1 Some basic set theory

1.1 Sets, ordinals, natural numbers, and the Peano Axioms [1]

- Definition. Informally, a set is any collection of objects. Formally, sets are defined via the Zermelo–Fraenkel axioms of set theory. This is usually denoted ZFC or ZF, depending on whether one is assuming/using the Axiom of Choice (treated in its own section) or not necessarily assuming/using it. In short, these axioms lay the ground rules for working with sets, they tell us which sets must exist, and they tell us how sets can be constructed from other sets. We will not list all of these axioms in detail.\footnote{This section more or less covered in §1 of J&P.}

- One important feature of ZF set theory is that it is formulated in such a way to avoid Russell’s paradox. Summary: Before set theory was formalized, Bertrand Russell wrote down the following paradoxical “set”:

\[ R = \{ x : x \notin x \} \]

The problem with this set is that \( R \in R \Leftrightarrow R \notin R \), so \( R \) should not be a set. ZF set theory avoids Russell’s paradox by only allowing a set above to be built as a subset of another set. That is, given a set \( y \), \( R' = \{ x \in y : x \notin x \} \) is allowed, but \( R = \{ x : x \notin x \} \) is not allowed. This is called restricted comprehension.

- ZF set theory explicitly asserts the existence of \( \emptyset \), the empty set, and it also explicitly asserts the existence of an infinite set.

- Under the ZF axioms, for any set \( x \), there is another set \( S(x) = x \cup \{x\} \) called its successor. Applying \( S \) repeatedly to \( \emptyset \) yields the finite ordinals:

\[
\begin{align*}
0 &= \emptyset \\
1 &= S(0) = 0 \cup \{0\} = \{0\} \\
2 &= S(1) = 1 \cup \{1\} = \{0, 1\} \\
3 &= S(2) = 2 \cup \{2\} = \{0, 1, 2\} \\
&\vdots
\end{align*}
\]

and ZF set theory guarantees \( \omega = \{ 0, 1, 2, 3, 4, \ldots \} \) is a valid set. This is called the set of finite ordinals, or natural numbers, and it is more commonly (in non-set theory settings) denoted \( \mathbb{N} \).

- The Peano axioms, which are formulated separately from ZF set theory, are axioms about the natural numbers\footnote{Similar material covered in §6 of J&P.} (equipped with the unary successor operator \( S : x \mapsto x + 1 \)).
– (Zero is not a successor.) For all \( x \in \mathbb{N} \), \( S(x) \neq 0 \).
– (Successor is injective.) For all \( x, y \in \mathbb{N} \), if \( S(x) = S(y) \) then \( x = y \).
– (Induction.) If \( A \subseteq \mathbb{N} \), \( 0 \in A \), and for all \( x \in A \) we have \( S(x) \in A \), then \( A = \mathbb{N} \).

• The induction axiom above is equivalent to the well-ordering principle for \( \mathbb{N} \): Every nonempty subset of \( \mathbb{N} \) has a least element. This also implies that any nonempty subset of \( \mathbb{Z} \) bounded below (resp. above) has a least (resp. greatest) element.

• Arithmetic in \( \mathbb{N} \).

– We can define \(+\) in \( \mathbb{N} \) inductively by \( x + 0 = x \) and \( x + S(y) = S(x + y) \).
– We can define \( \cdot \) in \( \mathbb{N} \) inductively by \( x \cdot 0 = 0 \) and \( x \cdot S(y) = x \cdot y + x \).
– We can define \( \leq \) in \( \mathbb{N} \) by \( x \leq y \) if there exists \( z \) with \( y = x + z \).

• One can prove the usual arithmetic properties of addition, multiplication, and order using the Peano axioms, but at great cost (of extreme tedium).

1.2 Cartesian products, functions, and relations [1, 2]

• Definition. If \( X \) and \( Y \) are sets, we can build a set \( X \times Y \) consisting of all ordered pairs \((x, y)\) where \( x \in X \) and \( y \in Y \). In the language of sets, the ordered pair \((x, y)\) might formally be identified with the set \( \{\{x\}, \{x, y\}\} \).\(^2\) The set \( X \times Y \) is called the Cartesian product (or just product) of \( X \) and \( Y \).\(^3\)

• Definition. A function \( f : X \to Y \) is a subset \( f \subseteq X \times Y \) with the following property: For all \( x \in X \) there exists exactly one \( y \) such that \((x, y) \in f \).\(^4\) Here \( X \) is called the domain and \( Y \) is called the codomain.\(^5\) We usually write \( f(x) = y \) instead of \((x, y) \in f \).

If \( U \subseteq X \), then \( f(U) = \{ f(x) : x \in U \} \) is a subset of \( Y \) called the image of \( U \) under \( f \). If \( V \subseteq Y \), then \( f^{-1}(V) = \{ x \in X : f(x) \in V \} \) is a subset of \( X \) called the preimage (or inverse image) of \( V \) under \( f \).\(^6\)

• Proofs in analysis (particularly when we get to topological analysis) often involve manipulating preimages in possibly confusing ways. J&P have collated a number of useful facts about how preimages interact with various set-theoretic operations in Theorem 2.6 (beware their nonstandard set-complement notation).

• Definition.\(^7\) Let \( f : X \to Y \) be a function.

– \( f \) is called injective (or an injection, or one-to-one) if for all \( x, x' \in X \), \( f(x) = f(x') \) only if \( x = x' \).

\(^2\)J&P Definition 2.1.
\(^3\)J&P Definition 2.3.
\(^4\)J&P Definition 2.4.
\(^5\)J&P Definition 2.5.
\(^6\)J&P Definition 2.5.
\(^7\)J&P Definition 2.5.
– $f$ is called surjective (or a surjection, or onto) if $f(X) = Y$. That is, if for every $y \in Y$ there exists $x \in X$ satisfying $f(x) = y$.
– $f$ is called bijective (or a bijection, or one-to-one and onto) if it is both injective and surjective.
– $f : X \to Y$ is bijective iff there is an inverse function $f^{-1} : Y \to X$ satisfying $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.

**Definition.** Let $X$ be a set. A relation on $X$ is any subset $R \subseteq X \times X$. We usually write $x_1 R x_2$ instead of $(x_1, x_2) \in R$ for relations (and $R$ is usually a symbol like $=$ or $\leq$ or $\sim$, etc.).

**Definition.** Let $X$ be a set and let $R$ be a relation on $X$.
– $R$ is called reflexive if $x R x$ for all $x \in X$.
– $R$ is called symmetric if $x R y$ guarantees $y R x$ for all $x, y \in X$.
– $R$ is called transitive if $x R y$ and $y R z$ guarantee $x R z$ for all $x, y, z \in X$.
– $R$ is called antisymmetric if $x R y$ and $y R x$ imply $x = y$ for all $x, y \in X$.
– $R$ is called comparable (or comprehensive, or dichotomous) if for all $x, y \in X$ either $x R y$ or $y R x$ (or both).

**Definition.** Let $X$ be a set and let $R$ be a relation on $X$.
– $R$ is called an equivalence relation if it is reflexive, symmetric, and transitive.
– $R$ is called a partial order if it is transitive and antisymmetric.
– $R$ is called a total order if it is transitive, antisymmetric, and comparable.

**Definition.** Let $X$ be a set and let $R$ be an equivalence relation on $X$.
For any $x \in X$, the $(R)$-equivalence class of $x$, denoted $[x]_R$, is the set $\{ y \in X : y R x \}$. Let $X/R = \{ [x]_R : x \in X \}$. Then $X = \bigcup_{E \in X/R} E$, and $E \cap E' = \emptyset$ for distinct $E, E' \in X/R$. That is, $X/R$ is a partition of $X$ (a way of breaking up the set into non-overlapping pieces). (Conversely, every partition of $X$ yields an equivalence relation.) The set $X/R$ is called the quotient of $X$ by $R$.

1.3 The Axiom of Choice

• The Axiom of Choice (AC) is an axiom usually included in Zermelo–Fraenkel set theory which asserts that for every set $\mathcal{F}$ of nonempty sets, one can create a new set $S_\mathcal{F}$ by choosing exactly one element from each member of $\mathcal{F}$.

• The trouble with AC is that it is non-constructive: While it asserts the existence of a particular set with certain properties, it does not give a method for constructing such a set. While most mathematicians accept and use AC to various degrees, it is good practice to point out when it is used in a proof (and to favor proofs that avoid it when available), as several results or examples that rely on it are considered intuitionally challenging or “pathological.”
• There are a number of weaker versions of AC. For example, the *Axiom of Countable Choice* restricts the axiom to countable $\mathcal{F}$. The *Axiom of Dependent Choice* is a more complicated weakening of AC, but it is sufficient to develop most of real analysis (surprisingly, ADC is equivalent to the Baire category theorem).

• As mentioned already, you encounter the abbreviations ZFC and ZF, which stand for Zermelo–Fraenkel set theory assuming AC or not assuming AC, respectively (note that ZF does not assume the *negation* of AC either).

1.4 Cardinality [11]

• **Definition.** The *cardinality* of a set $S$, denoted $|S|$ is, roughly speaking, the number of elements in $S$. This is straightforward for finite sets, but for infinite sets the situation is a bit more complicated. Let $A$ and $B$ be sets.

  - We say that $|A| = |B|$ if there is a bijective function $A \rightarrow B$.§J&P Definition 8.1.
  - We say $|A| \leq |B|$ if there is an injective function $A \rightarrow B$.
  - We say that $|A| < |B|$ if $|A| \leq |B|$ and $|A| \neq |B|$.

• Some notes:

  - It is easy to check that equality of cardinals as defined above is an equivalence relation (bijections are invertible and a composition of two bijections is again a bijection). This does not require AC.
  - On the other hand, we would like $\leq$ to be an ordering on cardinals. Transitivity does not require AC. Antisymmetry is called the *Schröder–Bernstein theorem*—this does not require AC. Comparability/dichotomy for $\leq$ among cardinals is actually equivalent to AC.
  - Suppose we define $|A| \leq^* |B|$ if either $A = \emptyset$ or there is a surjective function $B \rightarrow A$. The assertion “$|A| \leq |B|$ iff $|A| \leq^* |B|$” is called the *partition principle*. ZFC is sufficient to prove the partition principle, but it is an open problem if AC is necessary to prove it.

• **Definition.** Formally speaking, a *cardinal* is an equivalence class of sets under the cardinal equality relation. The finite cardinals correspond to $\mathbb{N}$ in the obvious way. The cardinality of $\mathbb{N}$ is called *countable infinity*, *omega* (denoted $\omega$), or *aleph-0* ($\aleph_0$).

• **Cardinal arithmetic.** Let $A$ and $B$ be sets. We define $|A| + |B|$, $|A| \cdot |B|$, and $|B|^{|A|}$ as follows:

  - $|A| + |B| = |A \sqcup B|$ where $A \sqcup B$ is the *disjoint union* of $A$ and $B$ (this can be defined by $A \sqcup B = \{(a,0) : a \in A\} \cup \{(b,1) : b \in B\}$—i.e., the union of $A$ and $B$ but you differentiate elements of $A$ and $B$, ignoring any overlap).
  - $|A| \cdot |B| = |A \times B|$ where $A \times B$ is the Cartesian product of $A$ and $B$.
- $|B|^{|A|}$ is the cardinality of $|^A B|$ where $^A B$ is the set of all functions $A \to B$.
- In particular, $2^{|A|} = ^A 2$ is the set of all functions $A \to 2 = \{0, 1\}$. Since defining such a function is equivalent to a choice of $S \subseteq A$, $2^{|A|} = |\mathcal{P}(A)|$ where $\mathcal{P}(A)$ is the power set of $A$.

You should convince yourself that these work as expected for finite cardinals.

- (AC) If one or both cardinals are infinite, $|A| + |B| = |A| \cdot |B| = \max\{|A|, |B|\}$.
- Cardinal addition, multiplication, and exponentiation obey expected (and useful) rules. For example, $(|C|^{|B|})^{|A|} = |C|^{|A||B|}$ (there’s a simple bijection $^A x^B C \to ^A (B C)$, see if you can figure it out!).

• **Definition.** A set $S$ is called **countable** if $|S| \leq \aleph_0$. $S$ is called **countably infinite** if $|S| = \aleph_0$.§J&P Definition 8.2.

• A more useful definition in practice: $S$ is countable if $S = \emptyset$ or there is a surjection $\mathbb{N} \to S$.§J&P Definition 9.3. That is, a nonempty set $S$ is countable if there is a sequence $(x_n)_{n=0}^\infty$ in $S$ such that for all $s \in S$ there exists $m \in \mathbb{N}$ such that $s = x_m$. (In general, that this is equivalent to the definition above requires a weak version of AC.) Such a sequence is called an enumeration of $S$.

• **Examples of countable sets:**
  - All finite sets are countable.
  - The family of countable sets is closed under (countable) unions,§J&P Theorem 9.5 (arbitrary) intersections, (finite) Cartesian products, and set quotients.
  - $\mathbb{N}$ is countable.
  - $\mathbb{Z}$ is countable, since it is the quotient of $\mathbb{N} \times \mathbb{N}$ by an equivalence relation.
  - $\mathbb{Q}$ is countable,§J&P Corollary 9.7. since it is the quotient of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by an equivalence relation.

• As you showed in the homework, there is no surjection $\mathbb{N} \to \mathcal{P}(\mathbb{N})$, so $\mathcal{P}(\mathbb{N})$ is uncountable, with cardinality $2^{\aleph_0}$, which is also called continuum and denoted $\mathfrak{c}$ (Gothic letter $c$). Furthermore, this means there are infinitely many distinct infinite cardinals: $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < \cdots$

2 The real numbers

2.1 Constructions of $\mathbb{Z}$ and $\mathbb{Q}$

- We are given the natural number system $\mathbb{N}$ from set theory (along with the Peano axioms for arithmetic). As a number system, $\mathbb{N}$ is useful because its elements can be used to **count** objects. We also would like to **measure** objects/quantities, which is why we need to construct more complicated number systems (eventually arriving at $\mathbb{R}$).
Construction of $\mathbb{Z}$. One obvious deficiency of $\mathbb{N}$ is that subtraction only works “sometimes.” To solve this issue, we can construct the larger number system of the integers as follows: Let $\sim$ be the equivalence relation on $\mathbb{N} \times \mathbb{N}$ given by $(a, b) \sim (c, d)$ if $a + d = b + c$ (in $\mathbb{Z}$ this relation is true precisely when $a - b = c - d$). Then $\mathbb{Z}$ can be realized as the quotient of $\mathbb{N} \times \mathbb{N}/ \sim$, with $[(a, b)]$ corresponding to the integer $a - b$.

- The natural numbers are included in $\mathbb{Z}$ by $n \mapsto [(n, 0)]$.
- Addition is given by $[(a, b)] + [(c, d)] = [(a + c, b + d)]$.
- Negation is given by $-[(a, b)] = [(b, a)]$.
- Multiplication is given by $[(a, b)] \cdot [(c, d)] = [(ac + bd, ad + bc)]$.

$\mathbb{Z}$ satisfies the axioms of an ordered ring (with identity).

Construction of $\mathbb{Q}$. Just as subtraction is only partially defined in $\mathbb{N}$, division is only partially defined in $\mathbb{Z}$ (which is why number theorists like me have jobs). We can construct the larger system of rational numbers as follows: Let $\sim$ be the equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ given by $(a, b) \sim (c, d)$ if $ad = bc$. Then $\mathbb{Q}$ can be realized as the quotient of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by $\sim$ with $[(a, b)]$ corresponding to the fraction $a/b$.

- The integers are included in $\mathbb{Q}$ by $n \mapsto [(n, 1)]$.
- Addition is given by $[(a, b)] + [(c, d)] = [(ad + bc, bd)]$.
- Negation is given by $-[(a, b)] = [(-a, b)]$.
- Inversion is given by $[(a, b)]^{-1} = [(b, a)]$ (provided $b \neq 0$).
- Multiplication is given by $[(a, b)] \cdot [(c, d)] = [(ac, bd)]$.

$\mathbb{Q}$ satisfies the axioms of an ordered field.

The construction of $\mathbb{Q}$ from $\mathbb{Z}$ is a general one in abstract algebra: The formation of the field of fractions (or quotient field) of an integral domain, or more generally, the localization of a ring.

2.2 The ordered field axioms

- A field is an algebraic structure consisting of a set $F$, two constants $0, 1 \in F$, and two binary operations $+, \cdot : F \times F \to F$, satisfying the following axioms:$§J&P §3.$

  - (Non-triviality.) $0 \neq 1$.
  - (Addition is associative.) For all $x, y, z \in F$, $(x + y) + z = x + (y + z)$.
  - (Addition is commutative.) For all $x, y \in F$, $x + y = y + x$.
  - (0 is the identity for addition.) For all $x \in F$, $x + 0 = x$.
  - (All elements have additive inverses.) For all $x \in F$ there exists $y \in F$ such that $x + y = 0$. 

6
- (Multiplication is associative.) For all \( x, y, z \in F \), \((xy)z = x(yz)\).
- (Multiplication is commutative.) For all \( x, y \in F \), \( xy = yx \).
- (1 is the identity for multiplication.) For all \( x \in F \), \( x \cdot 1 = x \).
- (Nonzero elements have multiplicative inverses.) For all \( x \in F \), if \( x \neq 0 \), then there exists \( y \in F \) such that \( xy = 1 \).
- (Multiplication distributes over addition.) For all \( x, y, z \in F \), \( x(y + z) = xy + xz \).

\[ \text{• } \mathbb{Q} \text{ is a field, as are } \mathbb{R} \text{ and } \mathbb{C}. \text{ We may encounter more exotic fields like } \mathbb{Q}_p \text{ in this class.} \]

\[ \text{• A ring (with identity) is an algebraic system that satisfies these axioms except for non-triviality (the trivial ring is the ring with one element, which is both } 0 \text{ and } 1) \text{ and the existence of multiplicative inverses. Every field is a ring, but not vice versa: } \mathbb{Z} \text{ is a ring but not a field, because only } 1 \text{ and } -1 \text{ have multiplicative inverses.} \]

\[ \text{• An ordered field is a field } F \text{ which has a total ordering relation } \leq \text{ that is compatible with arithmetic in } F \text{ as follows:} \]

\[ \begin{align*}
\text{If } x, y, z \in F, \text{ then } & x \leq y \text{ implies } x + z \leq y + z. \\
\text{If } x, y, z \in F \text{ and } z \geq 0, \text{ then } & x \leq y \text{ implies } xz \leq yz. \\
\end{align*} \]

\[ \text{• J&P use a different (but equivalent) definition for an ordered field: An ordered field is a field } F \text{ with a specified subset } P \text{ of positive elements which must satisfy} \]

\[ \begin{align*}
\text{For all } x, y & \in P, \text{ and } x + y \in P. \\
\text{For all } x & \in F, \text{ exactly one of } x \in P, \text{ } x = 0, \text{ or } -x \in P \text{ is true.} \\
\end{align*} \]

The order on the field is then defined by \( x \leq y \) when either \( x = y \) or \( y - x \in P \).\[ \text{§J&P Axiom 12, §4.} \]

\[ \text{• Every ordered field comes pre-equipped with an absolute value: } |\cdot| : F \to F, \]

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

One can prove\(^8\) that this absolute value has the following properties:

\[ \begin{align*}
\text{(Positivity.) } & |x| \geq 0 \text{ for all } x \in F, \\
\text{(Multiplicativity.) } & |xy| = |x||y| \text{ for all } x, y \in F, \\
\text{(Triangle inequality.) } & |x + y| \leq |x| + |y| \text{ for all } x, y \in F. \\
\end{align*} \]

In (subfields of) the real numbers, this absolute value induces a metric \( d(x, y) = |x - y| \).

\(^8\)J&P Theorem 4.5.
2.3 The least upper bound axiom

- It is much more difficult to transition from \( \mathbb{Q} \) to \( \mathbb{R} \), since the deficiencies of \( \mathbb{Q} \) for analysis are somewhat trickier to characterize precisely. We know that there are numbers that “should be there.” For example, we need \( \sqrt{2} \) to measure the hypotenuse of the right isosceles triangle, but \( \sqrt{2} \notin \mathbb{Q} \).

- Here is how to get a handle on how \( \sqrt{2} \) can be realized in terms only of rational numbers. We can partition \( \mathbb{Q} \) into two halves: \( A \cup B \) where \( A \) consists of rational numbers that are too small to be \( \sqrt{2} \) (they are negative or their square is < 2) and \( B \) consists of rational numbers that are too big to be \( \sqrt{2} \) (they are positive and their square is > 2). What we want from the real numbers is a number that sits between \( A \) and \( B \). In short, the issue with \( \mathbb{Q} \) is that it does not form a continuum of points—there are gaps in the rational numbers, when viewed as a “number line.”

- To fill these gaps we consider a new number system \( \mathbb{R} \) which is an ordered field that satisfies an additional axiom called the Least Upper Bound Axiom (or Dedekind Completeness Axiom): Every nonempty subset \( S \subseteq \mathbb{R} \) that is bounded above has a least upper bound or supremum (in \( \mathbb{R} \)).§J&P §5.

- Definition. If \( S \) is a set and \( M \in \mathbb{R} \), we say that \( M = \text{sup}(S) \) if:
  - \( M \) is an upper bound on \( S \): \( \forall x \in S : x \leq M \); and
  - If \( M' < M \), then \( M' \) is not an upper bound on \( S \): \( \forall \epsilon > 0, \exists y \in S : y > M - \epsilon \).
  We can extend the sup operator to all subsets of \( \mathbb{R} \) by taking \( \text{sup}(S) = +\infty \) if \( S \) is not bounded above, and \( \text{sup}(\emptyset) = -\infty \) (as all real numbers are an upper bound on \( \emptyset \)).

- Greatest lower bounds, or infima, are defined symmetrically. The existence of least upper bounds for nonempty sets bounded above is equivalent to the existence of greatest lower bounds for nonempty sets bounded below.\(^9\)

- The approach to analysis where we axiomatize \( \mathbb{R} \) but do not construct it explicitly is called the synthetic approach. It is a convenient approach because \( \mathbb{R} \) is rather annoying to construct directly. In the next section we will give some ideas of how this is done, but prove basically nothing.

2.4 Constructions of \( \mathbb{R} \)

We present four different ways of constructing the real numbers.

- As base-\( n \) expansions. Let \( n \geq 2 \) be an integer. A real number is a pair \( (N, (s_k)_k) \) where \( N \in \mathbb{Z} \) and \( (s_k)_{k=1}^\infty \) is a sequence over the set \( \{0, 1, 2, \ldots, n - 1\} \). This pair is identified with the real number
  \[
  N + \sum_{k=1}^\infty s_k \cdot n^{-k} = N.s_1s_2s_3s_4s_5\ldots
  \]

\(^9\)J&P Theorem 5.4.
This construction is the naïve grade school conception of the real numbers, but it is not particularly attractive from a mathematical point of view. Why? Try to define addition for such sequences (remembering that generally speaking they are infinite!) and get back to me. The only real advantage here is that \( \leq \) is easy to define.

• **As Dedekind cuts.** A *Dedekind cut* is a subset \( A \) of \( \mathbb{Q} \) such that
  
  - For all \( x \in A \) and all \( y \in A^c \), \( x < y \), and
  - \( A^c \) does not have a smallest element.

  The cut \( A \) is identified with the real number \( \sup A \), and if \( r \in \mathbb{R} \), the corresponding cut is \( \{ q \in \mathbb{Q} : q \leq r \} \). (There are many equivalent definitions of Dedekind cuts—in our earlier discussion of \( \sqrt{2} \), the partition \( \mathbb{Q} = A \cup B \) was essentially a Dedekind cut.)

  Advantages: Every real number corresponds precisely to one Dedekind cut and vice-versa (no equivalence relation is required). The order is easily defined, and the completeness axiom is easy to verify: If \( S \) is a nonempty set of Dedekind cuts bounded above, then \( \bigcup S \) is the Dedekind cut corresponding to \( \sup S \).

  Disadvantages: Arithmetic with Dedekind cuts is cumbersome at best.

• **As equivalence classes of Cauchy sequences.** This was covered in the homework.

  Advantages: Arithmetic is fairly easy to define and the field axioms are easy to check. A similar construction is also used to construct the completion of a general metric space.

  Disadvantages: Every equivalence class of rational Cauchy sequences is uncountably infinite. The least upper bound axiom is a pain in the butt to prove.

• **As equivalence classes of almost-additive functions** \( \mathbb{Z} \to \mathbb{Z} \). This was covered in the homework. (If you want more information, these are called the *Eudoxus reals*.)

  Advantages: Does not actually require one to construct \( \mathbb{Q} \) first! The construction only uses \( \mathbb{Z} \). Arithmetic is easy to define (though the axioms are not easy to check). Completeness is somewhat easier than with Cauchy sequences(?).

  Disadvantages: Field axioms are tedious to prove, and completeness is not much better.

All in all, it’s a mixed bag. I’d say Dedekind cuts give the most geometrically compelling picture of how \( \mathbb{R} \) arises from \( \mathbb{Q} \) (a real number is a way of chopping the rational line in half!), while the Cauchy sequences give the most algebraically compelling picture (a real number is something that can be arbitrarily well-approximated by rational numbers).

### 2.5 Basic consequences of the ordered field axioms

These include basic laws of arithmetic like \( xy = 0 \) only if \( x = 0 \) or \( y = 0 \); \( 0 < 1 \); etc. These can be proven easily on your own or they are in our standard references.\(^\text{10}\)

\(^\text{10}\)J&P §3 and exercises, §4 and exercises.
2.6 Relationship between $\mathbb{R}$ and its subsystems $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$

Every ordered field contains $\mathbb{N} = \{0, 1, 2, \ldots\}$, so because fields are closed under negation and inversion, every ordered field contains (copies of) $\mathbb{Z}$ and $\mathbb{Q}$ as well. The relationship between $\mathbb{R}$ and these other number systems is important to the development of analysis.

- **Theorem.** (Archimedean Property of $\mathbb{R}$.) Let $a, b \in \mathbb{R}$ with $a, b > 0$. There exists $n \in \mathbb{N}$ such that $na > b$.$^{11}$
  
  *Proof.* Let $S = \{ k \in \mathbb{N} : ka \leq b \}$. $S$ is nonempty because $0 \in S$, and $S$ is bounded above by $b/a$. Hence, $S$ has a least upper bound, $m = \sup S$. Since $m - 1$ is not an upper bound, there exists $k \in S$ such that $k > m - 1$, from which we obtain $k + 1 > m$. Because $m$ is an upper bound on $S$ it follows that $k + 1 \notin S$, so $(k + 1)a > b$, and therefore $n = k + 1$ satisfies our requirements. $\Box$

- **Lemma.** Suppose that $x, y \in \mathbb{R}$ with $x < y$ and $y - x > 1$. There exists $n \in \mathbb{Z}$ such that $x < n < y$.
  
  *Proof.* Let $S = \{ k \in \mathbb{Z} : k < x \}$. Then $S$ is nonempty (by induction, since it is not bounded below) and bounded above by $x$, so let $m = \sup S$. Then there is $k \in S$ such that $m - 1 < k \leq m$, so $k + 1 > m$, so $k + 1 \notin S$. Hence, $k + 1 \geq x$. Now, remembering that $k \in S$ so $k < x$, we have
  
  $$k + 1 < x + (y - x) = y$$

  and therefore, $n = k + 1$ satisfies our requirements. $\Box$

- **Theorem.** (Density of $\mathbb{Q}$ in $\mathbb{R}$.) If $x, y \in \mathbb{R}$ and $x < y$, then there exists $q \in \mathbb{Q}$ such that $x < q < y$.$^{12}$
  
  *Proof.* By the Archimedean property (with $a = y - x$ and $b = 1$), there exists $n \in \mathbb{N}$ such that $n(y - x) > 1$; note that $n > 0$. Then $nx < ny$ and $ny - nx > 1$ so by the lemma, there exists $m \in \mathbb{Z}$ such that $nx < m < ny$. Dividing through by $n$ yields $x < \frac{m}{n} < y$, so $q = m/n$ satisfies our requirements. $\Box$

2.7 The cardinality of $\mathbb{R}$ [11]

- While $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are all countable, $\mathbb{R}$ is uncountable. J&P prove that $|\mathbb{R}| > \aleph_0$ using the density of $\mathbb{Q}$ in $\mathbb{R}$ (Theorem 9.8), but we can show more.

- **Theorem.** The cardinality of the real numbers is $2^{\aleph_0}$.
  
  *Proof.* Recall that $2^{\aleph_0}$ is the cardinality of $\mathcal{P}(\mathbb{N})$, the set of all binary sequences. There is an injection $(0, 1) \rightarrow \mathcal{P}(\mathbb{N})$ given by binary expansion, choosing a sequence that ends in $0$ instead of a sequence ending in $1$ if $x$ has two binary expansions (which happens precisely when $x$ is a rational number whose denominator is equal to a power of 2). There are many bijections $(0, 1) \rightarrow \mathbb{R}$, such as $x \mapsto \frac{1}{2 + \frac{1}{x - 1}}$. This establishes $|\mathbb{R}| \leq 2^{\aleph_0}$.

$^{11}$J&P Corollary 6.12.

$^{12}$J&P Theorem 7.8.
On the other hand, there is an injection $\mathbb{N}^2 \to \mathbb{R}$ given by ternary expansion as follows:

$$t : (d_1, d_2, d_3, \ldots) \mapsto 2 \sum_{n=1}^{\infty} d_n \cdot 3^{-n}.$$ 

Here’s how to verify this is an injection: Suppose that $(d_n)_n$ and $(d'_n)_n$ are different binary sequences, and suppose that $m$ is the least index with $d_m \neq d'_m$. Assume without loss that $d_m = 1$ and $d'_m = 0$ and let $s = \sum_{1 \leq k < m} d_k \cdot 3^{-k} = \sum_{1 \leq k < m} d'_k \cdot 3^{-k}$. Note that $t((d_n)_n) - s \geq 2 \cdot 3^{-m}$, while

$$t((d'_n)_n) - s \leq 2 \sum_{n=m+1}^{\infty} 3^{-m} = 2 \cdot \frac{3^{-(m+1)}}{2/3} = 3^{-m}$$

from which it follows that $t((d_n)_n) > t((d'_n)_n)$.

The image of $t$ is the Cantor set (discussed in more detail in J&P Section 57). It follows that $2^{\aleph_0} \leq |\mathbb{R}|$, so by antisymmetry of $\leq$ among cardinals (Schröder–Bernstein theorem), we have $|\mathbb{R}| = 2^{\aleph_0}$.

- The uncountability of $\mathbb{R}$ is also a consequence of a corollary to the Baire category theorem, namely that any complete metric space without isolated points must be uncountable.

### 2.8 Sequences in $\mathbb{R}$

- **Definition.** A sequence $(x_n)_n$ of real numbers is a function $\mathbb{N} \to S : n \mapsto x_n$.

  - A sequence converges if there is some $L \in \mathbb{R}$ such that for all $\epsilon > 0$ there exists $N$ such that $n \geq N$ implies $|x_n - L| < \epsilon$. In this case, we say that $L = \lim(x_n)_n$.

  - A sequence is bounded if there is some $M \geq 0$ such that $|x_n| \leq M$ for all $n$.

  - A sequence is increasing if $x_{n+1} \geq x_n$ for all $n$, and decreasing if $x_{n+1} \leq x_n$ for all $n$. A sequence is monotonic (or monotone) if it is increasing or decreasing.

  - A sequence is called Cauchy if for all $\epsilon > 0$ there exists $N$ such that $i, j \geq N$ implies $|x_i - x_j| < \epsilon$.

  - If $(x_n)_n$ is a sequence, a subsequence of $(x_n)_n$ is any sequence of the form $(x_{n_i})_i$ with $(n_i)_i$ a strictly increasing sequence in $\mathbb{N}$.

- **Theorem.** Every bounded monotonic sequence in $\mathbb{R}$ converges.
Proof. We will show this is true for increasing sequences; the decreasing case is symmetric. Let \((x_n)_n\) be a sequence in \(\mathbb{R}\) that is increasing and bounded. The set \(\{x_n\}_n\) is nonempty and bounded above, so let \(L = \sup\{x_n\}\). We will prove that \(\lim(x_n) = L\).

Let \(\epsilon > 0\). Since \(L - \epsilon\) is not an upper bound on \(\{x_n\}_n\) there exists some index \(N\) such that \(L - \epsilon < x_N \leq L\). Since \((x_n)_n\) is increasing, it is actually the case that \(L - \epsilon < x_n \leq L\) for all \(n \geq N\). The inequality implies \(|x_n - L| < \epsilon\) for all \(n \geq N\), so we have shown convergence as desired.

\(\Box\)

• It’s worth noting that the above uses completeness in an essential way. As you showed in the homework, this theorem is actually equivalent to the Least Upper Bound Axiom in \(\mathbb{R}\).

• Lemma. Every sequence in \(\mathbb{R}\) has a monotonic subsequence.

Proof. Let \((x_n)_n\) be a sequence in \(\mathbb{R}\). We will say a term \(x_m\) is dominant if \(x_m \geq x_n\) for all \(n > m\).

If \((x_n)_n\) has infinitely many dominant terms, then the subsequence that consists of dominant terms is decreasing, and we obtain the conclusion.

On the other hand, suppose \((x_n)_n\) has only finitely many dominant terms. Choose \(n_1\) so that \(x_m\) is not dominant for any \(m \geq n_1\). Since \(x_{n_1}\) is not dominant, there exists \(n_2 > n_1\) such that \(x_{n_1} < x_{n_2}\). Repeating this process, we build a strictly increasing sequence of indices \((n_i)_i\) such that the corresponding subsequence \((x_{n_i})_i\) is strictly increasing. The lemma follows. \(\Box\)

• Theorem. (Bolzano–Weierstrass for \(\mathbb{R}\).) Every bounded sequence in \(\mathbb{R}\) has a convergent subsequence.\(^{20}\)

Proof. Follows from the preceding lemma and theorem. \(\Box\)

• Theorem. Cauchy sequences are bounded.\(^{21}\)

Proof. Let \((x_n)_n\) be a Cauchy sequence of real numbers. There exists \(N\) such that \(i, j \geq N\) implies \(|x_i - x_j| < 1\). In particular, \(|x_i - x_N| < 1\) for all \(i \geq N\), so

\[|x_i| = |x_i - x_N + x_N| \leq |x_i - x_N| + |x_N| < 1 + |x_N|.\]

Let \(M = \max\{|x_1|, |x_2|, \ldots, |x_{N-1}|, |x_N| + 1\}\). It is now easy to verify that \(|x_k| \leq M\) for all \(k\). \(\Box\)

• Theorem. (Cauchy criterion in \(\mathbb{R}\).) Every Cauchy sequence in \(\mathbb{R}\) converges.\(^{22}\)

Proof. Let \((x_n)_n\) be a Cauchy sequence. Then \((x_n)_n\) is bounded by the lemma, so it has a convergent subsequence \((x_{n_k})_k\). Let \(L\) be the limit of this subsequence; we will show that the whole sequence converges to \(L\).

\(^{20}\)J&P Theorem 18.1.
\(^{21}\)J&P Proof of Theorem 19.3.
\(^{22}\)J&P Theorem 19.3.
Let \( \epsilon > 0 \). Since \( (x_{nk})_k \to L \), there exists \( K \) such that \( |x_{nk} - L| < \frac{\epsilon}{2} \) whenever \( k \geq K \). Since \( (x_n)_n \) is Cauchy, there is \( N \) such that \( |x_i - x_j| < \frac{\epsilon}{2} \) whenever \( i, j \geq N \). Choose \( k \) so that \( k \geq K \) and \( n_k \geq N \) (this is possible since the indexing sequence \( (n_k)_k \) is strictly increasing). Let \( n \geq N \). We have

\[
|x_n - L| = |x_n - x_{nk} + x_{nk} - L| \leq |x_n - x_{nk}| + |x_{nk} - L|
\]

Since \( n, n_k \geq N \), the first of these is \( < \frac{\epsilon}{2} \), and since \( k \geq K \), the second of these is \( < \frac{\epsilon}{2} \). Thus, for all \( n \geq N \), we conclude \( |x_n - L| < \epsilon \), as desired. \( \square \)

### 2.9 Arithmetic of limits

Section 12 of J&P.

#### 2.10 Some basic limits and extraction of roots in \( \mathbb{R} \)

- Most of the results in this section are ultimately based on the binomial theorem:

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k
\]

where \( \binom{n}{k} \) is the \((n,k)\)th binomial coefficient. This is particularly useful in \( \mathbb{R} \) for constructing inequalities, since e.g., if \( x \geq 0 \), then \((1 + x)^n = 1 + nx + \cdots + x^n \geq 1 + nx \) (Bernoulli’s inequality).

- **Proposition.** If \(|r| < 1\) then \( \lim(r^n) = 0 \).

  **Proof.** We can find \( b > 0 \) such that \( |r| = \frac{1}{1+b} \), namely \( b = \frac{1}{|r|} - 1 \). Now, let \( \epsilon > 0 \) and fix \( N \) so that \( \frac{1}{Nb} < \epsilon \) (possible by the Archimedean principle). For any \( n \geq N \),

\[
|r^n - 0| = |r|^n = \frac{1}{(1+b)^n} \leq \frac{1}{1+nb} < \frac{1}{nb} \leq \frac{1}{Nb} < \epsilon
\]

which establishes the limit (the first \( \leq \) is Bernoulli’s inequality). \( \square \)

- **Proposition.** If \( \alpha > 0 \) and \( p \geq 1 \) is an integer, there exists a unique \( \beta > 0 \) such that \( \beta^p = \alpha \). We denote this real number \( \beta \) by \( \alpha^{1/p} \).

  **Proof.** Uniqueness is easy to prove once existence is established, and we may assume that \( 0 < \alpha < 1 \) (since \((1/\beta)^n = 1/\alpha \) will give us the result for \( \alpha > 1 \), and the result for \( \alpha = 1 \) is trivial).

  We define two sequences \((a_n)_n\) and \((s_n)_n\) as follows:

\[
a_n = \max \left\{ k \in \mathbb{Z} : \left( \frac{k}{2^n} \right)^p \leq \alpha \right\}
\]

---

23 J&P Theorem 16.3.
24 J&P Theorem 7.8.
and $s_n = a_n/2^n$. Why? $(s_n)_n$ consists of successively better binary approximations to $\alpha^{1/p}$. One must check that $a_n$ is well-defined, but that’s not too hard.

Next we claim that $(s_n)_n$ is increasing and bounded above. It is bounded above by 1, since $s_n > 1$ would imply $s_n^p = (a_n/2^n)^p > 1 > \alpha$, which contradicts the definition of $a_n$. It is increasing because

$$2a_n \in \left\{ k \in \mathbb{Z} : \left( \frac{k}{2^{n+1}} \right)^p \leq \alpha \right\}$$

and hence $2a_n \leq a_{n+1}$ (by definition of $a_{n+1}$) so dividing by $2^{n+1}$ yields $s_n \leq s_{n+1}$. Thus, $(s_n)_n$ converges to some limit $\beta$.

Finally, we will prove that $(s_n^p)_n \to \alpha$. By the arithmetic of limits, this will prove that the limit of $(s_n)_n$, namely $\beta$, must satisfy $\beta^p = \alpha$. By the definition of $(a_n)_n$ we have

$$\left( \frac{a_n}{2^n} \right)^p \leq \alpha < \left( \frac{a_n + 1}{2^n} \right)^p$$

and so it is enough to check that the difference between the left- and right-hand quantities tends to zero as $n \to \infty$. We have

$$(a_n + 1)^p - a_n^p = \sum_{k=0}^{p-1} a_n^k$$

so since $a_n \leq 2^n$ (because otherwise $s_n > 1$, which we already showed is not possible), we have $a_n^k \leq 2^{nk} \leq 2^n(p-1)$ when $k = 0, \ldots, p-1$, and so

$$\left( \frac{a_n + 1}{2^n} \right)^p - \left( \frac{a_n}{2^n} \right)^p \leq \frac{p \cdot 2^n(p-1)}{2^{np}} = \frac{p}{2^n}$$

From which we conclude that the left-hand quantity $\to 0$ as $n \to \infty$ (since $p$ is constant). □

- **Proposition.** If $p \geq 1$ is an integer, then $\lim(\frac{1}{n^p}) = 0$.
- **Proposition.** $\lim(n^{1/n}) = 1$.
- **Proposition.** If $a > 1$, then $\lim(a^{1/n}) = 1$.25

### 3 Metric spaces and their topology

#### 3.1 Metric spaces [7]

- **Definition.** A metric space is a pair $(X,d)$ where $X$ is a set of points and $d : X \times X \to \mathbb{R}$ is a metric (or distance function), satisfying

26 J&P Definition 35.1.
(i.) For all $x, y \in X$, $d(x, y) = 0$ iff $x = y$;
(ii.) For all $x, y \in X$, $d(x, y) = d(y, x)$;
(iii.) For all $x, y, z \in X$, $d(x, y) \leq d(x, z) + d(z, y)$.

(iii.) is the triangle inequality.

- (i.–iii.) together imply that $d(x, y) \geq 0$ for all $x, y \in X$. This is sometimes listed as one of the axioms of a metric, but it can be proven as a basic result.

- An ultrametric is a metric that satisfies a stronger version of the triangle inequality, namely that $d(x, y) \leq \max\{d(x, z), d(z, y)\}$ for all $x, y, z \in X$.

- A metric space is a space with some specified and sensible notion of distance.

### 3.2 Sequences in metric spaces [7]

- **Definition.** Let $X$ be a metric space. A sequence $(x_n)_n$ is a function $\mathbb{N} \to X : n \mapsto x_n$.
  - We say that $(x_n)_n$ converges if there is $L \in X$ such that for all $\epsilon > 0$ there exists $N$ such that $n \geq N$ implies $d(x_n, L) < \epsilon$. (We of course write $\lim(x_n)_n = L$.)\(^{27}\)
  - We say that $(x_n)_n$ is Cauchy if for every $\epsilon > 0$ there exists $N$ such that $i, j \geq N$ implies $d(x_i, x_j) < \epsilon$.\(^{28}\)
  - We say that $(x_n)_n$ is bounded if for all $o \in X$, the set $\{d(x_n, o)\}_n \subseteq \mathbb{R}$ is bounded.

- The next two propositions have similar proofs to the special case of $X = \mathbb{R}$.

- **Proposition.** Every convergent sequence in a metric space is a Cauchy sequence.\(^{29}\)

- **Proposition.** Every Cauchy sequence in a metric space is bounded.

- **Definition.** A metric space $X$ is called complete if every Cauchy sequence is a convergent sequence.\(^{30}\)

- **Theorem.** $\mathbb{R}$ is a complete metric space (with respect to the Euclidean metric).
  
  **Proof.** This is just a restatement of an earlier result. \(\Box\)

- **Definition.** Let $(X, d)$ be a metric space. A completion of $X$ is a metric space $(Y, d')$ such that $X \subseteq Y$, $d = d'|_{X \times X}$, $Y$ is complete, and $X$ is dense (defined shortly) in $Y$.

- **Theorem.** Every metric space has a completion.\(^{31}\)

  The proof, which is long and extremely tedious, is in J&P. The construction of a metric space’s completion is essentially the same as the construction of $\mathbb{R}$ from $\mathbb{Q}$ via equivalence classes of Cauchy sequences (a special case of this more general metric space construction).

\(^{27}\)J&P Definition 37.1.

\(^{28}\)J&P Definition 46.1.

\(^{29}\)J&P Theorem 46.2.

\(^{30}\)J&P Definition 46.3.

\(^{31}\)J&P Theorem 46.7.
3.3 \( \mathbb{R}^n \) as a metric space [7]

- The Cauchy–Schwarz inequality.\(^{32}\)
- Minkowski’s inequality (the Euclidean metric is a metric).\(^{33}\)
- Convergence in \( \mathbb{R}^n \) is equivalent to convergence in every coordinate.\(^{34}\)
- The Bolzano–Weierstrass theorem holds in \( \mathbb{R}^n \).
- \( \mathbb{R}^n \) is complete.\(^{35}\)

3.4 Topological spaces [8]

- **Definition.**\(^{36}\) A *topological space* is a pair \((X, \tau)\) where \(X\) is a set and \(\tau\) is a set of subsets of \(X\) (\(\tau\) is called the *topology* and its elements are called the *open sets*), satisfying
  
  i. \( \emptyset, X \in \tau \) (the empty set and whole space are open sets);
  
  ii. For any \( F \subseteq \tau, \bigcup F \in \tau \) (any union of open sets is open);
  
  iii. For any \( U_1, U_2 \in \tau, U_1 \cap U_2 \in \tau \) (a finite intersection of open sets is open).

- In topology a set is called *closed* if its complement is open.

  Note that we can define a topology by specifying the set of *closed sets* \( \tau' \) instead, verifying \( \emptyset, X \in \tau' \), and verifying that \( \tau' \) is closed under arbitrary intersections and closed under finite unions.

- If \( x \in X \) and \( U \in \tau \) such that \( x \in U \), we say that \( U \) is an *open neighborhood* of \( x \).

- A topological space is called *Hausdorff* if every pair of distinct points have disjoint open neighborhoods. That is, \( X \) is Hausdorff if for all \( x \neq y \) there exist \( U, U' \in \tau \) such that \( x \in U, y \in U' \), and \( U \cap U' = \emptyset \).

- We call a subset \( S \subseteq X \) of a topological space *dense* if every open set has nonempty intersection with \( S \).

3.5 Metric spaces and their structure [8]

- Let \( X \) be a metric space. For any \( x \in X \) and any \( r > 0 \) we define
  
  \[
  B_r(x) = \{ y \in X : d(x,y) < r \}
  \]

  the *open ball* of radius \( r \) around the point \( x \).\(^{37}\)

---

\(^{32}\)J&P Theorem 36.1.

\(^{33}\)J&P Theorem 36.2.

\(^{34}\)J&P Theorem 37.2.

\(^{35}\)J&P Theorem 46.4.

\(^{36}\)J&P end of section 39.

\(^{37}\)J&P Definition 39.1.
Let $X$ be a metric space let $x \in X$, and let $S \subseteq X$.

- We call $x$ an interior point of $S$ if there exists $r > 0$ such that $B_r(x) \subseteq S$.
- We call $x$ an adherent point of $S$ if for all $r > 0$ we have $B_r(x) \cap S \neq \emptyset$.
- We call $x$ a limit point of $S$ if for all $r > 0$ the set $B_r(x) \cap S$ contains some $y \in X$ such that $y \neq x$.
- We call $x$ an isolated point of $S$ if there exists $r > 0$ such that $B_r(x) \cap S = \{x\}$.

Notes: If $x$ is an interior point or an isolated point of $S$, then $x \in S$. If $x \in S$, then $x$ is adherent to $S$. If $x$ is an adherent point, then it is either a limit point or an isolated point of $S$.

Let $X$ be a metric space and let $S \subseteq X$.

- We call $S$ open if every $x \in S$ is an interior point of $S$.
- We call $S$ closed if, whenever $x$ is an adherent point of $S$, then $x \in S$.
  (Since isolated points already belong to $S$, is equivalent to require that $S$ contain all its limit points.)
- We call $S$ perfect if it is closed and has no isolated points.
- We call $S$ bounded if for any $o \in X$ (think of this as analogous to a choice of origin) there exists $M \geq 0$ such that $d(o, x) \leq M$ for all $x \in S$.
- We call $S$ dense (in $X$) if for all $x \in X$, $x$ is an adherent point of $S$.

In general topology, an isolated point is an $x$ such that $\{x\}$ is open. A dense set is one which has nonempty intersection with every open set.

If $X$ is a metric space, then the set of open sets (as defined relative to the metric $d$) is a Hausdorff topology on $X$.

If $X$ is a metric space and $U \subseteq X$ is open, then $U^c$ is closed (and vice versa).

If $X$ is a metric space, a subset $S \subseteq X$ is dense iff every $x \in X$ is adherent to $S$.

A metric space is called separable if it has a countable, dense subset.

$\mathbb{R}^n$ is separable.

---

41. J&P Definition 38.1.
42. J&P Definition 47.1.
44. J&P Theorem 39.5.
3.6 Compact sets [8]

- Let $X$ be a topological space. A subset $K \subseteq X$ is called compact if for all $\mathcal{G} \subseteq \tau$ satisfying $K \subseteq \bigcup \mathcal{G}$ there is a finite subset $\mathcal{G}' \subseteq \mathcal{G}$ such that $K \subseteq \bigcup \mathcal{G}'$.
- In this definition, $\mathcal{G}$ is called an open cover of $K$, and $\mathcal{G}'$ is a finite open subcover.\(^{45}\)
- Compact subsets of Hausdorff topological spaces (e.g., metric spaces) are closed.
- Compact subsets of metric spaces are bounded.

3.7 The Heine–Borel theorem [9]

- A subset of $\mathbb{R}^n$ is compact iff it is closed and bounded.\(^{46}\)

3.8 Sequential compactness theorem [10]

- **Lemma.** Suppose that $X$ is a metric space and $S$ is a subset of $X$. If $S$ is finite, then $S$ has no limit points (in $X$).

- **Lemma.** Let $X$ be a metric space and let $S \subseteq K \subseteq X$ with $K$ compact and $S$ infinite. Then there is $L \in K$ such that $L$ is a limit point of $S$.

  *Proof.* Suppose that $S \subseteq K \subseteq X$ with $S$ infinite, but that $K$ does not contain any limit points of $S$. Then, for all $x \in K$, there exists $r_x > 0$ such that $B_{r_x}(x) \cap S \subseteq \{x\}$. $\mathcal{G} = \{ B_{r_x}(x) : x \in K \}$ is an open cover of $K$. However, if $\mathcal{G}' = \{ B_{r_{x_1}}(x_1), \ldots, B_{r_{x_n}}(x_n) \}$ is any finite subset of $\mathcal{G}$, $S \cap \bigcup \mathcal{G}' \subseteq \{x_1, \ldots, x_n\}$, so since $S$ is infinite, $S \not\subseteq \bigcup \mathcal{G}'$ and therefore $K \not\subseteq \bigcup \mathcal{G}'$ (since $S \subseteq K$). Since $K$ has an open cover with no finite subcover, $K$ is not compact. \(\square\)

- **Theorem.** (Sequential compactness theorem.)\(^{47}\) Let $X$ be a metric space, let $K$ be a compact subset of $X$, and let $(s_n)$ be a sequence in $K$. Then, there is a subsequence $(s_{n_j})$ of $(s_n)$ that converges to a point $L \in K$.

  *Proof.* Let $S = \{s_n\}_n$. If $S$ is finite, we are done (essentially by the pigeonhole principle): There is $L \in S$ such that $I = \{ n : s_n = L \}$ is infinite, so indexing $I = \{ n_1 < n_2 < n_3 < \cdots \}$ yields the desired subsequence.

  On the other hand, if $S$ is infinite, then by the lemma, $K$ contains a limit point of $S$, which we will call $L$. Since $L$ is a limit point of $S = \{s_n\}$, we may choose $n_1$ so that $d(s_{n_1}, L) < 1$. We may then choose $n_2 > n_1$ such that $d(s_{n_2}, L) < \frac{1}{2}$ (we can guarantee $n_2 > n_1$ because $d(s_n, L) < \frac{1}{2}$ must be true for infinitely many values of $n$, since $L$ is a limit point of $\{s_n\}$). Continuing in this way, we find $n_1 < n_2 < n_3 < \cdots$ such that $d(s_{n_j}, L) < \frac{1}{j}$. Thus, $\lim(s_{n_j}) = L$, as desired. \(\square\)

\(^{45}\) J&P Definition 42.2.

\(^{46}\) J&P Theorem 43.9.

\(^{47}\) J&P Theorem 43.1.
3.9 Baire category theorem [10]

- **Theorem.** (Baire category theorem.)\(^{48}\) Let \(X\) be a complete metric space. Then,

  i. If \(\mathcal{U} = \{U_n\}_{n=1}^{\infty}\) is a sequence of dense open subsets of \(X\), their intersection \(\bigcap \mathcal{U} = \bigcap_{n=1}^{\infty} U_n\) is dense; and

  ii. If \(\mathcal{F} = \{F_n\}_{n=1}^{\infty}\) is a sequence of closed subsets of \(X\) such that \(X = \bigcup \mathcal{F} = \bigcup_{n=1}^{\infty} F_n\), then at least one element of \(\mathcal{F}\) has nonempty interior.

*Note that “sequence” in (i.) and (ii.) above can be replaced with “countable collection.”*

**Proof.** (i.) and (ii.) are equivalent; we will prove (i.).

Let \(W\) be a nonempty open subset of \(X\). We must show that \(W \cap \bigcap \mathcal{U}\) is nonempty, so it is enough to find a \(\xi \in X\) such that \(\xi \in W\) and \(\xi \in U_n\) for every \(n \geq 1\).

Since \(U_1\) is dense, it has nonempty intersection with \(W\). We may therefore find \(x_1 \in W \cap U_1\). Since \(W\) and \(U_1\) are open, \(W \cap U_1\) is open, so there is \(r_1 \in (0, 1)\) such that

\[
\overline{B}_{r_1}(x_1) \subseteq W \cap U_1
\]

where \(\overline{B}_{r_1}(x_1) = \{y \in X : d(x_1, y) \leq r_1\}\) is the closed ball of radius \(r_1\) around \(x_1\) (why? \(x_1\) is interior to \(W \cap U_1\), so there is \(\epsilon\) with \(B_{\epsilon}(x_1) \subseteq W \cap U_1\), and we can take \(r_1 = \epsilon/2\), for instance).

Now, \(U_2\) is dense and \(\overline{B}_{r_1}(x_1)\) is (nonempty and) open, so they have nonempty intersection. Thus, we can find an \(x_2 \in X\) and an \(r_2 \in (0, \frac{1}{2})\) such that \(\overline{B}_{r_2}(x_2) \subseteq \overline{B}_{r_1}(x_1) \cap U_2\).

Continuing in this way, we build a sequence \((x_n)\) of points in \(X\) and a sequence \((r_n)\) of real numbers such that

- \(0 < r_n < \frac{1}{n}\) for all \(n \geq 1\); and

- \(\overline{B}_{r_n}(x_n) \subseteq B_{r_{n-1}}(x_{n-1}) \cap U_n\).

- By the above, \(\overline{B}_{r_1}(x_1) \supseteq \overline{B}_{r_2}(x_2) \supseteq \overline{B}_{r_3}(x_3) \supseteq \ldots\)

The last of these implies that the sequence \((x_n)\) is Cauchy, since if \(i \geq j \geq N\), we have \(x_i, x_j \in \overline{B}_{r_j}(x_j)\) and therefore \(d(x_i, x_j) \leq r_j < \frac{1}{j} \leq \frac{1}{N}\). Because \(X\) is complete, \((x_n)_n\) converges to a limit \(\xi \in X\).

Since \((x_n, x_{n+1}, x_{n+2}, \ldots)\) is a convergent sequence (with limit \(\xi\)) in \(\overline{B}_{r_n}(x_n)\), which is closed, we may conclude that \(\xi \in \overline{B}_{r_n}(x_n)\) for all \(n\). Finally, since we have inclusions

- \(\overline{B}_{r_1}(x_1) \subseteq W\); and

- \(\overline{B}_{r_n}(x_n) \subseteq U_n\) for all \(n \geq 1\);

we may conclude that \(\xi \in W\) and that \(\xi \in U_n\) for all \(n \geq 1\). \(\square\)

---

\(^{48}\)J&P Theorem 47.2.
• Corollary. A complete metric space $X$ with no isolated points must be uncountable.

*Proof.* Note that if $X$ has no isolated points, then for every $x \in X$, the singleton set $\{x\}$ has empty interior. Now, if $X$ is a countable metric space, there is a sequence $(x_n)_n$ such that $X = \{x_n\}_n$. So, $F = \{x_n\}_n$ is a countable collection of closed sets with empty interior whose union is $X$, so $X$ is not complete. □

3.10 Small sets and big sets [11]

- One thing that makes analysis (even just in $\mathbb{R}$) challenging and interesting is the sometimes counterintuitive interplay between what set theory, topology, and measure theory have to say about the spaces we work in.

- For example, all three fields have different ways to address the “size” of $S \subseteq \mathbb{R}$.
  
  - A set theorist might say $S$ is small if it is countable and that it is big if it is uncountable.
  
  - A topologist might say $S$ is small if it is nowhere dense ($\overline{S} = \partial S$), and that it is big if it is somewhere dense ($\overline{S}$ has an interior point).
  
  - A measure theorist might say $S$ is small if it has zero Lebesgue measure (length), and that $S$ is large if it has positive or infinite measure.

- A minimal explanation of the Lebesgue measure: The Lebesgue measure is a function $\lambda : \Sigma \to [0, +\infty]$ that assigns lengths to the measurable subsets of $\mathbb{R}$, the set of which is denoted $\Sigma$ (the existence of non-measurable sets (sets whose length cannot be defined) will come up later). $\lambda$ satisfies the following:
  
  i. $\lambda(\varnothing) = 0$;
  
  ii. The length of an interval is what you would expect, e.g., $\lambda([a, b]) = b - a$ if $a \leq b$;
  
  iii. If $U, V \in \Sigma$ and $U \subseteq V$, then $\lambda(U) \leq \lambda(V)$;
  
  iv. If $U = \{U_n\}_{n=1}^\infty$ is a countable collection of measurable subsets of $\mathbb{R}$ such that $U_i \cap U_j = \varnothing$ whenever $i \neq j$ (pairwise disjoint), then $\lambda(\bigcup_n U_n) = \sum_n \lambda(U_n)$.

- Since $\lambda(\{a\}) = 0$ for all $a \in \mathbb{R}$, (iv.) implies that $\lambda(S) = 0$ if $S$ is countable.

- With one exception (countable implies zero measure), these notions turn out to be pretty much independent of each other:

<table>
<thead>
<tr>
<th>Example(s)</th>
<th>Set theory</th>
<th>Topology</th>
<th>Measure theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{R}$, $\mathbb{R} \setminus \mathbb{Q}$, intervals</td>
<td>big</td>
<td>big</td>
<td>big</td>
</tr>
<tr>
<td>$\mathbb{Q} \cup C_{\infty}$</td>
<td>big</td>
<td>big</td>
<td>small</td>
</tr>
<tr>
<td>Smith–Volterra–Cantor, $\bigcup_n B_{1/2^n}(q_n)^c$</td>
<td>big</td>
<td>small</td>
<td>big</td>
</tr>
<tr>
<td>$C_{\infty}$</td>
<td>big</td>
<td>small</td>
<td>small</td>
</tr>
<tr>
<td>$\mathbb{Q}$</td>
<td>small</td>
<td>big</td>
<td>small</td>
</tr>
<tr>
<td>$\varnothing$, finite sets, $\mathbb{Z}$</td>
<td>small</td>
<td>small</td>
<td>small</td>
</tr>
</tbody>
</table>

$^49$J&P Theorem 47.8.
• **Theorem.** $\mathbb{R}$ is not a countable union of “small” sets.

**Proof.** For set theory, this follows from the fact that $\mathbb{R}$ is uncountable and that a countable union of countable sets is countable.

For measure theory, this follows from the fact that the Lebesgue measure is countably additive, so $\mathbb{R}$ is not the union of countably many sets of measure zero (as $\lambda(\mathbb{R}) = \infty$). For topology, this is the Baire category theorem, (ii.)! □

4 Continuity

4.1 Metric and topological subspaces [12]

• Let $(X, \tau)$ be a topological space. If $S \subseteq X$, the *induced/relative/subspace topology* on $S$ is $\tau_S = \{ U \cap S : U \in \tau \}$. This makes $(S, \tau_S)$ a topological space in its own right.

• Let $(X, d)$ be a metric space. If $S \subseteq X$, the *induced/relative/subspace metric* on $S$ is $d_S = d_{|S \times S}$ (restricted to $S \times S$). This makes $(S, d_S)$ into a metric space in its own right.

• Note that if $U$ is open (or closed) in $S$, it need not be open (or closed) in $X$. For example, $[0, 1)$ is open in the subspace topology on $[0, \infty)$ inherited from $\mathbb{R}$ (since it is the intersection of $(-1, 1)$ and $[0, \infty)$), even though it is not open in $\mathbb{R}$ itself.

• On the other hand, if $K \subseteq S \subseteq X$ and $K$ is compact in the subspace topology of $S$, then it is compact in the topology on $X$.

• If $X$ is a metric space and $S \subseteq X$, then the subspace topology on $S$ is the topology that is induced by the subspace metric.

• We define these notions because when we talk about functions, our domains may be proper subsets of the usual metric spaces we consider. We will often not mention when we are using the subspace topology, as it is usually implicit.

• Probably self-explanatory: A topological space $X$ is called *compact* if $X$ is itself compact. For example, $[0, 1]$ is a compact topological space (with the topology inherited from the Euclidean topology on $\mathbb{R}$), but $\mathbb{R}$ itself is not a compact topological space.

4.2 Three definitions of continuity [12]

• Let $(X, d)$ and $(Y, d')$ be metric spaces, let $f : X \to Y$, and let $p \in X$.

We say that $f$ is *continuous at $p$*...

i. In the **analytic sense**\(^{51}\) if for all $\epsilon > 0$ there exists $\delta = \delta_{p,\epsilon} > 0$ such that

$$d(x, p) < \delta \Rightarrow d'(f(x), f(p)) < \epsilon.$$ 

\(^{50}\)J&P §41.

\(^{51}\)J&P Definition 40.1.
ii. In the **sequential sense**\(^{52}\) if for all \((x_n)_n \to p\) in \(X\) we have \((f(x_n))_n \to f(p)\) in \(Y\).

iii. In the **topological sense**\(^{53}\) if for all open sets \(V \subseteq Y\) such that \(f(p) \in V\), there is an open set \(U \subseteq X\) such that \(p \in U\) and \(f(U) \subseteq V\) (eq. \(U \subseteq f^{-1}(V)\)).

- **Notes:** In the topological sense, we can replace “open” with “closed.” The topological sense here is the definition of continuity at a point in topology (it does not require a metric!). In the analytic sense, \(\delta\) can depend on both \(p\) and \(\epsilon\) (important later).

- **Theorem.** The different senses of “continuous at \(p\)” (i–iii) above are equivalent.

  **Proof.** (i) \(\Rightarrow\) (ii) is easy. I personally trust (ii) \(\Rightarrow\) (i) less, so I’ll prove that one: Suppose \(f : X \to Y\) is not analytically continuous at \(p\); there exists \(\epsilon > 0\) such that

  \[
  \forall \delta > 0, \exists x \in X : d(x, p) < \delta \land d'(f(x), f(p)) \geq \epsilon.
  \]

  Specializing to \(\delta = \frac{1}{n}\), we can construct a sequence \((x_n)_n\) in \(X\) such that

  \[
  \forall n : d(x_n, p) < \frac{1}{n} \land d'(f(x_n), f(p)) \geq \epsilon
  \]

  The first inequality guarantees \((x_n)_n \to p\) while the second implies \((f(x_n))_n \not\to f(p)\), so \(f\) is not sequentially continuous.

  (iii) \(\Rightarrow\) (i). Suppose that \(f\) is topologically continuous at \(p\) and let \(\epsilon > 0\). \(V = B_r(f(p))\) is an open subset of \(Y\), so by topological continuity, there exists an open \(U\) such that \(p \in U \subseteq f^{-1}(V)\). Since \(U\) is open, \(p\) is interior to \(U\), so there exists \(\delta > 0\) satisfying \(B_\delta(p) \subseteq U\). Now, for all \(x \in X\), we have

  \[
  d(x, p) < \delta \iff x \in B_\delta(p) \implies x \in U \implies f(x) \in V = B_r(f(p)) \iff d'(f(x), f(p)) < \epsilon
  \]

  so \(f\) is analytically continuous at \(p\). The proof of (i) \(\Rightarrow\) (iii) is similar (possibly you can just carefully reverse all the implications above). \(\square\)

- **Definition.** We say that \(f : X \to Y\) is continuous if it is continuous at every \(p \in X\).

  This definition translates as follows for (ii) and (iii):

  - \(f : X \to Y\) is continuous if for every convergent sequence \((x_n)_n\) in \(X\), \((f(x_n))_n\) is convergent in \(Y\), and \(f(\lim x_n) = \lim f(x_n)\). In short “a continuous function is one that maps convergent sequences to convergent sequences.” This resembles the “calculus definition” of continuity (function and limit commute).

  - \(f : X \to Y\) is continuous if for all open (resp. closed) subsets \(V\) of \(Y\), \(f^{-1}(V)\) is an open (resp. closed) subset of \(X\).\(^{54}\)

- **Caution:** If \(f : X \to Y\) is continuous and \(U\) is an open (or closed) subset of \(X\), then \(f(U)\) need not be open (or closed). Functions that preserve open sets (resp. closed sets) are called open maps (resp. closed maps).

\(^{52}\)J&P Theorem 40.2.

\(^{53}\)J&P Definition 40.5.

\(^{54}\)J&P Definition 40.5.
4.3 Continuity and compactness [12]

- **Theorem.** If \( f : X \rightarrow Y \) and \( K \) is a compact subset of \( X \), then \( f(K) \) is a compact subset of \( Y \). That is, “Continuous images of compact sets are compact.”

- **Proof.** Let \( \mathcal{H} \) be an open cover of \( f(K) \). By topological continuity,
  \[
  \mathcal{G} = \{ f^{-1}(V) : V \in \mathcal{H} \}
  \]
  consists of open sets, and it is easy to check that \( K \subseteq \bigcup \mathcal{G} \), so \( \mathcal{G} \) is an open cover of \( K \). Since \( K \) is compact, there is a finite \( \mathcal{G}' \subseteq \mathcal{G} \) such that \( K \subseteq \bigcup \mathcal{G}' \). By construction, there is a finite \( \mathcal{H}' \) such that \( \mathcal{G}' = \{ f^{-1}(V) : V \in \mathcal{H}' \} \), and it is easy to check that \( f(K) \subseteq \bigcup \mathcal{H}' \). □ (Above, “easy to check” means use Theorem 2.6 or similar from J&P.)

- **Corollary.** (Extreme value theorem.) If \( f : X \rightarrow \mathbb{R} \) be a real-valued function on a (nonempty) compact metric space \( X \), then \( f \) achieves a minimum and a maximum value on \( X \). That is, there are \( p, q \in X \) such that \( f(p) \leq f(x) \leq f(q) \) for all \( x \in X \).
  
  **Proof.** \( f(X) \) is nonempty, closed, and bounded. By completeness in \( \mathbb{R} \), \( m = \inf f(X) \) and \( M = \sup f(X) \) are defined, and since \( f(X) \) is closed, \( m, M \in f(X) \) (it is a standard exercise to verify that the infimum and supremum of a bounded set are always adherent points). By definition, there exist \( p, q \in X \) such that \( f(p) = m \) and \( f(q) = M \). □

- **Corollary.** If \( f : X \rightarrow Y \) is continuous and injective, and \( X \) is compact, then \( f^{-1} : Y \rightarrow X \) is continuous.
  
  **Proof.** We will show that \( f^{-1} \) is continuous in the topological sense. Let \( U \) be an open set of \( X \); we need to show that \( (f^{-1})^{-1}(U) = f(U) \) is an open set of \( Y \). Since \( U^c \) is a closed subset of a compact space, it is compact, so \( f(U^c) \) is compact and therefore closed. Because \( f \) is injective, one can check that \( f(U^c) = f(U)^c \), so \( f(U^c)^c = (f(U)^c)^c = f(U) \) is open, completing the proof. □

---

55. J&P Theorem 44.1.
56. J&P Theorem 44.3.
4.4 Continuity and connectedness [13]

- **Definition.** Let $X$ be a topological space. $X$ is called *disconnected* if there are disjoint nonempty open $U, U' \subseteq X$ satisfying $X = U \cup U'$. Otherwise, $X$ is called *connected*.

A subset $S \subseteq X$ is called *connected* if $S$ is connected in the subspace topology inherited from $X$.

- Note that if $X$ is disconnected with $X = U \cup U'$, then $U$ and $U'$ are clopen subsets that are different from $\emptyset$ and $X$. That is, $X$ is disconnected iff it admits a nonempty proper clopen subset.\(^{57}\)

- **Lemma.**\(^ {58}\) If $S \subseteq \mathbb{R}$ then $S$ is connected (in the Euclidean topology) iff $S$ is an interval. That is, $S$ is connected iff it has the following property:

\[
\forall x, y, z \in \mathbb{R} : (x, y \in S) \land (x < z < y) \Rightarrow z \in S
\]

**Proof.** ($\Rightarrow$) Suppose that $S$ is not an interval. Then there exist $x, y \in S$ and $z \in \mathbb{R}$ such that $x < z < y$ but $z \notin \mathbb{R}$. Let $U = (−\infty, z)$ and $U' = (z, \infty)$. $S = (U \cap S) \cup (U' \cap S)$ is easily seen to be a disconnection of $S$.

($\Leftarrow$) Suppose that $S$ is an interval and that $S = U \cup U'$ for sets $U, U'$ which are nonempty and open in the subspace topology on $S$. We will prove that $U \cap U'$ is nonempty. Let $x \in U$ and let $y \in U$ and assume without loss that $x < y$. Construct two sequences $(x_n)_n$ and $(y_n)_n$ as follows:

- $x_0 = x$ and $y_0 = y$.
- For all $n$, let $\alpha_{n+1} = \frac{x_n + y_n}{2}$.
- By the interval property, $\alpha_{n+1} \in S$.
- If $\alpha_{n+1} \in U$, then let $x_{n+1} = \alpha_{n+1}$ and let $y_{n+1} = y_n$.
- Otherwise, $\alpha_{n+1} \in U'$ and we let $x_{n+1} = x_n$ and $y_{n+1} = \alpha_{n+1}$.

It is easy to see that $x_n \in U$, $y_n \in U'$, and $|x_n - y_n| = 2^{-n}|x - y|$ for all $n$, and that the sequences $(x_n)_n$ and $(y_n)_n$ are monotone and bounded. It follows that they converge to the same real number $L$ satisfying $x < L < y$. By the interval property, $L \in S$. If $L \in U$, then $L$ is interior to $U$, and there exists $\epsilon > 0$ such that $B_\epsilon(L) \subseteq U$ (still the case in the relative topology). By convergence $(y_n)_n \to L$, there exists $N$ such that $y_N \in B_\epsilon(L) \subseteq U$. It follows that $y_N \in U \cap U'$.

- **Theorem.** If $f : X \to Y$ is continuous and $X$ is connected, then $f(X)$ is connected.\(^ {59}\)

**Proof.** It is easy to check that if $f : X \to Y$ is continuous, then $f : X \to f(X)$ is continuous in the subspace topology on $f(X)$. If $f(X)$ is disconnected, then there are nonempty disjoint open sets $V, V'$ of $f(X)$ such that $V \cup V' = Y$. Let $U = f^{-1}(V)$ and $U' = f^{-1}(V')$. These are nonempty and open and $X = U \cup U'$. If $x \in U \cap U'$,

\(^{57}\)J&P Definition 45.1.

\(^{58}\)J&P Theorem 45.3.

\(^{59}\)J&P Theorem 45.5.
4.5 Uniform continuity

• Theorem. Let \( f : X \to Y \) be a continuous function of metric spaces. If \( X \) is compact, then \( f \) is uniformly continuous.

\[
\forall \epsilon > 0, \exists \delta > 0, \forall p, x \in X : d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \epsilon
\]

The salient differences between this and the analytic definition of continuity is that \( \delta \) no longer depends on the choice of \( p \)—for a given \( \epsilon \), the same \( \delta \) must work at all points \( p \in X \).

• If \( f \) is uniformly continuous then it is also (regular) continuous.

• Examples of functions that are continuous but not uniformly so:
  
  - \( f(x) = \frac{1}{x} \) on \((0, 1]\).
  - \( f(x) = \sin(\frac{1}{x}) \) on \((0, 1]\).
  - \( f(x) = x^2 \) on \(\mathbb{R}\).

To see why, pick a point \( p \) in the domain, and consider the inverse image of \((p-\delta,p+\delta)\) as we “slide” \( p \) around the \( x \)-axis. In the first example, as \( p \to 0^+ \), the interval \( f^{-1}((p-\delta,p+\delta)) \) becomes larger and larger (so for a given \( \epsilon, \delta \) cannot work for all \( p \)). In the second example, as \( p \to 0^+ \), eventually you have \( f^{-1}((p-\delta,p+\delta)) = [-1,1] \). The same idea works for the \( f(x) = x^2 \) example (seems like it should be tame! polynomials are usually so friendly!) with \( p \to \infty \).

• Theorem. Let \( f : X \to Y \) be a continuous function of metric spaces. If \( X \) is compact, then \( f \) is uniformly continuous.

\textit{Proof.} Suppose that \( f \) is not uniformly continuous. Then there exists an \( \epsilon > 0 \) such that for every \( n \geq 1 \), \( \delta = \frac{1}{n} \) “does not work.” That is, there exist \( p_n, q_n \in X \) such that

\[
d_X(p_n, q_n) < \frac{1}{n} \land d_Y(f(p_n), f(q_n)) \geq \epsilon
\]

By the sequential compactness theorem, since \( X \) is compact, \((p_n)_n\) has a convergent subsequence \((p_{n_i})_i \to L \in X\), and it is clear from \( d_X(p_{n_i}, q_{n_i}) < \frac{1}{n_i} \) that \((q_{n_i}) \to L\) as well. By sequential continuity of \( f \), we have

\[
\lim f(p_{n_i}) = f(\lim p_{n_i}) = f(L) = f(\lim q_{n_i}) = \lim f(q_{n_i})
\]

But this is impossible, since \( d_Y(f(p_{n_i}), f(q_{n_i})) \geq \epsilon \) for all \( i \). \( \square \)

---

\(^{60}\)J&P Theorem 45.6.
• This means that while the examples above are not uniformly continuous on those domains, they will be if we consider them on a closed interval (on which they are defined), instead.

4.6 Limits of functions, discontinuities

• Let $X, Y$ be metric spaces, $S \subseteq X$, $p \in \overline{S}$, $L \in Y$, and let $f : X \to Y$ be a function. We write

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

if the following is true: For all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in S$, if $d_X(x, p) < \delta$, then $d_Y(f(x), L) < \epsilon$. This is the limit of $f$ as $x$ tends to $p$ through $S$. Note that we do not assume $p \in S$ (and in fact, the utility of limits in calculus is that they are taken through sets that do not include $p$).

• A sequential definition of the limit is also often useful: $\lim_{x \to p} f(x) = L$ if for every sequence $(x_n)_{n \to p}$ with terms in $S$ we have $\lim(f(x_n))_n = L$.

• Let $E \subseteq \mathbb{R}$ and let $p \in E$. If $f : E \to \mathbb{R}$ is a real function (a real-valued function on a subset of the reals) we write

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

when those limits exist. If $f(p^-)$ and $f(p^+)$ both exist, their common value is called $\lim_{x \to p} f(x)$ (without further decoration).

• One can check that the “calculus definition” of continuity at a point matches our more theoretical one: A real function $f : E \to \mathbb{R}$ is continuous at $p \in E$ if $\lim_{x \to p} f(x)$ exists and is equal to $f(p)$.

• Definition. Let $f : E \to \mathbb{R}$ be a real function and let $p \in E$ (technical annoyance: with $p \neq \max(E), \min(E)$). Suppose that $f$ is discontinuous at $p$. We say that $f$ has a discontinuity of the first kind (or a simple discontinuity) at $p$ if $f$ is discontinuous at $p$ but $f(p^-)$ and $f(p^+)$ both exist. Otherwise, we say that $f$ has a discontinuity of the second kind at $p$.

If $p = \max(E)$ (resp $p = \min(E)$), then the discontinuity is of the first kind provided that $f(p^-)$ (resp $f(p^+)$) exists.

• Note: For a discontinuity of the first kind, we could have $f(p^-) = f(p^+) \neq f(p)$ or $f(p^-) \neq f(p^+)$.

• Theorem. Let $f : E \to \mathbb{R}$ be monotonic and suppose that $f$ is discontinuous at some $p \in E$. Then $f$ has a discontinuity of the first kind at $p$.

---

61 Limits are covered in J&P Sections 30–32.
62 Discontinuities are covered this way in Chapter 4 of Rudin, beginning around 4.25.
Proof. Conceptually the proof is not difficult: We just have to show that for an increasing function, \(f(p^-)\) and \(f(p^+)\) will always exist. This is a consequence of the least upper bound (and greatest lower bound) axiom. However, the details are a bit messy: See Theorem 4.29 and its corollary in Rudin. □

- **Corollary.** If \(f : E \to \mathbb{R}\) is a monotonic function, then

  \[
  \{ p \in E : f \text{ is discontinuous at } p \}
  \]

  is at most countably infinite.

*Proof. Assume \(f\) is increasing. Let \(D\) be the set of discontinuities (as above). We define (using AC) a function \(q : D \to \mathbb{Q}\) as follows: If \(p \in D\), then \(f(p^-)\) and \(f(p^+)\) both exist, but because \(f\) is increasing, we must have \(f(p^-) < f(p^+)\). Choose \(q(p)\) so that \(f(p^-) < q(p) < f(p^+)\) (by density of \(\mathbb{Q}\) in \(\mathbb{R}\)). One can check (because \(f\) is increasing) that \(q\) is injective. It follows that \(|D| \leq |\mathbb{Q}| = \aleph_0\). □

- The converse to the above is true (in some sense): For any countable \(T \subseteq \mathbb{R}\) we can construct an increasing function \(f_T : \mathbb{R} \to \mathbb{R}\) so that \(f_T\) is discontinuous at \(p\) iff \(p \in T\).

  Here’s how! Let \((t_n)_n\) be an enumeration of \(T\) (a bijection \(\mathbb{N} \to T\)). For any \(x \in \mathbb{R}\) set

  \[
  f_T(x) = \sum_{n \in \mathbb{N}} 2^{-n}.
  \]

  In the sum, \(n\) ranges over all indices \(n \in \mathbb{N}\) such that \(t_n < x\).

  This is well-defined because the sum on the right-hand side will always converge (by comparison with the convergent series \(\sum_{n \in \mathbb{N}} 2^{-n}\)). It is clearly increasing, since if \(x < x'\) we have \(\{t_n < x\} \subseteq \{t_n < x'\}\) and the terms of the series are all positive.

  One can show that \(f_T(t_n^+) - f_T(t_n^-) = 2^{-n}\) for all \(n \in \mathbb{N}\). Here’s how to show \(f_T\) is continuous at every \(p \notin T\): Let \(\epsilon > 0\) choose \(N\) large enough so that \(2^{-N} < \epsilon\), and choose \(\delta > 0\) small enough so that \(t_n \in (p - \delta, p + \delta)\) guarantees \(n > N\). If \(x \in (p - \delta, p + \delta)\) then \(f_T(x)\) and \(f_T(p)\) differ (in absolute value) by at most

  \[
  \sum_{n > N} 2^{-n} = 2^{-N} < \epsilon;
  \]

  hence \(f_T\) is continuous at \(p\).

5 Differentiation of real functions

We only did one lecture on derivatives (for now!). Our main purpose here is to prove the mean value theorem, which is an important step in the proof the fundamental theorem of calculus.
5.1 Definition and properties of the derivative

- **Definition.** Let $f : U \to \mathbb{R}$ be a real function on an open set $U \subseteq \mathbb{R}$ and let $p \in U$. We say that $f$ is **differentiable** at $p$ if the limit

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p}$$

exists. If it exists, we call it $f'(p)$, the **derivative** of $f$ at $x = p$. If $f$ is differentiable at every point of $U$, this yields a new function $f' : U \to \mathbb{R}$.63

- **Theorem.** If $f$ is differentiable at $p$, then it is continuous at $p$.64

- **Theorem.** The usual properties of the derivative you know and love from calculus are all true: Linearity, product rule, quotient rule, and chain rule.65

5.2 The critical points theorem and mean value theorem(s)

- **Definition.** Let $f : E \to \mathbb{R}$ be a function on some subset $E \subseteq \mathbb{R}$. We say that $f$ has a **local maximum** at $p \in E$ if there exists $\delta > 0$ such that $f(p) \geq f(x)$ for all $x \in (p - \delta, p + \delta)$. A local minimum is defined similarly. A local extremum is a local maximum or local minimum.66

- **Theorem.** (Critical points theorem.) If $f : U \to \mathbb{R}$ has a local extremum at $p \in U$ and $f$ is differentiable at $p$, then $f'(p) = 0$.67

*Proof.** We will assume that $f$ has a local maximum at $p$. Let $\delta > 0$ as in the definition of local maximum. Since $f(p) - f(x) \geq 0$ and $p - x \geq 0$ for all $x \in (p - \delta, p)$, we have $\frac{f(p) - f(x)}{p - x} \geq 0$ on this interval, so

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

is $\geq 0$. On the right half of $B_\delta(p)$, that is, on $(p, p + \delta)$ we have $f(x) - f(p) \leq 0$ and $x - p \geq 0$, so $\frac{f(x) - f(p)}{x - p} \leq 0$ and hence

$$f'(p) = \lim_{x \to p^+} \frac{f(x) - f(p)}{x - p}$$

is $\leq 0$. It follows that $f'(p) = 0$. $\square$

---

63J&P Definition 48.1.
64J&P Theorem 48.4.
65J&P Theorem 48.5 and 48.6.
66J&P Definition 49.1.
67J&P Theorem 49.3 and Corollary 49.4.
Theorem. (Rolle’s theorem.) If \( f : [a, b] \to \mathbb{R} \) is continuous and differentiable at every point in \((a, b)\), then if \( f(a) = f(b) \), there exists a \( p \in (a, b) \) such that \( f'(p) = 0 \).

Proof. By the extreme value theorem, \( f \) achieves a maximum and a minimum on \([a, b]\) say at \( p \) and \( q \), respectively. If \( \{a, b\} = \{p, q\} \) (that is, if the max/min both occur at endpoints), then \( f(p) = f(q) \) by hypothesis, so the function must be constant (as its max/min are identical). It is easy to show using the definition that the derivative of any constant function is 0, so \( f' = 0 \) on \((a, b)\).

Assume instead that \( p \) is not an endpoint (if \( q \) is not an endpoint the proof is similar). Then \( p \in (a, b) \) so \( f' \) is differentiable at \( p \). Since \( p \) is a global maximum for \( f \), it is also a local maximum, so by the critical points theorem, \( f'(p) = 0 \). □

Theorem. (Generalized mean value theorem.) Let \( f, g \) be continuous on \([a, b]\) and differentiable on \((a, b)\). There exists \( p \in (a, b) \) satisfying

\[
[f(b) - f(a)] \cdot g'(p) = [g(b) - g(a)] \cdot f'(p)
\]

Proof. Let \( h(x) = [f(b) - f(a)] \cdot g(x) - [g(b) - g(a)] \cdot f(x) \). This is continuous on \([a, b]\) and differentiable on \((a, b)\) (\( h \) is a linear combination of continuous/differentiable functions). It is easy to check that \( h(a) = h(b) \), so Rolle’s theorem applies: If \( h'(p) = 0 \), then rearranging gives the desired equality. □

Theorem. (Mean value theorem.) Let \( f \) be continuous on \([a, b]\) and differentiable on \((a, b)\). There exists \( p \in (a, b) \) satisfying

\[
f'(p) = \frac{f(b) - f(a)}{b - a}
\]

Proof. This is the generalized mean value theorem with \( g(x) = x \) (and some rearrangement). It can also be proven from Rolle’s theorem (which is a special case of this theorem with \( f(a) = f(b) \)) plus a shear transformation of the plane. □

6 Integration, part I (Riemann–Darboux theory)

(See separate notes on integration.)

\(^{68}\) J&P Theorem 49.5.

\(^{69}\) J&P Theorem 49.6.
7 Sequences and series of functions

7.1 Pointwise vs. uniform convergence

- Let $X$ and $Y$ be sets. A sequence of functions $X \to Y$ is a sequence $(f_n)_{n=0}^{\infty} = (f_0, f_1, f_2, \ldots)$ where every term is a function $f_n : X \to Y$.

- **Definition.** Let $X$ and $Y$ be metric spaces, let $(f_n)_{n=0}^{\infty}$ be a sequence of functions $X \to Y$ and let $f : X \to Y$. We say that $(f_n)_{n=0}^{\infty}$ converges to $f$ **pointwise** if for every $p \in X$, the sequence $(f_n(p))_{n=0}^{\infty}$ converges to $f(p)$ (in $Y$). That is, $(f_n)_{n=0}^{\infty}$ converges to $f$ pointwise if

$$\forall p \in X, \forall \epsilon > 0, \exists N, \forall n \geq N : d_Y(f_n(p), f(p)) < \epsilon$$

We simply write $f = \lim_{n \to \infty} f_n$ if $(f_n)_{n=0}^{\infty}$ converges pointwise.

- Pointwise convergence is, in some sense, the **wrong** notion of convergence for functions. This is because the “nice” properties of functions $f_n : X \to Y$ (boundedness, integrability, continuity, differentiability, etc.) may not be inherited by the pointwise limit $f : X \to Y$, as we will see in the following examples:

  - Let $f_n : [0, 1] \to \mathbb{R} : x \mapsto x^n$ (with $n \geq 1$). The sequence converges pointwise to the function

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

So, despite all terms of the sequence being continuous, the limit function is **not continuous**.

  - Let $f_n : [0, 1] \to \mathbb{R}$ be the function whose graph draws the legs of an isosceles triangle with base $[0, 2^{-n}]$ and height $2 \cdot 2^n$, with $f_n(x) = 0$ for any $x > 2^{-n}$. It is easy to check that this sequence converges pointwise to $f(x) = 0$. However, while

$$\int_0^1 f_n = \text{(area of a triangle with base } 2^{-n} \text{ and height } 2 \cdot 2^n) = 1$$

for all $n \geq 0$, we have $\int_0^1 f = 0$. So, while all $f_n$ are integrable and the pointwise limit $f = \lim_{n \to \infty} f_n$ is integrable,

$$\lim_{n \to \infty} \int_0^1 f_n \neq \int_0^1 \lim_{n \to \infty} f_n$$

  - Let $n \geq 1$ and let $f_n(x) = |x|^{1+1/n}$ on $[-1, 1]$. For $x \neq 0$, we have

$$f'_n(x) = (1 + \frac{1}{n}) |x|^{1/n} \sgn(x)$$

\[\text{J&P Definition 60.1.}\]
where \( \text{sgn}: \mathbb{R} \to \{-1, 0, 1\} \) is the *signum function* (tells you whether \( x < 0 \), \( x = 0 \), or \( x > 0 \)). Since the limit of the above exists and is zero as both \( x \to 0^- \) and \( x \to 0^+ \), \( f_n \) is differentiable at \( x = 0 \), with \( f'_n(0) = 0 \) for all \( n \geq 1 \). However,

\[
f(x) = \lim_{n \to \infty} f_n(x) = |x|
\]

which is famously non-differentiable at \( x = 0 \).

- Let \( (q_n)_{n=1}^\infty \) be an enumeration of the rational numbers in \([0, 1]\) and let \( f_n : [0, 1] \to \mathbb{R} \) be defined by
  \[
f_n(x) = \begin{cases} 
1 & \text{if } x = q_i \text{ for some } i \leq n, \\
0 & \text{otherwise}.
\end{cases}
\]

Then the pointwise limit is \( f(x) = 1_{Q \cap [0,1]} \), the Dirichlet function on \([0, 1]\). Since each one is nonzero at exactly \( n \) (finitely many) points, \( \int_0^1 f_n = 0 \) for all \( n \). However, \( \int_0^1 f \) does not exist because \( f \) is not Riemann integrable.

- The “right” definition of convergence for sequences of functions is *uniform convergence*. **Definition.** We say that \((f_n)_n \to f \) uniformly if
  \[
\forall \epsilon > 0, \exists N, \forall p \in X, \forall n \geq N : d_Y(f_n(p), f(p)) < \epsilon
\]
  In short: In the definition of piecewise convergence, \( N \) can depend on both \( p \) and \( \epsilon \). In uniform convergence, \( N \) can depend only on \( \epsilon \) (and the same \( \epsilon \) must work at every point). This is analogous to the difference between continuity and uniform continuity.

- It will turn out that if \((f_n)_n \to f \) uniformly, then \( f \) will typically inherit the nice properties of the functions in \((f_n)_n\). Furthermore, there is a sense in which this is the “natural” notion of convergence for functions.

- In the examples above, the sequence (or some other sequence?) does not converge uniformly. The second example fails uniform continuity for *every* \( \epsilon > 0 \). The first and fourth fail if \( 0 < \epsilon < 1 \).

The third is a bit trickier: \((f_n)_n \to f \) uniformly, but in fact \((f'_n)_n \) fails to converge (piecewise or uniformly). When we apply uniform convergent to differentiability, we will need both \((f_n)_n \to f \) and \((f'_n)_n \to f' \) to be uniform in order to guarantee that we can exchange the limit and derivative. In the third example, we have \((f'_n)_n \to \text{sgn} \) pointwise, but not uniformly.

### 7.2 Uniform convergence and the metric space of bounded functions

- For simplicity, let \( E \) be a subset of \( \mathbb{R} \). We denote by \( \mathcal{B}(E) \) the set of all *bounded* functions \( E \to \mathbb{R} \). This is an *algebra* over \( \mathbb{R} \): the set of bounded functions is closed under addition, multiplication, and (in particular) scaling by real constants.

- For any \( f \in \mathcal{B}(E) \) we define \( \|f\|_\infty = \sup_{x \in E} |f(x)| \). This is the \( \mathcal{L}^\infty \)-norm on \( \mathcal{B}(E) \):

\[71\text{J&P Theorem 60.3.} \]
i. $\|f\|_\infty = 0$ iff $f$ is the zero function,
ii. $\|\lambda f\|_\infty = |\lambda| \cdot \|f\|_\infty$ for any scalar $\lambda \in \mathbb{R}$,
iii. $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

- This gives the $L^\infty$-metric on $\mathcal{B}(E)$: $d(f, g) = \|f - g\|_\infty$.\(^{72}\)

- **Theorem.** Uniform convergence (of bounded functions $E \to \mathbb{R}$) is equivalent to convergence in the metric space $\mathcal{B}(E)$.\(^{73}\)

  **Proof.** ($\Rightarrow$) Suppose that $(f_n)_n \to f$ uniformly. Let $\epsilon > 0$. We can find $N$ such that

  $$|f_n(x) - f(x)| < \frac{\epsilon}{2}$$

  for all $x \in E$ and all $n \geq N$. Taking the supremum over all possible $x \in E$ yields

  $$\sup_{x \in E} |f_n(x) - f(x)| \leq \frac{\epsilon}{2} < \epsilon$$

  but the left-hand side is just $\|f_n - f\|_\infty$. Thus, we have found $N$ such that $n \geq N$ guarantees $\|f_n - f\|_\infty < \epsilon$, and so $(f_n) \to f$ in the metric space $\mathcal{B}(E)$.

  ($\Leftarrow$) Suppose that $(f_n)_n \to f$ in $\mathcal{B}(E)$. Let $\epsilon > 0$. We can find $N$ such that $n \geq N$ guarantees

  $$\|f_n - f\|_\infty = \sup_{x \in E} |f_n(x) - f(x)| < \epsilon$$

  from the second inequality, it follows that $|f_n(x) - f(x)| < \epsilon$ for all $x \in E$, and this proves convergence is uniform. \(\square\)

- **Theorem.** $\mathcal{B}(E)$ is a complete metric space.\(^{74}\)

  **Proof.** Suppose that $(f_n)_n$ is a Cauchy sequence in $\mathcal{B}(E)$. The first step is to show that $(f_n)_n$ converges to some $f \in \mathcal{B}(E)$ pointwise.

  - Cauchy sequences (in any metric space) are bounded, so $(f_n)_n$ is bounded. Therefore, there exists $M \geq 0$ such that $\|f_n\| \leq M$ for all $n$. It follows that $-M \leq f_n(x) \leq M$ for all $n$ and all $x \in E$.

  - From the fact that $(f_n)_n$ is Cauchy in $\mathcal{B}(E)$, we see that for each $p \in E$ the sequence $(f_n(p))_n$ is Cauchy in $\mathbb{R}$. Hence (since $\mathbb{R}$ is complete) $f(p) = \lim f_n(p)$ exists for every $p \in E$, and so $(f_n)_n$ converges pointwise to a function $f : E \to \mathbb{R}$.

  - Since $-M \leq f_n(p) \leq M$ for all $n$ and all $p \in E$, it follows that the same is true of the limit: $|f(p)| \leq M$ at every $p \in E$, so $\|f\|_\infty \leq M$ and this means that $f \in \mathcal{B}(E)$.

\(^{72}\)Rudin Definition 7.14; J&P Example 60.6, though they use lub for sup.
\(^{73}\)Rudin, discussion between 7.14 and 7.15; J&P Theorem 60.7.
\(^{74}\)Rudin Theorem 7.15; J&P Exercise 60.10.
Now we will show that \((f_n)_n \to f\) uniformly. Let \(\epsilon > 0\). By Cauchyness in \(\mathcal{B}(E)\), we can find \(N\) so that \(m, n \geq N\) implies \(\|f_m - f_n\|_\infty < \epsilon\). This implies that

\[|f_m(x) - f_n(x)| < \epsilon\]

Taking \(m \to \infty\) (and using pointwise convergence) we have \(|f(x) - f_n(x)| < \epsilon\) for all \(x \in E\) and all \(n \geq N\). Thus, \((f_n) \to f\) uniformly, and so it follows that \((f_n) \to f\) in \(\mathcal{B}(E)\). We have shown that every Cauchy sequence in \(\mathcal{B}(E)\) converges, so \(\mathcal{B}(E)\) is a complete metric space, as claimed. □

### 7.3 Uniform convergence and continuity

- **Theorem.** Let \(X\) and \(Y\) be metric spaces, let \(p \in X\), and suppose that \((f_n)_n\) is a sequence of functions \(X \to Y\) converging uniformly to some \(f : X \to Y\). If \(f_n\) is continuous at \(p\) for all \(n\), then \(f\) is continuous at \(p\).

**Strategy.** We need to show that we can force \(d_Y(f(p), f(x))\) to be small. The triangle inequality gives

\[
d_Y(f(p), f(x)) \leq d_Y(f(p), f_n(p)) + d_Y(f_n(p), f_n(x)) + d_Y(f_n(x), f(x))
\]

Uniform convergence will give us control over the first and third terms, while continuity of \(f_n\) will give us control over the second.

**Proof.** Let \(\epsilon > 0\). By uniform continuity, we can fix an index \(n\) so that \(d_Y(f(x), f_n(x)) < \frac{\epsilon}{3}\) for all \(x \in X\). By continuity of \(f_n\), we may choose \(\delta > 0\) so that \(d_Y(f_n(p), f_n(x)) < \frac{\epsilon}{3}\) whenever \(d_X(p, x) < \delta\). Combining these with the inequality in the strategizing above, we have, for all \(x \in X\)

\[
d_X(p, x) < \delta \Rightarrow d_Y(f(p), f(x)) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \quad \square
\]

- **Note:** In the proof above, it seems we only need \(f_n\) to be continuous at \(p\) for infinitely many \(n\) (?). However, the situation in applications is usually that all \(f_n\) are continuous, and the utility of the theorem above is proving that a uniform limit of continuous functions is also continuous.

### 8 Remainder of results, summarized and cited

#### 8.1 Uniform convergence and integrability/differentiability

- A *uniform* limit of (Riemann–Darboux-)integrable functions is integrable, and we can exchange the limit and the integral: If \((f_n)_n \to f\) uniformly and every \(f_n\) is integrable on \([a, b]\), then \(\lim_{n \to \infty} \int_a^b f_n = \int_a^b f\).\(^{76}\)

- If \((f_n)_n\) is a sequence of differentiable functions, \((f_n)_n \to f\) pointwise and \((f'_n)_n \to g\) uniformly, then \(f\) is differentiable and \(f' = g\).\(^{77}\)

\(^{75}\)J&P Theorem 60.4.

\(^{76}\)Rudin Theorem 7.16; J&P Theorem 61.1; both with \(\alpha = x\).

\(^{77}\)Rudin’s version is Theorem 7.17; J&P’s version is Theorem 61.2. Neither is used on the final exam.
8.2 Series of functions, Weierstrass $M$-test

- We also can define functions in terms of series: \( g = \sum_{n \geq 0} f_n \). In this case, we say that \( \sum f_n \) converges pointwise/uniformly to \( g \) if the sequence \( (g_n = \sum_{k=0}^{n} f_k)_n \) of its partial sums converges pointwise/uniformly to \( g \).\(^{78}\)

- A useful criterion for uniform convergence of series of functions is the Weierstrass $M$-test which says that if the summands \( f_n \) of a series of functions are dominated (in absolute value) by the summands of a convergent numerical series, then \( \sum f_n \) converges uniformly.\(^{79}\)

8.3 Power series and Taylor’s theorem

- A power series centered at \( \alpha \in \mathbb{R} \) is a function of the form
  \[
  F(x) = \sum_{n \geq 0} c_n (x - \alpha)^n
  \]
  where \( c_0, c_1, c_2, \ldots \in \mathbb{R} \). The domain of such a function is called the interval of convergence (it is always an interval). To determine the interval of convergence, one uses the ratio or root tests. Let \( I \) be the interval of convergence of the power series above.

  - Convergence is absolute in the interior \( I^\circ \) of the interval of convergence.
  - For every closed, bounded interval \( J \subseteq I^\circ \), convergence is uniform on \( J \).
  - It follows that \( F \) is continuous on \( I^\circ \): Uniform convergence on \( J \) implies that \( F \) is (uniformly) continuous on every closed bounded \( J \subseteq I^\circ \), and \( I^\circ = \bigcup_{J \subseteq I^\circ} J \), so \( F \) is continuous on \( I^\circ \) (though not necessarily uniformly so).
  - If \( J = [a, b] \) is a closed bounded subinterval of \( I^\circ \) then \( F \) is integrable on \([a, b]\) and integration can be performed “term-by-term”:
    \[
    \int_a^b F = \int_a^b \sum_{n=0}^{\infty} c_n (x - \alpha)^n \, dx = \sum_{n=0}^{\infty} c_n \int_a^b (x - \alpha)^n \, dx = \sum_{n=0}^{\infty} \frac{c_n (b^{n+1} - a^{n+1})}{n+1}
    \]
    - By the uniform-convergence-and-differentiability result, one can check that \( F'(x) = \sum_{n \geq 1} n c_n (x - \alpha)^{n-1} \). By the ratio or root tests, the power series that define \( F \) and \( F' \) have the same radii of convergence. It follows (by induction) that \( F \) is in fact smooth (infinitely many times differentiable) on \( I^\circ \).\(^{80}\)

- Let \( f \) be a real function that is smooth at \( \alpha \). The Taylor series for \( f \) at \( \alpha \) is
  \[
  T_{f, \alpha}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\alpha)}{n!} (x - \alpha)^k
  \]
  where \( f^{(k)} \) is the \( n \)th derivative of \( f \). The \( n \)th Taylor polynomial is \( T_{f, \alpha, n}(x) \) is the \( n \)th partial sum of the above (the sum from \( k = 0 \) to \( k = n \)).

\(^{78}\)J&P Definition 62.1.
\(^{79}\)J&P Theorem 62.6.
\(^{80}\)These are covered by Rudin Theorem 8.1 and J&P in Section 63.
Taylor’s theorem (aka the Lagrange bounds for Taylor series)\(^{81}\) give a bound on the error in Taylor polynomial approximations: For any \(p \in \mathbb{R}\),
\[
|T_{f,a,n}(p) - f(p)| \leq \frac{\sup_{x \in J} |f^{(n+1)}(x)|}{(n + 1)!} |p - \alpha|^{n+1}
\]
where \(J\) is the closed interval between \(\alpha\) and \(p\) (so either \([\alpha, p]\) or \([p, \alpha]\)).

If \(f = T_{f,a}\) on some nondegenerate interval around \(\alpha\), we say that \(f\) is real analytic at \(\alpha\). You saw an example of a non-real analytic function on the homework (the \(e^{-1/x}\) example).

### 8.4 The Stone–Weierstrass theorem

- Let \(X\) be a metric space and let \(\mathcal{C}(X)\) be the metric space of bounded, continuous, real-valued functions on \(X\) where the metric is \(d(f, g) = \|f - g\|_\infty = \sup_{x \in X} |f(x) - g(x)|\). The resulting topology is sometimes called the topology of uniform convergence or the \(L_\infty\)-topology.

- The Stone–Weierstrass theorem gives sufficient conditions for a set to be dense in \(\mathcal{C}(X)\) when \(X\) is compact.
  - A unital subalgebra of \(\mathcal{C}(X)\) is a subset that is closed under addition and multiplication of its elements (\(f, g \in A\) guarantees \(f + g, f \cdot g \in A\)), closed under multiplication by real scalars (\(f \in A\) and \(c \in \mathbb{R}\) guarantees \(cf \in A\)), and which satisfies \(1 \in A\).
  - \(A \subseteq \mathcal{C}(X)\) separates points if for all distinct \(p, q \in X\) we can find \(f \in A\) such that \(f(p) \neq f(q)\). (It is sufficient, but not necessary, that \(A\) contain an injective function.)

  The Stone–Weierstrass theorem says that if \(X\) is compact and \(A\) is a unital subalgebra of \(\mathcal{C}(X)\) that separates points, then \(A\) is dense in \(\mathcal{C}(X)\).\(^{82}\)

- The classical version of the theorem (Weierstrass’ approximation theorem\(^{83}\)) says that polynomials are dense in \(\mathcal{C}([a, b])\). This amounts to saying that for any continuous function \(f : [a, b] \to \mathbb{R}\), there is a sequence of polynomials \((f_n)_n\) that converges to \(f\) uniformly. That is, every continuous function on a closed bounded interval can be arbitrarily well-approximated by polynomials. It is easy to check that Stone–Weierstrass implies Weierstrass approximation, but it also implies more advanced results that relate to Fourier theory.

\(^{81}\)Rudin Theorem 5.15.
\(^{82}\)Rudin Theorem 7.32.
\(^{83}\)J&P Corollary 77.10; Rudin Theorem 7.26.
8.5 The Takagi function

- The Takagi function is a classic example of a function $\mathbb{R} \to \mathbb{R}$ that is continuous everywhere but differentiable nowhere. It is defined as follows: Let $t(x)$ be the function that measures the distance from $x$ to the nearest integer ($y = t(x)$ graphs a “sawtooth wave”). Let $t_n(x) = t(2^n x) \cdot 2^{-n}$ and let $T(x) = \sum_{n \geq 0} t_n(x)$.

- By the Weierstrass $M$-test (with $M_n = 2^{-n-1}$), $\sum t_n$ converges to $T$ uniformly, so $T$ is continuous on $\mathbb{R}$.

- The proof I presented that $T$ is not differentiable anywhere is essentially that of “The Takagi function: a survey” by Pieter C. Allaart and Kiko Kawamura (Theorem 2.1).

8.6 Measure spaces

- A measure space is a triple $(X, \Sigma, \mu)$ where $X$ is a set (the space), $\Sigma$ is a $\sigma$-algebra of subsets of $X$ (the measurable sets), and $\mu : \Sigma \to [0, +\infty]$ is a function (the measure), satisfying (i) $\mu(\emptyset) = 0$ and (ii) for all sequences $(E_n)_n$ in $\Sigma$ that satisfy $E_i \cap E_j = \emptyset$ whenever $i \neq j$, $\mu(\bigcup_n E_n) = \sum_n \mu(E_n)$.

  - A $\sigma$-algebra is a set of sets that contains $\emptyset$, is closed under complementation, and is closed under countable union (and therefore also countable intersection, by DeMorgan’s laws).

- Consequences of the definition:
  
  - If $A$ and $B$ are measurable and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
  
  - Generally for any sequence $(E_n)_n$ of measurable sets, we have $\mu(\bigcup E_n) \leq \sum \mu(E_n)$ (this is (ii) without the assumption of pairwise disjointness).
  
  - If $E_1 \subseteq E_2 \subseteq \cdots$ is an increasing sequence of measurable sets, then $\mu(\bigcup_n E_n) = \lim \mu(E_n) = \sup\{\mu(E_n)\}_n$.
  
  - If $E_1 \supseteq E_2 \supseteq \cdots$ is a decreasing sequence of measurable sets, and $\mu(E_k) < +\infty$ for at least one $k$, then $\mu(\bigcap_n E_n) = \lim \mu(E_n) = \inf\{\mu(E_n)\}_n$.

- Special kinds of measure spaces: A probability space is a measure space where $\mu(X) = 1$. A $\sigma$-finite measure space is one where $X$ is a countable union of subsets having finite measure ($\mathbb{R}$ is a $\sigma$-finite measure space).

- The Lebesgue measure on $\mathbb{R}$ is defined as follows. For any $S \subseteq \mathbb{R}$, we define $\lambda^*(S) = \inf \sum_{n=1}^\infty \text{length}(I_n)$ where the infimum is taken over all sequences $(I_n)_n$ of open intervals (possibly empty) that cover $S$ (that is, $S \subseteq \bigcup_n I_n$). This is the outer Lebesgue measure of $S$. It turns out that the outer Lebesgue measure is a true measure on $\mathbb{R}$.

\[
\Sigma = \{ S \subseteq \mathbb{R} : \lambda^*(S) = \lambda^*(S \cap A) + \lambda^*(S \setminus A) \text{ for any } A \subseteq \mathbb{R} \}
\]

And so $(\mathbb{R}, \Sigma, \lambda = \lambda^*|_{\Sigma})$ is a measure space.

\[^{84}\text{J&P discussion after 62.6.}\]
• The equation in the set is called Carathéodory’s condition.

• One can check that for all $p \in \mathbb{R}$ the set $\{ p \}$ is Lebesgue-measurable with measure zero, and it follows that all countable subsets of the reals have Lebesgue measure zero. (The converse is not true, since the Cantor set is uncountable yet measure zero.)

• An example of a non-Lebesgue measurable set is the Vitali set presented in class. The existence of non-measurable sets is related in a deep way with the axiom of choice—we used AC to construct the Vitali set, and there are versions of set theory in which all subsets of the real numbers are measurable.

8.7 Measurable functions and integration

• Let $(X, \Sigma, \mu)$ be a measure space. A measurable function $f : X \to \mathbb{R}$ is a function such that for all $a \in \mathbb{R}$, the set $\{ f > a \} = \{ x \in X : f(x) > a \}$ is measurable (that is, in $\Sigma$). We proved that replacing $\{ f > a \}$ with $\{ f \geq a \}, \{ f < a \}, \{ f \leq a \}$, or $f^{-1}(I)$ (for all intervals) in the definition give equivalent conditions.\footnote{Rudin Theorem 11.15.}

• A (nonnegative) simple function on $X$ is one that can be written as a finite nonnegative linear combination of characteristic functions of measurable sets: $s = \sum_{i=1}^{n} c_i \cdot 1_{E_i}$ where $c_i \geq 0$ and $E_i \in \Sigma$.

Simple functions are analogous to step functions in our treatment of Riemann–Darboux integration.

• Easy: Every simple function is measurable.

• The integral of the (nonnegative) simple function above (on $X$ with respect to $\mu$) is defined by

$$\int_{X} s \, d\mu = \sum_{i=1}^{n} c_i \mu(E_i).$$

And the integral of $s$ on $E$ (a measurable set) is $\int_{E} s \, d\mu = \int_{X} s \cdot 1_E \, d\mu$ (the integrand here is also simple, since $1_E \cdot 1_{E_i} = 1_{E \cap E_i}$).

• If $f$ is measurable and $f \geq 0$ we define $\int_{X} f \, d\mu$ by

$$\int_{X} f \, d\mu = \sup_{0 \leq s \leq f, s \text{ simple}} \int_{X} s \, d\mu$$

and $\int_{E} f \, d\mu$ (for measurable $E$) can either be defined as the supremum over $\int_{E} s \, d\mu$ or as $\int_{E} f \, d\mu = \int_{X} f \cdot 1_{E} \, d\mu$ (formally one must prove that these give the same definition).

• Using the equivalent definitions of measurable, one can easily show the following: If $f$ is measurable, then $|f|$ is measurable.\footnote{Rudin Theorem 11.16.} If $(f_n)_{n}$ is a uniformly bounded sequence
of functions (that is there is $M$ such that $|f_n(x)| \leq M$ for all $n$ and all $x$), then $f = \sup(f_n)_n$ is measurable.\footnote{Rudin Theorem 11.17.}

Furthermore, 0 is a measurable function. As a consequence, we conclude that if $f$ is measurable then so are

$$f^+ = \max(f, 0) \quad \text{and} \quad f^- = -\min(f, 0)$$

Note that $f = f^+ - f^-$ and $f^+, f^- \geq 0$.

- For a measurable function $f : X \to \mathbb{R}$ and a measurable set $E$,
  - If $\int_E f^+ \, d\mu$ and $\int_E f^- \, d\mu$ are both finite, we set
    $$\int_E f \, d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu$$
    and say that $f$ is integrable on $E$ with respect to $\mu$.
  - If exactly one of $\int_E f^+ \, d\mu$ or $\int_E f^- \, d\mu$ is infinite, we set $\int_E f \, d\mu = \pm\infty$ accordingly and say that $f$ is non-integrable (the integral takes an infinite value).
  - If both are infinite we do not define the integral, since $\infty - \infty$ is indeterminate.

- The integral as defined above satisfies:\footnote{Rudin, Remark 11.23.}
  - If $f \leq g$ on $E$ and both functions are integrable on $E$, we have $\int_E f \, d\mu \leq \int_E g \, d\mu$ (the integral is monotonically non-decreasing).
  - $\int_E cf \, d\mu = c \int_E f \, d\mu$ (part of, but conspicuously not all of, linearity).
  - If $\mu(E) = 0$ then $\int_E f \, d\mu = 0$.
  - If $f \geq 0$, then $\phi(E) = \int_E f \, d\mu$ is also a measure on $(X, \Sigma)$. In fact, in analogy with the fundamental theorem of calculus, we write $f = \frac{d\phi}{d\mu}$ in this case.\footnote{Rudin Theorem 11.24; J&P Theorem 88.8.}

- In the very last lecture, we will prove the last bullet point above and then use it to prove Lebesgue’s \textit{monotone convergence theorem}.\footnote{Rudin Theorem 11.28; J&P Theorem 88.6.} This allows us to finish the linearity proof for the integral above (weird!).\footnote{Rudin Theorem 11.29; J&P Lemma 88.7.} We will also state Lebesgue’s \textit{dominated convergence theorem} and a couple of consequences, time permitting.

- Remember that the Promises of Lebesgue Integration were that it (i) Would integrate more functions; (ii) Would generalize integration to functions on domains other than (subsets of Cartesian products of) $\mathbb{R}$, i.e., \textit{measure spaces}; and (iii) Allow us to exchange integrals and limits more easily. It does all three of these!
It also (iv) agrees with the Riemann–Darboux integral if \( f \) is R–D integrable.\(^{93}\)

Finally, (v) If \( f : [a, b] \to \mathbb{R} \), then \( f \) is Riemann–Darboux integrable on \([a, b]\) iff \( \lambda(\{ p \in [a, b] : f \text{ is discontinuous at } p \}) = 0.\(^{94}\)

\(^{93}\)Rudin Theorem 11.33a; J&P Theorem 90.3.

\(^{94}\)Rudin Theorem 11.33b.