consider the set \( E = \{ \frac{1}{n} : n \in \mathbb{N} \} \subseteq \mathbb{Q} \). This set has a maximal element: 1. So 1 is an upper bound. Moreover, if \( s \in \mathbb{Q} \) is < 1, then \( s \) is not an upper bound of \( E \) (since
1. ORDERED SETS, ORDERED FIELDS, AND Completeness

1 ∈ E). Thus, 1 = sup E. (This argument shows in general that, if E has a maximal element, then \( \max E = \sup E \).

On the other hand, E has no minimal element. But note that all elements of E are positive, so 0 is a lower bound for E. If s is any rational number > 0, there is certainly some \( n \in \mathbb{N} \) with \( 0 < \frac{1}{n} < s \) (this is the Archimedean property of the rational field). Hence, no such s is a lower bound for E. This shows that 0 is the greatest lower bound: \( 0 = \inf E \).

Example 1.5. It is well known that \( \sqrt{2} \) is not rational: in other words, there is no rational number \( p \) satisfying \( p^2 = 2 \). You probably saw this proof in high school. Suppose, for a contradiction, that \( p^2 = 2 \). Since \( p \) is rational, we can write it in lowest terms as \( p = m/n \) for \( m, n \in \mathbb{Z} \). So we have \( \frac{m^2}{n^2} = 2 \), or \( m^2 = 2n^2 \). Thus \( m^2 \) is even, which means that \( m \) is even (since the square of an odd integer is odd). So \( m = 2k \) for some \( k \in \mathbb{Z} \), meaning \( m^2 = 4k^2 \), and so \( 4k^2 = 2n^2 \), from which it follows that \( n^2 = 2k^2 \) is even. As before, this implies that \( n \) is even. But then both \( m \) and \( n \) are divisible by 2, which means they are not relatively prime. This contradicts the assumption that \( p = m/n \) is in lowest terms.

A finer analysis of this situation shows that \( \mathbb{Q} \) has “holes”. Let

\[
A = \{ r \in \mathbb{Q} : r > 0, r^2 < 2 \}, \quad \text{and} \quad B = \{ r \in \mathbb{Q} : r > 0, r^2 > 2 \}.
\]

The set \( A \) is bounded above: if \( q \geq \frac{3}{2} \) then \( q^2 \geq \frac{9}{4} > 2 \), meaning that \( q \notin A \); the contrapositive is that if \( q \in A \) then \( q < \frac{3}{2} \), so \( \frac{3}{2} \) is an upper bound for \( A \). In fact, take any positive rational number \( r \); then \( r^2 > 0 \) is also rational. By the total order relation, exactly one of the following three statements is true: \( r^2 < 2 \), \( r^2 = 2 \), or \( r^2 > 2 \). In other words, \( \mathbb{Q}_{>0} = A \cup \{ r \in \mathbb{Q} : r > 0, r^2 = 2 \} \cup B \). We just showed that the middle set is empty, so

\( \mathbb{Q}_{>0} = A \cup B \).

- Every element \( b \in B \) is an upper bound for \( A \). Indeed, if \( a \in A \) and \( b \in B \), then \( a^2 < 2 < b^2 \) so \( 0 < b^2 - a^2 = (b - a)(b + a) \), and dividing through by the positive number \( b + a \) shows \( b - a > 0 \) so \( a < b \). (This also shows that every element \( a \in A \) is a lower bound for \( B \).)

- On the other hand, if \( a \in A \), then \( a \) is not an upper bound for \( A \); i.e. given \( a \in A \), there exists \( a' \in A \) with \( a < a' \). To see this, we can just take

\[
a' = a + \frac{2 - a^2}{2 + a} = \frac{2a + 2}{a + 2}.
\]

Since \( a \in A \), we know \( a^2 < 2 \) so \( 2 - a^2 > 0 \), and the denominator \( 2 + a > 0 \); so \( a' > a \). But we also have

\[
2 - (a')^2 = \frac{2(a + 2)^2 - (2a + 2)^2}{(a + 2)^2} = \frac{2a^2 + 8a + 8 - 4a^2 - 8a - 4}{(a + 2)^2} = \frac{2(2 - a^2)}{(a + 2)^2} > 0;
\]

showing that \( a' \in A \), as claimed.

Thus, \( B \) is equal to the set of upper bounds of \( A \) in \( \mathbb{Q}_{>0} \), and similarly \( A \) is equal to the set of lower bounds of \( B \) in \( \mathbb{Q}_{>0} \).

But then we have the following strange situation. The set \( A \) of lower bounds of \( B \) has no greatest element: we just showed that, given any \( a \in A \), there is an \( a' \in A \) with \( a' > a \). Hence, \( B \) has no greatest lower bound: \( \inf B \) does not exist in \( \mathbb{Q}_{>0} \). Similarly, \( \sup A \) does not exist in \( \mathbb{Q}_{>0} \).

Example 1.5 viscerally demonstrates that there is a “hole” in \( \mathbb{Q} \): the fact that \( r^2 = 2 \) has no solution in \( \mathbb{Q} \) forces the ordered set to be disconnected into two pieces, each of which is very incomplete: not only does each fail to possess a max/min, they also fail to possess a sup/inf.
2. Lecture 2: January 7, 2016

We now set the stage for the formal study of the real numbers: it is the (unique) complete ordered field. To understand these words, we begin with fields.

**Definition 1.6.** A field is a set \( \mathbb{F} \) equipped with two binary operations \(+, \cdot : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F} \), called addition and multiplication, satisfying the following properties.

1. **Commutativity:** \( \forall a, b \in \mathbb{F}, a + b = b + a \) and \( a \cdot b = b \cdot a \).
2. **Associativity:** \( \forall a, b, c \in \mathbb{F}, (a + b) + c = a + (b + c) \) and \( (a \cdot b) \cdot c = a \cdot (b \cdot c) \).
3. **Identity:** there exists elements 0, 1 \( \in \mathbb{F} \) s.t. \( \forall a \in \mathbb{F}, 0 + a = a = 1 \cdot a \).
4. **Inverse:** for any \( a \in \mathbb{F} \), there is an element denoted \(-a \in \mathbb{F} \) with the property that \( a + (-a) = 0 \). For any \( a \in \mathbb{F} \setminus \{0\} \), there is an element denoted \( a^{-1} \) with the property that \( a \cdot a^{-1} = 1 \).
5. **Distributivity:** \( \forall a, b, c \in \mathbb{F}, a \cdot (b + c) = (a \cdot b) + (a \cdot c) \).

**Example 1.7.** Here are some examples of fields.

1. The field \( \mathbb{Z}_p = \{0, 1, \ldots, [p-1]\} \) for any prime \( p \), where the + and \( \cdot \) are the usual ones inherited from the + and \( \cdot \) on \( \mathbb{Z} \) (namely \( [a] + [b] = [a + b] \) and \( [a] \cdot [b] = [a \cdot b] \) – you studied this field in Math 109). All finite fields have this form.
2. \( \mathbb{Q} \) is a field.
3. \( \mathbb{Z} \) is not a field: it fails item (4), lacking multiplicative inverses of all elements other than \( \pm 1 \).
4. Let \( \mathbb{Q}(t) \) denote the set of rational functions of a single variable \( t \) with coefficients in \( \mathbb{Q} \):
   \[
   \mathbb{Q}(t) = \left\{ \frac{p(t)}{q(t)} : p(t), q(t) \text{ are polynomials with coefficients in } \mathbb{Q} \text{ and } q(t) \text{ is not identically 0} \right\}.
   \]
   With the usual addition and multiplication of functions, \( \mathbb{Q}(t) \) is a field. For example, \( \left( \frac{p(t)}{q(t)} \right)^{-1} = \frac{q(t)}{p(t)} \), which exists so long as \( p(t) \) is not identically 0 – i.e. as long as the original rational function \( \frac{p(t)}{q(t)} \) is not the 0 function.

Fields are the kinds of number systems that behave the way you’ve grown up believing numbers behave, as summarized in the following lemma.

**Lemma 1.8.** Let \( \mathbb{F} \) be a field. The following properties hold.

1. **Cancellation:** \( \forall a, b, c \in \mathbb{F}, if a + b = c \text{ then } b = c. If } a \neq 0 \text{, if } a \cdot b = a \cdot c \text{ then } b = c. \)
2. **Hungry Zero:** \( \forall a \in \mathbb{F}, 0 \cdot a = 0 \).
3. **No Zero Divisors:** \( \forall a, b \in \mathbb{F}, if a \cdot b = 0, then either } a = 0 \text{ or } b = 0. \)
4. **Negatives:** \( \forall a, b \in \mathbb{F}, (-a)\cdot b = -(ab), -(-a) = a, \text{ and } (-a)(-b) = ab. \)

**Proof.** We’ll just prove (2), leaving the others to the reader. For any \( a \in \mathbb{F} \), note that \( 0 \cdot a + a = 0 \cdot a + 1 \cdot a = (0 + 1) \cdot a = 1 \cdot a = a = 0 + a. \)

Hence, by (1) (cancellation), it follows that \( 0 \cdot a = 0 \). \( \square \)

**Example 1.9.** As in Example 1.7, we can consider \( \mathbb{Z}_n \) for any positive integer \( n \). This satisfies all of the properties of Definition 1.6 except (4): inverses don’t always exist. For example, if \( n \) can be factored as \( n = km \) for two positive integers \( k, m > 1 \), then we have two nonzero elements \( [k], [m] \in \mathbb{Z}_n \) such that \( [k] \cdot [m] = [km] = [n] = [0] \), which contradicts Lemma 1.8(3) – there are zero divisors. So \( \mathbb{Z}_n \) is not a field when \( n \) is composite.
Now, we combine fields with ordered sets.

**Definition 1.10.** An **ordered field** is a field $\mathbb{F}$ which is an ordered set $(\mathbb{F}, <)$, where the order relation also satisfies the following two properties:

1. $\forall a, b, c \in \mathbb{F}$, if $a < b$ then $a + c < b + c$.
2. $\forall a, b \in \mathbb{F}$, if $a > 0$ and $b > 0$, then $a \cdot b > 0$.

From here, all the usual properties mixing the order relation and the field operations follow. For example:

**Lemma 1.11.** Let $(\mathbb{F}, <)$ be an ordered field. Then

1. $\forall a \in \mathbb{F}$, $a > 0$ iff $-a < 0$.
2. $\forall a \in \mathbb{F} \setminus \{0\}$, $a^2 > 0$. In particular, $1 = 1^2 > 0$.
3. $\forall a, b \in \mathbb{F}$, if $a > 0$ and $b < 0$, then $a \cdot b < 0$.
4. $\forall a \in \mathbb{F}$, if $a > 0$ then $a^{-1} > 0$.

**Proof.** For (1), simply add $-a$ to both sides of the inequality. Note, by the properties of $<$, this means $\mathbb{F}$ is the union of three disjoint subsets: the positive elements $a > 0$, the negative elements $a < 0$, and the zero element $a = 0$; and the operation of multiplication by $-1$ interchanges the positive and negative elements. So, for (2), we note that our given $a \neq 0$ must be either positive or negative; if $a > 0$ then $a^2 = a \cdot a > 0$ by Definition 1.10(2), while if $a < 0$ then $a^2 = (-a)^2 > 0$ by the same argument. For (3), we then have $a > 0$ and $b < 0$, so $-(ab) = a \cdot (-b) > 0$, which means that $ab < 0$. Finally, for (4), suppose $a^{-1} < 0$, then by (3) we would have $1 = a \cdot a^{-1} < 0$; but by (2) we know $1 > 0$. This contradiction shows that $a^{-1} > 0$. □

**Example 1.12.**

1. $\mathbb{Q}$ is an ordered field, with its usual order: $\frac{m_1}{n_1} < \frac{m_2}{n_2}$ iff $m_1n_2 < m_2n_1$. In fact, this is the *unique* total order on the set $\mathbb{Q}$ which makes $\mathbb{Q}$ into an ordered field.

2. $\mathbb{Z}_p$ is not an ordered field for any prime $p$. For suppose it were; then by Lemma 1.11(2) we know that $[1] > [0]$. Then $[2] = [1] + [1] > [1] + [0] = [1]$, and so by transitivity $[2] > [0]$. Continuing this way by induction, we get to $[p - 1] > [0]$. But we also have $[0] = [1] + [p - 1] > [0] + [p - 1] = [p - 1]$. This is a contradiction.

3. Let $\mathbb{F}$ be an ordered field. Denote by $\mathbb{F}_c$ the following set of $2 \times 2$ matrices over $\mathbb{F}$:

$$\mathbb{F}_c = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a, b \in \mathbb{F} \right\}.$$ 

The determinant of such a matrix is $a^2 + b^2$. In an ordered field, we know that $a^2 > 0$ if $a \neq 0$, and thus we have the usual property that $a^2 + b^2 = 0$ iff $a = b = 0$. It follows that all nonzero matrices in $\mathbb{F}_c$ are invertible: we can easily verify that

$$(a^2 + b^2)^{-1} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ 

If we define

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

then $\mathbb{F}_c = \{ aI + bJ : a, b \in \mathbb{F} \}$. Note that $J^2 = -I$. It is now an easy exercise to show that $\mathbb{F}_c$ is a field, with $+, \cdot$ being given by matrix addition and multiplication, where $I$ is the multiplicative identity and the additive identity is the $2 \times 2$ zero matrix. (Note: this is *not generally true* if $\mathbb{F}$ is not an ordered field. For example, in $\mathbb{Z}_2$ we have $1^2 + (-1)^2 = 0$,
and as a result the matrix with \( a = b = 1 \) is not invertible in this case.) \( \mathbb{F}_c \) is the
complexification of \( \mathbb{F} \). We will later construct the complex numbers \( \mathbb{C} \) as \( \mathbb{C} = \mathbb{R}_c \).

3.5 If \( \mathbb{F} \) is any ordered field, then \( \mathbb{F}_c \) cannot be ordered – there is no order relation that makes
\( \mathbb{F}_c \) into an ordered field. This is actually what Problem 4 on HW1 asks you to prove.

Item 2 above noted that the finite fields \( \mathbb{Z}_p \) are not ordered fields. In fact, ordered fields must be
infinite. The next results shows why this is true.

**Lemma 1.13.** Let \((\mathbb{F}, \prec)\) be an ordered field. Then, for any \( n \in \mathbb{Z} \setminus \{0\} \), \( n \cdot 1_\mathbb{F} \neq 0_\mathbb{F} \).

Here \( n \cdot 1_\mathbb{F} = 1_\mathbb{F} + 1_\mathbb{F} + \cdots + 1_\mathbb{F} \). Note that this property is not automatic for fields: for example, in \( \mathbb{Z}_p \), \( p \cdot [1] = [0] \).

**Proof.** First, \( 1 \cdot 1_\mathbb{F} = 1_\mathbb{F} > 0_\mathbb{F} \) by Lemma 1.11(2). Proceeding by induction, suppose we’ve shown that \( n \cdot 1_\mathbb{F} \neq 0_\mathbb{F} \). Then \( (n+1) \cdot 1_\mathbb{F} = n \cdot 1_\mathbb{F} + 1_\mathbb{F} > 0_\mathbb{F} + 1_\mathbb{F} = 1_\mathbb{F} > 0_\mathbb{F} \). Thus, for every \( n > 0 \),
\( n \cdot 1_\mathbb{F} > 0_\mathbb{F} \), meaning it is \( \neq 0 \). If, on the other hand, \( n < 0 \) in \( \mathbb{Z} \), then \( n \cdot 1_\mathbb{F} = -(-n \cdot 1_\mathbb{F}) < 0_\mathbb{F} \), so also it is \( \neq 0_\mathbb{F} \). \( \square \)

**Corollary 1.14.** Let \( \mathbb{F} \) be an ordered field. The map \( \varphi : \mathbb{Q} \to \mathbb{F} \) given by \( \varphi\left(\frac{m}{n}\right) = (m \cdot 1_\mathbb{F}) \cdot (n \cdot 1_\mathbb{F})^{-1} \) is an injective ordered field homomorphism.

An **ordered field homomorphism** is a function which preserves the field operations: \( \varphi(a + b) = \varphi(a) + \varphi(b) \), \( \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b) \), and \( \varphi(0) = 0 \) and \( \varphi(1) = 1 \); and preserves the order relation: if \( a < b \) then \( \varphi(a) < \varphi(b) \). An injective ordered field homomorphism should be thought of as an
**embedding:** we realize \( \mathbb{Q} \) as a subset of \( \mathbb{F} \), in a way that respects all the ordered field structure.

**Proof.** First we must check that \( \varphi \) is well defined: if \( \frac{m_1}{n_1} = \frac{m_2}{n_2} \), then \( m_1 n_2 = m_2 n_1 \). It then
follows (by an easy induction) that \((m_1 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F}) = (m_2 \cdot 1_\mathbb{F}) \cdot (n_1 \cdot 1_\mathbb{F}) \). Dividing out on both sides then shows that \((m_1 \cdot 1_\mathbb{F})(n_1 \cdot 1_\mathbb{F})^{-1} = (m_2 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F})^{-1} \). Thus, \( \varphi \) is well-defined. It is similar and routine to verify that it is an ordered field homomorphism. Finally, to show it is one-to-one, suppose that \( \varphi(q_1) = \varphi(q_2) \) for \( q_1, q_2 \in \mathbb{Q} \). Using the homomorphism property, this means
\( \varphi(q_1 - q_2) = \varphi(q_1) - \varphi(q_2) = 0 \). Let \( q_1 - q_2 = \frac{m}{n} \); thus, we have \( \varphi\left(\frac{m}{n}\right) = (m \cdot 1_\mathbb{F}) \cdot (n \cdot 1_\mathbb{F})^{-1} = 0_\mathbb{F} \). But then, multiplying through by the non-zero (by Lemma 1.13) element \( n \cdot 1_\mathbb{F} \), we have \( m \cdot 1_\mathbb{F} = 0_\mathbb{F} \), and again by Lemma 1.13 it follows that \( m = 0 \). But this means \( q_1 - q_2 = \frac{m}{n} = 0 \), so \( q_1 = q_2 \).

Thus, \( \varphi \) is injective.

**Proof.** First we must check that \( \varphi \) is well defined: if \( \frac{m_1}{n_1} = \frac{m_2}{n_2} \), then \( m_1 n_2 = m_2 n_1 \). It then
follows (by an easy induction) that \((m_1 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F}) = (m_2 \cdot 1_\mathbb{F}) \cdot (n_1 \cdot 1_\mathbb{F}) \). Dividing out on both sides then shows that \((m_1 \cdot 1_\mathbb{F})(n_1 \cdot 1_\mathbb{F})^{-1} = (m_2 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F})^{-1} \). Thus, \( \varphi \) is well-defined. It is similar and routine to verify that it is an ordered field homomorphism. Finally, to show it is one-to-one, suppose that \( \varphi(q_1) = \varphi(q_2) \) for \( q_1, q_2 \in \mathbb{Q} \). Using the homomorphism property, this means
\( \varphi(q_1 - q_2) = \varphi(q_1) - \varphi(q_2) = 0 \). Let \( q_1 - q_2 = \frac{m}{n} \); thus, we have \( \varphi\left(\frac{m}{n}\right) = (m \cdot 1_\mathbb{F}) \cdot (n \cdot 1_\mathbb{F})^{-1} = 0_\mathbb{F} \). But then, multiplying through by the non-zero (by Lemma 1.13) element \( n \cdot 1_\mathbb{F} \), we have \( m \cdot 1_\mathbb{F} = 0_\mathbb{F} \), and again by Lemma 1.13 it follows that \( m = 0 \). But this means \( q_1 - q_2 = \frac{m}{n} = 0 \), so \( q_1 = q_2 \).

Thus, \( \varphi \) is injective.

In Lecture 1, we saw that \( \mathbb{Q} \) “has holes”. In example 1.15 we found two subsets \( A, B \subset \mathbb{Q} \) with
the property that \( B = \) the set of upper bounds of \( A \), \( A = \) the set of lower bounds of \( B \), and \( A \) has
no maximal element, while \( B \) has no minimal element. Thus, \( \text{sup} A \) and \( \text{inf} B \) do not exist. This
turns out to be a serious obstacle to doing the kind of analysis we’re used to in calculus, so we’d
like to fill in these holes. This motivates our next definition.

**Definition 1.15.** An ordered set \((S, <)\) is called complete if every nonempty subset \(\emptyset \neq E \subseteq S\) that is bounded above possesses a supremum \(\sup E \in S\). We also denote this by saying that \((S, <)\) has the least upper bound property.

We could also formulate things in terms of \(\inf\), with the greatest lower bound property. Example 1.15 demonstrates how these two are typically related. In fact, they are equivalent.

**Proposition 1.16.** An ordered set \((S, <)\) has the least upper bound property if and only if, for every nonempty subset \(\emptyset \neq F \subseteq S\) that is bounded below, \(\inf F \in S\) exists.

**Proof.** We will argue the forward implication: the least upper bound property implies the greatest lower bound property. The converse is very similar.

Let \(F \neq \emptyset\) be bounded below; then \(L \equiv \{\text{lower bounds for } F\}\) is a nonempty subset of \(S\). If \(x \in L\) and \(y \in F\), then \(x \leq y\), which shows that every \(y \in F\) is an upper bound for \(L\). Thus, \(L\) is bounded above and nonempty; by the least upper bound property of \(S\), \(\sigma = \sup L \in S\) exists. By definition of supremum, if \(x < \sigma\) then \(x\) is not an upper bound for \(L\); since every element of \(F\) is an upper bound for \(L\), this means that such \(x\) is not in \(F\). Taking contrapositives, this says that if \(z \in F\) then \(x \geq \sigma\). So \(\sigma\) is a lower bound for \(F\) – i.e. \(\sigma \in L\). This shows that \(\sigma = \max L\): i.e. \(\sigma\) is the greatest lower bound of \(F\): \(\sigma = \inf F\). So \(\inf F\) exists, as claimed. 

Let us now prove some important properties that complete ordered fields possess – properties that are critical for doing all of analysis.

**Theorem 1.17.** Let \(\mathbb{F}\) be a complete ordered field.

1. **Archimedean** Let \(x, y \in \mathbb{F}\) with \(x > 0\). Then there exists \(n \in \mathbb{N}\) so that \(nx > y\).
2. **Density of \(\mathbb{Q}\)** Let \(x, y \in \mathbb{F}\), with \(x < y\). Then there exists \(r \in \mathbb{Q}\) so that \(x < r < y\).

A field with property (1) is called Archimedean. It tells us (by setting \(x = 1\)) that the set \(\mathbb{N}\) is not bounded above in the field: there is no \(y \in \mathbb{F}\) that is \(\geq\) every integer. It also tells us (by setting \(y = 1\)) that there are no “infinitesimals” – that is, no matter how small a positive number \(x\) is, there is always a positive integer \(n\) such that \(0 < \frac{1}{n} < x\). This is an absolutely crucial property for a field to have if we want to talk about limits. And it does not hold in every ordered field.

**Example 1.18.** In the field \(\mathbb{Q}(t)\) of rational functions with rational coefficients, it is always possible to uniquely express a function \(f(t) \in \mathbb{Q}(t)\) in the form \(f(t) = \lambda \cdot \frac{p(t)}{q(t)}\) where \(\lambda \in \mathbb{Q}\) and \(p(t), q(t)\) are monic polynomials: their highest order terms have coefficient 1. This allows us to define an order on \(\mathbb{Q}(t)\): say \(f(t) < g(t)\) iff \(g(t) - f(t) = \lambda \frac{p(t)}{q(t)}\) where \(p(t), q(t)\) are monic and \(\lambda > 0\). (This is the same as insisting that the leading coefficients of the numerator and denominator of \(f(t) - g(t)\) have the same sign.) For example, \(\frac{t^2 - 2t + 7}{t^2 - 10t^2} > 0\) while \(-\frac{t^2 - 2t + 7}{t^2 - 10t^2} < 0\). Then it is easy but laborious to check that this makes \(\mathbb{Q}(t)\) into an ordered field. Note: \(t - n = 1 \cdot \frac{t - n}{1} > 0\) for any integer \(n\); this means that, in the ordered field \(\mathbb{Q}(t)\), the element \(t\) is greater than every integer. I.e. the set \(\mathbb{Z} \subset \mathbb{Q}(t)\) actually has an upper bound (e.g. \(t\)) in \(\mathbb{Q}(t)\). This means \(\mathbb{Q}(t)\) is a non-Archimedean field. In particular, by Theorem 1.17 \(\mathbb{Q}(t)\) is not a complete ordered field.

**Proof of Theorem 1.17.** (1) Suppose, for a contradiction, there there is no such \(n\): that is, \(nx \leq y\) for every \(n \in \mathbb{N}\). Let \(E = \{nx : n \in \mathbb{N}\}\). Then our assumption is that \(y\) is an upper bound for \(E\), so \(E\) is bounded above. It is also non-empty (it contains \(x\), for example). Thus, since \(\mathbb{F}\) is complete, it follows that \(\alpha = \sup E\) exists. In particular, since \(\alpha - x < \alpha\), this means that \(\alpha - x\) is
Combining these gives us
we must have
\[ m \]
Dividing through by (the positive) \( n \) shows that the set
\[ \{ k \in \mathbb{Z} : nx < k \leq m \} \]
is finite: it is contained in the finite set \( \{-m_2 + 1, -m_2 + 2, \ldots, m_1\} \). So, let \( m = \min\{k \in \mathbb{Z} : nx < k\} \). Then since \( m - 1 \in \mathbb{Z} \) and \( m - 1 < m \), we must have \( m - 1 \leq nx \).

Thus, we have two inequalities:
\[ n(y - x) > 1, \quad m - 1 \leq nx < m. \]
Combining these gives us
\[ nx < m \leq nx + 1 < ny. \]
Dividing through by (the positive) \( n \) shows that \( x < \frac{m}{n} < y \), so setting \( r = \frac{m}{n} \) completes the proof.

Here is another extremely important property that holds in ordered fields; this is crucial for doing calculus.

**Proposition 1.19.** Let \( \mathbb{F} \) be a complete ordered field. For each \( n \in \mathbb{N} \), let \( a_n, b_n \in \mathbb{F} \) satisfy
\[ a_1 \leq a_2 \leq \cdots \leq a_n \leq \cdots \leq b_n \leq \cdots \leq b_2 \leq b_1. \]
Further, suppose that \( b_n - a_n < \frac{1}{n} \). Then \( \bigcap_{n \in \mathbb{N}} [a_n, b_n] \) is nonempty, and consists of exactly one point.

This is sometimes called the **nested intervals property**. It is actually equivalent to the least upper bound property. On HW2, you will prove the converse.

**Proof.** By construction, \( b_1 \) is an upper bound for \( \{a_n : n \in \mathbb{N}\} \), which is a nonempty set. Thus, by completeness, \( \alpha = \sup a_n \) exists in \( \mathbb{F} \). Since \( \alpha \) is an upper bound for \( \{a_n\} \), we have \( a_n \leq \alpha \) for every \( n \). On the other hand, since \( b_n \geq a_n \) for every \( m, n, b_m \) is an upper bound for \( \{a_n\} \), and since \( \alpha \) is the least upper bound, it follows that \( \alpha \leq b_m \) as well. Thus \( \alpha \in [a_n, b_n] \) for every \( n \), and so it is in the intersection.

Now, suppose \( \beta \in \bigcap_n [a_n, b_n] \). Then either \( \alpha < \beta, \alpha > \beta \), or \( \alpha = \beta \). Suppose, for the moment, that \( \alpha < \beta \). Then we have \( a_n \leq \alpha < \beta \leq b_n \) for every \( n \), and since \( b_n - a_n < \frac{1}{n} \), it follows that \( 0 < \beta - \alpha < \frac{1}{n} \) for every \( n \). But this violates the Archimedean property of \( \mathbb{F} \). A similar contradiction arises if we assume \( \alpha > \beta \). Thus \( \alpha = \beta \), and so \( \alpha \) is the unique element of the intersection.

Note: in the setup of the lemma, it is similar to see that the intersection consists of \( \inf_n b_n \); so \( \sup_n a_n = \inf_n b_n \).
4. Lecture 4: January 14, 2014

We have now seen several properties possessed by complete ordered fields. We would hope to find some examples as well. Here comes the big punchline.

**Theorem 1.20.** There exists exactly one complete ordered field. We call this field \(\mathbb{R}\), the Real numbers.

We will talk about the proof of Theorem 1.20 as we proceed in the course. The textbook relegated an existence proof to the end of Chapter 1, through Dedekind cuts. This is an old-fashioned proof, and not very intuitive. We are not going to discuss it presently. Once we have developed a little more technology, we will prove the existence claim of the theorem using Cauchy’s construction of \(\mathbb{R}\) (through sequences).

We can, however, prove the uniqueness claim. To be precise, here is what uniqueness means in this case: suppose \(F\) and \(G\) are two complete ordered fields. Then there exists an ordered field isomorphism \(\varphi: F \rightarrow G\). That means \(\varphi\) is an ordered field homomorphism that is also a bijection. So, from the point of view of ordered fields, \(F\) and \(G\) are indistinguishable.

The first question is: given two complete ordered fields \(F\) and \(G\), how do we define \(\varphi: F \rightarrow G\)?

By Corollary 1.14, \(\mathbb{Q}\) embeds in each of \(F\) and \(G\) via \(\mathbb{Q} \cdot 1_F\) and \(\mathbb{Q} \cdot 1_G\). So we can define \(\varphi\) as a partial function by its action on \(\mathbb{Q}\):

\[
\varphi(r 1_F) = r 1_G, \quad r \in \mathbb{Q}.
\]

The question is: how should we define \(\varphi\) on elements of \(F\) that are not necessarily in \(\mathbb{Q} \cdot 1_F\)? Well, let \(x \in F \setminus \mathbb{Q}\). By Theorem 1.17(2), there are rationals \(a_n, b_n \in \mathbb{Q}\) such that

\[
x - \frac{1}{2n} 1_F < a_n 1_F < x < b_n 1_F < x + \frac{1}{2n} 1_F.
\]

In particular, \(b_n - a_n < \frac{1}{n}\). We should do this carefully and also make sure that \(a_1 \leq a_2 \leq \cdots \leq b_2 \leq b_1\) — this can be achieved by choosing the \(a_n\) and \(b_n\) successively, increasing the \(a_n\) or decreasing the \(b_n\) each step as needed. It follows from Proposition 1.19 that \(\bigcap_n [a_n 1_G, b_n 1_G]\) contains exactly one point, \(\alpha = \sup_n (a_n 1_G) = \inf_n (b_n 1_G)\). So we define

\[
\varphi(x) = \alpha.
\]

Note: if \(x \in \mathbb{Q}\), then \(x \cdot 1_G\) is the unique element in the intersection, meaning that we can take the above nested intervals definition as the formula for \(\varphi\) on all of \(F\), not just the irrational elements. This will be our starting point.

**Theorem 1.21.** If \(F\) and \(G\) are two complete ordered fields, then there exists an ordered field isomorphism \(\varphi: F \rightarrow G\).

**Proof.** Following our outline from above, we define \(\varphi\) as follows. To begin, using the denseness of \(\mathbb{Q}\) in \(F\), select \(a_1, b_1 \in \mathbb{Q}\) so that

\[
x - \frac{1}{2} 1_F < a_1 1_F < x < b_1 1_F < x + \frac{1}{2} 1_F.
\]

Now proceed inductively: once we’ve constructed \(a_1, \ldots, a_{n-1}\) and \(b_1, \ldots, b_{n-1}\), choose \(a_n\) and \(b_n\) so that

\[
\max \left\{ x - \frac{1}{2n} 1_F, a_{n-1} \right\} < a_n 1_F < x < b_n 1_F < \min \left\{ x + \frac{1}{2n} 1_F, b_{n-1} \right\}.
\]

(1.1)
Then we have \( a_1 < a_2 < \cdots < a_n < \cdots < b_n < \cdots < b_2 < b_1 \), and also 
\[
b_n - a_n < \left( x + \frac{1}{2n} \right) - \left( x - \frac{1}{2n} \right) = \frac{1}{n}.
\]

So by the nested intervals property Proposition 1.19 applied in the field \( \mathbb{G} \), we have 
\[
\bigcap_{n \in \mathbb{N}} [a_n \mathbb{G}, b_n \mathbb{G}] = \{ \alpha \}
\]
where \( \alpha = \sup_n a_n \mathbb{G} = \inf_n b_n \mathbb{G} \). We thus define \( \varphi(x) = \alpha \).

Now we must verify that:

- **\( \varphi \) is well-defined:** if \( a'_n, b'_n \) are some other rational elements satisfying (1.1) then \( \sup_n a_n \mathbb{G} = \sup_n a'_n \mathbb{G} \). In fact, this follows because we also then have the mixed inequalities 
  \[
x - \frac{1}{2n} = 1 < a'_n < x < b_n < x + \frac{1}{2n} = 1
\]
  and, as above, we have \( \sup_n a'_n \mathbb{G} = \inf_n b_n \mathbb{G} = \sup_n a_n \mathbb{G} \).

- **\( \varphi \) is an ordered field homomorphism.** This is laborious. Let’s check one of the field homomorphism properties: preservation of addition. Let \( x, y, \in \mathbb{F} \), and let \( a_n < x < b_n \) and \( c_n < y < d_n \) where \( b_n - a_n < \frac{1}{2n} < \frac{1}{n} \) and \( d_n - c_n < \frac{1}{2n} < \frac{1}{n} \). Then \( \varphi(x) = \sup_n a_n \mathbb{G} \) and \( \varphi(y) = \sup_n c_n \). Now, on the other hand, we have 
  \[
a_n + c_n < x + y < b_n + d_n, \quad \text{and} \quad (b_n + d_n) - (a_n + c_n) = (b_n - a_n) + (d_n - c_n) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}
\]
  It follows that \( \varphi(x + y) = \sup(a_n + c_n) \). So, to see that \( \varphi(x + y) = \varphi(x) + \varphi(y) \), it suffices to show that 
  \[
  \text{if } a_n \uparrow \& c_n \uparrow \text{ then } \sup_n (a_n + c_n) = \sup_n a_n + \sup_n c_n.
  \]
  This is also on HW2. The other ordered field homomorphism properties are verified similarly.

- **\( \varphi \) is a bijection.** First, suppose that \( x \neq y \in \mathbb{F} \). Then either \( x < y \) or \( x > y \); wlog \( x < y \). Since \( \varphi \) is an ordered field homomorphism, it follows that \( \varphi(x) < \varphi(y) \). In particular, \( \varphi(x) \neq \varphi(y) \). A similar argument in the case \( x > y \) shows that \( \varphi \) is one-to-one.

Now, fix \( y \in \mathbb{G} \). For each \( n \), choose \( a_n, b_n \in \mathbb{Q} \) nested so that \( b_n - a_n < \frac{1}{n} \) and \( a_n \mathbb{G} < y < b_n \mathbb{G} \). Mirroring the above arguments, we know that \( a = \sup_n a_n \mathbb{G} \in \bigcap_n [a_n \mathbb{G}, b_n \mathbb{G}] \). Since \( a < b_n \mathbb{G} \), we have \( a \mathbb{G} = \varphi(a_n \mathbb{G}) < \varphi(a) < \varphi(b_n \mathbb{G}) = b_n \mathbb{G} \). Thus \( \varphi(a) \in \bigcap_n [a_n \mathbb{G}, b_n \mathbb{G}] \), and this intersection consists of the singleton element \( y \), by Proposition 1.19. Hence, \( \varphi(a) = y \), and so \( \varphi \) is onto.

So, we see that there can be only one complete ordered field. (They’re like Highlanders.) A priori, that doesn’t preclude the possibility that there aren’t any at all. To prove that \( \mathbb{R} \) exists, we need to first start talking about convergence properties of sequences. That will be our next task.

Before proceeding, let’s return to our motivation for studying \( \sup \) and \( \inf \) and introducing completeness: we wanted to fill the “hole” in \( \mathbb{Q} \) where \( \sqrt{2} \) should be. To see that we’ve filled at least that hole, the next result shows that \( \mathbb{R} \) (the complete ordered field) contains square roots, and in fact \( n \)th roots, of all positive numbers. First, let’s state some standard results on “absolute value”.

Lemma 1.22. Let \( \mathbb{F} \) be an ordered field. For \( x \in \mathbb{F} \), define (as usual)
\[
|x| = \begin{cases} 
  x, & \text{if } x \geq 0 \\
  -x, & \text{if } x < 0
\end{cases}
\]
Then we have the following properties.

1. For all \( x \in \mathbb{F} \), \( |x| \geq 0 \), and \( |x| = 0 \) iff \( x = 0 \).
2. For all \( x, y \in \mathbb{F} \), \( |x + y| \leq |x| + |y| \).
3. For all \( x, y \in \mathbb{F} \), \( |xy| = |x||y| \).

All of these properties are straightforward but annoying to prove in cases. We will use the absolute value frequently in all that follows.

Theorem 1.23. Let \( n \in \mathbb{N} \), \( n \geq 1 \). For any \( x \in \mathbb{R} \), \( x > 0 \), there is a unique \( y \in \mathbb{R} \), \( y > 0 \), so that \( y^n = x \). We denote it by \( y = x^{1/n} \).

Proof of Theorem 1.23. First, for uniqueness: let \( y_1 \neq y_2 \) be two positive real numbers, wlog \( y_1 < y_2 \). Then \( y_1^2 = y_1y_1 < y_1y_2 < y_2y_2 = y_2^2 \); continuing by induction, we see that \( y_1^n < y_2^n \). That is: the function \( y \mapsto y^n \) is strictly increasing. In particular, it is one-to-one. It follows that there can be at most one \( y \) with \( y^n = x \).

Now for existence. Let \( E = \{y \in \mathbb{R} : y > 0, y^n < x\} \).

- \( E \neq \emptyset \) : note that \( t = \frac{x}{n+1} \in (0, 1) \). This means that \( 0 < t^n < t \), and so since \( \frac{x}{n+1} < x \), we have \( 0 < t^n < x \), meaning that \( t \in E \).
- \( E \) is bounded above: let \( s = 1 + x \). Then \( s > 1 \), and so \( s^n > s > x \). Thus, if \( y \in E \), then \( y^n < x < s^n \), and so \( s^n - y^n = (s - y)(s^{n-1} + s^{n-2}y + \cdots + y^{n-1}) \). The sum of terms is strictly positive, so we can divide out and find that \( s - y > 0 \). Thus \( s \) is an upper bound for \( E \).

Hence, by completeness of \( \mathbb{R} \), \( \alpha = \sup E \) exists. Since \( \alpha \) is the least upper bound, it follows that, for each \( k \), there is an element \( y_k \in E \) such that \( y_k > \alpha - \frac{1}{k} \). Since \( y_k^n < x \), we therefore have
\[
\left( \alpha - \frac{1}{k} \right)^n < y_k^n < x, \quad \text{for all } k \in \mathbb{N}.
\]
But we can expand
\[
\left( \alpha - \frac{1}{k} \right)^n = \sum_{j=0}^{n} \binom{n}{j} \alpha^{n-j} \left( -\frac{1}{k} \right)^j = \alpha^n - \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left( -\frac{1}{k} \right)^j.
\]
Thus, we have
\[
\alpha^n < x + \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left( -\frac{1}{k} \right)^j
\]
and so, applying the triangle inequality – Lemma 1.22 – repeatedly, we have
\[
\alpha^n < x + \frac{1}{k} \left| \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left( -\frac{1}{k} \right)^j \right| \leq x + \frac{1}{k} \cdot \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left( \frac{1}{k} \right)^j.
\]
Note that \( n \) is fixed, and \( \frac{1}{k} \leq 1 \), so for \( k \geq 1 \) we have \( \left( \frac{1}{k} \right)^j \leq 1 \). Let \( M = \sum_{k=1}^{n} \binom{n}{j} \alpha^{n-j} \); then we have
\[
\forall k \in \mathbb{N} \quad \alpha^n < x + \frac{M}{k} ; \quad \text{i.e. } \alpha^n - x < \frac{M}{k}.
\]
By the Archimedean property, it follows that $\alpha^n - x \leq 0$; thus, we have shown that $\alpha^n \leq x$.

On the other hand, let $y \in E$. Then for any $k \in \mathbb{N}$ we have, by similar calculations,

$$(y + \frac{1}{k})^n = y^n + \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} y^{n-j} \left(\frac{1}{k}\right)^{j-1} \leq y^n + \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} y^{n-j}.$$ 

Since $y \in E$, we know $y^n < x$, so $\epsilon = x - y^n > 0$. Let $L = \sum_{j=1}^{n} \binom{n}{j} y^{n-j}$, which is a positive constant; by the Archimedean property, there is some $k \in \mathbb{N}$ so that $\frac{1}{k} \cdot L < \epsilon$. Thus, for such $k$,

$$\left(y + \frac{1}{k}\right)^n \leq y^n + \frac{L}{k} < y^n + \epsilon = x.$$ 

That is: $y + \frac{1}{k} \in E$. But $y + \frac{1}{k} > y$. That is, for any $y \in E$, there is $y' > y$ with $y \in E$. So $E$ has no maximal element. This shows that $\alpha \notin E$, and hence $\alpha^n \geq x$.

In conclusion: we’ve shown that $\alpha^n \leq x$ and $x \leq \alpha^n$. It follows that $\alpha^n = x$. \hfill \square

On Homework 2, you will flesh out extending this argument to defining $x^r$ for $x > 0$ in $\mathbb{R}$ and $r \in \mathbb{Q}$, and then extending this further to define $x^y$ for $x > 0$ and $y \in \mathbb{R}$. One can use similar arguments to define $\log_b(x)$ for $x, b > 0$. We will wait a little while until we have a firm grounding in sequences and limits before rigorously developing the calculus of these well-known functions.
CHAPTER 2

Sequences and Limits

1. Lecture 5: January 19, 2016

**DEFINITION 2.1.** Let $X$ be a set. A **sequence** in $X$ is a function $a : \mathbb{N} \to X$. Instead of the usual notation $a(n)$ for the value of the function at $n \in \mathbb{N}$, we usually use the notation $a_n = a(n)$; accordingly, we often refer to the function as $(a_n)_{n \in \mathbb{N}}$ or $\{a_n\}_{n \in \mathbb{N}}$, or (when being sloppy) simply $(a_n)$ or $\{a_n\}$.

In ordered fields, we can talk about limits of sequences. The following definition took half a century to finalize; its invention (by Weierstraß) is one of the greatest achievements of analysis.

**DEFINITION 2.2.** Let $\mathbb{F}$ be an ordered field, and let $(a_n)$ be a sequence in $\mathbb{F}$. Let $a \in \mathbb{F}$. Say that $a_n$ **converges to** $a$, written $a_n \to a$ or $\lim_{n \to \infty} a_n = a$, if the following holds true:

$$\forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N |a_n - a| < \epsilon.$$  

Let’s decode the three-quantifier sentence here. What this says is, no matter how small a tolerance $\epsilon > 0$ you want, there is some time $N$ after which all the terms $a_n$ (for $n \geq N$) are within $\epsilon$ of $a$. Some convenient language for this is:

Given any $\epsilon > 0$, we have $|a_n - a| < \epsilon$ for **almost all** $n$.

Here we colloquially say that a set $S \subseteq \mathbb{N}$ contains almost all positive integers if the complement $\mathbb{N} \setminus S$ is finite. This is equivalent to saying that, after some $N$, all $n \geq N$ are in $S$. So, the limit definition is that, for any positive tolerance, no matter how small, almost all of the terms are within that tolerance of the limit.

If $(a_n)$ is a sequence and there exists $a$ so that $a_n \to a$, we say that $(a_n)$ **converges**; if there is no such $a$, we say that $(a_n)$ **diverges**. Here are some examples.

**EXAMPLE 2.3.** Consider each of the following sequences in an Archimedean field.

1. $a_n = 1$ converges to $1$. More generally, if $(a_n)$ is equal to a constant $a$ for almost all $n$, then $a_n \to a$.
2. $a_n = \frac{1}{n}$ converges to $0$.
3. $a_n = n + \frac{1}{n}$ diverges.
4. $a_n = (-1)^n$ diverges.
5. $a_n = 1 + \frac{1}{n}(-1)^n$ converges to $1$.
6. $a_n = \frac{4n+1}{7n-4}$ (defined for $n \geq 1$) converges to $\frac{4}{7}$.

In all these examples, we proved convergence (when the sequences converged) to a given value. However, a priori, it is not clear whether it might also have been possible to prove convergence to a different value as well. This is not the case: limits are unique.

**LEMMA 2.4.** Let $\mathbb{F}$ be an ordered field, and let $(a_n)$ be a sequence in $\mathbb{F}$. Suppose $a, b \in \mathbb{F}$ and $a_n \to a$ and $a_n \to b$. Then $a = b$. 

19
PROOF. Fix $\epsilon > 0$. We know that there is $N_1$ so that $|a_n - a| < \frac{\epsilon}{2}$ for all $n > N_1$, and there is $N_2$ so that $|a_n - b| < \frac{\epsilon}{2}$ for all $n > N_2$. Thus, for any $n > \max\{N_1, N_2\}$, we have

$$|a - b| = |a - a_n + a_n - b| \leq |a - a_n| + |a_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$ 

Now, suppose that $a \neq b$. Thus $a - b \neq 0$, which means that $|a - b| > 0$. So we can take $\epsilon = |a - b|$ above, and we find that $|a - b| < |a - b|$ — a contradiction. Hence, it must be true that $a = b$.  

REMARK 2.5. Note, in an Archimedean field, we are free to restrict $\epsilon = \frac{1}{k}$ for some $k \in \mathbb{N}$; that is, an equivalent statement of $a_n \to a$ is

Given any $k \in \mathbb{N}$, we have $|a_n - a| < \frac{1}{k}$ for almost all $n$.

In non-Archimedean fields, this does not suffice. For example, in the field $\mathbb{Q}(t)$, to show $a_n(t) \to a(t)$ it does not suffice to show that, for any $k \in \mathbb{N}$, $|a_n(t) - a(t)| < \frac{1}{k}$ for all sufficiently large $n$. Indeed, what if $a_n(t) - a(t) = \frac{1}{t}$? This does not go to 0, but it is $< \frac{1}{k}$ for all $k \in \mathbb{N}$. Similarly, the sequence $a_n = \frac{1}{n}$ diverges in a non-Archimedean field.
2. Lecture 6: January 21, 2016

Proposition 2.6. Let \( \mathbb{F} \) be a complete ordered field. Let \( (a_n) \) be a sequence in \( \mathbb{F} \), and suppose \( a_n \uparrow \) (i.e. \( a_n \leq a_{n+1} \) for all \( n \)) and bounded above. Then \( a_n \to \alpha \). Similarly, if \( b_n \downarrow \) and bounded below, then \( \beta = \inf\{b_n\} \) exists and \( b_n \to \beta \).

Proof. Since \( \mathbb{F} \) is a complete field, \( \alpha = \sup\{a_n\} \) exists in \( \mathbb{F} \). Let \( \epsilon > 0 \). Then \( \alpha - \epsilon < \alpha \), and so by definition there exists some element \( a_N \in \{a_n\} \) so that \( \alpha - \epsilon < a_N \leq \alpha \). Now, suppose \( n \geq N \); then \( a_n \leq \alpha \) of course, but also since \( a_n \uparrow \) we have \( a_n \geq a_N > \alpha - \epsilon \). Thus, we have shown that \( |a_n - \alpha| = \alpha - a_n < \epsilon \) for all \( n \geq N \), which is to say that \( a_n \to \alpha \).

The decreasing case is similar; alternatively, one can look at \( a_n = -b_n \), which is increasing and bounded above; then we have by the first part that \( -b_n = a_n \to \alpha \) where \( \alpha = \sup\{-b_n\} = -\inf\{a_n\} = -\beta \). It follows that \( b_n \to -\beta \), using the limit theorems below.

In the proposition, we needed \( (a_n) \) to be bounded (above or below); indeed, the sequence \( a_n = n \) is increasing, but not convergent. This is generally true: for any sequence to be convergent, it must be bounded (above and below). A sequence that is either increasing or decreasing is called monotone. So the proposition shows that monotone sequences either converge, or grow (in absolute value) without bound.

This gives us a new perspective on the motivating example that began our discussion of \( \sup \) and \( \inf \). Consider, again, the sets \( A = \{r \in \mathbb{Q}: r > 0, r^2 < 2\} \) and \( B = \{r \in \mathbb{Q}: r > 0, r^2 > 2\} \). We saw that the set of positive rationals is equal to \( A \cup B \), and therefore \( \sup A \) and \( \inf B \) do not exist in \( \mathbb{Q} \). Note that the sequence \( 1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots \) is in the set \( A \). We recognize the terms as the decimal approximations to \( \sqrt{2} \). This sequence looks like it’s going somewhere; but in fact the only place it can go is stuck in between \( A \) and \( B \), which is not in \( \mathbb{Q} \). The question is: why does it look like it’s going somewhere?

Definition 2.7. A sequence \( (a_n) \) in an ordered set is called Cauchy, or is said to be a Cauchy sequence, if

\[
\forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geq N \ |a_n - a_m| < \epsilon.
\]

That is: a sequence is Cauchy if its terms get and stay close to each other. That is: for any given tolerance \( \epsilon > 0 \), there is some time \( N \) after which all the terms are within distance \( \epsilon \) of \( a_N \). This notion is very close to convergence. Indeed:

Lemma 2.8. Any convergent sequence is Cauchy.

Proof. Let \( (a_n) \) be a convergent sequence, with limit \( a \). Fix \( \epsilon > 0 \), and choose \( N \) large enough so that \( |a_n - a| < \frac{\epsilon}{2} \) for \( n > N \). Then for any \( n, m \geq N \),

\[
|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a_m - a| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

Hence, \( (a_n) \) is Cauchy. \( \square \)

But the converse need not be true.

Example 2.9. In \( \mathbb{Q} \), the sequence \( 1, 1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots \) is Cauchy. Indeed, by the definition of decimal expansion, if \( a_n \) is the \( n \)-decimal expansion of a number, then \( a_{n+1} \) and \( a_n \) agree on the first \( n \) digits. This means exactly that \( |a_m - a_n| < \frac{1}{10^n} \) for any \( m > n \). So, fix \( \epsilon > 0 \). We can certainly find \( N \) so that \( \frac{1}{10^N} < \epsilon \) (since, for example, \( \frac{1}{10^{-2}} < \frac{1}{N} \)). Thus, for \( n, m > N \), we have \( |a_n - a_m| < \frac{1}{10^{\min(m,n)}} < \frac{1}{10^N} < \epsilon \).
Here are some more important facts about Cauchy sequences. Note that, by Lemma 2.8, any fact about Cauchy sequences is also a fact about convergent sequences.

**Proposition 2.10.** Let \((a_n)\) be a Cauchy sequences. Then \((a_n)\) is bounded: there is a constant \(M > 0\) so that \(|a_n| \leq M\) for all \(n\).

**Proof.** Taking \(\epsilon = 1\), it follows from the definition of Cauchy that there is some \(N \in \mathbb{N}\) so that \(|a_n - a_m| < 1\) for all \(n, m > N\). In particular, this shows that \(|a_n - a_{N+1}| < 1\) for all \(n > N\), which is to say that \(a_{N+1} - 1 < a_n < a_{N+1} + 1\). Hence \(|a_n| < \max\{|a_{N+1} - 1|, |a_{N+1} + 1|\}\) for \(n > N\). So, define \(M = \max\{|a_1|, \ldots, |a_N|, |a_{N+1} - 1|, |a_{N+1} + 1|\}\). If \(n \leq N\), then \(|a_n| \leq M\) since \(|a_n|\) appears in this list we maximize over; if \(n > N\) then, as just shown, \(|a_n| < \max\{|a_{N+1} - 1|, |a_{N+1} + 1|\}\) \(\leq M\). The result follows. □

Another useful concept when working with sequences is **subsequences**.

**Definition 2.11.** Let \(\{n_k: k \in \mathbb{N}\}\) be a set of positive integers with the property that \(n_k < n_{k+1}\) for all \(k\); that is \(n_k\) is an increasing sequence in \(\mathbb{N}\). Let \((a_n)\) be a sequence. The function \(k \mapsto a_{n_k}\) is called a **subsequence** of \((a_n)\), usually denoted \((a_{n_k})\).

**Example 2.12.**

(a) Let \(a_n = \frac{1}{n}\). Then \(a_{2n} = \frac{1}{2n}\) and \(a_{2n} = \frac{1}{2n}\) are subsequences. However

\[
b_n = \begin{cases} a_n & \text{if } n \text{ is odd} \\ a_n/2 & \text{if } n \text{ is even} \end{cases}
\]

is not a subsequence of \((a_n)\). Indeed, \(b_k = a_{n_k}\) where \((n_k)_{k=1}^\infty = (1, 1, 3, 2, 5, 3, 7, 4, 9, 5, \ldots)\), and this is not an increasing sequence of integers.

(b) Let \(a_n = (-1)^n\). Then \(a_{2n} = 1\) and \(a_{2n+1} = -1\) are subsequences.

Here is an extremely useful fact about the indices of subsequences: if \((n_k)\) is an increasing sequence in \(\mathbb{N}\), then \(n_k \geq k\) for every \(k\). (This follows by a simple induction.)

**Proposition 2.13.** Let \((a_n)\) be a sequence in an ordered set, and \((a_{n_k})\) a subsequence.

1. If \((a_n)\) is Cauchy, then \((a_{n_k})\) is Cauchy.
2. If \((a_n)\) is convergent with limit \(a\), then \((a_{n_k})\) is convergent with limit \(a\).
3. If \((a_n)\) is Cauchy, and \((a_{n_k})\) is convergent with limit \(a\), then \((a_n)\) is convergent with limit \(a\).

**Proof.** For (1): fix \(\epsilon > 0\) and let \(N \in \mathbb{N}\) be chosen so that \(|a_n - a_m| < \epsilon\) for \(n, m > N\). Then whenever \(k, \ell > N\), we have \(n_k \geq k > N\) and \(n_\ell \geq \ell > N\), so by definition \(|a_{n_k} - a_{n_\ell}| < \epsilon\). Thus \((a_n)\) is Cauchy. The proof of (2) is very similar. Item (3) is on HW3. □

Before proceeding with the theory of Cauchy sequences, here are some useful facts about convergent sequences sequences.

**Theorem 2.14.** Let \((a_n)\) and \((b_n)\) be convergent sequences in an ordered field \(\mathbb{F}\).

1. If \(a_n \leq b_n\) for all sufficiently large \(n\), then \(\lim_n a_n \leq \lim_n b_n\).
2. (Squeeze Theorem) Suppose also that \(\lim_n a_n = \lim_n b_n\). If \((c_n)\) is another sequence, and \(a_n \leq c_n \leq b_n\) for all sufficiently large \(n\), then \((c_n)\) is convergent, and \(\lim_n c_n = \lim_n a_n = \lim_n b_n\).
PROOF. Let \( a = \lim_n a_n \) and \( b = \lim_n b_n \). For (1), fix \( \epsilon > 0 \). There is \( N_a \in \mathbb{N} \) so that \( |a_n - a| < \frac{\epsilon}{2} \) for \( n > N_a \), and there is \( N_b \in \mathbb{N} \) so that \( |b_n - b| < \frac{\epsilon}{2} \) for \( n > N_b \). Thus, letting \( N = \max\{N_a, N_b\} \), we have \( a_n - a \geq -\frac{\epsilon}{2} \) and \( b_n - b \leq \frac{\epsilon}{2} \) for \( n > N \). But then

\[
a_n - b_n > a - \frac{\epsilon}{2} - b - \frac{\epsilon}{2} = a - b - \epsilon.
\]

Since \( a_n \leq b_n \) for all large \( n \), we therefore have \( 0 \geq a_n - b_n > a - b - \epsilon \) for such \( n \), and therefore \( a - b - \epsilon < 0 \). This is true for any \( \epsilon > 0 \), and therefore \( a - b \leq 0 \), as claimed.

For (2), we have \( a = b \). Choosing \( N_a, N_b, \) and \( N \) as above, we have \(-\frac{\epsilon}{2} < a_n - a \leq c_n - a \leq b_n - a < \frac{\epsilon}{2} \) for all \( n \geq N \). That is: \( |c_n - a| < \frac{\epsilon}{2} < \epsilon \) for all \( n \geq N \). This shows \( c_n \to a \), as claimed. \( \square \)

Cauchy sequences give us a way of talking about completeness that is not so wrapped up in the order properties. As discussed in Example 2.9 last lecture, the “hole” in \( \mathbb{Q} \) where \( \sqrt{2} \) should be is the limit of a sequence in \( \mathbb{Q} \) which is Cauchy, but does not converge in \( \mathbb{Q} \). Instead of filling in the holes by demanding bounded nonempty sets have suprema, we could instead demand that Cauchy sequences have limits.

**Definition 2.15.** Let \( S \) be an ordered set. Call \( S \) **Cauchy complete** if every Cauchy sequence in \( S \) actually converges in \( S \).

\( \mathbb{Q} \) is not Cauchy complete. But, as we will see, \( \mathbb{R} \) is. In fact, Cauchy completeness is equivalent to the least upper bound property in any Archimedean field. We can prove half of this assertion now.
### 3. Lecture 7: January 26, 2016

**Theorem 2.16.** Let $\mathbb{F}$ be an Archimedean field. If $\mathbb{F}$ is Cauchy complete, then $\mathbb{F}$ has the nested intervals property and hence is complete in the sense of Definition 1.15.\[1.15\]

**Proof.** That the nested intervals property implies the least upper bound property is the content of HW2 Exercise 3; so it suffices to verify that $\mathbb{F}$ has the nested intervals property. Let $(a_n)$ and $(b_n)$ be sequences in $\mathbb{F}$ with $a_n \uparrow, b_n \downarrow, a_n \leq b_n$, and $b_n - a_n < \frac{1}{n}$. Fix $\epsilon > 0$, and let $N \in \mathbb{N}$ be large enough that $\frac{1}{N} < \epsilon$ (here is where the Archimedean property is needed). Thus, for $n \geq N$, we have $b_n - a_n < \frac{1}{n} \leq \frac{1}{N} < \epsilon$. Then for $m, n > N$, wlog $m \geq n$, we have
\[
|a_m - a_n| = |a_m - a_n| \leq b_n - a_n < \epsilon.
\]
and so it follows that $|a_n - a_m| = a_m - a_n \leq b_n - a_n < \epsilon$. Thus $(a_n)$ is a Cauchy sequence. By the Cauchy completeness assumption on $\mathbb{F}$, we conclude that $a = \lim_n a_n$ exists in $\mathbb{F}$.

Now, fix $n_0$, and note that since $a_n \geq a_{n_0}$ for $n \geq n_0$, Theorem 2.14(1) shows that $a = \lim_n a_n \geq a_{n_0}$ (thinking of $a_{n_0}$ as the limit of the constant sequences $(a_{n_0}, a_{n_0}, \ldots)$). Similarly, since $a_n \leq b_{n_0}$ for all $n$, it follows that $a \leq b_{n_0}$. Thus $a \in \bigcap_n [a_n, b_n]$, proving this intersection is nonempty. As usual, it follows that the intersection consists only of $\{a\}$. Indeed, if $x, y \in \bigcap_n [a_n, b_n]$, without loss of generality label them so that $x \leq y$. Thus $a_n \leq x \leq y \leq b_n$ for every $n$. For given $\epsilon > 0$, choose $n$ so that $b_n - a_n < \epsilon$, then $y - x < \epsilon$. So $0 \leq y - x < \epsilon$ for all $\epsilon > 0$; it follows that $x = y$. This concludes the proof of the nested intervals property for $S$.\[2.16\]

**Remark 2.17.** The use of the Archimedean property is very subtle here. It is tempting to think that we can do without it. This is true if we replace the nested intervals property by a slightly weaker version: say an ordered $S$ satisfies the weak nested intervals property if, given $a_n \uparrow, b_n \downarrow, a_n \leq b_n$, and $b_n - a_n \to 0$, then $\bigcap_n [a_n, b_n]$ contains exactly one point. (This is weaker than the nested intervals property, because the assumption is stronger: we’re assuming $b_n - a_n \to 0$ here, while in the usual nested intervals property we assume that $b_n - a_n < \frac{1}{n}$, which does not imply $b_n - a_n \to 0$ in the non-Archimedean setting.) The trouble is: this weak nested intervals property does not imply the least upper bound property in the absence of the Archimedean property. In fact, there do exist non-Archimedean fields (which therefore do not have the least upper bound property), but are Cauchy complete. (We may explore this a little later.) This is a prime example of how counterintuitive analysis can be without the Archimedean property. Soon enough, we will once-and-for-all demand that it holds true (in the Real numbers), and dispense with these weird pathologies.

We would like to show the converse is true: that the least upper bound property implies Cauchy completeness. (This turns out to be true in any ordered set: after all, the least upper bound property implies the Archimedean property in an ordered field.) Then we could characterize the real numbers as the unique Archimedean field that is Cauchy complete. To do this, we need to dig a little deeper into the connection between limits and suprema / infima.

**Definition 2.18.** Let $S$ be an ordered set with the least upper bound property. Let $(a_n)$ be a bounded sequence in $S$. Define two new sequences from $(a_n)$:
\[
\overline{a}_k = \sup \{a_n : n \geq k\}, \quad a_k = \inf \{a_k : n \geq k\}.
\]
Since $(a_n)$ is bounded above (and nonempty), by the least upper bound property $\overline{a}_k$ exists for each $k$. Similarly, by Proposition 1.16 $a_k$ exists for each $k$.

Note that $\{a_n : n \geq k + 1\} \subseteq \{a_n : n \geq k\}$. Thus $\overline{a}_k$ is an upper bound for $\{a_n : n \geq k + 1\}$. It follows that $\overline{a}_k \geq \lim \{a_n : n \geq k + 1\}$, which is defined to be $\overline{a}_{k+1}$.\[1.16\]
This means that \( \bar{a}_k \geq \bar{a}_{k+1} \): the sequence \( \bar{a}_k \) is monotone decreasing. Similarly, the sequence \( a_k \) is monotone increasing.

By assumption, \( \{a_n\} \) is bounded. Thus there is a lower bound \( a_n \geq L \) for all \( n \). Since \( \bar{a}_1 \geq \bar{a}_k \geq a_k \geq L \) for all \( k \), the sequence \( \bar{a}_k \) is also bounded. Similarly, the sequence \( a_k \) is bounded.

Thus, \( \bar{a}_k \) is a decreasing, bounded-below sequence. By Proposition 2.6, \( \lim_{k \to \infty} \bar{a}_k \) exists, and is equal to \( \inf \{ \bar{a}_k \} \). Similarly, \( \lim_{k \to \infty} a_k \) exists, and is equal to \( \inf \{ a_k \} \).

We define
\[
\begin{align*}
\limsup_{n \to \infty} a_n &= \lim_{n \to \infty} \bar{a}_n = \lim_{k \to \infty} \sup \{ a_n : n \geq k \} = \inf \sup_{k \in \mathbb{N}} a_n \\
\liminf_{n \to \infty} a_n &= \lim_{n \to \infty} a_n = \lim_{k \to \infty} \inf \{ a_n : n \geq k \} = \sup \inf_{k \in \mathbb{N}} a_n.
\end{align*}
\]

**Example 2.19.** Let \( a_n = (-1)^n \). Note that \(-1 \leq a_n \leq 1 \) for all \( n \). Now, for any \( k \), there is some \( k' \geq k \) so that \( b_{k'} = 1 \). Thus \( \bar{b}_k = \sup_{n \geq k} a_k = 1 \). Similarly \( b_k = -1 \) for all \( k \). Thus \( \limsup_{n} b_n = 1 \) and \( \liminf_{n} b_n = -1 \).

Here are a few more examples computing \( \limsup \) and \( \liminf \).

**Example 2.20.**
1. Let \( a_n = \frac{1}{n} \). Since \( a_n \downarrow \), \( \bar{a}_k = \sup_{n \geq k} a_n = a_k = \frac{1}{k} \). Thus \( \limsup_{n} a_n = \lim_k a_k = 0 \). On the other hand, for any \( k \), \( \inf_k a_k = 0 \) (by the Archimedean property), and so \( \liminf_{n} a_n = \lim_k 0 = 0 \). In this case, the \( \limsup \) and \( \liminf \) agree.
2. Let \( b_n = \frac{(-1)^n}{n} \). Note that \(-1 \leq b_n \leq 1 \) for all \( n \), and more generally \( |b_n| \leq \frac{1}{n} \). For any \( k \), we therefore have \( \bar{b}_k = \sup \{ b_n : n \geq k \} \leq \sup \{ |b_n| : n \geq k \} = \frac{1}{k} \) and similarly \( b_k \geq -\frac{1}{k} \). Now, \( \bar{b}_k \leq b_k \) (the sup of any set is \( \geq \) its inf). Thus
\[
-\frac{1}{k} \leq \bar{b}_k \leq b_k \leq \frac{1}{k}.
\]
Since \( -\frac{1}{k} \to 0 \), it follows from the Squeeze Theorem that \( \lim_k \bar{b}_k = \lim_k b_k = 0 \). Thus \( \limsup_{n} b_n = \liminf_{n} b_n = 0 \).
3. The sequence \((1, 2, 3, 1, 2, 3, 1, 2, 3, \ldots)\) has \( \limsup = 3 \) and \( \liminf = 1 \).
4. Let \( c_n = n \). This is not a bounded sequence, so it doesn’t fit the mold for \( \limsup \) and \( \liminf \). Indeed, for any \( k \), \( \sup_{n \geq k} n \) does not exist for any \( k \), and so \( \limsup_n c_n \) does not exist. On the other hand, \( \inf_{n \geq k} a_n = k \) does exist, but this sequence is unbounded and has no limit, so \( \liminf c_n \) does not exist. This highlights the fact that we need both and upper and a lower bound in order for either \( \limsup \) or \( \liminf \) to exist.

In (1) and (2) in the example, \( \limsup \) and \( \liminf \) agree. This will always happen for a convergent sequence.

**Proposition 2.21.** Let \( (a_n) \) be a bounded sequence. Then \( \lim_n a_n \) exists iff \( \limsup_{n} a_n = \liminf_{n} a_n \), in which case all three limits have the same value.

**Proof.** Suppose that \( \limsup_{n} a_n = \liminf_{n} a_n \). Thus \( a_k \) and \( \bar{a}_k \) both converge to the same value. Since \( \bar{a}_k \leq a_k \leq \bar{a}_k \) for each \( k \), by the Squeeze Theorem, \( a_k \) also converges to this value, as claimed. Conversely, suppose that \( \lim_n a_n = a \) exists. Let \( \epsilon > 0 \), and choose \( N \in \mathbb{N} \) large enough that \( |a_n - a| < \epsilon \) for all \( n \geq N \). That is
\[
a - \epsilon < a_n < a + \epsilon, \quad n \geq N.
\]
It follows that
\[
a - \epsilon \leq \inf \hspace{1pt} a_n \leq \sup \hspace{1pt} a_n \leq a + \epsilon, \quad k \geq N
\]
which shows that both $\sigma_k$ and $\omega_k$ are in $[a - \epsilon, a + \epsilon]$ for $k \geq N$. Thus they both converge to $a$, as claimed.  

As with $\sup$ and $\inf$, there is a useful trick for transforming statements about $\limsup$ into statements about $\liminf$.

**Proposition 2.22.** Let $(a_n)$ be a bounded sequence. Then $\liminf_n (-a_n) = -\limsup_n a_n$.

**Proof.** Recall that, for any bounded set $A$, if $-A = \{-a: a \in A\}$, then $\sup(-A) = -\inf A$ and $\inf(-A) = -\sup A$. Now, let $b_n = -a_n$. Then $b_k = \inf \{b_n: n \geq k\} = \inf \{-a_n: n \geq k\} = -\sup \{a_n: n \geq k\} = -\sigma_k$. Thus

$$\liminf_{n \to \infty} b_n = \sup \{b_k: k \in \mathbb{N}\} = \sup \{-\sigma_k: k \in \mathbb{N}\} = -\inf \{\sigma_k: k \in \mathbb{N}\} = -\limsup_{n \to \infty} a_n.$$  

Here is a useful characterization of $\limsup$ and $\liminf$.

**Proposition 2.23.** Let $(a_n)$ be a bounded sequence in a complete ordered field. Denote $\bar{\sigma} = \limsup a_n$ and $\bar{\omega} = \liminf a_n$. Then $\bar{\sigma}$ and $\bar{\omega}$ are uniquely determined by the following properties: for all $\epsilon > 0$,

$$a_n \leq \bar{\sigma} + \epsilon \text{ for all sufficiently large } n, \text{ and}$$

$$a_n \geq \bar{\omega} - \epsilon \text{ for infinitely many } n,$$

and

$$a_n \leq \bar{\omega} + \epsilon \text{ for infinitely many } n, \text{ and}$$

$$a_n \geq \bar{\sigma} - \epsilon \text{ for all sufficiently large } n.$$  

**Proof.** This is an exercise on HW4.  

To put this into words: there are many “approximate eventual upper bounds” for the sequence: numbers $a$ large enough that the sequence eventually never gets much bigger than $a$. The $\limsup$, $\bar{\sigma}$, is the smallest approximate eventual upper bound: it is the unique number that the sequence eventually never strays far above, but also regularly gets close to from below. Similarly, the $\liminf$, $\bar{\omega}$, is the largest approximate eventual lower bound.
This brings us to an important understanding of \( \limsup \) and \( \liminf \): they are the maximal and minimal subsequential limits.

**Theorem 2.24.** Let \( (a_n) \) be a bounded sequence in a complete ordered field. There exists a subsequence of \( (a_n) \) that converges to \( \limsup_n a_n \), and there exists a subsequence of \( (a_n) \) that converges to \( \liminf_n a_n \). Moreover, if \( (b_k) \) is any convergent subsequence of \( (a_n) \), then

\[
\liminf_{n \to \infty} a_n \leq \lim_{k \to \infty} b_k \leq \limsup_{n \to \infty} a_n.
\]

**Proof.** Let \( \overline{a} = \limsup_n a_n \). By Proposition 2.23, for any \( k \in \mathbb{N} \) there are infinitely many \( n \) so that \( a_n \geq \overline{a} - \frac{1}{k} \). So, we proceed inductively: choose some \( n_1 \) so that \( a_{n_1} \geq \overline{a} - 1 \). Then, since there are infinitely many of them, we can find some \( n_2 > n_1 \) so that \( a_{n_2} \geq \overline{a} - \frac{1}{2} \). Proceeding, we find an increasing sequence \( n_1 < n_2 < \cdots < n_k < \cdots \) so that \( a_{n_k} \geq \overline{a} - \frac{1}{k} \) for each \( k \in \mathbb{N} \). We therefore have

\[
\overline{a} - \frac{1}{k} \leq a_{n_k} \leq \sup_{m \geq n_k} a_m = \overline{a}_{n_k}.
\]

Note that \( (\overline{a}_{n_k}) \) is a subsequence of \( (\overline{a}_n) \) which converges to \( \overline{a} \); thus, by Proposition 2.13, \( \lim_{k \to \infty} \overline{a}_{n_k} = \overline{a} \). Hence, by (2.1) and the Squeeze Theorem, it follows that \( a_{n_k} \to \overline{a} \), and we have constructed the desired subsequence. The proof for \( \liminf \) is very similar; alternatively, it can be reasoned using Proposition 2.22.

Now to prove the inequalities. Let \( (b_k) \) be a subsequence, so \( b_k = a_{m_k} \) for some \( m_1 < m_2 < m_3 < \cdots \). Then

\[
a_{m_k} = \inf_{n \geq m_k} a_n \leq b_k \leq \sup_{n \geq m_k} a_n = \overline{a}_{m_k}.
\]

Thus, applying the Squeeze theorem, it follows that

\[
\liminf_{n \to \infty} a_n = \lim_{k \to \infty} a_{m_k} \leq \lim_{k \to \infty} b_k \leq \lim_{k \to \infty} \overline{a}_{m_k} = \limsup_{n \to \infty} a_n
\]

as desired.

This allows us to immediately prove our first “named theorem” in Real Analysis: the **Bolzano-Weierstrass Theorem**.

**Theorem 2.25 (Bolzano-Weierstrass).** Let \( (a_n) \) be a bounded sequence in a complete ordered field, with \( a_n \in [\alpha, \beta] \) for all \( n \). Then \( (a_n) \) possesses a convergent subsequence, with limit in \( [\alpha, \beta] \).

**Proof.** Let \( \overline{a} = \limsup_n a_n \). By Theorem 2.24, there is a subsequence \( (a_{n_k}) \) of \( (a_n) \) that converges to \( \overline{a} \). Note, then, since \( \alpha \leq a_{n_k} \leq \beta \) for all \( k \), it follows from the Squeeze Theorem that \( \alpha \leq \lim_{k \to \infty} a_{n_k} = \overline{a} \leq \beta \), concluding the proof.

This finally leads us to the converse of Theorem 2.16.

**Theorem 2.26.** Let \( \mathbb{F} \) be a complete ordered field (i.e. possessing the least upper bound property). Then \( \mathbb{F} \) is Cauchy complete.

**Proof.** Let \( (a_n) \) be a Cauchy sequence in \( \mathbb{F} \). By Proposition 2.10, \( (a_n) \) is bounded. Thus, by the Bolzano-Weierstrass theorem, there is a subsequence \( a_{n_k} \) that converges. It then follows from Proposition 2.13 that \( (a_n) \) is convergent, concluding the proof.
To summarize: we now have three equivalent characterizations of the notion of “completeness” in an Archimedean field:

least upper bound property ⇐⇒ nested intervals property ⇐⇒ Cauchy completeness.

We also know, by the half of Theorem 1.20, we’ve proved, that such a field is unique. So, to finally prove the existence of $\mathbb{R}$, it will suffice to give a construction of a Cauchy complete field that is Archimedean. The supplementary notes “Construction of $\mathbb{R}$” describe how this is done in gory detail.

Henceforth, we will deal with the field $\mathbb{R}$, which satisfies all of the three equivalent completeness properties.

Now comfortably working in $\mathbb{R}$, let us state a few more (standard) limit theorems.

**Theorem 2.27 (Limit Theorems).** Let $(a_n)$ and $(b_n)$ be convergent sequences in $\mathbb{R}$, with $a_n \to a$ and $b_n \to b$.

1. The sequence $c_n = a_n + b_n$ converges to $a + b$.
2. The sequence $d_n = a_nb_n$ converges to $ab$.
3. If $b \neq 0$, then $b_n \neq 0$ for almost all $n$, and $e_n = \frac{a_n}{b_n}$ converges to $\frac{a}{b}$.

**Proof.** For (1), choose $N_a, N_b \in \mathbb{N}$ so that $|a_n - a| < \frac{\epsilon}{2}$ if $n \geq N_a$ and $|b_n - b| < \frac{\epsilon}{2}$ for $n \geq N_b$. For any $n \geq N = \max\{N_a, N_b\}$, we then have $|c_n - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$, proving that $\lim_n c_n = a + b$.

For (2), we need to be slightly more clever. Note that

$$|d_n - ab| = |a_nb_n - ab| = |a_nb_n - a_nb + a_nb - ab| \leq |a_n||b_n - b| + |a_n - a||b|.$$ 

By Proposition 2.10 there is some constant $M > 0$ so that $|a_n| \leq M$ for all $n$. So, for $\epsilon > 0$, fix $N_1$ large enough that $|b_n - b| < \frac{\epsilon}{2M}$ for all $n \geq N_1$, and fix $N_2$ large enough that $|a_n - a| < \frac{\epsilon}{2|b|}$ for all $n \geq N_2$. (If $b = 0$, we can take $N_2$ to be any number we like.) Then for $N = \max\{N_1, N_2\}$, if $n \geq N$ we have

$$|d_n - ab| \leq |a_n||b_n - b| + |a_n - a||b| < M \cdot \frac{\epsilon}{2M} + \frac{\epsilon}{2|b|} \cdot |b| = \epsilon,$$

proving that $\lim_n d_n = ab$.

For (3), first we need to show that $(e_n)$ even makes sense. Note that $e_n = \frac{a_n}{b_n}$ is not well-defined for any $n$ for which $b_n = 0$. But we’re only concerned about tails of sequences for limit statements, so once we’ve proven that $b_n \neq 0$ for almost all $n$, we know that $e_n$ is well-defined for all large $n$. For this, we use the assumption that $b \neq 0$, and so $|b| > 0$. Since $\lim_n b_n = b$, there is an $N_0 \in \mathbb{N}$ so that, for $n > N_0$, $|b_n - b| < \frac{|b|}{2}$; i.e. $-\frac{|b|}{2} < b_n - b < \frac{|b|}{2}$, and so $b_n < b + \frac{|b|}{2}$ and also $b_n > b - \frac{|b|}{2}$. Now, $b \neq 0$ so either $b < 0$ or $b > 0$. If $b < 0$, then $|b| = -b$ in which case $b_n < b + \frac{|b|}{2} = b - \frac{b}{2} = \frac{b}{2} < 0$; that is, for $n > N_0$, $b_n < 0$. If, on the other hand, $b > 0$, then $|b| = b$, and so $b_n > b - \frac{|b|}{2} = b - \frac{b}{2} = \frac{b}{2} > 0$; that is, for $n > N_0$, $b_n > 0$. Thus, in all cases, $b_n \neq 0$ for $n > N_0$, proving the first claim.

For the limit statement, note that $e_n = a_n \cdot \frac{1}{b_n}$. So, by (2), it suffices to show that $\frac{1}{b_n} \to \frac{1}{b}$.

Compute that

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b_n||b|}.$$
As shown above, there is \( N_0 \) so that, for \( n > N_0 \), then \( b_n > \frac{b}{2} = \frac{|b|}{2} \) if \( b > 0 \) and \( b_n < \frac{-b}{2} = -\frac{|b|}{2} \) if \( b < 0 \); i.e. this means that \( |b_n| > \frac{|b|}{2} \) for \( n > N_0 \). Hence, we have
\[
\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b_n||b|} < \frac{2|b_n - b|}{|b|^2}, \quad n > N_0.
\]

By assumption, \( b_n \to b \), and so we can choose \( N' \) large enough that \( |b_n - b| < \frac{|b|^2}{2} \varepsilon \) for \( n > N' \). Thus, letting \( N = \max\{N_0, N'\} \), we have
\[
\left| \frac{1}{b_n} - \frac{1}{b} \right| < 2 \frac{|b_n - b|}{|b|^2} < 2 \frac{|b|^2}{2} \varepsilon = \varepsilon, \quad n > N.
\]

This proves that \( \frac{1}{b_n} \to \frac{1}{b} \) as claimed. \( \square \)

One might hope that Theorem 2.27 carries over to \( \limsup \) and \( \liminf \); but this is not the case.

**Example 2.28.** Consider the sequences \( a_n = (-1)^n \) and \( b_n = -a_n = (-1)^{n+1} \). Then \( \limsup_n a_n = \limsup_n b_n = 1 \), \( \liminf_n a_n = \liminf_n b_n = -1 \), but \( a_n + b_n = 0 \) so \( \limsup_n (a_n + b_n) = \liminf_n (a_n + b_n) = 0 \). Hence, in this example we have
\[
-2 = \liminf_n a_n + \liminf_n b_n < \liminf_n (a_n + b_n) = 0 = \limsup_n (a_n + b_n) < \limsup_n a_n + \limsup_n b_n = 2.
\]

The inequalities in the example do turn out to be true in general.

**Proposition 2.29.** Let \( (a_n) \) and \( (b_n) \) be bounded sequences in \( \mathbb{R} \). The following always hold true.
\[
\liminf_{n \to \infty} (a_n + b_n) \geq \liminf_{n \to \infty} a_n + \liminf_{n \to \infty} b_n, \quad \text{and} \quad \limsup_{n \to \infty} (a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n.
\]

*If* \( a_n \geq 0 \) and \( b_n \geq 0 \) *for all sufficiently large* \( n \), *we also have the following.*
\[
\liminf_{n \to \infty} (a_n \cdot b_n) \geq \liminf_{n \to \infty} a_n \cdot \liminf_{n \to \infty} b_n, \quad \text{and} \quad \limsup_{n \to \infty} (a_n \cdot b_n) \leq \limsup_{n \to \infty} a_n \cdot \limsup_{n \to \infty} b_n.
\]

**Proof.** The proofs of the \( \limsup \) inequalities are exercises on HW4. Assuming these, the \( \liminf \) statements follow from Proposition 2.22. For example, we have
\[
\liminf_{n \to \infty} (a_n + b_n) = \liminf_{n \to \infty} [-(a_n - b_n)] = -\limsup_{n \to \infty} [(a_n) + (-b_n)].
\]

Since \( \limsup_{n \to \infty} [-(a_n) + (-b_n)] \leq \limsup_{n \to \infty} (-a_n) + \limsup_{n \to \infty} (-b_n) \) by HW4, taking negatives reverses the inequality, giving
\[
-\limsup_{n \to \infty} [-(a_n) + (-b_n)] \geq -\limsup_{n \to \infty} (-a_n) - \limsup_{n \to \infty} (-b_n).
\]

Now using Proposition 2.22 again on each term, we then have
\[
\liminf_{n \to \infty} (a_n + b_n) \geq -\limsup_{n \to \infty} (-a_n) - \limsup_{n \to \infty} (-b_n) = \liminf_{n \to \infty} a_n + \liminf_{n \to \infty} b_n
\]

as claimed. The proof of the inequality for products is very similar. \( \square \)

Let us close out our discussion (for now) of limits of real sequences with a rigorous treatment of the following special kinds of sequences.
PROPOSITION 2.30. Let $p > 0$ and $\alpha \in \mathbb{R}$.

1. \[ \lim_{n \to \infty} \frac{1}{n^p} = 0. \]
2. \[ \lim_{n \to \infty} n^{1/n} = 1. \]
3. \[ \lim_{n \to \infty} n^{1/n} = 1. \]
4. If $p > 1$ and $\alpha \in \mathbb{R}$, then \[ \lim_{n \to \infty} n^{\alpha p} = 0. \]
5. If $|p| < 1$, then \[ \lim_{n \to \infty} p^n = 0. \]

PROOF. For (1): fix $\epsilon > 0$, and choose $N \in \mathbb{N}$ large enough that $\frac{1}{N} < \epsilon^{1/p}$. Then for $n \geq N$, $\frac{1}{n} < \epsilon^{1/p}$, and so $0 < \frac{1}{n^p} = \left(\frac{1}{n}\right)^p < \epsilon$. This shows that $\frac{1}{n^p} \to 0$ as claimed.

For (2): as above, in the case $p = 1$ the sequence is constant $1^{1/n} = 1$ with limit 1. If $p > 1$, put $x_n = p^{1/n} - 1$. Since $p > 1$ we have $p^{1/n} > 1$ and so $x_n > 0$. From the binomial theorem, then,

\[ (1 + x_n)^n = \sum_{k=0}^{n} \binom{n}{k} x_n^k \geq 1 + nx_n. \]

By definition $(1 + x_n)^n = p$, and so

\[ 0 < x_n < \frac{(1 + x_n)^n - 1}{n} = \frac{p - 1}{n}. \]

Knowing that $\frac{p - 1}{n} \to 0$, it now follows from the Squeeze Theorem that $x_n \to 0$. This proves the limit in the case $p > 1$. If, on the other hand, $0 < p < 1$, then $r = \frac{1}{p} > 1$, and $p^{1/n} = \left(\frac{1}{r}\right)^{1/n} = \frac{1}{r^{1/n}}$. We have just proved that $r^{1/n} \to 1$, and so it follows from Theorem 2.27(3) that $p^{1/n} \to \frac{1}{1} = 1$.

For (3): we follow a similar outline. Let $x_n = n^{1/n} - 1$, which is $\geq 0$ (and $> 0$ for $n > 1$). We use the binomial theorem again, this time estimating with the quadratic term:

\[ n = (1 + x_n)^n = \sum_{k=0}^{n} \binom{n}{k} x_n^k \geq \binom{n}{2} x_n^2 = \frac{n(n-1)}{2} x_n^2. \]

Thus, we have (for $n \geq 2$)

\[ 0 \leq x_n \leq \sqrt{\frac{2}{n-1}} \]

and by the Squeeze Theorem $x_n \to 0$.

For (4): Choose a positive integer $\ell > \alpha$. Let $p = 1 + r$, so $r > 0$. Applying the binomial theorem again, when $n > \ell$ we have

\[ p^n = (1 + r)^n = \sum_{k=0}^{n} \binom{n}{k} r^k > \binom{n}{\ell} r^\ell = \frac{n(n-1) \cdots (n-\ell+1)}{\ell!} r^\ell. \]

Now, if we choose $n \geq 2\ell$, each term $n - \ell + j \geq \frac{n}{2}$ for $1 \leq j \leq \ell$, and so in this range

\[ p^n > \frac{1}{\ell!} \left(\frac{n}{2}\right)^\ell r^\ell. \]

Hence, for $n \geq 2\ell$, we have

\[ \frac{n^\alpha}{p^n} < \frac{n^\alpha}{\ell!2^\ell n^\ell r^\ell} = \ell!2^\ell \frac{r^\ell}{p^\ell} \cdot n^{\alpha-\ell}. \]
This is a constant \( \frac{\ell n^\alpha}{r} \) times \( n^{\alpha-\ell} \), where \( \alpha - \ell < 0 \); applying part (1) with \( p = \alpha - \ell \) proves the result.

Finally, for (5): the special case of (4) with \( \alpha = 0 \) yields \( \frac{1}{r^n} \to 0 \) when \( r > 1 \). Thus, with \( |p| < 1 \), setting \( r = \frac{1}{|p|} \) gives us \( |p^n| = |p|^n \to 0 \). The reader should prove (if they haven’t already) that \( |a_n| \to 0 \) iff \( a_n \to 0 \), so it follows that \( p^n \to 0 \) as claimed. \( \square \)
CHAPTER 3

Extensions of $\mathbb{R}$: the Extended Real Numbers $\overline{\mathbb{R}}$ and the Complex Numbers $\mathbb{C}$


Now that we have a good understanding of real numbers, it is convenient to extend them a little bit to give us language about certain kinds of divergent sequences.

**Definition 3.1.** Let $(a_n)$ be a sequence in $\mathbb{R}$. Say that $a_n$ **diverges to** $+\infty$ or $a_n \to +\infty$ if

$$\forall M > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N \ a_n > M.$$ 

That is: no matter how large a bound $M$ we choose, it is a lower bound for $a_n$ for all sufficiently large $n$. Similarly, we say that $a_n$ **diverges to** $-\infty$ if $-a_n \to +\infty$; this is equivalent to

$$\forall M > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N \ a_n < -M.$$ 

The expressions $a_n \to \pm \infty$ are also sometimes written as $\lim_{n \to \infty} a_n = \pm \infty$, and accordingly it is sometimes pronounced as $a_n$ **converges to** $\pm \infty$.

**Example 3.2.** The sequence $a_n = n^p$ diverges to $+\infty$ for any $p > 0$. Indeed, fix a large $M > 0$. Then $M^{1/p} > 0$, and by the Archimedean property there is an $N \in \mathbb{N}$ with $N > M^{1/p}$. Thus, for $n \geq N$, $n > M^{1/p}$, and so $a_n = n^p > (M^{1/p})^p = M$, as desired.

On the other hand, the sequence $(a_n) = (1, 0, 2, 0, 3, 0, 4, 0, \ldots)$ does not diverge to $+\infty$: no matter how large $N$ is, there is some integer $n \geq N$ with $a_n = 0$. (Indeed, we can either take $n = N$ or $n = N + 1$.) This sequence diverges, but it does not diverge to $+\infty$.

This suggests that we include the symbols $\pm \infty$ in the field $\mathbb{R}$. We must be careful how to do this, however. We have already proved that $\mathbb{R}$ is the **unique** complete ordered field, so no matter how we add $\pm \infty$, the resulting object cannot be a complete ordered field. In fact, it won’t be a field at all, for we won’t always be able to do algebraic operations.

**Definition 3.3.** Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. We make $\overline{\mathbb{R}}$ into an ordered set as follows: given $x, y \in \overline{\mathbb{R}}$, if in fact $x, y \in \mathbb{R}$ then we use the order relation from $\mathbb{R}$ to compare $x, y$. If one of the two (say $x$) is in $\mathbb{R}$, then we declare $-\infty < x < +\infty$. Finally, we declare $-\infty < +\infty$.

We make the following conventions. If $a \in \overline{\mathbb{R}}$ with $a > 0$, then $\pm \infty \cdot a = a \cdot \pm \infty = \pm \infty$; if $a \in \overline{\mathbb{R}}$ with $a < 0$ then $\pm \infty \cdot a = a \cdot \pm \infty = -\infty$. We also declare that $a + (\pm \infty) = \pm \infty$ for any $a \in \overline{\mathbb{R}}$, and that $(+\infty) + (+\infty) = +\infty$ while $(-\infty) + (-\infty) = -\infty$. We leave all the following expressions **undefined**:

$$(+\infty) + (-\infty), (-\infty) + (+\infty), \frac{\infty}{\infty}, 0 \cdot (\pm \infty), \text{ and } (\pm \infty) \cdot 0.$$
**Example 3.4.** Let $\alpha, \beta \in \mathbb{R}$ with $\alpha > 0$, and let $a_n = n$ while $b_n = -\alpha n + \beta$. Then

$$\lim_{n \to \infty} (a_n + b_n) = \begin{cases} +\infty, & \text{if } \alpha < 1 \\ \beta, & \text{if } \alpha = 1 \\ -\infty, & \text{if } \alpha > 1 \end{cases}$$

Hence the value of the limit of the sum depends on the value of $\alpha$. However, Example 3.2 shows that $a_n \to +\infty$ while a similar argument shows that $b_n \to -\infty$ for any $\alpha, \beta$. So we have to have

$$\lim_{n \to \infty} a_n + \lim_{n \to \infty} b_n = " = " (+\infty) + (-\infty).$$

This highlights why it is important to leave such expressions undefined: there is no way to consistently define them that respects the limit theorems.

We can also use these conventions to extend the notions of $\sup$ and $\inf$ to unbounded sets, and the notions of $\lim \sup$ and $\lim \inf$ to unbounded sequences.

**Definition 3.5.** Let $E \subseteq \mathbb{R}$ be any nonempty subset. If $E$ is not bounded above, declare $\sup E = +\infty$; if $E$ is not bounded below, declare $\inf E = -\infty$. We also make the convention that $\sup(\emptyset) = -\infty$ while $\inf(\emptyset) = +\infty$. (Note: this means that, in the one special case $E = \emptyset$, it is not true that $\inf E \leq \sup E$.)

Similarly, let $(a_n)$ be any sequence in $\mathbb{R}$. If $(a_n)$ is not bounded above, declare $\lim \sup_n a_n = +\infty$; if $(a_n)$ is not bounded below, declare $\lim \inf_n a_n = -\infty$.

With these conventions, essentially all of the theorems involving limits extend to unbounded sequences.

**Proposition 3.6.** Using the preceding conventions, Lemma 2.4, Proposition 2.6, Proposition 2.13(2), Squeeze Theorem 2.14, Proposition 2.21, Proposition 2.22, and Theorem 2.24 all generalize to the cases where the limits in the statements are allowed to be in $\mathbb{R}$ rather than just $\mathbb{R}$. Moreover, Theorem 2.27 and Proposition 2.29 also hold in this more general setting whenever the statements make sense: i.e. excluding the cases when the involved expressions are undefined (like $(+\infty) + (-\infty)$).

**Proof.** It would take many pages to prove all of the special cases of all of these results remain valid in the extended reals. Let us choose just one to illustrate: Theorem 2.27(1): if $\lim_n a_n = a$ and $\lim_n b_n = b$, then $\lim_n (a_n + b_n) = a + b$. We already know this holds true when $a, b \in \mathbb{R}$. If $a = +\infty$ and $b = -\infty$, or $a = -\infty$ and $b = +\infty$, the sum $a + b$ is undefined, and so we exclude these cases from the statement of the theorem. So we only need to consider the cases that $a \in \mathbb{R}$ and $b = \pm \infty$, $a = \pm \infty$ and $b \in \mathbb{R}$, or $a = b = \pm \infty$.

- $a \in \mathbb{R}$ and $b = +\infty$: Since $(a_n)$ is convergent in $\mathbb{R}$, it is bounded; thus say $|a_n| \leq L$. Then fix $M > 0$ and choose $N$ so that $b_n > M + L$ for $n \geq N$. Thus $a_n + b_n > -L + (M + L) = M$ for $n \geq N$, and so $a_n + b_n \to +\infty$. The argument is similar when $b = -\infty$.
- $a = \pm \infty$ and $b \in \mathbb{R}$: this is the same as the previous case, just reverse the roles of $a_n$ and $b_n$ and $a$ and $b$.
- $a = b = +\infty$: let $M > 0$, and choose $N_1$ so that $a_n > M/2$ for $n \geq N_1$; choose $N_2$ so that $b_n > M/2$ for $n \geq N_2$. Thus, for $n \geq N = \max\{N_1, N_2\}$, it follows that $a_n + b_n \geq M/2 + M/2 = M$, proving that $a_n + b_n \to +\infty$ as required. The argument when $a = b = -\infty$ is very similar.
Now we turn to a very different extension of \( \mathbb{R} \): the Complex Numbers. We’ve already discussed them a little bit, in Example 1.12(3-3.5) and HW1.4, so we’ll start by reiterating that discussion. We will rely on our knowledge of linear algebra.

**Definition 3.7.** Let \( \mathbb{C} \) denote the following set of \( 2 \times 2 \) matrices over \( \mathbb{R} \):

\[
\mathbb{C} = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a, b \in \mathbb{R} \right\}.
\]

Then \( \mathbb{C} = \text{span}_\mathbb{R} \{I, J\} \), where

\[
I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.
\]

As is customary, we denote \( I = 1 \) and \( J = i \). We can compute that \( J^2 = -I \), so \( i^2 = -1 \). Every complex number has the form \( a1 + bi \) for unique \( a, b \in \mathbb{R} \); we often suppress the 1 and write this as \( a + ib \). We think of \( \mathbb{R} \subset \mathbb{C} \) via the identification \( a \leftrightarrow a + i0 \) (so \( a \) is the matrix \( aI \)).

It is convenient to construct \( \mathbb{C} \) this way, since, as a collection of matrices, we already have addition and multiplication built in; and we have all the tools of linear algebra to prove properties of \( \mathbb{C} \).

**Proposition 3.8.** Denote \( 1_\mathbb{C} = I \) and \( 0_\mathbb{C} \) the \( 2 \times 2 \) zero matrix. Define + and \( \cdot \) on \( \mathbb{C} \) by their usual matrix definitions. Then \( \mathbb{C} \) is a field.

**Proof.** Most of the work is done for us, since + and \( \cdot \) of matrices are associative and distributive, and + is commutative, and \( 1_\mathbb{C} \) and \( 0_\mathbb{C} \) are multiplicative and additive identities. All that we are left to verify are the following three properties:

- \( \mathbb{C} \) is closed under + and \( \cdot \), i.e. we need to check that if \( z, w \in \mathbb{C} \), then \( z + w \in \mathbb{C} \) and \( z \cdot w \in \mathbb{C} \). Setting \( z = a + ib \) and \( w = c + id \), we simply compute

\[
z + w = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} + \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = \begin{bmatrix} a + c & -b - d \\ b + d & a + c \end{bmatrix} = (a + c) + (b + d)i \in \mathbb{C},
\]

\[
z \cdot w = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = \begin{bmatrix} ac - bd & -ad - bc \\ ad + bc & ac - bd \end{bmatrix} = (ac - bd) + (ad + bc)i \in \mathbb{C}.
\]

- \( \cdot \) is commutative: this follows from the computation above: if we exchange \( z \leftrightarrow w \), meaning \( a \leftrightarrow c \) and \( b \leftrightarrow d \), the value of the product \( z \cdot w \) is unaffected, so \( z \cdot w = w \cdot z \).

- If \( z \in \mathbb{C} \setminus \{0_\mathbb{C}\} \) then \( z^{-1} \) exists: here we use the criterion that a matrix \( z \) is invertible iff \( \det(z) \neq 0 \). We can readily compute that, for \( z \in \mathbb{C} \),

\[
\det(z) = \det \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a^2 + b^2
\]

and this \( = 0 \) iff \( a = b = 0 \) meaning \( a + ib = 0_\mathbb{C} \).

Now more notation.

**Definition 3.9.** Let \( z = a + ib \in \mathbb{C} \). We denote \( a = \Re(z) \) and \( b = \Im(z) \), the **Real** and **Imaginary** parts of \( z \). Define the **modulus** or **absolute value** of \( z \) to be

\[
|z| = \sqrt{\det(z)} = \sqrt{a^2 + b^2} = \sqrt{\Re(z)^2 + \Im(z)^2}.
\]
For \(z \in \mathbb{C}\), its complex conjugate \(\overline{z}\) is the complex number \(\overline{z} = \Re(z) - i\Im(z)\); in terms of matrices, this is just the transpose \(\overline{z} = z^\top\).

Note that \(z + \overline{z} = 2\Re(z)\) and \(z - \overline{z} = 2i\Im(z)\). Since \(i\) is invertible (indeed \(i^{-1} = -i\)), it follows that

\[
\Re(z) = \frac{z + \overline{z}}{2}, \quad \Im(z) = \frac{z - \overline{z}}{2i}.
\] (3.1)

Note that, if \(z \in \mathbb{C}\) happens to be in \(\mathbb{R}\) (meaning that \(\Im(z) = 0\) so \(z = \Re(z)\)), then \(|z| = \sqrt{\Re(z)^2 + 0} = \sqrt{z^2} = |z|\) corresponds to the absolute value in \(\mathbb{R}\); so the complex modulus generalizes the familiar absolute value.
Here are some important properties of modulus and complex conjugate.

**Lemma 3.10.** Let \( z, w \in \mathbb{C} \). Then we have the following.

1. \( \overline{z} = z \).
2. \( z + \overline{w} = z + w \) and \( z\overline{w} = \overline{z \cdot w} \).
3. \( z\overline{z} = |z|^2 \).
4. \( |\overline{z}| = |z| \).
5. \( |zw| = |z||w| \), and so \( |z^n| = |z|^n \) for all \( n \in \mathbb{N} \).
6. \( |\Re(z)| \leq |z| \) and \( |\Im(z)| \leq |z| \).
7. \( |z + w| \leq |z| + |w| \).
8. \( |z| = 0 \) iff \( z = 0 \).
9. If \( z \neq 0 \) then \( z^{-1} \) (which we also write as \( \frac{1}{z} \)) is given by
   \[ z^{-1} = \frac{\overline{z}}{|z|^2} \]
10. If \( z \neq 0 \) then \( |z^{-1}| = |z|^{-1} \), and so \( |z^n| = |z|^n \) for all \( n \in \mathbb{Z} \).

**Proof.** (1) is the familiar linear algebra fact that \( (z^\top)^\top = z \), and (2) follows similarly from linear algebra (and the commutativity of \( \cdot \) in \( \mathbb{C} \)): \( z + \overline{w} = (z + w)^\top = z^\top + w^\top = \overline{z + w} \), and \( z\overline{w} = (zw)^\top = w^\top z^\top = \overline{w}^\top \overline{z} = \overline{z\overline{w}} \).

(3) writing \( z = a + ib \) we have
   \[ zz = (a + ib)(a - ib) = a^2 + b^2 \]
   \( (4) \) then follows that from this and (1), and commutativity of complex multiplication: \( |\overline{z}|^2 = \overline{zz} = \overline{zz} = |z|^2 \); taking square roots (using the fact that \( |z| \geq 0 \)) shows that \( |\overline{z}| = |z| \).

(5) is a well-known property of determinants: \( |zw| = \det(zw) = \det(z) \det(w) = |z||w| \); taking \( z = w \) and doing induction shows that \( |z^n| = |z|^n \). (6) follows easily from the fact that \( |z| = \sqrt{\Re(z)^2 + \Im(z)^2} \). For (7), we have
   \[ |z + w|^2 = (z + w)(\overline{z} + \overline{w}) = (z + w)(\overline{z} + \overline{w}) = z\overline{z} + z\overline{w} + w\overline{z} + w\overline{w} = |z|^2 + \overline{z}\overline{w} + w\overline{z} + |w|^2 \]
   The two middle terms can be written as \( z\overline{w} + w\overline{z} = z\overline{w} + (\overline{z}w) \) and, by (3.1), this equals \( 2\Re(z\overline{w}) \).

Now, any real number \( x \) is \( \leq |x| \), and so
   \[ |z + w|^2 = |z|^2 + 2\Re(z\overline{w}) + |w|^2 \leq |z|^2 + 2|\Re(z\overline{w})| + |w|^2 \leq |z|^2 + 2|z\overline{w}|^2 + |w|^2 \]
   where we have used (6). From (4) and (5), \( |zw| = |z||w| = |z||w| \), and so finally we have
   \[ |z + w|^2 \leq |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2 \]
   Taking square roots proves the result.

For (8), it is immediate that \( |0| = 0 \); the converse was shown in the proof of Proposition 3.8: \( |z| = \det(z) = 0 \) iff \( (\Re(z))^2 + (\Im(z))^2 = 0 \) which happens only when \( \Re(z) = \Im(z) = 0 \), so \( z = 0 \). Part (9) follows similarly from the matrix representation; alternatively we can simply check from (3) that
   \[ z \cdot \frac{\overline{z}}{|z|^2} = \frac{z\overline{z}}{|z|^2} = \frac{|z|^2}{|z|^2} = 1 \]
   showing that \( z^{-1} = \frac{\overline{z}}{|z|^2} \). Finally, for (10), using (5) we have
   \[ |z^{-1}| = |z|^{-1} = 1 \]
so \(|z|^{-1} = |z^{-1}|\). An induction argument combining this with (5) shows that \(|z|^{-n} = |z^{-n}|\) for \(n \in \mathbb{N}\), and coupling this with the second statement of (5) concludes the proof.

Items (7) and (8) of Lemma 3.10 show that the complex modulus behaves just like the real absolute value: it satisfies the triangle inequality, and is only 0 at 0. These properties are all that were necessary to make most of the technology of limits of sequences in \(\mathbb{R}\) work, and so we can now use the complex modulus to extend these notions to \(\mathbb{C}\).

**Definition 3.11.** Let \((z_n)\) be a sequence in \(\mathbb{C}\). Given \(z \in \mathbb{C}\), say that

\[
\lim_{n \to \infty} z_n = z \text{ iff } \forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N \ |z_n - z| < \epsilon.
\]

Say that \((z_n)\) is a Cauchy sequence if

\[
\forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geq N \ |z_n - z_m| < \epsilon.
\]

Note that these are, symbolically, exactly the same as the definitions (6.1 and 2.7) of limits and Cauchy sequences of real numbers; the only difference is, we now interpret \(|z|\) to mean the modulus of the complex number \(z\) rather than the absolute value of a real number.

The properties of complex modulus mirroring those of real absolute value allow us to prove the results of Lemmas 2.4 and 2.8, Propositions 2.10 and 2.13, and the Limit Theorems (Theorem 2.27) with nearly identical proofs. To summarize:

**Theorem 3.12.** (1) Limits are unique: if \(z_n \to z\) and \(z_n \to w\), then \(z = w\).

(2) Every convergent sequence in \(\mathbb{C}\) is Cauchy.

(3) Every Cauchy sequence in \(\mathbb{C}\) is bounded.

(4) (a) If \(z_n \to z\) then any subsequence of \((z_n)\) converges to \(z\).

(b) If \((z_n)\) is Cauchy then any subsequence of \((z_n)\) is Cauchy.

(c) If \((z_n)\) is Cauchy and has a convergent subsequence with limit \(z\), then \(z_n \to z\).

(5) If \(z_n \to z\) and \(w_n \to w\), then \(z_n + w_n \to z + w\), \(z_n w_n \to zw\), and if \(z \neq 0\) then \(z_n \neq 0\) for sufficiently large \(n\) and \(1/z_n \to 1/z\).

To illustrate how to handle complex modulus in these proofs, let us look at the analog of Proposition 2.10 that Cauchy sequences are bounded. As before, we set \(\epsilon = 1\) and let \(N\) be large enough that \(|z_n - z_m| < 1\) whenever \(n, m > N\). Thus, taking \(m = N + 1\), for any \(n > N\) we have \(|z_n - z_{N+1}| < 1\). Now, \(z_n = (z_n - z_{N+1}) + z_{N+1}\), and so by the triangle inequality

\[
|z_n| = |(z_n - z_{N+1}) + z_{N+1}| \leq |z_n - z_{N+1}| + |z_{N+1}| < 1 + |z_{N+1}|, \quad \forall n > N.
\]

So, as in the previous proof, if we set \(M = \max\{|z_1|, |z_2|, \ldots, |z_N|, 1 + |z_{N+1}|\}\) then \(|z_n| \leq M\) for all \(n\).

In fact, convergence and Cauchy-ness of complex sequences boils down to convergence and Cauchy-ness of the real and imaginary parts separately.

**Proposition 3.13.** Let \((z_n)\) be a sequence in \(\mathbb{C}\). Then \((z_n)\) is Cauchy iff the two real sequences \((\Re(a_n))\) and \((\Im(b_n))\) are Cauchy, and \(z_n \to z\) iff \(\Re(z_n) \to \Re(z)\) and \(\Im(z_n) \to \Im(z)\).

**Proof.** Let \(z_n = a_n + ib_n\). Suppose \((a_n)\) and \((b_n)\) are Cauchy. Fix \(\epsilon > 0\) and choose \(N_1\) large enough that \(|a_n - a_m| < \frac{\epsilon}{2}\) for \(n, m > N_1\), and choose \(N_2\) large enough that \(|b_n - b_m| < \frac{\epsilon}{2}\) for
n, m > N_2. Then for n, m > N = \max\{N_1, N_2\}, we have
\[ |z_n - z_m| = |(a_n + i b_n) - (a_m + i b_m)| = |(a_n - a_m) + i(b_n - b_m)| \leq |a_n - a_m| + |i(b_n - b_m)| \]
\[ = |a_n - a_m| + |i(b_n - b_m)| \]
\[ < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \]

where, in the second last step, we used the fact that \[ |i(b_n - b_m)| = |i||b_n - b_m| \text{ and } |i| = 1 \]. Thus, \( (z_n) \) is Cauchy. For the converse, suppose that \( (z_n) \) is Cauchy. Fix \( \epsilon > 0 \), and choose \( N \) large enough that \( |z_n - z_m| < \epsilon \) for \( n, m > N \). Then we also have
\[ |\Re(z_n) - \Re(z_m)| = |\Re(z_n - z_m)| \leq |z_n - z_m| < \epsilon, \quad \text{and} \]
\[ |\Im(z_n) - \Im(z_m)| = |\Im(z_n - z_m)| \leq |z_n - z_m| < \epsilon \]
for \( n, m > N \). Thus, both \( (\Re(z_n)) \) and \( (\Im(z_n)) \) are Cauchy, as claimed.

The proof of the limit statements is very similar, and is left as a homework exercise (on HW6).

Now, \( \mathbb{C} \) is not an ordered field (you proved this on HW1), so it does not even make sense to ask if it has the least upper bound property (and likewise we cannot talk about a Squeeze Theorem, or \( \lim \sup \) and \( \lim \inf \)). This is one of the main reasons we gave an equivalent characterization of the least upper bound property – Cauchy completeness – that does not explicitly require an order relation.

**Theorem 3.14.** The field \( \mathbb{C} \) is Cauchy complete: any Cauchy sequence is convergent.

**Proof.** Let \( (z_n) \) be a Cauchy sequence in \( \mathbb{C} \). By Proposition 3.13, the two real sequences \( (\Re(z_n)) \) and \( (\Im(z_n)) \) are both Cauchy. Since \( \mathbb{R} \) is Cauchy complete, it follows that there are real numbers \( a, b \in \mathbb{R} \) so that \( \Re(z_n) \to a \) and \( \Im(z_n) \to b \). It then follows, again by Proposition 3.13, that \( z_n \to a + ib \).

In \( \mathbb{R} \), we proved the Bolzano-Weierstrass theorem (that bounded sequences have convergent subsequences) using the technology of \( \lim \sup \) and \( \lim \inf \). As noted, since \( \mathbb{C} \) is not ordered, there is no way to talk about \( \lim \sup \) and \( \lim \inf \) for a complex sequence. Nevertheless, the Bolzano-Weierstrass theorem holds true in \( \mathbb{C} \). We conclude our discussion of \( \mathbb{C} \) (for now) with its proof.

**Theorem 3.15 (Bolzano-Weierstrass).** Every bounded sequence in \( \mathbb{C} \) contains a convergent subsequence.

**Proof.** Let \( (z_n) \) be a bounded sequence. Letting \( z_n = a_n + ib_n \), since \( |a_n| \leq |z_n| \) and \( |b_n| \leq |z_n| \), it follows that \( (a_n) \) and \( (b_n) \) are bounded sequences in \( \mathbb{R} \). Now, by the Bolzano-Weierstrass theorem for \( \mathbb{R} \), there is a subsequence \( a_{n_k} \) of \( (a_n) \) that converges to some real number \( a \). Consider now the subsequence \( b_{n_k} \) of \( (b_n) \). Since \( (b_n) \) is bounded, so is \( (b_{n_k}) \), and so again applying the Bolzano-Weierstrass theorem for \( \mathbb{R} \), there is a further subsequence \( (b_{n_{k_{\ell}}}) \) that converges to some \( b \in \mathbb{R} \). The subsequence \( (a_{n_{k_\ell}}) \) is a subsequence of the convergent sequence \( a_{n_k} \) and hence also converges to \( a \). Thus, by Proposition 3.13, the subsequence \( z_{n_{k_\ell}} \) converges to \( a + ib \) as \( \ell \to \infty \).

**Remark 3.16.** This proof highlights an important technique with subsequences in higher dimensional spaces. We chose the second subsequence as a subsubsequence, not only a subsequence. Had we tried to select the subsequences of the real and imaginary parts independently, we could not have concluded anything about the two together. Indeed, the Bolzano-Weierstrass theorem gives us a convergent subsequence \( a_{n_k} \) and also gives us a convergent subsequence \( b_{m_k} \). But we need to
use the same index $n$ for both $a_n$ and $b_n$, which might not be possible with independent choices like this. A priori, the chosen convergent subsequence of $a_n$ might have been $(a_1, a_3, a_5, \ldots)$, while from $b_n$ we might have chosen $(b_2, b_4, b_6, \ldots)$, ne’er the ’tween shall meet.
CHAPTER 4

Series

1. Lecture 11: February 9, 2016

We now turn to a special class of sequences called \textit{series}.

\textbf{Definition 4.1.} Let \((a_n)\) be a sequence in \(\mathbb{R}\) or \(\mathbb{C}\). The \textit{series} associated to \((a_n)\) is the new sequence \((s_n)\) given by

\[ s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n. \]

It is a bit of a misnomer to refer to series as special kinds of sequences; indeed, any sequence is the series associated to some other sequence. For let \((a_n)\) be a sequence. Define a new sequence \((b_n)\) by

\[ b_1 = a_1, \quad b_n = a_n - a_{n-1} \text{ for } n > 1. \]

Then \(a_1 = b_1 = \sum_{k=1}^{1} b_k\), and for \(n > 1\) we compute that

\[ \sum_{k=1}^{n} b_k = b_1 + b_2 + \cdots + b_k = a_1 + (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_n - a_{n-1}) = a_n. \]

Thus, \((a_n)\) (the arbitrary sequence we started with) is the series associated to the sequence \((b_n)\).

We will see, however, that the concept of convergence is quite different when applied to the series associated to a sequence rather than the sequence itself.

\textbf{Definition 4.2.} Let \((a_n)\) be a sequence in \(\mathbb{R}\) or \(\mathbb{C}\), and let \(s_n = \sum_{k=1}^{n} a_k\) be its series. We say that the \textit{series converges} if the sequence \((s_n)\) converges. If the limit is \(s = \lim_{n \to \infty} s_n\), we denote it by

\[ s = \sum_{n=1}^{\infty} a_n = \lim_{n \to \infty} \sum_{k=1}^{n} a_k. \]

In this case, we will often use the cumbersome-but-standard notation “the series \(\sum_{n=1}^{\infty} a_n\) converges”.

\textbf{Example 4.3 (Geometric Series).} Let \(r \in \mathbb{C}\), and consider the sequence \(a_n = r^n\) (in this case it is customary to start at \(n = 0\)). We can compute the terms in the series exactly, following a trick purportedly invented by Gauss at age 10.

\[ s_n = \sum_{k=0}^{n} a_k = 1 + r + r^2 + \cdots + r^n. \]

\[ \therefore \quad r s_n = r + r^2 + \cdots + r^n + r^{n+1}. \]

So, subtracting the two lines, we have

\[ (1 - r) s_n = s_n - r s_n = 1 - r^{n+1}. \]
Now, if $r = 1$, this gives no information. In that degenerate case, we simply have $s_n = 1 + 1 + \cdots + 1 = n$, and this series does not converge. In all other cases, we have the explicit formula

$$s_n = \frac{1 - r^{n+1}}{1 - r}.$$

Using the limit theorems, we can decide whether this converges, and to what, just looking at the shifted sequence $a_{n+1} = r^{n+1}$. If $|r| < 1$, then this converges to 0. If $|r| \geq 1$, this sequence does not converge. (This is something you should work out.) Thus, we have

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1 - r}, \quad \text{if } |r| < 1$$

while the series diverges if $|r| \geq 1$.

**Example 4.4.** Consider the sequence $a_n = \frac{1}{n(n+1)}$. We can employ a trick here: the partial fractions decomposition:

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Thus, taking the series $s_n = \sum_{k=1}^{n} a_n$, we have

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left( \frac{1}{k} - \frac{1}{k+1} \right) = \left( 1 - \frac{1}{2} \right) + \left( \frac{1}{2} - \frac{1}{3} \right) + \cdots + \left( \frac{1}{n} - \frac{1}{n+1} \right).$$

This is a telescoping sum: all terms except for the first and the last cancel in pairs. Thus, we have a closed formula

$$s_n = 1 - \frac{1}{n+1}.$$

Hence, the series converges, and we have explicitly

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

**Example 4.5 (Harmonic Series).** Consider the series $s_n = \sum_{k=1}^{n} \frac{1}{k}$. That is

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

If you add up the first billion terms (i.e. $s_{10^{9}}$) you get about 21.3. This seems to suggest convergence; after all, the terms are getting arbitrarily small. However, this series does not converge. To see why, look at terms $s_N$ with $N = 2^m + 2^{m-1} + \cdots + 2 + 1$ for some positive integer $m$. (By the way, from the previous example, this could be written explicitly as $N = 2^{m+1} - 1$.) Then we can group terms as

$$s_N = (1) + \left( \frac{1}{2} + \frac{1}{3} \right) + \left( \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right) + \cdots + \left( \frac{1}{2m+1} + \frac{1}{2m+2} + \cdots + \frac{1}{2m+1} - 1 \right).$$

That is: we break up the sum into $m+1$ groups, the first group with 1 term, the second with 2, the third with 4, up to the last with $2^m$ terms. Now, $1 > \frac{1}{2}$. In the second group of terms, both $\frac{1}{2}$ and $\frac{1}{3}$
are $> \frac{1}{4}$. In the next group, each of the four terms is $> \frac{1}{8}$. That is, we have

$$s_N > \left( \frac{1}{2} \right) + \left( \frac{1}{4} + \frac{1}{4} \right) + \left( \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \right) + \cdots + \left( \frac{1}{2m+1} + \frac{1}{2m+1} + \cdots + \frac{1}{2m+1} \right)$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = \frac{m+1}{2}.$$  

Now, $s_{n+1} = s_n + \frac{1}{n+1} \geq s_n$, so $(s_n)$ is an increasing sequence. We’ve just shown that, for any integer $m$, we can find some time $N$ so that $s_N > \frac{m+1}{2}$, and so it follows that for all larger $n \geq N$, $s_n \geq s_N \geq \frac{m+1}{2}$. Since $\frac{m+1}{2}$ is arbitrarily larger, we’ve just proved that $s_n \to +\infty$ as $n \to \infty$. So the series diverges.

In Example 4.3, we were able to compute the $n$th term in the series as a closed formula, and compute the limit directly. It is rare that we can do this explicitly; more often, we will need to make estimates like we did in Example 4.5. So we now begin to discuss some general tools for attacking such limits.

**Proposition 4.6 (Cauchy Criterion).** Let $a_n$ be a sequence in $\mathbb{R}$ or $\mathbb{C}$. Then the series

$$\sum_{n=1}^{\infty} a_n$$

converges if and only if: for every $\epsilon > 0$, there is a natural number $N \in \mathbb{N}$ so that, for all $m \geq n \geq N$,

$$\left| \sum_{k=n+1}^{m} a_k \right| < \epsilon.$$

**Proof.** This is just a restatement of the Cauchy completeness of $\mathbb{R}$ and $\mathbb{C}$. Let $s_n = \sum_{k=1}^{n} a_k$. Then

$$\sum_{k=n+1}^{m} a_k = a_{n+1} + \cdots + a_m = s_m - s_n.$$

Thus, having decided to always use $m$ to denote the larger of $m, n$, the statement is that, for every $\epsilon > 0$ there is $N \in \mathbb{N}$ such that, for $m \geq n \geq N$, $|s_m - s_n| < \epsilon$; this is precisely the statement that $(s_n)$ is a Cauchy sequence. In $\mathbb{R}$ or $\mathbb{C}$, this is equivalent to $(s_n)$ being convergent, as desired. $\square$

**Corollary 4.7.** Let $(a_n)$ be a sequence such that the series $\sum_{n=1}^{\infty} a_n$ converges. Then $a_n \to 0$.

**Proof.** By the Cauchy Criterion (Proposition 4.6), given $\epsilon > 0$ we may find $N \in \mathbb{N}$ so that (letting $n = m - 1$) for $m > N$,

$$\epsilon > \left| \sum_{k=(m-1)+1}^{m} a_k \right| = |a_m|.$$  

This is precisely the statement that $a_m \to 0$ as $m \to \infty$. $\square$

As Example 4.5 points out, the converse to Corollary 4.7 is false: there are sequences, such as $a_n = \frac{1}{n}$, that tend to 0, but for which the series $\sum_{n=1}^{\infty} a_n$ diverges.

It is often impossible to compute the exact value of the sum $\sum_{n=1}^{\infty} a_n$ of a convergent series. More often, we use estimates to approximate the value. More basically, we use estimates to determine whether the series converges or not, without any direct knowledge of the value if it does converge. The most basic test for convergence is the comparison theorem.

**Theorem 4.8 (Comparison).** Let $(a_n)$ and $(b_n)$ be sequences in $\mathbb{C}$. If

$$a_n \leq b_n$$

for all $n$, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.
(1) If \( b_n \geq 0 \) and \( \sum n b_n \) converges, and if \( |a_n| \leq b_n \) for all sufficiently large \( n \), then \( \sum a_n \) converges, and \( |\sum a_n| \leq \sum b_n \).

(2) If \( a_n \geq b_n \geq 0 \) for all sufficiently large \( n \) and \( \sum b_n \) diverges, then \( \sum a_n \) diverges.

**Proof.** For item 1: by assumption \( \sum b_n \) converges, and so by the Cauchy criterion, for given \( \epsilon > 0 \) we can choose \( N_0 \in \mathbb{N} \) so that, for \( m \geq n \geq N_0 \),

\[
\sum_{k=n+1}^{m} b_k < \epsilon
\]

(here we have used the fact that \( b_n \geq 0 \) to drop the modulus). Now, let \( N_1 \) be large enough that \( |a_n| \leq b_n \) for \( n \geq N_1 \). Then for \( m \geq n \geq \max\{N_0, N_1\} \), we have

\[
\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k| \leq \sum_{k=n+1}^{m} b_k < \epsilon.
\]

So the series \( \sum a_n \) satisfies the Cauchy criterion, and therefore is convergent.

Item 2 follows from item 1 by contrapositive: if \( \sum a_n \) converges, then since \( b_n = |b_n| \leq a_n \) for all large \( n \), we have just proven that \( \sum b_n \) converges. Thus, if \( \sum b_n \) diverges, so must \( \sum a_n \).

**Example 4.9.** The series \( \sum \frac{1}{\sqrt{n}} \) diverges, since \( \frac{1}{\sqrt{n}} \geq \frac{1}{n} \) and, by Example 4.5 \( \sum \frac{1}{n} \) diverges. On the other hand, note that

\[
n^2 = \frac{1}{2} n^2 + \frac{1}{2} n^2 \geq \frac{1}{2} n^2 + \frac{1}{2} n = \frac{1}{2} n(n+1)
\]

for \( n \geq 1 \). Thus \( \frac{1}{n^2} \leq \frac{2}{n(n+1)} \). As we computed in Example 4.4 \( \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1 \), so by the limit theorems, \( \sum_{n=1}^{\infty} \frac{2}{n(n+1)} = 2 \). That is: this series converges. It follows from the comparison test that \( \sum_{n=1}^{\infty} \frac{1}{n^2} \) converges.

In showing that the harmonic series diverges, we broke the terms up into groups of exponentially increasing size. This is an important trick known as the lacunary technique, and it works well when the sequence of terms is positive and decreasing.

**Proposition 4.10 (Lacunary Series).** Suppose \( (a_n) \) is a sequence of non-negative numbers that is decreasing: \( a_n \geq a_{n+1} \geq 0 \) for all \( n \). Then \( \sum_{n=1}^{\infty} a_n \) converges if and only if the series

\[
\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots
\]

converges.

**Proof.** Since \( a_k \geq 0 \) for all \( k \), the series of partial sums \( s_n = \sum_{k=1}^{n} a_k \) is monotone increasing. Hence, convergence of \( s_n \) is equivalent to the boundedness of \( (s_n) \). Let \( t_k = a_1 + 2a_2 + 4a_4 + \cdots + 2^k a_k \). We will show that \( (s_n) \) is bounded iff \( (t_k) \) is bounded.

Note that \( 2^k \leq 2^{k+1} - 1 \), and so if \( n < 2^k \) then \( n < 2^{k+1} - 1 \). Then we have for such \( n \)

\[
s_n = a_1 + a_2 + a_3 + \cdots + a_n \leq a_1 + a_2 + a_3 + \cdots + a_{2^{k+1}-1}
\]

\[
= (a_1) + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1})
\]

\[
\leq a_1 + 2a_2 + \cdots + 2^k a_{2^k} = t_k.
\]
(In the last inequality, we used the fact that \( a_n \) is decreasing.) This shows that if \((t_k)\) is bounded, then so is \((s_n)\). For the converse, we just group terms slightly differently (exactly as we did in the proof of the divergence of the harmonic series): if \(n > 2^k\), then

\[
s_n = a_1 + a_2 + a_3 + \cdots + a_n \geq a_1 + a_2 + a_3 + \cdots + a_{2^k} \\
= (a_1) + (a_2) + (a_3 + a_4) + \cdots + (a_{2^k-1} + a_{2^k}) \\
\geq \frac{1}{2}a_1 + a_2 + 2a_4 \cdots + 2^{k-1}a_{2^k} = \frac{1}{2}t_k.
\]

Thus \(t_k \leq 2s_n\) whenever \(n > 2^k\). This shows that if \((s_n)\) is bounded then so is \((t_k)\), concluding the proof.

\[\square\]

**Example 4.11.** Let \(p \in \mathbb{R}\), and consider the series \(\sum_{n=1}^{\infty} \frac{1}{n^p}\). We’ve already seen that this series diverges when \(p = 1\). If \(p < 1\), then \(\frac{1}{n^p} \geq \frac{1}{n}\); it follows by the comparison theorem that the series \(\sum_{n=1}^{\infty} \frac{1}{n^p}\) diverges for \(p \leq 1\).

On the other hand, consider \(p > 1\). Here the sequence of terms \(a_n = \frac{1}{n^p}\) is positive and decreasing, so we may use the lacunary series test to determine whether the series converges. Compute that

\[
\sum_{k=0}^{\infty} 2^k a_{2^k} = \sum_{k=0}^{\infty} 2^k \frac{1}{2^kp} = \sum_{k=0}^{\infty} 2^{(1-p)k}.
\]

This is a geometric series, with base \(r = 2^{1-p}\). So \(0 < r < 1\) provided that \(p > 1\), in which case the series converges. Hence, by Proposition 4.10, the series \(\sum_{n=1}^{\infty} \frac{1}{n^p}\) converges if and only if \(p > 1\).
Two generally effective tools for deciding convergence, that you already saw in your calculus class, are the Root Test and the Ratio Test. Both of them are predicated on rough comparison with a geometric series, cf. Example 4.3. If $a_n = r^n$, then $\sum_n a_n$ converges iff $|r| < 1$. Now, note for this series that this important constant $|r|$ can be computed either as $|a_n|^{1/n}$ or as $|a_{n+1}/a_n|$. Even when these quantities are not constant, they still can give a lot of information about the convergence of the series.

**Theorem 4.12** (Root Test). Let $(a_n)$ be a sequence in $\mathbb{C}$. Define

$$\alpha = \limsup_{n \to \infty} |a_n|^{1/n}.$$  

If $\alpha < 1$, then $\sum_n a_n$ converges. If $\alpha > 1$, then $\sum_n a_n$ diverges.

**Remark 4.13.** It is important to note that the theorem gives no information when $\alpha = 1$. Indeed, consider Examples 4.5 and 4.9 showing that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ diverges, while $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ converges. But, in both cases, we have

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

(see Theorem 3.20(c) in Rudin). Thus, $\limsup_n |a_n|^{1/n} = 1$ can happen whether $\sum_n a_n$ converges or diverges.

**Proof.** Suppose $\alpha < 1$. Then choose any $r \in \mathbb{R}$ with $\alpha < r < 1$. That is, we have $\limsup_n |a_n|^{1/n} < r$. Let $b_n = |a_n|^{1/n}$; then the statement is that $\limsup_n b_n = \liminf_n b_n < r$. That means that, for all sufficiently large $n$, $b_n < r$, and so since $b_n \leq b_n$, we have $|a_n|^{1/n} < r$ for all sufficiently large $n$. That is: there is some $N$ so that $|a_n| < r^n$ for $n \geq N$. Since the series $\sum_n r^n$ converges (as $0 < r < 1$), it now follows that $\sum_n a_n$ converges by the comparison theorem.

Now, suppose $\alpha > 1$. As above, let $b_n = |a_n|^{1/n}$. Since $\alpha = \limsup_n b_n$, from Theorem 2.24 there is a subsequence $b^{nk}_{nk}$ that converges to $\alpha$. (This is even true of $\alpha = +\infty$; in this case, it is quite easy to see that the series diverges.) In particular, this means that $b^{nk}_{nk} > 1$ for all $k$, and so $|a^{nk}_{nk}| = b^{nk}_{nk} > 1$ as well. It follows that $a_n$ does not converge to 0, and so by Corollary 4.7 $\sum_n a_n$ diverges.

The Ratio Test, which we state and prove below, is actually weaker than the Root Test. Its proof is based on comparison with the Root Test, using the following result.

**Lemma 4.14.** Let $c_n$ be a sequence of positive real numbers. Then

$$\limsup_{n \to \infty} c_n^{1/n} \leq \limsup_{n \to \infty} \frac{c_{n+1}}{c_n}, \quad \text{and}$$

$$\liminf_{n \to \infty} c_n^{1/n} \geq \liminf_{n \to \infty} \frac{c_{n+1}}{c_n}.$$  

**Proof.** We prove the lim sup inequality, and leave the similar lim inf case as an exercise. Let $\gamma = \limsup_n c_n^{1/n}$. If $\gamma = +\infty$, there is nothing to prove, since every extended real number $x$ satisfies $x \leq +\infty$. So, assume $\gamma \in \mathbb{R}$. Then we can choose some $\beta > \gamma$, and as in the proof of the Root Test above, it follows that $c_n^{1/n} < \beta$ for all sufficiently large $n$, say $n \geq N$. But then, by induction, we have

$$\frac{c_{N+k}}{c_N} = \frac{c_{N+k}}{c_{N+k-1}} \cdots \frac{c_{N+1}}{c_N} < \beta^k.$$
Thus, for \( n \geq N \), letting \( k = n - N \), we have

\[
c_n = C_{N+k} < C_N \beta^k = C_N \beta^{n-N} = (c_N \beta^{-N}) \cdot \beta^n
\]

and so

\[
c_n^{1/n} < (c_N \beta^{-N})^{1/n} \cdot \beta.
\]

From the Squeeze Theorem, it follows that

\[
\limsup_{n \to \infty} c_n^{1/n} \leq \limsup_{n \to \infty} (c_N \beta^{-N})^{1/n} \cdot \beta = \beta \cdot \lim_{n \to \infty} (c_N \beta^{-N})^{1/n} = \beta.
\]

(Here we have used the fact that \( p = c_N \beta^{-N} \) is a positive constant, and \( \lim_{n \to \infty} p^{1/n} = 1 \) for any \( p > 0 \); this last well-known limit can be found as Theorem 3.20(b) in Rudin.) Thus, for any \( \beta > \gamma \), we have \( \limsup_n c_n^{1/n} \leq \beta \). It follows that \( \limsup_n c_n^{1/n} \leq \gamma \), as claimed. \( \square \)

**Theorem 4.15 (Ratio Test).** Let \((a_n)\) be a sequence in \( \mathbb{C} \).

1. If \( \limsup_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1 \), then \( \sum_n a_n \) converges.
2. If \( \liminf_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1 \), then \( \sum_n a_n \) diverges.

**Proof.** For (1): from Lemma 4.14, \( \limsup_n |a_n|^{1/n} \leq \limsup_n \left| \frac{a_{n+1}}{a_n} \right| < 1 \), and so by the Root Test, \( \sum_n a_n \) converges. For (2): from Lemma 4.14, \( \limsup_n |a_n|^{1/n} \geq \liminf_n |a_n|^{1/n} \geq \liminf_n \left| \frac{a_{n+1}}{a_n} \right| > 1 \), and so by the Root Test, \( \sum_n a_n \) diverges. \( \square \)

**Remark 4.16.** Once again, if the \( \limsup \) or \( \liminf \) of the ratio of successive terms = 1, the test cannot give any information: letting \( a_n = \frac{1}{n^2} \) and \( b_n = \frac{1}{n^3} \), in both cases we have \( \lim_n \left| \frac{a_{n+1}}{a_n} \right| = \lim_n \left| \frac{b_{n+1}}{b_n} \right| = 1 \), and yet \( \sum_n a_n \) converges while \( \sum_n b_n \) diverges.

**Example 4.17.** Consider the sequence \((a_n) = (\frac{1}{2}, \frac{1}{3}, \frac{1}{2^3}, \frac{1}{3^2}, \ldots)\). That is: \( a_{2n-1} = \frac{1}{2^n} \) and \( a_{2n} = \frac{1}{3^n} \) for \( n \geq 1 \). Thus \( |a_{2n-1}|^{1/(2n-1)} = (\frac{1}{2})^{-n/(2n-1)} \to \frac{1}{\sqrt{2}} \) while \( |a_{2n}|^{1/2n} = (\frac{1}{3})^{n/2n} = \frac{1}{\sqrt{3}} \). Thus \( \limsup_n |a_n|^{1/n} = \frac{1}{\sqrt{2}} \), and so by the Root Test, the series \( \sum_n a_n \) converges. But the Ratio Test is no use here. Note that

\[
a_{2n-1} = \frac{1}{2^n} = \left( \frac{2}{3} \right)^n \to 0,
\]

\[
a_{2n} = \frac{1}{3^n} = \frac{1}{2^n} \to +\infty.
\]

Thus \( \limsup_n \frac{a_{n+1}}{a_n} = +\infty > 1 \) while \( \liminf_n \frac{a_{n+1}}{a_n} = 0 < 1 \); so the Ratio Test gives no information.

**Remark 4.18.** You may remember the Ratio and Root Tests as being described as equivalent in your calculus class. This is only true if you restrict to the case when \( \lim_n \left| \frac{a_{n+1}}{a_n} \right| \) exists. In this case, the limit is equal to both the \( \liminf \) and the \( \limsup \), and then Lemma 4.14 shows that \( \lim_n |a_n|^{1/n} \) also exists. But this rules out series like the one above, that somehow “alternate” between different kinds of terms, all of which are shrinking fast enough for the series to converge.

**Example 4.19 (The number \( e \)).** Consider the series

\[
\sum_{n=0}^{\infty} \frac{1}{n!}.
\]
Note that the sequence of terms $a_n$ satisfies 
\[
\frac{a_{n+1}}{a_n} = \frac{1/(n+1)!}{1/n!} = \frac{1}{n+1} \to 0 \text{ as } n \to \infty,
\]
and so by the ratio test the series converges. Its exact value is called $e$. It is sometimes called Napier's constant, since it was first alluded to in a table of logarithms in an appendix of a book written by the Scottish mathematician/natural philosopher John Napier, circa 1618. It was first directly studied by Jacob Bernoulli, who used the letter $b$ to denote it. But, like everything else from that era, it was eventually Euler who proved much of what we know about it, and Euler called it $e$.

The approximate value is
\[
e \approx 2.71828182845904523536028747135266249775724709369995.
\]

Fun fact: when Google went public in 2004, their IPO (initial public offering) was $2,718,281,828. Nerrrrrrds.

**Lemma 4.20.** The number $e$ is given by
\[
e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.
\]

**Proof.** Let $s_n = \sum_{k=0}^{n} \frac{1}{k!}$ be the $n$th partial sum of the series defining $e$. Let $t_n = \left(1 + \frac{1}{n}\right)^n$. Now, for fixed $m$, we can use the binomial theorem to expand
\[
\left(1 + \frac{1}{n}\right)^m = \sum_{k=0}^{m} \binom{m}{k} \frac{1}{n^k} = \sum_{k=0}^{m} \frac{m(m-1) \cdots (m-k+1)}{k!} \cdot \frac{1}{n^k}.
\]
Write the $k$th term as
\[
\frac{1}{k!} \cdot \frac{m}{n} \cdot \frac{m-1}{n} \cdots \frac{m-k+1}{n}.
\]
Thus, specializing to the case $m = n$, we have
\[
t_n = \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{n} \frac{1}{k!} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} < \sum_{k=0}^{n} \frac{1}{k!} = s_n.
\]
We have yet to prove that $\lim_{n \to \infty} t_n$ exists, but since $s_n$ converges to the finite number $e$, it follows from $t_n < s_n$ that $t_n$ is bounded above, and so $\limsup_{n \to \infty} t_n$ exists, and (by HW5)
\[
\limsup_{n \to \infty} t_n \leq \limsup_{n \to \infty} s_n = e.
\]
Now, on the other hand, let $m$ be fixed. Then for $n \geq m$,
\[
t_n = \sum_{k=0}^{n} \frac{1}{k!} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \geq \sum_{k=0}^{m} \frac{1}{k!} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} = t^m_n.
\]
So, for fixed $m$, the two sequences $(t_n)$ and $(t^m_n)$ are comparable: $t_n \geq t^m_n$. Again by HW5, and using the limit theorems (for the finite sum with $m$ terms), we have
\[
\liminf_{n \to \infty} t_n \geq \liminf_{n \to \infty} t^m_n = \sum_{k=0}^{m} \frac{1}{k!} \lim_{n \to \infty} \frac{n-1}{n} \cdots \frac{n-k+1}{n} = \sum_{k=0}^{m} \frac{1}{k!} = s_m.
\]
As this holds true for every $m$, it follows from the squeeze theorem that
\[
\liminf_{n \to \infty} t_n \geq \lim_{m \to \infty} s_m = e.
\]
Thus
\[
e \leq \liminf_{n \to \infty} t_n \leq \limsup_{n \to \infty} t_n \leq e
\]
which implies that the lim sup and lim inf are both equal to $e$, as claimed. \qed
The second form of $e$, as a limit, is (one of the) reason(s) it is so important: this shows that $e$ shows up in many problems related to compound interest or exponential decay. However, as a means of approximating $e$, this limit is very slow: for example

$$t_{10} \approx 2.5937 \quad (4.6\% \text{ error}), \quad t_{100} \approx 2.7048 \quad (0.50\% \text{ error}).$$

That is: you need 100 terms in order to get 2 digits of accuracy. On the other hand, $s_{10}$ is within $10^{-8}$ of $e$, and $s_{100}$ is so close to $e$ my computer cannot compute the difference. But we can give a bound on this tiny error as follows. First note that $s_n$ is increasing, so $|e - s_n| = e - s_n$. Now

$$e - s_n = \sum_{k=0}^{\infty} \frac{1}{k!} - \sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=n+1}^{\infty} \frac{1}{k!}.$$ 

Now, we factor the terms (all of which have $k \geq n + 1$) as

$$\frac{1}{k!} = \frac{1}{(n+1)!} \cdots \frac{1}{n} \frac{1}{(n+2)!} \cdots \frac{1}{(n+k)!} < \frac{1}{(n+1)!} \cdot \frac{1}{(n+1)^{k-n-1}}.$$ 

Thus

$$e - s_n < \frac{1}{(n+1)!} \sum_{k=n+1}^{\infty} \frac{1}{(n+1)^{k-n-1}} = \frac{1}{(n+1)!} \sum_{j=0}^{\infty} \frac{1}{(n+1)^j}.$$ 

This is a geometric series, and $0 < \frac{1}{n+1} < 1$, so we know the sum is

$$\sum_{j=0}^{\infty} \frac{1}{(n+1)^j} = \frac{1}{1 - \frac{1}{n+1}} = \frac{n+1}{n}.$$ 

Thus, we have our estimate:

$$e - s_n < \frac{1}{(n+1)!} \cdot \frac{n+1}{n} = \frac{1}{n! \cdot n}.$$ 

This is a tiny number. Since $10! = 3,628,800$, this shows that $e - s_{10} < \frac{1}{3 \times 10^7}$ (and in fact it’s 3 times smaller than this). For $n = 100$, we have $100! \cdot 100 \approx 10^{160}$, so $s_{100}$ differs from $e$ only after the 160th decimal digit!

This is one of the rare occasions where a perfectly practical question of error approximation actually allows us to prove something entirely theoretical.

**Proposition 4.21.** The number $e$ is irrational.

**Proof.** For a contradiction, let us suppose $e \in \mathbb{Q}$. Since $e > 0$, this means there are positive integers $m, n$ so that $e = \frac{m}{n}$. Now, from the above estimate, we have

$$0 < e - s_n < \frac{1}{n! \cdot n}, \quad \therefore \quad 0 < n!e - n!s_n < \frac{1}{n}.$$ 

Now,

$$n!s_n = n! \sum_{k=0}^{\infty} \frac{1}{k!} = n! \sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=0}^{n} n(n-1) \cdots (n-k+1) \in \mathbb{N}.$$ 

Also, by assumption $e = \frac{m}{n}$, and so $n!e = m \cdot (n-1)! \in \mathbb{N}$. Thus $\ell = n!e - n!s_n \in \mathbb{N}$. But this means $0 < \ell < \frac{1}{n}$ for some $n \in \mathbb{N}$, and that is a contradiction (there are no integers between 0 and $\frac{1}{n}$).

Moving to our final topic on the subject of series, let us consider **absolute convergence**.
**Definition 4.22.** Let \((a_n)\) be a sequence in \(C\). We say that the series \(\sum_{n=1}^{\infty} a_n\) converges **absolutely** if, in fact, \(\sum_{n=1}^{\infty} |a_n|\) converges.

**Lemma 4.23.** If \(\sum_{n=1}^{\infty} a_n\) converges absolutely, then it converges.

**Proof.** This follows immediately from the Cauchy criterion. Fix \(\epsilon > 0\) and choose \(N \in \mathbb{N}\) large enough that, for \(m > n \geq N\), \(\sum_{k=n+1}^{m} |a_k| < \epsilon\). Then by the triangle inequality

\[
\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k| < \epsilon
\]

and so \(\sum_{n=1}^{\infty} a_n\) converges. \(\Box\)
The converse of Lemma 4.23 is quite false. To see why, let us study one particular class of real series known as alternating series.

Proposition 4.24 (Alternating Series). Let \( a_n \geq 0 \) be a monotone decreasing sequence with limit \( a_n \to 0 \). Then \( \sum_{n=1}^{\infty} (-1)^{n-1}a_n = a_1 - a_2 + a_3 - a_4 + \cdots \) converges.

Proof. Fix \( m > n \in \mathbb{N} \) and consider the tail sum
\[
\left| (-1)^{n+1}a_{n+1} + (-1)^{n+2}a_{n+2} + \cdots + (-1)^{m-1}a_m \right| = |a_{n+1} - a_{n+2} + \cdots \pm a_m|.
\]
We consider two cases: either \( m - n \) is even or odd. If it is even, then we can group the terms as
\[
|a_{n+1} - a_{n+2} + \cdots + (a_{m-1} - a_m)| = (a_{n+1} - a_{n+2}) + \cdots + (a_{m-1} - a_m),
\]
where we have used the fact that \( a_n \downarrow \). On the other hand, we may group terms as
\[
= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \cdots - a_m \leq a_{n+1}.
\]
On the other hand, if \( n - m \) is odd, then by similar reasoning
\[
\sum_{k=n+1}^{m} (-1)^{k-1}a_k = (a_{n+1} - a_{n+2}) + \cdots + (a_{m-2} - a_{m-1}) + a_m
\]
and we may group this as
\[
= a_{n+1} - (a_{n+2} - a_{n+3}) - \cdots - (a_{m-1} - a_m) \leq a_{n+1}.
\]
Hence, in all cases, we have
\[
\left| \sum_{k=n+1}^{m} (-1)^{k-1}a_k \right| \leq a_{n+1}.
\]
Thus, fix \( \epsilon > 0 \). Since \( a_n \to 0 \), we may choose \( N \in \mathbb{N} \) so that, for \( n \geq N \), \( a_n = |a_n| < \epsilon \). Since
\[
a_{n+1} \leq a_n,
\]
we therefore have
\[
|\sum_{k=n+1}^{m} (-1)^{k-1}a_k| = a_{n+1} < \epsilon
\]
whenever \( m > n \geq N \), which verifies the Cauchy criterion showing that \( \sum_{n=1}^{\infty} (-1)^{n-1}a_n \) converges. \( \square \)

Example 4.25. The sequence \( a_n = \frac{1}{n} \) is positive, decreasing, and satisfies \( a_n \to 0 \). Therefore, by Proposition 4.24
\[
\sum_{n=1}^{\infty} a_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots
\]
converges. (Remembering your calculus, it in fact converges to \( \ln 2 \).) This is known as the alternating harmonic series. Note that the absolute series \( \sum_{n=1}^{\infty} \frac{1}{n} \) diverges. So this is an example of a series that is convergent but not absolutely convergent. These are sometimes called conditionally convergent series.

Conditionally convergent series have strange properties, particularly with regard to rearrangements. That is: suppose we reorder the terms. Continuing Example 4.25, rearrange the terms as follows.
\[
1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \frac{1}{7} - \cdots + \frac{1}{2n-1} - \frac{1}{4n-2} - \frac{1}{4n} + \cdots
\]
Each of the terms in the alternating harmonic series appears exactly once in this sum. It is no longer alternating, so we cannot apply a theorem to tell whether it converges; but we can in fact sum it as follows: in each three-term group, simplify
\[
\frac{1}{2n-1} - \frac{1}{4n-2} - \frac{1}{4n} = \left( \frac{1}{2n-1} - \frac{1}{4n} \right) - \frac{1}{4n-2} - \frac{1}{2} \left( \frac{1}{2n-1} - \frac{1}{2n} \right).
\]
So, the sum of the whole rearranged series is
\[
\frac{1}{2} \left( 1 - \frac{1}{2} \right) + \frac{1}{2} \left( \frac{1}{3} - \frac{1}{4} \right) + \cdots + \frac{1}{2} \left( \frac{1}{2n-1} - \frac{1}{2n} \right) + \cdots = \frac{1}{2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \frac{1}{2} \ln 2.
\]
That is: this rearrangement produces half the value of the original series!

This is always possible for a conditionally convergent series of real numbers. Riemann proved this: if \(\sum_{n=1}^{\infty} a_n\) is conditionally convergent and \(a_n \in \mathbb{R}\), then there is a rearrangement \(a'_n\) of the terms so that the sequence \(s'_n = \sum_{k=1}^{n} a'_k\) has any tail behavior possible: given any \(\alpha, \beta\) with \(-\infty \leq \alpha \leq \beta \leq +\infty\), one can find a rearrangement so that \(\limsup s'_n = \beta\) and \(\liminf s'_n = \alpha\).

(This is proved as Theorem 3.54 in Rudin.) Fortunately, this kind of craziness is not possible for absolutely convergent series, as our final theorem in this section attests to.

**Theorem 4.26.** Let \((a_n)\) be a complex sequence such that \(\sum_{n=1}^{\infty} |a_n|\) converges. Then for any rearrangement \(a'_n\) of \(a_n\), \(\sum_{n=1}^{\infty} a'_n = \sum_{n=1}^{\infty} a_n\).

**Proof.** Fix \(\epsilon\), and choose \(N \in \mathbb{N}\) so that \(\sum_{k=n+1}^{\infty} |a_k| < \epsilon\) for \(m > n \geq N\). Let \(s_n = \sum_{k=1}^{m} a_k\) and \(s'_n = \sum_{k=1}^{n} a'_k\). The numbers \(1, 2, \ldots, N\) appear as indices in the rearranged sequence \((a'_n)\) each exactly once, so there must be some finite \(p\) so that they all appear by time \(p\) in \((a'_n)\). Thus, for \(m > n \geq p\), in the difference \(s_m - s'_m\), the \(N\) terms \(a_1, \ldots, a_N\) cancel leaving only with (original) indices \(> N\). Thus, by the choice of \(N\), this difference is \(\leq |\sum_{k=N+1}^{m} a_k| \leq \sum_{k=N+1}^{m} |a_k| < \epsilon\). This shows that the sequence \(s_n - s'_n\) converges to 0, and it follows, since we know \(s_n\) converges to \(\sum_{n=1}^{\infty} a_n\), that \(s'_n\) also converges to the sum. \(\square\)
CHAPTER 5

Metric Spaces

For the remainder of this course, we are going to generalize the concepts we’ve worked with (notably convergence) beyond the case of \( \mathbb{R} \) or \( \mathbb{C} \). The key to this generalization was already discussed in the generalization from \( \mathbb{R} \) to \( \mathbb{C} \): we replace the absolute value in \( \mathbb{R} \) (defined in terms of the order relation) with the complex modulus in \( \mathbb{C} \). For all of the same technology to work, only a few basic properties of the absolute value / modulus were needed: that \(|x| \geq 0\), that \(|x| = 0\) only when \(x = 0\), and finally the triangle inequality \(|x + y| \leq |x| + |y|\).

This last property requires a notion of addition, and we’d like to move beyond vector spaces. The trick is that, in the notion of convergence, the absolute value / modulus only ever comes up as a means of measuring distance between two elements: \(|x - y|\). Thinking of it this way, what do the three key properties say?

- For any two elements \(x, y\), the distance \(|x - y|\) is \(\geq 0\).
- If the distance \(|x - y|\) is 0, then actually \(x = y\).
- The triangle inequality: for any three elements \(x, y, z\), the distance \(|x - z|\) is bounded above by \(|x - y| + |y - z|\). (This is really why it’s called the triangle inequality: draw the associated picture.) Indeed, we have

\[
|x - z| = |(x - y) + (y - z)| \leq |x - y| + |y - z|.
\]

Interpreted in this light, we don’t need a notion of addition: everything can be stated purely in terms of the notion of distance (in this case given by \((x, y) \mapsto |x - y|\)). We generalize thus.

**Definition 5.1.** Let \(X\) be a nonempty set. A function \(d: X \times X \rightarrow \mathbb{R}\) is called a metric if it satisfies the following three properties.

1. For any \(x, y \in X\), \(d(x, y) = d(y, x) \geq 0\).
2. For any \(x, y \in X\), if \(d(x, y) = 0\), then actually \(x = y\).
3. For any \(x, y, z \in X\), \(d(x, z) \leq d(x, y) + d(y, z)\).

The pair \((X, d)\) is called a metric space.

**Example 5.2.** (1) As above, if we let \(d_{\mathbb{C}}(x, y) = |x - y|\), then \((\mathbb{C}, d_{\mathbb{C}})\) is a metric space. Same goes for \(\mathbb{R}\) equipped with the restriction of \(d_{\mathbb{C}}\) to \(\mathbb{R}\).

(2) More generally, fix \(n\), and consider the set \(\mathbb{C}^n\) of \(n\)-tuples of real numbers. Define the Euclidean norm on \(\mathbb{C}^n\) as follows:

\[
\|(x_1, \ldots, x_n)\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2}.
\]

It is a simple but laborious exercise to verify that the Euclidean metric \(d_2(x, y) = \|x - y\|_2\) is a metric on \(\mathbb{C}^n\). As above, the restriction to \(\mathbb{R}^n\) is also a metric.

(3) There are many other, different metrics on \(\mathbb{R}^n\). The best known are the \(p\)-metrics: for \(1 \leq p < \infty\),

\[
\|(x_1, \ldots, x_n)\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.
\]
There is also the $\infty$-norm, aka the sup norm
\[\|(x_1, \ldots, x_n)\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.\]

As above, all of these norms yield metrics in the usual way, $d_p(x, y) = \|x - y\|_p$. Note: the definition still makes sense when $p < 1$, but it no longer gives a metric: the triangle inequality is violated. For example, taking $p = \frac{1}{2}$, we have
\[
\|(9, 1) + (16, 0)\|_1/2 = 36
\]
\[
\|(9, 1)\|_1/2 + \|(16, 0)\|_1/2 = 32 < 36.
\]

(4) Let $B[0, 1]$ consist of all bounded functions $[0, 1] \to \mathbb{R}$. Then define a function $d_u : B[0, 1] \times B[0, 1] \to \mathbb{R}$ by
\[
d_u(f, g) = \sup_{x \in [0, 1]} |f(x) - g(x)|.
\]
This is well-defined: since $f$ and $g$ are bounded, the set $\{f(x) - g(x) : x \in [0, 1]\}$ is a bounded, nonempty set, so it has a sup. It is $\geq 0$, and moreover if $d_u(f, g) = 0$, then for every $x_0 \in [0, 1]$, $|f(x_0) - g(x_0)| \leq \sup_{x \in [0, 1]} |f(x) - g(x)| = 0$, which implies that $f(x_0) = g(x_0) = 0$ - i.e. $f = g$. This verifies the first two properties of Definition 5.1.

For the triangle inequality, we have
\[
d_u(f, h) = \sup_{x \in [0, 1]} |f(x) - h(x)| = \sup_{x \in [0, 1]} |f(x) - g(x) + g(x) - h(x)|
\]
\[
\leq \sup_{x \in [0, 1]} (|f(x) - g(x)| + |g(x) - h(x)|)
\]
\[
\leq \sup_{x \in [0, 1]} |f(x) - g(x)| + \sup_{x \in [0, 1]} |g(x) - h(x)|
\]
\[
d_u(f, g) + d_u(g, h)
\]
using the properties of $\sup$ we now know well. Thus the triangle inequality holds for $d_u$ as well, and so it is a metric. Note, like the above examples, it has the form $d_u(f, g) = \|f - g\|_u$ for a norm $\| \cdot \|_u$: a function on $B[0, 1]$ which has the properties $\|f\|_u \geq 0$ and $= 0$ only if $f = 0$, and satisfies the triangle inequality $\|f + g\|_u \leq \|f\|_u + \|g\|_u$.

Whenever we have a function like this defined on a vector space, it gives rise to a metric by subtraction.

(5) Not every metric is given in terms of a norm like this. For example, consider on $\mathbb{R}$ the function
\[d(x, y) = \min\{|x - y|, 1\}.\]

It is easy to verify that this satisfies properties (1) and (2) in Definition 5.1. The triangle inequality is also easy to see, by breaking into eight cases (depending whether $|x - y|$, $|x - z|$, and $|y - z|$ are $\leq 1$ or $> 1$); this boring proof is left to the reader.

(6) Given any nonempty set $X$, one can define a metric on $X$ by the silly rule
\[
d(x, y) = \begin{cases}
0, & x = y \\
1, & x \neq y.
\end{cases}
\]

This is known as the discrete metric. It says two points are close only if they are equal; otherwise they are far apart. It is again simple to verify this is a metric.
One important observation was made at several points in the examples: if \((X, d)\) is a metric space, and \(Y \subseteq X\), then \((Y, d|_Y)\) is a metric space – that is, the metric \(Y\) defined on all pairs \((x, y) \in X \times X\), also defines a metric when restricted only to pairs in \(Y \times Y\), as is straightforward to verify. Thus, the Euclidean metric on \(\mathbb{C}^n\) automatically gives us a metric (also called the Euclidean metric) on \(\mathbb{R}^n\). Similarly, the usual metric on \(\mathbb{R}\) restricts to a metric on \([0, 1]\).

Usually thinking of metric spaces using our intuition from \(\mathbb{R}^2\) and \(\mathbb{R}^3\), we introduce the following notation.

**Definition 5.3.** Let \((X, d)\) be a metric space, and let \(x_0 \in X\). For a fixed \(r > 0\), the **ball** of radius \(r\) centered at \(x_0\), denoted \(B_r(x_0)\), is the set

\[
B_r(x_0) = \{x \in X : d(x_0, x) < r\}.
\]

(Rudin calls this a **neighborhood** \(N_r(x_0)\).) With \(r = 1\), we refer to this as the **unit ball** centered at \(x_0\).

**Example 5.4.** (1) In \(\mathbb{R}^n\), using the definition of the Euclidean metric (and choosing the base point \(0\) to simplify things), we have

\[
B_r(0) = \{(x_1, \ldots, x_n) : x_1^2 + \cdots + x_n^2 < r^2\}
\]

which is what we usually know as a ball (in \(n\)-dimensions).

(2) Consider \((\mathbb{R}^2, d_p)\), with the \(p\)-metric of Example 5.2(3). Here are some pictures of the unit ball:

(3) In a discrete metric space \((X, d)\) as in Example 5.2(6), \(B_r(x_0) = X\) if \(r > 1\), and \(B_r(x_0) = \{x_0\}\) if \(r \leq 1\).
1. Lecture 14: February 18, 2016

Now, once more, we define convergence and Cauchy in the wider world of metric spaces.

**Definition 5.5.** Let \((X, d)\) be a metric space, and let \((x_n)\) be a sequence in \(X\), and let \(x \in X\). Say that \((x_n)\) **converges to** \(x\), or \(x_n \to x\), or \(\lim_{n \to \infty} x_n = x\), if
\[
\forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N \; d(x_n, x) < \epsilon.
\]
In other words, \(x_n \to x\) means that the real sequence \(d(x_n, x) \to 0\). Alternatively, we could state this as: for all sufficiently large \(n\), \(x_n \in B_\epsilon(x)\).

Similarly, say that \((x_n)\) is a **Cauchy sequence** in \(X\) if
\[
\forall \epsilon > 0 \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geq N \; d(x_n, x_m) < \epsilon.
\]

As discussed in the generalization from \(\mathbb{R}\) to \(\mathbb{C}\), limits are unique: if \(x_n \to x\) and \(x_n \to y\), then \(x = y\) (this follows from the fact that \(d(x, y) = 0\) implies that \(x = y\); it is primarily for this reason that this non-degeneracy property is required in the definition of a metric).

**Example 5.6.** Consider a discrete metric space: \((X, d)\) where \(d\) is given as in Example 5.2(6). Let \(x_n \to x\). In particular, this means that there is some \(N\) such that, for \(n \geq N\), \(d(x_n, x) < \frac{1}{2}\). But by the definition of \(d\), either \(d(x_n, x) = 0\) or \(d(x_n, x) = 1\); so if \(d(x_n, x) < \frac{1}{2}\) then \(d(x_n, x) = 0\), and so \(x_n = x\). Thus, if \(x_n \to x\), then \(x_n = x\) for all large \(n\). In a discrete metric space, convergence is the same thing as eventually constant. The same holds true for Cauchy.

In general, there is a fundamental difference between convergent sequences that are eventually constant and convergent sequences that are not eventually constant. We use this difference to define one of the most important topological concepts.

**Definition 5.7.** Let \((X, d)\) be a metric space, and let \(E \subseteq X\) be a subset. A point \(x \in X\) (not necessarily in \(E\)) is called a **limit point** of \(E\) if there is a sequence \(x_n \in E \setminus \{x\}\) that converges to \(x\), \(x_n \to x\). That is: a limit point of \(E\) is a limit of some not eventually constant sequence in \(E\). A point \(e \in E\) that is not a limit point of \(E\) is called an **isolated point** of \(E\).

**Example 5.8.** In \(\mathbb{R}\) with the usual metric, take \(E = (-1, 0] \cup \mathbb{N}\). Then \(-1\) is a limit point of \(E\): for example, \(-1 = \lim(-1 + \frac{1}{n})\) and \(-1 + \frac{1}{n} \in E\) for each \(n\). Also, any point \(x \in E\) is a limit point of \(E\): take \(x_n = x - \frac{1 + x}{n}\) as the sequence. This is in \(E\) since \(1 + x > 0\) and so \(x - \frac{1 + x}{n} < x \leq 0\), but also \(x - \frac{1 + x}{n} > x - (1 + x) = -1\).

On the other hand, the positive integers \(\mathbb{N}\) are isolated points of \(E\). For example, consider 1. If \(y_n\) is any sequence in \(\mathbb{R}\) that converges to 1, then we must have \(y_n \in (0.9, 1.1)\) for all large \(n\); but then if \(y_n \in E\) it follows that \(y_n = 1\) for all large \(n\), which isn’t allowed. Thus, no sequence in \(E \setminus \{1\}\) converges to 1, showing that the point 1 \(\in E\) is not a limit point of \(E\) – it is an isolated point.

The set of all limit points of a set \(E\) is denoted \(E’\). So \(E\) is closed iff \(E’ \subseteq E\).

**Definition 5.9.** A subset \(E\) of a metric space is called **closed** if it contains all of its limit points.

**Example 5.10.**
1. The set \(E = (-1, 0] \cup \mathbb{N}\) from Example 5.8 is not closed: \(-1\) is a limit point of \(E\), but \(-1 \notin E\).
2. The set \(F = [-1, 0] \cup \mathbb{N}\) is closed. The argument in Example 5.8 shows that each of the points in \([-1, 0]\) is a limit point of \(F\), while each of the points \(n \in \mathbb{N}\) is an isolated point.
of \( F \). On the other hand, if \( x \) is a real number not in \( F \), then either \( x < -1 \) or \( x > 0 \) and \( x \not\in \mathbb{N} \). In the former case, this means that no sequence in \( F \) can come within distance \( 1 + x > 0 \) of \( x \), and so cannot converge to \( x \); a similar argument with \( x > 0 \) shows that \( x \) is not a limit point of \( F \). Thus, the set of limit points of \( F \) consists exactly of the set \([-1, 0] \), and this set is contained in \( F \). So \( F \) is closed.

Definition \(5.9\) is stated in terms of limit points to make it clear that there are two kinds of points to consider in deciding whether a set is closed: isolated points and non-isolated points. For example, if one has a closed set, then adding to it a finite collection of isolated points will preserve closedness. But for the purposes of a concise definition, one need not be concerned about the distinction.

Proposition 5.11. A subset \( E \) of a metric space \((X, d)\) is closed if and only if, for any sequence \((x_n)\) in \( E \) that converges in \( X \), the limit \( \lim_{n \to \infty} x_n \) is actually in \( E \).

That is: closed means closed under limits of sequences.

Proof. Suppose \( E \) is closed, so \( E' \subseteq E \). Now, let \((x_n)\) be any sequence in \( E \) that converges to some point \( x \). If \( x_n \neq x \) for any \( n \), then by definition \( x \in E' \), and therefore by assumption \( x \in E \). If, on the other hand, there exists \( n \) with \( x_n = x \), then since \( x_n \in E \) for each \( n \), we have \( x \in E \). Thus, the \( E \) is closed under limits of sequences.

Conversely, suppose \( E \) is closed under limits of sequences. Let \( x \in E' \); so by definition there is a sequence \( x_n \in E \setminus \{x\} \) such that \( x_n \to x \). Well, since \( x_n \to x \) and \( x_n \in E \), by assumption \( x \in E \). Thus \( E' \subseteq E \), and \( E \) is closed. \( \square \)

There is a complementary notion to closed, called open.

Definition 5.12. A subset \( E \) of a metric space is called open if, for any point \( x \in E \), there is a ball \( B_r(x) \) (for some \( r > 0 \)) with \( B_r(x) \subseteq E \).

Example 5.13. (1) The set \( E = (-1, 0] \cup \mathbb{N} \) from Example \(5.8\) is not open. Consider the point \( 0 \in E \). For any \( 0 < r < 1 \), the ball \( B_r(0) = (-r, r) \) contains some points (for example \( \frac{r}{2} \)) that are in \((0, 1)\), and hence not in \( E \). Similarly, any of the points in \( \mathbb{N} \) are in \( E \) but none is contained in a ball contained in \( E \). So \( E \) is not open.

(2) On the other hand, the set \( U = (0, 1) \) is open. Indeed, let \( x \in U \). Let’s consider two cases: either \( 0 < x < \frac{1}{2} \) or \( \frac{1}{2} \leq x < 1 \). In the former case, the ball \( B_{x}(x) = (0, 2x) \) is contained in \( U = (0, 1) \); in the latter case, the ball \( B_{1-x}(x) = (2x - 1, 1) \) is contained in \( U \). So every point of \( x \) is contained in a ball inside \( U \), showing that \( U \) is open.

(3) Let \((X, d)\) be a discrete metric space. If \( x \in X \), then by Example \(5.4\), \( B_1(x) = \{x\} \).

Thus, every singleton point in a discrete metric space is an open set. On the other hand, by Example \(5.6\) there are no non-eventually-constant sequences converging to any point \( x \), which means every point is isolated. That is: \( X \) has no limit points, which means that (vacuously) \( X \) contains all its limit points. So \( X \) is also closed.

(4) Consider the empty set \( \emptyset \) in any metric space. It is both open and closed. Indeed, the definitions of “open” and “closed” each start with “for every point in the set...” and since there are no points in \( \emptyset \) to check the condition, it follows that the condition holds vacuously.

Example \(5.13\) has a nice, important generalization to any metric space. Not that \((0, 1)\) is itself a ball in \( \mathbb{R} \): it is the ball \( B_{1/2}(1/2) \). The fact is, any ball is open.
Proposition 5.14. Let \((X, d)\) be a metric space, let \(x \in X\), and let \(r > 0\). Then the ball \(B_r(x)\) is open in \(X\).

Proof. Let \(y \in B_r(x)\). This means \(d(x, y) < r\). Hence, there is some \(\epsilon > 0\) so that \(d(x, y) = r - \epsilon\). I claim that \(B_\epsilon(y) \subset B_r(x)\). Indeed, suppose that \(z \in B_\epsilon(y)\), meaning that \(d(z, y) < \epsilon\). Then
\[
d(x, z) \leq d(x, y) + d(y, z) = r - \epsilon + d(y, z) < r - \epsilon + \epsilon = r.
\]
so \(z \in B_r(x)\). We have thus shown that, for any \(y \in B_r(x)\), there is a ball \(B_\epsilon(y) \subset B_r(x)\). That is: \(B_r(x)\) is open. \(\square\)
We referred to open and closed as complementary properties. That doesn’t mean that any set is either open or closed: for example, the set \((-1,0]\) considered above is neither open nor closed. But they concepts are complementary, in the following precise sense.

**Proposition 5.15.** Let \((X,d)\) be a metric space. A subset \(E \subseteq X\) is open if and only if \(E^c = X \setminus E\) is closed.

Since \((E^c)^c = E\), it follows similarly that \(E\) is closed iff \(E^c\) is open. In the proof we will use the characterization of closed given in Proposition 5.11.

**Proof.** Suppose \(E\) is open. Let \((x_n)\) be a sequence in \(E^c\) that converges to some point \(x \in X\). We want to show that \(x \in E^c\); to produce a contradiction, we therefore assume that \(x \notin E^c\), meaning \(x \in E\). Since \(E\) is open, by definition there is some \(r > 0\) so that \(B_r(x) \subseteq E\). On the other hand, since \(x_n \to x\), there is certainly some \(N\) so that \(d(x_N, x) < r\). Thus \(x_N \in B_r(x) \subseteq E\), which means that \(x_N \in E\). But we assumed that \(x_N \in E^c\), so this is a contradiction. Therefore \(x \in E^c\). This shows that any convergent sequence in \(E^c\) has limit in \(E^c\), which shows that \(E^c\) is closed.

Conversely, suppose \(E^c\) is closed. Let \(x \in E\). We want to show that there is some \(r > 0\) so that \(B_r(x) \subseteq E\); to produce a contradiction, we therefore assume that there is no such \(r\). That means that, for say \(r = \frac{1}{n}\), \(B_{1/n}(x) \not\subseteq E\), which means precisely that there is some point \(x_n \notin E\) such that \(x_n \to x\). So, we have produced a sequence \(x_n \in E^c\) such that \(d(x_n, x) < \frac{1}{n}\), meaning that \(x_n \to x\). Thus \(x\) is the limit of a sequence in \(E^c\), and so by assumption \(x \in E^c\). This contradicts the assumption that \(x \in E\). Therefore there must be some \(r > 0\) so that \(B_r(x) \subseteq E\), and so \(E^c\) is closed.

Let us make a few more definitions that pertain the local properties of open and closed sets.

**Definition 5.16.** Let \((X,d)\) be a metric space, and let \(E \subseteq X\).

1. The **closure** of \(E\) is the set \(\overline{E} = E \cup E^c\).
2. A point \(x \in E\) is called an **interior point** if there is some \(r > 0\) with \(B_r(x) \subseteq E\). The set of all interior points of \(E\) is called the **interior** of \(E\), and is denoted \(\overset{•}{E}\).
3. The **boundary** of \(E\) is the set \(\partial E = \overline{E} \setminus \overset{•}{E}\).

**Remark 5.17.** Following the proof of Proposition 5.11, \(\overline{E}\) can alternatively be described as the set of all limits of convergent sequences in \(E\).

**Example 5.18.** Consider again the set \(E = (-1,0] \cup \mathbb{N}\) in \(\mathbb{R}\), considered in Examples 5.8 and 5.10. We’ve shown that the points in \((-\infty, -1)\) and \((0, \infty)\) are not limit points (the points \(1,2,3,\ldots\) are in \(E\) but are isolated); on the other hand, we’ve shown that the points \([-1,0]\) are all limit points. Thus \(E^c = [-1,0]\), and so \(\overline{E} = E \cup E^c = [-1,0] \cup \mathbb{N}\). We’ve also shown in Example 5.13(1) that there is no ball centered at 0 contained in \(\overset{•}{E}\); similarly, there are no balls centered at the points \(1,2,3,\ldots\) contained in \(E\), so these are not interior points. On the other hand, an argument very similar to Example 5.13(2) shows that the points in \((-1,0)\) are interior points. So \(\overline{E} = (-1,0)\). Finally, this shows that \(\partial E = \overline{E} \setminus \overset{•}{E} = \{-1,0,1,2,\ldots\}\).

**Example 5.19.** Let \(\mathbb{Q}\) denote the rational numbers as a subset of the metric space \(\mathbb{R}\). By Theorem 1.17(2) (the density of \(\mathbb{Q}\) in \(\mathbb{R}\)), given any real numbers \(a < b\) there is a rational number \(q \in \mathbb{Q}\) with \(a < q < b\). In particular, fix \(x \in \mathbb{R}\); then for \(n \in \mathbb{N}\) there is a rational number...
\(q_n\) with \(x + \frac{1}{2n} < q_n < x + \frac{1}{n}\). In particular, we have \(\frac{1}{2n} < |q_n - x| < \frac{1}{n}\). This shows that \(q_n \to x\) but \(q_n \neq x\) for any \(n\); thus \(x \in \mathbb{Q}'\). So every real number is a limit point of \(\mathbb{Q}\), and so \(\overline{\mathbb{Q}} = \mathbb{Q} \cup \mathbb{Q}' = \mathbb{Q} \cup \mathbb{R} = \mathbb{R}\). (This is another way of saying \(\mathbb{Q}\) is dense in \(\mathbb{R}\); in general, we say a subset \(E \subseteq X\) is \textit{dense} in a metric space \(X\) if \(\overline{E} = X\).)

On the other hand, let \(r > 0\), and let \(q \in \mathbb{Q}\). The number \(x = q + \frac{r}{\sqrt{2}}\) is \(< x + r\), which shows that \(x \in (q - r, q + r) = B_r(q)\). But \(x \notin \mathbb{Q}\): indeed, we can solve \(\sqrt{2} = \frac{x}{x-q}\), and so if \(x\) were rational \(\sqrt{2}\) would also be rational, which we know it is not. Thus \(B_r(q) \not\subseteq \mathbb{Q}\) for any \(r > 0\). This shows \(q\) is not interior to \(\mathbb{Q}\). This holds for any \(q \in \mathbb{Q}\), and so, in fact, \(\mathbb{Q} = \emptyset\).

**Theorem 5.20.** Let \((X, d)\) be a metric space, and let \(E \subseteq X\).

1. \(\overline{E}\) is closed; \(E\) is closed iff \(\overline{E} = \overline{E}\).
2. \(\overline{E}\) is open; \(E\) is open iff \(\overline{E} = \overline{E}\).

**Proof.** We begin with item 1. Let \((x_n)\) be a sequence in \(\overline{E}\) with limit \(x\). We wish to show \(x \in \overline{E}\). If \(x \in E \subseteq \overline{E}\) we are done, so assume \(x \notin E\). For each \(x_n\), either \(x_n \in E\) or \(x_n \in E'\). In the latter case, by definition of \(E'\) there is some other sequence \(y_k \in E\) such that \(y_k \to x_n\); in particular, we can choose some \(k\) large enough that \(d(y_k, x_n) < \frac{1}{n}\). So, we can define a new sequence \((x'_n)\) as follows: if \(x_n \in E\) then \(x'_n = x_n\); if \(x_n \notin E'\), then \(x'_n = y_k\) as above, so \(x'_n \in E\) and \(d(x_n, x'_n) < \frac{1}{n}\). Then we have \(d(x'_n, x) \leq d(x'_n, x_n) + d(x_n, x) < \frac{1}{n} + d(x_n, x) \to 0\), and so \(x'_n \to x\). As \(x'_n \in E\) and \(x \notin \overline{E}\), it follows that \(x\) is a limit point of \(E\), and so \(x \in E' \subseteq \overline{E} \cup E' = \overline{E}\). Thus \(\overline{E}\) is closed under limits; by Proposition 5.11, it follows that \(\overline{E}\) is closed, as claimed. Now, by definition \(\overline{E}\) is closed iff \(E' \subseteq \overline{E}\), and this happens iff \(\overline{E} = E \cup E' = E\), proving the second point.

For item 2, let \(x \in \overline{E}\); thus, there is some ball \(B_r(x)\) contained in \(E\). But by Proposition 5.14, the ball \(B_r(x)\) is open, which means all its points are interior points; thus \(B_r(x) \subseteq \overline{E}\). So, any point in \(\overline{E}\) is interior to \(\overline{E}\), which shows that \(\overline{E}\) is open. By definition \(\overline{E} \subseteq E\) for any set \(E\); thus \(\overline{E} = E\) iff \(E \subseteq \overline{E}\), which is the statement that every point of \(E\) is an interior point, which is precisely the definition of \(E\) being open. \(\square\)

**Example 5.21.** Let \(E \subseteq \mathbb{R}\) be nonempty and bounded above. Then \(\alpha = \sup E\) exists. By definition, \(\alpha - \frac{1}{n}\) is not an upper bound for \(E\) for any \(n \in \mathbb{N}\), which shows that there is an element \(x_n \in E\) with \(\alpha - \frac{1}{n} < x_n \leq \alpha\). This shows that \(x_n \to \alpha\). By Remark 5.17, it follows that \(\alpha \in \overline{E}\): the supremum is always in the closure. On the other hand, if there were some \(r > 0\) with \(B_r(\alpha) \subseteq \overline{E}\), then, for example, \(\alpha + \frac{r}{2} \in E\). Since \(\alpha + \frac{r}{2} > \alpha\), this contradicts \(\alpha\) being an upper bound for \(E\). Thus, \(\alpha\) is not in \(\overline{E}\). That is: \(\sup E \in \overline{E} \setminus \overline{E} = \partial E\).

Now we come to an important concept you may not have encountered before: compactness.

**Definition 5.22.** Let \((X, d)\) be a metric space. A subset \(K \subseteq X\) is called **compact** if every sequence \((x_n)\) in \(K\) has a convergent subsequence whose limit is in \(K\).

**Example 5.23.**

1. Let \(a < b\) be real numbers, and consider the set \(K = [a, b]\). The Bolzano-Weierstrass Theorem for \(\mathbb{R}\) (Theorem 2.25) is precisely the statement that \([a, b]\) is compact.

2. On the other hand, \(E = [a, b)\) is not compact: the sequence \(x_n = b - \frac{b-a}{n}\) is in \(E\), but converges to \(b \notin E\), therefore all of its subsequences converge to \(b\), and hence none of them converge in \(E\). Similarly, an unbounded interval like \([0, \infty)\) is not compact: for example the sequence \(x_n = n\) has no convergent subsequences at all.

3. Let \((X, d)\) be a discrete metric space. If \(K\) is a finite subset of \(X\), say \(K = \{y_1, y_2, \ldots, y_n\}\), then \(K\) is compact. Indeed, if \((x_n)\) is any sequence in \(K\), then there must be some (perhaps many) \(y_j\) so that \(x_n = y_j\) for infinitely many \(n\) (by the pigeonhole principle). That means exactly that there is an increasing sequence \(n_k\) with \(x_{n_k} = y_j\) for all \(k\), which means \(x_{n_k} \rightarrow y_j \in K\). Thus \(K\) is compact. On the other hand if \(E \subseteq X\) is infinite, it is not compact: for then we can find an infinite sequence \(x_1, x_2, x_3, \ldots \in E\) all distinct. Thus, any subsequence also has all distinct terms, which means it is not eventually constant. By Example 5.6 this means no subsequence converges.

Now, there is an alternate definition of compactness which is the only one used in Rudin; we refer to it as **topological compactness**, given in Definition 5.24 below. First, let us highlight the fact that Definition 5.22 was the **original** definition of compact, and predated the so-called “modern” definition by almost a century. Bolzano was already using our definition of compactness in 1817, although it would not be until 1906 that Definition 5.22 was written down formally (by Fréchet). It was around this time that Lebesgue proved (as a useful lemma) that Definition 5.24 also characterizes compactness; indeed, as we will see, it is a very useful tool. Much later, in 1929, the Russian school (led by Alexandrov and Urysohn) redefined compactness as what we are calling topological compactness. Our definition of the word **compact** is now often called **sequentially compact**.

**Definition 5.24.** Let \((X, d)\) be a metric space. Let \(K \subseteq X\) be a subset. An open cover of \(K\) is a collection (finite or infinite) of open set \(\mathcal{C}\) in \(X\) such that every point in \(K\) is in at least one \(U \in \mathcal{C}\): that is \(X \subseteq \bigcup \mathcal{C}\). We call \(K\) **topologically compact** if, given any open cover \(\mathcal{C}\) of \(K\), there is a finite sub cover: that is, there are finitely many \(U_1, \ldots, U_m \in \mathcal{C}\) such that \(K \subseteq U_1 \cup \cdots \cup U_m\).

**Example 5.25.** Consider the interval \((0, 1]\). We have already seen this is not compact. It is also not topologically compact. Indeed, consider the sets \(U_n = \left(\frac{1}{n}, 2\right)\) for \(n \in \mathbb{N}\). If \(x \in (0, 1]\) then \(x > 0\) and so there is some \(n \in \mathbb{N}\) with \(\frac{1}{n} < x\). Therefore \(x \in \left(\frac{1}{n}, 1\right) \subseteq \left(\frac{1}{n}, 2\right) = U_n\). This shows that the collection \(\mathcal{C} = \{U_n: n \in \mathbb{N}\}\) is an open cover of \((0, 1]\). Now, consider any finite collection of sets from \(\mathcal{C}\): \(U_{n_1}, U_{n_2}, \ldots, U_{n_k}\) for some \(k \in \mathbb{N}\). Note that \(\frac{1}{m} < \frac{1}{k}\) when \(m > k\), and so \(U_\ell \subseteq U_m\) in this case. What that means is that, if we let \(m = \max\{n_1, \ldots, n_k\}\) then \(U_{n_1} \cup \cdots \cup U_{n_k} = U_m = \left(\frac{1}{m}, 2\right)\). But then this does not cover \((0, 1]\): there are points \(x \in (0, 1]\) with \(x < \frac{1}{m}\). Thus, no finite subcover of \(\mathcal{C}\) will cover all of \((0, 1]\). The existence of such an open cover without any finite subcover shows that \((0, 1]\) is not topologically compact.
THEOREM 5.26. Let \( K \) be a set in a subset of a metric space. Then \( K \) is sequentially compact iff \( K \) is topologically compact.

We will not prove Theorem 5.26 here; this is the sort of thing that will be covered in an undergraduate topology course (such as Math 190). Rudin chooses to use the more abstract topological definition of compactness (for historical reasons that I find unsatisfactory), and this has the effect of both making everything more abstract, and also making all the proofs harder than necessary. We will stick exclusively with *sequential compactness*. This means all our proofs will be different from Rudin’s – and generally shorter and easier to understand!

In Example 5.23(2), the absence of the point \( b \) from \([a, b]\) makes the set non-compact. Note that \( b \) is in the closure of \([a, b]\). This highlights the following proposition.

**Proposition 5.27.** Compact sets are closed. Also, if \( K \) is compact and \( F \subseteq K \) is closed, then \( F \) is compact.

**Proof.** Suppose \( K \) is compact. Let \((x_n)\) be a sequence in \( K \) which converges. By compactness, there is some subsequence \((x_{n_k})\) that converges to a point in \( K \). But we know that every subsequence of \((x_n)\) converges to \(\lim x_n\), and hence \(\lim x_n \in K\). Thus, \(K\) is closed under limits, and so \(K\) is closed.

Now, let \( F \) be a closed subset of a compact set \( K \). Let \((y_n)\) be any sequence in \( F \). Then \((y_n)\) is a sequence in \( K \), and hence by compactness there is a subsequence \((y_{n_k})\) that converges in \( K \). Note that \( y_{n_k} \in F \) for each \( k \), and hence since \( F \) is closed it follows that \( \lim y_{n_k} \in F \). Thus, any sequence in \( F \) has a convergent subsequence with limit in \( F \); i.e. \( F \) is compact. \(\square\)

**Definition 5.28.** Let \( E \) be a subset of a metric space. The **diameter** of \( E \), denoted \( \text{diam}(E) \), is defined to be

\[
\text{diam}(E) = \sup\{d(x, y) : x, y \in E\}.
\]

**Note:** this might well be \(+\infty\). If \( \text{diam}(E) < +\infty \), we call \( E \) **bounded**; otherwise \( E \) is **unbounded**.

**Example 5.29.**

1. \( \text{diam}(B_r(x)) \leq 2r \) for any ball in a metric space. But it could be less: for example in a discrete metric space with at least two elements, \( \text{diam}(B_r(x)) = 0 \) if \( r \leq 1 \) and \( = 1 \) if \( r > 1 \).

2. In \( \mathbb{R} \), \( \text{diam}(0, 1) = \text{diam}(0, 1) = \text{diam}(0, 1) = 1 \).

3. In \( \mathbb{R} \), \( \text{diam}(\mathbb{N}) = \infty \). Indeed, \( d(0, n) = n \) so \( \sup\{d(x, y) : x, y \in \mathbb{N}\} \geq n \) for every \( n \).

**Note:** if \( E \) is a bounded set, with diameter \( \delta > 0 \), then for any point \( x \in E \), \( E \subseteq B_{2\delta}(x) \) (or \( B_{1.0001\delta}(x) \), or \( B_{0.0001\delta}(x) \), etc.) Conversely, suppose there is some \( x \) in the metric space and some \( r > 0 \) with \( E \subseteq B_r(x) \). Since \( \text{diam}(B_r(x)) \leq 2r \), it follows that \( \text{diam}(E) \leq 2r \). So, to say \( E \) is bounded is the same as saying it is contained in some ball.

**Proposition 5.30.** Compact sets are bounded.

**Proof.** We prove the contrapositive: unbounded sets are not compact. Let \( E \) be unbounded, and fix a point \( x_0 \in E \). Consider the set of balls \( B_n(x_0) \) for \( n \in \mathbb{N} \). By assumption, \( E \not\subseteq B_n(x_0) \) for any \( n \), so we can choose a point \( x_n \in E \) with \( d(x_0, x_n) \geq n \).

In fact, the sequence \((x_n)\) has no convergent subsequences. For let \( x \) be any point in the metric space. Let \( n \in \mathbb{N} \) be large enough that \( N > d(x_0, x) \). Then for \( n \geq N + 1 \), we have by the triangle inequality

\[
d(x_n, x) \geq d(x_n, x_0) - d(x_0, x) \geq n - d(x_0, x) \geq 1 + N - d(x_0, x) > 1.
\]
That is: for any point $x$ in the metric space, eventually $x_n$ never comes within distance 1 of $x$. It follows that no subsequence of $(x_n)$ can converge to $x$. Since this holds for any $x$, it follows that $(x_n)$ has no convergent subsequences. Since $(x_n)$ is a sequence in $E$, this means $E$ is not compact.

Thus, we have seen that compact sets are closed and bounded. One of the biggest theorems of this course, the Heine–Borel Theorem, states that the converse is true in Euclidean space.

**Theorem 5.31 (Heine–Borel).** Let $m \in \mathbb{N}$. A subset of $\mathbb{R}^m$ is compact iff it is closed and bounded.

**Proof.** Let $K \subset \mathbb{R}^m$. If $K$ is compact, then by Propositions 5.27 and 5.30 $K$ is closed and bounded. We must prove the converse. Suppose $K$ is a closed and bounded subset of $\mathbb{R}^m$. Let $(x_n)$ be a sequence in $K$. We may write it in terms of its components

$$x_n = (x_1^n, x_2^n, \ldots, x_m^n).$$

Consider first the sequence $(x_1^n)_{n=1}^\infty$ in $\mathbb{R}$. Note that

$$|x_1^n| \leq |x_n| = d(x_n, x_1) \leq \text{diam}(K).$$

So the sequence $(x_1^n)$ is a bounded sequence in $\mathbb{R}$. By Theorem 2.25 (the Bolzano-Weierstrass Theorem for $\mathbb{R}$), there is a subsequence $x_1^{n_k}$ that converges. Now we proceed as in the proof of Theorem 3.15 (the Bolzano-Weierstrass Theorem for $\mathbb{C}$). Consider the subsequence $x_2^{n_k}$. Again we have $|x_2^{n_k}| \leq \text{diam}(K)$ is bounded, so by the Bolzano-Weierstrass Theorem for $\mathbb{R}$, it possesses a further subsequence $x_2^{n_k\ell}$ that is convergent. Note that $x_1^{n_k\ell}$ is a subsequence of the convergent subsequence $x_1^{n_k}$, so it is also convergent. Now we proceed to select a further convergent subsubsequence that makes $x_3^{n_k\ell\ell}$ converge, and so forth. The notation becomes ridiculous, but in the end (after $m$ steps) we produce a single set of indices $1 \leq \ell_1 < \ell_2 < \cdots$ such that all of the components $(x_1^{\ell_n}, x_2^{\ell_n}, \ldots, x_m^{\ell_n})$ converge as $n \to \infty$. We now follow the proof of Proposition 3.13 exactly to see that convergence in $\mathbb{R}^m$ is equivalent to convergence of each component separately, and so we conclude that the subsequence $(x_{\ell_n})$ converges to some element $x \in \mathbb{R}^m$. Finally, note that $x_{\ell_n} \in K$ by assumption, and $K$ is closed; thus the point $x$ is also in $K$. This shows that every sequence in $K$ has a convergent subsequence with limit in $K$, concluding the proof that $K$ is compact. □
5. METRIC SPACES

4. Lecture 17: February 29, 2016

So, closed intervals \([a, b]\) are compact (as we already knew), as are sets like \([0, 1] \cup [2, 3] \cup \{4, 5, 6, 7, 8\}\). In fact, there are much more complicated closed and bounded sets in \(\mathbb{R}\) (e.g. the Cantor set of Example 5.34 below). Let’s emphasize that the Heine–Borel Theorem is exclusively about the metric spaces \(\mathbb{R}\); it does not apply in general.

Example 5.32. (1) Let \((X, d)\) be a discrete metric space. Then for any two points \(x, y \in X\), either \(d(x, y) = 0\) or \(d(x, y) = 1\). Thus, for any subset \(E \subseteq X\), \(\text{diam}(E) \leq 1\), so \(E\) is bounded. We have also shown that any subset \(E\) is closed. However, if \(E\) is an infinite set, it is not compact, cf. Example 5.23(3). So any infinite discrete metric spaces contains closed and bounded sets that are not compact.

(2) For a less contrived example, consider again \(B[0, 1]\), the set of all bounded, real-valued functions on \([0, 1]\), which is a metric space with respect to the metric \(d(f, g) = \sup_{x \in [0,1]} |f(x) - g(x)|\).

Consider the functions

\[
  f_n(x) = \begin{cases} 
    1, & x \geq \frac{1}{n} \\
    0, & x < \frac{1}{n}.
  \end{cases}
\]

All of these functions are in \(B[0, 1]\). We can also compute the function \(f_n - f_m\); assuming \(m > n\) we have

\[
  f_m(x) - f_n(x) = \begin{cases} 
    1, & \frac{1}{m} \leq x < \frac{1}{n} \\
    0, & x < \frac{1}{m}.
  \end{cases}
\]

This shows that \(d(f_n, f_m) = \sup_x |f_n(x) - f_m(x)| = 1\) for any \(m \neq n\)! Thus, the sequence \((f_n)\) cannot have a convergent subsequence: no two terms in the sequence are ever closer to (or farther from) each other than 1.

Here is an important property of compact sets. This is the generalization of the nested intervals property that we used in the construction of \(\mathbb{R}\).

Proposition 5.33. Let \(K_1, K_2, K_3, \ldots\) be nonempty compact sets in a metric space, and suppose they are nested: \(K_{n+1} \subseteq K_n\) for all \(n\). Then \(\bigcap_n K_n\) is a nonempty compact set. If, in addition, \(\text{diam}(K_n) \to 0\) as \(n \to \infty\), then \(\bigcap_n K_n\) consists of exactly one point.

Proof. Since \(K_n \neq \emptyset\) for any \(n\), we can choose a point \(x_n \in K_n\) for each \(n\). By the nested property, \(x_n \in K_1\) for each \(n\). Thus, \((x_n)\) is a sequence in the compact set \(K_1\), and therefore it has a convergent subsequence \(x_{n_k}\) with a limit \(x \in K_1\). Now, for any \(m \in \mathbb{N}\), the tail subsequence \((x_{n_k})_{k=m}^{\infty}\) also converges to \(x\); but this is a sequence of terms in \(K_{nm}\), which is closed, and so \(x \in K_{nm}\). This holds for every \(m\). Finally, for any \(n\), there is \(n_m > n\), and therefore \(K_{nm} \subseteq K_n\); thus, \(x \in K_n\) for every \(n\), which shows that \(x \in \bigcap_n K_n\). This intersection is therefore nonempty.

It is an intersection of compact sets, therefore it is compact (Exercise 1 on HW9).

For the second claim, let \(x, y \in \bigcap_n K_n\). Fix \(\epsilon > 0\); since \(\text{diam}(K_n) \to 0\), there is some \(n\) with \(\text{diam}(K_n) < \epsilon\). Thus, since \(x, y \in K_n\), \(d(x, y) \leq \text{diam}(K_n) < \epsilon\). So \(0 \leq d(x, y) < \epsilon\) for all \(\epsilon > 0\); it follows that \(d(x, y) = 0\) and so \(x = y\). That is: there is at most one point in the intersection. As we’ve shown the intersection is nonempty, this proves that it consists of exactly one point, as claimed. \(\square\)
**Example 5.34 (Cantor set).** The unit interval $K_0 = [0, 1]$ is compact. Now remove the “middle third” and let $K_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$; this set is also compact, and $K_1 \subset K_0$. Now repeat this: remove the middle third from each of the two intervals in $K_1$, producing $K_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. Again, this set is compact, and $K_2 \subset K_1$. We can repeat this “delete middle thirds” process indefinitely. Note a certain self-similarity: $K_2$ has two pieces, each of which looks like $K_1$ shrunk down by a factor of $\frac{1}{3}$. In fact, we can inductively define $K_n = \frac{1}{3}K_{n-1} \cup \left( \frac{2}{3} + \frac{1}{2}K_{n-1} \right)$.

All of these sets are finite collections of closed, bounded intervals, so they are all compact, and they are nested $K_{n+1} \subset K_n$. Hence, by Proposition 5.33, the set $C = \bigcap_n K_n$ is a nonempty compact set. This set is called the **Cantor set**.

What can we say about this set? Well, what is the length of the longest interval in it? Note that $K_n$ consists of $2^n$ intervals, each of length $\frac{1}{3^n}$. Since the length of the intersection of two intervals is $\leq$ the length of either interval, and since $C \subset K_n$ for every $n$, this means $C$ contains no intervals of length $\geq \frac{1}{3^n}$ for any $n \in \mathbb{N}$; but $\frac{1}{3^n} \to 0$ as $n \to \infty$, and therefore $C$ contains no intervals of length $> 0$. This proves that $\overset{\circ}{C} = \emptyset$. Indeed, if $x$ were an interior point of $C$, that would mean $B_r(x) \subset C$ for some $r > 0$; but $B_r(x) = (x - r, x + r)$ is an interval of length $2r > 0$, which we know is not contained in $C$. Thus no point is interior to $C$. At the same time, $C$ is compact, so it is closed. Thus $\overline{C} = C$, and so $\partial C = \overline{C} \setminus \overset{\circ}{C} = C$ – the Cantor set is its own boundary.

That also happens for discrete sets: if $K$ consists entirely of isolated points, then $\overset{\circ}{K}$ is closed and $\overline{K} = \emptyset$, so $\partial K = K$. But the Cantor set is the opposite of a discrete set: it contains no isolated points, so $C$ consists entirely of limit points, $C' = C$. To see this, fix $x \in C$; so $x \in K_n$ for every $n$. Now $K_n$ is a collection of disjoint closed intervals, so there is some interval $I_n \subset K_n$ with $x \in I_n$. Either $x$ is in the interior of this interval or it is one of the endpoints; either way, there is one endpoint $x_n$ of $I_n$ with $x_n \neq x$. Now, from the construction of $C$, the endpoints of all the intervals are in $C$, so $x_n \in C$. Also, as $\text{diam}(I_n) = \frac{1}{3^n} \to 0$ and $x, x_n \in I_n$, we have $d(x, x_n) \to 0$. This $x_n \to x$, but $x_n \neq x$ for any $n$, and $x_n \in C$; this proves that $x \in C'$. Since $x$ was an arbitrary element of $C$, this means $C \subset C'$, and since $C$ is closed, we have $C' \subset C$, so $C = C'$. 

CHAPTER 6

Limits and Continuity


We now begin to study functions. We have, of course, been studying functions (for example sequences, which are functions with domain \( \mathbb{N} \)); now we will concentrate on metric properties of functions. So we will set things up in terms of functions between metric spaces.

**Definition 6.1.** Let \( X \) and \( Y \) be metric spaces. Let \( E \subseteq X \), and let \( x_0 \in E' \) be a limit point of \( E \). Let \( L \in Y \). Now, for any function \( f: E \to Y \), we say \( f(x) \) tends to \( L \) as \( x \to x_0 \), or the limit as \( x \to x_0 \) of \( f(x) \) is \( L \), in symbols

\[
\lim_{x \to x_0} f(x) = L
\]

if: given any sequence \((x_n)\) in \( E \setminus \{x_0\}\) that converges \( x_n \to x_0 \), it follows that the sequence \((f(x_n))\) in \( Y \) converges \( f(x_n) \to L \).

This is a more general kind of limit than the limit of a sequence: we are letting the argument “tend to” a limit point through a set that may be quite different from \( \mathbb{N} \). Our definition makes use of our knowledge of limits of sequences. This is useful, for example, in establishing some of the basic properties of limits. For example:

**Lemma 6.2.** Limits are unique: if \( \lim_{x \to x_0} f(x) = L_1 \) and \( \lim_{x \to x_0} f(x) = L_2 \), then \( L_1 = L_2 \).

**Proof.** Since \( x_0 \) is a limit point, there is a sequence \((x_n)\) with \( x_n \neq x_0 \) for any \( n \) and \( x_n \to x_0 \). By definition of \( \lim_{x \to x_0} f(x) = L_1 \), this means that the sequence \( f(x_n) \) converges to \( L_1 \); by definition of \( \lim_{x \to x_0} f(x) = L_2 \), this means that \( f(x_n) \) converges to \( L_2 \). Thus, by uniqueness of limits of sequences, \( L_1 = L_2 \). \( \square \)

**Remark 6.3.** (1) If we had not included in the definition the fact that \( x_0 \) is a limit point, this argument would fail. Indeed, if \( x_0 \) is an isolated point, vacuously it holds that \( \lim_{x \to x_0} f(x) = L \) for all \( L \).

(2) On the other hand, we might try to modify the definition of limit so that this wouldn’t happen: we could, for example, insist that \( f(x_n) \to L \) for any sequence \( x_n \) that converges to \( x_0 \), even if it does hit \( x_0 \) at some times. But this would rule out some of our intuition about limits, as the following example shows.

**Example 6.4.** Consider the function \( f: \mathbb{R} \to \mathbb{R} \) given by

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

We know from our calculus intuition that \( \lim_{x \to 0} f(x) = 0 \). Indeed, we can verify this from Definition 6.1 if \( x_n \) is any sequence in \( \mathbb{R} \setminus \{0\} \), then \( f(x_n) = 0 \) for all \( n \), and the constant sequence 0 does indeed converge to 0.
On the other hand, suppose we had left out the \( x_n \neq x_0 \) clause in Definition 6.1 and insisted that \( f(x_n) \to L \) for every sequence \( x_n \to x_0 \). In this scenario, the function above would have no limit at 0. Indeed, we could take the sequence \( x_n = \frac{1}{n} \) if \( n \) is even and \( x_n = 0 \) if \( n \) is odd. Then the sequence \( f(x_n) = (0, 1, 0, 1, 0, 1, \ldots) \) has no limit.

This illustrates the fundamental idea of limits: a limit is where a function is going as you approach the limit point; it is unrelated to the actual value of the function at that point (if it is even defined).

We can use our theory of limits of sequences to calculate many limits. For example, if the range space for the function is the familiar \( \mathbb{C} \), we have the following echo of the limit theorems for \( \mathbb{C} \)-sequences:

**Theorem 6.5 (Limit Theorems).** Let \( f, g : X \to \mathbb{C} \), and let \( x_0 \) be a limit point in \( X \). If \( \lim_{x \to x_0} f(x) = L \) and \( \lim_{x \to x_0} g(x) = M \), then

\[
\lim_{x \to x_0} [f(x) + g(x)] = L + M, \quad \text{and} \quad \lim_{x \to x_0} [f(x) \cdot g(x)] = L \cdot M.
\]

**Proof.** Let \((x_n)\) be any sequence in \( X \setminus \{x_0\}\) that converges to \( x_0 \). By assumption, \( f(x_n) \to L \) and \( g(x_n) \to M \). Thus, by the limit theorems for sequences in \( \mathbb{C} \), cf. Theorem 2.27, \( f(x_n) + g(x_n) \to L + M \) and \( f(x_n) \cdot g(x_n) \to L \cdot M \). This is precisely what it means to say that \( \lim_{x \to x_0} [f(x) + g(x)] = L + M \) and \( \lim_{x \to x_0} [f(x) \cdot g(x)] = L \cdot M \).

There is an equivalent definition of limit which does not explicitly rely on sequences. This definition is one of the crowning achievements of 19th Century mathematics. The calculus was built on an intuitive understanding of limits in the minds of Newton and Liebnitz (and others), but it wasn’t until Weierstrass came up with this modern definition that analysis of functions was finally put on rigorous footing.

**Theorem 6.6.** Let \((X, d_X)\) and \((Y, d_Y)\) be metric spaces, let \( E \subseteq X \), let \( f : E \to Y \) be a function, let \( x_0 \in E' \), and let \( L \in Y \). Then \( \lim_{x \to x_0} f(x) = L \) if and only if the following holds true:

\[
\forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall x \in E, \ x \in B_\delta(x_0) \setminus \{x_0\} \implies f(x) \in B_\epsilon(L).
\]

I.e.

\[
\forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall x \in E, \ 0 < d_X(x, x_0) < \delta \implies d_Y(f(x), L) < \epsilon. \quad (6.1)
\]

In words: to say \( f(x) \) tends to \( L \) as \( x \) tends to \( x_0 \) means that, for any tolerance \( \epsilon > 0 \), no matter how small, there is some (potentially even smaller) tolerance \( \delta > 0 \) so that, if \( x \) is \( \delta \)-close to \( x_0 \) (but not equal to \( x_0 \)), then \( f(x) \) is \( \epsilon \)-close to \( L \).

**Proof.** First, suppose that (6.1) holds. Let \( x_n \in E \setminus \{x_0\} \) be a sequence converging to \( x_0 \). Fix \( \epsilon > 0 \), and let \( \delta > 0 \) be the corresponding \( \delta \). Now, as \( x_n \to x_0 \), there is some \( N \in \mathbb{N} \) so that, for \( n \geq N \), \( d_X(x_n, x_0) < \delta \). It follows from (6.1) that \( d_Y(f(x_n), L) < \epsilon \) for all \( n \geq N \). This proves that \( f(x_n) \to L \). Thus, we have shown that \( \lim_{x \to x_0} f(x) = L \) by definition.

Conversely, suppose (6.1) fails to hold. This means that there exists some \( \epsilon > 0 \) so that, for all \( \delta > 0 \), there is some point \( x_\delta \in B_\delta(x_0) \setminus \{x_0\} \) such that \( f(x_\delta) \) is not in \( B_\epsilon(L) \). In particular, do this with \( \delta = \frac{1}{2^n} \) for each \( n \in \mathbb{N} \), choose some \( x_n \in B_{1/2^n}(x_0) \) such that \( d_Y(f(x_n), L) \geq \epsilon \). On the one hand, since \( 0 < d_X(x_n, x_0) < \frac{1}{n} \to 0 \), we have \( x_n \to x_0 \) but \( x_n \neq x_0 \). On the other hand, since \( d_Y(f(x_n), L) \geq \epsilon \) for all \( n \), this means that the sequence \( f(x_n) \) does not converge to \( L \). By definition, this means that the statement \( \lim_{x \to x_0} f(x) = L \) is false. \( \square \)
EXAMPLE 6.7. Let us work directly from the \( \epsilon \)-\( \delta \) definition of (6.1) to show that \( \lim_{x \to 2} x^2 = 4 \). Here the domain and range metric spaces are both \( \mathbb{R} \). Fix \( \epsilon > 0 \). We want to guarantee that \( |x^2 - 4| < \epsilon \). Write this as \( |x - 2||x + 2| < \epsilon \). We want to choose \( \delta > 0 \) and force \( 0 < |x - 2| < \delta \), meaning \( 2 - \delta < x < 2 + \delta \). So, as long as we assure that \( \delta \leq 2 \), this means that \( 0 \leq x \leq 4 \), in which case \( |x - 2||x + 2| \leq 6|x - 2| \). Thus, it suffices to make sure that \( 6|x - 2| < \epsilon \), which is to say \( |x - 2| < \epsilon/6 \). This tells us how to choose \( \delta \).

So, starting fresh: Let \( \epsilon > 0 \). Choose \( \delta = \epsilon/6 \) if this is \( < 2 \), or \( \delta = 2 \) otherwise. Then, so long as \( 0 < |x - 2| < \delta \), we have \( 0 \leq 2 - \delta < x < 2 + \delta \leq 4 \), and so

\[
|x^2 - 4| = |x + 2||x - 2| \leq 6|x - 2| < 6 \cdot \frac{\epsilon}{6} = \epsilon.
\]

Thus, by (6.1), we have proven that \( \lim_{x \to 2} x^2 = 4 \).

On the other hand, if we refer to Theorem 6.5 we see that this follows from the fact that \( \lim_{x \to 2} x = 2 \) (which is easy to verify by either definition of limit) and therefore \( \lim_{x \to 2} x \cdot x = 2 \cdot 2 = 4 \). Similar considerations show that \( \lim_{x \to x_0} f(x) = f(x_0) \) holds for any point \( x_0 \in \mathbb{R} \) (or \( \mathbb{C} \)) if \( f \) is a polynomial, for example.
2. Lecture 19: March 8, 2016

In Example 6.7, what we showed is that the function \( f(x) = x^2 \) satisfies \( \lim_{x \to 2} f(x) = f(2) \). We should recognize this as saying that \( f \) is continuous at 2.

**Definition 6.8.** Let \( X, Y \) be metric spaces, \( E \subseteq X \), and \( f : E \to Y \). Let \( x_0 \in E' \) be a limit point. Say that \( f \) is continuous at \( x_0 \) if

\[
\lim_{x \to x_0} f(x) = f(x_0).
\]

Note that this only defines continuity at limit points: we have left undefined what it would mean for \( f \) to be continuous at an isolated point of its domain of definition. Indeed, what should we mean by saying that a function is continuous on the set \( \mathbb{N} \)? This is, to some degree, up to debate. The standard answer is to say this is a vacuous condition: every function is continuous on a discrete set.

Now, consider again Example 6.7. To use the definition of limit, we assumed that \( d(x, x_0) = |x - 2| > 0 \) (as limits are about where you’re going, not where you get to). However, observe that this requirement was never used in the proof. That is generically true in limits of continuous functions, as the next result demonstrates.

**Proposition 6.9.** Let \( X, Y \) be metric spaces and \( f : X \to Y \). Let \( x_0 \in X' \). Then \( f \) is continuous at \( x_0 \) if and only if for every sequence \((x_n)\) in \( X \) with \( \lim_{n \to \infty} x_n = x_0 \), it follows that \( \lim_{n \to \infty} f(x_n) = f(x_0) \). Similarly, \( f \) is continuous at \( x_0 \) if and only if

\[
\forall \epsilon > 0 \ \exists \delta > 0 \ \text{s.t.} \ \forall x \in E, \ d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \epsilon.
\]

That is: we need not assume that the sequence \((x_n)\) never hits \( x_0 \); and we need not remove \( x_0 \) from the \( \delta \)-ball in the \( \epsilon-\delta \) definition of the limit. In fact, with these assumption no longer required, there is no reason to assume \( x_0 \in X' \); this definition makes perfect sense for isolated points as well, so we take it more generally as the definition of continuity. From this more general definition, it follows that any function is continuous at an isolated point of its domain (as you should work out).

**Proof.** Suppose that \( x_n \to x_0 \) implies \( f(x_n) \to f(x_0) \) in general; then in particular this holds if we also assume that \( x_n \neq x_0 \) for any \( n \), which means that \( \lim_{x \to x_0} f(x) = f(x_0) \) by definition. Thus \( f \) is continuous at \( x_0 \). Conversely (in contrapositive form), suppose there is some sequence \( x_n \to x_0 \) such that \( f(x_n) \not\to f(x_0) \). This means there is some \( \epsilon > 0 \) so that \( d(f(x_n), f(x_0)) \geq \epsilon \) for infinitely many \( n \). So let \( n_1, n_2, n_3, \ldots \) be these infinitely many indices where, for each \( k \), \( d(f(x_{n_k}), f(x_0)) \geq \epsilon \); then \( (x_{n_k})_{k=1}^{\infty} \) is a sequence in \( X \) that converges to \( x_0 \) (as a subsequence of a sequence \( x_n \) which converge to \( x_0 \)), but \( f(x_{n_k}) \not\to f(x_0) \) and, moreover, since \( f(x_{n_k}) \neq f(x_0) \) for any \( k \), it follows that \( x_{n_k} \neq x_0 \) for any \( k \). This shows, from the definition, that it is false that \( \lim_{x \to x_0} f(x) = f(x_0) \), completing the proof of the first statement.

The proof of the equivalence of the \( \epsilon-\delta \) statement is similar and left to the reader. \( \square \)

The point is: when the putative limit is the value of the function at the limit point, there is no reason to exclude the limit point from consideration: where you are going and where you get to are the same in this case!

**Example 6.10.** Let \( (X, d) \) be a metric space, and let \( y \in X \). Then the function \( f(x) = d(x, y) \) is continuous at every point in \( X \). Indeed, fix \( x \in X \), and let \((x_n)\) be a sequence in \( X \) with \( x_n \to x \). Then

\[
d(x_n, y) \leq d(x_n, x) + d(x, y)
\]
and so \(d(x_n, y) - d(x, y) \leq d(x_n, x)\). But also
\[
d(x, y) \leq d(x, x_n) + d(x_n, y)
\]
and so \(d(x, y) - d(x_n, y) \leq d(x, x_n) = d(x_n, x)\). Together, these give
\[
0 \leq |d(x_n, y) - d(x, y)| \leq d(x_n, x).
\]
Since \(x_n \to x\), \(d(x_n, x) \to 0\) by definition, and so by the squeeze theorem \(|d(x_n, y) - d(x, y)| \to 0\), meaning that \(f(x_n) = d(x_n, y) \to d(x, y) = f(x)\). This shows that \(f\) is continuous at \(x_0\).

We would be remiss if we did not include some examples of discontinuous functions.

**Example 6.11.** Let \(f: \mathbb{R} \to \mathbb{R}\) be the function
\[
f(x) = \begin{cases} 
0, & x \not\in \mathbb{Q} \\
1, & x \in \mathbb{Q}.
\end{cases}
\]

Then \(f\) (sometimes called Dirichlet’s function) is not continuous at any point. Indeed, fix \(x \in \mathbb{R}\). For any \(\delta > 0\), the ball \(B_\delta(x) = (x - \delta, x + \delta)\) contains both rational and irrational numbers. So, if \(x \in \mathbb{Q}\), choose some \(y \not\in \mathbb{Q}\) in the ball, and we have \(|f(x) - f(y)| = |1 - 0| = 1\); if \(x \not\in \mathbb{Q}\), choose some \(y \in \mathbb{Q}\) in the ball, and we have \(|f(x) - f(y)| = |0 - 1| = 1\). In any case, we see that for any \(\delta > 0\) there are points \(y \in B_\delta(x)\) so that \(|f(y) - f(x)| = 1\), so we can never force \(f(y)\) to be in, for example, \(B_{\frac{1}{2}}(f(x))\). This shows \(f\) is discontinuous at \(x\), for any \(x\).

**Example 6.12.** Consider the following function \(f: [0, 1] \to [0, 1]\), sometimes called the popcorn function:
\[
f(x) = \begin{cases} 
0, & x \not\in \mathbb{Q} \\
\frac{1}{q}, & x = \frac{p}{q} \text{ in lowest terms}.
\end{cases}
\]

The graph of this function looks like this:

![Graph of the popcorn function]

In fact, \(f\) is discontinuous at all rational points, but it is actually continuous at all irrational points. Indeed, let \(x = \frac{p}{q}\) be rational, and let \(x_n = x + \frac{\sqrt{2}}{n}\) for all \(n\) large enough that this is in \([0, 1]\); then \(x_n \to x\). Then \(x_n \not\in \mathbb{Q}\) meaning that \(f(x_n) = 0\); but \(f(x) = \frac{1}{q} \neq 0\), so \(f(x_n) \neq f(x)\). On the other hand, let \(x \not\in \mathbb{Q}\); we want to show that \(f\) is continuous at \(x\), meaning
\[
\lim_{y \to x} f(y) = f(x) = 0. \text{ Fix } \epsilon > 0, \text{ and choose some integer } n \in \mathbb{N} \text{ with } \frac{1}{n} < \epsilon. \text{ As } f(x) \geq 0 \text{ for all } x, \text{ it suffices to show that } f(y) < \frac{1}{n} \text{ for all } y \text{ sufficiently close to } x. \text{ Well, if } y \text{ is a point in } [0, 1] \text{ where } f(y) \geq \frac{1}{n}, \text{ then } y \in \mathbb{Q} \text{ and, when written in lowest terms, } y = \frac{\xi}{\eta} \text{ with } \eta \leq n. \text{ There are only finitely many such rational numbers, and } x (\text{which is irrational}) \text{ is not one of them. Thus, we can define } \\
\delta = \min \{|x - y| : y = \frac{\xi}{\eta} \text{ in lowest terms, with } \eta \leq n\}; \text{ then for } |y - x| < \delta, \text{ it follows that } f(y) < \frac{1}{n} < \epsilon, \text{ proving that } \\
\lim_{y \to x} f(y) = 0, \text{ and so } f \text{ is continuous at } x. \]

In Examples 6.11 and 6.12, we looked at the set of points where a function is continuous. That is: if \( f : X \to Y \) is a function and \( E \subseteq X \), say that \( f \) is continuous on \( E \) if, for each \( x \in E \), \( f \) is continuous at \( x \).

**Example 6.13.** Let \( X = (0, 1) \), and let \( f : (0, 1) \to \mathbb{R} \) be the function \( f(x) = \frac{1}{x} \). Then \( f \) is continuous on its whole domain: for every \( x \in (0, 1) \), \( f \) is continuous at \( x \). We could see this by applying the limit theorems; but let’s use this as an opportunity to practice our \( \epsilon-\delta \) proofs. Fix \( \epsilon > 0 \). We want to guarantee that, when \( y \) is close to \( x \), we have \( \frac{1}{x} - \epsilon < \frac{1}{y} < \frac{1}{x} + \epsilon \).

We only need this to hold for all sufficiently small \( \epsilon > 0 \), so it’s fine to assume \( \epsilon \) is small enough that \( \frac{1}{x} - \epsilon > 0 \). Thus we can reciprocate to get

\[
\frac{1}{1/x - \epsilon} > y > \frac{1}{1/x + \epsilon}. 
\]

Now, subtract \( x \) from both sides and we have

\[
\frac{-\epsilon x}{1/x + \epsilon} = \frac{1}{1/x + \epsilon} - x < y - x < \frac{1}{1/x - \epsilon} - x = \frac{\epsilon x}{1/x - \epsilon}. 
\]

This shows us how to choose \( \delta \): we define

\[
\delta = \min \left\{ \frac{\epsilon x}{1/x + \epsilon}, \frac{\epsilon x}{1/x - \epsilon} \right\} = \frac{\epsilon x}{1/x + \epsilon}. \quad (6.2)
\]

Then, reversing the above steps, we have that for any \( y \in B_\delta(x) \), we have \( |y - x| < \frac{\epsilon x}{1/x + \epsilon} < \frac{\epsilon x}{1/x - \epsilon} \), and this gives in particular the above two inequalities that can be reversed to say \( |\frac{1}{x} - \frac{1}{y}| < \epsilon \). So we have proved that there is a \( \delta > 0 \) for any given \( \epsilon > 0 \) (as long as \( \epsilon < \frac{1}{x} \); otherwise, if \( \epsilon \geq \frac{1}{x} > 1 \), we could take \( \delta \) to be something silly and big), proving continuity at \( x \).
3. Lecture 20: March 10, 2016

In Example 6.13, we showed explicitly that the function \( f(x) = \frac{1}{x} \) is continuous at every point \( x \in (0, 1) \). But note: the \( \delta \) we had to choose for each \( \epsilon > 0 \) in (6.2) depends on \( x \) as well as \( \epsilon \). This will generically be true. Look at the \( \epsilon-\delta \) definition of continuity: a function \( f \) is continuous on a set \( E \) if

\[
\forall x \in E \forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall y \in E, \ d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \epsilon. \tag{6.3}
\]

Having chosen a \( x \) and \( \epsilon > 0 \), we must then find a suitable \( \delta = \delta(x, \epsilon) \). In Example 6.13, not only does \( \delta \) depend on \( x \), but it does so in a bad way: as \( x \to 0 \), for given \( \epsilon > 0 \), the \( \delta \to 0 \) as well (quite fast, in fact: the numerator is shrinking and the denominator is growing). The closer \( x \) is to 0, the smaller \( \delta \) must be to get the same control over the function. So, while the function is continuous, there is a lack of uniformity in how continuous it is. (Note: we have shown this \( \delta \) works; one might ask whether a larger, possibly more uniform \( \delta \) could work just as well. The answer is no: it is not hard to show in this example that the \( \delta \) in (6.2) is the largest possible \( \delta \) for the given \( x \) and \( \epsilon \); it is called the modulus of continuity of the function.)

**Definition 6.14.** Let \( X, Y \) be metric spaces, \( E \subseteq X \), and \( f: E \to Y \). Call \( f \) uniformly continuous on \( E \) if

\[
\forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall x, y \in E, \ d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \epsilon. \tag{6.4}
\]

Compare (6.4) with (6.3). The difference appears subtle: just the placement of the quantifier \( \forall x \). This makes a world of difference: (6.4) says that, not only is \( f \) continuous at each point \( x \), but one can choose a \( \delta = \delta(\epsilon) \) that is uniform: it need not depend on \( x \). This outlaws behavior like the function \( f(x) = \frac{1}{x^2} \) near 0.

**Example 6.15.** Let \( f(x) = 2x \) on \( \mathbb{R} \). Then \( |f(x) - f(y)| = 2|x - y| \), so for any \( \epsilon > 0 \), we may let \( \delta = \epsilon/2 \); then if \( |x - y| < \delta = \epsilon/2 \), it follows that \( |f(x) - f(y)| = 2|x - y| < 2\delta = \epsilon \). This shows that \( f \) is continuous at all points; moreover, we may choose \( \delta = \delta(\epsilon) = \epsilon/2 \) uniformly over all \( x, y \in \mathbb{R} \). Thus, \( f \) is uniformly continuous on \( \mathbb{R} \).

**Example 6.16.** Let \( f(x) = x^2 \) on \([0, \infty)\). We want to make \(|x^2 - y^2|\) small. We have \(|x^2 - y^2| = (x+y)|x - y| \). Thus, in order for \(|x^2 - y^2| < \epsilon \), we must have \(|x - y| < \frac{\epsilon}{x+y} \) (these are equivalent). But this shows that \( f \) is not uniformly continuous. Indeed, in order for \(|x - y| < \delta \) to imply that \(|x - y| < \epsilon \), we must have \( \delta \leq \frac{\epsilon}{x+y} \); and there is no positive number \( \delta = \delta(\epsilon) \) that is \( \leq \frac{\epsilon}{x+y} \) for all \( x, y > 0 \).

In Examples 6.13 and 6.16, we saw continuous functions on the intervals \((0, 1)\) and \([0, \infty)\) that are not uniformly continuous. In both cases, the non-uniformity was manifest by uncontrolled growth near the “edge”. As it turns out, if the domain of the continuous function is compact, this cannot happen. That will be our final big theorem of this class.

**Theorem 6.17.** Let \( X, Y \) be metric spaces, \( K \subseteq X \) compact, and \( f: K \to Y \) a continuous function. Then \( f \) is uniformly continuous.

**Proof.** Suppose, for a contradiction, that \( f \) is not uniformly continuous on \( K \). Negating Definition 6.14, this means

\[
\exists \epsilon > 0 \forall \delta > 0 \exists x, y \in K \text{ s.t. } d_X(x, y) < \delta, \text{ but } d_Y(f(x), f(y)) \geq \epsilon.
\]

That is: there is a positive number \( \epsilon > 0 \) so that, for every positive number \( \delta > 0 \), we can find two points \( x \) and \( y \) that are within distance \( \delta \) of each other, but such that \( f(x) \) and \( f(y) \) are at
least $\epsilon$ apart. So, let’s do this with $\delta = \frac{1}{n}$ for any given positive integer: we can find $x_n, y_n$ with $d_X(x_n, y_n) < \frac{1}{n}$, and yet $d_Y(f(x_n), f(y_n)) \geq \epsilon$.

Now, we use the compactness of the domain $K$: the sequence $(x_n)$ has a convergent subsequence $(x_{n_k})$ with a limit $x \in K$. Consider, now, the corresponding subsequence $y_{n_k}$; this has a convergent subsequence $y_{n_{k\ell}}$ with a limit $y \in K$. Now, $x_{n_{k\ell}}$ is a subsequence of $x_{n_k}$ which converges to $x$, hence $x_{n_{k\ell}} \to x$ as well. But we also have

$$d_X(x_{n_{k\ell}}, y_{n_{k\ell}}) < \frac{1}{n_{k\ell}} < \frac{1}{\ell} \to 0.$$  

Hence, it follows from the triangle inequality that $x = y$. On the other hand, by their very construction, the points $x_{n_{k\ell}}$ and $y_{n_{k\ell}}$ all satisfy

$$d_Y(f(x_{n_{k\ell}}), f(y_{n_{k\ell}})) \geq \epsilon. \quad (6.5)$$

But $x_{n_{k\ell}} \to x$ and so, since $f$ is continuous, $f(x_{n_{k\ell}}) \to f(x)$; similarly, $y_{n_{k\ell}} \to y = x$, and so by continuity $f(y_{n_{k\ell}}) \to f(y) = f(x)$. Thus, by Problem 4 on Exam 2,

$$d_Y(f(x_{n_{k\ell}}), f(y_{n_{k\ell}})) \to d_Y(f(x), f(x)) = 0.$$  

This contradicts (6.5). Thus, we have proven that $f$ is, in fact, uniformly continuous. \qed

Theorem 6.17 is typically the best way to prove uniform continuity of a function. For example: any polynomial is continuous on $\mathbb{R}$, but, as we saw in Example 6.16, they need not be uniformly continuous. By Theorem 6.17, polynomial functions on compact intervals $[a, b]$ are automatically uniformly continuous. What’s more: once you know a function is uniformly continuous on a set $K$, it is then automatically uniformly continuous on any subset $E \subseteq K$ (the same $\delta = \delta(\epsilon)$ that works on all of $K$ also works on all of $E \subseteq K$). So, for example, polynomials are uniformly continuous on all bounded intervals $(a, b), (a, b]$, etc. Similarly, the function $f(x) = \frac{1}{x}$ of Example 6.13 is uniformly continuous on $[\alpha, 1]$ for any $\alpha > 0$. We could see this directly from (6.2), since the modulus of continuity

$$\delta = \delta(x, \epsilon) = \frac{\epsilon x}{1/x + \epsilon}$$

decreases as $x$ decreases; it follows that the uniform $\delta = \delta(\alpha, \epsilon)$ will work for all $x \geq \alpha$. However, this gets smaller as $\alpha$ shrinks, and if we include all of $(0, 1]$ in the domain, there is no uniform $\delta$. For an alternate proof of the non-uniformity in this example, see HW10.4.

Here is another very useful property of continuous functions on compact sets.

**Proposition 6.18.** Let $X, Y$ be metric spaces, $K \subseteq X$ compact, and $f : K \to Y$ continuous (hence uniformly continuous). Then the image $f(K) \subseteq Y$ is compact.

To be clear: $f(K)$ denotes the image of $f$ on $K$:

$$f(K) = \{ f(x) \in Y : x \in K \} = \{ y \in Y : \exists x \in K \text{ s.t. } y = f(x) \}.$$  

**Proof.** Let $(y_n)$ be any sequence in $f(K)$. By definition of $f(K)$, for each $y_n$, there exists some (or potentially many) $x_n \in K$ such that $y_n = f(x_n)$. Since $K$ is compact, it then follows that the sequence $(x_n)$ has a convergent subsequence $(x_{n_k})$ with limit $x \in K$. Since $f$ is continuous, it then follows that $f(x_{n_k}) \to f(x)$ as $k \to \infty$. Since $x \in K$, $f(x) \in f(K)$. Thus, the subsequence $y_{n_k} = f(x_{n_k})$ of $y_n = f(x_n)$ converges in $K$. We have thus shown that every sequence in $f(K)$ has a convergent subsequence with limit in $f(K)$; that is, $f(K)$ is compact. \qed
Corollary 6.19 (Minimax Theorem). Let $K$ be a nonempty compact metric space, and $f : K \to \mathbb{R}$. Then $f$ attains its maximum and minimum values on $K$.

Corollary 6.19 is a standard result stated in calculus classes, usually in the special case that $K = [a,b]$ is a compact interval in $\mathbb{R}$.

Proof. By Proposition 6.18, $f(K)$ is compact. In particular, it is closed and bounded. It is also nonempty since $K$ is nonempty (so $f(K)$ contains $f(x)$ for any $x \in K$). Thus, by the least upper bound property of $\mathbb{R}$, the set $f(K) \subseteq \mathbb{R}$ has a supremum $M$ and and infimum $m$. Now, for any $n \in \mathbb{N}$, $M - \frac{1}{n} < M$, which means that $M - \frac{1}{n}$ is not an upper bound for $f(K)$; thus, there is some $y_n \in f(K)$ with $M - \frac{1}{n} < y_n \leq M$. Hence, by the Squeeze Theorem, $y_n \to M$. By definition of $f(K)$, there exists some $x_n \in K$ with $y_n = f(x_n)$. Since $K$ is compact, there is a convergent subsequence $(x_{n_k})$ of $(x_n)$, with limit $x \in K$. Since $f$ is continuous, $y_{n_k} = f(x_{n_k}) \to f(x)$. But $y_{n_k}$ is a subsequence of $y_n$ which converges to $M$; thus $f(x) = M$. We have therefore found a $x \in K$ for which $f(x) = M = \sup f(K)$. That is: $\sup f(K) = \max f(K)$, and the maximum is achieved at the point $x$. A very similar argument shows there is a point $x'$ with $f(x') = m = \inf f(K)$, completing the proof. \qed
Part 2

Math 140B
CHAPTER 7

More on Continuity

1. Lecture 1: March 29, 2016

Presently, we return to the definition(s) of continuity, and consider more purely topological characterizations. As such, we will not be dealing with continuity at a point, but instead continuity on a set (or most of the time on the whole domain metric space), which simply means pointwise continuity at each point in the domain. To begin, we note that continuity is well-behaved under composition.

**Proposition 7.1.** Let $X, Y, Z$ be metric spaces, and suppose $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are continuous functions. Then the composite function $g \circ f: X \rightarrow Z$ is continuous.

**Proof 1.** Here we see very clearly the power of the sequential definition of continuity (Definition 6.8). Let $x \in X$, and let $(x_n)$ be a sequence that converges to $x$. Since $f$ is continuous on $X$, hence at $x$, it follows that $f(x_n) \rightarrow f(x)$. Since $(f(x_n))$ is a sequence in $Y$ that converges to $f(x)$, and since $g$ is continuous on $Y$, hence at $f(x)$, it follows that $g(f(x_n)) \rightarrow g(f(x))$. But $g(f(x_n)) = (g \circ f)(x_n)$, and $g(f(x)) = (g \circ f)(x)$. It follows that $g \circ f$ is continuous at (an arbitrarily chosen) $x \in X$. □

**Proof 2.** It is also possible to prove the proposition using the $\varepsilon$-$\delta$ definition of continuity (Proposition 6.9). Fix $x \in X$ and $\varepsilon > 0$. Since $g$ is continuous on $Y$, hence at $f(x)$, there is some $\delta > 0$ so that $g(B_\delta(f(x))) \subseteq B_\varepsilon(g(f(x)))$. Now, since $f$ is continuous on $X$, hence at $x$, there is some $\delta' > 0$ so that $f(B_{\delta'}(x)) \subseteq B_\delta(f(x))$. Now, given two subsets $A, B \subseteq Y$ with $A \subseteq B$, it follows that $g(A) \subseteq g(B)$ by definition. Thus

$$(g \circ f)(B_{\delta'}(x)) = g(f(B_{\delta'}(x))) \subseteq g(B_\delta(f(x))) \subseteq B_\varepsilon(g(f(x))) = B_\varepsilon((g \circ f)(x)).$$

Thus, for each $\varepsilon > 0$ we can choose $\delta' > 0$ with $(g \circ f)(B_{\delta'}(x)) \subseteq B_\varepsilon((g \circ f)(x))$, which shows $g \circ f$ is continuous at (an arbitrarily chosen) $x$. □

In Proof 2 above, we used a property of set mappings (that $A \subseteq B \implies f(A) \subseteq f(B)$). In order to state and prove our next theorem, we need a more thorough understanding of the behavior of set mapping; the following discussion is purely set theoretic.

**Definition 7.2.** Let $X$ be a set. Denote by $2^X$ the **power set** of $X$: the set of all subsets of $X$. Let $Y$ be another set, and suppose $f: X \rightarrow Y$ is a function. Then there is an induced function (also denoted $f$) from $2^X$ to $2^Y$: for any subset $A \subseteq X$, $f(A) = \{ f(x) : x \in A \}$. It also induces a reverse map $f^{-1}: 2^Y \rightarrow 2^X$: for $B \subseteq Y$, $f^{-1}(B) = \{ x \in X : f(x) \in B \}$.

Note: it is not necessary for $f$ to be one-to-one in order for $f^{-1}$ to exist as a set mapping: it *always* exists. In fact, $f^{-1}$ generally has better properties than $f$ as a set mapping.

**Lemma 7.3.** Let $X$ and $Y$ be sets, and $f: X \rightarrow Y$.

1. For any $B \subseteq Y$, $f^{-1}(B^c) = f^{-1}(B)^c$.
7. MORE ON CONTINUITY

(2) For any $B_1, B_2 \subseteq Y$, $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$, and $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

(3) For any $A_1, A_2 \subseteq X$, $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$, while in general $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$.

**Proof.** (1) An element $x \in X$ is in $f^{-1}(B)^c$ if and only if $x \notin f^{-1}(B)$. By definition, this is the same as saying that $f(x) \notin B$, which is to say that $f(x) \in B^c$, or equivalently $x \in f^{-1}(B)^c$, as desired.

(2) An element $x \in X$ is in $f^{-1}(B_1 \cup B_2)$ iff $f(x) \in B_1 \cup B_2$. If $f(x) \in B_1$ then $x \in f^{-1}(B_1)$; if $f(x) \in B_2$ then $x \in f^{-1}(B_2)$; altogether, this means that $f(x) \in B_1 \cup B_2$ iff $x \in f^{-1}(B_1) \cup f^{-1}(B_2)$, as desired. The argument for intersections is similar.

(3) For the first statement, let $y \in f(A_1) \cup f(A_2)$. So either there is an $x_1 \in A_1$ with $f(x_1) = y$, or there is an $x_2 \in A_2$ with $f(x_2) = y$. In either case, $y \in f(A_1 \cup A_2)$, as desired. Conversely, if $y \in f(A_1 \cup A_2)$, then there is some $x \in A_1 \cup A_2$ with $y = f(x)$. Then either $x \in A_1$, in which case $y = f(x) \in f(A_1)$, or $x \in A_2$, in which case $y = f(x) \in f(A_2)$; thus $y \in f(A_1) \cup f(A_2)$, as desired.

For the second statement, let $y \in f(A_1 \cap A_2)$; so there is some $x \in A_1 \cap A_2$ with $f(x) = y$. Since $x \in A_1$, it follows that $y = f(x) \in f(A_1)$; since $x \in A_2$, it follows that $y = f(x) \in f(A_2)$; thus $y \in f(A_1) \cap f(A_2)$, as desired.

**Remark 7.4.** Nothing like item (1) holds for the forward set mapping in general. On the one hand, if $f$ is a bijection with inverse function $g$, then $f(A) = g^{-1}(A)$ (you should untwist the definitions to check this), and so in this case the forward set mapping has all the same nice properties as the inverse. On the other hand, suppose $X$ has more than one element, and $f$ is a constant map $f(x) = y_0$ for all $x \in X$. Then for any $A \subseteq X$, $f(A^c) = \{y_0\}$, while $f(A)^c = \{y_0\}^c$, so the two sets are not only not equal, they are complementary (meaning in general there are not even any consistent inclusions of one into the other).

Similarly, the inclusion $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$ is, in general, not an equality if $f$ is not a bijection. For example, take again a constant function $f(x) = y_0$, defined on a set $X$ containing at least two distinct points $x_1$ and $x_2$. Then $f(\{x_1\}) \cap f(\{x_2\}) = \{y_0\}$, but $\{x_1\} \cap \{x_2\} = \emptyset$, so $f(\{x_1\} \cap \{x_2\}) = \emptyset \subseteq \{y_0\}$. So, we see, in general the forward set mapping is not as well-behaved as the inverse set mapping with respect to the Boolean operations; i.e. the inverse set mapping is always a Boolean homomorphism.

The nice behavior of the inverse setting mapping helps to explain why it appears in the following topological characterization of continuity, instead of the forward set mapping.

**Theorem 7.5.** Let $X$ and $Y$ be metric spaces. A function $f : X \to Y$ is continuous if and only if the preimage of any open set is open; i.e. for all open sets $V \subseteq Y$, $f^{-1}(U)$ is open in $X$.

**Proof.** First, suppose $f$ is continuous. Let $V \subseteq Y$ be an open set, and let $x \in f^{-1}(V)$; this means that $f(x) \in V$. Since $V$ is open, there is some $\epsilon > 0$ with $B_\epsilon(f(x)) \subseteq V$. Since $f$ is continuous, there is some $\delta > 0$ with $f(B_\delta(x)) \subseteq B_\epsilon(f(x)) \subseteq V$. But that means that $B_\delta(x) \subseteq f^{-1}(V)$. Hence, every $x \in f^{-1}(V)$ is an interior point of $f^{-1}(V)$, and so $f^{-1}(V)$ is open.

Conversely, suppose we know that $f^{-1}(V)$ is open for every open $V \subseteq Y$. Let $x \in X$ and fix $\epsilon > 0$. By assumption, $f^{-1}(B_\epsilon(f(x)))$ is open. Since $x$ is a point in this open preimage, that
means there is some $\delta > 0$ with $B_\delta(x) \subseteq f^{-1}(B_\epsilon(f(x)))$. Hence $f(B_\delta(x)) \subseteq B_\epsilon(f(x))$, which shows that $f$ is continuous at (an arbitrarily chosen) $x \in X$. \qed

**Remark 7.6.** In the last step of the above proof, we essentially used the fact that $f(f^{-1}(B)) \subseteq B$, which is left as an exercise for you to verify. The reverse containment $f^{-1}(f(A)) \supseteq A$ also holds true in general; in both cases, the containments are generally not equalities.

The above proof is quite natural, and really shows that the $\epsilon$-$\delta$ definition of continuity fundamentally says that preimages of open sets are open. However, combining Theorem 7.5 with Lemma 7.3(1) gives an alternate characterization which is less clear (and therefore more interesting).

**Corollary 7.7.** Let $X$ and $Y$ be metric spaces. A function $f : X \to Y$ is continuous if and only if the preimage of any closed set is closed; i.e. for all closed sets $C \subseteq Y$, $f^{-1}(C)$ is closed in $X$.

**Proof.** By Theorem 7.5, it suffices to show the preimages of closed sets are closed if and only if preimages of open sets are open. But this follows from 7.3(1) together with Proposition 5.15: the complement of an open set is closed. Precisely: suppose the preimage of any open set is open. Let $C$ be a closed set in $Y$. Then $C^c$ is open, and so $f^{-1}(C^c)$ is open. But $f^{-1}(C^c) = f^{-1}(C)^c$, and so $f^{-1}(C)$ is a set whose complement is open, meaning that it is closed. The converse is very similar. \qed

**Example 7.8.** Consider the metric space $S = \{e^{i\theta} : \theta \in \mathbb{R}\} \subset \mathbb{C}$, the unit circle in the complex plane. As a subset of $\mathbb{C}$, it inherits the Euclidean $\mathbb{C}$-metric. But this might not be the most natural metric to use. For example, the distance between the points $(1, 0)$ and $(-1, 0)$ is 2 in the $\mathbb{C}$-metric; but if we are thinking of distance as the shortest path in the circle, then the distance should be $\pi$. To try to define this intrinsic metric, we’d like to have a correspondence between $e^{i\theta}$ and $\theta$. Of course, we cannot do this for all $\theta \in \mathbb{R}$, but we can restrict the allowed $\theta$ to be in $[0, 2\pi)$, and define

$$f(e^{i\theta}) = \theta, \quad 0 \leq \theta < 2\pi.$$ 

This is well-defined (by the polar decomposition: every nonzero complex number $z$ has a unique decomposition $z = |z|e^{i\theta}$ for some $\theta \in [0, 2\pi)$). But is this a continuous function $S \to \mathbb{R}$? Naively, we might write $f(u) = -i \ln u$ and expect this means it is continuous. However, consider the closed set $[\pi, 10] \subset \mathbb{R}$; the preimage of this set is

$$f^{-1}([\pi, 10]) = f^{-1}([\pi, 2\pi]) = f^{-1}([\pi, 2\pi] \cup [2\pi, 10]) = f^{-1}([\pi, 2\pi]) \cup f^{-1}([2\pi, 10]).$$

Since the image of $f$ does not intersect $[2\pi, 10]$, this second preimage is empty, and so $f^{-1}([\pi, 10]) = \{e^{i\theta} : \pi \leq \theta < 2\pi\}$, which is the bottom half of $S$, including the point $(-1, 0)$ but not including the point $(1, 0)$. This set is not closed in $S$: the sequence $e^{-\pi/n}$ lives in this bottom half, and converges to $(1, 0)$ as $n \to \infty$. Thus, $f$ is not continuous.

**Remark 7.9.** It is, indeed, possible to define a metric on $S$ that represents this “length of shortest path” intuition; but this example shows that one cannot draw a global correspondence between $S$ and an interval in $\mathbb{R}$ to make it work. Rather, one must work only locally. This is a topic for a course in differentiable manifolds.

Example 7.8 illustrates a very interesting point. The function $f$ in that example is a bijection; let’s consider its inverse $g : [0, 2\pi) \to S$, which is simply $g(\theta) = e^{i\theta}$. This function is continuous (although we will not prove this until we study sequences and series of functions, in the next chapter). So it is possible for a continuous bijection to have an inverse that is not continuous.
Indeed, in this example, $g$ “glues together” the points $0$ and $2\pi$ at the boundary of its domain, and so its inverse must “rip them apart”, which is discontinuous.

This kind of pathology does not happen, however, if the domain of the bijection is compact.

**Theorem 7.10.** Let $X$ and $Y$ be metric spaces, and suppose $X$ is compact. If $f: X \to Y$ is a continuous bijection, then $f^{-1}: Y \to X$ is continuous.

**Proof.** To avoid confusing notation, denote the inverse function $f^{-1} = g$. Let $C \subseteq X$ be closed. The preimage $g^{-1}(C)$ is equal to the forward image $f(C)$. Since $X$ is compact and $C \subseteq X$ is closed, $C$ is compact (cf. Proposition 5.27). Now, by Proposition 6.18 it follows that $f(C)$ is compact, and therefore closed (again by Proposition 5.27). Thus, $g^{-1}(C)$ is closed for every closed $C$, and so by Corollary 7.7, $g$ is continuous. □

A continuous bijection whose inverse is continuous is called a **homeomorphism**. The function $g(\theta) = e^{i\theta}$ from $[0, 2\pi)$ onto $S$ is a continuous bijection, but it is not a homeomorphism. In topology, homeomorphisms are the basic “isomorphisms”; they tell you when two structures are topologically indistinguishable. The circle $S$ and the half-open interval $[0, 2\pi)$ are topologically distinguishable (this is intuitively clear, but proving that there exists no homeomorphism between them requires developing tools beyond the scope of this course).
Before proceeding with more discussion on continuity, we return to a purely topological notion.

**Definition 7.11.** Let $X$ be a metric space. Two subsets $A$ and $B$ in $X$ are called **separated** if $\overline{A} \cap B = A \cap \overline{B} = \emptyset$; that is, no point in $A$ is in the closure of $B$, and no point in $B$ is in the closure of $A$. The metric space $X$ is called **connected** if it is not the union of two nonempty separated sets.

**Example 7.12.** The two intervals $(0, 1)$ and $(2, 3)$ are separated in $\mathbb{R}$; in fact, their closures $[0, 1]$ and $[2, 3]$ are disjoint (which is nominally stronger than separation). Similarly, the two intervals $(0, 1)$ and $(1, 2)$ are separated: $[0, 1]$ does not intersect $(1, 2)$, and $(0, 1)$ does not intersect $[1, 2]$. However, $(0, 1)$ and $[1, 2]$ are not separated: the closure $[0, 1]$ does intersect $[1, 2]$.

Two separated sets are, of course, disjoint, but as the example points out, disjointness is not sufficient to imply separation.

In $\mathbb{R}$, connected sets can be easily characterized; they are precisely the sets we’ve been most concerned with: intervals.

**Proposition 7.13.** A subset $E \subseteq \mathbb{R}$ is connected if and only if it is an interval: i.e. if and only if it has the property that, for all $x, y, z \in E$, $x < z < y$ and $x, y \in E$, then $z \in E$.

More precisely, the property in Proposition 7.13 should be called the **intermediate value property** (of subsets in an ordered set). It is equivalent to insisting that $E$ is an interval (with one or both endpoints included or not, or possibly infinite). Indeed, it is straightforward to see that intervals have the intermediate value property. The converse is a case analysis. For example, suppose $E \neq \emptyset$ has finite sup and inf. If $\sup E = \inf E$ then $E$ consists of a single point which is an interval. Otherwise, $\inf E < \sup E$; let $z$ be in between. Since $z < \sup E$, there exists some point $y \in E$ with $z < y \leq \sup E$. Similarly, since $z > \inf E$, there exists some point $x \in E$ with $\inf E \leq x < z$. Thus $x < z < y$, and by the intermediate value property, $z \in E$. This shows that $(\inf E, \sup E) \subseteq E$. Of course $E$ contains no points bigger than $\sup E$ or smaller than $\inf E$; so this shows $E$ is one of the four intervals whose closure is $(\inf E, \sup E)$ (with neither, either one, or both of the endpoints included). The argument is similar when one or both of the sup and inf are infinite.

**Proof of Proposition 7.13.** First, suppose that $E$ does not have the intermediate value property: there are points $x < z < y$ with $x, y \in E$ but $z \notin E$. Define $A = (-\infty, z) \cap E$ and $B = (z, \infty) \cap E$; since $z \notin E$, it follows that $E = A \cup B$. Since $x \in A$ and $y \in B$, both sets are nonempty. Now, $z$ is an upper bound for $A$, so $\sup E \leq z$. This means that $\overline{A} \subseteq (-\infty, z]$; this follows from the squeeze theorem: if $a_n \in E$ and $a_n \to a$ then, since $a_n < z$ for all $n$, $a \leq z$. Hence, since $B \subseteq (z, \infty)$, it follows that $\overline{A} \cap B = \emptyset$. A similar argument shows that $A \cap \overline{B} = \emptyset$. Thus $E = A \cup B$ with $A$ and $B$ nonempty and separated; thus $E$ is not connected.

For the converse, suppose that $E$ is not connected, and let $E = A \cup B$ with $A$ and $B$ nonempty and separated. Choose $x \in A$ and $y \in B$; we’ll assume $x < y$ (if they’re reversed order, just rename them). Set $z = \sup(A \cap [x, y])$. By Example 5.21, $z \in \overline{A}$, and therefore $z \notin B$ (since $A$ and $B$ are separated). Since $y \in B$, $z \neq y$. We now consider two cases.

- Suppose $z \notin A$. Since $x \in A$, $z \neq x$. But $z \in [x, y]$, therefore $x < z < y$. But $z \notin A$ and $z \notin B$, so $z \notin E = A \cup B$. Therefore, $E$ does not have the intermediate value property.
- Suppose $z \in A$. Since $A$ and $B$ are separated, $z \notin \overline{B}$, so $z \in \overline{B}^c$, which is an open set. Thus there is some $\epsilon > 0$ so that $B_{\epsilon}(z) \subseteq \overline{B}^c$, and so $z' = z + \frac{\epsilon}{2}$ is not in $\overline{B}$, therefore not
in $B$; therefore $z' \neq y$. Also $z' > z$ which is an upper bound for $A$, so $z' \notin A$. Thus we have $x \leq z < z' < y$, and $z' \notin E = A \cup B$. Therefore $E$ does not have the intermediate value property.

Connectedness is a topological property: it is invariant under homeomorphisms. In fact, it is invariant under any continuous map.

**THEOREM 7.14.** Let $X$ and $Y$ be metric spaces, and let $f : X \to Y$ be continuous. If $E \subseteq X$ is connected, then $f(E)$ is connected.

**PROOF.** We argue the contrapositive: suppose that $f(E)$ is not connected. That is, $f(E)$ a union of two nonempty separated sets: $f(E) = A \cup B$ where $A, B \neq \emptyset$ and $\overline{A} \cap B = A \cap \overline{B} = \emptyset$. Then let $G = E \cap f^{-1}(A)$ and $H = F \cap f^{-1}(B)$. Since $E \subseteq f^{-1}(f(E)) = f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$, it follows that $E = G \cup H$. We will show that $G$ and $H$ are nonempty and separated, and therefore $E$ is not connected.

First, note that $G$ and $H$ are nonempty. Indeed, if $G$ were empty, then $f^{-1}(A)$ would not intersect $E$, so no element of $E$ is mapped into $A$. Since $f(E) = A \cup B$, this would mean that $f(E) = B$, and since $A \cap B = \emptyset$, that would imply $A = \emptyset$, contradicting the hypothesis. A similar argument shows that $H$ is nonempty.

Now, since $A \subseteq \overline{A}$, $G \subseteq f^{-1}(A) \subseteq f^{-1}(\overline{A})$. Since $\overline{A}$ is closed, $f^{-1}(\overline{A})$ is closed (this is the only point in the proof where continuity is used!), and it follows that $\overline{G} \subseteq f^{-1}(\overline{A})$, which is to say that $f(\overline{G}) \subseteq \overline{A}$. On the other hand, using Lemma 7.3(3), we have $f(H) = f(E \cap f^{-1}(B)) \subseteq f(E) \cap f(f^{-1}(B)) \subseteq (A \cup B) \cap B = B$. Thus, applying Lemma 7.3(3) once more,

$$f(\overline{G} \cap H) \subseteq f(\overline{G}) \cap f(H) \subseteq \overline{A} \cap B = \emptyset.$$  

Since $\overline{G} \cap H$ is a set in the domain of $f$, it follows that $\overline{G} \cap H = \emptyset$. An entirely analogous argument demonstrates that $G \cap \overline{H} = \emptyset$.

Thus $E$ is a union of two nonempty separated sets, and so $E$ is not connected. \hfill \Box

Combining Theorem 7.14 with Proposition 7.13 (characterizing connected sets in $\mathbb{R}$) yields another big calculus theorem: the Intermediate Value Theorem.

**COROLLARY 7.15 (Intermediate Value Theorem).** Let $a < b$ be in $\mathbb{R}$ if $f : [a, b] \to \mathbb{R}$ be continuous. For any $y$ between $f(a)$ and $f(b)$, there exists and $x \in [a, b]$ where $f(x) = y$.

**PROOF.** By Proposition 7.13, the interval $[a, b]$ is connected. Thus, by Theorem 7.14, $f([a, b])$ is also connected. Note that $f(a)$ and $f(b)$ are two points in $f([a, b])$; hence, by Proposition 7.13, if $y$ is between $f(a)$ and $f(b)$, then $y \in f([a, b])$, which is precisely to say that there exists an $x \in [a, b]$ where $f(x) = y$. \hfill \Box
It is tempting to think that the Intermediate Value Theorem actually characterizes continuous functions, but this is not so. To be precise: say that a function \( f : [a, b] \to \mathbb{R} \) has the intermediate value property if, given any \( c, d \) with \( a < c < d < b \), and any value \( y \) between \( f(c) \) and \( f(d) \), there is a point \( x \in [c, d] \) with \( f(x) = y \). Corollary \ref{cor:intermediate-value-theorem} shows that any continuous function has the intermediate value property; but there are discontinuous function that do as well (e.g. Example \ref{example:jump-discontinuity} below). To see how this can happen, we now turn to further discuss what kinds of discontinuities functions can have, continuing the discussion from Examples \ref{example:jump-discontinuity} and \ref{example:jump-discontinuity-again}.

**Example 7.16.** A piecewise continuous function defined on an interval in \( \mathbb{R} \) can have a “jump discontinuity” (see Definition \ref{def:jump-discontinuity} below), where the two functions on the two pieces do not “match up” at the transition point. For example, take \( f : [0, 2] \to [0, 1] \) to be the functions defined piecewise by \( f(x) = x \) for \( 0 \leq x < 1 \) and \( f(x) = x - 1 \) for \( 1 \leq x \leq 2 \). Then \( f \) is not continuous as \( x = 1 \): consider the two sequences \( x_n = 1 - \frac{1}{n} \) and \( y_n = 1 + \frac{1}{n} \), both of which converge to \( 1 \). Then \( f(x_n) = x_n \to 1 \) as \( n \to \infty \), while \( f(y_n) = y_n - 1 \to 0 \) as \( n \to \infty \). It therefore cannot be true that all sequences converging to \( 1 \) have images that converge to \( f(1) \), since these two sequences, both converging to \( 1 \), have different image limits. (In this case \( f(y_n) \to 0 = f(1) \); we will see below that \( f \) is right-continuous.

Any function with a jump discontinuity will fail to have the intermediate value property as well. In the above example, consider the interval \([0.9, 1.1]\). The image of \( f \) here is \( f([0.9, 1.1]) = f([0.9, 1]) \cup f([1, 1.1]) = [0.9, 1] \cup [0, 1] \), which does not include any points in the interval \((0.1, 0.9)\). But \( f(0.9) = 0.9 \) and \( f(1.1) = 0.1 \). So in this extreme example, no point between \( f(0.9) \) and \( f(1.1) \) is in the image of \( f \) on \([0.9, 1.1]\).

**Example 7.17.** If we don’t restrict ourselves to a one dimensional domain, there is plenty of room for weirder discontinuities. For example, consider the function \( f : \mathbb{R}^2 \to \mathbb{R} \) defined by

\[
    f(x, y) = \begin{cases} 
        \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\
        0 & \text{if } (x, y) = (0, 0)
    \end{cases}
\]

This is a rational function, whose denominator only vanishes at \((0, 0)\); therefore, by the limit theorems, \( f \) is continuous on \( \mathbb{R}^2 \setminus \{(0, 0)\} \). But the function is not continuous at \((0, 0)\). Consider the two sequences \( a_n = (1/n, 1/n) \) and \( b_n = (-1/n, 1/n) \). Both converge to \((0, 0)\), but

\[
f(a_n) = \frac{\frac{1}{n} \cdot \frac{1}{n}}{(\frac{1}{n})^2 + (\frac{1}{n})^2} = \frac{1}{2}, \quad f(b_n) = \frac{-\frac{1}{n} \cdot \frac{1}{n}}{(-\frac{1}{n})^2 + (\frac{1}{n})^2} = -\frac{1}{2}.
\]

Hence, there is no limit, since these two sequences, both converging to \((0, 0)\), have different image limits. In fact, notice that for any \( m \in \mathbb{R} \), \( f \) is constant along the line \( y = mx \): \( f(x, mx) = \frac{x \cdot mx}{x^2 + (mx)^2} = \frac{m}{1 + m^2} \). The range of the function \( m \mapsto \frac{m}{1 + m^2} \) is \([\frac{-1}{2}, \frac{1}{2}]\) (the endpoints are achieved at \( m = \pm 1 \) as computed above; the function is continuous and so all values between \( \pm \frac{1}{2} \) are achieved by the Intermediate Value Theorem; and no other values are achieved since \( 0 \leq (1 - m)^2 = 1 + m^2 - 2m \), so \( 2m \leq 1 + m^2 \)). The graph of the function can be thought of as a “spiral / helical slide” wrapping around a vertical pole at the origin. While it is difficult to even state what an “intermediate value property” should mean in the context of a two-dimensional domain, it is clear that this function does not have a “jump” discontinuity: there are no gaps in the range of its values near \( 0 \). Below is a rendering of the graph of \( f \) on \([-1, 1]^2 \); you should use software like Maple or Mathematica to explore this function further.
Example 7.18. We can find examples of non-jump discontinuities for functions of a single real variable, too. The classic example is \( f(x) = \sin\left(\frac{1}{x}\right) \) for \( x \neq 0 \) and \( f(0) = 0 \) (or any value in \([-1, 1]\)). A rendering of the graph of \( f \) on \([-\frac{2}{\pi}, \frac{2}{\pi}]\) can be found on the following page.

We have not formally introduced the \( \sin \) function yet; if you want, you can replace it with \( f(x) = \varsigma\left(\frac{1}{x}\right) \) for any periodic continuous function \( \varsigma : \mathbb{R} \to [-1, 1] \) with \( \varsigma(0) = 0 \) to see a very similar picture (e.g. \( \varsigma(x) = x \) for \(-1 \leq x \leq 1\) and \( \varsigma(x) = 2 - x \) for \(1 \leq x \leq 3\), and then on any interval of the form \([4n - 1, 4n + 3]\) for \( n \in \mathbb{Z} \) define \( f(x) = \varsigma(x - 4n) \)). Since \( x \mapsto \frac{1}{x} \) is continuous on \( \mathbb{R} \setminus \{0\} \), and \( \sin \) is continuous on \( \mathbb{R} \), \( f \) is continuous on \( \mathbb{R} \setminus \{0\} \) by Proposition 7.1. However, \( f \) is not continuous at 0. For example, let \( x_n = \frac{1}{2n\pi + \pi/2} \) and \( y_n = \frac{1}{2n\pi + 3\pi/2} \). Then \( x_n \to 0 \) and \( y_n \to 0 \), but \( f(x_n) = \sin(2n\pi + \pi/2) = 1 \) and \( f(y_n) = \sin(2n\pi + 3\pi/2) = -1 \), so \( f \) does not have a limit at 0. (Both of these sequences approach 0 from within \((0, \infty)\), so \( f \) does not have a "right limit" either.)

However, \( f \) does have the intermediate value property on any interval in \( \mathbb{R} \). Since \( f \) is continuous on \( \mathbb{R} \setminus \{0\} \), this is only interesting for intervals that include 0 in their interior, so let \( a < 0 < b \). Consider the point \( c = \frac{1}{1/b + 2\pi} \). Then \( c < b \), and \( f(c) = \sin(1/b + 2\pi) = \sin(1/b) = f(b) \). What’s more, on the interval \([1/b, 1/b + 2\pi]\), the function \( \sin \) achieves its full range of values, \([-1, 1]\), and so \( f([c, b]) = [-1, 1] \). This shows that \( f([a, b]) \supset f([c, b]) = [-1, 1] \) as well. Since the range of \( f \) is \([-1, 1]\), any value \( y \) in between \( f(a) \) and \( f(b) \) is in \([-1, 1]\) and therefore is also in \( f([a, b]) \). Thus, \( f \) is a discontinuous function with the intermediate value property.

Let’s formalize the insights from the above examples into a characterization of different kinds of discontinuities. First we formally define left and right limits.
DEFINITION 7.19. Let $a < b$ be in $\mathbb{R}$, and let $f : [a, b] \to \mathbb{R}$ be a function. Let $x_0 \in (a, b)$. Say that

$$\lim_{x \to x_0^-} f(x) = L$$

if, for any sequence $(x_n)$ in $[a, x_0)$ with $x_n \to x_0$, $f(x_n) \to L$. For shorthand, we write $f(x_0-) = L$. Similarly, we say

$$\lim_{x \to x_0^+} f(x) = R$$

if, for any sequence $(y_n)$ in $(x_0, b]$ with $y_n \to x_0$, $f(y_n) \to R$. For shorthand, we write $f(x_0+) = R$.

Equivalently: $f(x_0-) = L$ means that the function $f|_{[a,x_0]}$ has limit $L$ as $x \to x_0$, and $f(x_0+) = R$ means that the function $f|_{[x_0,b]}$ has limit $R$ as $x \to x_0$. You should verify that $\lim_{x \to x_0} f(x) = L$ is equivalent to $f(x_0-) = f(x_0+) = L$.

DEFINITION 7.20. Let $a < b$ in $\mathbb{R}$, and let $x_0 \in (a, b)$. A function $f : [a, b] \to \mathbb{R}$ is said to have a jump discontinuity at $x_0$ if $f(x_0-) \text{ and } f(x_0+)$ exists, but $f(x_0-) \neq f(x_0+)$. In this case, we say that $f$ has a jump of size $f(x_0+) - f(x_0-)$. If $f$ is discontinuous at $x_0$, but does not have a jump discontinuity at $x_0$, we say $f$ has a non-jump discontinuity at $x_0$.

REMARK 7.21. Rudin uses the term “simple discontinuity” for jumps, and “discontinuity of the second kind” for non-jump discontinuities. This terminology is silly, for two reasons. First,
no working analyst has used that terminology in at least 60 years. Second, and more importantly, the distinction between the two kinds of discontinuities is very much like dividing the world into bananas and non-bananas; jump discontinuities are rare. Thus “jumps” vs. “non-jumps” is a more accurate division.

**Remark 7.22.** It is almost correct to characterize non-jump discontinuities as those where either the left or the right limit does not exist. The one kind of exception to this is a kind of degenerate discontinuity, where \( \lim_{t \to x_0} f(t) \) exists, but is not equal to \( f(x_0) \). (For example: \( f(t) = 0 \) for all \( t \neq x_0 \), but \( f(x_0) = 1 \).) Such a discontinuity does not count as a jump discontinuity by the above definition, and indeed it should not: the function is not jumping anywhere. (This is an important distinction, for example in the theory of paths of stochastic processes.) But it is also quite a bit milder than the discontinuity of Example 7.18. Such examples are called removable discontinuities. So non-jump discontinuities are partitioned into removable ones, and the truly bad ones, which we might call oscillatory discontinuities.

In example 7.16, the function has a jump discontinuity of size \(-1\) at 0. In example 7.18, the function has a non-jump discontinuity at 0. In Example 6.11, Dirichlet’s function, the indicator function of the rationals, every point is a non-jump discontinuity.

There is one important class of functions (which we will use extensively later in this course, when we study integration theory) in which only jump discontinuities can occur.

**Definition 7.23.** Let \( a < b \) be in \( \mathbb{R} \). A function \( f : [a, b] \to \mathbb{R} \) is said to be monotone increasing or nondecreasing if, for all \( x < y \) in \([a, b] \), \( f(x) \leq f(y) \); it is said to be monotone decreasing or nonincreasing if, for all \( x < y \) in \([a, b] \), \( f(x) \geq f(y) \). If \( f \) is either monotone increasing or monotone decreasing, \( f \) is called monotone.

The functions in Examples 6.11, 6.12 and 7.18 are not monotone; all of them oscillate wildly. In all cases, the discontinuities were not jump discontinuities. Example 7.16 is also not monotone, but it is milder (just one oscillation), and its discontinuity is a jump discontinuity. In general, monotone functions can be discontinuous (e.g. like Example 7.16, but with a positive jump: \( f(x) = x \) for \( 0 \leq x < 1 \) and \( f(x) = x + 1 \) for \( 1 \leq x \leq 2 \)), but that’s as bad as they get.

**Proposition 7.24.** Let \( a < b \) in \( \mathbb{R} \), and let \( f : (a, b) \to \mathbb{R} \) be a monotone function. Then \( f(x^-) \) and \( f(x^+) \) exists for each \( x \in (a, b) \); thus, all discontinuities of \( f \) are jump discontinuities. If \( f \) is monotone increasing, then for all \( x \in (a, b) \),

\[
\sup_{a < t < x} f(t) = f(x^-) \leq f(x) \leq f(x^+) = \inf_{x < t < b} f(t). \tag{7.1}
\]

Also \( f(x^+) \leq f(y^-) \) whenever \( a < x < y < b \). For monotone decreasing \( f \), the inequalities (and the \( \sup \) and \( \inf \)) are reversed.

**Proof.** We assume \( f \) is monotone increasing; the proof for the monotone decreasing case is analogous. But assumption, for any \( t < x \), \( f(t) \leq f(x) \). So the set \( \{ f(t) : a < t < x \} \) is bounded above by \( f(x) \), which shows that \( \alpha = \sup_{a < t < x} f(t) \leq f(x) \). Fix \( \epsilon > 0 \). Since \( \alpha \) is the least upper bound, there must exist some \( x' \in (a, x) \) with \( \alpha - \epsilon < f(x') \leq \alpha \). Since \( f \) is monotone increasing, if \( x' < t < x \), \( f(x') \leq f(t) \leq \alpha \). Combining these last two inequalities yields

\[ |f(t) - \alpha| < \epsilon \], \quad \text{for} \quad x' < t < x. \]

Hence, if \( t_n \in (a, x) \) and \( t_n \to x \), there is some \( N \) so that \( x' < t_n < x \) for \( n \geq N \), and so \( |f(t_n) - \alpha| < \epsilon \) for \( n \geq N \); this shows that \( f(x^-) = \lim_{t \to x^-} f(t) = \alpha \), as claimed. The proof that \( f(x) \leq f(x^+) = \inf_{x < t < b} f(t) \) is very similar; thus we have proven (7.1) true.
Now, if $a < x < y < b$, then since $f(s) \geq f(y)$ for all $s > y$, we have
\[ f(x+) = \inf_{x < t < b} f(t) = \inf_{x < t < y} f(t). \]
Applying the left-hand-side of (7.1) to the interval $(x, y)$ instead of $(a, b)$ shows that
\[ f(x+) = \inf_{x < t < y} f(t) \leq \sup_{x < t < y} f(t) = f(y-) \]
which concludes the proof of the second inequality. The proofs for monotone decreasing functions are analogous. □
CHAPTER 8

Differentiation of Functions of a Real Variable

1. Lecture 4: April 7, 2016

Almost everything that was discussed in Chapters 5, 6, and 7 applied quite generally to functions / sequences in metric spaces. In this chapter, we will focus on concepts that are exclusive to functions defined on (intervals in) \( \mathbb{R} \). Some of these concepts can be extended to more general spaces, but not the level of generality we’ve seen so far: taking derivatives involves more underlying structure than most metric spaces afford.

Let \( a < b \) in \( \mathbb{R} \), and suppose \( f: [a, b] \to \mathbb{R} \) is a function. For any \( x, y \in [a, b] \) with \( x \neq y \), we may define the difference quotient \( (DQf)(x, y) \) as follows:
\[
(DQf)(x, y) \equiv \frac{f(x) - f(y)}{x - y}.
\]

**Example 8.1.** If \( f(x) = x \), then \( (DQf)(x, y) = \frac{x-y}{x-y} = 1 \) for all \( x \neq y \). If \( f(x) = x^2 \), we can factor to find that
\[
(DQf)(x, y) = \frac{x^2 - y^2}{x - y} = \frac{(x - y)(x + y)}{x - y} = x + y, \quad x \neq y.
\]

In general, for power functions \( f(x) = x^k \), since \( x^k - y^k = (x - y)(x^{k-1} + x^{k-2}y + x^{k-3}y^2 + \cdots + xy^{k-2} + y^{k-1}) \), we see that \( (DQf)(x, y) \) is a polynomial (of homogeneous degree \( k - 1 \)) in \( x \) and \( y \). It is, a priori, only defined when \( x \neq y \).

We will use the difference quotient \( DQf \) to define the derivative of \( f \), by taking limits, as you no doubt recall. Before we do so, it is instructive to note a few nice properties that the difference quotient has. (Note that *every* function has a difference quotient.)

**Lemma 8.2.** Let \( a < b \) in \( \mathbb{R} \), and let \( f, g: [a, b] \to \mathbb{R} \) be functions. For any \( x \neq y \) in \( [a, b] \), we have the following.

1. \( DQ(f + g)(x, y) = DQf(x, y) + DQg(x, y) \).
2. \( DQ(fg)(x, y) = f(x)DQg(x, y) + DQf(x, y)g(y) \).
3. If \( g(x) \neq 0 \) for \( x \in [a, b] \), then \( DQg(x, y) = -\frac{1}{g(x)g(y)}DQf(x, y) \).

**Proof.** (1) is an immediate consequence of the distributive properties of addition and multiplication, and is left to the reader. For (2), we calculate as follows:
\[
f(x)g(x) - f(y)g(y) = f(x)g(x) - f(x)g(y) + f(x)g(y) - f(y)g(y)
= f(x)(g(x) - g(y)) + (f(x) - f(y))g(x).
\]
Dividing through by \( x - y \) yields the result. For (3), we compute
\[
DQg(x, y) = \frac{g(y) - g(x)}{x - y} = \frac{g(y) - g(x)}{g(x)g(y)(x - y)} = -\frac{1}{g(x)g(y)}DQf(x, y).
\]
\( \square \)
Remark 8.3. Since \( fg = gf \), (2) implies that we also have \( DQ(fg)(x,y) = f(y)DQg(x,y) + g(x)DQf(x,y) \). One can see this as well in the proof of (2), by introducing the opposite cross terms \(-f(y)g(x) + f(y)g(x)\).

You should recognize the statements of Lemma 8.2 as versions the usual rules of differentiation; e.g. (2) is the product rule. It is illuminating to note that these are simply algebraic rules that hold for difference quotients, having nothing to do with the limits involved in taking derivatives. In that vain, we can also prove a difference quotient version of the chain rule; this is particularly instructive, since it highlights a technical difficulty we will have to overcome in proving the chain rule for derivatives.

Lemma 8.4. Let \( a < b \) in \( \mathbb{R} \), and let \( g: [a, b] \to \mathbb{R} \) be a function. Suppose \( f \) is a real valued function defined on the range of \( g \); then \( \lim_{x \to y} f(y) \) is a function. If \( x \neq y \) in \( [a, b] \), and if \( g(x) \neq g(y) \), then
\[
DQ(f \circ g)(x, y) = DQ f(g(x), g(y)) \cdot DQ g(x, y).
\]

Proof. Since \( g(x) \neq g(y) \), we can multiply and divide through by it, to see that
\[
DQ(f \circ g)(x, y) = \frac{f \circ g(x) - f \circ g(y)}{x - y} = \frac{f(g(x)) - f(g(y))}{g(x) - g(y)} \cdot \frac{g(x) - g(y)}{x - y}
\]
and, by the definition of \( DQ \), this gives the desired result. \( \square \)

It is possible that \( g(x) = g(y) \), even if \( x \neq y \); in this case, \( DQ f(g(x), g(y)) \) is not defined, since \( DQ f(u, v) \) is only defined when \( u \neq v \). Of course, we would like to define it even on the diagonal \( u = v \) by taking the limit as \( v \to u \); we will presently do this to define the derivative. But the reader should be wary of the above proof: it is possible that \( g(x) = g(y) \) for \( x \) and \( y \) arbitrarily close to each other, in which case it will take some care to make sense of the limit; we will do this below in Proposition 8.12.

We now come to the definition of the derivative.

Definition 8.5. Let \( a < b \) in \( \mathbb{R} \), and let \( f: [a, b] \to \mathbb{R} \). For any point \( x \in [a, b] \), say that \( f \) is differentiable at \( x \) if the limit \( \lim_{t \to x} DQ f(x, t) \) exists; in this case, we call this limit the derivative of \( f \) at \( x \), and denote it
\[
f'(x) = \frac{df}{dx} = \lim_{t \to x} \frac{f(x) - f(t)}{x - t}.
\]

Remark 8.6. Some authors insist that \( x \in (a, b) \) for this definition. As stated above, we highlight the fact that differentiability (like continuity) depends on the domain of the function. At \( x = a \), the above limit is actually a right limit \( \lim_{t \to x} DQ f(x, t) \), and we might call \( f'(a) \) (if it exists) the right derivative of \( f \) at \( a \). If \( f \) is actually defined on a larger interval including \( a \) in its interior, then it is possible that \( f \) is not differentiable at \( a \), even if it is right differentiable there. Similar comments apply to left differentiability at \( b \).

Differentiability is a regularity property of a function at a point. It is stronger than continuity.

Proposition 8.7. Let \( a < b \) in \( \mathbb{R} \), and let \( f: [a, b] \to \mathbb{R} \). If \( x \in [a, b] \) and \( f \) is differentiable at \( x \), then \( f \) is continuous at \( x \).

Proof. Let \( (x_n) \) be a sequence in \( [a, b] \setminus \{x\} \) with \( x_n \to x \). Then we can multiply and divide by \( x_n - x \), and so
\[
f(x) - f(x_n) = \frac{f(x) - f(x_n)}{x - x_n} (x - x_n) = DQ f(x, x_n)(x - x_n).
\]
By assumption, \( f'(x) \) exists, which means that \( \lim_{n \to \infty} DQf(x, x_n) = f'(x) \). We also have \( \lim_{n \to \infty} (x - x_n) = 0 \). Thus, by the limit theorems,

\[
\lim_{n \to \infty} [f(x_n) - f(x)] = \lim_{n \to \infty} DQf(x, x_n) \cdot \lim_{n \to \infty} (x - x_n) = f'(x) \cdot 0 = 0.
\]

This shows \( \lim_{n \to \infty} f(x_n) = f(x) \), for any such sequence. By definition, this means \( \lim_{t \to x} f(t) = f(x) \), which is the definition of continuity of \( f \) at \( x \).

**Remark 8.8.** In the case \( x = a \), differentiability (which really means right differentiability) only implies right continuity; similarly differentiability at \( x = b \) (which really means left differentiability) only implies left continuity.

The converse of Proposition 8.7 is very false.

**Example 8.9.** Consider the function \( f(x) = |x| \) defined on \( \mathbb{R} \). This function is continuous on \( \mathbb{R} \), and in particular at 0: if \( x_n \to 0 \) then \( |x_n| \to 0 \). Now, for \( x \neq 0 \), \( DQf(0, t) = \frac{0 - |t|}{0 - t} = \frac{|t|}{t} \). This is either +1 if \( t > 0 \) or −1 if \( t < 0 \). So take, for example, the sequence \( t_n = \frac{(-1)^n}{n} \), which tends to 0. The sequence \( DQf(0, t_n) = (-1)^n \) does not converge as \( n \to \infty \). This means that the limit \( f'(0) \) does not exist, so \( f \) is not differentiable at 0.

**Remark 8.10.** In Example 8.9, the function \( f \) fails to be differentiable only at a single point. It was long thought (by Newton and others prior to the 19th Century) that non-differentiable points of continuous functions were all of this nature: that continuous functions would have to be differentiable except at a discrete set of points. This is extremely far from true. Later on, we will see examples of functions that are continuous everywhere on \( \mathbb{R} \), but differentiable nowhere. In a sense, such functions are the most important, to the modern theory of analysis and probability.

Let us now state the well-known differentiation rules that follow from Lemma 8.2.

**Proposition 8.11.** Let \( a < b \) in \( \mathbb{R} \), and let \( f, g : [a, b] \to \mathbb{R} \). Suppose \( f \) and \( g \) are differentiable at a point \( x \in [a, b] \). The following hold true.

1. \( f + g \) is differentiable at \( x \), and \( (f + g)'(x) = f'(x) + g'(x) \).
2. \( fg \) is differentiable at \( x \), and \( (fg)'(x) = f'(x)g(x) + f(x)g'(x) \).
3. If \( g(x) \neq 0 \), then \( \frac{1}{g} \) is differentiable at \( x \), and \( \left( \frac{1}{g} \right)'(x) = -\frac{g'(x)}{g(x)^2} \).

**Proof.** (1) follow immediately from Lemma 8.2(1), using the limit theorem (a limit of a sum is the sum of the limits). For (2), using Lemma 8.2(2), we have

\[
(fg)'(x) = \lim_{t \to x} DQ(fg)(x, t) = \lim_{t \to x} [f(x)DQg(x, t) + DQf(x, t)g(t)]
\]

\[
= f(x) \lim_{t \to x} DQg(x, t) + \lim_{t \to x} DQf(x, t) \cdot \lim_{t \to x} g(t)
\]

The limits of the difference quotients are, by definition, \( f'(x) \) and \( g'(x) \). By Proposition 8.7, \( g \) is continuous at \( x \), and therefore \( \lim_{t \to x} g(t) = g(x) \), yielding the desired formula. For (3), we note (again by Proposition 8.7) that \( g \) is continuous at \( x \). Since \( g(x) \neq 0 \), it follows that, for some \( \delta > 0 \), \( g(t) \neq 0 \) for all \( t \in (x - \delta, x + \delta) \cap [a, b] \); indeed, from the \( \epsilon - \delta \) definition of continuity at \( x \), this follows by taking any \( \epsilon < |g(x)| \). Restricting to this neighborhood of \( x \), the function \( \frac{1}{g} \) is well defined, and we may apply Lemma 8.2(3) and the limit theorems to find that

\[
\left( \frac{1}{g} \right)'(x) = \lim_{t \to 0} DQg(x, t) = \lim_{t \to 0} \left( -\frac{1}{g(x)g(t)} DQg(x, t) \right) = -\frac{1}{g(x)} \cdot \lim_{t \to 0} \frac{1}{g(t)} \cdot \lim_{t \to x} DQg(x, t).
\]
Since $g$ is continuous and nonvanishing on a neighborhood of $x$, $\frac{1}{g(x)}$ is continuous at $x$, and the first limit is $\frac{1}{g(x)}$; the second limit is $g'(x)$ by definition. This concludes the proof.

Part (3) above is a special case of the so-called *quotient rule*, which states that

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

under the conditions stated above. This follows immediately by combining (2) and (3), noting that $\frac{\delta f}{\delta g} = f \cdot \frac{1}{g}$; the details are left to the reader.

We now come to the chain rule.

**Proposition 8.12.** Let $a < b$ in $\mathbb{R}$, and let $g: [a, b] \to \mathbb{R}$. Suppose $x \in [a, b]$ and $g$ is differentiable at $x$. Let $f$ be defined on the range of $g$, and in particular on a neighborhood of $g(x)$, and suppose that $f$ is differentiable at $g(x)$. Then $f \circ g$ is differentiable at $x$, and $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$.

As mentioned above, the proof is technically more challenging than might at first seem necessary. The straightforward approach would be to take the statement of Lemma 8.4 and apply the limit theorems to derive the stated formula:

$$\lim_{t \to x} DQ(f \circ g)(x, t) = \lim_{t \to x} DQ f(g(x), g(t)) \cdot \lim_{t \to x} Dg(x, t) = \lim_{t \to x} DQ f(g(x), g(t)) \cdot g'(x)$$

For the first limit, we would like to say that since $g$ is continuous at $x$, $g(t) \to g(x)$ as $t \to x$, and so this limit is the same as $\lim_{s \to g(x)} Df(g(x), s) = f'(g(x))$. The problem with this is that the quantity $Df(g(x), g(t))$ is only well-defined if $g(x) \neq g(t)$, and it is perfectly possible for $g$ to have the property that $g(t) = g(x)$ for many $t$, arbitrarily close to $x$ (e.g. $g$ could be constant near $x$; or, technically worse, $g$ could oscillate fast near $x$). Therefore, $\lim_{t \to x} DQ f(g(x), g(t))$ may not be defined since, for $\lim_{t \to x} u(t)$ to make sense, $u$ must be defined on $(x - \delta, x + \delta) \setminus \{x\}$ (intersected with $[a, b]$). Therefore, we must be much more careful in proving the chain rule.

**Proof.** By definition, $\lim_{t \to x} DQ g(x, t) = g'(x)$; so if we set $u_x(t) = DQ g(x, t) - g'(x)$, we have $u_x(t) \to 0$ as $t \to x$. Similarly, setting $y = g(x)$, we have $\lim_{s \to y} DQ f(y, s) = f'(y)$; so if we set $v_y(s) = DQ f(y, s) - f'(s)$, we have $v_y(s) \to 0$ as $s \to y$. Now, unraveling the definition of the difference quotient, we have

$$g(x) - g(t) = (x - t) DQ g(x, t) = (x - t) [g'(x) + u_x(t)]$$

$$f(y) - f(s) = (y - s) DQ f(y, s) = (y - s) [f'(y) + u_y(s)].$$

Here $t$ is in a neighborhood of $x$ (where $g$ is defined) and $s$ is in a neighborhood of $y$ (where $f$ is defined). Composing, we have $f \circ g(x) - f \circ g(t) = f(y) - f(g(t))$. Since $g$ is continuous, when $t$ is in a small enough neighborhood of $x$, $s = g(t)$ is in the given neighborhood of $y = g(x)$, and so

$$f \circ g(x) - f \circ g(t) = f(y) - f(s) = (y - s) [f'(y) + u_y(s)]$$

$$= (g(x) - g(t)) [f'(y) - u_y(s)]$$

$$= (x - t) [g'(x) + u_x(t)] [f'(y) + u_y(g(t))].$$

For all $t \neq x$, we can then divide through by $x - t$ and we see that

$$DQ(f \circ g)(x, t) = [g'(x) + u_x(t)] [f'(y) + u_y(s)] = [g'(x) + u_x(t)] [f'(y) + u_y(g(t))].$$

As $t \to x$, $u_x(t) \to 0$. Also, $g$ is continuous at $x$, so as $t \to x$, $g(t) \to g(x) = y$, and so $v_y(g(t)) \to 0$. It now follows from the limit theorems that $\lim_{t \to x} DQ(f \circ g)(x, t) = [g'(x) + 0][f'(y) + 0] = g'(x) f'(g(x))$, as desired.
As you'll recall from calculus, one of the main applications of derivatives is in the study of extrema.

**Definition 8.13.** Let \( X, Y \) be metric spaces, and let \( f : X \to Y \). A point \( x \in X \) is called a local maximizer of \( f \) if there is a positive radius \( \delta > 0 \) so that, for all \( t \in B_\delta(x) \), \( f(t) \leq f(x) \); in this case the value \( f(x) \) is called a local maximum. Similarly, \( x \) is a local minimizer if there is a \( \delta > 0 \) so that, for all \( t \in B_\delta(x) \), \( f(t) \geq f(x) \); in this case the value \( f(x) \) is called local minimum. The local maxima and minima of \( f \) are called its local extrema, and the points at which they occur are local extremizers.

For a \( f : \mathbb{R} \to \mathbb{R} \), local extrema of \( f \) can be determined by locating the points \( x \) where \( f'(x) = 0 \).

**Theorem 8.14.** Let \( a < b \) in \( \mathbb{R} \), and suppose \( f : (a, b) \to \mathbb{R} \) be a function. If \( f \) has a local extremum at \( x \in (a, b) \), and if \( f'(x) \) exists, then \( f'(x) = 0 \).

**Proof.** We will assume \( x \) is a local maximizer; the argument for a local minimizer is analogous. By assumption, there is a \( \delta > 0 \) such that \( a < x - \delta < x + \delta < b \) and \( f(x) \geq f(t) \) for all \( t \in (x - \delta, x + \delta) \). In particular, this means that if \( x - \delta < t < x \), then \( x - t > 0 \) while \( f(x) - f(t) \geq 0 \), and so \( DQ f(x, t) \geq 0 \) in this interval; likewise, if \( x < t < x + \delta \), then \( x - t < 0 \) while \( f(x) - f(t) \geq 0 \), so here \( DQ f(x, t) \leq 0 \). Thus, using the squeeze theorem, we have

\[
\lim_{t \to x^-} DQ f(x, t) \leq 0 \quad \text{and} \quad \lim_{t \to x^+} DQ f(x, t) \geq 0.
\]

By assumption \( f'(x) = \lim_{t \to x} DQ f(x, t) = \lim_{t \to x^-} DQ f(x, t) = \lim_{t \to x^+} DQ f(x, t) \); thus, this common value is both \( \geq 0 \) and \( \leq 0 \). It follows that \( f'(x) = 0 \), as claimed.

As you’ll recall, points \( x \) where \( f'(x) = 0 \) are called critical points. The content of Theorem 8.14 is that, if a function is known to be differentiable at all points in its domain, then its local extrema all occur at critical points. This is one of the core tools in calculus, and we will use it in many theoretical applications as well. In order for it to be useful, of course, we must know that our function is differentiable not just at certain points but everywhere: for example, the function \( f(x) = |x| \) attains is global minimum (which is its only local extemum) at the point \( x = 0 \), where \( f \) is not differentiable. We will thus be most interested, for now, in functions that are differentiable everywhere.

Like continuity, differentiability is a local property: we speak of differentiability of a function at a point \( x \) in its domain (and this only depends on the behavior of \( f \) in an arbitrarily small neighborhood of \( x \)). Also like continuity, we can then boost this to a global property, by talking about \( f \) being differentiable on a set (i.e. differentiable at all points of a given set). In the case of derivatives, this produces a new function.

**Definition 8.15.** Let \( a < b \) in \( \mathbb{R} \), and let \( f : [a, b] \to \mathbb{R} \) be differentiable (at all points of its domain). We let \( f' \) denote the new function \( f' : [a, b] \to \mathbb{R} \) whose value at \( x \) is \( f'(x) = \lim_{t \to x} DQ f(x, t) \).

Here are two examples that show how \( f' \) can be poorly behaved (at points) even when \( f \) is quite well behaved. Both examples take for granted the functions \( \sin \) and \( \cos \), which are differentiable on \( \mathbb{R} \) and satisfy \( \sin' = \cos \) and \( \cos' = -\sin \). We will formally develop these later this quarter.
8. DIFFERENTIATION OF FUNCTIONS OF A REAL VARIABLE

EXAMPLE 8.16. Define a function $f : \mathbb{R} \to \mathbb{R}$ as follows:

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

First of all, note that $f$ is continuous at all $x \neq 0$ since it is a composition of continuous functions there; and at $x = 0$, note that $| \sin \frac{1}{x} | \leq 1$ for all $x$, so $|f(x)| \leq |x| \to 0 = f(0)$ as $x \to 0$. So $f$ is continuous on $\mathbb{R}$. What’s more, using the rules of differentiation, we can compute that $f$ is differentiable at all points $x \neq 0$, and for such points

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

However, at 0, we have

$$DQf(0, t) = \frac{0 - f(t)}{0 - t} = \sin \frac{1}{t}$$

and this function has no limit as $t \to 0$ (cf. Example 7.18). Thus, $f'(0)$ does not exist. Also, the formula for $f'(x)$ for $x \neq 0$ involves terms like $\sin \frac{1}{x}$ and $\frac{1}{x} \cos \frac{1}{x}$ which are at least as badly behaved as $\sin \frac{1}{x}$. So there is no way to extend the function $f'$ to be defined at 0 in such a way that it will be continuous: $f'$ has a non-jump discontinuity at 0.

![Figure 1](image-url). The graph of $x \sin(\frac{1}{x})$; the envelope is given by the lines $y = \pm x$.

EXAMPLE 8.17. Now consider the following function:

$$g(x) = \begin{cases} 
  x^2 \sin \frac{1}{x} & x \neq 0 \\
  0 & x = 0 
\end{cases}$$

An argument very similar to the one in Example 8.16 shows that $g$ is continuous on $\mathbb{R}$, and differentiable everywhere except possibly at 0; here

$$g'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}.$$
Now, looking at \( x = 0 \) specifically, we have
\[
DQg(0, t) = \frac{0 - f(t)}{0 - t} = \frac{t^2 \sin \frac{1}{t}}{t} = t \sin \frac{1}{t} = f(t) \to 0 \text{ as } t \to 0
\]
where \( f \) is the function from Example 8.16, which is continuous at 0 as shown above. Thus, \( g \) is actually differentiable on all of \( \mathbb{R} \), with \( f'(0) = 0 \). But the formula for \( g' \) at points other than 0 shows that \( g \) is not continuous on \( \mathbb{R} \): it has a non-jump discontinuity at 0. Ergo: it is possible for a function to be continuous and differentiable everywhere, but for its derivative to have a (bad) discontinuity somewhere.

\[\text{Figure 2. The graph of } x^2 \sin \left(\frac{1}{x}\right); \text{ the envelope is given by the curves } y = \pm x^2.\]

If we want to use tools like Theorem 8.14, functions like the one in Example 8.16 are out; functions like the one in Example 8.17 on the other hand, are eligible, despite the bad behavior of the derivative as a function.

We now come to the most important theorem on all of calculus: the Mean Value Theorem.

**Theorem 8.18 (Mean Value Theorem).** Let \( a < b \) in \( \mathbb{R} \), and let \( f : [a, b] \to \mathbb{R} \) be a continuous function. Suppose also that \( f \) is differentiable on \((a, b)\). Then there is a point \( x \in (a, b) \) such that \( f'(x) = DQf(a, b) \); i.e.
\[
f'(x) = \frac{f(b) - f(a)}{b - a}.
\]
Notice that we do not even need to assume that \( f \) is (one-sided) differentiable at the endpoints; it need only be continuous there.

**Proof.** Define a new function \( h(t) = (b - a)f(t) - t[f(b) - f(a)] \). First, notice that
\[
h(a) = (b - a)f(a) - a[f(b) - f(a)] = bf(a) - af(b),
\]
\[
h(b) = (b - a)f(b) - b[f(b) - f(a)] = bf(a) - af(b).
\]
So \( h(a) = h(b) \).

Now, by the differentiation rules, \( h \) is differentiable on \((a, b)\) and \( h'(t) = (b - a)f'(t) - [f(b) - f(a)] \), and so our goal is to show that \( h'(x) = 0 \) for some \( x \in (a, b) \). First, consider the case that \( h \) is constant: then \( h'(x) = 0 \) for all \( x \in (a, b) \), and we’re done. Otherwise, there is some point \( t \in (a, b) \) such that \( h(t) \neq h(a) = h(b) \). Thus, either \( h(t) > h(a) \) or \( h(t) < h(a) \). For the moment, we assume the former: there is some point \( t \in (a, b) \) where \( h(t) > h(a) \).

Since \( f \) is continuous on \([a, b]\), so is \( h \) by the limit theorems. Since \([a, b]\) is compact, by the Minimax theorem, there is a point \( x \in [a, b] \) where \( h(x) = \max\{f(t): t \in [a, b]\} \). As \( h(a) = h(b) \) and this value is not the maximum (by assumption that \( h(t) > h(a) \) for some \( t \in (a, b) \)), it follows that \( x \in (a, b) \). As \( h \) is differentiable in \((a, b)\), by Theorem 8.14 it follows that \( h'(x) = 0 \), and this shows the desired conclusion in this case.

For the case that we only know that \( h(t) < h(a) \) for some \( t \in (a, b) \), the argument is analogous, using the minimum, rather than the maximum, of \( h \). \( \Box \)

To demonstrate why we called the Mean Value Theorem the most important theorem in calculus, note the following corollary which encapsulates most of the material (on curve sketching, etc.) which follows immediately from it.

**Corollary 8.19.** Let \( a < b \) in \( \mathbb{R} \), and let \( f: [a, b] \to \mathbb{R} \) be continuous and differentiable on 
\((a, b)\).

1. If \( f' \geq 0 \) then \( f \) is monotone increasing.
2. If \( f' \leq 0 \) then \( f \) is monotone decreasing.

**Proof.** All three parts of the proof follow from the fact that, for any \( x_1, x_2 \in [a, b] \), by the Mean Value Theorem there is a point \( x \in (x_1, x_2) \) such that
\[
f(x_2) - f(x_1) = (x_2 - x_1)f'(x).
\]
For example, in case (2) where \( f'(x) = 0 \) for all \( x \), this shows \( f(x_2) - f(x_1) = 0 \) for all \( x_1, x_2 \in [a, b] \), which is precisely to say that \( f \) is constant. The other two cases are similar. \( \Box \)

In addition to practical applications like curve sketching, the mean value theorem allows us to see that, while derivative functions \( f' \) can be quite irregular (e.g. Example 8.17) there are constraints. While derivative functions need not be continuous, they always have the intermediate value property.

**Proposition 8.20 (Darboux).** Let \( a < b \) in \( \mathbb{R} \), and suppose \( f: (a, b) \to \mathbb{R} \) is differentiable. Then \( f \) has the intermediate value property on \((a, b)\): for any \( x_1 < x_2 \) in \((a, b)\), if \( y \) be a real number between \( f'(x_1) \) and \( f'(x_2) \), then there is a point \( x \in (x_1, x_2) \) such that \( f'(x) = y \).

**Proof.** Without loss of generality, we suppose \( f'(x_1) < y < f'(x_2) \); the reverse ordering case is very similar. Set \( g(t) = f(t) - yt \). Since \( f \) is differentiable on \((a, b)\), it is continuous on \([x_1, x_2]\), and so therefore is \( g \). As \([x_1, x_2]\) is compact, by the minimax theorem, \( g \) attains its minimum value at some point \( x \in [x_1, x_2] \). Now, \( g \) is also differentiable on \((a, b)\), and \( g'(t) = f'(t) - y \). By assumption, \( g'(x_1) = f'(x_1) - y < 0 \) and \( g'(x_2) = f'(x_2) - y > 0 \).

We claim that \( x_1 < x < x_2 \). To prove this, we just need to show that neither \( x_1 \) nor \( x_2 \) is a minimizer for \( g \). Indeed, consider \( x_1 \): we have \( \lim_{t \to x_1} DQg(x_1, t) = g'(x_1) < 0 \); i.e.
\[
\lim_{t \to x_1} \frac{g(t) - g(x_1)}{t - x_1} < 0.
\]
It follows that, for all $t$ sufficiently close to $x_1$, $DQg(x_1, t) < 0$. In particular, for $t > x_1$ and sufficiently small, multiplying through by the positive $t - x_1$ yields $g(t) - g(x_1) < 0$. This shows that $g(x_1)$ is not the minimum of $g$ on $[x_1, x_2]$. An analogous argument at $x_2$, using the fact that $g'(x_2) > 0$, shows that $x_2$ is not a minimizer for $g$. Thus, $x \in (x_1, x_2)$ as claimed.

Thus, by the Mean Value Theorem (or in fact by Theorem 8.14, which is equivalent to it), $g'(x) = 0$. This shows that $0 = g'(x) = f'(x) - y$, and so $f'(x) = y$ as claimed. □

Remark 8.21. This theorem was proved by Darboux in the late 19th Century. Until that point, it was widely believed that only continuous functions could have the intermediate value property. But Darboux proved Proposition 8.20 and then gave examples (like Example 8.17) to show that this is not the case.

Corollary 8.22. If $f$ is differentiable, then $f'$ cannot have any jump discontinuities.

Proof. As mentioned just after Example 7.16, any function with a jump discontinuity must fail to have the intermediate value property (in a neighborhood of the jump). □

As we saw in the previous lecture, a differentiable function \( f \) can have a derivative \( f' \) which is not very regular. The function \( f' \) cannot be arbitrary (as Darboux’s theorem shows, \( f' \) must at least have the intermediate value property), but \( f' \) certainly need not be continuous. In many circumstances, it is natural to assume that \( f' \) is continuous. For example, in many calculus applications, we use the second derivative, meaning we actually assume that the function \( f' \) is differentiable (ergo continuous).

**Definition 8.23.** Let \( a < b \) in \( \mathbb{R} \), and let \( f: [a, b] \to \mathbb{R} \) be differentiable. If \( f' \) is a continuous function on \([a, b]\), we say \( f \) is continuously differentiable, and write \( f \in C^1[a, b] \). More generally, for a positive integer \( k \), we say \( f \in C^k[a, b] \) if \( f^{(j)} \) is continuous for \( 0 \leq j \leq k \), where \( f^{(j)} \) is defined recursively by \( f^{(j)} = (f^{(j-1)})' \), and \( f^{(0)} = f \).

Example [8,17] shows that a function \( f \) can be differentiable (on all of \( \mathbb{R} \), even) without being \( C^1 \). The difference is subtle but important. It is analogous to the difference between continuous and uniformly continuous functions: continuity is a local property boosted to a global property pointwise, while uniform continuity is a truly global property. Similarly, being differentiable is a local property boosted to a global property pointwise, while being \( C^1 \) is a truly global property. (Warning: this is not a perfect analogy. It is, for example, not true that a differentiable function is continuously differentiable on a compact interval.)

Recall our motivation for introducing differentiability. Continuity of \( f \) at a point \( x \) is the statement that \( f(x + t) - f(x) \) tends to 0 as \( h \to 0 \), but at what rate? If \( f \) is differentiable, the answer is that there is a correction factor, the linear function \( h \mapsto f'(x) t \), so that

\[
f(x + t) - f(x) - f'(x) t = o(t)
\]

where the notation \( \alpha(t) = \beta(t) + o(t) \) means that \( \lim_{t \to 0} \frac{\alpha(t) - \beta(t)}{t} = 0 \). Thus, differentiability implies that, up to a linear correction, \( f(x + t) \) is closer to \( f(x) \) than any linear function: the difference goes to 0 faster than \( t \) as \( t \to 0 \). This leads us to ask what happens if there are higher derivatives? Is the difference even smaller, modulo higher order corrections? The answer is yes, which is the statement of Taylor’s theorem.

**Theorem 8.24.** [Taylor’s Theorem] Let \( a < b \) in \( \mathbb{R} \), let \( k \in \mathbb{N} \), and let \( f: (a, b) \to \mathbb{R} \) be \( C^{k-1} \), such that \( f^{(k)} \) exists (but need not be differentiable) on \((a, b)\). Define the Taylor polynomial \( T_{x}^{k-1} f \) by

\[
(T_{x}^{k-1} f)(t) = f(x) + f'(x) t + \frac{1}{2} f''(x) t^2 + \cdots + \frac{1}{(k-1)!} f^{(k-1)}(x) t^{k-1} = \sum_{j=0}^{k-1} \frac{f^{(j)}(x)}{j!} t^j.
\]

For each \( t \) such that \( x + t \in (a, b) \), there exists a point \( \xi \) between \( x \) and \( x + t \) such that

\[
f(x + t) = (T_{x}^{k-1} f)(t) + \frac{1}{k!} f^{(k)}(\xi) t^k.
\]

The theorem is often stated using the variable \( y = x + t \) as \( f(y) = (T_{x}^{k-1} f)(y - x) + \frac{1}{k!} f^{(k)}(\xi)(y - x)^k \) for some point \( \xi \) between \( x \) and \( y \). This is natural in the sense that the polynomial \( y \mapsto (T_{x}^{k-1} f)(y - x) \) can be described as the unique degree \( k - 1 \) polynomial whose derivatives of orders \( \leq k - 1 \) at \( x \) match those of \( f \). Note also that, in the special case \( k = 1 \), the statement is just

\[
f(y) = f(x) + f'(\xi) (y - x)
\]

for some \( \xi \) between \( x \) and \( y \).
which is exactly the statement of the Mean Value Theorem. We will see below that the proof of Taylor’s theorem is just repeated application of Taylor’s theorem.

The last term involving \( f^{(k)}(\xi) \) is called the \( k \)-remainder term, often written as \((R^k f)(x, y)\), so we have

\[
f(y) = \sum_{j=0}^{k-1} \frac{f^{(j)}(x)}{j!} (y - x)^j + (R^k f)(x, y).
\]

This says that \( f \) is well-approximated by the Taylor polynomial \( T_x^{k-1} f \), so long as the remainder term \( R^k f(x, y) \) can be shown to be small. From the form of the remainder term in the theorem, this amounts to having good control on the \( k \)th derivative \( f^{(k)} \) (at arbitrary points, since all we know is that \( \xi \) is between \( x \) and \( y \)). Indeed, we have

\[
\frac{f(x + t) - (T_x^{k-1} f)(t)}{t^k} = \frac{1}{k!} f^{(k)}(\xi)
\]

and so, as long as we have good uniform control on \( f^{(k)} \), this gives us the alluded-to generalization of (8.1): if, for example, \( |f^{(k)}(\xi)| \leq M \) for all \( \xi \), then we’ll have

\[
f(x + t) = T_x^{k-1} f(t) + o(t^k).
\]

**Proof of Theorem 8.24.** For any \( t_0 \) such that \( x + t_0 \in (a, b) \), define \( \alpha_0 \) by the equation

\[
f(x + t_0) = (T_x^{k-1} f)(t_0) + \frac{t_0^k}{k!} \alpha_0.
\]

Our goal is to show that \( \alpha_0 = f^{(k)}(\xi) \) for some \( \xi \) between \( x \) and \( x + t_0 \). Well, on the interval between 0 and \( t_0 \), consider the function

\[
g(t) = (T_x^{k-1} f)(t) + \frac{t_0^k}{k!} \alpha_0 - f(x + t), \quad \text{for } t \text{ between } 0 \text{ and } t_0.
\]

As \( f \) and \( T_x^{k-1} f \) are \( C^{k-1} \) and have \( k \)th derivatives, the same holds true of \( g \). By definition of \( \alpha_0 \), \( g(t_0) = 0 \). Also, \((T_x^{k-1} f)(0) = f(x)\) by definition, so \( g(0) = 0 \). Therefore, by the mean value theorem there is a point \( t_1 \) between 0 and \( t_0 \) such that \( g'(t_1) = \frac{g(t_0) - g(0)}{t_0} = 0 \). But by construction of Taylor polynomials, and the fact that \( \frac{d}{dt} t^k k! t^{k-1} = 0 \) at \( t = 0 \), we have \( g'(0) = (T_x^{k-1} f)'(0) - f'(x) = 0 \), and so we can apply the Mean Value Theorem to \( g' \) to find a point \( t_2 \) between 0 and \( t_1 \) where \( g''(t_2) = 0 \). We may continue this way finding \( t_1, t_2, t_3, \ldots, t_k \), with \( t_j \) between 0 and \( t_{j-1} \), such that \( g^{(j)}(t_j) = 0 \). At the last step, since \( T_x^{k-1} f \) is a polynomial of degree \( k - 1 \), its \( k \)th derivative is 0, and also \( \frac{d^k}{dt^k} \frac{1}{k!} t^k = 1 \), so

\[
0 = g^{(k)}(t_k) = (T_x^{k-1} f)^{(k-1)}(t_k) + \alpha_0 - f^{(k)}(x + t_k) = \alpha_0 - f^{(k)}(x + t_k).
\]

Setting \( \xi = x + t_k \) concludes the proof.

Taylor’s theorem is one of the most ubiquitously useful results in analysis, and we will use it frequently. As a first application, we now use it to understand another very powerful computational tool: l’Hôpital’s rule. This deals with limits of the form

\[
\lim_{y \to x} \frac{f(y)}{g(y)}
\]

where \( \lim_{y \to x} f(y) = \lim_{y \to x} g(y) = 0 \). (There are some other cases as well, which we will discuss below.) To understand what happens to such a limit, let us suppose for the moment that \( f \) and \( g \) are \( C^2 \) in a neighborhood of \( x \). These constraints are much stronger than required, as we’ll see in the
proof below; but with these assumptions, we can use Taylor’s theorem to understand what happens. Since \( f \) and \( g \) are continuous at \( x \), we have \( f(x) = \lim_{y\to x} f(y) = 0 \) and \( g(x) = \lim_{y\to x} g(y) = 0 \). We then have, for sufficiently small \( t \),

\[
\begin{align*}
    f(x + t) &= f(x) + f'(x)t + \frac{1}{2}f''(\xi)t^2 = f'(x)t + \frac{1}{2}f''(\xi)t^2 \\
    g(x + t) &= g(x) + g'(x)t + \frac{1}{2}g''(\eta)t^2 = g'(x)t + \frac{1}{2}g''(\eta)t^2
\end{align*}
\]

for some \( \xi \) and \( \eta \) between \( x \) and \( x + t \). Thus

\[
\frac{f(x + t)}{g(x + t)} = \frac{f'(x) + \frac{1}{2}f''(\xi)t}{g'(x) + \frac{1}{2}g''(\eta)t}.
\]

Since we assumed \( f, g \in C^2 \), the functions \( f'' \) and \( g'' \) are continuous. As \( t \to 0 \), since \( \xi \) and \( \eta \) are between \( x \) and \( x + t \), we see that \( \xi \to x \) and \( \eta \to x \), and so \( f''(\xi) \to f''(x) \) and \( g''(\xi) \to g''(x) \). By the limit theorems, then we see that if \( g'(x) \neq 0 \), then

\[
\lim_{y\to x} \frac{f(y)}{g(y)} = \lim_{t\to 0} \frac{f(x + t)}{g(x + t)} = \frac{f'(x)}{g'(x)}.
\]

This is the result of L’Hôpital’s rule: the limit of a ratio of functions that is indeterminate is equal to the limit of ratio of derivatives (provided it is not indeterminate). If \( f'(x) = g'(x) = 0 \), we could then apply the same reasoning with higher derivatives (assuming the functions are \( C^3 \), for example) to get \( \frac{f''(x)}{g''(x)} \), and so forth.
4. Lecture 7: April 19, 2016

Having used Taylor’s theorem to explore why L'Hôpital’s rule makes sense (and prove it for nice enough functions), we now go about proving that it holds much more generally: we do not need the functions to be $C^2$, or even $C^1$, merely differentiable; and they need not be differentiable at the point $x$, only in a (one-sided) neighborhood of $x$.

**Theorem 8.25 (L'Hôpital’s Rule).** Let $a < b$ in $\mathbb{R}$, and let $f, g: (a, b) \to \mathbb{R}$ be differentiable functions, with $g(x) \neq 0$ and $g'(x) \neq 0$ for all $x \in (a, b)$. Suppose that

$$\lim_{x \to a+} \frac{f'(x)}{g'(x)} = L$$

exists in $\mathbb{R}$. If $\lim_{x \to a+} f(x) = \lim_{x \to a+} g(x) = 0$, or if $\lim_{x \to a+} |g(x)| = \infty$, then

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

The analogous statement holds with $\lim_{x \to a+}$ replaced with $\lim_{x \to b-}$.

To prove L'Hôpital’s rule in this general form, we first need an extended version of the Mean Value Theorem that deals with ratios of functions.

**Lemma 8.26 (Extended Mean Value Theorem).** Let $a < b$ in $\mathbb{R}$, and let $f, g: [a, b] \to \mathbb{R}$ be continuous functions, such that $f$ and $g$ are differentiable on $(a, b)$. Assume that $g(a) \neq g(b)$. Then there is a point $x \in (a, b)$ where $g'(x) \neq 0$, and

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

The requirement that $g(b) \neq g(a)$ is simply so that the ratio on the left makes sense; the lemma can be stated more generally without this assumption, and without the conclusion that $g'(x) \neq 0$, in the form $[f(b) - f(a)]g'(x) = [g(b) - g(a)]f'(x)$. It is tempting to try to prove the lemma by applying the Mean Value Theorem to $f$ and $g$ separately, but this will only show that the ratio on the left is equal to $f'(x)/g'(y)$ for two (not necessarily equal) points $x, y$. Instead, we just follow the precise outline of the proof of the Mean Value Theorem [8.18].

**Proof.** Define $h(t) = [f(b) - f(a)]g(t) - [g(b) - g(a)]f(t)$. If $h$ is constant, then its derivative is 0, and the result is that the conclusion holds at every $x$. So assume $h$ is not constant, and wlog assume there is some point $t \in (a, b)$ where $h(t) > h(a)$. The function $h$ is continuous on the compact interval $[a, b]$, so achieves its maximum at some point $x$ in this closed interval; because there is some $t$ with $h(t) > h(a)$, we know $x \neq a$. Also, a quick calculation shows that $h(b) - h(a) = f(b)g(a) - f(a)g(b)$, and so $x \neq b$ either. Since $h$ is differentiable in $(a, b)$, and $h$ achieves its maximum at $x \in (a, b)$, it follows that $h'(x) = 0$, which yields the desired conclusion. \hfill $\square$

We can now proceed to prove L'Hôpital’s rule.

**Proof of Theorem 8.25.** We first treat the case $\lim_{x \to a+} f(x) = \lim_{x \to a+} g(x) = 0$. Here, we may extend the functions to be defined on $[a, b]$ by defining $f(a) = g(a) = 0$, and this (by definition) makes them continuous on this interval. Let $(x_n)$ be a sequence in $(a, b)$ such that $x_n \to a$ as $n \to \infty$. Then $f$ and $g$ are continuous on $[a, x_n]$, and differentiable on $(a, x_n)$,
with $g'(x) \neq 0$ for all $x$ in this interval. By the extended Mean Value Theorem, there is a point $t_n \in (a, x_n)$ such that

$$\frac{f'(t_n)}{g'(t_n)} = \frac{f(x_n) - f(a)}{g(x_n) - g(a)} = \frac{f(x_n)}{g(x_n)}.$$  

By the Squeeze theorem, $t_n \to a$, and so by assumption $\lim_{n \to \infty} f'(t_n)/g'(t_n) \to L$. Hence, the function $r(x) = f(x)/g(x)$ satisfies $r(x_n) \to L$ for every sequence $(x_n)$ in $(a, b)$ for which $x_n \to a$; by definition of right limit, the conclusion follows.

Now consider the case that $\lim_{x \to a+} |g(x)| = \infty$. The idea here is related, but a little more complicated, and is more amenable to an $\epsilon$-$\delta$ proof. By assumption $\lim_{x \to a+} \frac{f(x)}{g(x)} = L$; thus, for any $\epsilon > 0$, there is some $\delta_0 > 0$ so that

$$\left| \frac{f'(t)}{g'(t)} - L \right| < \frac{\epsilon}{4} \quad \text{for all} \quad t \in (a, a + \delta_0). \tag{8.2}$$  

(We freely assume $\delta_0$ is small enough that $a + \delta_0 < b$, of course.) Fix some $x_0 \in (a, a + \delta_0)$. Since $|g(x)| \to \infty$ as $x \to a$, there is some $\delta_1 > 0$ so that, for $x \in (a, a + \delta_1)$, $|g(x)| > |g(x_0)|$; in particular, $g(x) \neq g(x_0)$ for $x \in (a, a + \delta_1)$. Now, both $f$ and $g$ are continuous on $[a, x_0]$, and differentiable on $(a, x_0)$. It follows from the Extended Mean Value Theorem that for each $x \in (a, a + \delta_1)$, there is a point $t_0 = t(x, x_0)$ in $(x, x_0)$ (and therefore in $(a, a + \delta_0)$) such that

$$\frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(t_0)}{g'(t_0)}.$$  

We are interested in $\frac{f(x)}{g(x)}$, not the more complicated quotient above; but since $|g(x)|$ grows without bound, the two are very close. Dividing through top and bottom by $g(x)$ gives

$$\frac{f'(t_0)}{g'(t_0)} = \frac{f(x)}{g(x)} = \left(1 - \frac{g(x_0)}{g(x)}\right) \frac{f'(t_0)}{g'(t_0)} + \frac{f(x_0)}{g(x)}. \tag{8.3}$$  

We now have all the pieces in place. Before proceeding formally, let’s outline how this works. From (8.2), $\frac{f'(t_0)}{g'(t_0)}$ is close to $L$. Since $|g(x)| \to \infty$, for $x$ close to $a$, $1 - \frac{g(x_0)}{g(x)}$ is close to 1, and $\frac{f(x_0)}{g(x)}$ is close to 0. This shows that for such $x$, $\frac{f(x)}{g(x)}$ is close to $L$, as desired.

Now let’s make this precise. Working from (8.3), if $a < x < a + \min\{\delta_0, \delta_1\}$, then $t_0 \in (a, a + \delta_0)$, and we have

$$\left| \frac{f(x)}{g(x)} - L \right| \leq \left| \left(1 - \frac{g(x_0)}{g(x)}\right) \frac{f'(t_0)}{g'(t_0)} - L \right| + \left| \frac{f(x_0)}{g(x)} \right|$$

$$= \left| \left(1 - \frac{g(x_0)}{g(x)}\right) \left( \frac{f'(t_0)}{g'(t_0)} - L \right) - L \frac{g(x_0)}{g(x)} \right| + \left| \frac{f(x_0)}{g(x)} \right|$$

$$\leq \left| 1 - \frac{g(x_0)}{g(x)} \right| \left| \frac{f'(t_0)}{g'(t_0)} - L \right| + |L| \left| \frac{g(x_0)}{g(x)} \right| + \left| \frac{f(x_0)}{g(x)} \right|$$

$$< \left(1 + \frac{g(x_0)}{g(x)}\right) \frac{\epsilon}{4} + |L| \left| \frac{g(x_0)}{g(x)} \right| + \left| \frac{f(x_0)}{g(x)} \right|, \tag{8.4}$$
where we used (8.2) in (8.4). As \( |g(x)| \to \infty \) as \( x \to a^+ \), and since \( |f(x_0)| \) and \( |g(x_0)| \) are fixed finite numbers, we can find some \( \delta_2 > 0 \) so that, for all \( x \in (a, a + \delta_2) \),

\[
\frac{|g(x_0)|}{|g(x)|} < 1 \quad \text{and} \quad |L| \frac{|g(x_0)|}{|g(x)|} + |f(x_0)| \frac{|g(x)|}{|g(x_0)|} < \frac{\epsilon}{2}.
\]

The first bound shows that the first term in (8.4) is \( < \frac{\epsilon}{2} \), and the second bound shows that the last terms in (8.4) are \( < \frac{\epsilon}{2} \). Thus

\[
\left| \frac{f(x)}{g(x)} - L \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad \text{for} \quad x \in (a, a + \min\{\delta_0, \delta_1, \delta_2\}).
\]

As \( \epsilon > 0 \) was chosen arbitrarily, this shows that \( \lim_{x \to a^+} \frac{f(x)}{g(x)} = L \), as desired. \( \square \)

**Remark 8.27.** The same holds true with \( a = -\infty \) or \( b = +\infty \), and the proof is extremely similar; the details are left to the bored reader.

Let us now conclude our discussion of differentiation by considering how much of what we’ve developed applies to vector-valued functions of a real variable. First, the definitions are essentially the same.

**Definition 8.28.** Let \( a < b \) in \( \mathbb{R} \), let \( d \in \mathbb{N} \), and let \( f : (a, b) \to \mathbb{R}^d \) be a functions. Call \( f \) differentiable at \( t_0 \in [a, b] \) if

\[
f'(t_0) = \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}
\]

exists. Say that \( f \) is differentiable if \( f'(t_0) \) exists for each \( t_0 \) in the domain of \( f \); in this case, \( f' \) is a function \([a, b) \to \mathbb{R}^d \) as well. If this function is continuous, we say \( f \in C^1 \). More generally, for \( k \in \mathbb{N} \), we say \( f \in C^k \) if \( f \) has \( k \) continuous derivatives. If \( f \in C^k \) for all \( k \), we say \( f \in C^\infty \), and call \( f \) smooth.

It is often customary to use the variable name \( t \) for vector-valued functions, as we often think of \( f \) as tracing out a curve in space, with \( t \) tracking the flow of time. It is then very important to note that smoothness properties of \( f \) are different from smoothness properties of the range \( f[a, b] \): it is perfectly possible to construct a smooth function \( f : [0, 1] \to \mathbb{R}^2 \) such that the image \( f[0, 1] \) is a curve with a right angle in it. We may discuss this later, time permitting.

**Proposition 8.29.** A function \( f = (f_1, f_2, \ldots, f_d) \) is differentiable at \( t_0 \) if and only if each of its component functions \( f_1, f_2, \ldots, f_d \) is differentiable at \( t_0 \), in which case

\[
f'(t_0) = \left( f'_1(t_0), f'_2(t_0), \ldots, f'_d(t_0) \right).
\]

More generally, a function \( f \) is \( C^k \) iff its component functions are \( C^k \).

**Proof.** We just write out that

\[
f(t) - f(t_0) = (f_1(t) - f_1(t_0), f_2(t) - f_2(t_0), \ldots, f_d(t) - f_d(t_0))
\]

and so, dividing by \( t - t_0 \), we see that the difference quotient is a vector whose \( j \)th component is \( \frac{f_j(t) - f_j(t_0)}{t - t_0} \). Since a limit of a vector sequence exists iff the limits of all its components exist, in which case the limit is the vector of component limits (as proved, e.g., in the case \( d = 2 \) in Proposition 3.13), the first result follows. The same applies to \( C^k \) functions by noting that continuity obeys the same structure (a vector-valued function is continuous iff its component functions are all continuous), and then iterating. \( \square \)
Example 8.30. As a special case, consider functions \( f : \mathbb{R} \to \mathbb{R}^2 \), where we identify \( \mathbb{R}^2 \cong \mathbb{C} \); then we may write
\[
    f(t) = f_1(t) + if_2(t).
\]
For example, taking it on faith that \( \cos \) and \( \sin \) are differentiable functions satisfying \( \sin' = \cos \) and \( \cos' = \sin \), consider the function
\[
    e(t) = \cos t + i \sin t.
\]
Both component functions are smooth, and therefore \( e \) is a smooth function. Indeed, this function traces out the unit circle at unit angular speed in the counter-clockwise direction. Note that
\[
    e'(t) = -\sin t + i \cos t = i(i \sin t + \cos t) = ie(t).
\]

Thinking back to your knowledge of differential equations, if \( u \) is a smooth functions satisfying \( u'(t) = au(t) \) for some constant \( a \), then \( u(t) = e^{at}u(0) \). In our case \( e(0) = 1 + 0i = 1 \), so we should expect that \( e(t) = e^{it} \). This is, in fact, true, and we will prove it more satisfactorily later this quarter.

While the definitions and notation of derivatives apply equally well to vector-valued functions, the major theorems do not apply. For example, consider the Mean Value Theorem. One might think that one could show that \( \frac{f(b)-f(a)}{b-a} = f'(\xi) \) for some \( \xi \in (a, b) \), the usual way one might approach this (componentwise) fails. The Mean Value Theorem applied to each component \( f_j \) asserts that there is some point \( \xi_j \in (a, b) \) where \( \frac{f_j(b)-f_j(a)}{b-a} = f_j'(\xi_j) \); but the point \( \xi_j \) can certainly depend on \( j \), and there is no obvious way to guarantee that a single point \( \xi \) will work for all \( j \). In fact, this is just not true.

Example 8.31. Continuing with Example 8.30 notice that \( e(2\pi) - e(0) = 1 - 1 = 0 \). However, \( e'(t) = ie(t) \), and so \( ||e'(t)||^2 = \sin^2 t + \cos^2 t = 1 \) for any \( t \). Thus, there is no point \( \xi \) where \( e'(\xi) = 0 \), and so
\[
    \frac{e(2\pi) - e(0)}{2\pi - 0} = e'(\xi) \quad \text{for any } \xi \in (0, 2\pi).
\]

The Mean Value Theorem also featured prominently in the proof of L'Hôpital's Rule. In general for vector-valued functions \( f \) and \( g \), it does not even make sense to take the ratio \( \frac{f}{g} \) (one cannot divide by a vector). In the case of \( \mathbb{C} \)-valued functions, we can divide (in the same sense that we can for real-valued functions: so long as the denominator itself does not vanish at points where we are dividing). Even so, L'Hôpital's Rule fails here.

Example 8.32. Consider the two complex-valued functions \( f(t) = t \) and \( g(t) = te^{i/t} \). That is, \( f \) is nominally complex-valued, \( f(t) = t + i0 \), while \( g(t) = t(\cos \frac{1}{t} + i \sin \frac{1}{t}) \). Of course \( \lim_{t \to 0} f(t) = 0 \). In Example 8.16 we showed that \( t \sin \frac{1}{t} \to 0 \) as \( t \to 0 \); a very similar proof shows that \( t \cos \frac{1}{t} \to 0 \) as \( t \to 0 \). Thus \( \lim_{t \to 0} g(t) = 0 \) as well. So \( \frac{f(t)}{g(t)} \) is a \( 0/0 \) type indeterminate form as \( t \to 0 \), where both top and bottom are differentiable away from 0. If L'Hôpital's Rule held precisely as in Theorem 8.25 we would expect to see \( \lim_{t \to 0} \frac{f(t)}{g(t)} = \lim_{t \to 0} \frac{f'(t)}{g'(t)} \). We can compute these derivatives (for \( t \neq 0 \)): \( f'(t) = 1 \), while
\[
    g'(t) = \frac{d}{dt} \left( t \cos \frac{1}{t} \right) + i \frac{d}{dt} \left( t \sin \frac{1}{t} \right) = \left( \cos \frac{1}{t} + \frac{1}{t} \sin \frac{1}{t} \right) + i \left( \sin \frac{1}{t} - \frac{1}{t} \cos \frac{1}{t} \right).
\]
Notice that

\[ |g'(t)|^2 = \left( \cos \frac{1}{t} + \frac{1}{t} \sin \frac{1}{t} \right)^2 + \left( \sin \frac{1}{t} - \frac{1}{t} \cos \frac{1}{t} \right)^2 = \left( \cos^2 \frac{1}{t} + \sin^2 \frac{1}{t} \right) \left( 1 + \frac{1}{t^2} \right) = 1 + \frac{1}{t^2} > 0 \]

and hence \( g'(t) \) is never equal to 0. The reciprocal can be computed: since \( \frac{1}{z} = \frac{\overline{z}}{|z|^2} \),

\[ \frac{f'(t)}{g'(t)} = \frac{1}{g'(t)} = \frac{1}{1 + 1/t^2} \left[ \left( \cos \frac{1}{t} + \frac{1}{t} \sin \frac{1}{t} \right) - i \left( \sin \frac{1}{t} - \frac{1}{t} \cos \frac{1}{t} \right) \right]. \]

Let's look at the real part for now. Simplifying, we have

\[ \frac{t^2}{t^2 + 1} \left( \cos \frac{1}{t} + \frac{1}{t} \sin \frac{1}{t} \right) = \frac{1}{t^2 + 1} \left( t^2 \cos \frac{1}{t} + t \sin \frac{1}{t} \right). \]

At \( t \to 0 \), \( \frac{1}{t^2 + 1} \to 1 \), and both terms inside the brackets tend to 0 (cf. Example 8.17). Similar arguments show that the second term in \( \frac{f'(t)}{g'(t)} \) tends to 0; so \( \lim_{t \to 0} \frac{f'(t)}{g'(t)} = 0 \).

However, we can compute directly (using computations just like the ones above, or by following our nose with the exponential notation) that

\[ \frac{f(t)}{g(t)} = \frac{t}{te^{i/t}} = e^{-i/t} = \cos \frac{1}{t} - i \sin \frac{1}{t} \]

and as we know (cf. Example 7.18), this limit does not exists. So the statement of L'Hôpital’s rule is simply false for complex-valued functions.

We have claimed that the Mean Value Theorem is they key tool of calculus; and L’Hôpital’s rule is a very powerful computational tool. Are we then lost when it comes to calculus of vector-valued functions of a real variable? No. The saving grace is that, while the Mean Value Theorem does not hold, it does hold as an inequality in general, and that is enough for most theoretical applications.

**Theorem 8.33 (Mean Value Theorem for Vector-Valued Functions).** Let \( a < b \) in \( \mathbb{R} \), \( d \in \mathbb{N} \), and let \( f: [a, b] \to \mathbb{R}^d \) be continuous, and differentiable on \((a, b)\). Then there exists a point \( \xi \in (a, b) \) where

\[ \|f(b) - f(a)\| \leq (b - a)\|f'(\xi)\|. \]

Note that this result does not contradict Example 8.31; in that example, the left-hand-side of the inequality is 0, which is certainly \( \leq \) the non-negative right-hand-side.

To prove Theorem 8.33 we need one key inequality which is one of the most important inequalities in all of analysis.

**Lemma 8.34 (The Cauchy-Schwarz Inequality).** Let \( d \in \mathbb{N} \), and let \( v, w \in \mathbb{R}^d \). Then \( |v \cdot w| \leq \|v\|\|w\| \).

**Proof.** In this finite-dimensional setting, it is possible to prove the inequality by squaring both sides and comparing terms directly. But it is instructive to give a cleaner proof even here. Consider the function \( p(t) = \|v - tw\|^2 \). This function is always \( \geq 0 \). Expanding it out, we have

\[ p(t) = (v - tw) \cdot (v - tw) = v \cdot v - 2tv \cdot w + t^2w \cdot w = \|v\|^2 - 2tv \cdot w + t^2\|w\|^2 = at^2 + bt + c \]
where $a = \|w\|^2$, $b = -2v \cdot w$, and $c = \|v\|^2$. Thus $p$ is a quadratic polynomial, with leading coefficient $a > 0$. Since we know $p(t) \geq 0$ for all $t$, it follows that $p$ cannot have two distinct real roots (if so, the function would be strictly negative between those roots). Thus, either $p$ has no real roots, or a double-root. From the quadratic formula, it thus follows that the discriminant $b^2 - 4ac \leq 0$. That is:

$$0 \geq b^2 - 4ac = (-2v \cdot w)^2 - 4\|w\|^2\|v\|^2.$$ This shows that $(v \cdot w)^2 \leq \|v\|^2\|w\|^2$; taking square roots yields the result. □

**Remark 8.35.** This proof relied on nothing more than the fact that the norm $N(v) = \|v\|$ “polarizes” in terms of the bilinear form $B(v, w) = v \cdot w$: $N(v)^2 = B(v, v)$. Indeed, the theorem holds in this level of generality: if $B$ is a symmetric bilinear form on any vector space for which $B(v, v) \geq 0$ for all $v$, then

$$|B(v, w)| \leq B(v, v)^{1/2}B(w, w)^{1/2}.$$ 

This generalization is especially handy in many infinite-dimensional settings.

**Proof of Theorem 8.33.** Set $v = f(b) - f(a) \in \mathbb{R}^d$. Consider the real-valued function

$$\varphi(t) = v \cdot f(t) = \sum_{j=1}^{d} v_j f_j(t), \quad t \in [a, b].$$

As each component $v_j f_j$ is a continuous function on $[a, b]$, differentiable on $(a, b)$, the same applies to the sum, and so $\varphi$ is the kind of function to which the Mean Value Theorem applies. Thus, there is a point $\xi \in (a, b)$ where $\varphi'(\xi) = \frac{\varphi(b) - \varphi(a)}{b - a}$. That is to say

$$(b - a)\varphi'(\xi) = \varphi(b) - \varphi(a) = v \cdot f(b) - v \cdot f(a) = v \cdot [f(b) - f(a)] = v \cdot v = \|v\|^2.$$ 

Also, note that

$$\varphi'(\xi) = \frac{d}{dt}\left| \sum_{j=1}^{d} v_j f_j(t) \right|_{t=\xi} = \sum_{j=1}^{d} v_j f_j'(\xi) = v \cdot f'(\xi).$$

Thus we have

$$(b - a) v \cdot f'(\xi) = \|v\|^2.$$ 

Taking absolute values and applying the Cauchy-Schwarz inequality,

$$\|v\|^2 = |\|v\|^2| = |(b - a) v \cdot f'(\xi)| \leq (b - a) \|v\|\|f'(\xi)\|.$$ 

If $v = f(b) - f(a) = 0$, there is nothing to prove; otherwise we can divide out one factor of $\|v\|$ to yield

$$\|f(b) - f(a)\| = \|v\| \leq (b - a)\|f'(\xi)\|$$

which is the desired conclusion. □

**Corollary 8.36.** If $f \in C^1(a, b)$, then

$$\|f(s) - f(t)\| \leq \sup_{s \leq \xi \leq t} \|f'(\xi)\| |s - t|, \quad \text{for all } s, t \in (a, b).$$

This allows for a vector-valued version of Lipschitz functions (cf. Homework 3): we define the Lipschitz norm of such a function to be

$$\|f\|_{\text{Lip}} = \sup_{s \neq t} \frac{\|f(s) - f(t)\|}{|s - t|}.$$
and we call a vector-valued function $f$ Lipschitz if $\|f\|_{\text{Lip}} < \infty$. Corollary 8.36 shows that if $f \in C^1$ with a bounded derivative then $f$ is Lipschitz, and $\|f\|_{\text{Lip}} \leq \sup_{\xi} \|f'(\xi)\|$; in fact, these are equal for the same reason they are in the scalar-valued case covered on the homework exercise (since $\|f'(t)\|$ is very close to $\frac{\|f(s) - f(t)\|}{|s-t|}$ when $s$ is close to $t$). One can then follow with a comparable theory of Hölder continuous vector-valued functions as well; this is important, but for now it is left to the reader’s imagination.

One final remark: since Taylor’s theorem was proved simply by iterating the Mean Value Theorem, in the vector-valued setting, one cannot expect a Taylor expansion with a remainder term analogous to the one in Theorem 8.24. One can, instead, formulate a Taylor inequality, where

$$\|f(x + t) - T^{k-1}f(x)\| \leq \frac{1}{k!} \|f^{(k)}(\xi)\| |t|^k,$$

for some point $\xi$ between $x$ and $x + t$.

This is just as useful for approximations; but in truth, most applications of Taylor’s theorem in the vector-valued setting are just as amenable to applying the Theorem to each component of $f$ separately and working from there.
CHAPTER 9

Integration

1. Lecture 8: April 21, 2016

You likely saw some version of the actual definition of the integral in your calculus class: it was presented as a “limit of Riemann sums”. This is not quite accurate: it is not a limit in the sense that we’ve defined in this course. (It is a much more complicated kind of limit.) There are several approached to making sense of this rigorously. We are going to present the largely historically-accurate version here developed by Riemann somewhat non-rigorously, and really properly developed by Darboux (indeed, in some sources it is called the Darboux integral).

**Definition 9.1.** Let \( a < b \) in \( \mathbb{R} \). A **partition** \( \Pi \) of \([a, b]\) is a finite set of points \( \{t_0, t_1, \ldots, t_n\} \) where

\[
a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b.
\]

For \( 1 \leq j \leq n \), denote \( \Delta t_j = x_j - x_{j-1} \).

A word on notation: we will usually have a fixed partition around, and the letter \( n \) will be used consistently to mean the index of the largest partition point (unless otherwise stated).

**Definition 9.2.** Let \( a < b \) in \( \mathbb{R} \). Given a bounded function \( f : [a, b] \to \mathbb{R} \), and a partition \( \Pi = \{t_0, t_1, \ldots, t_n\} \) of \([a, b]\), we can define **upper and lower sums** of \( f \) on \( \Pi \):

\[
U(f, \Pi) \equiv \sum_{j=1}^{n} \sup_{t_{j-1} \leq t \leq t_j} f(t) \cdot \Delta t_j
\]

\[
L(f, \Pi) \equiv \sum_{j=1}^{n} \inf_{t_{j-1} \leq t \leq t_j} f(t) \cdot \Delta t_j.
\]

Note that this only makes sense for bounded \( f \); if \( f \) is not bounded, then on at least one partition interval \([t_{j-1}, t_j]\) \( f \) is unbounded, and so at least one of the terms in either \( U(f, \Pi) \) or \( L(f, \Pi) \) will be \( \pm \infty \); but there could be more than one term that is \( \pm \infty \), perhaps with opposite signs, and so the sums may not even be defined. Thus, we restrict our attention to bounded functions for now; we will later discuss how to extend integration to some unbounded functions.

**Definition 9.3.** Let \( a < b \) in \( \mathbb{R} \), and let \( f : [a, b] \to \mathbb{R} \). Define the **upper and lower Darboux integrals** of \( f \) as follows:

\[
U(f) = \inf_{\Pi} U(f, \Pi), \quad L(f) = \sup_{\Pi} L(f, \Pi).
\]

The \( \inf \) and \( \sup \) are taken over all partitions of \([a, b]\). If \( U(f) = L(f) \), we say that \( f \) is **Riemann integrable** (or **Darboux integrable**), and denote this common value by

\[
\int_a^b f = \int_a^b f(t) \, dt \equiv U(f) = L(f).
\]
Remark 9.4. The textbook uses the notation

\[ U(f) = \int_a^b f(t) \, dt \quad \text{and} \quad L(f) = \int_a^b f(t) \, dt. \]

We will probably never use this notation.

Before continuing, let’s discuss where these definitions come from. \( U(f, \Pi) \) and \( L(f, \Pi) \) are approximations of the “area under the curve” of the graph of \( f \): you partition the domain \([a, b]\) into a finite collection of adjacent intervals, and on each interval you approximate the function \( f \) by a constant one. Which constant should you choose? The upper sum has you choose the largest (or more precisely supral) value on that interval, while the lower sum chooses the smallest (or more precisely infimal) value. There are other possibilities: the original scheme by Riemann is to choose another set of points \( t_1^*, t_2^*, \ldots, t_n^* \) with \( t_{j-1} \leq t_j^* \leq t_j \), and then approximate \( f \) by the constant \( f(t_j^*) \) on the interval \([t_{j-1}, t_j]\). If the points \( \{t_j^*\} \) are denoted together as \( \Pi^* \), we might denote this sum as

\[ R(f, \Pi, \Pi^*) = \sum_{j=1}^n f(t_j^*) \Delta t_j \]

which is the kind of “Riemann sum” you saw in calculus. In all cases, we then define an approximate integral to be the (signed) area of the rectangles with base length \( \Delta t_j \) and height given by the appropriate constant approximation of the function. Notice that, by definition of sup and inf on each interval,

\[ L(f, \Pi) \leq R(f, \Pi, \Pi^*) \leq U(f, \Pi) \]

and so if \( U(f) = L(f) \), the Riemann sums must also “converge” to this common value.

How do we compute the area under the actual curve, assuming that even makes sense? In the Riemann scheme, we then start changing the partition \( \Pi \) (and the associated points \( \Pi^* \)) to make the maximum width of any \( \Delta t_j \) smaller; this is supposed to give a better approximation, and then we take the “limit” as this width goes to 0. In the Darboux approach (which we are following), notice that \( U(f, P) \) is definitely an over-estimate for the “actual” area: if you change the partition \( \Pi \) (for example by refining it to add a new point splitting one of the intervals into two), you can only decrease the value of the sum, since the sup of \( f \) on the two sub-intervals cannot be larger than the sup on the whole interval: for \( t_{j-1} \leq s \leq t_j \),

\[ \sup_{t_{j-1} \leq t \leq t_j} f(t) \cdot (t_j - t_{j-1}) \geq \sup_{t_{j-1} \leq t \leq s} f(t) \cdot (s - t_{j-1}) + \sup_{s \leq t \leq t_j} f(t) \cdot (t_j - s). \] (9.1)

(You should draw a quick picture to see why this is true.) Hence, replacing \( \Pi \) with a refined partition \( \Pi' \) will yield \( U(f, \Pi') \leq U(f, \Pi) \). A similar argument shows that \( L(f, \Pi') \geq L(f, \Pi) \).

This is why we define the upper Darboux integral to be the infimum of the upper sums over all partitions, and likewise the lower Darboux integral is the supremum of the lower sums over all partitions.

Of course, we need to make sure these infima and suprema make sense. Let’s summarize this with the following lemma.

Lemma 9.5. Let \( a < b \) in \( \mathbb{R} \), and let \( f : [a, b] \to \mathbb{R} \) be a bounded function, with \( |f(t)| \leq M \) for all \( t \in [a, b] \). Then for any partitions \( \Pi_1 \) and \( \Pi_2 \) of \([a, b]\),

\[-M(b-a) \leq L(f, \Pi_1) \leq U(f, \Pi_2) \leq M(b-a).\]

Ergo

\[-M(b-a) \leq L(f) \leq U(f) \leq M(b-a).\]
PROOF. Let $\Pi_1 = \{s_0 < s_1 < \cdots < s_m\}$ and $\Pi_2 = \{t_0 < t_1 < \cdots < t_n\}$. Since $f(t) \leq M$ for all $t$, $M$ is an upper bound for $\{f(t): t_{j-1} \leq t \leq t_j\}$ for each $j$, and hence $\sup_{t_{j-1} \leq t \leq t_j} f(t) \leq M$. Thus

$$U(f, \Pi_2) = \sum_{j=1}^n \sup_{t_{j-1} \leq t \leq t_j} f(t) \cdot \Delta t_j \leq \sum_{j=1}^n M \Delta t_j = M \sum_{j=1}^n (t_j - t_{j-1}) = M(b - a).$$

(The last sum is telescoping.) A similar proof, using the fact that $f(t) \geq -M$ for all $t \in [a, b]$, shows that $L(f, \Pi_1) \geq -M(b - a)$. For the middle inequality, we use an important trick: we introduce a new partition $\Pi$ which is the \textit{common refinement} of $\Pi_1$ and $\Pi_2$: $\Pi = \Pi_1 \cup \Pi_2 = \{s_0, s_1, \ldots, s_m, t_0, t_1, \ldots, t_n\}$ (not written in order here). For this common partition (whose points we’ll refer to as $u_j$), we note that, for each $j$,

$$\inf_{u_{j-1} \leq t \leq u_j} f(t) \leq \sup_{u_{j-1} \leq t \leq u_j} f(t).$$

Multiplying both sides by the positive number $\Delta u_j$ and summing up yields $L(f, \Pi) \leq U(f, \Pi)$. Then, using (9.1), we see (by induction) that

$$L(f, \Pi_1) \leq L(f, \Pi) \leq U(f, \Pi) \leq U(f, \Pi_2).$$

This concludes the proof of the first chain of inequalities. For the second: since $U(f, \Pi_2) \leq M(b - a)$ for all $\Pi_2$, it follows that $U(f) \leq M(b - a)$; a similar argument shows that $L(f, \Pi_1) \geq -M(b - a)$. For the middle inequality, we hold $\Pi_2$ fixed: since $L(f, \Pi_1) \leq U(f, \Pi_2)$ for all $\Pi_1$, it follows that $L(f) = \sup_{\Pi_1} L(f, \Pi_1) \leq U(f, \Pi_2)$. Thus $L(f)$ is a lower bound for $\{U(f, \Pi_2): \Pi_2\}$, and so $U(f) = \inf_{\Pi_2} U(f, \Pi_2) \geq L(f)$, as desired. \qed

Thus, $U(f)$ and $L(f)$ are well-defined for any bounded function $f$, and ordered $L(f) \leq U(f)$. The question is whether they’re equal. The answer is: certainly not always.

**Example 9.6.** Consider Dirichlet’s function from Example [6.1] the indicator function of the rationals.

$$f(x) = \begin{cases} 0, & x \notin \mathbb{Q} \\ 1, & x \in \mathbb{Q} \end{cases}.$$  

Taking this function for $x \in [0, 1]$ let us compute the upper and lower integrals. First, fix any partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = 1\}$ of $[0, 1]$. On any interval $[t_{j-1}, t_j]$, since $t_{j-1} < t_j$, we know there are both rational and irrational points in the interval. It follows that

$$\sup_{t_{j-1} \leq t \leq t_j} f(t) = 1, \quad \inf_{t_{j-1} \leq t \leq t_j} f(t) = 0.$$

Thus

$$U(f, \Pi) = \sum_{j=1}^n 1 \cdot \Delta t_j = (1 - 0) = 1, \quad L(f, \Pi) = \sum_{j=1}^n 0 \cdot \Delta t_j = 0.$$

So $U(f, \Pi)$ and $L(f, \Pi)$ do not depend on $\Pi$, and therefore taking appropriate $\sup$ and $\inf$, we see that $U(f) = 1$ while $L(f) = 0$. Ergo, $f$ is not Riemann integrable: $U(f) \neq L(f)$.

The question of when it is true that $U(f) = L(f)$ is a delicate one having to do with the continuity properties of the function $f$. We will explore and answer this question fully. Before we do, it pays to be a little more general right away (without adding any abstraction), and talk about the \textit{Riemann-Stieltjes} integral. That is our next topic.
2. Lecture 9: April 26, 2016

The generalization of the Riemann / Darboux integral we will now develop allows for a “weighting” of the domain. In the construction of the integral, when approximating the area under the graph of a function by rectangles, we may compute the “area” of each partition rectangle over \([t_{j-1}, t_j]\) by declaring the length of this interval is not necessarily \(\Delta t_j\) but instead some (well-behaved) function of this width. In fact, a wide variety of weight functions are possible to produce meaningful theories of integration, but to mimick precisely the upper and lower integral construction outlined in the previous lecture, we restrict ourselves to monotone increasing weight functions.

**Definition 9.7.** Let \(a < b\) in \(\mathbb{R}\), and let \(\alpha: [a, b] \rightarrow \mathbb{R}\) be a monotone increasing function. (In particular, since \(\alpha([a, b]) = [\alpha(a), \alpha(b)]\), \(\alpha\) is also bounded.) Given a partition \(\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}\), define \(\Delta \alpha_j = \alpha(t_j) - \alpha(t_{j-1})\) (which is \(\geq 0\) since \(\alpha\) is monotone increasing).

Let \(f: [a, b] \rightarrow \mathbb{R}\) be a bounded function. Define the **upper sum** and **lower sum** of \(f\) relative to \(\Pi\) and \(\alpha\) as

\[
U(f, \Pi, \alpha) \equiv \sum_{j=1}^{n} \sup_{t_{j-1} \leq t \leq t_j} f(t) \cdot \Delta \alpha_j \\
L(f, \Pi, \alpha) \equiv \sum_{j=1}^{n} \inf_{t_{j-1} \leq t \leq t_j} f(t) \cdot \Delta \alpha_j
\]

An argument exactly like the one in Lemma 9.5 shows that, if \(|f(t)| \leq M\) for all \(t\) then for any partitions \(\Pi_1\) and \(\Pi_2\), with \(\Pi^* = \Pi_1 \cup \Pi_2\), we have

\[
-M(\alpha(b) - \alpha(a)) \leq L(f, \Pi_1, \alpha) \leq L(f, \Pi^*, \alpha) \leq U(f, \Pi_2, \alpha) \leq M(\alpha(b) - \alpha(a)). \tag{9.2}
\]

In particular, the definition / proposition makes sense.

**Proposition 9.8.** Let \(a < b\) in \(\mathbb{R}\), let \(\alpha: [a, b] \rightarrow \mathbb{R}\) be monotone increasing, and let \(f: [a, b] \rightarrow \mathbb{R}\) be bounded. Define

\[
U(f, \alpha) \equiv \inf_{\Pi} U(f, \Pi, \alpha), \quad \text{and} \quad L(f, \alpha) = \sup_{\Pi} L(f, \Pi, \alpha).
\]

Then

\[
-M(\alpha(b) - \alpha(a)) \leq L(f, \alpha) \leq U(f, \alpha) \leq M(\alpha(b) - \alpha(a)).
\]

If \(L(f, \alpha) = U(f, \alpha)\), we say that \(f\) is **Riemann-Stieltjes integrable** with respect to \(\alpha\), and write \(f \in \mathcal{R}(\alpha)\). In this case, we denote the common value by

\[
\int_{a}^{b} f \, d\alpha = \int_{a}^{b} f(t) \, d\alpha(t) \equiv U(f, \alpha) = L(f, \alpha).
\]

**Remark 9.9.** We will tend not to use the second notation \(\int_{a}^{b} f(t) \, d\alpha(t)\), since the \(t\) is just a dummy variable and carries no independent meaning here.

Taking \(\alpha(x) = x\), this reduces to the Riemann integral theory discussed in the previous lecture. In this case, we write simply \(f \in \mathcal{R}\). We might write the integral in this case as \(\int_{a}^{b} f(t) \, dt\), but more likely just as \(\int_{a}^{b} f\).

We now give a quantitative characterization of Riemann-Stieltjes integrability.
Lemma 9.10. Let $a < b$ in $\mathbb{R}$, let $f: [a, b] \to \mathbb{R}$ be bounded, and let $\alpha: [a, b] \to \mathbb{R}$ be monotone increasing. Then $f \in \mathcal{R}(\alpha)$ if and only if for each $\epsilon > 0$ there is a partition $\Pi$ of $[a, b]$ such that

$$U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \epsilon. \quad (9.3)$$

Proof. First, suppose that (9.3) holds; so fix $\epsilon > 0$ and let $\Pi$ be a partition verifying that inequality. By definition $L(f, \Pi, \alpha) \leq L(f, \alpha) \leq U(f, \alpha) \leq U(f, \Pi, \alpha)$, and so (9.3) implies that $0 \leq U(f, \alpha) - L(f, \alpha) < \epsilon$ for each $\epsilon > 0$; thus $U(f) = L(f)$, as desired.

Conversely, suppose $f \in \mathcal{R}(\alpha)$, and let $\epsilon > 0$. Since $\int f \, d\alpha = \sup_{\Pi} L(f, \Pi, \alpha)$, there is some partition $\Pi_1$ so that $L(f, \Pi_1, \alpha) > \int f \, d\alpha - \frac{\epsilon}{2}$, and there is some partition $\Pi_2$ so that $U(f, \Pi_2, \alpha) < \int f \, d\alpha + \frac{\epsilon}{2}$. Let $\Pi^* = \Pi_1 \cup \Pi_2$ be the common refinement. Then by (9.2), we have

$$U(\Pi^*, f, \alpha) \leq U(\Pi_2, f, \alpha) < \int f \, d\alpha + \frac{\epsilon}{2} \leq L(f, \Pi_1, \alpha) + \frac{\epsilon}{2} \leq L(f, \Pi^*, \alpha) + \epsilon.$$

Thus (9.3) is verified by the partition $\Pi^*$, concluding the proof.

With the characterization of Lemma 9.10, we can now easily see that continuous functions are always Riemann-Stieltjes integrable.

Theorem 9.11. Let $a < b$ in $\mathbb{R}$, let $f: [a, b] \to \mathbb{R}$ be continuous, and let $\alpha: [a, b] \to \mathbb{R}$ be monotone increasing. Then $f \in \mathcal{R}(\alpha)$.

Proof. First, if $\alpha$ is constant on $[a, b]$, then $\Delta \alpha_j = 0$ on any interval $[t_{j-1}, t_j]$, and so $U(f, \Pi, \alpha) = L(f, \Pi, \alpha) = 0$ for all $\Pi$; thus $U(f, \alpha) = L(f, \alpha) = \int_a^b f \, d\alpha = 0$. Otherwise, we must have $\alpha(a) < \alpha(b)$. Now, since $f$ is continuous on the compact interval $[a, b]$, it is uniformly continuous there, so we may choose $\delta > 0$ such that, for any $x, y \in [a, b]$ with $|x - y| < \delta$, it follows that $|f(x) - f(y)| < \frac{\epsilon}{\alpha(b) - \alpha(a)}$.

Now, fix a partition $\Pi = \{a = a_0 < a_1 < \cdots < a_n = b\}$ of $[a, b]$ for which $\Delta \alpha_j = t_j - t_{j-1} < \delta$ for all $j$; for example, fix $n$ with $\frac{b-a}{n} < \delta$ and use the even partition $\Pi = \{a, a + \frac{b-a}{n}, a + 2\frac{b-a}{n}, \cdots, a + (n-1)\frac{b-a}{n}, b\}$. Since $f$ is continuous on each interval $[t_{j-1}, t_j]$, there are points $x_j$ and $y_j$ such that

$$\sup_{t_{j-1} \leq t \leq t_j} f(t) = f(x_j) \quad \text{and} \quad \inf_{t_{j-1} \leq t \leq t_j} f(t) = f(y_j).$$

Since $t_j - t_{j-1} < \delta$ it follows that $|x_j - y_j| < \delta$, and so $f(x_j) - f(y_j) = |f(x_j) - f(y_j)| < \frac{\epsilon}{\alpha(b) - \alpha(a)}$. Thus

$$U(f, \Pi, \alpha) - L(f, \Pi, \alpha) = \sum_{j=1}^{n} [f(x_j) - f(y_j)] \Delta \alpha_j < \sum_{j=1}^{n} \frac{\epsilon}{\alpha(b) - \alpha(a)} \Delta \alpha_j.$$

Factoring out the constant $\frac{\epsilon}{\alpha(b) - \alpha(a)}$, we have just the telescoping sum $\sum_{j=1}^{n} \Delta \alpha_j = \alpha(b) - \alpha(a)$, and so we see that with this partition, $U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \epsilon$. As we can do this for any $\epsilon > 0$, by Lemma 9.10 it follows that $f \in \mathcal{R}(\alpha)$, as desired.

So, we now know how to integrate continuous functions. In particular, taking $\alpha(x) = x$, this gives us the usual Riemann integral of continuous functions, which is the main object of study in integral calculus. However, the class of functions that can be integrated is much larger than continuous functions. For example:

Theorem 9.12. Let $a < b$ in $\mathbb{R}$, let $f: [a, b] \to \mathbb{R}$ be bounded, and let $\alpha: [a, b] \to \mathbb{R}$ be monotone increasing. Suppose that the set of points in $[a, b]$ where $f$ is discontinuous is finite, and at each such point $\alpha$ is continuous. Then $f \in \mathcal{R}(\alpha)$.
PROOF. As in the proof of Theorem 9.11, if \( \alpha \) is constant there is nothing to prove, so we freely assume \( \alpha(b) < \alpha(a) \). Fix \( \epsilon > 0 \), and let \( M = \sup |f| \). Let \( D = \{x_1, x_2, \ldots, x_d\} \) be the set of points in \( [a, b] \) where \( f \) is not continuous. By assumption \( \alpha \) is continuous at each \( x_j \), so there is some \( \eta_j > 0 \) such that \( |\alpha(s) - \alpha(t)| < \frac{\epsilon}{4Md} \) for all \( s, t \in (x_j - \eta_j, x_j + \eta_j) \cap [a, b] \). Let \( \eta \leq \min\{\eta_1, \ldots, \eta_d\} \) be small enough that \( (x_j - \eta, x_j + \eta) \subset (a, b) \) for all \( j \) (except if possibly \( x_j = a \) or \( x_j = b \)). Now define points \( u_j < v_j \) as follows: if \( x_j = a \) then \( u_j = a \) and \( v_j = a + \frac{\eta}{2} \); if \( x_j = b \) then \( u_j = b - \frac{\eta}{2} \); and if \( x_j \in (a, b) \) then \( u_j = x_j - \frac{\eta}{2} \) and \( v_j = x_j + \frac{\eta}{2} \). By definition of \( \eta \), it follows that \( \alpha(v_j) - \alpha(u_j) < \frac{\epsilon}{4Md} \) for \( 1 \leq j \leq d \), and so

\[
\sum_{j=1}^{d} [\alpha(v_j) - \alpha(u_j)] < \frac{\epsilon}{4M}.
\]

Now, let \( K = [a, b] \setminus \bigcup_{j=1}^{d}(u_j, v_j) \). This is a closed subset of \( [a, b] \) and so is compact. Because no point in \( D \) is in \( K \), \( f \) is continuous on \( K \), and hence uniformly continuous there. So we may choose \( \delta > 0 \) so that, whenever \( s, t \in K \) with \( |s - t| < \delta \), it follows that

\[
|f(s) - f(t)| < \frac{\epsilon}{2(\alpha(b) - \alpha(a))}.
\]

Now form a partition \( \Pi \) of \( [a, b] \) as follows: \( \Pi \) contains all the points \( u_j \) and \( v_j \) for \( 1 \leq j \leq d \); it contains no points in any of the intervals \( (u_j, v_j) \); and for any point \( t_i \) in \( \Pi \) not of the form \( v_j \), we have \( t_i - t_{i-1} < \delta \).

We now expand \( U(f, \Pi, \alpha) - L(f, \Pi, \alpha) \):

\[
U(f, \Pi, \alpha) - L(f, \Pi, \alpha) = \sum_{i}^{d} [\sup_{t_{i-1} \leq t \leq t_i} f(t) - \inf_{t_{i-1} \leq t \leq t_i} f(t)]\Delta \alpha_i.
\]

We break this sum into two kinds of terms, dividing the indices into \( i \in I_1 \) and \( i \in I_2 \): \( I_1 \) consists of those \( i \) for which \( t_i - v_j \) for some \( j \), and \( I_2 \) consists of all the others. For \( i \in I_1 \), by construction \( t_{i-1} = u_j \). For these terms we make the estimate that \( \sup_{t_{i-1} \leq t \leq t_i} f(t) - \int_{t_{i-1} \leq t \leq t_i} f(t) \leq 2M \) and so we get

\[
\sum_{i \in I_1} [\sup_{t_{i-1} \leq t \leq t_i} f(t) - \inf_{t_{i-1} \leq t \leq t_i} f(t)]\Delta \alpha_i \leq 2M \sum_{i \in I_1} \Delta \alpha_i = 2M \sum_{i \in I_1} [\alpha(t_i) - \alpha(t_{i-1})]
\]

\[
= 2M \sum_{i \in I_1} [\alpha(t_i) - \alpha(t_{i-1})]
\]

\[
= 2M \sum_{j=1}^{d} [\alpha(v_j) - \alpha(u_j)]
\]

\[
< 2M \cdot \frac{\epsilon}{4M} = \frac{\epsilon}{2}.
\]

Now, for those \( i \in I_2 \), we have constructed \( \Pi \) so that \( t_i - t_{i-1} < \delta \), which means that for any \( s, t \in [t_{i-1}, t_i] \)

\[
|f(s) - f(t)| \leq \frac{\epsilon}{2|\alpha(b) - \alpha(a)|},
\]

and so this also holds true for the difference between the supral and infimal values of \( f \) on the interval. Summing up these terms yields

\[
\sum_{i \in I_2} [\sup_{t_{i-1} \leq t \leq t_i} f(t) - \inf_{t_{i-1} \leq t \leq t_i} f(t)]\Delta \alpha_i \leq \sum_{i \in I_2} \frac{\epsilon}{2|\alpha(b) - \alpha(a)|} \cdot [\alpha(t_i) - \alpha(t_{i-1})]
\]

\[
\leq 2 \frac{\epsilon}{|\alpha(b) - \alpha(a)|} \sum_{i} [\alpha(t_i) - \alpha(t_{i-1})] = \frac{\epsilon}{2}.
\]
Thus \( U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \), and so by Lemma 9.10 it follows that \( f \in \mathcal{R}(\alpha) \). □

**Remark 9.13.** If both \( f \) and \( \alpha \) are discontinuous at a point, it is possible that \( f \notin \mathcal{R}(\alpha) \), regardless of how continuous \( f \) is elsewhere. You will work with such an example on this week’s homework.

**Remark 9.14.** It is important to note that Theorem 9.12 applies to discontinuities in general, not just jump discontinuities. In particular, we now know that the function \( f(t) = \sin \frac{1}{t} \) of Example 7.18 is Riemann-Stieltjes integrable (with respect to any monotone increasing integrator \( \alpha \)) on any compact interval, including 0 or not. And the proof shows us why. Even though the function oscillates wildly near 0, it is bounded by 1 in absolute value, and so for any small \( \delta > 0 \), the total contribution of the upper or lower sums over any partition from points in \((-\delta, \delta)\) is no bigger than \( \alpha(\delta) - \alpha(-\delta) \), which is very small so long as \( \alpha \) is continuous at 0.

**Remark 9.15.** It is natural to wonder how much further this can be taken, in terms of allowing irregular \( f \) to be integrated. For example, the above proof does not immediately generalize to the case where \( f \) has countably infinitely many discontinuities. It is a delicate matter in general to settle which \( f \) are in \( \mathcal{R}(\alpha) \) for a particular \( \alpha \). The most important case where \( \alpha(x) = x \) (the Riemann integral), however, is completely understood. In that case, even if \( f \) has (at most) countably infinitely many discontinuities, it is still Riemann integrable. In that case, the exact criterion for Riemann integrability is *continuous almost everywhere*: a bounded function \( f \) is in \( \mathcal{R} \) iff, for every \( \epsilon > 0 \), there is a countable collection of open intervals \( I_1, I_2, I_3, \ldots \) with \( \sum_j \text{length}(I_j) < \epsilon \) such that the set of discontinuities of \( f \) is contained in \( \bigcup_j I_j \). (I.e. the set of discontinuities has “measure 0”)

We now turn to the basic properties of the Riemann-Stieltjes integral, all of which reflect its nature as a kind of limit sum. We state them as a sequence of lemmas. In all cases, \( a < b \) are real numbers, \( f, f_1, f_2 : [a, b] \to \mathbb{R} \) are bounded functions, and \( \alpha, \alpha_1, \alpha_2 : [a, b] \to \mathbb{R} \) are monotone increasing functions.

**Lemma 9.16.** If \( f_1, f_2 \in \mathcal{R}(\alpha) \) and \( c \in \mathbb{R} \), then \( f_1 + f_2 \in \mathcal{R}(\alpha) \) and \( cf_1 \in \mathcal{R}(\alpha) \), with

\[
\int_a^b (f_1 + f_2) \, d\alpha = \int_a^b f_1 \, d\alpha + \int_a^b f_2 \, d\alpha \quad \text{and} \quad \int_a^b cf_1 \, d\alpha = c \int_a^b f_1 \, d\alpha.
\]

**Proof.** Fix a partition \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \). To save notation, let \( I_j = [t_{j-1}, t_j] \). Since

\[
\inf_{I_j} f_1 + \inf_{I_j} f_2 \leq \inf_{I_j} (f_1 + f_2) \leq \sup_{I_j} (f_1 + f_2) \leq \sup_{I_j} f_1 + \sup_{I_j} f_2
\]

multiplying by \( \Delta \alpha_j \) and summing yields

\[
L(f_1, \Pi, \alpha) + L(f_2, \Pi, \alpha) \leq L(f_1 + f_2, \Pi, \alpha) \leq U(f_1 + f_2, \Pi, \alpha) \leq U(f_1, \Pi, \alpha) + U(f_2, \Pi, \alpha).
\]

(9.4)

Since \( f_j \in \mathcal{R}(\alpha) \) for \( j = 1, 2 \), there are partitions \( \Pi_j \) such that \( U(f_j, \Pi_j, \alpha) - L(f_j, \Pi_j, \alpha) < \frac{\epsilon}{2} \). Letting \( \Pi^* = \Pi_1 \cup \Pi_2 \) as usual, we then have \( U(f_j, \Pi^*, \alpha) \leq U(f_j, \Pi_j, \alpha) \) and \( L(f_j, \Pi^*, \alpha) \geq L(f_j, \Pi_j, \alpha) \), so \( U(f_j, \Pi^*, \alpha) - L(f_j, \Pi^*, \alpha) < \frac{\epsilon}{2} \). Adding up and applying (9.4) (with \( \Pi^* \) in place of \( \Pi \)) yields

\[
U(f_1 + f_2, \Pi^*, \alpha) - L(f_1 + f_2, \Pi^*, \alpha) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\]

Hence \( f_1 + f_2 \in \mathcal{R}(\alpha) \). What’s more, since \( L(f_j, \Pi^*, \alpha) \leq \int f_j \, d\alpha \leq U(f_j, \Pi^*, \alpha) \), it follows that \( U(f_j, \Pi^*, \alpha) < \int f_j \, d\alpha + \frac{\epsilon}{2} \) and \( L(f_j, \Pi^*, \alpha) > \int f_j \, d\alpha - \frac{\epsilon}{2} \). Thus, applying (9.4) again, we have

\[
\int f_1 \, d\alpha + \int f_2 \, d\alpha - \epsilon < L(f_1 + f_2, \Pi^*, \alpha) \leq U(f_1 + f_2, \Pi^*, \alpha) < \int f_1 \, d\alpha + \int f_2 \, d\alpha + \epsilon.
\]

Taking sup and inf as appropriate shows that \( \int (f_1 + f_2) \, d\alpha \) is distance \( < \epsilon \) away from \( \int f_1 \, d\alpha + \int f_2 \, d\alpha \) for each \( \epsilon > 0 \), and this establishes the first equality.

For the second, we note that for \( c > 0 \), we simply have \( U(cf_1, \Pi, \alpha) = cU(f_1, \Pi, \alpha) \) and \( L(cf_1, \Pi, \alpha) = cL(f_1, \Pi, \alpha) \) for any partition \( \Pi \), and hence if we choose a \( \Pi \) so that \( U(f_1, \Pi, \alpha) - L(f_1, \Pi, \alpha) < \frac{\epsilon}{c} \), then \( U(cf_1, \Pi, \alpha) - L(cf_1, \Pi, \alpha) = c[U(f_1, \Pi, \alpha) - L(f_1, \Pi, \alpha)] < \epsilon \), so \( cf_1 \in \mathcal{R}(\alpha) \), and moreover \( \int cf_1 \, d\alpha = c \int f_1 \, d\alpha \) which is what we wanted. 

**Lemma 9.17.** If \( f_1 \leq f_2 \) on \([a, b]\) and \( f_1, f_2 \in \mathcal{R}(\alpha) \), then \( \int f_1 \, d\alpha \leq \int f_2 \, d\alpha \).

**Proof.** The inequality \( f_1(t) \leq f_2(t) \) implies that \( \sup_{I_j} f_1 \leq \sup_{I_j} f_2 \) for all intervals \( I_j \), and hence \( U(f_1, \Pi, \alpha) \leq U(f_2, \Pi, \alpha) \) for any partition \( \Pi \). Taking \( \inf_{\Pi} \) now yields \( \int f_1 \, d\alpha = U(f_1, \alpha) \leq U(f_2, \alpha) = \int f_2 \, d\alpha. \)
LEMMA 9.18. If \( a < c < b \) and \( f \in \mathcal{R}(\alpha) \) on \([a, b]\), then it is in \( \mathcal{R}(\alpha) \) on \([a, c]\) and on \([c, b]\), and
\[
\int_a^b f \, d\alpha = \int_a^c f \, d\alpha + \int_c^b f \, d\alpha.
\]

PROOF. Set \( f_1 = f\mathbf{1}_{[a,c]} \) and \( f_2 = f\mathbf{1}_{[c,b]} \). Then both \( f_1 \) and \( f_2 \) are in \( \mathcal{R}(\alpha) \). Indeed, if \( \Pi \) is a partition for which \( U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \epsilon \), then we may (possibly) refine it by taking \( \Pi^* = \Pi \cup \{c\} \) and noting that we still have \( U(f, \Pi^*, \alpha) - L(f, \Pi^*, \alpha) < \epsilon \). Let \( m \) be the index in \( \Pi^* = \{a = t_0 < t_1 < \cdots < t_n = b\} \) with \( t_m = c \), and let \( I_j = [t_{j-1}, t_j] \); then
\[
U(f_1, \Pi^*, \alpha) - L(f_1, \Pi^*, \alpha) = \sum_{j=1}^m [\sup_{I_j} f - \inf_{I_j} f] \Delta \alpha_j
\]
\[
= \sum_{j=1}^m [\sup_{I_j} f - \inf_{I_j} f] \Delta \alpha_j \leq U(f, \Pi^*, \alpha) - L(f, \Pi^*, \alpha) < \epsilon
\]
since the omitted terms are all \( \geq 0 \). A similar calculation shows that \( U(f_2, \Pi^*, \alpha) - L(f_2, \Pi^*, \alpha) \) is the sum of terms of index \( \geq m \) which is \( \leq \) the full sum, hence \( < \epsilon \). So both \( f_1, f_2 \) are in \( \mathcal{R}(\alpha) \), and by Lemma 9.16
\[
\int_a^b f \, d\alpha = \int_a^c f \, d\alpha + \int_c^b f \, d\alpha.
\]
It is left to the reader to establish (using very similar arguments) that \( \int_a^b f \, d\alpha = \int_a^c f \, d\alpha \) and \( \int_a^b f \, d\alpha = \int_a^c f \, d\alpha \), concluding the proof.

LEMMA 9.19. If \( f \in \mathcal{R}(\alpha_1) \) and \( f \in \mathcal{R}(\alpha_2) \) then \( f \in \mathcal{R}(\alpha_1 + \alpha_2) \) and \( f \in \mathcal{R}(c\alpha_1) \) for \( c > 0 \), with
\[
\int_a^b f \, d(\alpha_1 + \alpha_2) = \int_a^b f \, d\alpha_1 + \int_a^b f \, d\alpha_2 \quad and \quad \int_a^b f \, d(c\alpha_1) = c \int_a^b f \, d\alpha_1.
\]

PROOF. Here we simply note that
\[
\Delta(\alpha_1 + \alpha_2)_j = (\alpha_1 + \alpha_2)(t_j) - (\alpha_1 + \alpha_2)(t_{j-1})
\]
\[
= [\alpha_1(t_j) - \alpha_1(t_{j-1})] + [\alpha_2(t_j) - \alpha_2(t_{j-1})] = \Delta(\alpha_1)_j + \Delta(\alpha_2)_j,
\]
and similarly \( \Delta(c\alpha_1)_j = c \Delta(\alpha_1)_j \), for each \( j \). Since all these increments are \( \geq 0 \), it then follows that
\[
U(f, \Pi, \alpha_1 + \alpha_2) = U(f, \Pi, \alpha_1) + U(f, \Pi, \alpha_2) \quad and \quad U(f, \Pi, c\alpha_1) = c U(f, \Pi, \alpha_1).
\]
Taking \( \inf_{\Pi} \) yields we get \( U(f, \alpha_1 + \alpha_2) = U(f, \alpha_1) + U(f, \alpha_2) = \int f \, d\alpha_1 + \int f \, d\alpha_2 \) and \( U(f, c\alpha_1) = c U(f, \alpha_1) = c \int f \, d\alpha_1 \). Similar considerations with lower sums show that these two are equal to the given linear combinations, concluding the proof.

LEMMA 9.20. Let \( f \in \mathcal{R}(\alpha) \). Let \( M = \sup f \) and \( m = \inf f \), and suppose \( \phi : [m, M] \to \mathbb{R} \) is continuous. Then \( h = \phi \circ f \) is in \( \mathcal{R}(\alpha) \).

PROOF. Fix \( \epsilon > 0 \). Denote by \( C = \sup |\phi| \); we assume it is \( > 0 \), otherwise \( \phi = 0 \) and the statement of the lemma is silly; similarly, we assume \( \alpha(b) > \alpha(a) \). As \( \phi \) is uniformly continuous on \([m, M]\), there is some \( \delta > 0 \) so that \( |\phi(x) - \phi(y)| < \frac{\epsilon}{2|\alpha(b) - \alpha(a)|} \) whenever \( |x - y| < \delta \). For later convenience, we will make sure to select \( \delta \leq \frac{1}{4C} \).
Since \( f \in \mathcal{R}(\alpha) \), there is a partition \( \Pi \) of \([a, b]\) such that \( U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \delta^2 \). Now the difference \( U(h, \Pi, \alpha) - L(h, \Pi, \alpha) \) is equal to

\[
\sum_j \left[ \sup_{I_j} h - \inf_{I_j} h \right] \Delta \alpha_j.
\]

We break up the sum into two parts: \( j \in J_1 \sqcup J_2 \), where \( j \in J_1 \) iff \( \sup_{I_j} f - \inf_{I_j} f < \delta \), and \( j \in J_2 \) iff this difference is \( \geq \delta \). The choice of \( \delta \) shows that, for \( j \in J_1 \), \( \sup_{I_j} h - \inf_{I_j} h < \frac{\epsilon}{2[\alpha(b) - \alpha(a)]} \). For \( j \in J_2 \), on the other hand, the best we can say in general is that \( \sup_{I_j} h - \inf_{I_j} h \leq 2 \sup |h| = 2C \). But we’ve arranged things so that the total \( \alpha \)-length of the intervals indexed by \( J_2 \) is small: because \( \sup_{I_j} f - \inf_{I_j} f \geq \delta \) for \( j \in J_2 \),

\[
\delta \sum_{j \in J_2} \Delta \alpha_j \leq \sum_{j \in J_2} \left[ \sup_{I_j} f - \inf_{I_j} f \right] \Delta \alpha_j = U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \delta^2
\]

and so \( \sum_{j \in J_2} \Delta \alpha_j < \delta \). Thus, adding up,

\[
U(h, \Pi, \alpha) - L(h, \Pi, \alpha) = \sum_{j \in J_1} \left[ \sup_{I_j} h - \inf_{I_j} h \right] \Delta \alpha_j + \sum_{j \in J_2} \left[ \sup_{I_j} h - \inf_{I_j} h \right] \Delta \alpha_j
\]

\[
< \frac{\epsilon}{2[\alpha(b) - \alpha(a)]} \sum_{j \in J_1} \Delta \alpha_j + 2C \sum_{j \in J_2} \Delta \alpha_j
\]

\[
\leq \frac{\epsilon}{2[\alpha(b) - \alpha(a)]} \sum_j \Delta \alpha_j + 2C \delta
\]

\[
< \frac{\epsilon}{2} + 2C \cdot \frac{\epsilon}{4C} = \epsilon.
\]

This shows that \( h \in \mathcal{R}(\alpha) \), as desired. \( \square \)

**Remark 9.21.** The above proof shows no simple connection between the value of \( \int \phi \circ f \, d\alpha \) and \( \int f \, d\alpha \), and indeed there is no simple connection.

**Lemma 9.22.** If \( f, g \in \mathcal{R}(\alpha) \), then so is \( fg \).

**Proof.** By Lemma 9.16, \( f \pm g \) are both in \( \mathcal{R}(\alpha) \). The function \( \phi(x) = x^2 \) is continuous, and so by Lemma 9.20 \( (f + g)^2 \) and \( (f - g)^2 \) are both in \( \mathcal{R}(\alpha) \). Thus, by Lemma 9.16 again, so is

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

\( \square \)

**Lemma 9.23.** If \( f \in \mathcal{R} \), then \( |f| \in \mathcal{R} \), and \( \left| \int f \, d\alpha \right| \leq \int |f| \, d\alpha \).

**Proof.** The function \( \phi(x) = |x| \) is continuous, and so by Lemma 9.20 \( |f| \in \mathcal{R}(\alpha) \). Now, let \( \sigma = \text{sgn} \left( \int f \, d\alpha \right) \) (so \( \sigma = 1 \) if the integral is \( \geq 0 \) and \( \sigma = -1 \) if the integral is \( < 0 \)). Then by Lemma 9.16

\[
\left| \int f \, d\alpha \right| = \sigma \int f \, d\alpha = \int (\sigma f) \, d\alpha.
\]
For each $t$, we have $f(t) \leq |f(t)|$ and $-f(t) \leq |f(t)|$; thus $\sigma f \leq f$. Thus, by Lemma 9.17,

$$\left| \int f \, d\alpha \right| = \int (\sigma f) \, d\alpha \leq \int |f| \, d\alpha.$$

We now know how to integrate (or more precisely that we can integrate) a reasonably large class of functions, against arbitrary monotone increasing integrators. One may then reasonably ask: what benefit have we gained by including more general integrator weights \( \alpha \)? To partly answer this, consider the following example.

**Example 9.24.** Let \( a < b \) in \( \mathbb{R} \), and fix some \( s \in (a, b) \). Let \( \alpha_s \) be the monotone function \( \alpha_s(t) = 1_{(s,b)}(t) \), i.e. \( \alpha_s(t) = 0 \) if \( t \leq s \) and \( \alpha_s(t) = 1 \) if \( t > s \). If \( f \) is continuous on \([a, b]\) then

\[
\int_a^b f \, d\alpha_s = f(s).
\]

To see this, fix any partition \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \). Let \( m \geq 1 \) be the unique index such that \( t_{m-1} < s \leq t_m \). For any \( j \neq m \), \( \alpha_s \) is constant on \([t_{j-1}, t_j]\) and so \( \Delta \alpha_j = 0 \), while \( \Delta \alpha_m = 1 - 0 = 1 \). Thus

\[
U(f, \Pi, \alpha_s) = \sup_{t_{m-1} \leq t \leq t_m} f(t), \quad L(f, \Pi, \alpha_s) = \inf_{t_{m-1} \leq t \leq t_m} f(t).
\]

Since \( s \in [t_{m-1}, t_m] \), \( \inf_{t_{m-1} \leq t \leq t_m} f(t) \leq f(s) \leq \sup_{t_{m-1} \leq t \leq t_m} f(t) \). It follows that \( L(f, \alpha_s) \leq f(s) \leq U(f, \alpha_s) \). Since \( f \) is continuous, Theorem 9.11 implies that \( L(f, \alpha_s) = U(f, \alpha_s) \), and so this common value must be \( \int f \, d\alpha_s = f(s) \) as claimed. (A more careful argument shows that this holds true even if we only assume that \( f \) is continuous at \( s \).)

Example 9.24 gives a rigorous treatment of a “delta function”. Physicists love to use delta functions: a “function” \( \delta(t) \) with the property that, for \( s \in (a, b) \),

\[
\int_a^b f(t)\delta(t - s) \, dt = f(s).
\]

In fact, there is no such function \( \delta(t) \) which Riemann integrates a function by evaluating it at a point. Instead, \( \delta(t - s) \, dt \) must be interpreted as the Riemann–Stieltjes integrator \( d\alpha_s(t) \).

We can use the additivity of the integral to generalize Example 9.24, and put discrete infinite sums on the same footing as integrals, and treat them all as one kind of object. If \( (s_n)_{n=1}^\infty \) is an increasing (possibly finite) sequence in \((a, b)\), and if \( (c_n) \) is a nonnegative sequence such that \( \sum c_n < \infty \), we can define a pure step function

\[
\alpha(t) = \sum_{n=1}^\infty c_n \alpha_{s_n}(t).
\]

This is a monotone increasing function on \([a, b]\). It is constant on \([a, a + s_1]\), taking value \( c_1 \); its value on \((a + s_1, a + s_2]\) is \( c_1 + c_2 \); and so forth.

**Lemma 9.25.** Let \( \alpha \) be the pure step function above. If \( f \) is continuous on \([a, b]\), then

\[
\int_a^b f \, d\alpha = \sum_{n=1}^\infty c_n f(s_n).
\]

**Proof.** Since \( |\alpha_s(t)| \leq 1 \) for any \( s \), by the comparison test, the series \( \alpha(t) = \sum_{n=1}^\infty c_n \alpha_{s_n}(t) \) converges for all \( t \). The function \( \alpha \) defined is monotone increasing, since \( c_n \geq 0 \) for all \( n \), and \( \alpha_s(t) \leq \alpha_s(t') \) whenever \( t < t' \). Hence, it makes sense to compute \( \int_a^b f \, d\alpha \). (Note also that \( \alpha(a) = 0 \) while \( \alpha(b) = \sum c_n \).) Similarly, since \( f \) is bounded on \([a, b]\), the sum \( \sum_{n=1}^\infty c_n f(s_n) \) also converges by the comparison test, so all quantities presented make sense.
Now, let $\epsilon > 0$, and let $M = \sup |f|$ (and assume $M > 0$ to avoid silliness). By the Cauchy criterion for convergence of $\sum_{n} c_n$, there is some $N \in \mathbb{N}$ so that $\sum_{n=N} c_n < \frac{\epsilon}{M}$. We then break up $\alpha$ accordingly: $\alpha(t) = \alpha_1(t) + \alpha_2(t)$, where

$$\alpha_1(t) = \sum_{n=1}^{N} c_n \alpha_{s_n}(t), \quad \alpha_2(t) = \sum_{n=N+1}^{\infty} c_n \alpha_{s_n}(t).$$

By Lemma 9.19 we have

$$\int f \, d\alpha = \int f \, d\alpha_1 + \int d\alpha_2.$$

The first integral, by induction on Lemma 9.19 has the value

$$\int f \, d\alpha_1 = \sum_{n=1}^{N} c_n \int f \, d\alpha_s = \sum_{n=1}^{N} c_n f(s_n).$$

Thus, by Lemma 9.23 and Lemma 9.17,

$$|\int f \, d\alpha - \int f \, d\alpha_1| = |\int f \, d\alpha_2| \leq \int |f| \, d\alpha_2 \leq M \int d\alpha_2 = M[\alpha_2(b) - \alpha_2(a)].$$

Since $(s_n)$ is an increasing sequence, $s_{N+1} > a$, and so $\alpha_{s_n}(a) = 0$ for $n > N$; thus $\alpha_2(a) = 0$. On the other hand $\alpha_{s_n}(b) = 1$ for each $n$, and so $\alpha_2(b) = \sum_{n=N+1}^{\infty} c_n < \frac{\epsilon}{M}$ by constructions. Thus the above inequality shows that

$$|\int f \, d\alpha - \sum_{n=1}^{N} c_n f(s_n)| = |\int f \, d\alpha - \int f \, d\alpha_1| < \frac{\epsilon}{M} \cdot M = \epsilon.$$

Since we can do this for each $\epsilon > 0$, and since we know $\sum_{n=1}^{N} c_n f(s_n)$ converges as $N \to \infty$, it follows that it must converge to $\int f \, d\alpha$ as claimed. \qed

Thus, if we use a pure step function as our integrator, we unify the theory of infinite series and integration. On the flip side, what happens if $\alpha$ is the opposite of a purely discrete function: what if $\alpha$ is differentiable? In that case, we will see that integration with respect to $\alpha$ actually involved the derivative $\alpha'$ (and this will motivate the connection afterward to the Fundamental Theorem of Calculus). As you should be used to by now, anything involving derivatives will involve the Mean Value Theorem, which will mean evaluating the derivative of the integrand on $[t_{j-1}, t_j]$ at a (random) point $\eta_j \in (t_{j-1}, t_j)$. As such, it will be convenient in the next several results to connect our formulation of the integral, in terms of upper and lower sums, to Riemann(–Stieltjes) sums.

**Lemma 9.26.** Let $f$ be in $\mathcal{R}(\alpha)$ on $[a, b]$, let $\epsilon > 0$, and let $\Pi$ be a partition for which $U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \epsilon$ (cf. Lemma 9.10). With $\Pi = \{a = t_0 < t_1 < t_2 < \cdots < t_n = b\}$ and $I_j = [t_{j-1}, t_j]$, let $\eta_j \in I_j$. Then

$$\left| \sum_{j=1}^{n} f(\eta_j) \Delta \alpha_j - \int_{a}^{b} f \, d\alpha \right| < \epsilon.$$

**Proof.** By definition $\inf_{I_j} f \leq f(\eta_j) \leq \sup_{I_j} f$ for each $j$. Multiplying by the non-negative numbers $\Delta \alpha_j$ and summing up shows that

$$L(f, \Pi, \alpha) \leq \sum_{j=1}^{n} f(\eta_j) \Delta \alpha_j \leq U(f, \Pi, \alpha). \quad (9.5)$$
Of course, we also have
\[ L(f, \Pi, \alpha) \leq \int_a^b f \, d\alpha \leq U(f, \Pi, \alpha). \] (9.6)

So the Riemann–Stieltjes sum in question and the integral \( \int f \, d\alpha \) are both in the interval from \( L(f, \Pi, \alpha) \) up to \( U(f, \Pi, \alpha) \). By assumption this interval has width \( \epsilon \). Thus, the distance between the Riemann–Stieltjes sum and the integral is \( \epsilon \), as claimed.

**Theorem 9.27.** Let \( \alpha: [a, b] \to \mathbb{R} \) be monotone increasing, and suppose that \( \alpha \) is differentiable on \( [a, b] \) with \( \alpha' \in \mathcal{R} \). Then for any bounded \( f: [a, b] \to \mathbb{R} \), \( f \in \mathcal{R}(\alpha) \) if and only if \( f \alpha' \in \mathcal{R} \), and
\[
\int f \, d\alpha = \int f \alpha'.
\]

This theorem is often summarized by writing \( d\alpha(t) = \alpha'(t) \, dt \).

**Proof.** Fix \( \epsilon > 0 \), and let \( M = \sup |f| \) (assume \( M > 0 \) to avoid silliness). Since \( \alpha' \in \mathcal{R} \), there is a partition \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \) of \( [a, b] \) such that \( U(\alpha', \Pi) - L(\alpha', \Pi) < \frac{\epsilon}{3M} \).

By the Mean Value Theorem (applied to \( \alpha \)), for each \( j \) there is a point \( \xi_j \in (t_{j-1}, t_j) \) such that \( \Delta \alpha_j = \alpha(t_j) - \alpha(t_{j-1}) = \alpha' \alpha_j \). Now, fix any points \( \eta_j \in I_j \); then the Riemann–Stieltjes sum at the points \( \eta_j \) can nearly be expressed as a Riemann sum:
\[
\sum_{j=1}^n f(\eta_j) \Delta \alpha_j = \sum_{j=1}^n f(\eta_j) \alpha'(\xi_j) \Delta t_j.
\]

We would like to convert this to a proper Riemann sum at the points \( \eta_j \); to do so, we need to compare the points \( \eta_j \) to \( \xi_j \), which results in a correction factor
\[
\left| \sum_{j=1}^n f(\eta_j) \Delta \alpha_j - \sum_{j=1}^n f(\eta_j) \alpha'(\eta_j) \Delta t_j \right| = \left| \sum_{j=1}^n f(\eta_j) [\alpha'(\xi_j) - \alpha'(\eta_j)] \Delta t_j \right| 
\leq \sum_{j=1}^n |f(\eta_j)| |\alpha'(\xi_j) - \alpha'(\eta_j)| \Delta t_j
\leq M \sum_{j=1}^n |\alpha'(\xi_j) - \alpha'(\eta_j)| \Delta t_j.
\]

Now, since \( \xi_j, \eta_j \in I_j \), it follows that \( |\alpha'(\xi_j) - \alpha'(\eta_j)| \leq \sup I_j \alpha' - \inf I_j \alpha' \), and thus
\[
\sum_{j=1}^n |\alpha'(\xi_j) - \alpha'(\eta_j)| \Delta t_j \leq \sum_{j=1}^n \left[ \sup I_j \alpha' - \inf I_j \alpha' \right] \Delta t_j = U(\alpha', \Pi) - L(\alpha', \Pi) < \frac{\epsilon}{3M}.
\]

Thus, we see that, for any points \( \eta_j \in I_j \),
\[
\left| \sum_{j=1}^n f(\eta_j) \Delta \alpha_j - \sum_{j=1}^n f(\eta_j) \alpha'(\eta_j) \Delta t_j \right| < \frac{\epsilon}{3}, \quad (9.7)
\]

In particular, this shows that
\[
\sum_{j=1}^n f(\eta_j) \Delta \alpha_j < \sum_{j=1}^n (f \alpha')(\eta_j) \Delta t_j + \frac{\epsilon}{3} \leq \sum_{j=1}^n \sup I_j (f \eta') \Delta t_j + \frac{\epsilon}{3} = U(f \alpha', \Pi) + \frac{\epsilon}{3}.
\]
Writing out the definitions, we see that partitions is a bijection of partitions, \( \varphi \) is an inverse map \( f \) increasing, and let \( f \in \Pi \). Thus integrals are equal as claimed.

Thus, we have shown that, for any partition \( \Pi \) for which \( U(f') - L(f') < \frac{\epsilon}{3M} \), it follows that \( |U(f, \Pi, \alpha) - U(f', \Pi, \alpha)| < \frac{\epsilon}{3} \) and \( |L(f, \Pi, \alpha) - L(f', \Pi, \alpha)| < \frac{\epsilon}{3} \). If we find such a partition \( \Pi \), the antecedent will also hold true for any refinement of \( \Pi \).

Now, to conclude the proof: suppose that \( f \in \mathcal{R}(\alpha) \); then let \( \Pi' \) be a partition for which \( U(f, \Pi', \alpha) - L(f, \Pi', \alpha) < \frac{\epsilon}{3} \). If \( \Pi \) is a partition for which \( U(f', \Pi) - L(f', \Pi) < \frac{\epsilon}{3M} \), take \( \Pi^* = \Pi \cup \Pi' \); then we have \( U(f', \Pi^*) - L(f', \Pi^*) < \frac{\epsilon}{3} \). Thus \( |U(f, \Pi^*, \alpha) - U(f', \Pi^*)| < \frac{\epsilon}{3} \) and \( |L(f, \Pi^*, \alpha) - L(f', \Pi^*)| < \frac{\epsilon}{3} \). Since \( \Pi^* \) refines \( \Pi' \), it is also true that \( U(f, \Pi^*, \alpha) - L(f, \Pi^*, \alpha) < \frac{\epsilon}{3} \). Thus

\[
U(f, \Pi') - L(f, \Pi') < |U(f, \Pi') - U(f', \Pi')| + |U(f', \Pi') - U(f', \Pi^*)| + |L(f', \Pi^*) - L(f', \Pi^*)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\]

Thus \( f \in \mathcal{R} \) as well; and the fact that we can find a partition \( \Pi^* \) such that \( |U(f, \Pi^*, \alpha) - U(f', \Pi^*)| < \frac{\epsilon}{3} \) for each \( \epsilon > 0 \) shows that \( \int f d\alpha - \int f' d\alpha' < \frac{\epsilon}{3} \) for all \( \epsilon > 0 \), thus the two integrals are equal as claimed.

An entirely analogous argument beginning from the assumption that \( f \alpha' \in \mathcal{R} \) shows that \( f \in \mathcal{R}(\alpha) \) and that the two integrals are the same, concluding the proof. \( \square \)

This brings us to one of the most important tool for actually computing integrals.

**THEOREM 9.28 (The Change of Variables Formula).** Let \( a < b \) and \( c < d \) in \( \mathbb{R} \), and let \( \varphi: [c, d] \to [a, b] \) be a strictly increasing surjective function. Let \( \alpha: [a, b] \to \mathbb{R} \) be monotone increasing, and let \( f \in \mathcal{R}(\alpha) \). Define \( \beta, g: [c, d] \to \mathbb{R} \) by \( \beta = \alpha \circ \varphi \) and \( g = f \circ \varphi \). Then \( \beta \) is monotone increasing, \( g \in \mathcal{R}(\beta) \), and

\[
\int_{c}^{d} g \, d\beta = \int_{a}^{b} f \, d\alpha.
\]

**PROOF.** Since \( \varphi \) is strictly increasing, it is one-to-one; since it is surjective onto \([a, b] \), there is an inverse map \( \varphi^{-1}: [a, b] \to [c, d] \). This gives a bijection between partitions \( \Pi \) of \([a, b] \) and partitions \( \Theta \) of \([c, d] \); the correspondence is \( t \in \Theta \iff \varphi(t) \in \Pi \), and so we write \( \varphi(\Theta) = \Pi \). Writing out the definitions, we see that

\[
U(g, \Theta, \beta) = U(f, \Pi, \alpha) \quad \text{and} \quad L(g, \Theta, \beta) = L(f, \Pi, \alpha).
\]

Thus, finding a partition \( \Pi \) so that \( U(f, \Pi, \alpha) - L(f, \Pi, \alpha) < \epsilon \) is equivalent to finding a partition \( \Theta = \varphi^{-1}(\Pi) \) for which \( U(g, \Theta, \beta) - L(g, \Theta, \beta) < \epsilon \), and thus \( g \in \mathcal{R}(\beta) \). Because the map \( \Pi \to \Theta \) is a bijection of partitions, \( \int f \, d\alpha = \inf_{\Pi} U(f, \Pi, \alpha) = \inf_{\Theta} U(g, \Theta, \beta) = \int g \, d\beta \). \( \square \)
Theorem 9.28 may not look like the change of variables theorem you remember from calculus, but it is actually the generalization of it to the Riemann–Stieltjes integral. To restore the theorem you recall exactly, take the special case $\alpha(x) = x$. Then $\beta = \alpha \circ \varphi = \varphi$. If we further assume that $\varphi$ is differentiable and $\varphi' \in \mathcal{R}$, then Theorem 9.27 shows that $d\beta(x) = \varphi'(x) \, dx$, and so
\begin{equation}
\int_a^b f(u) \, du = \int_a^b f \, d\alpha = \int_c^d f \circ \varphi \, d\beta = \int_c^d f(\varphi(x)) \varphi'(x) \, dx \tag{9.8}
\end{equation}
which is the statement you learned in calculus.

Remark 9.29. In calculus, you may have stated this theorem without the requirement that $\varphi$ is strictly increasing. It is possible to generalize this, by reinterpreting the symbol $\int_a^b$ to include orientation. Suppose, for example, that $\varphi$ is strictly decreasing. Then $\beta = \alpha \circ \varphi$ is monotone decreasing, and therefore not the kind of integrator we know how to use. We could redo everything in this chapter so far for monotone decreasing integrators, and it would all work similarly, with appropriate minus signs thrown in. This can be accounted for by introducing the new (familiar from calculus) notation that if $a < b$ then $\int_a^b \equiv - \int_b^a$; this is what we mean by adding an orientation to the integral. With this in hand, everything works the same for monotone decreasing integrators, including Theorem 9.28. What’s more, by employing Lemma 9.18 suitably reinterpreted in terms of the new orientation concept, we can even handle the case that $\varphi$ is piecewise strictly monotone: there are finitely many points $a = x_0 < x_1 < x_2 < \cdots < x_p = b$ such that $\varphi$ is strictly monotone on each interval $(x_j-1, x_j)$. We can even allow $\varphi$ to be flat on some of the intervals (since the integral will just be 0 there). These are actually the kinds of functions for which the calculus change of variables formula works as stated above. Alternatively, one can restrict a little further to functions $\varphi \in C^1$ such that $\varphi'$ has only finitely many zeroes (this is true for all the usual functions studied in calculus).
We now come to the central result of calculus, appropriately called the Fundamental Theorem of Calculus: the Riemann integral is (more or less) the inverse of the derivative.

**Theorem 9.30 (Fundamental Theorem of Calculus).** Let $a < b$ in $\mathbb{R}$, and let $f \in \mathcal{R}$ on $[a, b]$.

(a) For $a \leq x \leq b$, define $F(x)$ by

$$F(x) = \int_a^x f(t) \, dt.$$ 

Then $F$ is Lipschitz continuous on $[a, b]$. Moreover, if $f$ is continuous at a point $x_0 \in [a, b]$, then $F$ is differentiable at $x_0$, with $F'(x_0) = f(x_0)$.

(b) If there exists a differentiable function $G : [a, b] \rightarrow \mathbb{R}$ such that $G' = f$, then $\int_a^b f(t) \, dt = G(b) - G(a)$.

**Proof.** For part (a), since $f \in \mathcal{R}$ it is bounded, say $|f| \leq M$. Then for any $x, y \in [a, b]$, say $x < y$; then by Lemma 9.18 and 9.23,

$$F(y) = \int_a^y f = \int_a^x f + \int_x^y f = F(x) + \int_x^y f$$

and so by Lemma 9.17

$$|F(y) - F(x)| = \left|\int_x^y f\right| \leq \int_x^y |f| \leq \int_x^y M = M(y - x).$$

This shows that $F$ is Lipschitz continuous, with Lipschitz constant $\leq M$. (In fact, the Lipschitz constant is precisely $\sup |f|$.)

Now, suppose $f$ is continuous at $x_0$. Fix $\epsilon > 0$, and choose $\delta > 0$ so that $|f(t) - f(x_0)| < \epsilon$ whenever $|t - x_0| < \delta$. Then for $x_0 \leq t < x_0 + \delta$, we have

$$\frac{F(t) - F(x_0)}{t - x_0} - f'(x_0) = \frac{1}{t - x_0} \int_{x_0}^t f - f'(x_0) = \frac{1}{t - x_0} \int_{x_0}^t [f - f(x_0)].$$

where we have used the fact that $\int_{x_0}^t f(x_0) = f(x_0)(t - x_0)$. Because $|t - x_0| < \delta, |f - f(x_0)| < \epsilon$ on $[x_0, t]$, and so

$$\left|\frac{F(t) - F(x_0)}{t - x_0} - f(x_0)\right| \leq \frac{1}{t - x_0} \int |f - f'(x_0)| \leq \frac{1}{t - x_0} \int_0^\epsilon \epsilon = \epsilon.$$

An analogous argument shows that the difference quotient $DQF(t, x_0)$ is distance less than $\epsilon$ from $f(x_0)$ in the case $x_0 - \delta < t < x_0$ as well. Thus, we have shown that $F'(x_0) = \lim_{t \to x_0} DQF(t) = f(x_0)$, as claimed.

For part (b), fix $\epsilon > 0$, and let $\Pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a partition for which $U(f, \Pi) - L(f, \Pi) < \epsilon$. By the Mean Value Theorem, for each $j$ there is a point $\xi_j \in (t_{j-1}, t_j)$ such that $G(t_j) - G(t_{j-1}) = F'(\xi_j) \Delta t_j = f(\xi_j) \Delta t_j$. Thus, reversing the telescoping sum,

$$G(b) - G(a) = \sum_{j=1}^n [G(t_j) - G(t_{j-1})] = \sum_{j=1}^n f(\xi_j) \Delta t_j.$$
By Lemma 9.26 by the choice of $\Pi$, we have
\[
\left| \int_a^b f - [G(b) - G(a)] \right| = \left| \int_a^b f - \sum_{j=1}^n f(\xi_j) \Delta t_j \right| < \epsilon.
\]
As this holds for every $\epsilon > 0$, the result follows. \hfill \square

In summary: if $f$ is continuous, then $F(x) = \int_a^x f(t) \, dt$ is an anti-derivative (a differentiable function with $F' = f$), and moreover if $G$ is any anti-derivative, then $\int_a^b f = G(b) - G(a)$. Any two antiderivatives differ by a constant: if $F - G$ is a differentiable function whose derivative is 0, which means it is constant by Corollary 8.19(2). So this is consistent: if $F = G + c$, then $F(b) - F(a) = G(b) - G(a)$.

In light of Theorem 9.30, the special case of the Change of Variables Formula in (9.8) is a straightforward consequence of the chain rule (under continuity assumptions). Indeed, let $F$ be continuous and suppose $\phi$ is differentiable with $\phi' \in \mathcal{R}$ on $[c,d]$. Let $F$ be an anti-derivative of $f$ (as in Theorem 9.30(a)), so that $F' = f$. Then $F \circ \phi$ is differentiable, and by the chain rule $(F \circ \phi)' = (F' \circ \phi) \phi' = (f \circ \phi) \phi'$. Employing the Fundamental Theorem of Calculus, we get
\[
\int_c^d f(\phi(x)) \phi'(x) \, dx = \int_c^d (F' \circ \phi)'(x) \, dx = F \circ \phi(d) - F \circ \phi(c).
\]
Note that this holds regardless of whether $\phi$ is increasing (or even piecewise monotone). So as long as $\phi(c) = a$ and $\phi(d) = b$, we then have
\[
\int_c^d f(\phi(x)) \phi'(x) \, dx = F(b) - F(a) = \int_a^b f(u) \, du
\]
again by the Fundamental Theorem of Calculus. This condition certainly holds if $\phi$ is strictly increasing from $[c,d]$ onto $[a,b]$, but this is not required; it could oscillate infinitely often as it fills out the interval. But this approach required the assumption that $f$ is continuous, rather than just Riemann integrable; if one wants more general Riemann integrable functions $f$, the previous approach (which required $\phi$ be at least piecewise monotone) is required.

In the same light, let us now use the Fundamental Theorem of Calculus to turn the product rule into a powerful computational (and theoretical) tool for Riemann integration.

**Theorem 9.31 (Integration by Parts).** Let $a < b$ in $\mathbb{R}$, and let $f, g : [a, b] \to \mathbb{R}$ be differentiable functions with $f', g' \in \mathcal{R}$. Then $fg'$ and $f'g$ are both in $\mathcal{R}$, and
\[
\int_a^b f(t)g'(t) \, dt = f(b)g(b) - f(a)g(a) - \int_a^b f'(t)g(t) \, dt.
\]

**Proof.** The chain rule gives $(fg)' = fg' + f'g$. By assumption $f'$ and $g'$ are both in $\mathcal{R}$, and so are $f$ and $g$ since they are differentiable (hence continuous). Thus, by Lemma 9.22, $fg'$ and $f'g$ are both in $\mathcal{R}$, and by the Fundamental Theorem of Calculus and 9.16
\[
f(b)g(b) - f(a)g(a) = \int_a^b (fg)'(t) \, dt = \int_a^b f(t)g'(t) \, dt + \int_a^b f'(t)g(t) \, dt.
\]
Subtracting yields the result. \hfill \square
As with the Change of Variables formula, we can wonder if there is a version for more general integrators. Indeed, suppose that \( g \) is differentiable, and \( g' \) is strictly increasing. Then Theorem 9.27 shows that \( g'(t) \, dt = dg(t) \), and so we can rewrite Integration by Parts in the form

\[
\int_a^b f \, dg = f(b)g(b) - f(a)g(a) - \int_a^b f'(t)g(t) \, dt.
\]

In fact, this formula holds true even if \( g \) is not differentiable, as you will prove on your homework.

To conclude this chapter on integration, we consider an application to curves.

**Definition 9.32.** Let \( a < b \) in \( \mathbb{R} \) and let \( d \in \mathbb{N} \). A (parametrized) curve in \( \mathbb{R}^d \), with parameter domain \([a, b]\), is a continuous function \( \gamma: [a, b] \to \mathbb{R}^d \). If \( \gamma(a) = \gamma(b) \), it is called a closed curve. If \( \gamma \) is one-to-one on \((a, b)\), it is called a simple curve.

Note: a curve is more than just a path traced out in space (which is the image \( \gamma([a, b]) \)); it also includes the information of the parametrization \( t \mapsto \gamma(t) \), which is usually thought of as specifying the position of a particle at time \( t \) as it moves through space.

Our present goal is to measure the **length** of a curve. To that end, we begin by approximating the curve by pieces whose lengths we know how to measure: line segments. Fix a partition \( \Pi = \{ a = t_0 < t_1 < \cdots < t_n = b \} \) of \([a, b]\). We replace \( \gamma \) by the path which passes through the points \( \gamma(t_0), \gamma(t_1), \gamma(t_2), \ldots, \gamma(t_n) \), and is a straight line segment between successive points. The length of each such line segment is just the Euclidean length of the difference vector: \( |\gamma(t_j) - \gamma(t_{j-1})| \).

Hence, we approximate the length of \( \gamma \) by

\[
\Lambda(\gamma, \Pi) = \sum_{j=1}^n |\gamma(t_j) - \gamma(t_{j-1})|.
\]

Note that if we refine the partition used, the triangle inequality makes this length shrink: for any point \( s \in (t_{j-1}, t_j) \), then

\[
|\gamma(t_j) - \gamma(t_{j-1})| \leq |\gamma(t_j) - \gamma(s)| + |\gamma(s) - \gamma(t_{j-1})|.
\]

It follows (by induction) that if \( \Pi^* \) is a refinement of \( \Pi \) then \( \Lambda(\gamma, \Pi) \leq \Lambda(\gamma, \Pi^*) \). Indeed, this matches up with the fact (known as the isoperimetric inequality) that the shortest path between two points is the straight-line path. Thus, if we want to take a “limit” making partitions finer and finer, we ought to define the length of the curve \( \gamma \) by

\[
\Lambda(\gamma) = \sup_{\Pi} \Lambda(\gamma, \Pi).
\]

This \( \sup \) may well be infinite. A first guess at such an example would be something like the graph of the function \( f(x) = \frac{1}{x} \), which has a vertical asymptote. But there is no way to parametrize this curve continuously on a closed interval, which is needed in our definition of curve. Nevertheless, there are continuous curves on closed intervals that have infinite length.

**Example 9.33.** Consider the curve \( \gamma: [0, 1] \to \mathbb{R}^2 \) which traces out the graph of the function \( f(x) = x \sin \frac{1}{x} \) (with \( f(0) = 0 \)) of Example 8.16

\[
\gamma(t) = (t, f(t)).
\]

Since \( f \) is continuous on \([0, 1]\), as is the identity function, it follows that \( \gamma \) is continuous on \([0, 1]\), and hence is a curve according to the above definition. Now, for each \( n \), consider the partition (with points written in the reverse of the usual order) \( \Pi_n = \{ 1 = t_0 > t_1 > \cdots > t_{n-1} > t_n = 0 \} \) where
for \(0 < j < n\), \(t_j = (\frac{n}{2} + (n - 1)\pi)^{-1}\). So \(\frac{1}{t_1} = \frac{1}{t_2} = \frac{3\pi}{2}\), and so forth through \(\frac{1}{t_{n-1}} = \frac{(2n-3)\pi}{2}\). At all of these points, \(\sin(\frac{1}{t_{j-1}}) = \pm 1\), with the sign changing from one term to the next. Hence, the lengths of adjacent increments (when neither \(j\) nor \(j - 1\) is 0 or \(n\)) are given by
\[
|\gamma(t_j) - \gamma(t_{j-1})| = |(t_j, \pm t_j) - (t_{j-1}, \mp t_{j-1})| = \sqrt{(t_j - t_{j-1})^2 + (t_j + t_{j-1})^2} > t_j + t_{j-1} > t_j.
\]
Now, \(t_j = \frac{2}{(2j-1)\pi} > \frac{1}{4(j-1)}\) for \(0 < j < n\), and so in the range \(2 \leq j \leq n - 1\) (where neither \(j\) nor \(j - 1\) is 0 or \(n\)), we have
\[
\begin{align*}
\Lambda(\gamma, \Pi_n) > \sum_{j=2}^{n-1} |\gamma(t_j) - \gamma(t_{j-1})| &> \frac{1}{4} \sum_{j=2}^{n-1} \frac{1}{j - 1} = \frac{1}{4} \sum_{k=1}^{n-2} \frac{1}{k}.
\end{align*}
\]
Since \(\sum_{k=1}^{\infty} \frac{1}{k} = +\infty\), it follows that \(\Lambda(\gamma) = \sup_\Pi \Lambda(\gamma, \Pi) = +\infty\).

So the length of the curve traced out by the graph of our favorite pathological continuous function is infinite.

We are interested in curves whose length is finite; these are called rectifiable. (The act of rectifying a curve is to “unravel” it into a straight line without stretching.) There is a large class of curves that are rectifiable, and with (nominally) computable lengths: \(C^1\) curves. To prove this, we first need to briefly extend the integral to curves (i.e. vector-valued functions of a real variable).

**Definition 9.34.** Let \(f : [a, b] \to \mathbb{R}^d\) be a function \(f = (f_1, \ldots, f_n)\) where each component \(f_j\) is Riemann integrable on \([a, b]\); we still denote this by \(f \in \mathcal{R}\). The integral \(\int_a^b f\) is the vector defined by componentwise integration:
\[
\int_a^b f \equiv \left(\int_a^b f_1, \ldots, \int_a^b f_d\right).
\]

As with derivatives, any theorem about integrals of scalar-valued functions extends immediately to vector-valued functions, so long as it applies separately to the components. For example: the Fundamental Theorem of Calculus still holds: if \(f \in \mathcal{R}\) and \(F' = f\) (meaning that \(F = (F_1, \ldots, F_n)\) is differentiable and \(F'_j = f_j\) for each \(j\)), then \(\int_a^b f = F(b) - F(a)\). One result about integrals that does not obviously carry over to the vector-valued case is Lemma [9.23](#) \(|\int f| \leq \int |f|\). If we try to apply this componentwise, we get
\[
|\int f| = \sqrt{\sum_{j=1}^n |\int f_j|^2} \leq \sqrt{\sum_{j=1}^n \left(\int |f_j|\right)^2}
\]
but this is not related in any clear way to
\[
\int |f| = \int \sqrt{\sum_{j=1}^n |f_j|^2}.
\]
Nonetheless, these two are comparable in precisely the same manner.

**Lemma 9.35.** Let \(f \in \mathcal{R}\) on \([a, b]\). Then \(|f| \in \mathcal{R}\) as well, and
\[
\int_a^b |f| \leq \int_a^b |f|.
\]
Proof. By Lemma 9.22, \( f_j^2 \in R \) for each \( j \), and then by induction on Lemma 9.22, \(|f| = f_1^2 + \cdots + f_d^2 \) is in \( R \). Since \( x \mapsto \sqrt{x} \) is continuous on \([0, \infty)\), it then follows from Lemma 9.20 that \(|f| = \sqrt{f_1^2 + \cdots + f_d^2} \) is in \( R \) as claimed. To prove the inequality, we take a hint from the proof of the vector Mean Value inequality (Theorem 8.33); let \( v = \int f \), and note that

\[
|v|^2 = v \cdot v = v \cdot \int f = \sum_{j=1}^d v_j \int f_j = \int \left( \sum_{j=1}^d v_j f_j \right) = \int v \cdot f
\]

where we used Lemma 9.16 in the penultimate equality. Applying the Cauchy-Schwarz inequality, we have \( v \cdot f \leq |v||f| \). Since we know that \(|f| \in R\), it then follows from Lemma 9.17 that \( \int v \cdot f \leq \int |v||f| = |v| \int |f| \). Thus \( |v|^2 \leq |v| \int |f| \); so either \( v = 0 \) (in which case \( |v| = 0 \leq \int |f| \)) or we can cancel one \(|v| > 0\) to find \(|v| \leq \int |f| \), which is the desired inequality.

Remark 9.36. We can define \( \int f \, dx \) for general increasing \( \alpha \) in the same manner, and the above proof shows that the inequality of Lemma 9.35 holds in this general setting; but we will not have occasion to use it beyond the standard Riemann integral case.

With these facts about integration of vector-valued functions at hand, we can now prove that \( C^1 \) curves are rectifiable.

Theorem 9.37. Let \( \gamma \) be a \( C^1 \) curve on \([a, b]\), meaning \( \gamma' \) is continuous on \([a, b]\). Then \( \gamma \) is rectifiable, and

\[
\Lambda(\gamma) = \int_a^b |\gamma'(t)| \, dt.
\]

Proof. Let \( \Pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \) be a partition. Applying the Fundamental Theorem of Calculus, we have

\[
\gamma(t_j) - \gamma(t_{j-1}) = \int_{t_{j-1}}^{t_j} \gamma'(t) \, dt
\]

and thus, applying Lemma 9.35, we have

\[
\Lambda(\gamma, \Pi) = \sum_{j=1}^n |\gamma(t_j) - \gamma(t_{j-1})| = \sum_{j=1}^n \left| \int_{t_{j-1}}^{t_j} \gamma'(t) \, dt \right| \leq \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |\gamma'(t)| \, dt = \int_a^b |\gamma'(t)| \, dt.
\]

(The last equality follows from Lemma 9.18, collapsing the telescoping sum.) Hence, \( \int_a^b |\gamma'(t)| \, dt \) is an upper bound for \( \Lambda(\gamma, \Pi) \) over all \( \Pi \), and thus \( \Lambda(\gamma) = \sup_\Pi \Lambda(\gamma, \Pi) \leq \int_a^b |\gamma'(t)| \, dt \). We are left only to prove the reverse inequality.

Fix \( \epsilon > 0 \). Since \( \gamma' \) is continuous on the compact interval \([a, b]\), it is uniformly continuous, and so there is some \( \delta > 0 \) so that \( |\gamma'(s) - \gamma'(t)| < \frac{\epsilon}{2(b-a)} \) whenever \( |s - t| < \delta \). Let \( \Pi \) be any partition with \( \Delta t_j < \delta \) for each \( j \). Then for any \( t \in [t_{j-1}, t_j] \), since \( |t - t_j| < \delta \), it follows that \( |\gamma'(t_j)| - |\gamma'(t)|| < \frac{\epsilon}{2(b-a)} \); in particular, \( |\gamma'(t_j)| \leq |\gamma'(t_j)| + \frac{\epsilon}{2(b-a)} \). Thus

\[
\int_{t_{j-1}}^{t_j} |\gamma'(t)| \, dt \leq \int_{t_{j-1}}^{t_j} \left( |\gamma'(t_j)| + \frac{\epsilon}{2(b-a)} \right) \, dt = (|\gamma'(t_j)| + \frac{\epsilon}{2(b-a)}) \Delta t_j.
\]
For the first term here, we make the following clever estimate:

\[
|\gamma'(t_j)|\Delta t_j = \left| \int_{t_{j-1}}^{t_j} \gamma'(t_j) \, dt \right| = \left| \int_{t_{j-1}}^{t_j} [\gamma'(t) + \gamma'(t_j) - \gamma'(t)] \, dt \right| \\
\leq \left| \int_{t_{j-1}}^{t_j} \gamma'(t) \, dt \right| + \left| \int_{t_{j-1}}^{t_j} [\gamma'(t_j) - \gamma'(t)] \, dt \right| \\
\leq |\gamma(t_j) - \gamma(t_{j-1})| + \frac{\epsilon}{2(b-a)} \Delta t_j
\]

where we’ve applied the Fundamental Theorem of Calculus again to the first term, and in the second term used Lemma 9.35 and then the fact that \(|\gamma'(t_j) - \gamma'(t)| < \frac{\epsilon}{2(b-a)}\) again. Combining this with (9.9) and summing (using Lemma 9.18) gives

\[
\int_a^b |\gamma'(t)| \, dt = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} |\gamma'(t)| \, dt \leq \sum_{j=1}^n \left( |\gamma(t_j) - \gamma(t_{j-1})| + 2 \frac{\epsilon}{2(b-a)} \Delta t_j \right) = \Lambda(\gamma, \Pi) + \epsilon
\]

Since this holds true for all \(\epsilon > 0\), it follows that \(\int_a^b |\gamma'(t)| \, dt \leq \Lambda(\gamma)\), as desired. \(\square\)

**Remark 9.38.** In the above proof, the continuity of \(\gamma'\) was not needed to show that \(\gamma\) is rectifiable; indeed, this holds whenever \(\gamma' \in \mathcal{R}\), and the inequality \(\Lambda(\gamma) \leq \int_a^b |\gamma'(t)| \, dt\) holds true. But this inequality may be strict if \(\gamma'\) is Riemann integrable but not continuous.