MATH 320/321 (Real Analysis) Notes

Rio Weil

This document was typeset on July 25, 2021

Introduction:
This set of notes is transcribed from UBC's MATH 320/321 (Real Variables I/II) sequence. The course covers the first 9 chapters of Rudin's "Principles of Mathematical Analysis" with occasional omissions & additions. The numbering of the definitions/theorems/examples will follow that used in Rudin for convenience. The structure of these notes is such that they are split into main text (the boxed elements) and side text (everything else). It is possible to solely read the main text for all of the material, but the additional discussion provided by the side text may be useful. If any errors are found in the notes, feel free to email me at `ryoweil6@student.ubc.ca`.

Contents

1 The Real and Complex Number Systems

1.1 The Naturals, Integers, and Rationals

We begin by a review of number systems which are already familiar.

Definition: The Natural Numbers

The **Naturals**, denoted by $\mathbb{N}$, is the set $\{1, 2, 3, \ldots\}$.

For $x, y, \in \mathbb{N}$, we have that $x + y \in \mathbb{N}$ and $xy \in \mathbb{N}$, so the naturals are closed under addition and multiplication. However, we note that it is not closed under subtraction; take for example $2 - 4 = -2 \notin \mathbb{N}$.

Definition: The Integers

The **Integers**, denoted by $\mathbb{Z}$, is the set $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

The integers are closed under addition, multiplication, and subtraction. However, it is not closed under division; for example, $1/2 \notin \mathbb{Z}$.

Definition: The Rationals (informal)

The **Rationals**, denoted by $\mathbb{Q}$, can be defined as $\left\{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{N}\right\}$, where $\frac{m_1}{n_1}$ and $\frac{m_2}{n_2}$ are identified if $m_1 n_2 = m_2 n_1$.

We note that unlike the naturals/integers, the rationals do not have as obvious of a denumeration. This above is a good definition if we already have the same rigorous idea of what a rational number is in our mind; i.e. it works because we have a shared preconceived understanding of a rational number.

If this is not the case, it may help to define the rational numbers more rigorously/formally (even if the definition may be slightly harder to parse). As a second attempt at a definition, we can say that $\mathbb{Q}$ is the set of ordered pairs $\{(m, n) : m \in \mathbb{Z}, n \in \mathbb{N}\}$. However, this is not quite enough as we need a notion of equivalence between two rational numbers (e.g. $(1, 2) = (2, 4)$). Hence, a complete and rigorous definition would be:

Definition: The Rationals (formal)

The **Rationals**, denoted by $\mathbb{Q}$, is the set $\{(m, n) : m \in \mathbb{Z}, n \in \mathbb{N}\} / \sim$ where $(m_1, n_1) \sim (m_2, n_2)$ if $m_1 n_2 = m_2 n_1$.

Under the formal definition, the rationals are a set of equivalence classes of ordered pairs, under the equivalence relation $\sim$. We note that the rationals are closed under addition, subtraction, multiplication, and division.

This formal definition might be slightly harder to parse, so it might be useful to consider an example with a similar flavour. Consider the set $X = \{m \in \mathbb{Z}\} / \sim$ such that $m_1 \sim m_2$ if $m_1 - m_2$ is divisible by 12. This is "clock arithmetic", with equivalence classes $[0], [1], [2], \ldots$ for each hour on an analog clock. A fun side note: If instead of 12 we picked a prime number, we would get a field (we will discuss what this is in a later lecture)!

Note that under this definition, $(1, 2)$ and $(2, 4)$ are different representations of the same rational number. With this definition, we would define addition such that $(m_1, n_1) + (m_2, n_2) = (m_1 n_2 + m_2 n_1, n_1 n_2)$. Note that $(2m_1, 2n_2) + (m_2, n_2) = (2m_1 n_2 + 2m_2 n_1, 2n_1 n_2)$ and we can identify $(m_1 n_2 + m_2 n_1, n_1 n_2)$ with $(2m_1 n_2 + 2m_2 n_1, 2n_1 n_2)$. If we choose different representations when we do addition, we might get a different representation in our result, but it will represent the same rational number regardless of the choice of representations we originally chose to do the addition.

A natural question then becomes if the rationals are sufficient for doing all of real analysis. Certainly, it seems as we have a number system that is closed under all our basic arithmetic operations; but is this enough? For example, are we able to take limits just using the rationals? The answer turns out to be no (they are insufficient!) and the following example will serve as one illustration of this fact.

Example 1.1: Incompleteness of the Rationals

There exists no $p \in \mathbb{Q}$ such that $p^2 = 2$.

We proceed via proof by contradiction. Recall in that these types of proof, we start with a certain wrong assumption, follow a correct/true line of reasoning, reach an eventual absurdity, and therefore conclude that the original assumption was mistaken.

Proof

Let us then suppose for the contradiction that there exists $p = \frac{m}{n}$ with $p^2 = 2$. We then have that not both m, n are even, and hence at least one is odd. Then, we have that $2 = p^2 = \frac{m^2}{n^2}$ and hence $m^2 = 2n^2$, so m^2 is even, implying m is even. So, let us write $m = 2k$ for $k \in \mathbb{Z}$. Then, $(2k)^2 = 4k^2 = 2n^2$, and hence $2k^2 = n^2$. Therefore, n^2 is even and hence n is even. m and n are therefore both even, a contradiction. We conclude that no such p exists. □

Why can we conclude that not both m, n are even in the above proof? This is the case as if m, n we both even, then we could write $m = 2m'$, $n = 2n'$ for some m', n', and then $p = \frac{m}{n} = \frac{2m'}{2n'} = \frac{m'}{n'}$ which we can continue until either the numerator or denominator is odd. A natural question to consider is how to prove that this process of reducing fractions will eventually conclude. The resolution is to invoke the fundamental theorem of arithmetic, and write m, n in terms of their unique prime factorization. We are then able to cancel out factors of 2 from the numerator/denominator until at least one is odd.

We note that this example leads us to conclude that the rationals have certain "holes" in them. This is concerning, as there are sequences of rational numbers that tend to $\sqrt{2}$. Conversely, its not as concerning that there is no rational number x such that $x^2 = -1$, as there is no such sequence of rational numbers that is "close to" i (note that both $\sqrt{2}$ and i have not yet been defined, but this will come shortly).

Example 1.1: Incompleteness of the Rationals

Let $A = \left\{p \in \mathbb{Q} : p > 0, p^2 < 2\right\}$, and $B = \left\{p \in \mathbb{Q} : p > 0, p^2 > 2\right\}$. Then, $\forall p \in A, \exists q \in A$ such that $p < q$, and $\forall p \in B, \exists q \in B$ such that $q < p$.

Figure 1: Visualization of sets A and B. We note that $\sqrt{2}$ has not been defined in our formalism yet, but from our prior mathematical intuition it would be what goes in the "hole" of the rationals.

For the proof of this statement, we consider playing a 2 person game. One person is $\forall$, one person is $\exists$, and we consider if one person has a winning strategy. $\forall$ goes first, and then $\exists$ goes next, having seen the choice that $\forall$ has made. Then, we check if indeed $p < q$. If $p < q$, then $\exists$ wins. If $p \not< q$, then $\forall$ wins.

Proof

Let $p \in A$. Then, let $q = \frac{2p+2}{2+p}$. Since $p \in \mathbb{Q}$, it follows that $2p+2 \in \mathbb{Q}$ and $2+p \in \mathbb{Q}$ so $q \in \mathbb{Q}$. Furthermore, we have that $2p+2 > 0$ and $2+p > 0$, so $q > 0$. We also have that:

$$q^2 = \frac{(2p+2)^2}{(2+p)^2} = 2 + \frac{2(p^2-2)}{(p+2)^2} < 2$$

Where the inequality follows from the fact that $p^2 < 2$ and hence $(p^2-2) < 0$. It therefore follows that $q \in A$. Finally, we have that:

$$q = p + \frac{2-p^2}{2+p} > p$$

so $q > p$, completing the proof of the first part of the claim. The second part is left as an exercise (we note that the same q can be used). □

The number $q = \frac{2p+2}{2+p}$ seems to be pulled out of a hat, but actually comes from a fairly geometric picture (the secant method of approximating roots). Discussion on this topic can be found here: `https://math.stackexchange.com/questions/141774/choice-of-q-in-baby-rudins-example-1-1`.

1.2 Ordered Sets

Over the next couple sections, we will be discussing certain properties of sets that will give us a better understanding of the real numbers, and allow us to construct them.

Definition 1.5: Order

An **order** $<$ on a set S is a relation with the following properties:

(i) For every pair $x, y \in S$, exactly one of $x < y$, $x = y$, or $y < x$ is true.

(ii) For $x, y, z \in S$, if $x < y$ and $y < z$, then $x < z$.

A point on notation; We note that $x > y$ means $y < x$, and $x \leq y$ means $x < y$ or $x = y$.

Definition 1.6: Ordered Sets

An **ordered set** is a pair $(S, <)$. We may write just S if the order can be inferred by the context.

A familiar (and useful) set of examples is $S = \mathbb{N}$ or $S = \mathbb{Z}$ or $S = \mathbb{Q}$. For these three sets, we have that $x < y$ if $y - x$ is positive. For another example, consider the set S of english words; then the order $<$ can be the dictionary/lexographic order.

Definition 1.7: Upper & Lower Bounds

Let S be an ordered set and $E \subset S$ (for the duration of these notes, we will follow Rudin's notation, with $E \subset S$ as a non-strict subset, and $E \subsetneq S$ as a strict subset). E is **bounded above** if there exists an element $\beta \in S$ such that $\forall x \in E$, $x \leq \beta$. Any such β is an **upper bound** of E. Similarly, we say that E is **bounded below** if there exists an element $\alpha \in S$ such that $\forall x \in E$, $\alpha \leq x$. In this case, α is a **lower bound** of E.

As an example, one can take $S = \mathbb{Q}$, $E = A = \left\{p \in \mathbb{Q} : p > 0, p^2 > 2\right\}$ (as in Example 1.1(b)). Here, E is

bounded above, with $\beta = 2$ as one possible upper bound. to see this is the case, consider that if $p \in E$:

$$2 - p = \frac{4 - p^2}{2 + p} > \frac{4 - 2}{2 + p} > 0$$

However, if we take $S = A$, $E = A$, then E is not bounded above as we saw in the example. There is no upper bound of A in A. In general, this example reveals the subtle point that "the upper bound of a set" is ill-defined; we need to specify $E \subset S$.

1.3 The Least Upper Bound Property

Definition 1.8: Least Upper Bound & Greatest Lower Bound

Let S be an ordered set, and let $E \subset S$ with E bounded above. If $\exists \alpha \in S$ such that:

(i) α is an upper bound for E

(ii) If $\gamma < \alpha$, then γ is not an upper bound for E

The α is the **least upper bound**, or **supermum** of E. This can be denoted as $\alpha = \sup(E)$. Analogously, the **greatest lower bound**, or **infimum** of E (denoted $\alpha = \inf(E)$) is an element $\alpha \in S$ (if it exists) such that:

(i) α is a lower bound for E

(ii) If $\gamma > \alpha$, then γ is not an upper bound of E.

Theorem

If the supremum/infimum of $E \subset S$ exist, they are unique.

Proof

Let $E \subset S$. Suppose that there exist α_1, α_2 such that $\alpha_1 = \sup(E)$ and $\alpha_2 = \sup(E)$. If $\alpha_1 < \alpha_2$, as α_1 is an upper bound of E, this contradicts the fact that α_2 is the least upper bound of E. We reach an identical contradiction if $\alpha_2 < \alpha_1$. Therefore we conclude that $\alpha_1 = \alpha_2$ and the supremum of E is unique (if it exists). The proof for the infimum is analogous. □

Theorem

If $E \subset S$ has a maximum element α (that is, an element such that $x < \alpha$ for all $x \in E$) then $\alpha = \sup(E)$. Similarly, if E has a minimum element α, then $\alpha = \inf(E)$.

Proof

Let $E \subset S$ and $\alpha = \max(E)$. By definition α is an upper bound of E, and if $x < \alpha$ for some $x \in E$ then x is not an upper bound of E as it is not greater than $\alpha \in E$. The claim follows (with an identical proof for the minimum). □

Example 1.9

(a) Consider again the sets $A, B \subset \mathbb{Q}$ from example 1.1. A is bounded above by any element in B, and the upper bounds of A are exactly the elements of B. Since B has no smallest member, A does not have a least upper bound in $\mathbb{Q}$.

(b) Let $E_1, E_2 \subset \mathbb{Q}$ such that $E_1 = \{r : \mathbb{Q}, r < 0\}$ and $E_2 = \{r : \mathbb{Q}, r \leq 0\}$. Then $\sup(E_1) = \sup(E_2) = 0$. Note that this example shows that the supremum can either be contained or not contained in the set; $0 \notin E_1$ but $0 \in E_2$.

(c) Let $E \subset \mathbb{Q}$ such that $E = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. Then $\sup(E) = 1$ and $\inf(E) = 0$. This is proven below.

Proof

$\sup(E) = 1$ immediately follows from the equivalence of the maximum and supremum as proven above. To see that $\inf(E) = 0$, first note that 0 is a lower bound for E as all of the elements of E are positive. To see that it is the lower bound, take any $x > 0$. Then, we have that for any $n > \frac{1}{x}$, $\frac{1}{n} < x$ and hence x is not an upper bound of E. This proves the claim. □

Definition 1.10: The LUB/GUB Property

An ordered set S has the **least upper bound property** if for every $E \subset S$, if $E \neq \emptyset$ and E is bounded above, then E has a least upper bound (that is, $\sup(E)$ exists in S). Similarly, an ordered set S has the **greatest lower bound property** if for every $E \subset S$, if $E \neq \emptyset$ and E is bounded below, then E has a greatest lower bound.

We will show in the next theorem that these properties are actually equivalent; before then, we briefly consider two examples.

Example

$\mathbb{Z}$ has the least upper bound property, while $\mathbb{Q}$ does not.

Proof

For the first claim, consider any nonempty $E \subset \mathbb{Z}$ that is bounded above. Choose any $x \in E$. Since $\mathbb{Z}$ is bounded above, there exist finitely many elements that are greater than x. Take the maximum of these finitely many elements. This maximum is also the maximum of E, so it is the supremum of E. Therefore $\mathbb{Z}$ has the LUB property as claimed.
The second claim immediately follows from Example 1.9(a). □

Theorem 1.11

Let S be an ordered set. Then S has the LUB property if and only if it has the GUB property.

Proof

$\boxed{\Longrightarrow}$ Let S be an ordered set with the LUB property. Let $E \subset S$ with $E \neq \emptyset$, with E bounded below. Let $L = \{x \in S : x \text{ is a lower bound of } E.\}$. $L \neq \emptyset$ as E is bounded below (and hence has at least one lower bound). If $y \in E$, then y is an upper bound for L. Since E is nonempty, L is therefore bounded above. Since S has the LUB property, then $\sup(L)$ must exist. Let us call this α. Then, $\alpha \leq x \ \forall x \in E$ (as if $\gamma < \alpha$, then γ is not an upper bound of L and hence $\gamma \neq E$). Hence, α is a lower bound for E and hence $\alpha \in L$. Since $\alpha = \sup(L)$ and α is an upper bound for L, we have that $\alpha \geq \gamma \ \forall \gamma \in L$. Thus, $\alpha = \inf(E)$.
$\boxed{\Longleftarrow}$ Left as an exercise. □

1.4 Fields and Ordered Fields

Definition 1.12: Fields

A **field** F is a set with two binary operations, $+$ and $\cdot$ (addition and multiplication) such that the following axioms are satisfied:

(A1): If $x, y \in F$, then $x + y \in F$. (Closure under addition)

(A2): $x + y = y + x$ for all $x, y \in F$. (Commutativity of addition)

(A3): $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$. (Associativity of addition)

(A4): $\exists 0 \in F$ such that $\forall x \in F$, $0 + x = x$. (Additive identity)

(A5): $\forall x \in F$, $\exists y$ such that $x + y = 0$. We can denote $y = -x$. (Additive inverse)

(M1): If $x, y \in F$, then $x \cdot y \in F$. (Closure under multiplication)

(M2): $x \cdot y = y \cdot x$ for all $x, y \in F$.

(M3): $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in F$. (Associativity under multiplication)

(M4): $\exists 1 \in F$ such that $1 \neq 0$ and $\forall x \in F$, $1 \cdot x = x$. (Multiplicative identity)

(M5): $\forall x \in F$, exists $y \in F$ such that $x \cdot y = 1$. We can denote $y = \frac{1}{x}$. (Multiplicative inverse)

(D): $x \cdot (y + z) = x \cdot y + x \cdot z$, $\forall x, y, z \in F$. (Distributive law)

Note that A3/M3 show that $x + y + z$ and $x \cdot y \cdot z$ are well defined in a mathematical sense; however, associativity may not hold for computers that do math with finite precision!

Theorem

The additive/multiplicative identities given by (A4)/(M4) and the additive/multiplicative inverses given by (A5)/(M5) are unique.

Proof

Let F be an ordered field. Suppose that there exist $0_1, 0_2 \in F$ such that $0_1 + x = x$ and $0_2 + x = x$ for all $x \in F$. We then have that:

$$\begin{aligned} 0_1 + 0_2 &= 0_1 + 0_2 \\ 0_1 + 0_2 &= 0_2 + 0_1 && \text{(A2)} \\ 0_2 &= 0_1 && \text{(Property of additive identity)} \end{aligned}$$

Which shows that the additive identity is unique. The remaining proofs are left as an exercise. □

Some easy (and familiar) consequences of the field axioms can be found in Rudin 1.14-1.16. Instead of repeating those here, we will discuss some examples.

The rationals form a field (under the usual notions of addition/multiplication), but the integers do not, as there are no multiplicative inverses (e.g. there exists no integer $x \in \mathbb{Z}$ such that $2 \cdot x = 1$). The simplest example of a field is $F = \{0, 1\}$, with the relations:

$$\begin{aligned} 0 + 0 &= 0 \quad 0 \cdot 0 = 1 \\ 0 + 1 &= 0 \quad 0 \cdot 1 = 0 \\ 1 + 1 &= 0 \quad 1 \cdot 1 = 1 \end{aligned}$$

This field is often called $\mathbb{F}_2$ or F_2, and is useful in computer science (where bits can take on two states, 0 or 1). As a slight tangent, a byte (8 bits) can be considered an element of an 8-dimensional vector space over the field $\mathbb{F}_2$, where $+$ would be the XOR operator and $\cdot$ would be the AND operation.

A generalization of the above example is $\mathbb{F}_p$ or F_p, for a prime number p. This field would consist of the elements $0, 1, \ldots, p-1$. The addition and multiplication are carried out mod p. An interesting result is that in general, finite fields must have cardinality of some prime power.

Note that a field cannot have a single element; the field axioms (A4) and (M4) require the existence of distinct additive and multiplicative identities, which a singleton set cannot satisfy.

Although algebra is not the focus of this course, it may be interesting to briefly think about sets with less structure than a field. We start by considering a group.

A **group** G is a set with a binary operation $(a, b) \mapsto a \cdot b$ such that the following axioms are satisfied:

(M1): If $a, b \in G$, then $a \cdot b \in G$ (Closure)

(M3): For $a, b, c \in G$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (Associativity)

(M4): There exists $1 \in G$ such that $\forall x \in G$, $1 \cdot x = x$. (Identity)

(M5): $\forall x \in G$, there exists $y \in G$ such that $x \cdot y = 1$. (Inverse)

We note that $\mathbb{Z}$ is a group under addition, but not under multiplication (due to lack of multiplicative inverses). We can also consider the set of 2x2 matrices with integer entries:

$$G = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{Z} \right\}$$

G is again a group under matrix addition, but not under matrix multiplication (as not every matrix in G is invertible). If we restricted G to be the set of 2×2 invertible matrices, in this case it could form a group under matrix multiplication. A set with slightly more structure than a group (though not quite as structured as a field) is a ring:

A **ring** R is a set with two binary operations $(a,b) \mapsto a+b$ and $(a,b) \mapsto a \cdot b$ such that the following axioms are satisfied:

(A1): If $x, y \in R$, then $x+y \in R$. (Closure under addition)

(A2): $x+y = y+x$ for all $x, y \in R$. (Commutativity of addition)

(A3): $(x+y)+z = x+(y+z)$ for all $x, y, z \in R$. (Associativity of addition)

(A4): $\exists 0 \in R$ such that $\forall x \in R$, $0+x = x$. (Additive identity)

(A5): $\forall x \in R$, $\exists y$ such that $x+y = 0$. We can denote $y = -x$. (Additive inverse)

(M1): If $x, y \in R$, then $x \cdot y \in R$. (Closure under multiplication)

(M3): $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for all $x, y, z \in R$. (Associativity under multiplication)

(M4): $\exists 1 \in R$ such that $1 \neq 0$ and $\forall x \in R$, $1 \cdot x = x$. (Multiplicative identity)

(D1): $x \cdot (y+z) = x \cdot y + x \cdot z$, $\forall x, y, z \in R$. (Left distributivity)

(D2): $(y+z) \cdot x = y \cdot x + z \cdot x$, $\forall x, y, z \in R$. (Right distributivity)

Rings have the same axioms as fields under addition, but multiplication is not necessarily commutative (this is why an additional distributivity axiom is added), and multiplicative inverses are not required. We note that $\mathbb{Z}$ and G are both rings under their respective operations of addition and multiplication.

For the remainder of this course, we will really only be discussing fields; however, they will be the objects of interest in abstract algebra courses!

Definition 1.17: Ordered Field

An **Ordered field** is a field F that is also an ordered set, such that the following axioms are satisfied:

(i) If $x, y, z \in F$ and $y < z$, then $x+y < x+z$.

(ii) If $x, y \in F$ and $x > 0, y > 0$, then $x \cdot y > 0$.

Some properties of ordered fields are discussed in Rudin 1.18. We will again refer the reader to the discussion in the textbook for these properties, and here consider some examples.

$\mathbb{Q}$ is an ordered field, with the familiar order of $a > b$ if $a - b > 0$. A question may arise if $\mathbb{F}_2$ is an ordered field. A priori fields do not have order, but is it possible to impose an order on this set such that it is an ordered field? The answer turns out to be no.

Proof. It suffices to show that both possible orderings leads to a contradiction. Suppose $0 < 1$. Then, $1 = 0+1 < 1+1 = 0$ which is a contradiction. Suppose instead that $1 < 0$. Then, $0 = 1+1 < 1+0 = 1$ which again is a contradiction. □

Theorem 1.19: Existence of $\mathbb{R}$

There exists an ordered field $\mathbb{R}$ which has the LUB property and contains $\mathbb{Q}$ as a subfield.

What does it mean for $\mathbb{Q}$ to be a subfield? It means that there exists an injective function $\mathbb{Q} \mapsto \mathbb{R}$ that respects the properties of an ordered field.

This field $\mathbb{R}$ happens to be exactly the set of real numbers we are familiar with. However, a natural question is "what does it mean that there exsits a field?" It turns out that we can define the reals based on the definitions we have made already. One further question might be that could there not exists several fields with the above property; however, taking the appropriate view, we will find that there is a unqiue such field.

1.5 Consequences of the LUB Property

We will use the least upper bound property and the fact that $\mathbb{R}$ has $\mathbb{Q}$ as a subfield to derive its properties.

Theorem 1.20: Archimedian Property, Density of Rationals/Irrationals in $\mathbb{R}$

(a) If $x, y \in \mathbb{R}$ and $x > 0$, then $\exists n \in \mathbb{N}$ such that $nx > y$.

(b) If $x, y \in \mathbb{R}$, and $x < y$, then $\exists p \in \mathbb{Q}$ such that $x < p < y$. ($\mathbb{Q}$ is dense in $\mathbb{R}$)

(c) If $x, y \in \mathbb{R}$, and $x < y$, then $\exists \alpha \in \mathbb{R} \setminus \mathbb{Q}$ such that $x < \alpha < y$. ($\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$)

Proof

(a) Let $A = \{nx : n \in \mathbb{N}\}$. Suppose for the sake of contradiction that the conclusion was false; then y is an upper bound of A. Then, $\alpha = \sup(A)$ exists by the LUB property of $\mathbb{R}$. Since $x > 0$, we then have that $\alpha - x < \alpha$ by the property of an ordered field. Hence, $\alpha - x$ is not an upper bound for A. Therefore, there exists some $m \in \mathbb{N}$ such that $mx > \alpha - x$. It then follows that $(m+1)x > \alpha$. We therefore have found $m + 1 = k \in \mathbb{N}$ such that $kx > \alpha$, contradicting α being the least upper bound of A. □

In order to prove (b) and (c), we first prove a stronger version of 1.20(a):

Lemma

If $x, y \in \mathbb{R}$ and $x > 0$, then there exists $n \in \mathbb{Z}$ such that $(n-1)x \leq y < nx$.

Proof

Suppose $y \geq 0$. Let $A = \{m \in \mathbb{N} : y < mx\} \subset \mathbb{N}$. By Theorem 1.20 (a), we have that $A \neq \emptyset$. Every non-empty subset of $\mathbb{N}$ has a smallest element (to see this, let $x \in A$, and define $A' = \{y \in A : y \leq x\}$. This is finite and nonempty and so has a smallest element, and the minimum element of this set will also be a lower bound and hence the minimum element of all of A), so let $n = \min(A)$. The claim holds for this n. The case for $y < 0$ is left as an exercise. □

Proof

(b) Since $y - x > 0$, by (a), $\exists n \in \mathbb{N}$ such that $1 < n(y - x)$. Furthermore, by the Lemma we have that $\exists m \in \mathbb{Z}$ such that $m - 1 \leq nx < m$ and hence $m \leq nx + 1$. From these inequalities we obtain that $nx < m \leq nx + 1 < ny$, and therefore $x < \frac{m}{n} < y$ for some $m \in \mathbb{Z}$, $n \in \mathbb{N}$. □

For the proof of part (c), we will use the result of Theorem 1.21 from the next section, specifically that there exists $s \in \mathbb{R} \setminus \mathbb{Q}$ such that $s > 0$ and $s^2 = 2$. We will call this $\sqrt{2}$.

Proof

(c) First, we have that $\sqrt{2} < 2$ as if $\sqrt{2} = 2$ then $(\sqrt{2})^2 = 2 = 2^2 = 4$ which is a contradiction, and if $\sqrt{2} > 2$ then $2 = \sqrt{2} \cdot \sqrt{2} > 2 \cdot 2 = 4$ by Rudin 1.18 which is yet again a contradiction. Thus, $\frac{\sqrt{2}}{2} < 1$.
Let $x, y \in \mathbb{R}$ such that $x < y$. By Theorem 1.20(b), there exists $p, q \in \mathbb{Q}$ such that $x < p < q < y$. Let $\alpha = p + \frac{\sqrt{2}}{2}(q - p)$. Then, we have that $p < \alpha < p + 1(q - p) < q$ and hence $x < p < \alpha < q < y$.
If $\alpha \in \mathbb{Q}$, then $\sqrt{2} = 2\left(\frac{\alpha - p}{q - p}\right) \in \mathbb{Q}$, which is a contradiction, so it follows that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. □

1.6 Integer Roots of the Reals

In this section, we will prove that $\sqrt{2}$ exists and is an irrational number, but we will not use the fact that $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$; this would of course be circular reasoning. The more general idea will be to prove that for any $n \in \mathbb{N}$, there exists $y \in \mathbb{R}$ such that $y = x^{1/n}$. Before this, we prove a lemma.

Lemma

If $0 < a < b$ and $n \in \mathbb{N}$, then $0 < b^n - a^n \leq nb^{n-1}(b - a)$

Note that a "Calculus proof" of this Lemma would be to let $f(x) = x^n$, and then

$$f(b) - f(a) = f'(c)(b - a) = nc^{n-1}(b - a) \leq nb^{n-1}(b - a)$$

Where we invoke the mean value theorem. But this obviously doesn't work as we have neither defined a derivative nor proven the mean value theorem. A proper proof would be:

Proof

Let $0 < a < b$. Then, we may factor $b^n - a^n$ such that:

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

The second factor is a sum of n terms, each positive, and in between 0 and b^{n-1}. ThereforE:

$$b^n - a^n \leq nb^{n-1}(b - a)$$

which proves the claim. □

We will now state the theorem formally:

Theorem 1.21: Integer Roots of the Reals

Let $x \in \mathbb{R}$, $x > 0$, and $n \in \mathbb{N}$. Then, there exists a unique $y \in \mathbb{R}$ such that $y > 0$ and $y^n = x$.

Note that somewhere in the proof, we will use the fact that $y \in \mathbb{R}$; this statement doesn't hold for rationals (see Example 1.1) so some property of the reals must come into play somewhere.

Proof

If $n = 1$, then the unique solution is $y = x$; we may therefore assume that $n \geq 2$.
Uniqueness: Suppose there exist two distinct numbers y_1, y_2 with $y_1 > 0, y_2 > 0$, and $y_1^n = y_2^n = x$. WLOG, suppose $0 < y_1 < y_2$. We then have that $0 < y_1^n < y_2^n$ which is a contradiction.
Existence: We prove existence in three steps.

1. We show that $E \neq \emptyset$. Let $E = \{t \in \mathbb{R} : t > 0, t^n < x\}$. If $x < 1$, then $x^n < x$, so $x \in E$. If $x \geq 1$, then $\left(\frac{1}{2}\right)^n < \frac{1}{2} < x$, so $\frac{1}{2} \in E$. Therefore, $E \neq \emptyset$.

2. We show that E is bounded above and has a supremum in $\mathbb{R}$. If $t > 1 + x$, then it follows that $t^n > t > x$, so $t \neq E$. Hence, $1 + x$ is an upper bound of E. By Theorem 1.19 (the LUB property of $\mathbb{R}$), it follows that $\sup(E) \in \mathbb{R}$ exists.

3. We show that $y = \sup(E)$ satisfies $y^n = x$. As $\mathbb{R}$ is an ordered field, one of $y^n < x$, $y^n = x$, or $y^n > x$ must be true; we show that the first and third are impossible.

   (a) Suppose $y^n < x$. We will obtain a contradiction by finding $h > 0$ such that $(y + h)^n < x$. (Why is this a contradiction? $y + h > y$, so if $(y + h)^n < x$, then $y + h \in E$, contradicting the fact that $y + h$ would be an upper bound of E). WLOG, suppose that $h < 1$. By the above Lemma, we have that:

   $$(y + h)^n - y^n \leq n(y + h)^{n-1}h \leq n(y + 1)^{n-1}h$$

   By choosing h sufficiently small, that is:

   $$h < \min\left\{1, \frac{x - y^n}{n(y + 1)^{n-1}}\right\}$$

   Then $n(y + 1)^{n-1}h < x^n - y^n$ from which it follows that $(y + h)^n - y^n < x^n - y^n$ and so $y + h < x$, which is the desired contradiction.

   (b) Suppose $y^n > x$. We will obtain a contradivction by finding $h > 0$ such that $(y - h)^n > x$. If this is true, then $y - h$ is an upper bound for E, contradicting the fact that y is the least upper bound for E. WLOG suppose that $h < 0$. Again applying the Lemma, we have that:

   $$y^n - (y - h)^n \leq ny^{n-1}h$$

   By choosing h sufficiently small, that is:

   $$h < \min\left\{1, \frac{y^n - x}{ny^{n-1}}\right\}$$

   It then follows that:

   $$y^n - (y - h)^n \leq ny^{n-1}h < y^n - x$$

   and hence $(y - h)^n > x$, which is the desired contradiction. □

1.7 Construction of the Reals

Theorem 1.19 says that there exists an ordered field that contains $\mathbb{Q}$ as a subfield. We now go about proving this statement. The construction is fairly technical and hence will be carried out in multiple steps. Some of the steps are left as exercises (one can refer to Rudin for the fully complete construction).

Step 1: Defining the elements of $\mathbb{R}$

The members of $\mathbb{R}$ will be proper subsets of $\mathbb{Q}$, called cuts. $\mathbb{R} = \{\text{all cuts}\}$.

Definition: Cuts

A **cut** is a proper subset $\alpha \subsetneq \mathbb{Q}$ with the three properties:

(I) $\alpha \neq \emptyset$

(II) If $p \in \alpha$, then $q \in \alpha \ \forall q < p$.

(III) If $p \in \alpha$, then $\exists r \in \alpha$ such that $p < r$.

Figure 2: Visualization of a cut α. The real number being described of this cut can be thought of as the number at the right boundary (the arrow).

In a sense, a cut gives us a way of discussing the real numbers (in the way we are familiar with them already) without referring to them directly; much like we could formally define/refer to rationals as equivalence classes of ordered pairs.
As a note, we could very well define cuts to be bounded below rather than above, and the following construction would still work out.

Step 2: $\mathbb{R}$ is an ordered set

We define $\alpha < \beta$ to mean $\alpha \subsetneq \beta$. We show that this makes $\mathbb{R}$ into an ordered set. First checking transitivity, we have that if $\alpha < \beta$ and $\beta < \gamma$ then $\alpha < \gamma$ by the fact that set inclusion is transitive. Furthermore, at most one of $\alpha < \beta$, $\alpha = \beta$, and $\beta < \alpha$ hold; to see this is the case, suppose the first two fail. Then, $\alpha \nsubseteq \beta$. Hence, $\exists p \in \alpha$ with $p \notin \beta$. If $q \in \beta$, $q < p$ and hence $q \in \alpha$ by (II), so $\beta \subset \alpha$, and since $\beta \neq \alpha$ it follows that $\beta \subsetneq \alpha$.

Step 3: $\mathbb{R}$ has the LUB property

We show that $\mathbb{R}$ has the LUB property. To see this is the case, let $A \subset \mathbb{R}$ with $A \neq \emptyset$, and suppose that there exists $\beta \in \mathbb{R}$ that is an upper bound for A. We will now define $\gamma = \bigcup_{\alpha \in A} \alpha$ and prove that $\gamma \in \mathbb{R}$ and $\gamma = \sup A$ (hence A has a supremum and $\mathbb{R}$ has the LUB property).
Since $A \neq \emptyset$, $\exists \alpha_0 \in A$, and since $\alpha_0 \neq \emptyset$ (as it is a cut) and $\alpha \subset \gamma$, it follows that $\gamma \neq \emptyset$. Next, we have that $\gamma \subset \beta$, since $\alpha \subset \beta$ for every $\alpha \in A$, and hence $\gamma \neq \mathbb{Q}$, that is, $\gamma \subsetneq \mathbb{Q}$. Hence γ satisfies property (I) of a cut.
Take $p \in \gamma$. Then $p \in \alpha_1$ for some $\alpha_1 \in A$. If $q < p$, then $q \in \alpha$ (as α is a cut) so $q \in \gamma$, satisfying property (II).
Next, choose $r \in \alpha_1$ such that $r > p$, then $r \in \gamma$ (as $\alpha_1 \subset \gamma$) and hence γ satisfies property (III). Hence γ is a cut, and $\gamma \in \mathbb{R}$.
Finally, we show that $\gamma = \sup A$. Clearly, $\alpha \leq \gamma$ for all $\alpha \in A$, as $\gamma = \bigcup_{\alpha \in A} \alpha$, so γ is an upper boun dof A. To show that it is the least upper bound, let $\delta < \gamma$ be a cut. Then, $\exists s \in \gamma$ such that $s \notin \delta$. Therefore, $\exists \alpha_2 \in A$ such that $s \in \alpha_2$; hence $\delta < \alpha_2$, so δ is not an upper bound for A, giving the desired result.

Step 4: Addition on $\mathbb{R}$

Definition: Addition

If $\alpha, \beta \in \mathbb{R}$, we define $\alpha + \beta = \{s + t : s \in \alpha, t \in \beta\}$. Showing that this is a cut is left as an exercise.

Definition: Zero

$0^* = \{s \in \mathbb{Q}\}$. Showing that this is a cut is left as an exercise.

We leave it as an exercise to show that the addition axioms (A1)-(A5) of a field are satisfied under this definition of addition on $\mathbb{R}$, with the 0 element as 0^* defined above.

Step 5: $\mathbb{R}$ satisfies the Ordered Field Property (i)

We verify that if $\alpha, \beta, \gamma \in \mathbb{R}$ and $\beta < \gamma$, then $\alpha + \beta < \alpha + \gamma$.
For every $s \in \alpha, t \in \beta$, we have that $t \in \gamma$ as β is a subset of γ by the definition of order on $\mathbb{R}$. Hence, $s + t \in \alpha + \beta$ implies $s + t \in \alpha + \gamma$. Therefore, $\alpha + \beta \subset \alpha + \gamma$ and hence $\alpha + \beta \leq \alpha + \gamma$.
We are then left to check that $\alpha + \beta \neq \alpha + \gamma$. To see that this is the case, if $\alpha + \beta = \alpha + \gamma$, then $\beta = \alpha + \beta - \alpha = \alpha + \gamma - \alpha = \gamma$ by the field axioms for addition. Therefore we obtain that $\beta = \gamma$, contradicting that $\beta < \gamma$. Hence the claim is proven.
As a remark, note that $0^* < \alpha \iff -\alpha < 0^*$.

Next we will define multiplication on $\mathbb{R}$. A first attempt would be $\alpha \cdot \beta = \{s \cdot t : s \in \alpha, t \in \beta\}$. However, this definition is incosistent with negative numbers from what we require multiplication to accomplish. $-1 \cdot -1$ would fail to be a cut (it would not contain any negative numbers and hence fail criteria (II)) and $-1 \cdot 1$ would yield the entirety of the rationals (again not a cut!)

Step 6: Positive Multiplication on $\mathbb{R}$

Definition: Positive Reals

We define $\mathbb{R}^+ = \{\alpha \in \mathbb{R} : \alpha > 0^*\}$

DefinitionMultiplication of Positive Reals

If $\alpha, \beta \in \mathbb{R}^+$, we define $\alpha \cdot \beta = \{r \cdot s : r \in \alpha, r > 0, s \in \beta, s > 0\} \cup \{t \in \mathbb{Q}, t \leq 0\}$. Equivalently, $\alpha \cdot \beta = \{p \in \mathbb{Q} :\leq r \cdot s : r \in \alpha, r > 0, s \in \beta, s > 0\}$. We leave it as an exercise to show that $\alpha \cdot \beta \in \mathbb{R}$, and moreover, $\alpha \cdot \beta \in \mathbb{R}^+$. Showing this second fact proves ordered field property (ii).

Definition: One

$1^* = \{r \in \mathbb{Q} : r < 1\}$. We again leave showing $1^* \in \mathbb{R}^+$ as an exercise.

Step 7: Multiplication on all of $\mathbb{R}$

Definition: Multiplication by zero

$\alpha \cdot 0^* = 0^* = 0^* \cdot \alpha$

Definition: Multiplication

We define general multiplication as below, where the $\cdot$ on the RHS represents the multiplication of positive reals as outlined in Step 5.

$$\alpha \cdot \beta = \begin{cases} (-\alpha) \cdot (-\beta) & \text{if } \alpha < 0^* \text{ and } \beta < 0^* \\ -\left((-\alpha) \cdot \beta\right) & \text{if } \alpha < 0^* \text{ and } \beta > 0^* \\ -\left(\alpha \cdot (-\beta)\right) & \text{if } \alpha > 0^* \text{ and } \beta < 0^* \end{cases}$$

We leave it as an exercise to show that the multiplicative axioms (M1)-(M5), as well as the distributive law (D) of a field are satisfied under this definition of multiplication on $\mathbb{R}$.

Up until this point, we have shown $\mathbb{R}$ is an ordered field with the LUB property; we last check that it contains $\mathbb{Q}$ as a subfield. Note that we do have to be a bit careful with what we mean here; $\mathbb{R}$ does not literally contain $\mathbb{Q}$; $\mathbb{R}$ is indeed a set of proper subsets of $\mathbb{Q}$. What we really mean is to associate every element of $\mathbb{Q}$ to an element of $\mathbb{R}$ such that the field structure is preserved.

Step 8: $\mathbb{R}$ contains $\mathbb{Q}$ as a subfield

For each $r \in \mathbb{Q}$, associate the cut $r^* = \{p \in \mathbb{Q}, p < r^*\}$. We then leave as an easy exercise to verify that $r^* < s^* \iff r < s$, $r^* + s^* = r + s$, and $r^* \cdot s^* = r \cdot s$. This concludes the construction of the reals. □

Note that later on in the course, we will construct the real numbers in a different fashion; by considering Cauchy sequences modulo an equivalence relation. Also note that from here on out, it will suffice to have the standard/traditional picture of a "real number" in mind (i.e. infinite decimal expansions) and we will not have to really think about the real numbers as cuts; this was just necessary for the formal construction.

1.8 The Complex Field

Definition 1.24: The Complex Numbers

We define the set of **complex numbers** to be $\{(a,b) : a,b \in \mathbb{R}\}$, denoted by $\mathbb{C}$. For $x = (a,b) \in \mathbb{C}$ and $y = (c,d) \in \mathbb{C}$, we write $x = y$ if and only if $a = c$ and $b = d$ (note that this is a very different notion of equality compared to the rationals). We define the zero element to be $(0,0)$ and the one element to be $(1,0)$. We define addition of complex numbers such that:

$$x + y = (a,b) + (c,d) = (a + c, b + d)$$

And multiplication of complex numbers such that:

$$x \cdot y = (a,b) \cdot (c,d) = (ac - ba, ad + bc)$$

Theorem 1.25

The operations of $+$ and $\cdot$, as well as the zero/one elements defined above turn $\mathbb{C}$ into a field.

Proof

It suffices to verify the field axioms (A1)-(A5), (M1)-(M5), and (D) as discussed in 1.12. We will here show (M3), (M4), and (M5) and leave the rest as exercises.

(M3): Let $x, y, z \in \mathbb{C}$. We show that $(x \cdot y) \cdot z = x \cdot (y \cdot z)$. Let $x = (a, b), y = (c, d)$, and $z = (e, f)$. We then have that:

$$\begin{aligned}(x \cdot y) \cdot z &= (ac - bd, ad + bc) \cdot (e, f) \\ &= ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e)\end{aligned}$$

We also have that:

$$\begin{aligned}x \cdot (y \cdot z) &= (a, b) \cdot (ce - df, cf + de) \\ &= (a(ce - df) - b(cf + de), a(cf + de) + b(ce - df)) \\ &= (ace - adf - bcf - bde, acf + ade + bce - bdf) \\ &= ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e)\end{aligned}$$

So the claim is proven.

(M4): $(a, b)(1, 0) = (a \cdot 1 - b \cdot 0, a \cdot 0 + b \cdot 1) = (a, b)$

(M5): Let $x \in \mathbb{C}$ such that $x \neq 0$. Then, $x = (a, b)$ where either $a \neq 0$ or $b \neq 0$ or both. Hence, $a^2 + b^2 > 0$. Then, let $\frac{1}{x} = (\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2})$. We then have that:

$$\begin{aligned}x\frac{1}{x} &= (a, b)\left(\frac{a}{a^2 + b^2}, -\frac{b}{a^2 + b^2}\right) \\ &= \left(a\frac{a}{a^2 + b^2} - b\left(-\frac{b}{a^2 + b^2}\right), a\left(-\frac{b}{a^2 + b^2}\right) + b\left(\frac{a}{a^2 + b^2}\right)\right) \\ &= \left(\frac{a^2 + b^2}{a^2 + b^2}, -\frac{ab}{a^2 + b^2} + \frac{ab}{a^2 + b^2}\right) \\ &= (1, 0)\end{aligned}$$

Which proves the claim. □

Much like $\mathbb{Q}$ was a subfield of $\mathbb{R}$, $\mathbb{R}$ is a subfield of $\mathbb{C}$, and there exists a map ϕ from $\mathbb{R}$ to $\mathbb{C}$ that respects the field axioms, namely:

$$\begin{array}{rccc}\phi & : & \mathbb{R} & \longrightarrow & \mathbb{C} \\ & & x & \longmapsto & (x, 0)\end{array}$$

The theorem below shows that ϕ preserves the field structure:

Theorem 1.26

For $a, b \in \mathbb{R}$ we have that $(a, 0) + (b, 0) = (a + b, 0)$ and $(a, 0)(b, 0) = (ab, 0)$.

Definition 1.27: i

$i = (0, 1)$.

Theorem 1.28

$i^2 = -1$.

Theorem 1.29

If $a, b \in \mathbb{R}$, then $(a, b) = a + bi$.

Proof

Below are the trivial proofs for the above three theorems.

$$\begin{aligned}(a,0) + (b,0) &= (a+b, 0+0) = (a+b, 0)\\ (a,0) \cdot (b,0) &= (a \cdot b - 0 \cdot 0, a \cdot 0 + 0 \cdot b) = (ab, 0)\\ i^2 = i \cdot i &= (0,1) \cdot (0,1) = (-1, 0) = -1\\ a + bi &= (a,0) + b(0,1) = (a,0) + (0,b) = (a,b)\end{aligned}$$

A slightly odd question may be to ask whether $\mathbb{C}$ is a subfield of $\mathbb{R}$, i.e. does there exist $\psi : \mathbb{C} \mapsto \mathbb{R}$ such that $\psi(a + b) = \psi(a) + \psi(b)$ and $\psi(a \cdot b) = \psi(a) \cdot \psi(b)$. As we will prove in Chapter 2, we do have that $|\mathbb{C}| = \left|\mathbb{R}^2\right| = |\mathbb{R}|$ (where $||$ denotes cardinality of the set, to be defined shortly), so there does exist a bijection (i.e. a function that is both injective/one-to-one and surjective/onto; we will define these terms precisely in the next chapter) between the two sets.

As a Lemma, we have that the only injective function $f : \mathbb{Q} \mapsto \mathbb{R}$ that satisfies $f(a + b) = f(a) + f(b)$ and $f(a \cdot b) = f(a) \cdot f(b)$ is $f(x) = x$. The proof of this is left as a homework problem (HW2). Therefore, it follows that the only injective function $g : \mathbb{Q} \times \{0\} \mapsto \mathbb{R}$ (where $\times$ denotes the Cartesian product) is given by $g((x, 0)) = x$. We now give a proof that $\mathbb{C}$ is not a subfield of $\mathbb{R}$.

Proof. Suppose then for the sake of contradiction that there exists an injective function $\psi : \mathbb{Q} \times \{0, 1\} \mapsto \mathbb{R}$. Such a function then must satisfy$\psi(i \cdot i) = \psi(-1) = -1$, and $\psi(i \cdot i) = \psi(i) \cdot \psi(i) = \psi((0, 1)) \cdot \psi((0, 1)) = 0 \cdot 0 = 0$ which is a contradiction. Hence, no such injection exists from $\mathbb{Q} \times \{0, 1\}$ to $\mathbb{R}$ and hence no such injection could exist from $\mathbb{C}$ ($\mathbb{R}^2$) to $\mathbb{R}$. Hence $\mathbb{C}$ is not a subfield of $\mathbb{R}$. □

Definition 1.30: Real/Imaginary Parts and Complex Conjugates

Let $z = a + bi \in \mathbb{C}$. Then, $\text{Re}(z) = a$ is the **real part** of z and $\text{Im}(z) = b$ is the **imaginary part** of z. The **complex conjugate** of z, denoted by $\bar{z}$, is defined as $\bar{z} = a - bi$.

Theorem 1.31

Let $z, w \in \mathbb{C}$. It then follows that:

(a) $\overline{z+w} = \bar{z} + \bar{w}$.

(b) $\overline{zw} = \bar{z} \cdot \bar{w}$.

(c) $z + \bar{z} = 2\,\mathrm{Re}(z)$, $z - \bar{z} = 2i\,\mathrm{Im}(z)$.

(d) $z\bar{z}$ is real and positive (except when $z = 0$).

Proof

We prove (d). We have that:

$$z\bar{z} = (a+bi)(a-bi) = a^2 + b^2$$

$a^2 + b^2 \geq 0$, and $a^2 + b^2 = 0 \iff a = 0, b = 0$ which proves the claim. □

Definition 1.32: Absolute Value

We define the **absolute value** $|z|$ of a complex number z as $|z| = \sqrt{z\bar{z}}$. Note that if $a \in \mathbb{R}$ and $z = (a, 0)$, then

$$|z| = \sqrt{a^2} = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

Hence if $a \in \mathbb{R}$, we can define $|a| = \big|(a, 0)\big|$.

Theorem 1.33

Let $z, w \in \mathbb{C}$.

(a) $|z| \geq 0$, $|z| = 0 \iff z = 0$.

(b) $|\bar{z}| = |z|$.

(c) $|z||w| = |zw|$.

(d) $\big|\mathrm{Re}(z)\big| \leq |z|$, $\big|\mathrm{Im}(z)\big| \leq |z|$.

(e) $|z + w| \leq |z| + |w|$.

Proof

We prove (d) and (e). Let $z, w \in \mathbb{C}$, with $z = a + bi$. For (d) we have that $\mathrm{Re}(z) = a$, so

$$\left|\mathrm{Re}(a)\right| = \sqrt{a^2} \leq \sqrt{a^2 + b^2} = |z|$$

And an equivalent proof follows for $\mathrm{Im}(z)$. For (e), we have that:

$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) \\
&= z\bar{z} + z\bar{w} + w\bar{z} + w\bar{w} \\
&= |z|^2 + 2\,\mathrm{Re}(z\bar{w}) + |w|^2 \\
&\leq |z|^2 + 2\left|\mathrm{Re}(z\bar{w})\right| + |y|^2 && (|x| > x) \\
&= |z|^2 + 2|z\bar{w}| + |w|^2 && (1.33(\mathrm{d})) \\
&= |z|^2 + 2|z||\bar{w}| + |w|^2 && (1.33(\mathrm{c})) \\
&= |z|^2 + 2|z||w| + |w|^2 && (1.33(\mathrm{b})) \\
&= (|z| + |w|)^2
\end{aligned}$$

The claim follows by taking square roots on both sides. □

1.9 The Cauchy-Shwartz Inequality

Recall the summation notation:

$$x_1 + x_2 + \ldots + x_n = \sum_{j=1}^{n} x_i$$

Theorem 1.35: Cauchy-Shwartz Inequality

Let $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{C}$. We then have that:

$$\left|\sum_{j=1}^{n} a_j \bar{b}_j\right|^2 \leq \left(\sum_{j=1}^{n} \left|a_j\right|^2\right)\left(\sum_{j=1}^{n} \left|b_j\right|^2\right)$$

Note that in the above theorem, both the RHS and the LHS are real numbers (check!) so the equality makes sense (recall that there is no order on $\mathbb{C}$; in fact, it is impossible to define one).

A geometric interpretation of the above inequality is as follows. Let $\mathbf{a}, \mathbf{b}$ be vectors in $\mathbb{C}^n$. Then, $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{j=1}^{n} a_j \bar{b}_j$ is the inner product of $\mathbf{a}$ and $\mathbf{b}$. Then, the inequality says that $\left|\langle \mathbf{a}, \mathbf{b} \rangle\right|^2 \leq \langle \mathbf{a}, \mathbf{a} \rangle \cdot \langle \mathbf{b}, \mathbf{b} \rangle$.

Proof

Define $A = \sum_{j=1}^{n} \left|a_j\right|^2$, $B = \sum_{j=1}^{n} \left|b_j\right|^2$, and $C = \sum_{j=1}^{n} a_j \bar{b}_j$. If $B = 0$ (that is, all of the b_js are zero) then the LHS/RHS are both zero and we are done. So, let us assume that $B > 0$. Let $\lambda \in \mathbb{C}$, and we then have that:

$$\begin{aligned}
0 &\leq \sum_{j=1}^{n} \left|a_j + \lambda b_j\right|^2 \\
&= \sum_{j=1}^{n} (a_j + \lambda b_j)(\bar{a}_j + \bar{\lambda}\bar{b}_j) \\
&= \sum_{j=1}^{n} \left|a_j\right|^2 + \bar{\lambda} \sum_{j=1}^{n} a_j \bar{b}_j + \lambda \sum_{j=1}^{n} \bar{a}_j b_j + |\lambda|^2 \sum_{j=1}^{n} \left|b_j\right|^2 \\
&= A + \bar{\lambda} C + \lambda \bar{C} + |\lambda|^2 B
\end{aligned}$$

This inequality holds for any λ; it therefore holds for $\lambda = -\frac{C}{B}$, so:

$$\begin{aligned}
0 &\leq A - \frac{\bar{C}}{B} C - \frac{C}{B} \bar{C} + \frac{C\bar{C}}{B^2} B \\
&= A - \frac{|C|^2}{B}
\end{aligned}$$

So we therefore obtain that $|C|^2 \leq AB$ which is the desired inequality. □

A natural question given any inequality is when does equality hold; the answer turns out to be if the vectors are linearly independent, that is, at least one of $\mathbf{a} = \alpha\mathbf{b}$ and $\mathbf{b} = \beta\mathbf{a}$ ($\alpha, \beta \in \mathbb{C}$) hold. Note that we only require one of the two relations to hold; in the case that one of $\mathbf{a}, \mathbf{b}$ are $\mathbf{0}$ (the vector of all zeros) both equalities cannot be true. It is left as a homework problem to verify equality in the Cauchy-Shwartz inequality if and only if at least one of the two conditions holds (HW3).

1.10 Euclidean Space

Definition 1.36: Euclidean k-space

If $k \in \mathbb{N}$, define $\mathbb{R}^k$ as the set of k-tuples of real numbers:

$$\mathbb{R}^k = \{\mathbf{x} = (x_1, x_2, \ldots, x_k) : x_1, x_2, \ldots, x_k \in \mathbb{R}\}$$

We can then define vector addition as:

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_k + y_k)$$

And scalar multiplication (for $\alpha \in \mathbb{R}$) to be:

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_k)$$

These operations make $\mathbb{R}^k$ into a vector space over the real field. We can define the inner product over $\mathbb{R}^k$ to be:

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{k} x_j y_j$$

This allows us to define the norm of $\mathbf{x}$ to be:

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \left(\sum_{j=1}^{n} x_j^2 \right)^{1/2}$$

$\mathbb{R}^k$ with the above inner product and norm is called **Euclidean k-space**.

We briefly remark that the above inner product we defined agrees with the inner product we defined over $\mathbb{C}^k$; we can identify $r \in \mathbb{R}$ with $(r, 0) \in \mathbb{C}$, and hence recognize that $\mathbb{R}^k \subset \mathbb{C}^k$ where the imaginary part of each coordinate is zero. Then, for the inner product we get the exact same result, as $\bar{b}_j = b_j$ for any complex numbers with imaginary part zero. From this we can conclude that the Cauchy-Shwartz inequality also holds in $\mathbb{R}^k$.

Note that although the field $\mathbb{C}$ is $\mathbb{R}^2$ with multiplication defined as in Definition 1.24, in general vector multiplication on $\mathbb{R}^n$ is not well defined. That is, we cannot make $\mathbb{R}^n$ into a field in general; though we can make it into a vector space, which has slightly less structure.

One possibly familiar notion of vector multiplication in $\mathbb{R}^3$ is the cross product. For $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$, the cross product is defined as:

$$\mathbf{x} \times \mathbf{y} = (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1)$$

However, the cross product does not satisfy properties that would be necessary to make $\mathbb{R}^3$ a field. For one, it is not commutative, but anticommutative; $\mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x}$. One might ask whether vectors in $\mathbb{R}^3$ have well-defind inverses, but even before that, there does not exist an identity vector in $\mathbb{R}^3$ under the cross product! In fact, $\mathbb{R}^3$ under vector addition and cross product multiplication can be viewed as a noncommutative ring without an identity.

Note that there is a more general notion of a "wedge product" between vectors in $\mathbb{R}^n$. We are in a sense very "lucky" that in $\mathbb{R}^3$, the wedge product of two vectors returns another vector in $\mathbb{R}^3$.

Theorem 1.37

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^k$, and $\alpha \in \mathbb{R}$. Then:

(a) $|\mathbf{x}| \geq 0$

(b) $|\mathbf{x}| = 0 \iff \mathbf{x} = (0, \ldots, 0)$. This is often denoted as $\mathbf{0}$, the "zero vector".

(c) $|\alpha \mathbf{x}| = |\alpha||\mathbf{x}|$

(d) $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$

(e) $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$

(f) $|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}|$

(e) and (f) are often called "triangle inequalities"; a visual intuition for these inequalities is given in the following figure:

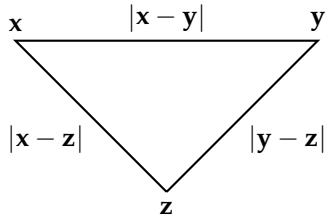

Figure 3: Visual picture for Theorem 1.37(f), drawn in $\mathbb{R}^2$. Suppose we started at $\mathbf{x}$ and wanted the shortest path to $\mathbf{z}$; we could try walking directly to $\mathbf{z}$, or we could try walking somewhere else first ($\mathbf{y}$) and then to $\mathbf{z}$. However, the theorem tells us that the direct path will always be shorter in Euclidean space.

Note that equality in part (f) arises if and only if $\mathbf{y}$ lies on the line segment between $\mathbf{x}$ and $\mathbf{z}$.

Proof

(a)-(c) are immediate, and (d) immediately follows from Theorem 1.35 (Cauchy-Shwartz). For (e), we have that:

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y})(\mathbf{x} + \mathbf{y}) \\ &= |\mathbf{x}|^2 + 2\mathbf{x} \cdot \mathbf{y} + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + |2\mathbf{x} \cdot \mathbf{y}| + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2 \quad \text{(1.37(d))} \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2 \end{aligned}$$

And the claim follows by taking square roots on both sides. For (f), substitute $\mathbf{x} \mapsto \mathbf{x} - \mathbf{y}$ and $\mathbf{y} \mapsto \mathbf{y} - \mathbf{z}$ into (e). □

Though we discuss the Euclidean norm here, it may also be of interest to consider/discuss other norms. One example is the L_1 norm (c.f. the norm discussed in Definition 1.36, which is the L_2 norm), which is the sum of the absolute values of each of the components. For $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ we have that:

$$|\mathbf{x}|_1 = |x_1| + |x_2| + \ldots + |x_n|, \quad |\mathbf{x} - \mathbf{y}|_1 = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n|$$

The L_1 norm is often called the "Taxicab norm" or the "Manhattan norm" as the way it quantifies distance is akin to walking in discrete NSEW chunks; much like a taxi running through a grid-like New York City!

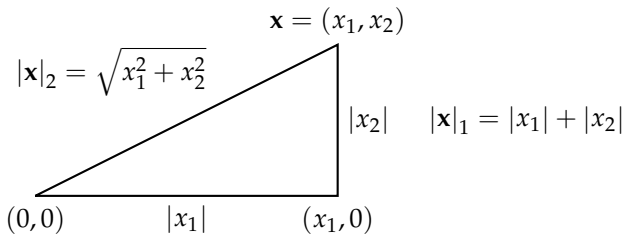

Figure 4: Visual comparison of the L_1 and L_2 norms in $\mathbb{R}^2$.

We are free to generalize this notion to the L_n norm, and we may also define the L_∞ norm, which for $\mathbf{x} \in \mathbb{R}^n$ is defined as:

$$|\mathbf{x}|_\infty = \max_i |x_i|$$

In general for any $\mathbf{x} \in \mathbb{R}^n$, we have that $|x|_1 \geq |x|_2 \geq |x|_3 \geq \ldots \geq |x|_\infty$. We note that we that we can generalize these norms to the cases where we have infinite components:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} < \infty \quad \|f\|_p \equiv \left(\int_S |f|^p \, \mathrm{d}\mu \right)^{1/p} < \infty$$

Which allow us to define norms for function spaces. However, a detailed discussion of these are beyond the scope of this course (to be covered in a later course in functional analysis!) Moreover, we haven't even defined what an infinite sum or integral are yet, which we will get to in later chapters.

2 Basic Topology

2.1 Finite and Countable Sets

This chapter is split into two portions; the first looks at counting, what it means for us to say that two sets have the same number of elements, and concludes with a classic theorem of Cantor concerning uncountable sets. The second part looks at the topology of metric spaces, before moving onto the topology of the real numbers.

Let us then begin with our discussion of counting. If we consider counting how many bananas there are on a table (say there are 10 bananas), then what we are formally doing is establishing a correspondence between each ball on the table with an element in the set $\{1, \ldots, 10\}$. When we refer to the number of elements in a set, it will be good to keep in mind that we are establishing functions between sets. Although we have been discussing functions with some frequency in the course already, we give a definition below for completeness.

Definition 2.1: Functions

Let A, B be sets. Then, a map that associates each element $x \in A$ with a unique element denoted as $f(x) \in B$ is a **function** $f : A \to B$. We then define A as the **domain** of f and the set $\{f(x) : x \in A\}$ as the **range**. For $E \subset A$, we call $f(E) = \{f(x) : x \in E\}$ the **image** of E under f. For $F \subset B$, we call $f^{-1}(B) = \{x \in A : f(x) \in B\}$ the **preimage** of F.

Definition 2.2: Injective/Surjective Functions

Let $f : A \mapsto B$ be a function. If for $x_1, x_2 \in A$ we have that $f(x_1) = f(x_2) \implies x_1 = x_2$ (or equivalently, $x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$), then we say that f is **injective**, or **one-to-one**. If for all $y \in B$ there exists $x \in A$ such that $y = f(x)$, then we say that f is **surjective**, or **onto**. If a function is both injective and surjective, it is **bijective**.

Intuitively, we can think of injectivity as implying each element in B being reached at most once, and surjectivity implying that each element in B is reached at least once.

Definition 2.3: Cardinality & Equivalence

Let A, B be sets. We say that A, B have the same **cardinality** if there exists $f : A \mapsto B$ such that f is bijective. We can denote this as $A \sim B$ where $\sim$ indicates an **equivalence relation**. An equivalence relation has three properties:

(a) Reflexivity: $A \sim A$.

(b) Symmetry: If $A \sim B$ then $B \sim A$.

(c) Transitivity: If $A \sim B$ and $B \sim C$ then $A \sim C$.

As a point of notation, $|S|$ denotes the cardinality of the set S.

We get (a) from each set having a bijection to itself (i.e. the identity function), (b) from the fact that if there exists a bijection $f : A \mapsto B$, then there must exist an inverse $f^{-1} : B \mapsto A$, and (c) from if there exist bijections $f : A \mapsto B$ and $g : B \mapsto C$ then the composition $g \circ f : A \mapsto C$ will also be a bijection.

Definition 2.4: Countability

First, we denote $J_n = \{1, 2, 3, \ldots, n\}$ and $J = \mathbb{N} = \{1, 2, 3, \ldots\}$. Let A be a set. We say that A is **finite** if it has a finite number of elements, that is, there exists $n \in \mathbb{N}$ such that $A \sim J_n$. A set A is **infinite** if it is not finite, and we cannot put A in bijection with J_n for any $n \in \mathbb{N}$. We say that A is **countable**if $A \sim \mathbb{N}$, and **uncountable** otherwise.

Note that the above definition gives us a useful notion for what sets we can consider countable; if we can enumerate a set with the naturals, this yields a bijection with $\mathbb{N}$ and hence the set must be countable.

We here give some additional properties concerning cardinalities of sets, which may be useful:

- $|A| \leq |B| \iff \exists f : A \mapsto B$ such that f is injective
- $|A| \geq |B| \iff \exists f : A \mapsto B$ such that f is surjective
- $|A| \leq |B|$ and $|A| \geq |B| \implies |A| = |B|$

Example 2.5

$\mathbb{Z}$ is countable. To see this, consider the function:

$$f = \begin{cases} \frac{n}{2} & n \text{ is even} \\ -\frac{n-1}{2} & n \text{ is odd} \end{cases}$$

f is a bijection (check!) and hence $\mathbb{N} \sim \mathbb{Z}$.

The above example serves as a bit of a warning sign. Even though $\mathbb{N} \subsetneq \mathbb{Z}$ and $\mathbb{Z}$ has "more elements", we still find that the two sets have the same cardinality. A similar example is given by $\mathbb{N}$ and the set of all even natural numbers (which we may denote $2\mathbb{N}$); the bijection $f(n) : n \mapsto 2n$ between these two sets shows that $\mathbb{N} \sim 2\mathbb{N}$, even though $2\mathbb{N}$ is a strict subset of $\mathbb{N}$.

Theorem 2.8

A subset of a countable set is either finite or countable.

Proof

(Sketch) The countability of A implies that $A = \{a_1, a_2, a_3, a_4, a_5 \ldots\}$ (in other words, we can enumerate the elements using $\mathbb{N}$). Let $S \subset A$. Then, $S = \{a_1, \cancel{a_2}, a_3, a_4, \cancel{a_5}, \ldots\}$, that is, A with some (or none) of the elements removed. Now, we can rename all the elements with $a_1, a_2, \ldots$; what we have left is again an enumeration, so it is yet again (at most) countable.

One potentially useful fact is that if we have a set S and a function $f : \mathbb{N} \mapsto S$ such that f is surjective, then S is at most countable.

Proof. Let $T = \{n \in \mathbb{N} : f(n) \neq f(m), \forall m = 1, 2, \ldots, m\}$. We restrict $f : T \mapsto S$, then f is injective by constructive. It is still surjective, hence $T \sim S$. Since $T \subset \mathbb{N}$, by Theorem 2.8, S is finite or countable. □

Theorem 2.12

Let $E_1, E_2, \ldots$ be countable sets (i.e. we have a countable number of countable sets). Define $S = \bigcup_{n=1}^{\infty} E_n$. Then, S is countable.

Proof

Write $E_n = \{x_{n1}, x_{n2}, x_{n3}, \ldots\}$ (which we can do as each of the E_ns are countable). Then, we form an array:

$$\begin{array}{lcl} E_1 & = & x_{11} \quad x_{12} \quad x_{13} \quad \cdots \\ E_2 & = & x_{21} \quad x_{22} \quad x_{23} \quad \cdots \\ E_3 & = & x_{31} \quad x_{32} \quad x_{33} \quad \cdots \\ \ldots & & \end{array}$$

Then, we can re-number the elements along the diagonal lines (i.e. $x_{11}, x_{21}, x_{12}, x_{31}, x_{22}, x_{13}, \ldots$). This new enumeration corresponds to a countable set. From there, we let $T \subset \mathbb{N}$ be the remaining labels in the enumeration after removing the repeated elements from the sequence. Then, $T \sim S$, and hence S is at most countable. S cannot be finite as $E_1 \subset S$ and E_1 is not finite. Hence S is countable. □

Corollary 2.13: $\mathbb{Q}$ is Countable.

- If A is countable, the set of n-tuples of $(a_1, \ldots a_n)$ is also countable for any $n \in \mathbb{N}$.
- $\mathbb{Q}$ is countable.

We defined $\mathbb{Q}$ as pairs of integers, but by the first part of the corollary (which follows immediately by application of Theorem 2.12) $\mathbb{Z}^2$ (the set of pairs of integers) has equal cardinality to $\mathbb{Z}$, and since $\mathbb{Q}$ is a subset of the set of pairs of integers, $\mathbb{Q}$ is countable.

Another way we can formalize this argument: if we let $X_t = \left\{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{N}, |m| \leq t, n \leq t\right\}$, then we have that $|X_t|$ is at most $t \cdot (2t+1)$ (the first term being the choises for n, the second term being the choices for m). Of course, the cardinality of X_t is actually less than that as we would have to remove repeats. Nontheless, we have that $\mathbb{Q} = \bigcap_{t=1}^{\infty} X_t$ and hence $\mathbb{Q}$ would be countable by Theorem 2.12.

From the discussion of today, we have established that $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$. Does $\mathbb{R}$ also have equal cardinality to these sets? Do infinite sets in general have the same cardinality? The answer turns out to be no for both of these questions. We will answer the first question in the next lecture (when we discuss Cantor diagonalization, a highlight of the course), but we can discuss the second statement now. First, we make a definition:

Definition: Power Sets

Let A be a set. Then, the **power set** of A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

An interesting theorem then follows:

Theorem: Power Set Cardinality

Let A be a set. Then, $|A| < \left|\mathcal{P}(A)\right|$.

Proof

Suppose for the sake of contradiction that there exists a surjection $f : A \to \mathcal{P}(A)$ (this would imply that $|A| \geq \left|\mathcal{P}(A)\right|$, so by showing this is false, we obtain the desired result). Then, each element $x \in A$ gets mapped to some subset of A. We either have that x belongs to the subset that it gets mapped to, or it doesn't. Therefore, we can define a new subset $B \subset A$, such that:

$$B = \{x \in A : x \notin f(x)\}$$

In other words, the set of all xs that are not in the subset that they get mapped to by f. Since f is surjective, there must be an element $y \in A$ such that $f(y) = B$. One of $y \in B$, $y \notin B$ must be true. If $y \in B$, by construction of B we have that y is not in the subset that it gets mapped to by f, which is a contradiction. If $y \notin B$, by definition of B, $y \in B$ as it is not in the subset that it gets mapped to, yet again a contradiction. Therefore, we conclude that no $y \in A$ exists such that $f(y) = B$, and hence, no surjective f exists such that $f : A \to \mathcal{P}(A)$. Hence, $|A| < \left|\mathcal{P}(A)\right|$. □

An interesting consequence of this theorem is that for a countable set A, we then have that $\mathcal{P}(A)$ is an infinite set which has greater cardinality! For example, $|\mathbb{N}| < \left|\mathcal{P}(\mathbb{N})\right|$. Moreover, this gives rise to an infinite number of cardinalities in ascending order; $|\mathbb{N}| < \left|\mathcal{P}(\mathbb{N})\right| < \left|\mathcal{P}(\mathcal{P}(\mathbb{N}))\right| < \cdots$ and so on.

2.2 Uncountable Sets

Theorem 2.14: Existence of Uncountable Sets

Let $A = \{(b_1, b_2, \ldots), b_n \in \{0, 1\}\}$ be the set of binary sequences. Then, A is uncountable.

Proof

It suffices to show that every countable subset of A is a proper subset of A. Let $E \subset A$ be countable, and let $E = \left\{S^{(1)}, S^{(2)}, S^{(3)}, \ldots\right\}$. To show that E is a proper subset, we show that there exists a sequence $S \in A \setminus E$. To construct such an S, let us put the elements of E in an array.

$$\begin{aligned}
S^{(1)} &= \boxed{b_1^1} \quad b_2^1 \quad b_3^1 \quad \cdots \\
S^{(2)} &= b_1^2 \quad \boxed{b_2^2} \quad b_3^2 \quad \cdots \\
S^{(3)} &= b_1^3 \quad b_2^3 \quad \boxed{b_3^3} \quad \cdots
\end{aligned}$$

Then, define:

$$\tilde{b}_n^n = \begin{cases} 1 & \text{if } b_n^n = 0 \\ 0 & \text{if } b_n^n = 1 \end{cases}$$

I.e. $\tilde{b}_n^n$ is the bit flip of b_n^n. Then, let $S = (\tilde{b}_1^1, \tilde{b}_2^2, \tilde{b}_3^3, \ldots)$, that is, S is the sequence of bit-flipped diagonal elements of the original array. By construction, $S \neq S^{(k)}$ as for any $S^{(k)} \in E$ as S differs at the kth position. Hence, $S \notin E$ and therefore $E \subsetneq A$. □

The above proof is a very famous argument, invented by the mathematician George Cantor. The discovery that there exist sets with greater cardinality than $\mathbb{N}$ was initially quite controversial in the math community!

Corollary: $\mathbb{R}$ is Uncountable

$\mathcal{P}(\mathbb{N})$ (the power set of $\mathbb{N}$) is uncountable. $\mathbb{R}$ is uncountable.

Proof

Although we showed that $\mathcal{P}(\mathbb{N})$ was uncountable last lecture, we show this in an alternative way by considering a bijection between $\mathcal{P}(\mathbb{N})$ and the set A of binary sequences. To do this, consider that we can associate a subset $T \subset \mathbb{N}$, $T \in \mathcal{P}(\mathbb{N})$ with the sequence corresponding to:

$$b_n = \begin{cases} 1 & \text{if } n \in T \\ 0 & \text{if } n \notin T \end{cases}$$

Since A is uncountable, it follows that $\mathcal{P}(\mathbb{N})$ is uncountable. The second statement in the corollary follows (roughly) by considering $\mathbb{R}$ represented in binary, though this requires more justification than what we present here (we will show below that a subset of $\mathbb{R}$ is uncountable). □

Theorem

$[0,1] \subset \mathbb{R}$ is uncountable.

Proof

(Sketch) We construct a bijection from $[0,1]$ to A. Let $x \in [0,1]$, and let b_1 be the largest integer such that $N_1 = \frac{b_1}{2} \leq x$. Then, let $b_2 \in \{0,1\}$ be the largest integer such that $N_2 = \frac{b_1}{2} + \frac{b_2}{2^2} \leq x$. We can continue dividing $[0,1]$ in half in this way, approximating x by powers of 2 (a decimal expansion in binary). Then, let $E(x) = \{N_1, N_2, N_3, \ldots\}$. By construction, $E(x)$ is bounded above by x and nonempty. Hence, $\sup(E(x))$ exists and is unique, and in fact is equal to x (note that we are in essence constructing an "infinite series" where the sequence of partial sums is increasing and bounded, approaching x from the left). Doing this we can associate $(b_1, b_2, b_3, \ldots) \in A$ with every number $x \in [0,1]$ and therefore $[0,1] \sim A$. Hence $[0,1]$ is uncountable. □

It might be worth investigating why a more naive approach to the above argument may fail. Let D be the sets of all decanary sequences, i.e. all sequeces of the form $(d_1, d_2, d_3, \ldots)$ where $d_n \in \{n \in \mathbb{Z}, 0 \leq n \leq 9\}$. Then, it might be tempting to say that the function:

$$\begin{array}{rccc} f & : & D & \longrightarrow & [0,1] \\ & & (d_1, d_2, d_3, \ldots) & \longmapsto & 0.d_1 d_2 d_3 \ldots \end{array}$$

Is a bijection; however, it is actually not the case! To see this, consider that $0.1000\ldots = 0.0999\ldots$. The sequences $(1,0,0,0,\ldots)$ and $(0,9,9,9,\ldots)$ are distinct, but the real number they would map to would be the same. This map is not an injection but a surjection. To show that $\mathbb{R}$ is uncountable, we require an injection from D to $[0,1]$. So instead, what we have done with the above proof is map the set of binary sequences B to $[0,1]$:

$$\begin{array}{rccc} g & : & B & \longrightarrow & [0,1] \\ & & (b_1, b_2, b_3, \ldots) & \longmapsto & \sum_{j=1}^{\infty} b_j 3^{-j} = \sup\left\{\sum_{j=1}^{N} b_j 3^{-j} : N \in \mathbb{N}\right\} \end{array}$$

Note that in general, an infinite sum is not equal to the supremum of its partial sums (as a counterexample, consider the sum with first term 1, the second term -1, and all the remaining terms 0; evidently the value

of the infinite sum is 0, but the supremum of the partial sums would be 1.) We will define an infinite sum more generally in chapter 3, but in our case here, this definition suffices as in the case where all terms of the sum are non-negative, the supremum of the partial sums is indeed the value of the infinite sum. We leave it as an exercise to show that g is an injection. Also, one neat remark; the above set is equivalent to the Cantor set, which we will more formally define/discuss in Chapter 2.

We take note a theorem that if Y is a set, and there exists $X \subset Y$ such that X is in bijective correspondence with an uncountable set, then Y is also uncountable. Thus we can use the above theorem to conclude that $\mathbb{R}$ is uncountable.

2.3 Topology of Metric Spaces

In our investigation of topology, we will try to better understand distances between and neighbourhoods of points. To do so, we first introduce the notion of a metric space.

Definition 2.15: Metric Spaces

A set X is a **metric space** (whose elements we call points) with a **metric** $d : X \times X \mapsto \mathbb{R}$ such that for $x, y, z \in X$:

(a) $d(x,y) > 0$ if $x \neq y$; $d(x,x) = 0$;

(b) $d(x,y) = d(y,x)$ (d is symmetric);

(c) $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality).

Example 2.16

$\mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^n, \mathbb{C}$ are all metric spaces with $d(x,y) = |x-y|$. Any subset $Y \subset X$ of a metric space is also a metric space, with the same metric.

Another example of a metric space is given below; note that this example can be generalized, in that any connected graph can be made into a metric space.

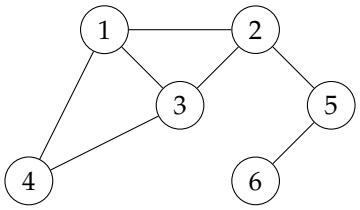

Figure 5: A graph-theoretic example of a metric space. Let $X = \{1,2,3,4,5,6\}$. Then, for $x, y \in X$, let $d(x,y)$ be the number of edges if the shortest path between x and y. Properties (a), (b) of a metric are immediately satisfied, and (c) follows from the property of the shortest path.

It may be of interest to consider other possible metrics. The discrete metric is defined such that $d(x,y) = 1$ if $x \neq y$ and $d(x,x) = 0$. The L1 Norm is defined such that $d(f,g) = \int_0^1 |f(x) - g(x)| dx$ where f, g are functions. In an inner product space, we have that $d(v,w) = \|v - w\| = \langle v - w, v - w \rangle^{1/2}$; if v, w are functions, then we have that $d(v,w) = \left(\int_0^1 |v(x) - w(x)|^2 dx \right)^{1/2}$. In analysis, we are often interested in the notion of things being close to other things (e.g. with limits, continuity) so a notion of distance is extremely important to define.

Definition 2.18: Neighbourhoods

A **neighbourhood** in a metric space X is a set $N_r(p) = \{q \in X : d(p,q) < r\}$ with $r > 0$.

Example

- In $\mathbb{R}$, $N_r(p)$ is the interval $(p-r, p+r)$ about midpoint p.
- In $\mathbb{R}^2$, $N_r(p)$ is the open disk about center p.
- In $\mathbb{R}^3$, $N_r(p)$ is the open ball about center p.
- In $\mathbb{R}^n$, $N_r(p)$ is the open hyperball about center p.

As another example, consider that in $\mathbb{Z}$, $N_1(0) = \{0\}, N_{3/2} = \{-1, 0, 1\}$ and $N_2(0) = \{-1, 0, 1\}$.

Definition 2.18: Interior Points

Let $E \subset X$. Then, p is an **interior point** of S if there is a neighbourhood $N_r(p)$ such that $N \subset E$.

Intuitively, an interior point of E is a point that is not on the boundary of E. As an example, in $\mathbb{R}^n$, if $E = \left\{y : |x-y| \leq 1\right\}$, then the interior points of E (which we can denote as E°) are $E^\circ = \left\{y : |x-y| < 1\right\}$. The idea is that there is always some finite distance to the boundary, so we can always fit a (perhaps small) open ball in. But this doesn't hold at the boundary!

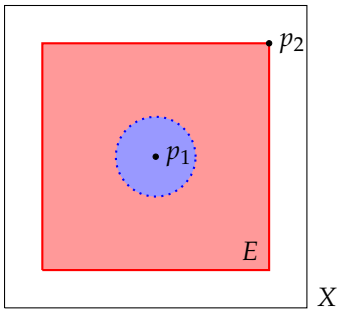

Figure 6: A visualation of an interior point. A set $E \subset X$ is pictured. p_1 is an interior point as there exists $N_r(p_1) \subset E$. p_2 is not an interior point as there does not exist a neighbourhood of p_2 that is entirely contained in E (it is on the boundary).

Definition 2.18: Open Sets

A set $E \subset X$ is **open** if every point of E is an interior point of E.

Theorem 2.19

Every neighbourhood is an open set.

Proof

Consider a neighbourhood $E = N_r(p) \subset X$. Let $q \in E$. We will show that q is an interior point of E. Choose $s < r - d(p,q)$. Then, let $x \in N_s(q)$. By the triangle inequality:

$$d(x,p) \leq d(x,q) + d(q,p) < s + d(q,p) < r - d(p,q) + d(p,q) = r.$$

Hence, $d(x,p) < r$ and it follows that $x \in N_r(p)$. Hence, $N_s(q) \subset N_r(p)$ and q is an interior point of E. □

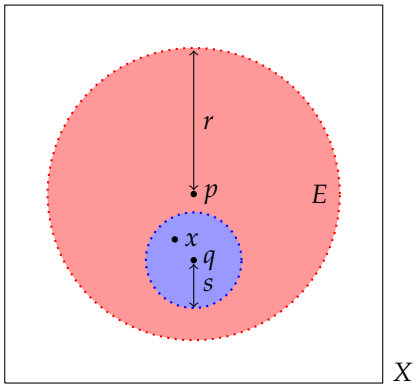

Figure 7: Visualization of the Sets/Points in Theorem 2.19

Definition 2.18: Limit Points/Isolated Points

Let $E \subset X$ and $p \in X$. Then, p is a **limit point** of E if every neighbourhood of p contains $q \in E$, $q \neq p$. If $p \in E$ and p is not a limit point of E, then p is an **isolated point** of E.

Example

Let $E = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. Then, $\frac{1}{2}$ is not a limit point of E, as for $r < \frac{1}{4}$ $N_r(\frac{1}{2})$ does not contain any other points of E. On the other hand, 0 is a limit point of E. For any neighbourhood $N_r(0)$ of 0, $\frac{1}{N} \in N_r(0)$ for $N > \frac{1}{r}$. Note that 0 is the only limit point of E, and is not contained in E (indeed, there is no requirement that a limit point be contained in the set).

Theorem 2.20

If p is a limit point of E, then every neighbourhood of p contains an infinite number of $q \in E$.

Proof

(Sketch) Let $r_1 = 1$. Then, there exists $q_1 \in N_{r_1}(p)$ such that $q_1 \in E$ and $q_1 \neq p$ as p is a limit point of E by assumption. Let $r_2 = d(q_1, p)$. Then, there exists $q_2 \in N_{r_2}(p)$ such that $q_2 \in E$ and $q_2 \neq p$. We can repeat this process to get a (countably infinite) sequence of distinct points $q \in N_{r_1}(p)$, which proves the claim. □

Corollary

If $E \subset X$ is finite, then E has no limit points.

Definition 2.18: Dense Sets

$E \subset X$ is dense in X if every point of X is a limit point of E, or a point of E.

By consequence of Theorem 1.20, we have that $\mathbb{Q}$ is dense in $\mathbb{R}$ and $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$. As a general method to show that a set is dense in another set, take $x \in X \setminus E$, and show x must be a limit point of E.

Definition 2.18: Bounded Sets

$E \subset X$ is bounded if there exists $M \in \mathbb{R}$ and a point $q \in X$ such that $d(p, q) < M$ for all $p \in E$.

For example, $(0, 1), [0, 1], [0, 1] \times [0, 1]$ are bounded, and $(0, \infty), \mathbb{N}, \mathbb{R}$ are unbounded (with the usual metric on $\mathbb{R}$).

Definition 2.18: Closed Sets

A set $E \subset X$ is closed if every limit point of E is in E.

Note that with the above Corollary, we find that every finite set is (trivially) closed.

Definition 2.18: Complement

Let $E \subset X$. Then, the complement of E, denoted E^c is $E^c = \{x \in X : x \notin E\}$.

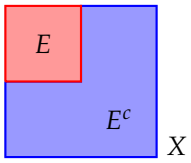

Figure 8: Visualization of a set E and its complement.

Theorem 2.23

A set $E \subset X$ is open if and only if E^c is closed.

Note that this theorem does not imply that all sets are closed or open; it is possible to have a set that is neither closed or open (such as $[0, 1) \subset \mathbb{R}$) and then its complement (with the former example, $(-\infty, 0) \cup [1, \infty)$) which is also neither closed nor open (indeed, most sets are neither closed nor open). As an

additional note, we have that X (the entire metric space) and $\emptyset$ are both open and closed (which we may affectionately label as "clopen").

Proof

$\boxed{\Longrightarrow}$ Assume E is open. If E^c has no limit points, it is trivially closed, so suppose that there exists a limit point x of E^c. Suppose for the sake of contradiction that $x \notin E^c$. Then, $x \in E$. As E is open, x is an interior point of E, so there exists a neighbourhood $N_r(x) \subset E$. In particular, $N_r(x) \cap E^c = \emptyset$, contradicting the fact that x is a limit point of E^c. Hence, $x \in E^c$ and E^c is closed.
$\boxed{\Longleftarrow}$ Assume E^c is closed. Let $x \in E$. In particular, $x \notin E^c$, so x is not a limit point of E^c. So, there exists a neighbourhood $N_r(x)$ which contains no point of E^c, i.e. $N_r(x) \cap E^c = \emptyset$. It follows that $N_r(x) \subset E$, and hence x is an interior point of E. This argument applies to all points of E, hence E is open. □

Corollary

A set $F \subset X$ is closed if and only if F^c is open.

Let $F = E^c$ in Theorem 2.23 to realize the above Corollary.

Theorem 2.24

(a) For any collection $\{E_\alpha\}$ of open sets, $\bigcup_\alpha E_\alpha$ is open.

(b) For any collection $\{F_\alpha\}$ of closed sets, $\bigcap_\alpha F_\alpha$ is closed.

(c) For any finite collection $E_1, \ldots, E_n$ of open sets, $\bigcap_{i=1}^n E_i$ is open.

(d) For any finite collection $F_1, \ldots, F_n$ of closed sets, $\bigcup_{i=1}^n F_i$ is closed.

A point of notation; $\{E_\alpha\}$ can be finite, countable, or uncountable; the indices α are taken from an index set A which can be chosen to be of any cardinality.

Proof

(a) Suppose all sets in $\{E_\alpha\}$ are closed. Let $x \in \bigcup_\alpha E_\alpha$. Then, there exists α_0 such that $x \in E_{\alpha_0}$. Since E_{α_0} is open, there exists a neighbourhood $N_r(x)$ of x such that $N_r(x) \subset E_{\alpha_0} \subset \bigcup_\alpha E_\alpha$. Hence, $\bigcup_\alpha E_\alpha$ is open.

(b) Suppose all sets in $\{F_\alpha\}$ are open. To show that $\bigcap_\alpha F_\alpha$ is closed, we show that $\left(\bigcap_\alpha F_\alpha\right)^c$ is open (by Theorem 2.23). We have that $\left(\bigcap_\alpha F_\alpha\right)^c = \bigcup_\alpha F_\alpha^c$. As all F_α^c are open, by part (a) we have that $\bigcup_\alpha F_\alpha^c$ is also open. Hence $\bigcap_\alpha F_\alpha$ is closed.

(c) Suppose $E_1, \ldots, E_n$ are open. Let $x \in \bigcap_{i=1}^n E_i$, and then we have that $x \in E_i$ for all $i \in \{1, \ldots, n\}$. Hence, there exists r_i such that $N_{r_i}(x) \subset E_i$ as each of the E_is are open. Let $r = \min\{r_1, \ldots, r_n\}$ and then we have that $N_r(x) \subset N_{r_i}(x) \subset E_i$ for all E_i. Therefore, $N_r(x) \subset \bigcap_{i=1}^n E_i$ and $\bigcap_{i=1}^n E_i$ is open.

(d) Suppose $F_1, \ldots, F_n$ are closed. By Theorem 2.23 we have that $\bigcup_{i=1}^n F_i$ is closed if and only if $\left(\bigcup_{i=1}^n F_i\right)^c = \bigcap_{i=1}^n F_i^c$ is open. Since all F_i^cs are open, by part (c) $\bigcap_{i=1}^n F_i^c$ is open, and hence $\bigcup_{i=1}^n F_i$ is closed. □

Example 2.25

We consider some examples to see why the finiteness of the collections in parts (c)/(d) of the theorem are essential. Suppose $E_n = \left(-\frac{1}{n}, \frac{1}{n}\right) \subset \mathbb{R}$. These sets form a countably infinite collection of subsets of $\mathbb{R}$. We then consider that $\bigcap_{n=1}^{\infty} E_n = \{0\}$, which is not open; showing that openness is not preserved under infinite intersections. Next, consider $F_n = [0, 1 - \frac{1}{n}] \subset \mathbb{R}$, which form a countably infinite collection of closed sets in $\mathbb{R}$. We then have that $\bigcup_{n=1}^{\infty} F_n = [0, 1)$ which is not closed as 0 is not an interior point of the set. Hence, closedness is not preserved under infinite unions.

2.4 Closure and Relative Topology

As a review of some definitions of the previous section, we call a set open if every point of the set is an interior point, and closed if it contains all of its limit points. A complement of an open set is closed and vise versa. We also note that openness is preserved under infinite unions and finite intersections, and closedness is preserved under infinite intersections and finite unions.

Definition 2.26: Closure

Let X be a metric space. Let $E \subset X$, and denote E' as the set of all limit points of E. Then, the set $\overline{E} = E \cup E'$ is the **closure** of E.

Theorem 2.27

(a) $\overline{E}$ is closed.

(b) $E = \overline{E}$ if and only if E is closed.

(c) $\overline{E} \subset F$ for every closed set $F \subset X$ such that $E \subset F$.

Proof

(a) Let $p \in \overline{E}^c = E^c \cap (E')^c$. Then, $p \notin E$, and furthermore $p \notin E'$ so p is not a limit point of E. Hence, there exists a neighbourhood $N_r(p)$ such that $N_r(p) \cap E = \emptyset$. Moreover, no point of $N_r(p)$ is in E' (if there existed $q \in E'$ such that $q \in N_r(p)$, then there would exist some $N_{r'}(q)$ (which contains points of E) such that $N_{r'}(q) \subset N_r(p)$ which contradicts the fact that $N_r(p)$ contains no points of E). Hence, $N_r(p) \subset E^c \cap (E')^c = \overline{E}^c$. Hence, $\overline{E}^c$ is open and $\overline{E}$ is closed.

(b) $\boxed{\Longrightarrow}$ If $E = \overline{E}$, then E is closed by (a).

$\boxed{\Longleftarrow}$ If E is closed, then E contains all of its limit points, so $\overline{E} = E \cup E' \subset E$. By definition $\overline{E} \supseteq E$, so $E = \overline{E}$.

(c) Let $E \subset F$ with F closed. Then, $F' \subset F$. Also, $E' \subset F'$. So, by (b) we have $F = \overline{F} = F \cup F' \supseteq E \cup E' = \overline{E}$. □

We will soon define the notion of being relatively open, but before doing so, we consider a motivating example.

In $\mathbb{R}$, we know (a, b) to be in an open set. However, embedded in $\mathbb{R}^2$, (a, b) has no interior points; any neighbourhood about any $x \in (a, b)$ extends into the plane and will inevitably contain points $y \in \mathbb{R}^2$, $y \notin (a, b)$. Hence, (a, b) is open as a subset of $\mathbb{R}$, but not as a subset of $\mathbb{R}^2$.

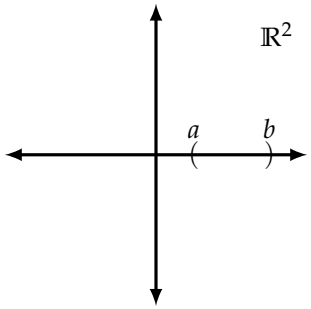

Figure 9: The interval (a, b) embedded in $\mathbb{R}^2$.

Definition 2.29: Relative Openness

Let $E \subset Y \subset X$. Then, E is **relatively open** with respect to Y if it is an open set in the metric space Y.

However, we note that the above definition is not very useful. Is there a better notion of relative openness? Going back to our prior example, consider that (a, b) can be viewed as the intersection of an open disk with $\mathbb{R}$. This gives an equivalent definition!

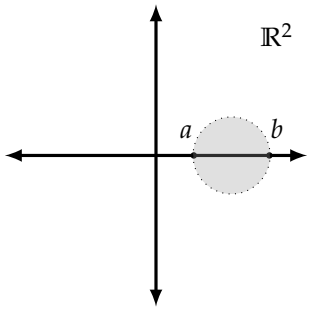

Figure 10: The interval (a, b) can be viewed as the intersection of $\mathbb{R}$ with the open disk $N = \left\{(x, y) \in \mathbb{R}^2 : \sqrt{(x - \frac{b+a}{2})^2 + y^2} < \frac{b-a}{2}\right\}$, giving rise to a more useful notion of relative openness.

Theorem 2.30

Let $E \subset Y \subset X$. Then, E is open relative to Y if and only if there exists an open set $G \subset X$ such that $E = G \cap Y$.

Proof

$\boxed{\Longrightarrow}$ Let $p \in E$. Then, there exists a neighbourhood $N_r^Y(p) \subset E$ (Note that here, $N_r^Y(p) = \{y \in Y : d(p, y) < r\}$). By Theorem 2.24, we have that $G = \bigcup_{p \in E} N_{r_p}^X(p)$ is open. Then, we have that $G \cap Y = \bigcup_{p \in E} N_{r_p}^X(p) \cap Y = \bigcup_{p \in E} N_{r_p}^Y(p) = E$.

$\boxed{\Longleftarrow}$ Suppose $G \subset X$ is open and $E = G \cap Y$. Let $p \in E$. Then, there exists a neighbourhood $N_r^X(p) \subset G$ as G is open, so $N_r^Y(p) = N_r^X(p) \cap Y \subset G \cap Y = E$. Hence, p is an interior point of E in the metric space of Y, and hence E is relatively open in Y. $\square$

2.5 Compactness

Definition 2.31: Open Covers

An **open cover** of a subset $E \subset X$ is a collection $\{G_\alpha\}$ of open sets of X such that $E \subset \bigcup_\alpha G_\alpha$.

Note that the open cover can be either a finite or infinite (countable or uncountable) collection.

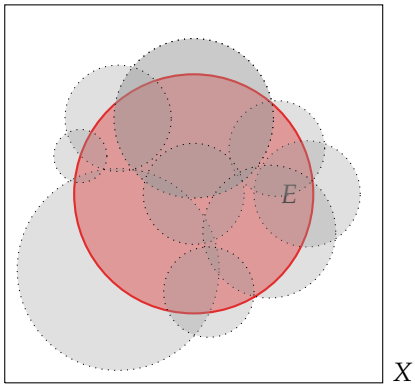

Figure 11: Visualization of a set E and a collection of open disks that form a (finite) open cover of E.

Definition 2.32: Compactness

A subset $K \subset X$ is **compact** if every open cover of K has a finite subcover. Explicitly, if $\{G_\alpha\}$ is an open cover of K, then there exist indices $\alpha_1, \ldots, \alpha_n$ such that $K \subset G_{\alpha_1} \cup \ldots \cup G_{\alpha_n}$.

While the above definition might seem strange/esoteric, it turns out to be very useful and important to analysis. We saw in the last lecture that open sets in a sense do not behave "nicely"; a given set could be open with respect to a subspace Y but not to X (and we had to introduce the notion of relative openness to take care of this fact). Compactness behaves more nicely, as we will see with the following theorem.

Theorem 2.33

Suppose $K \subset Y \subset X$. Then, K is compact with respect to X if and only if K is compact with respect to Y.

While openness depends on the choice of subspace, this theorem states that compactness is independent of what subspace we choose to look at. It is in a sense an intrinsic property of the set.

Proof

$\boxed{\Longrightarrow}$ Suppose K is compact with respect to X. Let $\{V_\alpha\}$ be an open (in Y) cover of K. Then, by Theorem 2.30 we have that there exists $\{G_\alpha\}$ of open sets in X such that $V_\alpha = G_\alpha \cap Y$. Then, $\{G_\alpha\}$ is an open (in X) cover of K, and hence there exists a finite subcover $K \subset \bigcup_{i=1}^n G_{\alpha_i}$. Then, we have that $K \subset \bigcup_{i=1}^n (G_{\alpha_i} \cap Y) = \bigcup_{i=1}^n V_{\alpha_i}$ which is a finite (sub)cover of K in Y. Hence, K is compact with respect to Y.

$\boxed{\Longleftarrow}$ Suppose K is compact with respect to Y. Let $\{G_\alpha\}$ be an open (in X) cover of K. Let $V_\alpha = G_\alpha \cap Y$. These are relatively open in Y by Theorem 2.30, and $\{V_\alpha\}$ still cover K. Hence, as K is compact in Y, there exists a finite subcover $K \subset \bigcup_{i=1}^n V_{\alpha_i} \subset \bigcup_{i=1}^n G_{\alpha_i}$. Therefore, we have found a finite subcover in X, so K is compact in X. □

Theorem 2.34

Let $K \subset X$ be compact. Then, K is closed.

A priori, it might seem like the notions of compactness and closedness are quite different, but this theorem shows that they are indeed quite closely related.

Proof

We show that K^c is open. Let $p \in K^c$. For any $q \in K$. Let $r_q = \frac{1}{2}d(p,q)$, and define $W_q = N_{r_q}(q)$. So, taking the collection $\left\{W_q\right\}$ we have that $\bigcup_{q\in K} W_q$ is an open cover (as it clearly contains every $q \in K$). K is compact, so there exists a finite subcover $\bigcup_{i=1}^{n} W_{q_i}$. Then, let $r = \min\left\{r_{q_1}, \ldots, r_{q_n}\right\}$. Then, $N_r(p) \subset K^c$, hence p is an interior point of K^c. Therefore K^c is open, and K is closed by Theorem 2.23. □

Theorem 2.35

If $F \subset X$ is closed, $K \subset X$ is compact, and $F \subset K$, then F is compact.

Proof

Let $\{V_\alpha\}$ be an open cover of F. Then, $\{V_\alpha\} \cup F^c$ is an open cover of K (F^c is open as F is closed). Hence, as K is compact, there exists a finite subcover. Dropping F^c (if it is part of the finite subcover of K), we obtain a finite subcover of F. Hence, F is compact. □

Corollary

If $F \subset X$ is closed and $K \subset X$ is compact, then $F \cap K$ is compact. In other words, compactness is preserved under intersection (with closed sets).

Proof

If K is compact, it is closed by Theorem 2.34. $F \cap K$ is closed by Theorem 2.24, and since $F \cap K \subset K$, by Theorem 2.35 $F \cap K$ is compact. □

In a previous course, we may have introduced the definition that compact sets are closed and bounded. Here, we started with a significantly different definition, and work up to prove the Heine-Borel Theorem (coming soon!) which tells us that compactness is equivalent to being closed and bounded in $\mathbb{R}^k$. Before we move onto such further properties and theorems of compact sets, we may find it useful to consider some examples of sets that are and are not compact.

The singleton set $\{x\}$ is compact in any metric space (out of any open cover, simply pick a single open set that contains x; this yields the desired subcover). Moreover, any finite set is compact in any topology (consider that for any open cover of a set of n elements, we can just choose at most n open sets corresponding to each element to form our finite subcover).

Next, consider any infinite set Y in a metric space X with the discrete metric:

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

Then is $Y \subset X$ compact? The answer is no; consider the open cover of singleton sets, $U = \{\{x\} : x \in Y\}$. First, note that this is indeed an open cover, as $\{x\}$ is open in X; consider that $N_r(x) \subset \{x\}$ if $r < 1$, and hence x is an interior point of $\{x\}$, making it open. However, there exists no finite subcover of this open cover as we need to pick an infinite number of singletons to cover each point in Y.

$[1, \infty) \subset \mathbb{R}$ is not compact as we could consider the open cover $\{N_1(n) : n \in \mathbb{N}\}$, which has no finite subcover. For the sake of contradiction, suppose such a finite subcover existed. Then, there exists some maximum m for which there would be a neighbourhood around in this subcover. Then, all real numbers greater than $m + 1$ would not be in this subcover, which is a contradiction. In a similar way, we can show that $\mathbb{R} \subset \mathbb{R}, \mathbb{R}^2 \subset \mathbb{R}^2$ are not compact.

Another interesting example to consider is $(0, 1)$. To see that this interval is not compact, consider the open cover $\left\{(\frac{1}{n}, 1) : n \in \mathbb{N}\right\}$. This open cover has no finite subcover. Suppose for the sake of contradiction that a finite subcover existed. Then, there would be some minimum $\frac{1}{N}$ such that $(\frac{1}{N}, 1)$ would be in the subcover. But then no $0 < x < \frac{1}{N}$ would be contained in the subcover, which is a contradiction.

From the Heine-Borel theorem, it is clear here that $[1, \infty), \mathbb{R}, \mathbb{R}^2$ fail due to the sets being unbounded, and the $(0, 1)$ fails due to the set not being closed. We will now move onto a sequence of Theorems that will eventually lead to Heine-Borel (the climax of this chapter)!

Theorem 2.36

Let $\{K_\alpha\}$ is a collection of compact sets such that $\bigcap_{i=1}^n K_{\alpha_i} \neq \emptyset$ for any subcollection/any choice of indices $\alpha_1, \ldots, \alpha_n$. Then, $\bigcap_\alpha K_\alpha \neq \emptyset$.

Proof

Suppose for the sake of contradiction that the assumptions hold but $\bigcap_\alpha K_\alpha \neq \emptyset$. Pick α_0. Then, $K_{\alpha_0} \cap (\bigcap_{\alpha \neq \alpha_0} K_\alpha) = \emptyset$. Hence, $K_{\alpha_0} \subset \left(\bigcap_{\alpha \neq \alpha_0} K_\alpha\right)^c = \bigcup_{\alpha \neq \alpha_0} K_\alpha^c$. Hence, $\{K_\alpha^c\}$ is an open cover of K_{α_0}. K_{α_0} is compact, so there exists a finite subcover $\bigcup_{i=1}^n K_{\alpha_i}^c = \left(\bigcap_{i=1}^n K_{\alpha_i}\right)^c$. But then we have that $K_{\alpha_0} \cap \left(\bigcap_{i=1}^n K_{\alpha_i}\right) = \emptyset$, which is a contradiction as a finite intersection should be nonempty by assumption. □

Corollary

Let $\{K_1, K_2, \ldots\}$ be a collection of nonempty and compact sets such that $K_{i+1} \subset K_i$. Then, $\bigcap_{i=1}^\infty K_i \neq \emptyset$.

Proof

If $n_1 < \ldots < n_m$, then $\bigcap_{i=1}^m K_{n_i} = K_{n_m} \neq \emptyset$. Then, $\bigcap_{i=1}^\infty K_{n_i} \neq \emptyset$ by the above theorem. □

Theorem 2.37

Let K be compact and $E \subset K$ be an infinite set. Then, E has a limit point in K.

Proof

Suppose for the sake of contradiction that no point of K is a limit point of E. Then, for all $q \in K$, then there exists $V_q = N_{r_q}(q)$ such that:

$$V_q \cap E = \begin{cases} \{q\} & \text{if } q \in E \\ \emptyset & \text{if } q \notin E \end{cases}$$

Trivially, $\left\{V_q\right\}$ is an open cover of K. So, there exists a finite subcover $\left\{V_q\right\}_{i=1}^{n}$ of K. So:

$$E \subset K \cap E \subset \left(\bigcup_{i=1}^{n} V_{q_i}\right) \cap E = \bigcup_{i=1}^{n} (V_{q_1} \cup E)$$

But the RHS is a finite set, contradicting the assumption that E is infinite. □

2.6 Compactness in $\mathbb{R}^k$ and the Cantor Set

Theorem 2.38

Let $I_n = [a_n, b_n] \subset \mathbb{R}$ such that $I_{n+1} \subset I_n$ for all n. Then, $\bigcap_{i=1}^{\infty} I_i \neq \emptyset$.

This theorem is very reminiscent of Theorem 2.37. However, we cannot apply it directly as we do not know if $[a, b]$ is compact (though we will show that this is indeed the case by the end of the lecture, we want to avoid circular reasoning)!

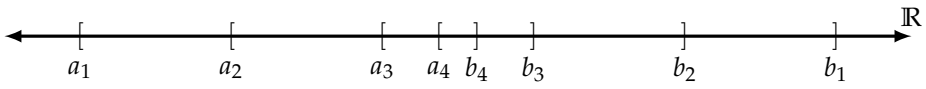

Figure 12: Visualization of the first few sets I_n in Theorem 2.38. Note that this is just an example, and the sets do not need to by "symmetrically shrinking" around a point as pictured.

Proof

We have that $a_n \leq a_{n+m} \leq a_{n+m} \leq b_n$ for all $n, m \in \mathbb{N}$. Let $E = \{a_1, a_2, \ldots\}$. E is nonempty and bounded by any b_n, so by the least upper bound property of $\mathbb{R}$, there exists $x = \sup E \in \mathbb{R}$. We then have that $a_n \leq x \leq b_n$ for all n as x is the least upper bound. Hence, $x \in I_n$ for all n, and hence $x \in \bigcap_{i=1}^{\infty} I_i$. We conclude that $\bigcap_{i=1}^{\infty} I_i \neq \emptyset$. □

Definition: k-cells

A **k-cell** is $I \subset \mathbb{R}^k$ such that $I = \left\{(x_1, x_2, \ldots x_k) \in \mathbb{R}^k : a_j \leq x_j \leq b_j\right\}$ for some $\left\{a_1, \ldots a_j, b_1, \ldots b_j\right\}$.

A k-cell can be viewed as the generalization of a rectangle for $\mathbb{R}^k$.

Theorem 2.39

Theorem 2.38 holds for general k-cells.

The above theorem follows naturally by applying Theorem 2.38 to each coordinate.

Theorem 2.40

Let $I \subset \mathbb{R}^k$ be a k-cell. Then, I is compact.

Proof

Lt $\{G_\alpha\}$ be an open cover of $I = I_0$. Suppose for the sake of contradiction that $\{G_\alpha\}$ has no finite subcover. Let $c_j = \frac{1}{2}(a_j + b_j)$. Splitting up I_0 along each coordinate (i.e. $[a_j, c_j], [c_j, b_j]$) we obtain 2^k k-cells Q_i, with $I_0 = \bigcup_{i=1}^{2^k} Q_i$. At least one of these Q_ns has no finite subcover by assumption. Call this I_1. Then, repeat the same division process using I_1 (and so on). This yields a sequence of k-cells $I_0, I_1, I_2, \ldots$. This sequence has the following properties:

(a) $I_0 \supset I_1 \supset I_2 \supset \ldots$

(b) None of I_ns have a finite subcover from $\{G_\alpha\}$ by construction.

(c) Let $\delta = \sqrt{\sum_{j=1}^{k}(b_j - a_j)^2}$ be the diameter of I. Then, for $x, y \in I_n$, $|x - y| < 2^{-n}\delta$.

By (a) and Theorem 2.39, we have that there exists $\mathbf{x}^*$ such that $\mathbf{x}^* \in \bigcap_{n=1}^{\infty} I_n \subset I$. There exists α_0 such that $\mathbf{x} \in G_{\alpha_0}$, and as G_{α_0} is open, there exists some $r > 0$ such that $N_r(\mathbf{x}^*) \subset G_{\alpha_0}$. By (c), we have that $I_n \subset N_{2^{-n+1}\delta}(\mathbf{x}^*)$. For sufficiently large n, $I_n \subset N_{2^{-n+1}\delta}(\mathbf{x}^*) \subset N_r(\mathbf{x}^*) \subset G_{\alpha_0}$, but this contradicts (b). Hence, there must exist a finite subcover of $\{G_\alpha\}$ for I. □

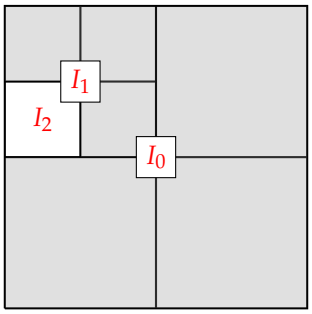

Figure 13: Visualization of the division process (first three iterations shown) in the proof of Theorem 2.40, for $I \subset \mathbb{R}^2$.

Theorem 2.41: Heine-Borel

If a set in $\mathbb{R}^k$ has the following three properties, then it has the other two.

(a) E is bounded and closed.

(b) E is compact.

(c) Every infinite subset of E has at least one limit point in E

In particular, the equivalence of (a) and (b) is what is commonly referred to as the Heine-Borel theorem.

Proof

$\boxed{\text{(a)} \implies \text{(b)}}$ If E is bounded, then there exists a k-cell I such that $E \subset I$. I is compact by Theorem 2.40, and since E is closed, by Theorem 2.35 we have that E is compact.

$\boxed{\text{(b)} \implies \text{(c)}}$ See Theorem 2.37.

$\boxed{\text{(c)} \implies \text{(a)}}$ Suppose first for the sake of contradiction that E is not bounded. Then, E has an infinite subset $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ with $|\mathbf{x}_n| > n$ for all n. Hence, S has no limit points in $\mathbb{R}^k$, and therefore no limit points in E. But this contradicts (c).
Suppose next that E is not closed. Then, there exists $\mathbf{x}_0 \in \mathbb{R}^k$ which is a limit point of E but is not in E. Form $S = \{x_1, x_2, \ldots\} \subset E$ with $|x_n - x_0| < \frac{1}{n}$ for all n. We now show that S has no limit point in $\mathbb{R}^k$ except for $\mathbf{x}_0$. To see this, let $\mathbf{y} \in \mathbb{R}^k, \mathbf{y} \neq \mathbf{x}_0$. Then:

$$|\mathbf{x}_n - \mathbf{y}| \geq |\mathbf{x}_0 - \mathbf{y}| - |\mathbf{x}_n - \mathbf{x}_0| \geq |\mathbf{x}_0 - \mathbf{y}| - \frac{1}{n} \geq |\mathbf{x}_0 - \mathbf{y}| - \frac{n}{2}|\mathbf{x}_0 - \mathbf{y}| = \frac{1}{2}|\mathbf{x}_0 - \mathbf{y}|$$

Where the last inequality holds for sufficiently large n. As every neighbourhood of $\mathbf{y}$ must contain an infinite number of points in S if $\mathbf{y}$ is to be a limit point of S (Theorem 2.20) we find that $\mathbf{y}$ is not a limit point of S. Hence, S has no limit point in E, again contradicting (c). We conclude that E must be bounded and closed. □

We have already shown that compactness implies closed in any metric space (see Theorem 2.34). It also turns out that compactness implies boundedness in any metric space, as well!

Proof. Let $E \subset X$ be compact. Suppose for the sake of contradiction that E was unbounded. Pick any $x_0 \in E$, then E is unbounded, so $\big(N_n(x_0)\big)^c \cap E \neq \emptyset$ for all $n \in \mathbb{N}$. But, $E \subset \bigcup_{n\in\mathbb{N}} N_n(x_0)$, and hence $\{N_n(x_0)\}$ forms an (infinite) subcover of E. By the compactness of E, there exists a finite subcover; but then, there is a maximal radius neighbourhood of x_0 that is contained in this subcover. However, this neighbourhood would not contain all $x \in E$ as E is unbounded, which is a contradiction. □

From this, we have shown that compact implies closed and bounded in any metric space (i.e. the forwards direction of Heine-Borel holds in general). The converse is not always true, however. For one example, consider any infinite set E in a metric space equipped with the discrete metric. E is closed, since E has no limit points (the neighbourhood of radius $r \leq 1$ around any point contains no other point of E). E is bounded as a neighbourhood of radius $r > 1$ around any point contains all points of E. However, as we have discussed previously, E is not compact. Another example would be $\mathbb{R}^\infty$ (i.e. the set of all infinite sequences of real numbers), but the argument for this is left as homework (HW5).

Theorem 2.42: Weierstrauss

Suppose $E \subset \mathbb{R}^k$ is bounded and infinite. Then, E has a limit point in $\mathbb{R}^k$.

Proof

Since E is bounded, $E \subset I \subset \mathbb{R}^k$ for some k-cell I. By Theorem 2.40 E is compact, so by Theorem 2.37 E has a limit point in I (and hence in $\mathbb{R}^k$). □

Having proven the landmark (Heine-Borel) theorem of this section, we close this chapter with discussion of perfect sets and the Cantor set.

Definition 2.18: Perfect Sets

A set $P \subset X$ is **perfect** if P is closed and every point of P is a limit point of P.

Theorem 2.43

Let $P \subset \mathbb{R}^k$ be nonempty and perfect. Then, P is uncountable.

Proof

Since P has limit points, it is not finite by Theorem 2.20. Assume for the sake of contradiction that P is countable. Then, $P = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$. Let $V_1 = N_r(\mathbf{x_1})$ be any neighbourhood of $\mathbf{x}_1$. Suppose we construct B_n such that $V_n \cap P \neq \emptyset$ (note that we don't assume that $\mathbf{x}_n \in V_n$, just that some point $\mathbf{p} \in P$ is in V_n). Since P is perfect, we may construct V_{n+1} such that:

(i) $\overline{V}_{n+1} \subset V_n$

(ii) $\mathbf{x}_n \notin \overline{V}_{n+1}$

(iii) $V_{n+1} \cap P \neq \emptyset$

Note that $\overline{V}_n$ is closed and bounded, and hence is compact by Theorem 2.41. Hence, by the corollary to Theorem 2.35 we have that $K_n = \overline{V}_n \cap P$ is also compact. By (ii), we have that $\mathbf{x}_n \notin \bigcap_{n=1}^{\infty} K_n$, and this holds for all $n \in \mathbb{N}$. Hence, $P \cap \bigcap_{n=1}^{\infty} K_n = \emptyset$, but $K_n \subset P$, so it follows that $\bigcap_{n=1}^{\infty} K_n = \emptyset$. But K_n is non-empty, compact, and $K_{n+1} \subset K_n$, so this contradicts the corollary to Theorem 2.36. We conclude that P is uncountable. □

Corollary

Every interval $[a, b]$ is uncountable. It follows that $\mathbb{R}$ is uncountable.

Definition 2.44: The Cantor Set

Let $E_0 = [0, 1]$. Then, remove the middle third (that is, $(\frac{1}{3}, \frac{2}{3})$) to obtain $E_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Then remove the middle thirds of each of these parts to get $E_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. This yields a sequence of sets $E_0 \supset E_1 \supset E_2 \supset E_3 \supset \ldots$.
Then, we may define the **Cantor set** as $P = \bigcap_{n=1}^{\infty} E_n$.

Although we will rigorously define a measure in this course, each E_n in the above construction is the union of 2^n intervals of measure $\frac{1}{3^n}$, for a total measure of $\frac{2^n}{3^n}$ for each E_n. Taking the limit of $n \to \infty$, we

can therefore see that the Cantor set has measure zero. However, the Cantor set is uncountable (as we will see shortly), making it an example of an uncountable set of measure zero.

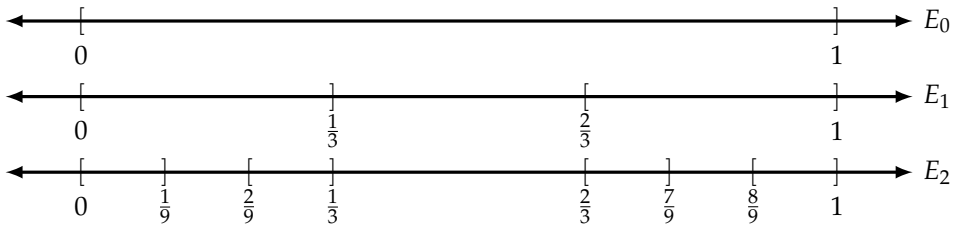

Figure 14: Visualization of the first few sets E_0, E_1, E_2 used in the construction of the Cantor set.

Theorem

The Cantor set contains no interval (a, b).

Proof

(Sketch) Any interval (a, b) contains some interval $\left(\frac{3k+1}{3^m}, \frac{3k+2}{3^m}\right)$ which are all removed in the construction of the set. □

If the Cantor set contains no intervals, then how is it uncountable? It might help to consider that by construction, the endpoints of each E_n belong to P; in the limit, this yields an uncountable number of points contained in P.

Theorem

The Cantor set is uncountable. Moreover, it is perfect.

Proof

Let $X \in P$ and S be an interval such that $x \in S$. Let I_n be the interval of E_n containing x. The length of I_n is given by $\frac{1}{3^n}$. For sufficiently large n, $I_n \subset S$. Let x_n be an endpoint of I_n such that $x_n \neq x$. By construction, $x_n \in P$, and hence x is a limit point of P by construction. Additionally, P is closed by Theorem 2.24 as it is an infinite intersection of closed sets. Hence P is perfect, and by Theorem 2.43 it is uncountable. □

2.7 Connected Sets

Though not covered in lecture, we here briefly discuss connectedness as it comes up later in Chapter 4.

Definition 2.45: Connected Sets

Two subsets A, B of a metric space X are separated if $A \cap \overline{B} = \overline{A} \cap B = \emptyset$. A set $E \subset X$ is connected if E is not a union of two nonempty separated sets.

Separated sets are disjoint, but disjoint sets are not necessarily separated; for example, $A = [0, 1]$ and $B = (1, 2)$ are not separated as $\overline{B} = [1, 2]$ and hence $A \cap \overline{B} = \{1\}$. However, $A = (0, 1)$ and $B = (1, 2)$ are separated. The following theorem characterizes connected subsets of $\mathbb{R}$.

Theorem 2.47

A subset $E \subset \mathbb{R}$ is connected if and only if it has the property that if $x, y \in E$ and $x < z < y$, then $z \in E$.

Proof

Not covered in lecture, see Rudin. □

3 Numerical Sequences and Series

3.1 Sequences

We begin by formally defining a sequence.

Definition: Sequences

Let X be a metric space. A **sequence** is a function $f : \mathbb{N} \mapsto X$. We can denote a term in the sequence as $f(n) = x_n$, or the entire sequence as $\{x_n\}_{n=1}^{\infty}$, $\{x_n\}$, (x_n), or $\{x_1, x_2, x_3 \dots\}$.

We now discuss the notion of convergence of a sequence. Intuitively, we can equate convergence with the notion of points getting closer together.

Definition 3.1: Convergence of Sequences

A sequence $\{p_n\}_{n=1}^{\infty}$ **converges** to $p \in X$ if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $d(p_n, p) < \epsilon$. In this case, we say that $\{p_n\}$ converges to p, or that p is the limit of $\{p_n\}$, and denote this as $p_n \to p$ or $\lim_{n\to\infty} p_n = p$. If $\{p_n\}$ does not converge, we say it **diverges**.

To phrase this definition in another way, we fix some $\epsilon > 0$, and then we have that all points in the sequence past some $N \in \mathbb{N}$ are contained in the neighbourhood $N_\epsilon(p)$. In practice, it can be difficult to apply this definition of convergence if we don't know what the limiting p is, as the definition implicitly uses the value of the limit. We will later discuss another definition of convergence (in $\mathbb{R}^k$) that does not use the value of the limit.

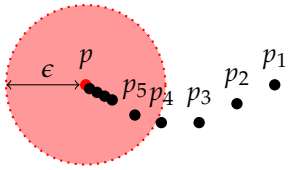

Figure 15: Visualization of a sequence $\{p_n\} \subset \mathbb{R}^2$ converging to a point p. For the $\epsilon > 0$ shown in the picture, we have that all points of the sequence past $N = 5$ lie in the open disk of radius ϵ around p.

As a remark, consider that convergence can depend on our choice of metric space; for example, $\left\{\frac{1}{n}\right\}$ as a sequence in $\mathbb{R}$ converges to 0, but the same sequence in the strictly positive reals ($\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$) does not converge.

Another interesting example (that again shows us the importance of the choice of metric space). Is $\mathbb{R}$ equipped with the discrete metric. A question we can ask is "given some points $p \in \mathbb{R}$, what sequences converge to p?" The answer turns out to be eventually constant sequences only; that is, sequences for which $p_n = p$ for $n \geq N$ for some N.

Proof. If $p_n \to p$, then setting $\epsilon = \frac{1}{2}$, we have that there exists $N \in NN$ such that for all $n \geq N$, $d(p_n, p) < \epsilon = \frac{1}{2}$. Under the discrete metric, this is only possible if $p_n = p$. □

This of course is a strikingly different picture for $\mathbb{R}$ with the standard metric of $d(x, y) = |x - y|$. For example, the sequnce $p_n = \frac{1}{n}$ has no term equal to zero, but converges to $p = 0$. The takeaway message here can be that in the Euclidean metric, points can "get closer" but in the discrete metric, they cannot.

Theorem 3.3

Suppose $\{s_n\}, \{t_n\}$ are complex sequences that converge, with $\lim_{n\to\infty} s_n \to s$ and $\lim_{n\to\infty} t_n \to t$. Then:

(a) $\lim_{n\to\infty}(s_n + t_n) = s + t$.

(b) $\lim_{n\to\infty} cs_n = cs$ and $\lim_{n\to\infty}(c + s_n) = c + s$ for all $c \in \mathbb{C}$.

(c) $\lim_{n\to\infty} s_n t_n = st$

(d) $\lim_{n\to\infty} \frac{1}{s_n} = \frac{1}{s}$ provided $s \neq 0$ and $s_n \neq 0$ for all n.

Proof

(a) Let $\epsilon > 0$. There exist $N_1, N_2 \in \mathbb{N}$ such that $\left|s_{n_1} - s\right| < \frac{\epsilon}{2}$ for $n_1 \geq N_1$ and $\left|t_{n_2} - t\right| < \frac{\epsilon}{2}$ for $n_2 \geq N_2$. Take $N = \max N_1, N_2$, and using the triangle inequality, it follows that for $n \geq N$:

$$\left|(s_n + t_n) - (s + t)\right| \leq |s_n - s| + |t_n - t| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

We conclude that $\lim_{n\to\infty}(s_n + t_n) = s + t$.

(b) Let $\epsilon > 0$. If $c = 0$ then the first sequence trivially converges to 0, so suppose that $c \neq 0$. There exists N such that $|s_n - s| < \frac{\epsilon}{|c|}$ for $n \geq N$, so it follows that:

$$|cs_n - cs| = |c||s_n - s| < |c|\frac{\epsilon}{|c|} = \epsilon.$$

For the second identity, we have that $c_n \to c$ for any constant sequence $c_n = c$ so we may apply (a).

(c) Let $\epsilon > 0$. There exist N_1, N_2 such that $\left|s_{n_1} - s\right| < \sqrt{2}$ for $n_1 \geq N_1$ and $\left|t_{n_2} - t\right| < \sqrt{2}$ for $n_2 \geq N_2$. We then consider that:

$$s_n t_n - st = (s_n - s)(t_n - t) + s(t_n - t) + t(s_n - s)$$

For $n \geq N = \max N_1, N_2$, we have that:

$$(s_n - s)(t_n - t) < \epsilon$$

And we hence observe that $\lim_{n\to\infty}(s_n - s)(t_n - t) = 0$. We can then use (a) and (b) to find that:

$$\lim_{n\to\infty} s(t_n - t) = 0, \quad \lim_{n\to\infty} t(s_n - s) = 0$$

So we conclude that $\lim_{n\to\infty}(s_n t_n - st) = 0$ and hence $s_n t_n \to st$.

Proof (Continued)

(d) Choose m such that $|s_n - s| < \frac{1}{2}|s|$ if $n \geq m$. We then have that $|s_n| > \frac{1}{2}|s|$ for $n \geq m$. Let $\epsilon >)0$. Then, there exists N with $N > m$ such that for $n \geq N$:

$$|s_n - s| < \frac{1}{2}|s|^2\epsilon$$

Hence, for $n \geq N$:

$$\left|\frac{1}{s_n} - \frac{1}{s}\right| = \left|\frac{s_n - s}{s_n s}\right| < \frac{2}{|s|^2}|s_n - s| < \epsilon$$

□

Lemma: Squeeze Lemma

Let $\{x_n\}$, $\{s_n\}$ be real-valued sequences. Then, if $0 \leq x_n \leq s_n$ for all n, and $\lim_{n\to\infty} s_n = 0$, then $\lim_{n\to\infty} x_n = 0$.

Proof

Let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that $n \geq N$ implies $0 \leq s_n < \epsilon$. Then, we have that for $n \geq N$, $0 \leq x_n \leq s_n < \epsilon$ and hence $x_n \to 0$ as claimed. □

Note that we can prove a more generalized version of the Squeeze Lemma.

Lemma: Generalized Squeeze Lemma

Suppose we have sequences $\{l_n\}, \{x_n\}, \{u_n\}$ such that $l_n \leq x_n \leq u_n$ for all n and $\lim_{n\to\infty} l_n = \lim_{n\to\infty} u_n = L \in \mathbb{R}$. Then, $\lim_{n\to\infty} x_n = L$.

Proof

We have that $0 \leq a_n - l_n \leq u_n - l_n$. We have that $\lim_{n\to\infty} u_n - l_n = 0$ by Theorem 3.3(a), so by the (original) Squeeze Lemma we have that $\lim_{n\to\infty} a_n - l_n = 0$. It then follows that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} l_n = L$ as claimed. □

Theorem 3.20

(a) Let $p > 0$. Then, $\lim_{n\to\infty} \frac{1}{n^p} = 0$.

(b) Let $p > 0$. Then, $\lim_{n\to\infty} \sqrt[n]{p} = 1$.

(c) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.

(d) Let $p > 0$ and $\alpha \in \mathbb{R}$. Then, $\lim_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$.

(e) Let $|x| < 1$. Then, $\lim_{n\to\infty} x^n = 0$.

Proof

(a) Let $\epsilon > 0$. Choose N such that $\frac{1}{N^p} < \epsilon$, namely $N > \left(\frac{1}{\epsilon}\right)^{1/p}$. Then, for $n \geq N$, $\frac{1}{n^p} < \frac{1}{N^p} < \epsilon$.

(b) If $p = 1$, the sequence is constant and the conclusion immediate.
If $p > 1$, then let $x_n = \sqrt[n]{p} - 1$. We then have that:

$$p = (x_n + 1)^n = \sum_{k=0}^{n} \binom{n}{k} x_n^k \geq n x_n$$

Where the second equality follows from the binomial theorem (where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$), and the inequality follows by considering that we just keep the $k = 1$ term (and the series is non-negative). Hence, we have that $x_n \leq \frac{p}{n}$, and $x_n \to 0$ by (a).

If $p < 1$, then let $q = \frac{1}{p} > 1$. Then, $\sqrt[n]{q} \to 1$ by the argument above. By Theorem 3.3(d), we then have that $\sqrt[n]{p} = \frac{1}{\sqrt[n]{q}} \to \frac{1}{1} = 1$.

(c) Let $x_n = \sqrt[n]{n} - 1$. Then, we have that:

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

Where the inequality follows from keeping the $k = 2$ term only. We then have that $x_n \leq \sqrt{\frac{2}{n-1}}$ and hence $x_n \to 0$ by the Squeeze Lemma.

(d) We want to show $\frac{n^\alpha}{(1+p)^n} \to 0$; we therefore want an upper bound on the expression, and hence a lower bound on $(1+p)^n$. Applying the Binomial Theorem we have that:

$$(1+p)^n = \sum_{k=0}^{n} \binom{n}{k} p^k = \left((n)(n-1)(n-2)\cdots(n-k+1)\right) \frac{p^k}{n!}$$

Now, we pick $k > \alpha$. For $2n > k$, we then have that:

$$(1+p)^n \geq \left(\frac{n}{2}\right)^k \frac{p^k}{k!}$$

We therefore have that:

$$\frac{n^\alpha}{(1+p)^k} \leq \frac{2^k k!}{p^k} n^{\alpha - k} \to 0$$

And the claim follows by the Squeeze Lemma.

(e) Taking $\alpha = 0$ in (d), the claim follows by setting $|x| = \frac{1}{1+p} < 1$ (as $p > 0$) and recognizing that $x_n \to 0 \iff |x^n| = |x|^n \to 0$. □

3.2 Subsequences

Theorem 3.2

Let $\{p_n\}$ be a sequence in X.

(a) $p_n \to p$ in X if and only if for all $r > 0$, $N_r(p)$ contains all but finitely many points of $\{p_n\}$.

(b) If $p_n \to p$ and $p_n \to p'$ then $p = p'$. In other words, the limit is unique.

(c) If $\{p_n\}$ is convergent, then it is bounded (that is, for any $q \in X$ there exists $M \in \mathbb{R}$ such that $d(q, p_n) \leq M$ for all $n \in \mathbb{N}$).

(d) If $E \subset X$ has a limit point p, then there exists $\{p_n\}$ in E such that $p_n \to p$.

Proof

(a) The claim follows immediately from the definition of convergence; for any $r = \epsilon > 0$, there exists $N \in \mathbb{N}$ such that $N_r(p)$ contains $\{p_n : n \geq N\}$.

(b) There exist N_1, N_2 such that $d(p, p_{n_1}) < \frac{\epsilon}{2}$ if $n_1 \geq N_1$ and $d(p, p_{n_2}) < \frac{\epsilon}{2}$ if $n_2 \geq N_2$. Then for $n \geq N = \max N_1, N_2$ we have (using the triangle inequality) that:

$$d(p, p') \leq d(p, p_n) + d(p_n, p') < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Since ϵ is arbitrary, $d(p, p') = 0$ and hence $p = p'$.

(c) If $p_n \to p$, there exists N such that $d(p_n, p) < 1$ for all $n \geq N$. Set:

$$r = \max\{1, d(p_1, p), d(p_2, p), \ldots, d(p_{N-1}, p)\}$$

For any $q \in X$, we then have that:

$$d(q, p_n) \leq d(q, p) + d(p, p_n) \leq d(q, p) + r$$

so the claim follows with $M = r + d(q, p) + 1$.

(d) Pick $p_n \in E$ such that $d(p_n, p) < \frac{1}{n}$. Let $\epsilon > 0$, and $N > \frac{1}{\epsilon}$. Then, $n \geq N$ implies $\frac{1}{n} \leq \frac{1}{N} < \epsilon$ and hence $d(p_n, p) < \epsilon$ for all $n \geq N$, and hence $p_n \to p$ as desired. □

Definition 3.5: Subsequences

Given $\{p_n\}$ and $n_1 < n_2 < n_3 < \ldots$, we say that $\left\{p_{n_j}\right\}$ is a **subsequence** of $\{p_n\}$.

We first consider some examples. Let $p_n = n$. Then some valid subsequences of $\{p_n\}$ are $\{1, 2, 3, 4, 5, \ldots\}$ (the original sequence), $\{1, 3, 5, 7, \ldots\}$ (the odds), $\{2, 3, 5, 7, 11, 13, \ldots\}$ (the primes). Next, let $p_n = i^n$. We have that $\{p_n\} = \{i, -1, -i, 1, i, -1, -i, 1, \ldots\}$ which is clearly divergent. However, the subsequences $\{i, i, i, \ldots\}$, $\{-1, -1, -1, \ldots\}$, $\{-i, -i, -i, \ldots\}$ and $\{1, 1, 1, \ldots\}$ are all convergent! It is hence possible for a divergent sequence to have a convergent subsequence.

Lemma

If $p_n \to p$, then every subsequence of $\{p_n\}$ converges to p.

Proof

Suppose $p_n \to p$ and let $\{p_{n_j}\}$ be a subsequence of p_n. Let $\epsilon > 0$. Then, there exists some $N \in \mathbb{N}$ such that $d(p, p_n) < \epsilon$ if $n \geq N$. Hence, $d(p, p_{n_j}) < \epsilon$ if $n_j \geq N$ and hence $p_{n_j} \to p$. □

Theorem 3.6: Bolzano–Weierstrass

(a) If $\{p_n\} \subset X$ with X compact, then $\{p_n\}$ has a convergenct subsequence.

(b) If $\{p_n\} \subset \mathbb{R}^k$ and $\{p_n\}$ is bounded, then $\{p_n\}$ has a convergent subsequence.

Proof

(a) Let E be the range of $\{p_n\}$. If E is finite, then there exists $x \in X$ and $n_1 < n_2 < n_3 < \ldots$ such that $p_{n_j} = x$ for all j. Therefore $p_{n_j} \to x$ and we are done. If E is infinite, then by compactness, $E \subset X$ has a limit point in X by Theorem 2.37. By Theorem 3.2(d) there exists a sequence $\{p_{n_j}\}$ in E such that $p_{n_j} \to p$.

(b) By Theorem 2.41, E (being bounded) lies in a compact subset of $\mathbb{R}^k$. We then apply (a). □

3.3 Cauchy Sequences and Completeness

Definition 3.8: Cauchy Sequences

A sequence $\{p_n\} \subset X$ is a **Cauchy sequence** if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$, $d(p_n, p_m) < \epsilon$.

Note the fact that this definition does not refer to a particular p that the sequence may converge to! It instead formalizes the notion of the points of a sequence getting "closer together" as the sequence goes on. It is therefore easier to check if a sequence is Cauchy than if it converges, as we don't need to know the value of the limit. To this end, it is useful to know in what situations a sequence being Cauchy implies that the sequence is convergent. We will soon arrive at a theorem that addresses this question, but first we establish a little more machinery.

Definition 3.9: Diameter

Let $E \subset X$. Then the **diameter** of E, denoted $\operatorname{diam} E$ is defined as $\operatorname{diam} E = \sup\{d(p, q) : p, q \in E\}$. It follows from the definition that a sequence $\{p_n\}$ is Cauchy if and only if $\lim_{N\to\infty} \operatorname{diam} E_n = 0$ where $E_n = \{p_n\}_{n=N}^{\infty}$ (the tail of the sequence).

Example

(a) If $E = (a,b) \subset \mathbb{R}$ or $E = [a,b] \subset \mathbb{R}$, then $\operatorname{diam} E = b - a$.

(b) If $E = (0,1) \times (0,1) \subset \mathbb{R}^2$, then $\operatorname{diam} E = \sqrt{2}$ (the diagonal of the open square).

Theorem 3.10

(a) Let $E \subset X$. Then, $\operatorname{diam} \overline{E} = \operatorname{diam} E$.

(b) If $K_n \subset X$ are compact, $K_{n+1} \subset K_n$ for all n, and $\lim_{n\to\infty} \operatorname{diam} K_n = 0$, then $\bigcap_{n=1}^{\infty} K_n$ consists of exactly one point.

Proof

(a) Since $E \subset \overline{E}$, it is clear that $\operatorname{diam} \overline{E} \geq \operatorname{diam} E$. Next, let $\epsilon > 0$ and $p, q \in \overline{E}$. Choose $p', q' \in E$ such that $d(p,p') < \frac{\epsilon}{2}$, $d(q,q') < \frac{\epsilon}{2}$ (this choice is possible as either p, q are in E, or p, q are limit points of E). Then, we have that:

$$d(p,q) \leq d(p,p') + d(p',q) \leq d(p,p') + d(p',q') + d(q',q) < \frac{\epsilon}{2} + \operatorname{diam} E + \frac{\epsilon}{2} = \operatorname{diam} E + \epsilon$$

ϵ, p, and q are arbitrary, so it follows that $\operatorname{diam} \overline{E} \leq \operatorname{diam} E + \epsilon$ from the definition of the diameter. It then follows that $\operatorname{diam} \overline{E} \leq \operatorname{diam} E$. We conclude that $\operatorname{diam} \overline{E} = \operatorname{diam} E$.

(b) Let $K = \bigcap_{n=1}^{\infty} K_n$. By the corollary to Theorem 2.36, we have that $K \neq \emptyset$, so K contains at least one point. Since $K \subset K_n$, it follows that $\operatorname{diam} K \leq \operatorname{diam} K_n$ for any n, and since $\operatorname{diam} K_n \to 0$, $\operatorname{diam} K = 0$. If there were $p, q \in K$ such that $p \neq q$, then $\operatorname{diam} K \neq 0$, so it must follow that K has at most one point. We conclude that K has exactly one point. □

Lemma

If a sequence $\{p_n\}$ is Cauchy, then it is bounded.

Proof

If $\{p_n\}$ is Cauchy, then we have that $\lim_{N\to\infty} \operatorname{diam} E_N = \lim_{N\to\infty} \operatorname{diam} \{p_n\}_{n=N}^{\infty} = 0$. Then for some $N \in \mathbb{N}$, $\operatorname{diam} E_N < 1$. The range of $\{p_n\}$ is the union of E_N and the finite set $\{p_1, \ldots, p_{N-1}\}$ and hence $\{p_n\}$ is bounded. □

Theorem 3.11

(a) If a sequence $\{p_n\} \subset X$ converges, then it is Cauchy.

(b) If a sequence $\{p_n\} \subset X$ is Cauchy and X is compact, then $\{p_n\}$ converges to some $p \in X$.

(c) In $\mathbb{R}^k$, every Cauchy sequence is convergent.

Proof

(a) Let $p_n \to p$ and let $\epsilon > 0$. There exists $N \in \mathbb{N}$ such that $d(p_n, p) < \frac{\epsilon}{2}$ if $n \geq N$. Then, for $n, m \geq N$, we have that:

$$d(p_n, p_m) \leq d(p_n, p) + d(p, p_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

so $\{p_n\}$ is Cauchy.

(b) Let $E_N = \{p_n\}_{n=N}^{\infty}$. Then, $\overline{E}_N \subset X$ is closed, so by the compactness of X we have that $\overline{E}_N$ is compact by Theorem 2.35. Since $E_{N+1} \subset E_N$, we have that $\overline{E}_{N+1} \subset \overline{E}_N$, and additionally we have that $\lim_{N\to\infty} \overline{E}_N = \lim_{N\to\infty} E_N = 0$ where the first equality follows from Theorem 3.10(a) and the second equality follows from the fact that $\{p_n\}$ is Cauchy and Definition 3.9. Thus, Theorem 3.10(b) says that there exists a unique point $p \in \bigcap_{n=1}^{\infty} \overline{E}_N$. Next, let $\epsilon > 0$. Then, there exists N_0 such that $\operatorname{diam} \overline{E}_N < \epsilon$ for all $N \geq N_0$. So, $d(p, q) < \epsilon$ for all $q \in \overline{E}_N$, so the same holds for all $q \in E_N$. Hence, $d(p, p_n) < \epsilon$ for all $n \geq N_0$, which shows that $p_n \to p$ and proves the claim.

(c) By the above Lemma, Cauchy sequences are bounded. Hence, $\{p_n\} \subset I$ for some k-cell $I \subset \mathbb{R}^k$. Since I is compact in $\mathbb{R}^k$, the claim follows from (b). □

Definition 3.12: Completeness

A metric space X is called **complete** if every Cauchy sequence converges in X.

It might be tempting at first to think that every space would be complete, but this is not the case. For example, something that can go wrong is a sequnece can be Cauchy, but the limit can lie "outside" of the space. To see this, consider again the sequence $\left\{\frac{1}{n}\right\}$ in the metric space $\mathbb{R}^+ = \mathbb{R} \setminus \{x \in \mathbb{R} : x \leq 0\}$. The sequence is Cauchy, but does not converge in $\mathbb{R}^+$ (as it converges to 0, which lies outside of the space).

Example

(i) Compact sets are complete by Theorem 3.11(b).

(ii) $\mathbb{R}^k$ (and $\mathbb{C}$) are complete by Theorem 3.11(c).

(iii) $\mathbb{Q}$ is not complete. We can make a sequence of rational points that converges to an irrational number in $\mathbb{R}$ which is Cauchy, but does not converge in $\mathbb{Q}$ (Example 1.1 gives a way one might construct such a sequence).

Note that $\mathbb{Q}$ can be completed to $\mathbb{R}$, and in general for any (X, d) which is not complete, there exists (X^*, d^*) that is complete such that $|X| = X^*$. Indeed, this is another way we can construct the real numbers! $\mathbb{R}$ can be viewed as equivalence classes of Cauchy sequences in $\mathbb{Q}$. The idea is to define an equivalnence relation $\sim$ such that $p_n \sim q_n$ if $\lim_{n\to\infty} d(p_n, q_n) = 0$. X^* is then defined as the set of equivalence classes under that equivalence relation, equipped with the metric $d^*([p], [q]) = \lim_{n\to\infty} d(p_n, q_n)$. It can then be checked that d^* is a valid metric and that X^* is complete. For the full proof, see HW7, or exercises 3.23-3.25 in Rudin (note: this proof is quite technical/difficult).

To motivate the next theorem, consider that all convergent sequences in $\mathbb{R}$ (and in general) are bounded (as we saw in Theorem 3.2(c)). However, this is not always true; for example consider $p_n = (-1)^n$ which is clearly bounded but divergent. What then are conditions that a bounded sequence may converge?

Definition 3.13: Monotonic Sequences

A sequence $\{p_n\} \subset \mathbb{R}$ is **monotonically increasing** if $p_{n+1} \geq p_n$ for all n, and **montonically decreasing** if $p_{n+1} \leq p_n$.

Theorem 3.14

Suppose $\{p_n\} \subset \mathbb{R}$ is montonic. Then, $\{p_n\}$ is convergent if and only if it is bounded.

Proof

$\boxed{\Longrightarrow}$ See Theorem 3.2(c).
$\boxed{\Longleftarrow}$ We show the proof for the increasing case as the decreasing case is analogous. Let $p = \sup p_n : n \in \mathbb{N}$ which exists as $\{p_n\}$ is bounded and $\mathbb{R}$ has the LUB property. Then, $p_n \leq p$ for all n. Let $\epsilon > 0$. Then, there exists $N \in \mathbb{N}$ such that $p - \epsilon < p_N < p_{N+1}$. By the monotonicity of $\{p_n\}$, it follows that $|p_n - p| < \epsilon$ for all $n \geq N$. Hence, $p_n \to p$. □

Definition 3.15: Limits to Infinity

Let $\{p_n\} \subset \mathbb{R}$. If for all $M \in \mathbb{R}$, there exists $N \in \mathbb{N}$ such that $p_n > M$ for all $n \geq N$, then we write $p_n \to \infty$. If instead for all $M \in \mathbb{R}$ there exists $N \in \mathbb{N}$ such that $p_n < M$ for all $n \geq N$, then we write $p_n \to -\infty$.

3.4 Limit Supremum and Limit Infimum

As a motivating question, how would we say something about the largest and smallest accumulation points of a sequence? This leads us to the following definition.

Definition 3.16: limsup and liminf

Let $\{s_n\} \subset \mathbb{R}$, then, we define the **limit supremum** as:

$$\limsup_{n\to\infty} s_n = \inf_{n\geq 1} \sup_{m\geq n} s_m = \lim_{n\to\infty} \sup_{m\geq n} s_m$$

And the **limit infimum** as:

$$\liminf_{n\to\infty} s_n = \sup_{n\geq 1} \inf_{m\geq n} s_m = \lim_{n\to\infty} \inf_{m\geq n} s_m$$

Note that unlike the limit of a real-valued sequence, the limsup and liminf always exist.

In the above definition, the equivalence of $inf_{n\geq 1} \sup_{m\geq n} s_m$ and $\lim_{n\to\infty} \sup_{m\geq n} s_m$ may be slightly confusing. To see this, consider the fact that the sequence $q_n = \sup\{p_n : n \geq 1\}$ is a strictly decreasing sequence in n (with increasing n, we take the supremum over less terms each time), so taking the limit of $n \to \infty$ or taking the infimum over n are equivalent.

Note that Rudin defines the limsup/liminf differently, but perfectly equivalently. Namely, if $\{p_n\} \subset \mathbb{R}$ is a sequence, then E is the set of all subsequential limits (i.e. there is a subsequence of $\{p_n\}$ with a given limit). Then, $\limsup^* p_n = \sup E$ (we use the $*$ to denote Rudin's definition). The equivalence is not immediately obvious, so we here give a sketch to show that the two definitions coincide. We will use the

general technique of showing that the two expressions are ϵ close to one another. Namely, for any $\epsilon > 0$, we show that the following two statements hold:

1) $\limsup^* p_n \leq \limsup p_n + \epsilon$

2) $\limsup^* p_n + \epsilon \geq \limsup p_n$

To show 1), for each N, we let n_N be an index such that p_{n_N} satisfies:

$$p_{n_N} \geq \sup\{p_n : n \geq N\} - \frac{\epsilon}{N}$$

And then we claim that $\lim_{n\to\infty} p_{n_N} = \limsup p_n$ (Exercise). There is one slight technical issue in that $\{p_{n_N}\}$ may not be a subsequence of $\{p_n\}$, in particular we don't know that $n_{N_1} < n_{N_2} < n_{N_3} < \dots$ just by the above construction (but in order for this to be a valid subsequence, we need this to be the case). Fortunately, this is a fixable issue. We do know that $n_{N_1} \geq 1$, $n_{N_2} \geq 2$, $n_{N_3} \geq 3$ by construction, so if it turns out to be the case that $n_{N_1} < n_{N_2}$ doesn't hold, we can skip ahead into the sequence until we find the first j for which the equality holds. Concretely, if $n_{N_1} = 1000$ (as an example), then if we skip ahead to $n_{N_{1001}}$, it is guaranteed that $n_{N_1} < n_{N_{1001}}$ and we can from there construct a valid sequence. The sketch for 2) is left as an exercise.

Example

(a) Consider $s_n = (-1)^n\left(1 + \frac{1}{n^2}\right)$. Then, we have that $1 \leq \sup_{m\geq n} s_n \leq 1 + \frac{1}{1^2} = 2$, so $\limsup_{n\to\infty} s_n = 1$. Similarly, $\limsup_{n\to\infty} s_n = -1$. Note that this sequence has no limit (it oscillates indefinitely and does not converge) but these quantities are well defined. We notice that the limsup is greater than the liminf in this case, and indeed it is true in general that $\liminf_{n\to\infty} s_n \leq \limsup_{n\to\infty} s_n$.

(b) If $\{s_n\}$ is not bounded above, then $\sup_{m\geq n} s_n = \infty$ for all n and we write $\limsup_{n\to} s_n = \infty$. Similarly, if $\{s_n\}$ is not bounded below, then $\inf_{m\geq n} s_n = -\infty$ for all n and we write $\limsup_{n\to\infty} s_n = -\infty$.

One difficulty with discussing the convergence of a sequence is that the definition is difficult to apply; we need to know what the sequence converges to. The notion of a Cauchy sequence then begins helpful (as Cauchy and convergent are equivalent in complete metric spaces). In addition, it is helpful to consider the limsup and liminf, as we can bound the limit above and below with these quantities respectively. In particular, the limit of the supremum of the tail of the sequence equals the limit of the infinimum of the tail of the sequence equals the limit of the sequence if the sequence is convergent.

Theorem 3.18

Let $\{s_n\} \subset \mathbb{R}$. Then, $\lim_{n\to\infty} s_n = L$ if and only if $\limsup_{n\to\infty} s_n = \liminf_{n\to\infty} = L$.

Proof

$\boxed{\Longrightarrow}$ Let $\epsilon > 0$. Then, there exists $N \in \mathbb{N}$ such that $s_m \in (L-\epsilon, L+\epsilon)$ for all $m \geq N$¿ Then, we have that:

$$L - \epsilon \leq \inf_{m \geq N} s_m \leq \sup_{m \geq N} \leq L + \epsilon$$

Taking limits we have:

$$L - \epsilon \leq \liminf_{n \to \infty} \leq \limsup_{n \to \infty} s_n \leq L + \epsilon$$

ϵ is arbitrary, so we have that $\limsup_{n\to\infty} s_n = \liminf_{n\to\infty} = L$.
$\boxed{\Longleftarrow}$ We have that:

$$\inf_{m \geq n} s_n \leq s_n \leq \sup_{m \geq n} s_m$$

for all $n \in \mathbb{N}$. By assumption we have $\limsup_{n\to\infty} s_n = \liminf_{n\to\infty} = L$ so by the Generalized Squeeze Lemma we conclude that $\lim_{n\to\infty} s_n = L$. □

3.5 Series

Definition 3.21: Infinite Series

Let $\{a_n\} \subset \mathbb{C}$. We then form a new sequence of $s_n = \sum_{j=1}^{n} a_j$ (the sequence of partial sums). Then, if $s_n \to s$. We say that the series $\sum_{j=1}^{\infty} a_j$ converges, and write $\sum_{j=1}^{\infty} a_j = s$. If s_n does not converge, we say that $\sum_{j=1}^{\infty} a_j$ diverges. As a notational point, we will sometimes omit the bounds of summation and write $\sum a_j$ to denote an infinite series, where the meaning is clear from context.

Note that series are just a specific subset of sequences. Although the above definition states that $\{a_n\} \subset \mathbb{C}$, it is in general possible to define series over general vector spaces, with $\{a_n\} \in V$ and $\{s_n\} \in V$.

Note that this generalization allows us to state an equivalent notion of completeness for vector spaces, namely that V is compelte if and only if for all sequences $\{a_n\} \subset V$, the sum $\sum_{n=0}^{\infty} \|a_n\|$ converges to a point in V. As before, $\mathbb{R}, \mathbb{R}^k$ (both over the field $\mathbb{R}$), are complete vector spaces. An example of a vector space that is not complete is the set of functions $f : \mathbb{R} \mapsto \mathbb{C}$ such that $\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$ (where the integral is the familiar Riemann integral from first year calculus, to be defined more rigorously in Chapter 6). For example, the sequence $f_n \subset V$ such that:

$$f_n = \begin{cases} 1 & \text{for the first } n \text{ rationals} \\ 0 & \text{elsewhere} \end{cases}$$

does not converge to a function in V. In fact, Lebesgue integration (which is not covered in this course, but will be the primary focus of a course in measure theory) deals with this issue.

We will now restate the Cauchy criterion (Theorem 3.11) for series.

Theorem 3.22

A series $\sum a_j$ converges if and only if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\left|\sum_{j=m}^{n} a_j\right| < \epsilon$ for all $n \geq m \geq N$.

Proof

$\sum a_j$ converges if and only if $\{s_n\}$ converges if and only if $\{s_n\}$ is Cauchy. So by definition there exists N such that for $n \geq m-1 \geq N$:

$$|s_n - s_m| = \left| \sum_{j=1}^{n} a_j - \sum_{j=1}^{m-1} a_j \right| = \left| \sum_{j=m}^{n} a_j \right| < \epsilon$$

which proves the claim. □

Theorem 3.23: Divergence Test

If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

Proof

Choose $n = m$ in Theorem 3.22. □

Note that this criteria gives us the ability to easily check if a series diverges; if a_j does not converge to 0, then $\sum a_j$ diverges. However, note that the reverse implication does NOT hold! $\sum \frac{1}{n}$ diverges (as we will show next lecture) even though clearly $\frac{1}{n} \to 0$.

Theorem 3.24

Let $\{a_n\} \subset \mathbb{R}$ and $a_n \geq 0$ for all n. Then, we have that $\sum a_n$ converges if and only if the sequence of partial sums $\{s_n\}$ is bounded.

Proof

If $a_n \geq 0$, then $\{s_n\}$ is monotonically increasing. Then by Theorem 3.14, $\{s_n\}$ converges if and only if it is bounded. □

Theorem 3.25: Comparison Test

(a) Let $\{a_n\} \subset \mathbb{C}$ and $\{c_n\} \subset \mathbb{R}$. Then, if $|a_n| \leq c_n$ for all $n \geq N_0$ for some $N_0 \in \mathbb{N}$, and $\sum c_n$ converges, then $\sum a_n$ converges.

(b) Let $\{a_n\} \subset \mathbb{R}$ and $\{d_n\} \subset \mathbb{R}$. If $a_n \geq d_n \geq 0$ for all $n \geq N_0$ for some $N_0 \in \mathbb{N}$ and $\sum d_n$ diverges, then $\sum a_n$ diverges.

Proof

(a) Let $\epsilon > 0$. Then, there exists $N \in \mathbb{N}$ such that $\sum_{j=m}^{n} c_n < \epsilon$ for all $n \geq m \geq N$. Then, take $N \geq N_0$, and we have that:

$$\left| \sum_{j=m}^{n} a_j \right| \leq \sum_{j=m}^{n} \left| a_j \right| \leq \sum_{j=m}^{n} c_j < \epsilon$$

Where in the first inequality we apply the triangle inequality (Theorem 1.37). We conclude that $\sum a_j$ converges by the Cauchy criterion (Theorem 3.22).

(b) The claim follows by considering the contrapositive of (a). □

Theorem 3.26: The Geometric Series

If $0 \leq x < 1$, then $\sum_{j=0}^{\infty} x^j = \frac{1}{1-x}$. If $x \geq 1$, then $\sum_{j=0}^{\infty} x^j$ diverges.

Proof

Suppose $x = 1$. Then, $x_n = 1$ does not converge to zero, so by Theorem 3.24 $\sum_{j=0}^{\infty} x^j$ diverges. Suppose then that $x \neq 1$. We have that $s_n = 1 + x + x^2 + \ldots + x^n$. Hence, $xs_n = x + x^2 + x^3 + \ldots + x^n + x^{n+1}$. Hence, $(1-x)s_n = 1 + x^{n+1}$ and therefore a closed form expression for s_n is:

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

If $x > 1$, we have that s_n diverges (again) by Theorem 3.24. If $x < 1$, then $x^{n+1} \to 0$ by 3.20(e), so $s_n \to \frac{1}{1-x}$. □

Note that by the comparison test, the above result can be generalized to see that $\sum z^j$ for $z \in \mathbb{C}$ converges for $|z| < 1$ and diverges for $|z| \geq 1$.

Theorem 3.27: Cauchy Condensation Test

Suppose $\{a_n\} \subset \mathbb{R}$ and $a_1 \geq a_2 \geq a_3 \geq \ldots \geq 0$. Then, $\sum a_j$ converges if and only if $\sum_{n=1}^{\infty} 2^n a_{2^n}$ converges.

Proof

$\boxed{\Longrightarrow}$ For $2^k < n$, using the fact that the sequence is decreasing, we have that:

$$\begin{aligned} a_1 + a_2 + \ldots + a_n &\geq a_1 + a_2 + (a_3 + a_4) + \ldots + (a_{2^{k-1}+1} + \ldots + a_{2^k}) \\ &\geq \frac{1}{2}a_1 + a_2 + 2a_4 + \ldots + 2^{k-1}a_{2^k} \\ &= \frac{1}{2}(a_1 + 2a_2 + 4a_4 + \ldots + 2^k a_{2^k}) \end{aligned}$$

Hence by the comparison test (Theorem 3.25) we have that $\sum 2^n a_{2^n}$ converges if $\sum a_j$ converges.
$\boxed{\Longleftarrow}$ We show the contrapositive. For $2^k > n$, We have that:

$$\begin{aligned} a_1 + a_2 + \ldots + a_n &\leq a_1 + (a_2 + a_3) + \ldots + (a_{2^k} + \ldots + a_{2^{k+1}-1}) \\ &\leq a_1 + 2a_2 + 4a_4 + \ldots + 2^k a_{2^k} \end{aligned}$$

Hence by the comparison test (Theorem 3.25) we have that $\sum 2^n a_{2^n}$ diverges if $\sum a_j$ diverges. □

3.6 p-Series and Euler's Number

We now use the result of Theorem 3.27 to prove a result about a familiar subset of series.

Theorem 3.28: p-Series

$\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \leq 1$.

Proof

If $p \leq 0$, then n^p does not converge to 0 and hence $\sum \frac{1}{n^p}$ diverges by Theorem 3.23. If $p > 0$, then $\frac{1}{n^p}$ is a monotonically decreasing sequence of positive terms. Hence, we can apply the result of Theorem 3.27. $\sum \frac{1}{n^p}$ converges if and only if $\sum_k 2^k \frac{1}{2^{kp}} = \sum_k \left(\frac{1}{2^{p-1}}\right)^k$ converges. By Theorem 3.26, this last expression is convergent if and only if $0 < \frac{1}{2^{p-1}} < 1$, i.e. if $p > 1$, proving the claim. □

From the above result, we have that the p-series converges if $p > 1$ and diverges otherwise. Is there something "in between" these two regions?

Theorem 3.29

$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$ converges if $p > 1$ and diverges if $p \leq 1$.

Proof

$\log n$ is monotonically increasing for $p > 0$. So, $\frac{1}{n(\log n)^p}$ is monotonically decreasing. Hence, $\sum \frac{1}{n(\log n)^p}$ converges if and only if $\sum_k 2^k \frac{1}{2^k} \frac{1}{(\log 2^k)^p} = \sum_k \frac{1}{k^p}$ converges, and hence the claim follows from Theorem 3.28. □

Definition 3.30: Euler's Number

$e = \sum_{n=0}^{\infty} \frac{1}{n!}$, where $0! = 1$ and $n! = n \cdot (n-1)! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ for $n \geq 1$.

We should check that this expression is well defined (namely, that the series converges). We first observe that $\frac{1}{n!} \leq \frac{1}{2^{n-1}}$ for $n \geq 1$. Therefore, we have that:

$$s_n = \sum_{j=0}^{n} \frac{1}{j!} \leq 1 + \sum_{j=1}^{n} \frac{1}{2^{j-1}} \leq 1 + \sum_{j=0}^{\infty} \frac{1}{2^j} = 1 + \frac{1}{1 - \frac{1}{2}} = 3$$

Where we use Theorem 3.26 in the second last equality. Hence, we conclude that $\sum_{j=0}^{n} \frac{1}{j!}$ converges by comparison, and moreover, that $0 < e < 3$.

It will also be of interest to investigate the rate of convergence of this series. To this end, we observe:

$$\begin{aligned} 0 < e - s_q = \frac{1}{(q+1)!} + \frac{1}{(q+2)!} + \ldots &\leq \frac{1}{(q+1)!} \left(1 + \frac{1}{q+1} + \frac{1}{(q+1)^2} + \ldots \right) \\ &= \frac{1}{(q+1)!} \frac{1}{1 - \frac{1}{q+1}} = \frac{1}{(q+1)!} \frac{q+1}{q} = \frac{1}{q!q} \end{aligned}$$

Hence we have that the error goes to zero extremely quickly! Moreover, we can use this fact to show that e is irrational.

Theorem 3.32: Irrationality of e

$e \notin \mathbb{Q}$.

Proof

Suppose $e = \frac{p}{q}$ for $p, q \in \mathbb{N}$. Then, by the argument above, we have that $0 < e - s_q < \frac{1}{q!q}$. Hence, we have that $0 < q!q - q!s_q < \frac{1}{q} \leq 1$. But, $q!e = q!\frac{p}{q} = (q-1)!p \in \mathbb{N}$, and $q!s_q = \sum_{j=0}^{q} q!\frac{1}{j!} \in \mathbb{N}$. Hence, $q!e - q!s_q \in \mathbb{Z}$, but this is a contradiction as $0 < q!e - q!s_q < 1$. □

There is another familiar definition of e involving a limit that one may recall from first year calculus. These definitions are equivalent, as we will show in the next theorem.

Theorem 3.31

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

Proof

Let $t_n = \left(1 + \frac{1}{n}\right)^n$. By the Binomial theorem, we have that:

$$t_n = \sum_{j=0}^{n} \binom{n}{j} \left(\frac{1}{n}\right)^j = \sum_{j=0}^{n} \frac{1}{j!} \left(\frac{n}{n} \cdot \frac{(n-1)}{n} \cdots \frac{(n-j+1)}{n}\right) \leq s_n$$

Where the last inequality follows from the fact tha the term in brackets is less than (or equal to) 1. Hence, we have that $\limsup_{n\to\infty} t_n \leq \limsup_{n\to\infty} s_n = \lim_{n\to\infty} s_n = e$.
On the other hand, fix $m \in \mathbb{N}$ and let $n \geq m$. Then, we have that:

$$t_n \geq \sum_{j=0}^{m} \binom{n}{j} \frac{1}{n^j} = \sum_{j=0}^{m} \frac{1}{j!} \left(\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{j-1}{n}\right)\right)$$

The infimum of the term in the brackets is just 1, so we therefore have that:

$$\inf_{n\geq m} t_n = \sum_{j=0}^{m} \frac{1}{j!} = s_m$$

Now, we take $m \to \infty$ to find that $\liminf_{m\to\infty} t_m \geq \liminf_{m\to\infty} s_m = \lim_{m\to\infty} s_m = e$.
Having shown that $\liminf_{n\to\infty} t_n \geq e \geq \limsup_{n\to\infty} t_n$, we conclude that $\limsup_{n\to\infty} = \liminf_{n\to\infty} = e$ and hence $\lim_{n\to\infty} t_n = e$. □

3.7 The Ratio and Root Tests

Theorem 3.33: The Root Test

Let $\sum a_n$ be a series, and put $\alpha = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$. Then,

(i) $\sum a_n$ converges if $\alpha < 1$.

(ii) $\sum a_n$ diverges if $\alpha > 1$.

(iii) If $\alpha = 1$, the test is inconclusive.

Proof

(i) Suppose $\limsup_{n\to\infty} \sqrt[n]{|a_n|} = \alpha < 1$. Take β such that $\alpha < \beta < 1$ and $N \in \mathbb{N}$ such that $\sqrt[n]{|a_n|} < \beta$ for all $n \geq N$. Hence, for $n \geq N$, $|a_n| < \beta^n$, and $\beta < 1$. The result follows by using the comparison Test with the geometric series.

(ii) Suppose $\limsup_{n\to\infty} \sqrt[n]{|a_n|} = \alpha > 1$. Then, there exists a subsequence such that $\sqrt[n_j]{\left|a_{n_j}\right|} \to \alpha$. Therefore, there exists N such that for $j \geq N$, $\sqrt[n_j]{\left|a_{n_j}\right|} > 1$, that is to say, $\sqrt[n_j]{\left|a_{n_j}\right|} > 1$ for infinitely many terms. Hence, $\sqrt[n]{|a_n|}$ does not converge to zero, and hence the series does not converge by the divergence test.

(iii) Consider $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$. $\alpha = 1$ for both sums, but by Theorem 3.28 the former diverges and the latter converges. □

Theorem 3.34: The Ratio Test

Let $\sum a_n$ be a series such that $a_n \neq 0$ for all n. Then,

(i) $\sum a_n$ converges if $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$.

(ii) Diverges if there exists N_0 such that $\left|\frac{a_{n+1}}{a_n}\right| \geq 1$ for all $n \geq N_0$.

In other cases, the test is inconclusive.

Proof

(i) By assumption, there exists $\beta < 1$ such that for some N, $\left|\frac{a_{n+1}}{a_n}\right| < \beta$ for all $n \geq N$. We then have that $|a_{N+1}| < \beta|a_N|$, that $|a_{N+1}| < \beta|a_{N+1}| < \beta^2|a_N|$ and inductively we obtain that $\left|a_{N+p}\right| < \beta^p|a_N|$. In other words, we have that for $n \geq N$, $|a_n| < |a_N|\beta^{-N}\beta^n$. Since $\sum \beta^n$ converges (convergent geometric series), $\sum a_n$ converges by the comparison test.

(ii) For $n \geq N_0$, we have that $|a_n| \leq |a_{N+1}|$. Hence, a_n does not converge to 0, and the claim follows by the divergence test. □

As a remark, the the ratio test is less powerful than the root test. For any series for which the ratio test is conclusive, the root test is also conclusive. But the converse is not true. However, the ratio test is easier to apply in practice. We also note that the above ratio test implies the (perhaps more familiar) version from first year calculus:

Corollary

Let $\sum a_n$ be a series such that $a_n \neq 0$ for all n. Then:

(i) $\sum a_n$ converges if $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$.

(ii) $\sum a_n$ diverges if $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| > 1$.

(iii) $\sum a_n$ diverges if $\liminf \left|\frac{a_{n+1}}{a_n}\right| > 1$.

Example 3.35

Consider the series:

$$\frac{1}{2} + 1 + \frac{1}{8} + \frac{1}{4} + \frac{1}{32} + \frac{1}{16} + \dots$$

We then have that the ratio $\frac{a_{n+1}}{a_n}$ is the sequence $2, \frac{1}{8}, 2, \frac{1}{8}, \dots$. Therefore, $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = 2 > 1$ and $\liminf_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \frac{1}{8} < 1$ and the ratio test is inconclusive/tells us nothing. We then consider the root test. For $n \geq 3$, we have that:

$$a_n = \begin{cases} \left(\frac{1}{4}\right)^k = \left(\frac{1}{4}\right)^{\frac{n-1}{2}} = 2\left(\frac{1}{4}\right)^{\frac{n}{2}} & n = 2k+1 \\ \frac{1}{2}\left(\frac{1}{4}\right)^k = \frac{1}{2}\left(\frac{1}{4}\right)^{\frac{n}{2}} & n = 2k \end{cases}$$

Since $\lim_{n\to\infty} \sqrt[n]{p} = 1$ for $p > 0$ (Theorem 3.20(b)), we have that:

$$\lim_{n\to\infty} \sqrt[n]{a_n} = \lim_{n\to\infty} \sqrt[n]{c\left(\frac{1}{4}\right)^{\frac{n}{2}}} = \lim_{n\to\infty} \sqrt[n]{c}\frac{1}{2} = \frac{1}{2} < 1$$

where the above limit holds for either $c = \frac{1}{2}$ or $c = 2$. Hence, we conclude that the series converges by the root test. This example demonstrates how the root test is "sharper" than the ratio test (though harder to apply).

3.8 Power Series

Definition 3.38: Power Series

For $z \in \mathbb{C}$ and a sequence $\{c_n\}$, $\sum_{n=0}^{\infty} c_n z^n$ is called a **power series**.

Theorem 3.39: Radius of Convergence

Let $R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}}$, with the convention $R = \infty$ if $\limsup_{n\to\infty} \sqrt[n]{|c_n|} = 0$ and $R = 0$ if $\limsup_{n\to\infty} \sqrt[n]{|c_n|} = \infty$. Then, $\sum_{n=0}^{\infty} c_n z^n$ converges if $|z| < R$ and diverges if $|z| > R$. R is called the **radius of convergence** of $\sum c_n z_n$. We note that on the circle $|z| = R$, the behavior is varied; the series can be divergent or convergent, and it can also depend on the particular choice of z on the circle.

Proof

We have that $\limsup_{n\to\infty} \sqrt[n]{c_n z^n} = \limsup_{n\to\infty} \sqrt[n]{c_n}|z| = \frac{|z|}{R}$. Therefore, by the root test (Theorem 3.33) the series converges if $|z| < R$ and diverges if $|z| > R$ (and nothing can be said if $|z| = R$). □

Note that we can use the ratio test to determine R as well, as we will see in the next few examples.

Example 3.40

(a) $\sum n! z^n$. By the ratio test, we have that $\lim_{n\to\infty} \left|\frac{(n+1)!z^{n+1}}{n!z^n}\right| = \lim_{n\to\infty}(n+1)|z| = \infty$ for all $z \neq 0$. Hence the diverges for all $z \in \mathbb{C} \neq \{0\}$, and we conclude that $R = 0$.

(b) $\sum \frac{z^n}{n^n}$. By Theorem 3.39, we have that:

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{\frac{1}{n^n}}} = \frac{1}{\limsup_{n\to\infty} \frac{1}{n}} = \frac{1}{\lim_{n\to\infty} \frac{1}{n}} = \infty$$

(c) $\sum \frac{z^n}{n!}$ also has $R = \infty$ (as can be checked easily with the ratio test). For $R = 1$, the series is equal to e. As we will define later in Chapter 8, this series is equal to e^z.

(d) $\sum \frac{z^n}{n^p}$ with $p > 1$. By Theorem 3.39, we have that:

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{\frac{1}{n^p}}} = \frac{1}{\limsup_{n\to\infty} \left(\frac{1}{\sqrt[n]{n}}\right)^p} = \frac{1}{1^p} = 1$$

where for the second last equality we apply Theorem 3.20(c). Note that this series converges for all $|z| \leq 1$ (although the above calculation does not show convergence on the boundary).

Before moving on, let us consider some further examples. Suppose $a_n = 1$ for all n. Then, our power series if just $\sum_{n=0}^{\infty} z^n$, which is just the Geometric series. By Theorem 3.26, we have that the series converges if $|z| < 1$ to $\frac{1}{1-z}$.

As another example, consider the series $\sum_{n=1}^{\infty} \frac{1}{n} z^n$. By the ratio test, we find that $R = 1$, as:

$$\limsup_{n\to\infty} \left|\frac{\frac{1}{n+1}z^{n+1}}{\frac{1}{n}z^n}\right| = \limsup_{n\to\infty} \left|\frac{n}{n+1}z\right| = |z| \limsup_{n\to\infty} \left|1 - \frac{1}{n+1}\right| = |z|$$

So this series converges if $|z| < 1$, diverges if $|z| > 1$. What happens for $|z| = 1$? At $z = 1$, we have that the series is just the standard harmonic series and diverges (by Theorem 3.28). At $z = -1$, we have an alternating series (a series whose terms are decreasing and tend to zero, and alternate in sign with

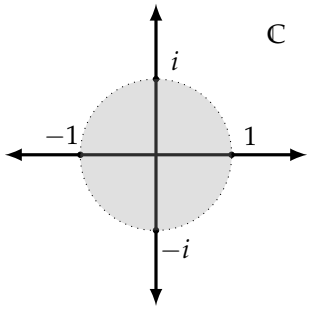

Figure 16: Visualization of the radius of convergence of the Geometric Series. The series converges in the shaded region, and diverges outside of it (and on the boundary).

each term), so we can apply the Alternating series test (below) to conclude that it converges. What about elsewhere on the circle? Let's look at $z = i$. We then have that the terms look like:

$$1 + \frac{i}{2} - \frac{1}{3} - \frac{i}{4} + \frac{1}{5} + \frac{i}{6} - \dots$$

we then have that the real and imaginary parts of the series are separately alternating series that decrease in magnitude; hence both parts are convergent, and the series as a whole is convergent at $z = i$. The same argument can be applied to conclude convergence of the series at $z = -i$. In fact, this series converges everywhere on the unit circle except at $z = 1$, which is a conclusion that follows from Theorem 3.44 (not covered in lecture, but feel free to refer to Rudin)!

Theorem 3.43: The Alternating Series Test

Let $\{a_n\} \subset \mathbb{C}$ and suppose that:

(a) $|a_1| \geq |a_2| \geq |a_3| \dots$.

(b) $a_{2m-1} \geq 0, a_{2m} \leq 1$ for $m \in \mathbb{N}$

(c) $\lim_{n\to\infty} a_n = 0$.

Then, $\sum a_n$ converges.

Proof

Rudin establishes a partial summation formula (Theorem 3.41) and proves a more general theorem (Theorem 3.42) to prove this claim. However, an alternative proof in the case where $\{a_n\}$ is real is left as homework (HW7). □

3.9 Absolute Convergence

Definition: Absolute Convergence

$\sum a_n$ is **absolutely convergent** if $\sum |a_n|$ converges. Note that if $\sum a_n$ is convergenct but $\sum |a_n|$ diverges, then $\sum a_n$ is **conditionally convergent**. Note that for real series with strictly positive terms, absolute convergence and conditional convergence are equivalent. Also, note that the root and ratio tests test for absolute convergence, and hence do not yield any information for conditional convergence.

As an example, consider that $\sum \frac{(-1)^n}{n}$ converges, but $\sum \frac{1}{n}$ diverges, so $\sum \frac{(-1)^n}{n}$ is conditionally convergent.

Theorem 3.45

If $\sum a_n$ converges absolutely, then $\sum a_n$ converges.

Proof

We have that:

$$\left| \sum_{j=m}^{n} a_j \right| \leq \sum_{j=m}^{n} \left| a_j \right| < \epsilon$$

For all $n \geq m \geq N$ for some N by the fact that $\sum \left| a_j \right|$ converges. Hence, $\sum a_j$ is convergent by the Cauchy Criterion. □

For absolutely convergent series, we can freely change the order of the additions without affecting the value of the sum (as we will soon see). However, for series that are not absolutely convergent, this turns out to not be the case!

Definition 3.52: Rearrangements

Given a bijection $K : \mathbb{N} \to \mathbb{N}$, the series:

$$\sum_n a'_n = \sum_n a_{K(n)}$$

is called a rearrangement of $\sum_n a_n$.

Theorem 3.55

If $\sum a_n$ is absolutely convergent, every rearrangement $\sum a'_n$ converges to the same limit.

Proof

Let $\{s'_n\}$ beth sequence of partial sums of the rearrangement $\sum a'_n$. Let $\epsilon > 0$. By the absolute convergent of the original series, there exists $N \in \mathbb{N}$ such that for all $n \geq m \geq N$, $\sum_{j=m}^{n} \left| a_j \right| < \epsilon$. Then, pick p such that $\{1, 2, \ldots N\} \subset \{K(1), K(2), K(3), \ldots K(p)\}$. Then, the summands $a_1, a_2, \ldots a_N$ cancel out in $s_n - s'_n$ for $n \geq p$, leaving only terms $a_{K(j)}$ past a_N. Hence, $\left| s_n - s'_n \right| < \epsilon$ for $n \geq p \geq N$, and we conclude that $\sum a'_n$ converges to the same limit as $\sum a_n$. □

Theorem 3.54: Riemann Rearrangment Theorem

If $\sum a_n$ is a conditionally convergent series of real numbers, and $-\infty \leq \alpha \leq \beta \leq \infty$, then there is a rearrangement $\sum a'_n$ such that $\liminf_{n\to\infty} s'_n = \alpha$ and $\limsup_{n\to\infty} a'_n = \beta$. Taking $\alpha = \beta$, we have that for any real number, there exists a rearrangement that converges to it.

Proof

Not covered in lecture, see Rudin. □

For example, given $\sum(-1)^n a_n$ with $a_n \geq 0$ and $\lim_{n\to\infty} a_n = 0$, we can rearrange this series to converge to any point in $\mathbb{R}$ that we like (for example, π). The idea is to select positive terms from the series until we overshoot π, then choose a sequnece of alternating negative/positive terms of decreasing magnitude until the ϵ distance from π decreases to zero.

3.10 Addition and Multiplication of Series

Theorem 3.47: Series Addition

Let $\sum a_n = A$ and $\sum b_n = B$. Then, $\sum(a_n + b_n) = A + B$ and $\sum ca_n = cA$ for any fixed $c \in \mathbb{C}$.

Proof

Let $A_n = \sum_{j=0}^{n} a_j$ and $B_n = \sum_{j=0}^{n} b_j$. Then, $A_n + B_n = \sum_{j=0}^{n} a_j + b_j$ and since $\lim_{n\to\infty} A_n = A$ and $\lim_{n\to\infty} B_n = B$, it follows that:

$$\lim_{n\to\infty} (A_n + B_n) = A + B.$$

For the second assertion, we have that $\lim_{n\to\infty} cA_n = c \lim_{n\to\infty} A_n = cA$. □

Definition 3.48: Series Multiplication

Let $\sum a_n$ and $\sum b_n$ be two series. Then, the **product** of of the two series is the series $\sum c_n$ where:

$$c_n = \sum_{j=0}^{n} a_j b_{n-j} = \sum_{j=0}^{n} a_{n-j} b_j$$

Any student who has studied Fourier Series prior to this course will notice the similarity of the above definition to the convolution of two functions.

Theorem 3.50

Suppose that $\sum a_n$ converges absolutely and $\sum b_n$ converges. Let $\sum a_n = A$ and $\sum b_n = B$. Then, $\sum c_n$ converges and $\sum c_n = AB$.

Proof

Let $A_n = \sum_{j=0}^{n} a_j$ and $B_n = \sum_{j=0}^{n} b_j$. Let $\beta_n = B_n - B$ (note that $\beta_n \to 0$). Now, we have that:

$$C_n = \sum_{k=0}^{n} c_k = \sum_{k=0}^{n} \sum_{j=0}^{k} a_j b_{k-j} = \sum_{j=0}^{n} \sum_{k=j}^{n} a_j a_{k-j} = \sum_{j=0}^{n} a_j \sum_{k=j}^{n} b_{k-j} = \sum_{j=0}^{n} a_j B_{n-j} = \sum_{j=0}^{n} a_j (B + \beta_{n-j})$$

From here, we split the sum and then we have that:

$$C_n = \sum_{j=0}^{n} a_j B + \sum_{j=0}^{n} a_j \beta_{n-j}$$

Defining $\gamma_n = \sum_{j=0}^{n} a_j \beta_{n-j}$ and taking the limit of $n \to \infty$, we have:

$$\lim_{n\to\infty} C_n = C = \lim_{n\to\infty} \left(\sum_{j=0}^{n} a_j B + \gamma_n \right) = \lim_{n\to\infty} \sum_{j=0}^{n} a_j B + \lim_{n\to\infty} \gamma_n = AB + \lim_{n\to\infty} \gamma_n$$

So the claim is proven if $\lim_{n\to\infty} \gamma_n = 0$. Let $\alpha = \sum |a_n| < \infty$ (by assumption of absolute convergence). Let $\epsilon > 0$. Then, tere exists $N \in \mathbb{N}$ such that $\left|\beta_j\right| < \frac{\epsilon}{\alpha}$ for all $j \geq N$ (as $\beta_n \to 0$). Hence,

$$|\gamma_n| \leq \left| \sum_{j=0}^{n} a_{n-j} \beta_j \right| \leq \left| \sum_{j=0}^{N} a_{n-j} \beta_j \right| + \left| \sum_{j=N+1}^{n} a_{n-j} \beta_j \right| < \left| \sum_{j=0}^{N} a_{n-j} \beta_j \right| + \sum_{j=N+1}^{n} \left| a_{n-j} \right| \frac{\epsilon}{\alpha} \leq \left| \sum_{j=0}^{N} a_{n-j} \beta_j \right| + \epsilon$$

Letting $n \to \infty$ with N fixed, we have the first term goes to 0 as $a_n \to 0$ as $n \to 0$. Hence, We have that:

$$\lim_{n\to\infty} |\gamma_n| < \epsilon$$

And as ϵ is arbitrary, $\lim_{n\to\infty} |\gamma_n| = 0$ and the claim follows. □

4 Continuity

4.1 Limits and Continuity

Definition 4.1: Limits

Let X, Y be metric spaces. Let $E \subset X$, and let $f : E \mapsto Y$. Let $p \in X$ be a limit point of E. Then, we say that $\lim_{x \to p} f(x) = q$ or $f(x) \to q$ as $x \to p$ if there exists $q \in Y$ such that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in E$ with $0 < d_X(x, p) < \delta$ we have that $d_Y(f(x), q) < \epsilon$.

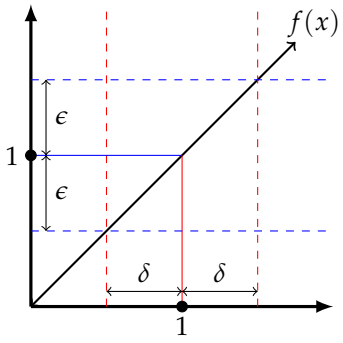

Figure 17: Visualization of the limit $\lim_{x \to 1} f(x) = 1$ for $f(x) = x$. For any $\epsilon > 0$, we can take $\delta = \epsilon$ and then we have that $|f(x) - 1| < \epsilon$ if $|x - 1| < \delta$.

Note in the above definition that we do not care about $f(p)$, that is, the actual value of f at p. In particular, if $p \notin E$, then $f(p)$ is not even necessarily defined. This distinction between the limit and the actual value of a function at a point becomes crucial later on when we want to define a derivative. Although we will discuss this in more detail in Chapter 5, the definition of a derivative of a function g at a point $p \in \mathbb{R}$ involves the function $f : \mathbb{R} \to \mathbb{R}$ such that:

$$f(x) = \frac{g(x) - g(p)}{x - p}$$

Evidently, the domain of f does not contain the point p, but we are interested in the value of f in the limit of $x \to p$ (which, if it exists, is the value of the derivative).

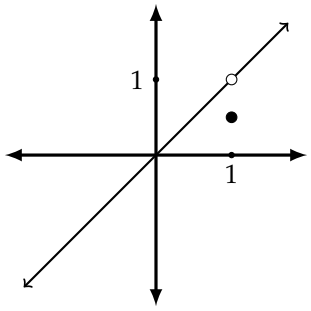

Figure 18: Visualization of the function $f(x) = x$ for $x \in \mathbb{R} \setminus \{1\}$, $f(x) = 0$ for $x = 1$. In this case, we have that $f(1) = 0$ but $\lim_{x \to 1} f(x) = 1$, demonstrating that the actual value of the function is irrelevant when defining the limit.

Theorem 4.2

Let X, Y be metric spaces. Let $E \subset X$ and $f : E \mapsto Y$. Suppose that for all sequences $\{p_n\} \subset E$ with $p_n \to p$ and $p_n \neq p$, we have that $f(p_n) \to q \in Y$. Then, this is equivalent to saying that $\lim_{x \to p} f(x) = q$.

Proof

$\boxed{\implies}$ Suppose that $\lim_{x \to p} f(x) = q$, and let $\{p_n\}$ be a sequence in E with $p_n \to p$ and $p_n \neq p$ for all n. We wish to show that $f(p_n) \to q$. Let $\epsilon > 0$. We show that there exists $N \in \mathbb{N}$ such that $d_Y(f(p_n), q) < \epsilon$ for all $n \geq N$. Since $\lim_{x \to p} f(x) = q$, there exists $\delta > 0$ such that for all $x \in E$ with $d_X(p, x) < \delta$, $d_Y(f(x), q) < \epsilon$. Since we know that $p_n \to p$, there exists some N such that $0 < d(p_n, p) < \delta$ for all $n \geq N$, so we have that $d_Y(f(p_n), q) < \epsilon$ as required.
$\boxed{\impliedby}$ We show the contrapositive. Suppose that $\lim_{x \to p} f(x) \neq q$. We wish to find a sequence $\{p_n\} \subset E$ with $p_n \to p$ and $p_n \neq p$ for all n such that $f(p_n)$ does not converge to q. Since $\lim_{x \to p} f(x) \neq q$, then there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists $x \in E$ such that $0 < d_X(x, p) < \delta$ but $d_Y(f(x), q) \geq \epsilon$. For each δ of the form $\frac{1}{n}$, let $p_n \in E$ be the corresponding value of x. Then, $p_n \to p$, $p_n \neq p$ for all n, and $f(p_n)$ does not converge to q as $d(f(p_n), q) \geq \epsilon$ for all n. □

Theorem 4.4

When $Y = \mathbb{C}$ (i.e. the functions we consider are complex), then limits respect sums, differences, products, and functions. That is, let X is a metric space, $E \subset X$, and $f, g : E \mapsto \mathbb{C}$ with p a limit point of E. If $\lim_{x \to p} f(x) = q$ and $\lim_{x \to p} g(x) = r$, then $\lim_{x \to p} (f + g)(x) = q + r$. The same holds for subtraction, multiplication, and division (provided we do not divide by zero).

Proof

By Theorem 4.2, these properties of limits follow from the analogous properties of sequences (Theorem 3.3).

Definition 4.5: Continuity

Let X, Y be metric spaces, and $E \subset X$. Let $p \in E$, and define $f : E \mapsto Y$. We say that f is **continuous** at p if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in E$ with $d_X(x, p) < \delta$, we have that $d_Y(f(x), f(p)) < \epsilon$. Equivalently, $f(N_\delta^E(p)) \subset N_\epsilon^Y(f(p))$. If f is continuous at p for all $p \in E$, we say that f is continuous.

Note that this definition of continuity is heavily reliant on the particular metric of X and Y; in particular, there can be functions that are continuous for some choices of metric but not others.

Let us consider some examples of continuous functions (while thinking about different possible metric spaces).

First, let us take onsider $X = E = \mathbb{Z}$ and $Y = \mathbb{R}$. What functions $f : E \to Y$ are continuous at $p = 0$? The answer turns out to be all functions! To see this, fix $n \in \mathbb{Z}$ and let $\epsilon > 0$. We then have that if $|n - m| < \frac{1}{2}$, then $\left|f(n) - f(m)\right| < \epsilon$ as the only point m contained in $N_{1/2}^{\mathbb{Z}}(n)$ is n itself (and hence $\left|f(n) - f(m)\right| = \left|f(n) - f(n)\right| = 0$). This argument applies to every $n \in \mathbb{Z}$ and hence all functions $f : \mathbb{Z} \mapsto \mathbb{R}$ are continuous.

As further examples (that work for arbitrary metric spaces), If we have that $f : X \mapsto X$, $f(x) = x$, we have that f is continuous (pick $\delta = \epsilon$ in the definition of continuity). If we have that $f : X \mapsto Y$, $f(x) = c$

for some $c \in Y$, then f is also continuous (pick any $\delta > 0$ in the definition).

Finally, the above definition doesn't make a distinction between limit points and isolated points. However, it turns out that according to the definition, if $p \in E$ is isolated, then every function f with E as its domain is continuous. To see this, consider that for any $\epsilon > 0$, we can pick $\delta > 0$ such that the only point $x \in E$ for which $d_X(x,p) < \delta$ is $x = p$ (such a choice of δ is possible as p is isolated). It then follows that $d_Y(f(x), f(p)) = 0 < \epsilon$.

We now consider a Theorem which gives us a familiar notion of continuity (that may have been encountered in first year calculus).

Theorem 4.6

Suppose that p is a limit point of E in Definition 4.5. Then, f is continuous at p if and only if $\lim_{x \to p} f(x) = f(p)$.

Proof

The claim immediately follows by comparing Definitions 4.1 and 4.5. □

Theorem 4.7

Let X, Y, Z be metric spaces, and $E \subset X, F \subset Y$. Let $f : E \mapsto Y$, $g : F \mapsto Z$, and suppose $f(E) \subset F$. Let $p \in E$. If f is continuous at p and g is continous at $f(p)$, then $g \circ f : E \mapsto Z$ is continuous at p.

Proof

Let $\epsilon > 0$. Since g is continuous at $f(p)$, there exists $\gamma > 0$ such that $d_Z(g(y), g(f(p))) < \epsilon$ if $d_Y(y, f(p)), < \gamma$. Since f is continuous at p, there exists $\delta > 0$ such that $d_Y(f(x), f(p)) < \gamma$ if $d_X(x,p) < \delta$. Hence, we have that $d_Z(h(x), h(p)) = d_Z(g(f(x)), g(f(p))) < \epsilon$ if $d_X(x,p) < \delta$ and $x \in E$. We conclude that h is continuous at p. □

4.2 Topological Characterization of Continuity

Theorem 4.8

Let X, Y be metric spaces, and $f : X \mapsto Y$. Then, f is continuous if and only if $f^{-1}(V) \subset X$ is open for every open set $V \subset Y$.

Proof

$\boxed{\Longrightarrow}$ Suppose f is continuous, and $V \subset Y$ is open. Let $p \in f^{-1}(V)$, so $f(p) \in V$. V is open, so it follows that $f(p)$ is an interior point of V. So, there exists $r > 0$ such that $N_r^Y(f(p)) \subset V$. Next, f is continuous, so there exists $\delta > 0$ such that for all $x \in X$ with $d_X(x,p) < \delta$, $d_Y(f(x), f(p)) < r$. Hence, we obtain that $f(x) \in N_r^Y(f(p)) \subset V$. In particular, $N_\delta^X(p) \subset f^{-1}(V)$, so p is an interior point of $f^{-1}(V)$. Hence every point of $f^{-1}(V)$ is an interior point, and $f^{-1}(V)$ is open.

$\boxed{\Longleftarrow}$ Suppose $f^{-1}(V)$ is open for every open set $V \subset Y$. Let $p \in X$ and $\epsilon > 0$. Let $V = N_\epsilon^Y(f(p))$ which is open, so by assumption $f^{-1}(V)$ is open. $p \in f^{-1}(V)$, so p is an interior point. Hence, there exists $\delta > 0$ such that $N_\delta^X(p) \subset f^{-1}(V)$. In other words, if $d_X(x,p) < \delta$, then $f(x) \in V$, so $d_Y(f(x), f(p)) < \epsilon$. f is then continuous by definition. □

Corollary

$f : X \mapsto Y$ is continuous if and only if $f^{-1}(F) \subset X$ is closed for every closed set $F \subset Y$.

Proof

Let $V = F^c$ with V open. Then, the above statement is equivalent to saying that a function $f : X \mapsto Y$ is continuous if and only if $f^{-1}(V^c) = \left(f^{-1}(V)\right)^c$ is closed for every open set $V \subset Y$. This is equivalent to the statement that $f^{-1}(V)$ is open for every open set $V \subset Y$, and the claim follows by the previous theorem. □

Note that the above topological characterization says properties of sets are preserved under taking the preimage; it does not say anything about preserving properties under the image. That is to say, images of open/closed sets are not necessarily open/closed.

Example

Consider $X = \mathbb{R}^+ = (0, \infty)$ and $Y = \mathbb{R}$. Then, the function:

$$\begin{array}{rccc} f & : & X & \longrightarrow & Y \\ & & x & \longmapsto & \frac{1}{x} \end{array}$$

is continuous. Then defining $A = [1, \infty)$ we have that A is closed, but $f(A) = (0, 1]$ is not closed.

Theorem 4.9

Let $f : X \mapsto \mathbb{C}$ and $g : X \mapsto \mathbb{C}$ be continous functions. Then, $f + g$, fg are continuous, and $\frac{f}{g}$ is continuous if $g(x) \neq 0$.

Proof

At isolated points there is nothing to prove (as any choice of function that is defined at an isolated p will be continuous there). For limit points, the claim follows from Theorems 4.4 and 4.6. □

Theorem 4.10

(a) Let $f_1, f_2, \ldots, f_k : X \mapsto \mathbb{R}$, and define $\mathbf{f} : X \mapsto \mathbb{R}^k$ by:

$$\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_k(x))$$

Then, $\mathbf{f}$ is continuous if and only if every f_i is continuous.

(b) If $\mathbf{f} = (f_1, \ldots f_k) : X \mapsto \mathbb{R}^k$ and $\mathbf{g} = (g_1, \ldots g_k) : X \mapsto \mathbb{R}^k$ are continous, then the functions:

$$\mathbf{f} + \mathbf{g} = (f_1 + g_1, \ldots, f_k + g_k)$$

$$\mathbf{f} \cdot \mathbf{g} = f_1 g_1 + \ldots + f_k g_k$$

are continous.

Proof

(a) $\boxed{\Longrightarrow}$ Suppose $\mathbf{f}$ is continuous. Then, let $\epsilon > 0$. For each $p \in X$, there exists $\delta > 0$ such that $d_X(x,p) < \delta$ implies $\left|\mathbf{f}(x) - \mathbf{f}(p)\right| < \epsilon$. We then observe that for any $i \in \{1, \dots k\}$:

$$\left|\mathbf{f}(x) - \mathbf{f}(p)\right| = \left(\sum_{j=1}^{k} \left|f_j(x) - f_j(p)\right|^2\right)^{1/2} \geq \left|f_i(x) - f_i(p)\right|$$

Where the inequality follows by just keeping one term from the sum of non-negative terms. So for any $d_X(x,p) < \delta$ we therefore have that $\left|f_i(x) - f_i(p)\right| < \epsilon$, and hence each f_i is continuous.

$\boxed{\Longleftarrow}$ Suppose each of $f_1, \dots f_k$ is continuous. Then let $\frac{\epsilon}{\sqrt{k}} > 0$. For each $p \in X$, there exists δ_i such that $d_X(x,p) < \delta_i$ implies that $\left|f_i(x) - f_i(p)\right| < \frac{\epsilon}{\sqrt{k}}$. Take $\delta = \min\{\delta_1, \dots \delta_k\}$. Then, if $d_X(x,p) < \delta$, we have that:

$$\left|\mathbf{f}(x) - \mathbf{f}(p)\right| = \left(\sum_{j=1}^{k} \left|f_j(x) - f_j(p)\right|^2\right)^{1/2} < \left(\sum_{j=1}^{k} \left(\frac{\epsilon}{\sqrt{k}}\right)^2\right)^{1/2} = \epsilon$$

So it follows that $\mathbf{f}$ is continuous.

(b) The claim follows from (a) and Theorem 4.9. □

Example 4.11

We will now explore some interesting examples of continuous functions.

(a) For each index $i = 1, \dots k$, define $\phi_i : \mathbb{R}^k \mapsto \mathbb{R}$ by $\phi_i(\mathbf{x}) = x_i$. Then, ϕ_i is continuous.

Proof. Let $\epsilon > 0$. Then, for $\mathbf{p} \in \mathbb{R}^k$, if $|\mathbf{x} - \mathbf{p}| < \delta$ with $\delta = \epsilon$, we have that:

$$\epsilon > |\mathbf{x} - \mathbf{p}| = \left(\sum_{j=1}^{k} \left|x_j - p_j\right|^2\right)^{1/2} \geq \left(\left|x_i - p_i\right|^2\right)^{1/2} = \left|x_i - p_i\right|$$

So for $|\mathbf{x} - \mathbf{p}| < \delta$, we have that $\left|\phi_i(\mathbf{x}) - \phi_i(\mathbf{p})\right| < \epsilon$. We conclude that ϕ_i is continous. □

We could also use the topological characterization of continuity to prove this claim. If $V \subset \mathbb{R}$ is open, then $\mathbb{R} \times \mathbb{R} \times \ldots \times V \times \mathbb{R} \times \ldots \times \mathbb{R}$ is open, showing again that ϕ_i is continuous (a much easier proof)!

(b) Let $f : \mathbb{R}^k \mapsto \mathbb{R}$ be given by $f(\mathbf{x}) = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ with $n_i \in \mathbb{N} \cup \{0\}$. f is continous, and hence so is any polynomial $P : \mathbb{R}^k \mapsto \mathbb{R}$.

(c) Rational functions P/Q where P, Q are polynomials are continous everywhere except where Q is zero.

(d) $f : \mathbb{R}^k \mapsto \mathbb{R}$ given by $f(\mathbf{x}) = |\mathbf{x}|$ is continous.

(e) Suppose $f : X \mapsto \mathbb{R}^k$ is continous. Then so is $g : X \mapsto \mathbb{R}^k$ with $g(x) = \left|f(x)\right|$.

4.3 Continuity and Compactness

Theorem 4.14

Let $f : X \mapsto Y$ be continuous, and suppose X is compact. Then, the image $f(X)$ is compact.

Proof

Let $\{V_\alpha\}$ be an open cover of $f(X)$. For each α, Let $O_\alpha = f^{-1}(V_\alpha)$. Then, $\{O_\alpha\}$ is an open cover of X. to see this, we recognize that each O_α is open due to the continuity of f (Theorem 4.8) and each $p \in X$ satisfies V_β for some β, so $p \in O_\beta$. X is compact, so there exists a finite subcover $\{O_{\alpha_1}, \dots O_{\alpha_n}\}$. But it then follows that $\{V_{\alpha_1}, \dots V_{\alpha_n}\}$ is a finite subcover of $f(X)$. Hence $f(X)$ is compact. □

We previously discussed how an image of a closed set (under a continuous function) does not have to be closed. However, the above Theorem shows that the image of a compact set (under a continuous function) must be compact.

Taking a slight tangent, a question of interest may be to consider the topology of the extended real numbers; namely, $\mathbb{R} \cup \{\infty, -\infty\}$ (where $-\infty < x < \infty$ for all $x \in \mathbb{R}$). In particular, a natural question to ask is whether $[1, \infty) \cup \{\infty\}$ would be closed/compact.

We start then by defining a reasonable metric on $X = \mathbb{R} \cup \{\infty, -\infty\}$. A "reasonable" definition of a metric will satify the requirement that $[1, \infty) \cup \{\infty\}$ is compact or not. We then define the metric:

$$d(x, y) = \left|\arctan(x) - \arctan(y)\right|$$

where $\arctan(\infty) = \frac{\pi}{2}$ and $\arctan(-\infty) = -\frac{\pi}{2}$. We then have that (X, d) is a metric space (check!)

Note that the extended reals have been made into a metric space here, but it is no longer a field; in particular, we have that $(-\infty + \infty) + \infty = 0 + \infty = \infty$ but $-\infty + (\infty + \infty) = -\infty + \infty = 0$ so associativity no longer holds.

With a reasonable metric defined on the extended reals, let us now return to our previous example in Section 4.2 where we considered the function $f(x) = \frac{1}{x}$. We now extend the domain of the function such that:

$$\begin{array}{rccc} f & : & X \setminus \{0\} & \longrightarrow & \mathbb{R} \\ & & x & \longmapsto & \frac{1}{x} \end{array}$$

With $f(\infty) = f(-\infty) = 0$. We leave it as an exercise to verify that f is continuous on $X \setminus \{0\}$ using the continuity of the arctan function.

Now, consider $A = [1, \infty) \subset X$. A is not closed, as ∞ is a limit point of A (this can be verified by checking that for all $\epsilon > 0$, there exists $x \in [1, \infty)$ such that $|\arctan x - \arctan \infty| = \frac{\pi}{2} - \arctan x < \epsilon$). However, we do have that $\overline{A} = [1, \infty]$ is closed and compact, and that $f(\overline{A}) = [0, 1]$ is compact as Theorem 4.14 states. In a sense, we have "compactified" the reals through our construction, as X is compact while $\mathbb{R}$ was not. We also have that X is bounded (unlike $\mathbb{R}$), as all distances between points are bounded by at most π. In particular, $\operatorname{diam} X = \pi$.

Definition 4.13: Bounded functions

Let X be a metric space, and $E \subset X$. We say that $f : E \mapsto Y$ is **bounded** if $f(E)$ is a bounded set. In particular, if $Y = \mathbb{R}^k$, then f is bounded if and only if there exists $M \in \mathbb{R}$ such that $\left|f(x)\right| \leq M$ for all $x \in E$.

Theorem 4.15

Let X be a compact metric space, and $f : X \mapsto Y$ be continuous. Then, $f(X)$ is bounded.

Proof

By Theorem 4.14, we have that $f(X)$ is compact. $f(X)$ is therefore bounded (see discussion proceeding Theorem 2.41). □

An interesting special case in the above theorem to think about is when $Y = \mathbb{R}$; this yields the Extreme Value Theorem, below.

Theorem 4.16: Extreme Value Theorem

Let X be a compact metric space, and let $f : X \mapsto \mathbb{R}$ be continuous. Then, there exist $p, q \in X$ such that $f(p) \leq f(x) \leq f(q)$ for all $x \in X$. In other words, f attains its minimum (infimum) and maximum (supremum) on X.

Proof

By Theorems 4.15 and 4.16, $f(X)$ is closed and bounded. In particular, $m = \inf f(X)$ and $M = \sup f(X)$ exist (as the set is bounded), and $m, M \in f(X)$ as $f(X)$ contains all of its limit points. Hence, there exist $p, q \in X$ such that $f(p) = m$, $f(q) = M$. □

We consider some examples which demonstrate the necessity of the compactness of the domain. Let $E = (0,1) \subset \mathbb{R}$ and consider $f : (0,1) \mapsto \mathbb{R}$. Let $f(x) = \frac{1}{x^2}$. $(0,1)$ is bounded but not closed, and we observe that $f((0,1)) = [1, \infty)$ which is unbounded (and hence f does not attain a maximum). Instead let $f(x) = x$. Then, we have that $f((0,1)) = (0,1)$ which is bounded, but f does not attain a maximum/minimum on its domain.

Theorem 4.17

Let X be a compact metric space, and $f : X \mapsto Y$ be continuous and a bijection. Define $f^{-1} : Y \mapsto X$ by $f^{-1}(y) = x$ if $f(x) = y$ (this is well-defined due to the bijectivity of f). Then, f^{-1} is continuous.

Proof

By Theorem 4.8, it suffices to show that for every open set $V \subset X$, $\left(f^{-1}\right)^{-1}(V) = f(V)$ is open. By the openness of V, V^c is closed, and hence compact (as closed subsets of compact sets are compact by Theorem 2.35). So, $f(V^c) \subset Y$ is compact by Theorem 4.14. Therefore, $f(V^c) = \left(f(V)\right)^c$ is closed (as compact sets are closed by Theorem 2.34). Therefore, $\left(\left(f(V)\right)^c\right)^c = f(V)$ is open as desired. Hence f^{-1} is continuous. □

Example 4.21

This example shows the necessity of compactness of X in Theorem 4.17. Let $S = [0, 2\pi)$ which is not closed and hence not compact (Heine Borel/Theorem 2.41). Let $Y = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\}$. Define $f : X \mapsto Y$ where $f(\theta) = (\cos\theta, \sin\theta)$. f is continuous and a bijection (check!) but f^{-1} is not continuous. To see this is the case, consider first that $[0, \frac{\pi}{2}) \subset X$ is open in X. At everywhere except 0 it should be evident that all points of $[0, \frac{\pi}{2})$ are interior to X, and at 0, we have that $N_{1/2}(0) \subset [0, \frac{\pi}{2})$ showing that 0 is also an interior point. But, $(f^{-1})^{-1}([0, \frac{\pi}{2})) = f([0, \frac{\pi}{2}))$ is not open, because not every point is an interior point; namely, $(1, 0) = f(0) \in f([0, \frac{\pi}{2}))$ is not an interior point. So, f^{-1} is not continuous.

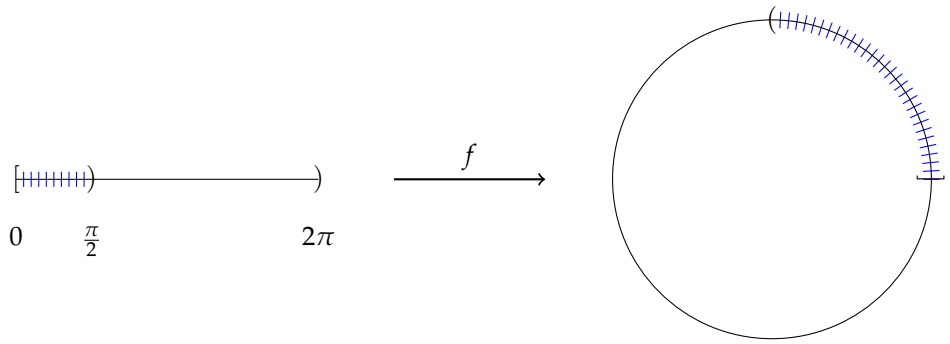

Figure 19: Visual of the map f which maps a point $\theta \in [0, 2\pi)$ to the point $(\cos\theta, \sin\theta)$ on the unit circle. Pictured is the map from $[0, \frac{\pi}{2})$ to $f([0, \frac{\pi}{2}))$, the former which is open in $[0, 2\pi)$, and the latter which is not open in $Y = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\}$.

4.4 Uniform Continuiity, Connectedness, and IVT

Definition 4.18: Uniform Continuity

Let X, Y be metric spaces and $f : X \mapsto Y$. Then, f is **uniformly continuous** if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $p, q \in X$ with $d_X(p, q) < \delta$, we have that $d_Y(f(p), f(q)) < \epsilon$. It is clear from the definition that every uniformly continuous function is continuous.

Note that this definition is very similar to the definition of continuity, but with a difference in the quantifiers. With continuity, for each ϵ there is a δ that we can find for a given point $p \in X$. For uniform continuity, we have that for each ϵ, there exists a δ that works uniformly for all p.

We now consider some examples to see the difference concretely. Take $X = (0, 1), Y = \mathbb{R}$, and $f(x) = \frac{1}{x}$. We then have that f is continuous, but not uniformly continuous, showing that in general continuity does not imply uniform continuity.

Next, consider $X = Y = \mathbb{R}$ and $f(x) = \sin x$. Then, f is uniformly continuous. Given any ϵ, we can pick $\delta = \epsilon$ and this proof will work. The "calclus-inspired" proof would use that $\frac{d}{dx} \sin x = \cos x$ and $|\cos x| \leq 1$, so we can always pick $\delta = \epsilon$ with the MVT. However, we have yet to define the derivative or prove the mean value theorem, so an alternate appraoch would be to invoke trigonometric identities.

Finally, let $X = [0, 10], Y = \mathbb{R}$, and $f(x) = x^2$. Then, f is uniformly continuous. However, if $X = Y = \mathbb{R}$ and $f(x) = x^2$, then f is not uniformly continuous. The uniform continuity of f depends on the domain; in particular, $[0, 10]$ is closed/compact, while $\mathbb{R}$ is not compact. This motivates our next theorem.

Theorem 4.19

Let X, Y be metric spaces with X compact. Let $f : X \mapsto Y$ be continuous. Then, f is uniformly continuous.

Proof

Let $\epsilon > 0$. By the continuity of f, for each $p \in X$, there exists $\delta_p 0$ such that for all $q \in X$ with $d_X(p,q) < \delta_p$, we have that $d_Y(f(p), f(q)) < \frac{\epsilon}{2}$. Define $U_p = N_{\delta_p/2}(p)$. We then have that $\left\{U_p\right\}_{p \in X}$ is an open cover for X. By the compactness of X, it has a finite subcover. Let $\left\{U_{p_1}, \ldots, U_{p_n}\right\}$ be a finite subcover. Let $\delta = \frac{1}{2} \min \left\{\delta_{p_1}, \ldots, \delta_{p_n}\right\} > 0$ (note the importance of taking the minimum over a *finite* number of δ_{p_i}s; if we took an infimum over an infinite number, it could be possible that the infimum could be zero). Then, take $p, q \in X$ with $d_X(p,q) < \delta$. Then, $p \in U_{p_i}$ for some i. So, $d_X(p, p_i) < \frac{\delta_{p_i}}{2}$, and then by the triangle inequality we have that:

$$d_X(q, p_i) \leq d_X(q,p) + d_X(p, p_i) < \delta + \frac{\delta_{p_i}}{2} \leq \frac{\delta_{p_i}}{2} + \frac{\delta_{p_i}}{2} = \delta_{p_i}$$

Therefore, we have that:

$$d_Y(f(p), f(q)) \leq d_Y(f(p), f(p_i)) + d_Y(f(p_i), f(q)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

And we conclude that f is uniformly continuous. □

Theorem 4.22

Let X, Y be metric spaces, and let $f : X \mapsto Y$ be continuous. If $E \subset X$ is connected (see Definition 2.45) then its image $f(E)$ is also connected.

Proof

Suppose for the sake of contradiction that we can write $f(E) = A \cup B$ with $A, B \neq \emptyset$ and $A \cap \overline{B} = \overline{A} \cap B = \emptyset$ (i.e. $f(E)$ is not connected). Then, define $G = f^{-1}(A) \cap E$ and $H = f^{-1}(B) \cap E$. Then, $E = G \cup H$, with $G \neq \emptyset$, $H \neq \emptyset$, and $G \cap H = \emptyset$. We wish to show that $\overline{G} \cap H = G \cap \overline{H} = \emptyset$ as this will contradict the connectedness of E. Since $A \subset \overline{A}$, We have that $G \subset f^{-1}(A) \subset f^{-1}(\overline{A})$. f is continuous, so by Theorem 4.8, we have that $f^{-1}(\overline{A})$ is closed. Hence by Theorem 2.27 we have that $\overline{G} \subset f^{-1}(\overline{A})$. Since $f(H) = B$ and $\overline{A} \cap B = \emptyset$, we have that:

$$f(\overline{G}) \cap f(H) = \emptyset$$

And therefore $\overline{G} \cap H = \emptyset$. By an identitcal argument, $G \cap \overline{H} = \emptyset$. Hence, $E = G \cup H$ is not connected, which is a contradiction. □

Theorem 4.23: Intermediate Value Theorem

Let $f : [a,b] \mapsto \mathbb{R}$ be continuous and $f(a) < f(b)$. Then, for all $\alpha \in (f(a), f(b))$, there exists $c \in (a,b)$ with $f(c) = \alpha$.

Proof

$[a,b]$ is connected, so by Theorem 4.22, $f([a,b])$ is conected. Hence by 2.47, it contains every point between $f(a)$ and $f(b)$. □

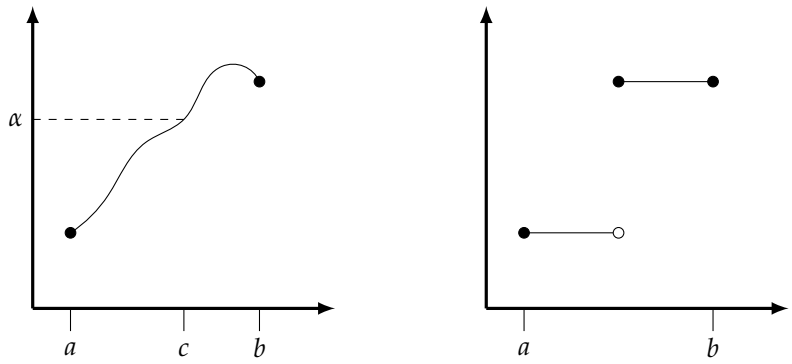

Figure 20: Visualization of a function for which the IVT applies (left) and a function where it does not (right).

As a closing remark for this chapter, consider that continuity over a closed interval implies the intermediate value theorem. Does this then imply that a connected space with the intermediate value theorem implies continuity? The answer turns out to be no. Consider discontinuities other than jumps (IVT only tells us a function does not "jump" on the domain of interest), that is, a discontinuity that arises from the non-existence of a limit. As an example, consider:

$$\begin{array}{rccl} f & : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & & x & \longmapsto & \begin{cases} \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} \end{array}$$

This function is not continuous at $x = 0$ (the limit does not even exist there!) but satisfies the IVT.

4.5 Topological Spaces

Up until this point, we have been discussing metric spaces (X,d), where $U \subset X$ is open if every point of U is an interior point according to the metric d. In this picture, $\emptyset$ is open, X is open, and if $\{U_\alpha\}$ are open, then $\bigcup_\alpha U_\alpha$ (arbitrary unions) and $\bigcap_{i=1}^n U_{\alpha_i}$ (finite intersections) are open. Although this metric space picture is the one we have been using (and will be using going forwards in the course), it could be interesting to take a temporary tangent, and consider a generalization of these notions to toplogical spaces:

Definition: Topological Spaces

A pair (X,τ) is a **topological space** if $\tau \subset \mathcal{P}(X)$ such that $\emptyset \in \tau$, $X \in \tau$, and if $\{U_\alpha\} \subset \tau$, then $\bigcup_\alpha U_\alpha \in \tau$ and $\bigcap_{i=1}^n U_{\alpha_i} \in \tau$. In this definition, τ is the set of "open sets" of X.

In metric spaces, we defined openness in terms of a metric, but in topological spaces, we throw away the metric.

If (X,τ) and (Y,ρ) are topological spaces, how do we define a continuous function $f : X \mapsto Y$? Evidently, the $\epsilon - \delta$ definition no longer makes sense (in the absense of a metric), but our topological characterization of continutiy still holds:

Definition: Continuity on Topological Spaces

Let (X, τ) and (Y, ρ) be topological spaces. Then, $f : X \mapsto Y$ is **continuous** if for every $U \in \rho$, $f^{-1}(U) \in \tau$.

A natural question that arises given a topological space is whether it is possible or not to find a metric that gives us this topology.

Definition: Metriziability

Let (X, τ) be a topological space. If there exists a metric $d : X \times X \mapsto \mathbb{R}$ such that every set $U \in \tau$ is open with respect to the metric d, then we say that X is **metrizable**.

However, not all topological spaces are metrizable. For example, take (X, τ) with $\tau = \{\emptyset, X\}$. If X has more than 1 element, then there is no metric that gives rise to this trivial topology.

Proof. We show the proof for the case of 2 elements (the proof for any finite $|X|$ follows analogously, and the argument for the infinite case is left as an exercise). Let $X = \{a, b\}$ and $\tau = \{\emptyset, X\}$. if d is a metric on X, then $d(a, b) = r > 0$. Then, let $0 < s < r$. Then, $N_s(a)$ has to be open, as does $N_s(b) = \{b\}$. The set of open sets of X under d is therefore $\{X, \emptyset, \{a\}, \{b\}\} \neq \tau$ and hence (X, τ) is not metrizable. □

The original definition we had of convergent sequences (and Cauchy sequences) also does not generalize well to topological spaces (as both invoke the notion of a metric). We can define convergence in topological spaces (in a way that does not explicitly use a metric) as follows:

Definition: Convergence in Topological Spaces

Let (X, τ) be a topological space, and suppose $\{x_n\} \subset X$. We say that x_n **converges** to x if for every set $U \in \tau$ such that $x \in \tau$, there exists N such that $x_n \in U$ for all $n \geq N$.

Note that it is impossible to define the notion of a Cauchy sequence without a metric (it cannot be defined solely in terms of open sets). Convergence is topological, but the metric matters for Cauchy.

Recall that the compactness of a set K was defined in terms of every cover having a finite subcover. The sequential compactness of a set K can be defined by saying that every sequence in K has a convergent subsequence. These definitions have no explicit reference to a metric given our above two definitions of open sets/convergence, so the definition of compactness holds in the same way in the picture of topological spaces. Note that in metric spaces, compactness and sequentially compactness are equivalent, but they can be different in topological spaces (this is another fact that tells us that there exist topological spaces that are not metrizable)!

So reviewing what we can define meaningfully in the picture of topological spaces, we see that we can define continuity, convergence, compactness, and sequential compactness; unfortunately, Cauchy sequences are left on the cutting room floor.

5 Differentiation

5.1 Derivatives

Definition 5.1: Derivatives

Let $f : [a,b] \mapsto \mathbb{R}$, and $x \in [a,b]$. We then define the **derivative** of f at x as:

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$

If the limit exists. Alternative notations for the derivative are given by:

$$\frac{\partial f}{\partial x}(x) \text{ or } \frac{\mathrm{d}}{\mathrm{d}x} f(x) \text{ or } \left.\frac{\mathrm{d}}{\mathrm{d}y} f(x)\right|_{y=x}$$

As an interpretation of the derivative, take $[a,b]$ to be a metric space, with x a limit point of $[a,b] \setminus \{x\}$. Then, $g(t) = \frac{f(t)-f(x)}{t-x}$ is a function from $[a,b] \setminus \{x\} \mapsto \mathbb{R}$. If $x \in (a,b)$, then the above definition of the derivative agrees with the definition of $f'(x)$ from first year calculus. If $x = a$ or $x = b$, then the above definition agrees with the definition of the one-sided derivative from first year calculus. Note that we will not discuss in this class cases where the domain gets more complicated (i.e. not just closed intervals of $\mathbb{R}$).

Theorem 5.2

Let $f : [a,b] \mapsto \mathbb{R}$, let $x \in [a,b]$, and suppose $f'(x)$ exists. Then, f is continuous at x.

Proof

For $t \neq x$, we can write:

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

Taking the limit of $t \to x$, we then have that:

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

Where in the last line we invoke Theorem 4.4. Evaluating the limits on the RHS by using the existence of the derivative of f at x, we have

$$\lim_{t \to x} f(t) = f(x) + f'(x) \cdot (0) = f(x)$$

So we conclude that f is continuous at x by Theorem 4.6. □

The interpretation is that differentiability at $x \in (a,b)$ implies continuity of f at x, and the left/right differentiability of f at a/b implies the left/right continuity of f at a/b. We have wrapped the proof of all these cases into one!

Note that the converse of the above theorem is not true. As a simple example, take $f(x) = |x|$ on $[-1,1]$, which is continuous at $x = 0$ (it can be verified that $\lim_{x \to 0} f(x) = f(0) = 0$) but is not differentiable there (the left/right handed limits of the difference quotient do not agree and hence the

derivative does not exist). In Chapter 7, we will construct a function that is continuous everywhere and differentiable nowhere!

NWe will now proceed to prove a series of theorems that have been seen in first year, but using our new/rigorous definitions.

Theorem 5.3: Sum, Product, and Quotient Rules

Let $f, g : [a,b] \mapsto \mathbb{R}$. Let $x \in [a,b]$ and suppose f and g are differentiable at x. Then, $f+g$, $f-g$, $f \cdot g$ are differentiable at x, and so is $\frac{f}{g}$ provided $g(x) \neq 0$. Furthermore:

(a) $(f+g)'(x) = f'(x) + g'(x)$

(b) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$

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

Proof

(a) Follows immediately from the additive property of limits (Theorem 4.4).

(b) Let $h = fg$. We then have that:

$$h(t) - h(x) = f(t)\left[g(t) - g(x)\right] + g(x)\left[f(t) - f(x)\right]$$

For $t \neq x$, we can divide both sides by $t - x$ to obtain:

$$\frac{h(t) - h(x)}{t - x} = f(t)\frac{g(t) - g(x)}{t - x} + g(x)\frac{f(t) - f(x)}{t - x}$$

Taking the limit of $t \to x$ on both sides, we obtain:

$$h'(x) = f(x)g'(x) + f'(x)g(x)$$

as desired.

(c) Let $h(t) = \frac{f(t)}{g(t)}$. Then:

$$\begin{aligned} h(t) - h(x) &= \frac{f(t)}{g(t)} - \frac{f(x)}{g(x)} \\ &= \frac{1}{g(t)g(x)}\left(f(t)g(x) - g(t)f(x)\right) \\ &= \frac{1}{g(t)g(x)}\left[g(x)\left(f(t) - f(x)\right) - f(x)\left(g(t) - g(x)\right)\right] \end{aligned}$$

For $t \neq x$, we can divide both sides by $t - x$ to get:

$$\frac{h(t) - h(x)}{t - x} = \frac{1}{g(t)g(x)}\left[g(t)\frac{f(t) - f(x)}{t - x} - f(x)\frac{g(t) - g(x)}{t - x}\right]$$

Taking the limit as $t \to x$ on both sides, we obtain the desired expression. □

As an exercise, one can prove by induction (applying 5.3(b)) that $(f_1 f_2 f_3 \dots f_n)'(x)$ (where $f_i : [a,b] \mapsto \mathbb{R}$ and each $f_i'(x)$ exists) is given by:

$$f_1'(x) f_2(x) \dots f_n(x) + \dots + f_1(x) f_2(x) \dots f_n'(x).$$

Note that as a corollary of this, we get that if $f(x) = x^n$, then $f'(x) = nx^{n-1}$ and we hence recover the familiar power rule from first year calculus!

Theorem 5.5: Chain Rule

Let $f : [a,b] \mapsto \mathbb{R}$, $x \in [a,b]$, and suppose f is differentiable at x. Suppose furthermore that $f([a,b])$ is contained in some interval I. Let $g : I \mapsto \mathbb{R}$ and suppose g is differentiable at $f(x)$. Then, $g \circ f : [a,b] \mapsto \mathbb{R}$ is differentiable at x, and furthermore:

$$(g \circ f)'(x) = g'(f(x)) f'(x)$$

Proof

Define $h(t) = g \circ f(t)$ for $a \leq t \leq b$, $t \neq x$. We cna then write:

$$f(t) - f(x) = (t - x) \left[f'(x) + u(t) \right]$$

For a function $u(t)$ with $\lim_{t \to x} u(t) = 0$. Now defining $y = f(x)$, we write:

$$g(s) - g(y) = (s - y) \left[g'(y) + r(s) \right]$$

For a function $r(s)$ with $\lim_{s \to y} r(s) = 0$. Hence, we have that:

$$\begin{aligned} h(t) - h(x) &= g(f(t)) - g(f(x)) \\ &= \left(f(t) - f(x) \right) \left(g'(y) + r(s) \right) \\ &= (t - x) \left[f'(x) + u(t) \right] \left(g'(y) + r(s) \right) \end{aligned}$$

Dividing both sides by $t - x$, we obtain:

$$\frac{h(t) - h(x)}{t - x} = \left[f'(x) + u(t) \right] \left(g'(y) + r(s) \right)$$

We now take the limit of $t \to x$ on both sides. $\lim_{t \to x} u(t) = 0$, and f is differentiable and hence continuous at x, so $s = f(t) \to y$ as $t \to x$. Thus, $r(s) \to 0$ as $t \to x$, and in conclusion:

$$h'(x) = (g \circ f)'(x) = f'(x) g'(y) = g'(f(x)) f'(x)$$

as desired. $\square$

5.2 MVT

Definition 5.7: Local Maxima/Minima

Let X be a metric space. Let $f : X \mapsto \mathbb{R}$, and let $x \in X$. We say that x is a **local maximum** of f if there exists $\delta > 0$ such that $f(y) \leq f(x)$ for all $y \in N_\delta(x)$. A **local minimum** is defined similarly, with $f(y) \geq f(x)$ instead.

For a metric space X equipped with the discrete metric, all points $x \in X$ are simultaneously local maxima and minima. To see this, take any $0 < \delta \leq 1$.

Theorem 5.8

Let $f : [a, b] \mapsto \mathbb{R}$. Let $x \in [a, b]$ and suppose that $f'(x)$ exists, and f is either a local maximum or local minimum of f. Then, $f'(x) = 0$.

Proof

Suppose x is a local minimum. Then, there exists $\delta > 0$ such that $N_\delta(x) \subset [a, b]$, and $f(y) \geq f(x)$ for all $y \in N_\delta(x)$. Thus, if $x < y < x + \delta$, then:

$$\frac{f(y) - f(x)}{y - x} \geq 0 \implies f'(x) \geq 0$$

Conversely, if $x - \delta < y < x$, then:

$$\frac{f(y) - f(x)}{y - x} \leq 0 \implies f'(x) \leq 0$$

So taken together we obtain that $f'(x) = 0$. An identical argument is used for the case of a local maximum. □

Theorem: Rolle's Theorem

Let $f : [a, b] \mapsto \mathbb{R}$ be continuous, and suppose f is differentiable on (a, b). If $f(a) = f(b)$, then there exists $x \in (a, b)$ such that $f'(x) = 0$.

Proof

Since $[a, b]$ is compact and f is continuous, by the EVT (Theorem 4.16) f attains its maximum on $[a, b]$, that is, there exists $c \in [a, b]$ such that $f(y) \leq f(x)$ for all $y \in [a, b]$. If $c \in (a, b)$, then by Theorem 5.8, $f'(c) = 0$ and we are done. Next, suppose $c = a$ or $c = b$. Again by the EVT, f attains its minumum on $[a, b]$, that is, there exists $d \in [a, b]$ such that $f(y) \geq f(d)$ for all $y \in [a, b]$. If $d \in (a, b)$, then by Theorem 5.8, $f'(d) = 0$ and we are done. Suppose then that $d = a$ or $d = b$. Since $f(a) = f(b)$, we therefore obtain that $f(a) = f(b) = f(c) = f(d)$ and the maximum/minimum values agree. Hence, $f(y) = f(a)$ for all $y \in [a, b]$, so $f'(y) = 0$ for all $y \in [a, b]$. So, the desired x may be any point in $[a, b]$. □

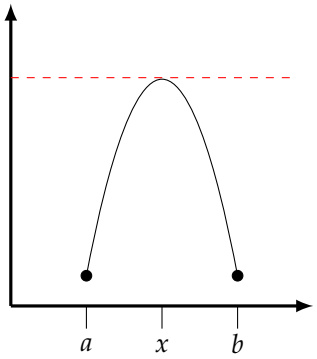

Figure 21: A simple parabolic function that demonstrates Rolle's Theorem.

Theorem 5.10: Mean Value Theorem

Let $f : [a, b] \mapsto \mathbb{R}$ be continuosu and differentiable on (a, b). Then, there exists $x \in (a, b)$ such that $f(b) - f(a) = f'(x)(b - a)$.

The visual interpretation of this theorem is that there exists $x \in (a, b)$ such that the slope of the tangent line to f at x is equal to the secant line slope between $(a, f(a))$ and $(b, f(b))$. The idea of the proof is to rotate one's head such that the sectant line is horizontal; one is then able to apply Rolle's Theorem!

Proof

Define $h(y) = f(y) = \frac{f(b)-f(a)}{b-a}(y - a)$. h is continuous on $[a, b]$ and differentiable on (a, b) (being a sum of continuous/differentiable functions). We have that $h(a) = f(a) - 0 = f(a)$, and $h(b) = f(b) - \frac{f(b)-f(a)}{b-a}(b - a) = f(a)$. Applying Rolle's Theorem to h, there exists $x \in (a, b)$ such that $h'(x) = 0$. Therefore, $h'(x) = 0 = f'(x) - \frac{f(a)-f(b)}{b-a} = 0$, and we conclude that $f(b) - f(a) = f'(x)(b - a)$ for some $x \in (a, b)$. □

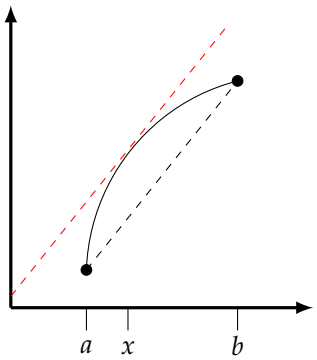

Figure 22: A simple continuous function that demonstrates the MVT.

Theorem 5.11

Let $f : [a, b] \mapsto \mathbb{R}$ be differentiable on (a, b). Then:

(a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.

(b) If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.

(c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.

Proof

If $a < x < y < b$, by the mean value theorem, there exists $z \in (x, y)$ such that:

$$f(y) - f(x) = f'(z)(y - x)$$

Note that $y - x > 0$ by construction.

(a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then $f'(z) \geq 0$, showing that $f(y) - f(x) \geq 0$ and hence that f is monotonically increasing.

(b) If $f'(x) = 0$ for all $x \in (a, b)$, then $f'(z) = 0$, showing that $f(y) - f(x) = 0$ and hence that f is constant on (a, b).

(c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then $f'(z) \leq 0$, showing that $f(y) - f(x) \leq 0$ and hence that f is monotonically decreasing. □

5.3 Taylor's Theorem

Definition 5.14: Higher Order Derivatives

If f is differentiable in a neighbourhood of x, then we may compute a second order derivative:

$$\lim_{t \to x} \frac{f'(t) - f'(x)}{t - x} = (f')'(x) = f''(x)$$

We can then continue this process to obtain $f^{(3)}(x), f^{(4)}(x), \ldots, f^{(n)}(x)$.

Definition: $C^n(I, \mathbb{R})$

If f is continuous in I, we can write $f \in C^0(I, \mathbb{R})$. If f is differentiable in a neighbourhood I and the derivative f' is continuous in I, then we write $f \in C^1(I, \mathbb{R})$. In general, $f \in C^n(I, \mathbb{R})$ denotes the nth derivative of f is continuous in I. Note that where it is clear from context, we may drop the $\mathbb{R}$ and just write $C^n(I)$.

Recall that for a function f continuous and differentiable on (x_0, x), the Mean Value Theorem proved the existence of some $\tilde{x}$ such that:

$$f(x) = f(x_0) + f'(\tilde{x})(x - x_0)$$

This gives us a natural method to build up approximations for functions; we can start with a constant approximation $f(x_0)$, then add a linear term $f'(\tilde{x})(x - x_0)$, then add on a quadratic term $(x - x_0)^2$ and so on. The following theorem gives us a way to construct these approximations and bound their error.

Theorem 5.15: Taylor's Theorem

Let I be a neighbourhood of x_0, and $f \in C^p(I)$. Then, for any $n < p$, we have that:

$$f(x) = \sum_{j=0}^{n} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j + \frac{f^{(n+1)}(\tilde{x})}{(n+1)!}(x - x_0)^{n+1}$$

Where $\tilde{x} = x_0 + \lambda(x - x_0)$ for some $\lambda \in (0, 1)$ (i.e. $\tilde{x} \in (x_0, x)$). Note that $\tilde{x}$ depends on x, x_0, and n.

Proof

For $n = 0$, the claim reduces to the Mean Value Theorem. Then, let $n \geq 1$. Let A be a constant that depends on x, x_0, and n, and let $P_n(x) = \sum_{j=0}^{n} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j$. Then, we can write:

$$f(x) = P_n(x) + A(x - x_0)^{n+1}$$

We need to show that we can express A as relating to the derivative, namely, that there exists $\tilde{x}$ such that $A = \frac{f^{(n+1)}(\tilde{x})}{(n+1)!}$. Let $g(t) = f(t) - P_n(t) - A(t - x_0)^{n+1}$ with $t \in I$. Then, $g \in C^p(I)$. For $n < p$, we then have that:

$$g^{(n+1)}(t) = f^{(n+1)}(t) - 0 - A(n+1)!(t - x_0)$$

We claim that there exists $\tilde{x} \in (x_0, x)$ such that $g^{(n+1)}(\tilde{x}) = 0$. To see this, consider that $P^{(j)}(x_0) = f^{(j)}(x_0)$ for $j = 0, 1, \ldots, n$, so $g(x_0) = 0$, and furthermore:

$$g'(x_0) = g''(x_0) = g^{(3)}(x_0) = \ldots = g^{(n)}(x_0) = 0$$

Moreover y the choice of n, we have that $g(x) = 0$. Hence, by Rolle's Theorem, there exists a point x_1 between x_0 and x such that $g'(x_1) = 0$. Similarly, repeating the argument above, there exists an x_2 between x_0 and x_1 such that $g''(x_2) = 0$. Repeating this process up to $g^{(n)}$, we have that $g^{(n)}(x_n) = 0$, for some $x_0 < x_n < x_{n-1} < \cdots < x$ and in turn, there exists $x_{n+1} \in (x_0, x_n)$ such that $g^{(n+1)}(x_{n+1}) = 0$. Setting $\tilde{x} = x_{n+1}$, the claim is shown. □

As an example, we consider the Taylor series of the function $f(x) = \cos(x)$ (We will formally define this function later on, but for now, let us assume its familar properties and derivatives). We then have that:

$$f^{(j)}(0) = \begin{cases} (-1)^m & \text{if } j = 2m \\ 0 & \text{if } j = 2m + 1 \end{cases}$$

If we have the sum run from $j = 0$ to some $j = n$, let us then try to estimate the rest. Let $\tilde{x} \in (0, x)$. Then, letting the error term be represented by ϵ, we have that:

$$\epsilon = \left| \frac{f^{(n+1)}(\tilde{x})}{(n+1)!}(x - x_0)^{n+1} \right| \leq \frac{|x|^{n+1}}{(n+1)!}$$

And we observe that $\lim_{n\to\infty} \frac{|x|^{n+1}}{(n+1)!} = 0$ and hence the error ϵ goes to zero in the $n \to \infty$ limit. Therefore, the difference between $\cos(x)$ and its Taylor polynomial vanishes quickly for any x, and the Taylor series

converges for all x. Taking the limit of the sum, we have that:

$$f(x) = \cos(x) = \lim_{n\to\infty}\left(P_n(x) + \frac{f^{(n+1)}(x_0)}{(n+1)!}(x-x_0)^{n+1}\right) = \sum_{j=0}^{\infty}\frac{(-1)^{2m}}{m!}x^{2m} = 1 - \frac{x^2}{2} + \frac{x^4}{4} - \frac{x^6}{6} + \dots$$

A question of interest might be how many terms do we need in the Polynomial such that our error is less than 10^{-6}, say, for estimating the value of $\cos\left(\frac{\pi}{12}\right)$. In other words, we want to find the m such that:

$$\frac{1}{2m!}\left(\frac{\pi}{12}\right)^{2m} \leq 10^{-6}$$

Rearranging, we require:

$$(2m)!\left(\frac{12}{\pi}\right)^{2m} > 10^{6}$$

Making a table of the value of the LHS as a function of m, we have: So we see that three terms are sufficient

m	$(2m)!\left(\frac{12}{\pi}\right)^{2m}$
1	≈ 29
2	≈ 5110
3	$\approx 2.23 \times 10^6$

for a good approximation in this case (and as stated before, the series converges very quickly)!

A natural question of interest is the convergence of the sum in the $N \to \infty$ limit, that is, the convergence of the power series $\sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$. We are also interested are interested for when $f(x) = \sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$ holds (that is, when is a function equal to its Taylor series)? These turn out to be distinct questions; in particular, there are functions whose power series converge for all x but are equal to their power series nowhere (except at x_0 where equality must hold). This motivates the following definition:

Definition: Analyticity

A function f is **analytic** if $f(x) = \sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$ in a neighbourhood of x_0.

To motivate this definition, it will help to study a function which is *not* analytic. Consider the function:

$$f(x) = \begin{cases} \exp\left(-\frac{1}{x}\right) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

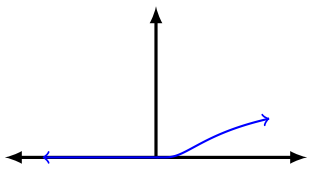

Figure 23: Plot of f.

f is continuous everywhere by construction. It is infinitely differentiable at $x = 0$, but it is only equal to its Taylor series around $x_0 = 0$ for $x \leq 0$. To see this, we observe that:

$$f'(x) = \frac{1}{x^2} \exp\left(-\frac{1}{x}\right)$$

$$f^{(n)}(x) = \frac{P_n(x)}{x^{2n}} \exp\left(-\frac{1}{x}\right)$$

We have that $f^{(n)}(x) \to 0$ as $x \to 0$ for all n as the exponential dominates the polynomial singularity. We hence have that $f \in C^\infty$. The taylor polynomial at $x_0 = 0$ however, as we have that $f(0) = 0$ and $f^{(n)}(0) = 0$, leading to:

$$\sum_{n=0}^{N} \frac{f^{(n)}(0)}{n!} x^n = 0$$

for all N. We therefore have that the series converges for all x, but is only equal to f for $x \leq 0$; hence it is not analytic (as there exists no neighbourhood around $x_0 = 0$ for which f is equal to its Taylor series).

We now consider a function $\chi(x)$ defined as $\chi(x) = \frac{f(x)}{f(x)+f(1-x)}$. This function is alsoz continuous, its denominator is never zero, and $\chi \in C^\infty$. We observe that $\chi(x) = 0$ for $x \leq 0$ and $\chi(x) = 1$ for $x \geq 1$, and overall the function looks much like a step function:

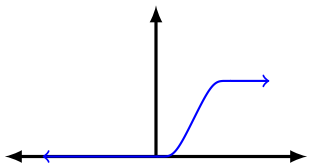

Figure 24: Plot of χ.

Indeed, this function can be used as a "cutoff"/"switch" function that behaves much like a step function (except it is infinitely differentiable).

A question becomes whether such a function could be analytic. The answer turns out to be no, and the proof we leave as an exercise. As a sketch, consider an analytic function g such that $g(x) = 0$ for $x \leq x_0$ and $g(x) \neq 0$ for $x > x_0$. One can derive a contradiction by considering the Taylor series expansion about x_0 and then using the assumed analyticity of g.

Note that in a sense, Taylor's Theorem is the culmination of a sequence of theorems we have proven in the course. Roughly, the sequence was as follows:

1) $f : X \mapsto Y$ and $K \subset X$, then $f(K)$ compact (Theorem 4.14)

2) Extreme Value Theorem: If $f : X \mapsto \mathbb{R}$ with f continu(X, τ) and (Y, ρ) ous and K compact, then f realizes its supremum and infimum on K. (Theorem 4.16)

3) Rolle's Theorem

4) Mean Value Theorem (Theorem 5.10)

5) Taylor's Theorem (Theorem 5.15)

A question that arises is could we have gone through this sequence of proofs with just the rational numbers ($\mathbb{Q}$)? The intuitive answer is no, but it may be interesting to see where along this chain the logic breaks down.

The first step that looks immediately questionable is step 2; the supremum/infimum is not well-defined for all subsets of $\mathbb{Q}$, so we might be able to find a breakdown there. To this end, we consider the set $A = \left\{q \in \mathbb{Q} : q > 0, q^2 < 2\right\}$ that arises in Example 1.1 and try to find a function $f : K \mapsto \mathbb{Q}$ with $K \subset \mathbb{Q}$ compact such that $f(K) = A$. An idea would be to try $f(x) = \sqrt{x}$, with $K = \left\{q \in \mathbb{Q} : \exists r : r^2 = q\right\} \cap [0,2]$ but this doesn't work as K is not compact. Trying another attempt, $K = [1,2] \cap \mathbb{Q}$ with $f(x) = \sin(x)$ does not work either as $\sin(x)$ is not necessarily rational, and moreover, $[1,3] \cap \mathbb{Q}$ is not a compact set as not all Cauchy sequences in S converge! Another attempt would be $K = [1,2] \cap \mathbb{Q} = X$ with $f(q) = \left|q^2 - 2\right|$ where it would seem as though $f(K) = \{r \in \mathbb{Q} : 0 < r \leq 2\}$ provides a good counterexample, but this fails for the same reason as K is not compact. Finding a valid counterexample for the EVT is therefore difficult.

An easier break along the chain to find is with Rolle's Theorem. One can consider the function $f(x) = x^2 - 2$ which can break the Intermediate value theorem as 0 is not contained in the image if the domain is $\mathbb{Q}$. We can dress this up to construct a counterexample for Rolle's Theorem.

5.4 Local Behavior of Functions

Theorem: Second Derivative Test

Suppose $f \in C^3(I)$ where I is a neighbourhood of x_0. Furthermore, suppose that $f'(x_0) = 0$. If $f''(x_0) > 0$, then x_0 is a local minimum. Conversely, if $f''(x_0) < 0$, then x_0 is a local maximum.

Proof

By Taylor's Theorem (Theorem 5.15), if we let $x = x_0 + h$ for $h > 0$, there exists some $\tilde{x} = x_0 + \lambda h$ with $\lambda \in (0,1)$ such that:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{6}f^{(3)}(\tilde{x})(x - x_0)^3$$

Using that $f'(x_0) = 0$, we have:

$$f(x) - f(x_0) = h^2\left(\frac{1}{2}f''(x_0) + \frac{1}{6}f^{(3)}(x_0 + \lambda h)h\right)$$

Let $0 < \epsilon < \left|\frac{1}{2}f''(x_0)\right|$. Then, by the assumed continuity of $f^(3)$, we have that there exists $\delta > 0$ such that $|h| < \delta$ implies $\left|\frac{1}{6}f^{(3)}(x_0 + \lambda h)h\right| < \epsilon$. Hence, for sufficiently small h, we have that:

$$\operatorname{sgn}\left(f(x) - f(x_0)\right) = \operatorname{sgn}(f''(x_0))$$

So we conclude that if $f'(x_0) = 0$ and $f''(x_0) > 0$ then x_0 is a local minimum, and if $f''(x_0) < 0$, then x_0 is a local maximum. □

Definition: Convex Functions

Let $f : (a,b) \mapsto \mathbb{R}$ is **convex** if for all $x, y \in (a,b)$ with $a < x < y < b$ and for all $\lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Note that an alternative definition of convexity is that for any x, y with $x < y$, the function evaluated at some point $z \in (x, y)$ will always be below the average of the function at x, y. That is:

$$f(z) \leq \frac{f(x) + f(y)}{2}$$

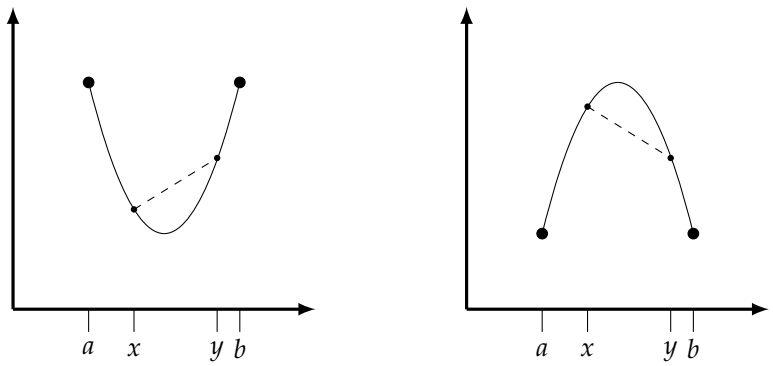

Figure 25: Visualization of a convex and non-convex function. For the upwards facing parabola, the function (and hence all points $(\lambda x + (1-\lambda)y, f(\lambda x + (1-\lambda)y))$ for $\lambda \in (0, 1)$) always lies below the line connecting $(x, f(x))$ and $(y, f(y))$ for $a < x < y < b$ and hence it is convex. For the downwards facing parabola, this is no longer true and the function is not convex (it is instead*concave*).

Another equivalent way of defining convexity is to say that f is convex if and only if for all $a < s < t < y < b$:

$$\frac{f(t) - f(s)}{t - s} \leq \frac{f(u) - f(t)}{u - t}$$

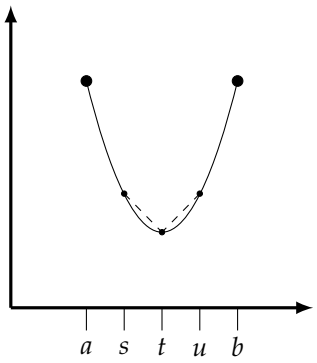

Figure 26: Visualization of the alternative definition of convexity. For any $a < s < t < u < v$, the slope of the line segment joining s and t is less than the slope of the line segment joining t and u.

Theorem

(a) Assume that $f : (a, b) \mapsto \mathbb{R}$ is convex. Then, f is continuous.

(b) Assume that $f \in C^1(a, b)$. Then, if f is convex, f' is increasing.

(c) Assume that $f \in C^2(a, b)$. Then f convex implies $f'' \geq 0$.

Proof

(a) Let $[c, d] \subset (a, b)$ and $a < c_1 < c < x < y < d < d_1 < b$. By convexity, we have that:

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(d) - f(y)}{d - y} \leq \frac{f(d_1) - f(d)}{d_1 - d}$$

and also that:

$$\frac{f(y) - f(x)}{y - x} \geq \frac{f(x) - f(c)}{x - c} \geq \frac{f(c) - f(c_1)}{c - c_1}$$

We therefore have that:

$$\left\{ \left| \frac{f(y) - f(x)}{y - x} \right| : c < x < y < b \right\} < M$$

for some $M \in \mathbb{R}$. Therefore, $\left|f(y) - f(x)\right| < M\left|y - x\right|$ for all $x, y \in (c, d)$. This holds for all $[c, d] \subset (a, b)$, showing the continuity of f.

(b) Let f be convex, and let $a < c < x < y < d < b$. Then, by convexity we have that:

$$\frac{f(x) - f(c)}{x - c} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(d) - f(y)}{d - y}$$

Ignoring the central term in the inequality, and taking the limit as $x \to c$ and $y \to b$, we have that:

$$f'(c) \leq f'(b)$$

So we conclude that f' is increasing on (a, b).

(c) If f is convex, f' is increasing by (b). Then, we have that for any $a < x < y < b$:

$$\frac{f(y) - f(x)}{y - x} \geq 0$$

And hence taking $y \to x$ we have that $f'(x) \geq 0$. □

Corollary

If $f \in C^3(I)$ and $f'(x_0) = 0$ and $f''(x_0) \neq 0$, x_0 is a local minimum if f is convex in a neighbourhood of x_0.

6 The Riemann-Stieltjes Integral

6.1 Definition of the Integral

Definition 6.1: Partition

A **partition** of $[a, b] \subset \mathbb{R}$ is a set $\{x_0, x_1, \ldots, x_n\}$ (for some $n \in \mathbb{N}$) such that:

$$a = x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n = b$$

We can then write:

$$\Delta x_i = x_i - x_{i-1}$$

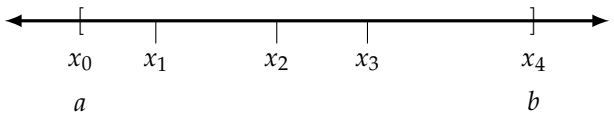

Figure 27: Visualization of a partition $\{x_0, x_1, x_2, x_3, x_4\}$ of $[a, b]$. Note that the points in the partitions need not be equally spaced.

Definition 6.1: Upper and Lower Sums

Given $f : [a, b] \mapsto \mathbb{R}$ and a partition P of $[a, b]$ let:

$$M_i = \sup \{f(x) : x_{i-1} \leq x \leq x_i\}$$
$$m_i = \inf \{f(x) : x_{i-1} \leq x \leq x_i\}$$

Then, we can define the **upper** and **lower sums**:

$$U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i$$
$$L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i$$

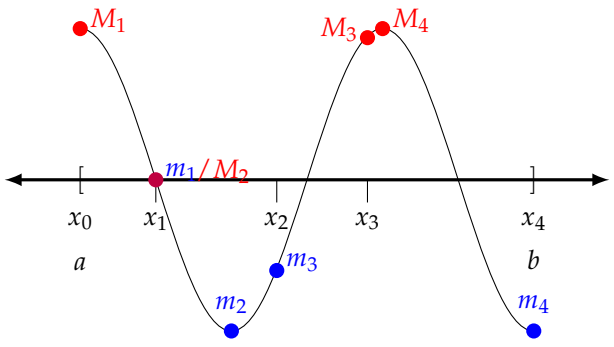

Figure 28: Example of a function f, a partition P of $[a, b]$, and the M_i, m_is for this choice of partition.

By construction, it should be evident that $L(P,f) \leq U(P,f)$ for all P, f.

A natural question that arises from the form of the above expression is whether these are Riemann sums or not. Recall from first year calculus that we would choose the left endpoint, right endpoint, or some other arbitrary choice of a point in the subinterval. Here, we in a sense use a "special case" of the supremum/infimum. We will see that this choice is musch easier to use in proofs due to monotonicity properties. Namely, if we have a partition and add another point, then $U(P,f)$ can only decrease, and $L(P,f)$ can only increase (we will see this in a theorem soon)!

Definition 6.1: Upper/Lower Integrals and Riemann Integrability

We define the **upper Riemann integral** to be:

$$\overline{\int_a^b} = \inf_P U(P,f).$$

and the **lower Riemann integral** to be:

$$\underline{\int_a^b} = \sup_P L(P,f).$$

Here, the infimum/supremum is taken over all partitions P. We say that f is **Riemann integrable** on $[a,b]$, and write $f \in \mathcal{R}[a,b]$ if:

$$\overline{\int_a^b} = \underline{\int_a^b}$$

which we can write as:

$$\int_a^b f dx \text{ or } \int_a^b f(x)dx$$

Note that the choice of variable in the above definition is totally arbitrary.

Also, note that while f is not required to be continuous in the above definition, it is required to be bounded; else, M_i and m_i may not exist. Since f is bounded, $U(P,f)$, $L(P,f)$ are bounded for all P, f and hence we have a set of real numbers for which we may consider the supremum/infimium of by the LUB/GLB property of the reals. Since the upper/lower sums lie in a bounded interval, there is no questions about whether the lower/upper integrals exist. The question becomes whether they are equal or not. Before getting into further discussion on this topic, we discuss a bound:

Theorem 6.1: ML Bounds

Let $m = \inf\{f(x) : a \leq x \leq b\}$ and $M = \sup\{f(x) : a \leq x \leq b\}$ (which exist by the boundedness of f). Then,

$$m(b-a) \leq L(P,f) \leq U(P,f) \leq M(b-a)$$

For any choice of partition P.

Proof

For any i, we have that:

$$m \leq m_i \leq M_i \leq M$$

Therefore:

$$\sum_{i=1}^{n} m\Delta x_i \leq \sum_{i=1}^{n} m_i \Delta x_i \leq \sum_{i=1}^{n} M_i \Delta x_i \leq \sum_{i=1}^{n} M\Delta x_i$$

So we conclude that:

$$m(b-a) \leq L(P,f) \leq U(P,f) \leq M(b-a)$$

□

Now that we have established the Riemann integral, a natural question is how can we extend this notion. In order to do so, we will use a montonically increasing function $\alpha : [a,b] \mapsto \mathbb{R}$ (that is, $\alpha(x) \leq \alpha(y)$ for all $x \leq y$). Note that α need not be continuous. Indeed, compared to the Riemann integral where $\alpha(x) = x$ and was continuous, in this general setting, α is allowed to have jumps. This allows for certain benefits, as we will soon discuss. However, we note that α can only have a finite number of jumps.

Theorem 4.30

Let $\alpha : [a,b] \mapsto \mathbb{R}$ be monotonic. Then, it can only have finitely many discontinities.

Proof

Assign a rational number $r(x)$ to each of the discontinuities of α. Then, we have that:

$$\lim_{x \to r(x)^-} \alpha(x) < \alpha(x) < \lim_{x \to r(x)^+} \alpha(x)$$

Since $x_1 < x_2$ implies $\lim_{x \to r(x_1)^+} \alpha(x) \leq lim_{x \to r(x_2)^-} \alpha(x)$, we have that $r(x_1) \neq r(x_2)$ if $x_1 \neq x_2$. We therefore have established a function r from the set of discontinuities of α to the rationals. As the rationals are countable, the set of discontiuities of α are also countable. □

With this established, we now define the generalized Riemann-Stieltjes integral.

Definition 6.2: Riemann-Stieltjes Integral

Let $\alpha : [a,b] \mapsto \mathbb{R}$ be increasing, and given a partition P of $[a,b]$, define:

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}) (\geq 0)$$

For bounded $f : [a,b] \mapsto \mathbb{R}$, Let:

$$U(P,f,\alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i$$
$$L(P,f,\alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i$$

We then take the infimum/supremum over partitions P to get:

$$\overline{\int_a^b} f d\alpha = \inf_P U(P,f,\alpha)$$
$$\underline{\int_a^b} = \sup_P L(P,f,\alpha)$$

If equal, we write their value as:

$$\int_a^b f d\alpha \text{ or } \int_a^b f(x) d\alpha(x)$$

and we write that $f \in \mathcal{R}_\alpha[a,b]$. In the case where $\alpha(x) = x$, we recover the Riemann integral.

Why is this definition useful? What does it accomplish for us that the original Riemann integral does not? We consider a physically motivating example. Suppose we have a thin wire with varying mass density $\rho(x)$. If we wanted to calculate the mass density of the wire, we would integrat the density $\rho(x)$ over the length of the wire. Now, suppose our wire consists of steel of continuously varying mass density, as well as beads/point masses placed on certain locations of the wire. The Riemann integral cannot handle these point masses, but the Riemann-Stieltjes integral can deal with this case if we use an α with discontinuities in it. Hence, the Riemann-Stieltjes integral allows us to handle cases where we both have continuous and discrete masses to integrate over. It acts as a bridge between Riemann and Lebesgue integration (the latter of which will be the subject of a later course in measure theory).

We now will answer the question: "for what choices of f, α is f Riemann-Stieltjes integrable?"

6.2 Criterion for Integrability

Definition 6.3: Refinements and Common Refinemnet

P^* is a **refinement** of P if $P \subset P^*$ and P, P^* are partitions. The common refinemnet of P_1, P_2 is $P^* = P_1 \cup P_2$.

Theorem 6.4

If P^* is a refinemne tof P, then:

$$L(P,f,\alpha) \leq L(P^*,f,\alpha) \leq U(P^*,f,\alpha) \leq U(P,f,\alpha)$$

As a remark, when we take infimums/supremums over partitions P to obtain the upper/lower Riemann-Stieltjes integals, we are taking refinements.

Also, note that the above theorem does *not* apply to (right-hand, left-hand, midpoint, arbitrary) Riemann sums, and is a consequence of the choice of upper/lower sums with supremums/infimums taken over the subintervals.

Proof

It suffices to consider P^* with a single extra point $x_{i-1} < x^* < x_i$, and then the genral case follows by induction.
For the case where $\alpha(x) = x$, the refinement adds $(m^* - m_i)(x^* - x_{i-1}) \geq 0$.
For the general case, we have that:

$$\begin{aligned} L(P^*, f, \alpha) - L(P, f, \alpha) &= \big(m^*(\alpha(x^*) - \alpha(x_{i-1})) + m_i(x^* - x_{i-1})\big) - m_i(\alpha_{x_i} - \alpha_{x_{i-1}}) \\ &= \big(m^*(\alpha(x^*) - \alpha(x_{i-1})) + m_i(x^* - x_{i-1})\big) \\ &\quad - m_i\left[(\alpha(x^*) - \alpha(x_{i-1})) + (\alpha(x_i) - \alpha(x^*))\right] \\ &= (m^* - m_i)(\alpha(x^*) - \alpha(x_i)) \end{aligned}$$

α is montonically increasing, so $x^* \geq x_i$ implies that the second term is positive. Furthermore, $m^* \geq m_i$ as $\inf\{f(x) : x \in [x_{i-1}, x^*]\} \geq \inf\{f(x) : x \in [x_{i-1}, x_i]\}$. It follows that:

$$L(P^*, f, \alpha) - L(P, f, \alpha) \geq 0$$

and the proof for $U(P^*, f, \alpha) - U(P, f, \alpha) \leq 0$ follows analogously. □

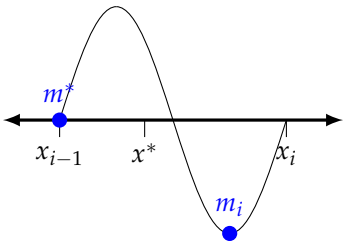

Figure 29: Visualization of the effect of adding an extra point x^* to the partition P. We can see that this has the net effect of increasing $L(P, f, \alpha)$ as $m^* \geq m_i$.

Note that as a point of notation, if f, α are fixed, we sometimes can write $L(P), U(P)$ in place of $L(P, f, \alpha)$ and $U(P, f, \alpha)$. Sometimes where the context is clear, it is also common to write $\mathcal{R}(\alpha)$ in place of $\mathcal{R}_\alpha[a, b]$.

Theorem 6.5

$\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$.

Proof

For partitions P_1, P_2, let $P^* = P_1 \cup P_2$ be the common refinement. By Theorem 6.4, we have that:

$$L(P_1) \leq L(P^*) \leq U(P^*) \leq U(P_2)$$

And in particular, $L(P_1) \leq U(P_2)$. Therefore, for any fixed P_2, $U(P_2)$ is an upper bound on the set of all lower sums. As the supremum is the least upper bound, we have that:

$$\sup_{P_1} L(P_1) \leq U(P_2)$$

Therefore, as $\sup_{P_1} L(P_1)$ is a lower bound on the set of all upper sums, and the infimum is the greatest lower bound, we have that:

$$\sup_{P_1} L(P_1) \leq \inf_{P_2} U(P_2)$$

So therefore:

$$\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$$

as claimed. □

Theorem 6.6: ϵ-Criterion for Integrability

$f \in \mathcal{R}_\alpha[a, b]$ if and only if for all $\epsilon > 0$, there exists a partition P_ϵ of $[a, b]$ such that $U(P_\epsilon) - L(P_\epsilon) < \epsilon$.

Proof

$\boxed{\Longrightarrow}$ By hypothesis, $\sup_P L(P) = \int_a^b f d\alpha = \inf_P U(P)$. Let $\epsilon > 0$. Then by the property of sup/inf, there exist P_1, P_2 such that:

$$\int_a^b f d\alpha - L(P_1) < \frac{\epsilon}{2}$$
$$U(P_2) - \int_a^b f d\alpha < \frac{\epsilon}{2}$$

Adding the first inequality to the second, we get:

$$U(P_2) - L(P_1) < \epsilon$$

Letting $P^* = P_1 \cup P_2$, by Theorem 6.4 we have that:

$$U(P^*) - L(P^*) \leq U(P_2) - L(P_1) < \epsilon$$

which proves the claim.
$\boxed{\Longleftarrow}$ Let $\epsilon > 0$ Then by Theorem 6.5 we have that:

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha$$

and furthermore:

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq U(P_\epsilon) - L(P_\epsilon) < \epsilon$$

Where the second inequality is true for any choice of partition. ϵ is arbitrary, so we conclude that:

$$\overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha = 0$$

And therefore $\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$, and $f \in \mathcal{R}_\alpha[a, b]$. □

Theorem 6.7 is a little technical, so we shall skip it for now.

Theorem 6.8: Continuity implies integrability

If f is continuous on $[a, b]$, then $f \in \mathcal{R}_\alpha[a, b]$.

Note in the above theorem that we make no assumptions on α, only that (of course) it is monotonic.

Proof

By definition, $U(P) - L(P) = \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i$. The idea will be to choose small intervals to make these differences small. Since $[a, b]$ is compact, f is uniformly continuous by Theorem 4.19. So, for all $\eta > 0$, there exists $\delta > 0$ such that $\left|f(x) - f(t)\right| < \eta$ if $|x - t| < \delta$. Thus, if P is constructed such that $\Delta x_i < \delta$ then $M_i - m_i < \eta$. We then have that:

$$U(P) - L(P) \leq \sum_{i=1}^{n} \eta\Delta\alpha_i = \eta \sum_{i=1}^{n} (\alpha(x_i) - \alpha(x_{i-1})) = \eta\left(\alpha(b) - \alpha(a)\right)$$

Where in the last equality we use the fact that we have a telescoping sum. Given $\epsilon > 0$, we choose $\eta < \frac{\epsilon}{\alpha(b)-\alpha(a)}$. With this choice of partition with $\Delta x_i < \delta = \delta(\eta)$, we have that:

$$U(P) - L(P) < \epsilon$$

and we conclude that $f \in \mathcal{R}_\alpha[a, b]$ by Theorem 6.6. □

A natural question is "what fs are Riemann integrable in general?" The answer turns out to be "if f is continuous almost everywhere". This sounds handwavy, but has a precise definition; although it is not covered in this course, one can refer to Rudin 11.33(b) for details.

Theorem 6.9

If f is monotone on $[a, b]$ and α is continuous on $[a, b]$m then $f \in \mathcal{R}_\alpha[a, b]$.

In the proof of Theorem 6.8, we used the continuity of f to bound the maximum/minimum on each subinterval. Here, f is no longer continuous, so we cannot control the maximum/minimum. Instead, we will use the continuity of α to control the size of the $\Delta\alpha$s.

Proof

Given $n \in \mathbb{N}$, choose P such that $\Delta\alpha_i = \frac{\alpha(b)-\alpha(a)}{n}$ for all $i \in 1, \ldots, n$. Note that such a choice is possible by the continuity of α and the IVT (Theorem 4.23). We then have that:

$$U(P) - L(P) = \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i = \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} (M_i - m_i)$$

Suppose (WLOG) that f is an increasing function. Then, $M_i = f(x_i)$ and $m_i = f(x_{i-1})$ due to the monotone increasing property. Hence:

$$U(P) - L(P) = \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} f(x_i) - f(x_{i-1}) = \frac{\alpha(b) - \alpha(a)}{n} \left[f(b) - f(a)\right]$$

Where in the last equality we use the fact that the sum telescopes. If $\alpha(b) = \alpha(a)$, then $U(P) - L(P) = 0$ and the claim immediately follows. If $\alpha(b) > \alpha(a)$, then the claim follows by choosing $n > \frac{1}{\epsilon} \frac{f(b)-f(a)}{\alpha(b)-\alpha(a)}$ (that is, selecting a partition P with sufficiently large n), from which it follows that:

$$U(P) - L(P) < \epsilon$$

and hence $f \in \mathcal{R}_\alpha[a, b]$ by Theorem 6.6. □

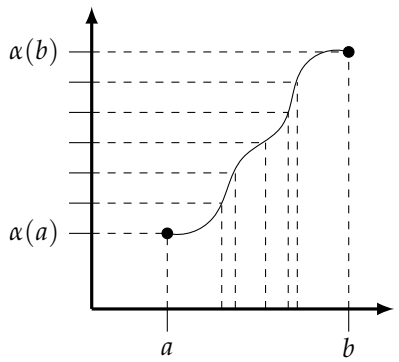

Figure 30: Visualization of the idea of the proof of Theorem 6.9. We chop up $[\alpha(a), \alpha(b)]$ into equal sized subintervals.

Theorem 6.10

Suppose $f : [a, b] \mapsto \mathbb{R}$ is bounded and has finitely many discontinuities. Suppose α is continuous at every point where f is not. Then, $f \in \mathcal{R}_\alpha[a, b]$.

Proof

As has become standard, we will be applying Theorem 6.6. We have that:

$$U(P) - L(P) = \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i$$

Let $\epsilon > 0$, and let $E = \{e_1, \dots e_k\}$ be the set of points where f is not continuous. α is continuous at each e_j by hypothesis. Therefore, there exists disjoint intervals (u_j, v_j) that cover points in E, such that $u_j < e_j < v_j$ and $\alpha(v_j) - \alpha(v_j) < \epsilon$ (by the continuity of α). Let $K = [a, b] \cap \left(\bigcup_{j=1}^{k} (u_j, v_j)\right)^c$. K is a compact set as it is a finite union of closed intervals. f is continuous on K by hypothesis, so f is uniformly conitnuous on K by Theorem 4.19. Hence, there exists $\delta > 0$ such that for $s, t \in K$, $|s - t| < \delta \implies \left|f(s) - f(t)\right| < \epsilon$. We then form a partition $P = \{x_1, \dots, x_n\}$ to consist of $\{u_1, v_1, \dots, u_k, v_k\}$ and additional points in K with $\delta x_i < \delta$. For such i, f is continuous on the subinterval and hence $M_i - m_i < \epsilon$. For the other intervals $[u_j, v_j]$, we have that $M_j - m_j \leq 2M$ where $M = \sup\left\{\left|f(x)\right| : x \in [a, b]\right\}$ and that $\Delta\alpha_j < \epsilon$. Therefore, we have that:

$$\begin{aligned} 0 \leq U(P) - L(P) &= \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i \\ &\leq 2M\epsilon + \epsilon\left(\alpha(b) - \alpha(a)\right) \end{aligned}$$

Where the first term comes from the $[u_j, v_j]$s and the second term is the maximum possible value from the subintervals of K. By choosing ϵ small enough, the RHS is small as desired for some choice of P. Hence, $U(P) - L(P) < \epsilon$ for some P and hence $f \in \mathcal{R}_\alpha[a, b]$. □

A natural question that arises is "what if f and α are discontinuous at the same point?" In this case, we can construct functions such that $f \notin \mathcal{R}_\alpha[a, b]$ (see HW1Q5).

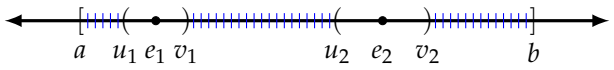

Figure 31: Visualization of how $[a, b]$ gets split up in the proof of Theorem 6.10. K consists of the union of the closed intervals marked in blue.

Theorem 6.11

If $f \in \mathcal{R}_\alpha[a, b]$, $m \leq f(x) \leq M$ for all $x \in [a, b]$, and ϕ is continuous on $[m, M]$, then $\phi \circ f \in \mathcal{R}_\alpha[a, b]$.

Proof

Not covered in lecture, see Rudin. □

As an example of the above theorem, suppose we have that $f \in \mathcal{R}(\alpha)$. Then, we have that $f^2 \in \mathcal{R}(\alpha)$ and $|f| \in \mathcal{R}(\alpha)$ as $\phi(x) = x^2$ and $\phi(x) = |x|$ are both continuous functions.

6.3 Properties of the Integral

Theorem 6.12: Linearity and Related Properties

(a) Suppose $f_1 \in \mathcal{R}_\alpha[a, b]$ and $f_2 \in \mathcal{R}_\alpha[a, b]$. Then, $f_1 + f_2 \in \mathcal{R}_\alpha[a, b]$ and $cf_1 \in \mathcal{R}_{[a,b]}(\alpha)$ for all $c \in \mathbb{R}$. Furthermore:

$$\int_a^b (f_1 + f_2)d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha \text{ and } \int_a^b cf_1 d\alpha = c\int_a^b f_1 d\alpha$$

(b) Suppose $f_1, f_2 \in \mathcal{R}_\alpha[a, b]$ and $f_1(x) \leq f_2(x)$ for all $x \in [a, b]$. Then,

$$\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha$$

(c) If $f \in \mathcal{R}_\alpha[a, b]$ and $c \in (a, b)$, then $f \in \mathcal{R}_\alpha[a, c] \cap \mathcal{R}_\alpha[c, b]$ and furthermore:

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha$$

(d) If $f \in \mathcal{R}_\alpha[a, b]$ and $|f(x)| \leq M$ for all $x \in [a, b]$, then:

$$\left| \int_a^b f d\alpha \right| \leq M(\alpha(b) - \alpha(a))$$

(e) If $f \in \mathcal{R}_{\alpha_1}[a, b]$ and $f \in \mathcal{R}_{\alpha_2}[a, b]$, then $f \in \mathcal{R}_{\alpha_1+\alpha_2}[a, b]$ and $f \in \mathcal{R}_{c\alpha_1}[a, b]$ for all $c \geq 0$. Furthermore:

$$\int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 = \int_a^b f d(\alpha_1 + \alpha_2) \text{ and } \int_a^b f d(c\alpha_1) = c\int_a^b f d\alpha_1$$

Proof

(a) See Rudin and HW2Q1.

(b) We have that $L(P, f_1) \leq L(P, f_2)$ for all partitions P as $\inf\{f_1(x) : x \in [x_{i-1}, x_i]\} \leq \inf\{f_2(x) : x \in [x_{i-1}, x_i]\}$ for all i (as $f_1(x) \leq f_2(x)$). Therefore, we have that:

$$\int_a^b f_1 d\alpha = \sup_P L(P, f_1) \leq \sup_P L(P, f_2) = \int_a^b f d\alpha.$$

(c) Exercise (see HW2Q1 for the $\overline{\int_a^b}$ case)

(d) Consider any partition P of $[a, b]$. Let $M = \sup\left\{|f(x) : x \in [a, b]|\right\}$. We then have that (using the ML bound from Theorem 6.1):

$$\begin{aligned} -M[\alpha(b) - \alpha(a)] = \sum_{i=1}^n (-M)\Delta\alpha_i \leq \sum_{i=1}^n M_i \Delta\alpha_i = L(P, f, \alpha) \leq \int_a^b f d\alpha \\ \leq U(P, f, \alpha) = \sum_{i=1}^n M_i \Delta\alpha_i \leq \sum_{i=1}^n M \Delta\alpha_i = M[\alpha(b) - \alpha(a)]. \end{aligned}$$

(e) We prove the first (additivity) statement. Let $\epsilon > 0$. Choose $P_i, i \in \{1, 2\}$ (By Theorem 6.6) such that

$$U(P_i, f, \alpha_i) - L(P_i, f, \alpha_i) < \frac{\epsilon}{2}$$

Let $P = P_1 \cup P_2$. Then, by Theorem 6.4, we have that:

$$U(P^*, f, \alpha_i) - L(P^*, f, \alpha_i) < \frac{\epsilon}{2}$$

Adding the two expressions (for $i = 1, 2$) together, we have:

$$U(P^*, f, \alpha_1 + \alpha_2) - L(P^*, f, \alpha_1 + \alpha_2) < \epsilon$$

So $f \in \mathcal{R}_{\alpha_1+\alpha_2}[a, b]$ by Theorem 6.6. Furthermore, We have that:

$$\int_a^b f d(\alpha_1 + \alpha_2) \leq U(P^*, f, \alpha_1 + \alpha_2) = U(P^*, f, \alpha_1) + U(P^*, f, \alpha_2) < \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 + \epsilon$$

where we apply the inequality of $U(P^*, f, \alpha_i) < \int_a^b f d\alpha_i + \frac{\epsilon}{2}$ in the last line. Similarly, we have that:

$$\int_a^b f d(\alpha_1 + \alpha_2) \geq L(P^*, f, \alpha_1 + \alpha_2) = L(P^*, f, \alpha_1) + L(P, f, \alpha_2) > \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 - \epsilon$$

Since ϵ is arbitrary, we therefore conclude that:

$$\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$$

□

Theorem 6.13

(a) Suppose $f, g \in \mathcal{R}(\alpha)$. Then, $fg \in \mathcal{R}(\alpha)$.

(b) Suppose $f \in \mathcal{R}(\alpha)$. Then, $|f| \in \mathcal{R}(\alpha)$ and $\left|\int_a^b f d\alpha\right| \leq \int_a^b |f| d\alpha$.

Proof

(a) By Theorem 6.12(a), $f \pm g \in \mathcal{R}(\alpha)$, so by Theorem 6.11, $(f \pm g)^2 \in \mathcal{R}(\alpha)$. Since $(f+g)^2 - (f-g)^2 = 4fg$, we then have that:

$$\frac{(f+g)^2 + (f-g)^2}{4} = fg \in \mathcal{R}(\alpha)$$

(b) By Theorem 6.11, $|f| \in \mathcal{R}(\alpha)$ letting $\phi(t) = |t|$. Let $c = \operatorname{sgn}\left(\int_a^b f d\alpha\right) \in \{-1, 0, 1\}$. Then, we have that:

$$\left|\int_a^b f d\alpha\right| = c \int_a^b d\alpha = \int_a^b cf d\alpha \leq \int_a^b |f| d\alpha$$

Where we use Theorem 6.12(a) in the second equality, and Theorem 6.12(b) in the last inequality (as $cf \leq |f|$). □

Theorem 6.15

Suppose f is bounded on $[a, b]$, $s \in (a, b)$, and f is continuous at s. Let:

$$\alpha(x) = \begin{cases} 0 & x \leq s \\ 1 & x > s \end{cases}$$

Then, $\int_a^b f d\alpha = f(s)$ (and in particular, $f \in \mathcal{R}(\alpha)$).

This result is interesting, and some remarks on the nature of this theorem are in order. Firstly, by Theorem 4.29 (not covered in Lecture, see Rudin), since α is monotonically increasing, we have that $\lim_{t \to x^+} \alpha(t) = \alpha(x^+)$ and $\lim_{t \to x^-} \alpha(t) = \alpha(x^-)$ exist at every $x \in (a, b)$ and $\alpha(x^-) \leq \alpha(x) \leq \alpha(x^+)$. Secondly, note that Rudin defines α in the above Theorem to be left contiuous, but in probability theory, it is conventional to use a right continuous (i.e. $\alpha(x) = \alpha(x^+)$ for all $x \in (a, b)$). We leave it as an exercise to prove the theorem for the case of a right continuous α (the strategy and result are identitcal).

Any physicists looking at the result of the Theorem will note that it looks very much like the "Dirac Delta" function; indeed, this choice of α is making rigorous the notion of a δ function where:

$$\int_a^b f(x)\delta(x - s) = f(s)$$

We leave it as a homework exercise (HW2Q2) to prove that no such function can actually exist. This step-function method is one way to make the δ function well-defined; other methods include taking the limit of a bell curve, or bringing in the theory of distributions (the latter which is most definitely outside the scope of this course).

Finally, we note that this example really shows off something that the Riemann integral cannot reproduce; the incorporation of "discrete" points.

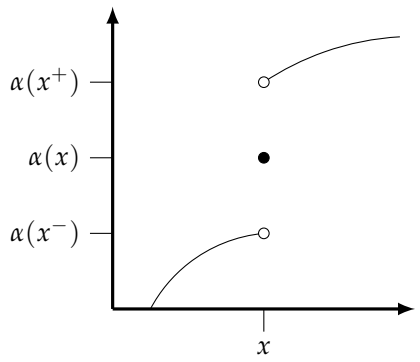

Figure 32: Visualization of Theorem 4.29. For any strictly increasing function, we have that $\alpha(x^-) \leq \alpha(x) \leq \alpha(x^+)$ at a discontinuity x.

Proof

Choose $P = \{x_0, x_1 = s, x_2, x_3 = b\}$. We then have that:

$$U(P) = \sum_{i=1}^{3} M_i \Delta\alpha_i = M_2 \Delta\alpha_2 = M_2 = \sup\{f(x) : x \in [x_1, x_2]\}$$
$$L(P) = \sum_{i=1}^{3} m_i \Delta\alpha_i = m_2 \Delta\alpha_2 = m_2$$

So, we have that $m_2 \leq \int_a^b f d\alpha \leq M_2$ (assuming the integral exists). Taking the limit as $x_2 \to s^+$, by the continuity of f at s we have that $M_2 \to f(s)$ and $m_2 \to f(s)$. Therefore, $\int_a^b f d\alpha$ exists and equals $f(s)$. □

Note that the above proof works exactly the same way if $s = a$. If $s = b$, then because we take α to be left continuous, $\alpha = 0$ everywhere on $[a, b]$ and the integral just equals zero.

Definition 6.14: Unit Step Function

We define the (left continuous) **unit step function** as:

$$I(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

In Theorem 6.15, we have that $\alpha(x) = I(x - s)$.

Theorem 6.16

Let $c_n \geq 0$ be such that $\sum_{n=1}^{\infty} c_n < \infty$. Take $s_n \in (a, b)$ distinct (that is, $s_n \neq s_m$ if $n \neq m$). Define $\alpha(x) = \sum_{n=1}^{\infty} c_n I(x - s_n)$ (which converges/exists by comparison, as $\sum c_n$ converges). Let f be continuous on $[a, b]$. Then,

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} c_n f(s_n)$$

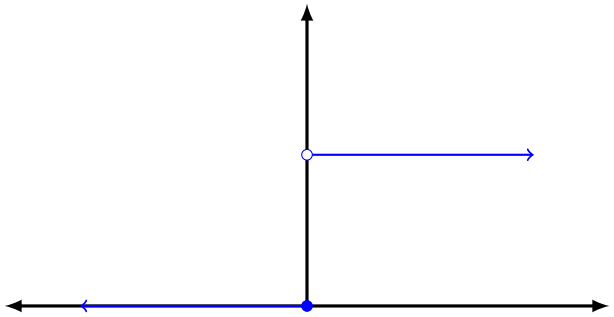

Figure 33: Plot of the (left continuous) unit step function. To obtain the α used in Theorem 6.15, move the step to $x = s$ instead of $x = 0$.

Note that the resultant series in the theorem above is convergent/well defined as f is bouded on $[a, b]$ and hence $\sum_{n=1}^{\infty} c_n f(s_n) \leq M \sum_{n=1}^{\infty} c_n$ where $M = \max\{f(x)\}$. Hence the series converges by comparison. If all the c_ns are zero except for one, this is just Theorem 6.15. Note that the above result holds for a finite sum as well (we would use induction to prove it in this case).

Proof

Let $R_N = \int_a^b f d\alpha - \sum_{n=1}^{N} c_n f(s_n)$. We show that $R_n \to 0$ as $N \to \infty$ (i.e. given $\epsilon > 0$, we show there exists N_0 such that $|R_N| < \epsilon$ if $N \geq N_0$). We define:

$$\alpha_1(x) = \sum_{n=1}^{N} c_n I(x - s_n) \quad \alpha_2(x) = \sum_{n=N+1}^{\infty} c_n I(x - s_n)$$

Then, by Theorem 6.12(e) we have that:

$$\int_a^b f d\alpha = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$$

Then, we have that:

$$\begin{aligned} \int_a^b d\alpha_1 &= \sum_{i=1}^{N} \int_a^b f(x) d(c_n I(x - s_n)) \\ &= \sum_{n=1}^{N} c_n \int_a^b f(x) d(I(x - s_n)) \\ &= \sum_{n=1}^{N} c_n f(s_n) \end{aligned}$$

Wherein the first two equalities we again apply Theorem 6.12(e) and in the last equality we use Theorem 6.15. We therefore obtain that $R_N = \int_a^b d\alpha_2$. We then have that:

$$|R_n| \leq M \left[\alpha_2(b) - \alpha_2(a)\right] = M \sum_{n=N+1}^{\infty} c_n$$

Then by the convergence of $\sum c_n$, we can choose N_0 such that $\sum_{n=N_0+1}^{\infty} c_n < \frac{\epsilon}{M}$. We therefore have that $R_N < \epsilon$ if $N \geq N_0$, proving the claim. □

Theorem 6.17

Suppose:

(i) $\left|f(x)\right| \leq M$ for all $x \in [a,b]$.

(ii) α is continuous and increasing on $[a,b]$ and differentable on (a,b).

(iii) $\alpha' \in \mathcal{R}[a,b]$.

Then, $f \in \mathcal{R}_\alpha[a,b]$ if and only if $f\alpha' \in \mathcal{R}[a,b]$ and in this case:

$$\int_a^b f d\alpha = \int_a^b f(x)\alpha'(x)dx$$

Proof

It suffices to show that:

$$\overline{\int_a^b} f d\alpha = \overline{\int_a^b} f\alpha' dx \text{ and } \underline{\int_a^b} f d\alpha = \underline{\int_a^b} f\alpha' dx$$

We prove the first equality and the second equality follows analogously. Let $\epsilon > 0$. Since $\alpha' \in \mathcal{R}$, there exists a partition P such that $U(P,\alpha') - L(P,\alpha') < \epsilon$ (this also holds for any refinement of P). Letting $A_i = \sup \alpha', a_i = \inf \alpha' on [x_{i-1}, x_i]$ we have that:

$$\sum_{i=1}^n (A_i - a_i)\Delta x_i < \epsilon$$

By the Mean Value Theorem (Theorem 5.10) we have that there exists $t_i \in [x_{i-1}, x_i]$ such that $\Delta\alpha_i = \alpha'(t_i)\Delta x_i$. We also have that for all $s_i \in [x_{i-1}, x_i]$, $\left|\alpha'(s_i) - \alpha'(t_i)\right| \leq A_i - a_i$ (α' can only vary as much as the maximum minus the minimum on the interval). We therefore have that:

$$\sum_{i=1}^n \left|\alpha'(s_i) - \alpha'(t_i)\right|\Delta x_i \leq \sum_{i=1}^n A_i - a_i \Delta x_i < \epsilon$$

Recall that we skipped Theorem 6.7, but it might be worth comparing this result to Theorem 6.7(c). Let $M = \sup f(x) : x \in [a,b]$. Now, for any choice of $s_i \in [x_{i-1}, x_i]$, we have that:

$$\left|\sum_{i=1}^n f(s_i)\Delta\alpha_i - \sum_{i=1}^n \alpha'(s_i)\Delta x_i\right| \leq \sum_{i=1}^m M\left|\alpha'(t_i) - \alpha'(s_i)\right|\Delta x_i < M\epsilon \quad (*)$$

Therefore:

$$\sum_{i=1}^n f(s_i)\Delta\alpha_i \leq \sum_{i=1}^n f(s_i)\alpha'(s_i)\Delta x_i + M\epsilon \leq U(P, f\alpha') + M\epsilon$$

We now take the supremum over each $[s_{i-1}, s_i]$. Since s_i is arbitrary, we have that $\overline{\int_a^b} f d\alpha \leq U(P,f,\alpha) \leq U(P, f\alpha') + M\epsilon$, and hence $\overline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f\alpha' dx + M\epsilon$. $(*)$ gives $\overline{\int_a^b} f\alpha' dx \leq \overline{\int_a^b} f d\alpha + M\epsilon$. Since ϵ is arbitrary, we have that $\overline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f\alpha' dx$ and $\overline{\int_a^b} f\alpha' dx \leq \overline{\int_a^b} f d\alpha$, so $\overline{\int_a^b} f\alpha' dx = \overline{\int_a^b} f d\alpha$. $\square$

This theorem tells us something that looks very familiar from Calculus... namely, a change of variables!

Recall the COV formula from first year:

$$\int_a^b f(x)dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(y))\phi'(y)dy$$

Where $x = \phi(y), y = \phi^{-1}(y)$. We now prove this notion rigorously.

Theorem 6.19: Change of Variable

Let $f : [a,b] \mapsto \mathbb{R}$. Suppose $\phi : [A,B] \mapsto [a,b]$ is continuous and strictly increasing. Suppose α is increasing on $[a,b]$ and $f \in \mathcal{R}_\alpha[a,b]$. Let $g = f \circ \phi : [A,B] \mapsto \mathbb{R}$ and $\beta = \alpha \circ \phi : [A,B] \mapsto \mathbb{R}$. Then, we have that $g \in \mathcal{R}_\beta[A,B]$, and:

$$\int_A^B g d\beta = \int_a^b f d\alpha$$

As an example, consider the integral $\int_a^b \sin x^2 dx$ for $0 \leq a < b$. We then have that $f(x) = \sin x^2$ and $\alpha(x) = x$. We make the substitution $x^2 = y$, so $\phi(x) = \sqrt{x}$ and $\phi^{-1}(y) = y^2$. Then, A = $\phi^{-1}(a) = a^2$, $B = \phi^{-1}(b) = b^2$. We then have that $g(y) = f \circ \phi(y) = f(\sqrt{y}) = \sin(y)$, and $\beta(y) = \alpha \circ \phi(y) = \alpha(\sqrt{y}) = \sqrt{y}$. We then have that:

$$\int_a^b \sin x^2 dx = \int_{a^2}^{b^2} \sin y d\beta = \int_{a^2}^{b^2} \frac{\sin y}{2\sqrt{y}} dy$$

Where the first equality follows by Theorem 6.19 and the second equality follows by Theorem 6.17.

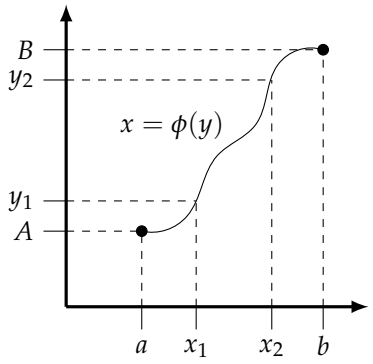

Figure 34: Plot of a (monotonically increasing and continuous) function ϕ and a demonstration of how it puts partitions of $[a,b]$ and $[A,B]$ in one-to-one correspondence. This gives the intuition for the proof of Theorem 6.19.

Proof

We make the observation that partitions P of $[a,b]$ and parititons Q of $[A,B]$ can be put in 1-1 correspondence via $x_i = \phi(y_i)$. Additionally, we make the observation that the set of g values on $[y_{i-1}, y_i]$ is equal to the set of f values on $[xi-1, x_i]$. Finally, we observe that $\alpha(x_i) = (\alpha \circ \phi)(y_i) = \beta(y_i)$. With these three observations, we have that:

$$U(P,f,\alpha) = U(Q,g,\beta) \text{ and } L(P,f,\alpha) = U(Q,g,\beta)$$

Let $\epsilon > 0$. Since $f \in \mathcal{R}_\alpha[a,b]$, there exists a P such that $U(P,f,\alpha) - L(P,f,\alpha) < \epsilon$ by Theorem 6.6. For this partition, we have that $U(Q,g,\beta) - L(Q,g,\beta) < \epsilon$ for the corresponding partition Q. Hence, $g \in \mathcal{R}_\beta[A,B]$. Finally, we have that:

$$\int_A^B g d\beta = \inf_Q U(Q,g,\beta) = \inf_P U(P,f,\alpha) = \int_a^b f d\alpha$$

□

6.4 The Fundamental Theorem of Calculus

Theorem 6.20: Fundamental Theorem of Calculus I

Let $f \in \mathcal{R}[a,b]$ and for $x \in [a,b]$, define $F(x) = \int_a^x f(t)dt$. Then, F is continuous on $[a,b]$. If f is continuous at $x_0 \in [a,b]$ then $F'(x_0)$ exists and $F'(x_0) = f(x_0)$.

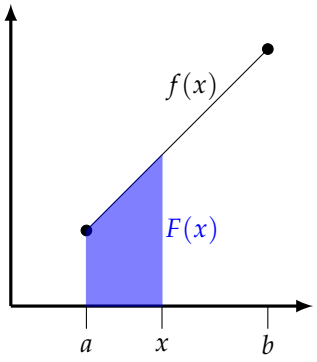

Figure 35: The Fundamental Theorem of Calculus/Theorem 6.20 gives us a way to relate a curve (f) with the culmulative area beneath it (F) by means of the derivative.

To get intuition for Theorem 6.20, we think about what happens when we "zoom in" to a function f. Over an interval (x_0, x_0+h), if x_0 is continuous at x_0 and h is small, then f will be roughly constant over the interval. Hence, $F(x_0+h) - F(x_0) \approx f(x_0)h$. In the limit of $h \to 0$, this approximation becomes exact.

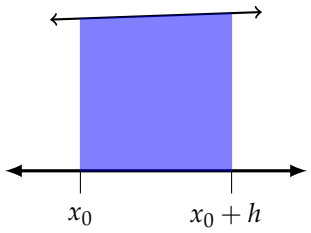

Figure 36: Visual Intuition for Theorem 6.20 and its proof. As f is continuous at x_0, for small h, f is roughly constant on $(x_0, x_0 + h)$. Hence, the increase in culmulative area below the curve is roughly $f(x_0) \cdot h$ (height of $f(x_0)$ times width h of the approximate rectangle).

Proof

We first show the continuity of F. Choose M such that $\left|f(t)\right| \leq M$ for all $t \in [a, b]$. For $a \leq x < y \leq b$, we then have that:

$$\left|F(x) - F(y)\right| = \left|\int_a^x f(t)dt - \int_a^y f(t)dt\right| = \left|\int_x^y f(t)dt\right| \leq M(y - x)$$

Where we use Theorem 6.12(c) for the second equality and Theorem 6.12(d) for the inequality at the end. Let $\epsilon > 0$. If $\left|x - y\right| < \delta = \frac{\epsilon}{M}$, then $\left|F(x) - F(y)\right| < \epsilon$ so F is continuous.
We next show the differentiability of F.We wish to show that:

$$\lim_{h \to 0} \frac{F(x_0 + h) - F(x_0)}{h} = f(x_0)$$

We show the case for $h > 0$, that is, taking the limit of $h \to 0^+$. We have that:

$$\frac{1}{h}\left[F(x_0 + h) - F(x_0)\right] - f(x_0) = \frac{1}{h}\int_{x_0}^{x_0+h} f(t)dt - f(x_0) = \frac{1}{h}\int_{x_0}^{x_0+h} \left[f(t) - f(x_0)\right] dt$$

Where in the last line we use the observation that $\frac{1}{h}\int_{x_0}^{x_0+h} f(x_0)dt = f(x_0)$. Since f is continuous at x_0, for any $\epsilon > 0$, there exists $\delta > 0$ such that:

$$|t - x_0| < \delta \implies \left|f(t) - f(x_0)\right| < \epsilon$$

Thus, if $h < \delta$, then:

$$\left|\frac{1}{h}\left[F(x_0 + h) - F(x_0)\right] - f(x_0)\right| \leq \frac{1}{h}\int_{x_0}^{x_0+h} \left|f(t) - f(x_0)\right|dt < \frac{1}{h}\epsilon h = \epsilon$$

So we conclude that:

$$\lim_{h \to 0^+} \frac{F(x_0 + h) - F(x_0)}{h} = F'(x_0) = f(x_0)$$

□

Theorem 6.21: Fundamental Theorem of Calculus II

If $f \in \mathcal{R}[a,b]$ and there exists F on $[a,b]$ such that $F' = f$, then:

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

Proof

For any partition P, we have that

$$\begin{aligned} F(b) - F(a) &= \sum_{i=1}^{n} \left[F(x_i) - F(x_{i-1})\right] \\ &= \sum_{i=1}^{n} F'(t_i)\Delta x_i \\ &= \sum_{i=1}^{n} f(t_i)\Delta x_i \end{aligned}$$

Where the first equality follows by the fact that the sum telescopes, the second equality follows from the Mean Value Theorem/Theorem 5.10 (By the continuity of F, there exists $t_i \in [x_{i-1}, x_i]$ such that $F(x_i) - F(x_{i-1}) = F'(t_i)(x_i - x_{i-1})$) and the third equality follows by Theorem 6.20. Now, we ahve that $f(t_i) \in [m_i, M_i]$ for each i, so we therefore have that:

$$\sum_{i=1}^{n} m_i\Delta x_i \leq \sum_{i=1}^{n} f(t_i)\Delta x_i \leq \sum_{i=1}^{n} M_i\Delta x_i$$

Hence:

$$F(b) - F(a) \in [L(P,f), U(P,f)]$$

Additionally, by definition of the integral we have that:

$$\int_a^b f(x)dx \in [L(P,f), U(P,f)]$$

Let $\epsilon > 0$. Choose P such that $U(P,f) - L(P,f) < \epsilon$. Then, we have that:

$$\left|F(b) - F(a) - \int_a^b f(x)dx\right| < \epsilon$$

Since ϵ is arbitrary, we have that:

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

□

Theorem 6.22: Integration By Parts

If F, G are differentiable on $[a,b]$ with $F' = f \in \mathcal{R}[a,b]$ and $G' = g \in \mathcal{R}[a,b]$, then:

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx$$

Note that the above integration by parts formula generalizes to different α (not just $\alpha(x) = x$. The proof of this is left as an exercise (Rudin Chapter 6 Problem 17).

Proof

Let $H(x) = F(x)G(x)$. Then, $H'(x) = F'(x)G(x) + F(x)G'(x) = f(x)G(x) + F(x)g(x) \in \mathcal{R}[a,b]$ by the product rule (Theorem 5.3) and the FTC I (Theorem 6.20). Therefore, we have that:

$$H(b) - H(a) = \int_a^b H'(x)dx = \int_a^b f(x)G(x)dx + \int_a^b F(x)g(x)dx$$

Where the first equality follows by FTC II (Theorem 6.21). We have that $H(b) - H(a) = F(b)G(b) - F(a)G(a)$, so rearranging the above expression, we find that:

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx$$

Which is the desired expression. □

Definition: Improper Integrals

If $f \in \mathcal{R}[a,b]$ for all $b > a$, then we define:

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

if the integral exists in $\mathbb{R}$. Then, we say that the **improper integral** $\int_a^\infty f(x)dx$ converges.

Definition: Absolute Convergence of Integrals

If $\int_a^\infty |f(x)|dx$ converges, then we say that $\int_a^\infty f(x)dx$ **converges absolutely**.

To finish off this chapter, we will work through a comprehensive problem together; we prove that $\int_0^\infty \sin t^2 dt$ converges but not absolutely.

Proof. The proof that the integral does not converge absolutely is left as a homework exercise (HW3Q1). We will prove that the integral converges here. Let $p_n = \int_0^n \sin t^2 dt$ where $n \in \mathbb{N}$. We will show that:

(i) $\{p_n\}$ is a Cauchy sequence and hence has a limit in $\mathbb{R}$.

(ii) This is enough to show that the improper integral converges.

We start by showing (i), namely that $\{p_n\}$ is Cauchy. For $x < y$, we have that:

$$\int_x^y \sin t^2 dt = \int_{x^2}^{y^2} \sin u \frac{1}{2\sqrt{u}} du$$

by a change of variable (Theorem 6.19). Now performing an integration by parts (Theorem 6.22) with $F(u) = \frac{1}{2\sqrt{u}}$ and $G'(u)du = \sin u du$, we have that $F'(u) = -\frac{1}{4u^{3/2}}$ and $G(u) = -\cos u$, so:

$$\int_x^y \sin t^2 dt = \int_{x^2}^{y^2} \sin u \frac{1}{2\sqrt{u}} du = -\frac{\cos u}{2\sqrt{u}}\bigg|_{x^2}^{y^2} - \int_{x^2}^{y^2} \frac{\cos u}{4u^{3/2}} du$$

Hence:

$$\int_x^y \sin t^2 dt = \frac{\cos x^2}{2x} - \frac{\cos y^2}{2y} - \int_{x^2}^{y^2} \frac{\cos u}{4u^{3/2}} du$$

Suppose $n \geq m$. Then, we have that:

$$\begin{aligned} \left|p_n - p_m\right| &= \left|\int_m^n \sin t^2 dt\right| \\ &= \left|\frac{\cos m^2}{2m} - \frac{\cos n^2}{2n} - \int_{m^2}^{n^2} \frac{\cos u}{4u^{3/2}} du\right| \end{aligned}$$

Using the fact that $\left|\cos(x)\right| \leq 1$ and applying the Triangle inequality, we have that:

$$\left|p_n - p_m\right| \leq \frac{1}{2m} + \frac{1}{2n} - \int_{m^2}^{n^2} \frac{1}{4u^{3/2}} du = \frac{1}{2m} + \frac{1}{2n} + \frac{-1}{2u^{1/2}}\bigg|_{m^2}^{n^2} = \frac{1}{2m} + \frac{1}{2n} - \frac{1}{2n} + \frac{1}{2m} = \frac{1}{m} \quad (*)$$

Let $\epsilon > 0$. Choose N_0 such that $\frac{1}{N_0} < \epsilon$. Then, for $m \geq n \geq N_0$ we have that:

$$\left|p_n - p_m\right| \leq \frac{1}{m} \leq \frac{1}{N_0} < \epsilon$$

and we conclude that $\left\{p_n\right\}$ is Cauchy. Since $\mathbb{R}$ is complete, $p_n \to p$ for some $p \in \mathbb{R}$. To finish the proof, we show (ii); that this is sufficient. For $b \geq N_0$, choose $N \geq N_0$ such that $b \in [N, N+1)$. Then, we have that:

$$\int_0^b \sin t^2 dt = p_N + \int_N^b \sin t^2 dt \leq p_N + \frac{1}{N}$$

Where the inequality follows by $(*)$. We therefore have that:

$$\left|p - \int_0^b \sin t^2 dt\right| \leq \left|p - p_N\right| + \left|\int_N^b \sin t^2 dt\right| < \epsilon + \frac{1}{N} \leq \epsilon + \frac{1}{N_0} < \epsilon + \epsilon = 2\epsilon$$

Since ϵ is arbitrary, we conclude that $\lim_{b\to\infty} \int_0^b \sin t^2 dt = p$ and hence the improper integral converges. $\square$

7 Sequences and Series of Functions

7.1 Motivating Examples

Example

For $m, n \in \mathbb{N}$, let $p_{n,m} = \frac{m}{n}$. Then,

$$\lim_{m\to\infty} p_{m,n} = \infty, \quad \lim_{n\to\infty} p_{m,n} = 0$$

In particular,

$$\lim_{m\to\infty}\lim_{n\to\infty} p_{m,n} = 0, \quad \lim_{n\to\infty}\lim_{m\to\infty} p_{m,n} = \infty.$$

Which demonstrates that the order of which limits are taken in can affect the value.

Example

Define the sequence of functions:

$$f_n(x) = \begin{cases} 1 & x \geq 0 \\ 1 + nx & -\frac{1}{n} < x < 0 \\ 0 & x \leq -\frac{1}{n} \end{cases}$$

Since f_n is piecewise linear, it is continuous. However, looking at the $n \to \infty$ limit, we have:

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Which is the right continuous step function, which is evidently discontinuous at $x = 0$. Hence, the limit of continuous functions can be discontinuous. Another way of viewing this problem is:

$$\lim_{n\to\infty}\lim_{x\to 0} f_n(x) = 0, \quad \lim_{x\to 0}\lim_{n\to\infty} f_n(x) = \text{D.N.E.}$$

so again we see the order of taking our limits can be important.

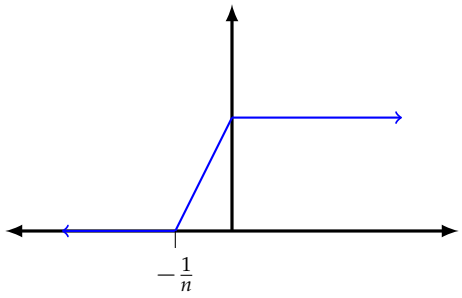

Figure 37: Plot of f_n in the above example.

Example 7.4

For $m \in \mathbb{N}$ and $x \in \mathbb{R}$, let $f_m(x) = \lim_{n\to\infty} \left[\cos(m!\pi x)\right]^{2n}$. Since $\left|\cos(k\pi)\right| = 1$ if $k \in \mathbb{Z}$, we see that $f_m(x) = 1$ when $m!x \in \mathbb{Z}$. Conversely, since $\left|\cos(k\pi)\right| < 1$ if $k \neq \mathbb{Z}$, $f_m(x) = 0$ when $m!x \notin \mathbb{Z}$. Some plots of $f_m(x)$ on $[0,1]$ for $m = 1,2,3$ are below as a visualization. We now define $f(x) = \lim_{m\to\infty} f_m(x)$. If $x = \frac{p}{q} \in \mathbb{Q}$, then $m!x = \frac{m!p}{q} \in \mathbb{Z}$ for m large enough (for $m \geq q$, as the denominator cancells). Therefore, we have that $f(x) = 1$ for $x \in \mathbb{Q}$. Conversely, if $x \notin \mathbb{Q}$, then $m!x \notin \mathbb{Z}$ for all $m \in \mathbb{N}$. So, $f_m(x) = 0$ for all m, and $f(x) = 0$. Therefore, we have that:

$$f(x) = \lim_{m\to\infty} f_m(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

In other words, f is the Dirchlet function. The interesting part is that each of the $f_m(x)$ are Riemann integrable on $[0,1]$ by Theorem 6.10 (as f has finitely many discontinuities for any $m \in \mathbb{N}$). However, the limit is not Riemann integrable, as we prove below. Hence, the limit of Riemann integrable functions is not necessarily Riemann integrable.

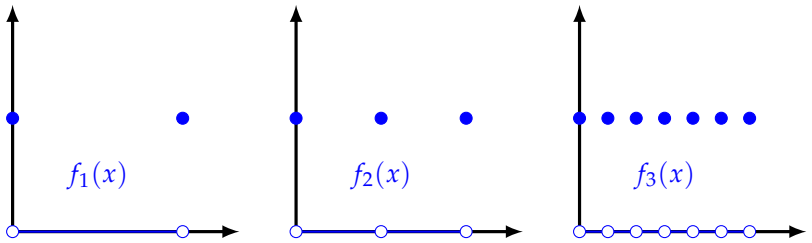

Figure 38: Plot of $f_m(x)$ over the interval $[0,1]$ for $m = 1,2,3$. For $m = 1$, only $x = 0,1$ satisfy $m!x = x \in \mathbb{Z}$. For $m = 2$, we have that $x = 0, \frac{1}{2}, 1$ satisfy $m!x = 2x \in \mathbb{Z}$. Finally, for $m = 3$, we have that $x = 0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1$ satisfy $m!x = 6x \in \mathbb{Z}$.

We now show that f defined in the above example is not Riemann integrable on $[0,1]$.

Proof. Consider any partition P of $[0,1]$. Due to the density of rational and irrational numbers in $\mathbb{R}$ (Theorem 1.20) we have that $M_i = \sup f(x) : x \in [x_{i-1}, x_i] = 1$ and $m_i = \inf f(x) : x \in [x_{i-1}, x_i] = 0$ for all i. Therefore, we have that $U(P,f) = \sum_{i=1}^{N} M_i \Delta x_i = 1$ and $L(P,f) = \sum_{i=1}^{N} m_i \Delta x_i = 0$ for all partitions P. Therefore, $\sup_P U(P,f) = 1$ and $\inf_P L(P,f) = 0$, and we conclude that f is not Riemann integrable on $[0,1]$. □

Example

Define f_n such that:

$$f_n(x) = \begin{cases} 0 & |x| \geq \frac{1}{n} \\ n(nx+1) & -\frac{1}{n} < x < 0 \\ -n(nx+1) & 0 < x < \frac{1}{n} \\ 0 & x = 0 \end{cases}$$

Then, we have that $f(x) = \lim_{n\to\infty} f_n(x) = 0$ for all x. Furthermore, we have that $\int_{-1}^{1} f_n(x)dx = 1$ for all n, but $\int_{-1}^{1} f(x)dx = 0$. Hence, we have that:

$$\lim_{n\to\infty} \int_{-1}^{1} f_n(x)dx = 1 \neq 0 = \int_{-1}^{1} \lim_{n\to\infty} f_n(x)dx$$

showing that problems can arise when we interchange the order of an integral with a limit.

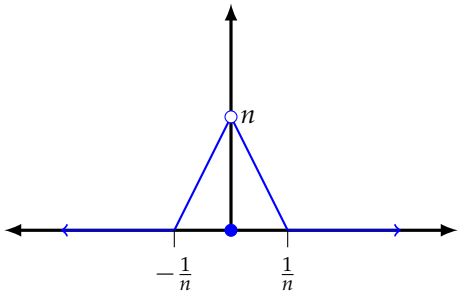

Figure 39: Plot of f_n in the above example.

Example 7.5

Let $f_n(x) = \frac{\sin nx}{\sqrt{n}}$ for $n \in \mathbb{N}, x \in \mathbb{R}$. Then, let $f(x) = \lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$, so $f'(x) = 0$. However, $f_n'(x) = \frac{1}{\sqrt{n}} n \cos nx = \sqrt{n}\cos nx$ and $\lim_{n\to\infty} \sqrt{n}\cos nx$ does not exist. For example, $f_n'(\pi) = \sqrt{n}(-1)^n$ which is a divergent sequence. So:

$$f'(\pi) = \left(\lim_{n\to\infty} f_n\right)'(\pi) = 0 \neq \lim_{n\to\infty} f_n'(\pi)$$

whcih shows us that problems can arise when interchanging a derivative (which is just a type of limit) with a limit.

With the above five examples, we have seen examples of bad behaviour that can occur under interchange of limits. Namely:

1. An interchange of the order of limits can change the limiting value for a double sequence.
2. The limit of a sequence of continuous functions is not necessarily continuous.
3. The limit of a sequence of Riemann integrable functions is not necessarily Riemann integrable.

4. The limit of a sequence of Riemann integrals can differ from the Riemann integral of the limit of a sequence.

5. The limit of a sequence of derivatives can differ from the derivative of a limit of a sequence.

The good news is that in all of these examples, the sequences we looked at had a "weak" form of convergence, where we fix x and then take the $n \to \infty$ limit. We will now proceed to look at a stronger version of convergence, which looks at "all x at once", ensuring that this bad behaviour does not (for the most part) occur.

7.2 Uniform Convergence

Definition 7.7: Uniform Convergence

Let E be any set and $f_n : E \mapsto \mathbb{R}$ or $f_n : E \mapsto \mathbb{C}$ for $n \in \mathbb{N}$. Then, f_n **converges uniformly** to f on E if for all $\epsilon > 0$, there exists N such that $n \geq N$ implies that $\left|f_n(x) - f(x)\right| < \epsilon$ for all $x \in E$.

Note the lack of x dependence in the above definition. We give a useful visual intuition of uniform convergence below:

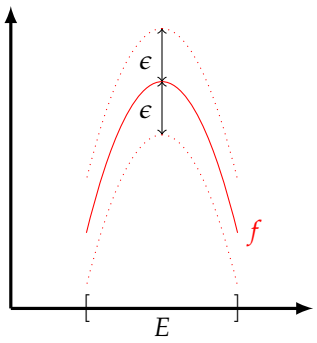

Figure 40: Visualization of the intuition behind uniform convergence. If $f_n \to f$, uniformly, for any $\epsilon > 0$, we can find N such that for $n \geq N$, $f_n(x)$ lies in the ϵ-tube (pictured above) around f.

Example

Let us return to Example 7.5. We have that:

$$\left|f_n(x) - f(x)\right| = \left|\frac{\sin nx}{\sqrt{n}} - 0\right| \leq \frac{1}{\sqrt{n}}$$

So taking n large enough such that $\frac{1}{\sqrt{n}} < \epsilon$, we can see that $f_n(x)$ converges uniformly to $f(x) = 0$. Note that this example does show that uniform convergence is *not* sufficient for:

$$\lim_{n\to\infty} f_n' = \left(\lim_{n\to\infty} f_n\right)'$$

to hold. We will return to the relation of uniform convergence and differentiation in a later theorem.

Example

Let us return to our second example from our section on motivating examples. Recall we had:

$$f_n(x) = \begin{cases} 1 & x \geq 0 \\ 1+nx & -\frac{1}{n} < x < 0 \\ 0 & x \leq -\frac{1}{n} \end{cases} \quad f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We then have that:

$$f_n(x) - f(x) = \begin{cases} 1+nx & -\frac{1}{n} < x < 0 \\ 0 & \text{otherwise} \end{cases}$$

So for $x = -\frac{1}{2n}$, we have that:

$$f_n\left(-\frac{1}{2n}\right) - f\left(-\frac{1}{2n}\right) = 1 + n\left(-\frac{1}{2n}\right) - 0 = \frac{1}{2}$$

Which will never be less than ϵ for $\epsilon < \frac{1}{2}$. Hence, we conclude that f_n does not converge uniformly to f on $\mathbb{R}$.

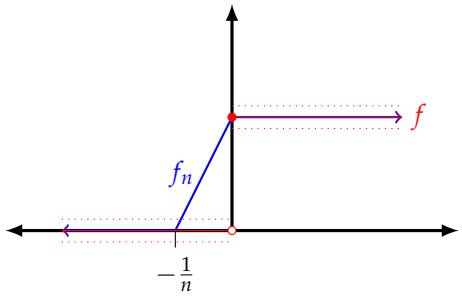

Figure 41: Visualization of why the convergence of $f_n \to f$ in the above example is not uniform. We can see that if we draw a small enough ϵ tube (i.e. $\epsilon \leq 1$), there is no way to choose n large enough to make all of $f_n(x)$ lie in the tube.

Theorem 7.8: Cauchy Criterion for Uniform Convergence

f_n converges uniformly on E if and only if for all $\epsilon > 0$, thre exists N such that if $m, n \geq N$, then $\left|f_m(x) - f_n(x)\right| < \epsilon$ for all $x \in E$.

Again, note the lack of x dependence in the above theorem.

Proof

$\boxed{\implies}$ Suppose $f_n \to f$ uniformly on E. Then, there exists some N such that for $m, n \geq N$:

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

for all $x \in E$. Therefore by the triangle inequality, we have that:

$$|f_m(x) - f_n(x)| \leq |f_m(x) - f(x)| + |f(x) - f_n(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Hence $|f_m(x) - f_n(x)| < \epsilon$ for all $x \in E$.
$\boxed{\impliedby}$ Let $x \in E$. By assumption, $\{f_n(x)\}_{n \in \mathbb{N}}$ is a Cauchy sequence, and hence has a limit $f(x)$ (as both $\mathbb{R}$ and $\mathbb{C}$, the possible codomains of f, are complete). We then let $f(x) = \lim_{n\to\infty} f_n(x)$, so we have pointwise convergence. To see that the convergence is uniform, let $\epsilon > 0$. We know that $|f_m(x) - f_n(x)| < \epsilon$ for $m, n \geq N$ and for all x. Then, let $m \to \infty$. Then, $|f(x) - f_n(x)| \leq \epsilon$. for all $n \geq N$ and all $x \in E$, so the convergence is uniform. □

Theorem 7.9

Suppose $\lim_{n\to\infty} f_n(x) = f(x)$ for $x \in E$, and let:

$$M_n = \sup_{x \in E} |f_n(x) - f(x)|$$

Then, $f_n \to f$ uniformly on E if and only if $M_n \to 0$ as $n \to \infty$.

Proof

$\boxed{\implies}$ suppose $f_n \to f$ uniformly. Then, for any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ and all $x \in E$:

$$|f_n(x) - f(x)| < \epsilon$$

Since this holds for all $x \in E$, taking the supremum of $|f_n(x) - f(x)|$ we have that:

$$\sup_{x \in E} |f_n(x) - f(x)| = M_n \leq \epsilon$$

We then have that $M_n \leq \epsilon$ for $n \geq N$ for some N, and hence $M_n \to 0$.
$\boxed{\impliedby}$ Suppose that $M_n \to 0$. Then, for any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $n \geq N$:

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

We then have that for any $x \in E$:

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

so we conclude that $f_n \to f$ uniformly. □

Definition: Uniform Convergence of Series

We say that $\sum_{n=1}^{\infty} f_n(x)$ **converges uniformly** on E if $S_n(x) = \sum_{i=1}^{n} f_i(x)$ is a uniformly convergent sequence of functions.

Theorem 7.10: Weierstrauss M-Test

Suppose $\left|f_n(x)\right| < M_n$ for all $n \geq N_0$ and for all $x \in E$. Suppose also that $\sum_{n=N_0}^{\infty} M_n < \infty$. Then, $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on E.

Proof

Let $S_n(x) = \sum_{i=1}^{n} f_i(x)$. For $n > m \geq N_0$, we have that:

$$\left|S_n(x) - S_m(x)\right| = \left|\sum_{i=m+1}^{n} f_i(x)\right| \leq \sum_{i=m+1}^{n} \left|f_i(x)\right| \leq \sum_{i=m+1}^{n} M_i$$

Let $\epsilon > 0$. Choose $N \geq N_0$ such that $\sum_{i=N+1}^{\infty} M_i < \epsilon$ (which we can choose as the series converges by assumption). We then have that $\left|S_n(x) - S_m(x)\right| < \epsilon$ for all $n > m \geq N$ for all $x \in E$. Hence, $S_n(x)$ converges uniformly on E. □

Theorem 7.11

Let $E \subset X$ and $f_n : E \mapsto \mathbb{R}$ or $\mathbb{C}$, $n \in \mathbb{N}$. Suppose $f_n \to f$ uniformly on E, and let $x \in E$ (where x is a limit point of E). Suppose $\lim_{t \to x} f_n(t) = A_n$ exists for each $n \in \mathbb{N}$. Then, $A_n \to A$ for some A And $\lim_{t \to x} f(t) = A$. In other words:

$$\lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{n \to \infty} \lim_{t \to x} f_n(t)$$

showing that the interchange of limits is valid when we have uniform convergence.

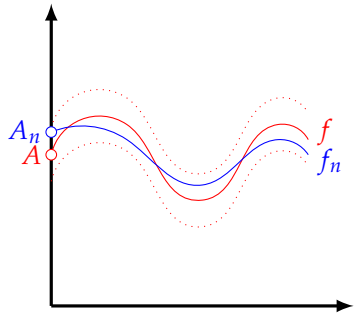

Figure 42: Visualization of Theorem 7.11, with $E = (0, \infty)$ and $x = 0$. Eventually, the graph of f_n lies in the ϵ tube around f (no matter how skinny the tube is). But, A_n is being determined by f_n near 0, so there is nowhere for A_n to go except to the limiting value. That is, $A_n \to A$ as the ϵ tube gets compressed.

Proof

We first show that $A_n \to A$ for some A. Since $\mathbb{R}, \mathbb{C}$ are complete metric spaces, it suffices to show that $\{A_n\}$ is Cauchy. Given $\epsilon > 0$, choose N such that for $m, n \geq N$, $\left|f_n(t) - f_m(t)\right| < \epsilon$ for all t (such an N exists by Theorem 7.8). Letting $t \to x$, we therefore obtain that $|A_n - A_m| \leq \epsilon$ for all $m, n \geq N$, showing that $\{A_n\}$ is Cauchy. Hence, the sequence converges to some limit A.
Now, we show that $\lim_{t \to x} f(t) = A$. We show this by the common "$\epsilon/3$ argument". For all $t \in E$ and $n \in \mathbb{N}$, we have by the triangle inequality that:

$$\left|f(t) - A\right| \leq \left|f(t) - f_n(t)\right| + \left|f_n(t) - A_n\right| + |A_n - A| (*)$$

Which is a good move, as we know that we can make each of the three terms on the RHS arbitrarily small (they are "close"). Let $\epsilon > 0$. Since $f_n \to f$ uniformly, there exists N_1 such that $\left|f(t) - f_n(t)\right| < \frac{\epsilon}{3}$ for all $n \geq N_1$ and all $t \in E$. Since $A_n \to A$, there exists some N_2 such that $|A_n - A| < \frac{\epsilon}{3}$ for all $n \geq N_2$. Letting $N = \max\{N_1, N_2\}$ and taking $n = N$ in $(*)$, we have that:

$$\left|f(t) - A\right| < \frac{\epsilon}{3} + \left|f_N(t) - A_N\right| + \frac{\epsilon}{3}$$

Since $\lim_{t \to x} f_N(t) = A_N$, we can choose $\delta > 0$ such that $t \in N_\delta(x)$ implies $\left|f_N(t) - A_N\right| < \frac{\epsilon}{3}$ (Note a subtle point here that this choice of δ depends on N!). Therefore, if $t \in N_\delta(x)$, we have that:

$$\left|f(t) - A\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

Hence, as $t \to x$, $f(t) \to A$. □

Theorem 7.12

Suppose f_n is continuous on E for all $n \in \mathbb{N}$, and $f_n \to f$ uniformly on E. Then, f is continuous.

Proof

Every f_n is continuous at isolated points of E, so it suffices to consider limit points $x \in E' \cap E$. For these points, we have that:

$$f(x) = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{t \to x} f_n(t) = \lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{t \to x} f(t)$$

Where the third equality (the interchange of the two limits) follows from Theorem 7.11. We conclude that f is continuous by Theorem 4.6 (as $f(x) = \lim_{t \to x} f(t)$). □

Theorem 7.13

Suppose K is compact, and:

(a) f_n is continuous on K for each $n \in \mathbb{N}$

(b) $f_n \to f$ pointwise (that is, for each $x \in K$, $f_n(x) \to f(x)$) and f is continuous

(c) $f_n(x) \geq f_{n+1}(x)$ for all $x \in K$ and all $n \in \mathbb{N}$ (note that the opposite inequality also works, just multiply by -1).

Then, $f_n \to f$ uniformly on K.

Note that this theorem is not super useful, being that it requires so many specific assumptions; however, we will find that it does have an interesting proof. Before we move to that, let us show some counterexamples for when the assumptions do not hold.

Example

Let $K = [-1, 0)$, and define:

$$f_n(x) = \begin{cases} 0 & -1 \leq x \leq -\frac{1}{n} \\ 1 + nx & -\frac{1}{n} < x < 0 \end{cases}$$

We then have that f_n is continuous on K, that $f_n \to 0$ pointwise on K, that f is continuous (the zero function), and f_n is decreasing with n. However, we note that f_n does not converge uniformly to f on K, with points close to zero being problem points (for example, take $x = -\frac{1}{2n}$, and then $f_n(x) - f(x) = \frac{1}{2}$ for all n). We note that K is *not* compact, showing the importance of compactness of the domain in the above Theorem.

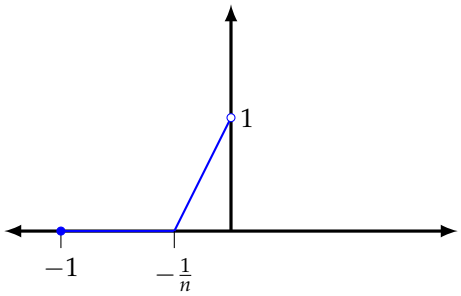

Figure 43: Plot of f_n on $K = [-1, 0)$ from the above example.

Example

Let $K = [0, 1]$, and define:

$$f_n(x) = \begin{cases} 2nx & 0 \leq x < \frac{1}{2n} \\ 2 - 2nx & \frac{1}{n} \leq x \leq \frac{1}{n} \\ 0 & \frac{1}{n} < x \leq 1 \end{cases}$$

We then have that f_n is continuous, $f_n \to f = 0$ pointwise (which is continuous), and K is compact. However, f_n does not converge to f uniformly. In this case, condition (c) of the above Theorem fails; f_n is not monotonic in n.

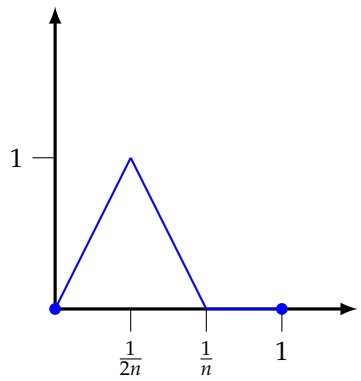

Figure 44: Plot of f_n on $K = [0, 1]$ from the above example.

Proof

Let $g_n = f_n - f$. We can then see that:

(a) g_n is continuous (the difference of two continuous functions is continuous by Theorem 4.9)

(b) $g_n \to 0$ pointwise for all $x \in K$

(c) $g_n \geq g_{n+1} \geq 0$ for all $x \in K$.

The goal will be to show that $g_n \to 0$ uniformly on K. We will use the finite intersection property of compact sets to show this. Let $\epsilon > 0$. We will show that there exists N such that $0 \leq g_n(x) < \epsilon$ for all $n \geq N$ and for all $x \in K$. Note that it suffices to show that $g_N(x) < \epsilon$ for some N, as g is monotone decreasing in n. Define $K_n = g_n^{-1}([\epsilon, \infty))$ (i.e. the set of "bad xs"). We are done if we are able to show that there exists a N with $K_N = \emptyset$. Since g_n is conitnuous, K_n is closed as $[\epsilon, \infty)$ is closed. Since $K_n \subset K$, K is therefore compact as a closed subset of a compact set (Theorem 2.35). Additionally, we have that $K_{n+1} \subset K_n$, as $g_{n+1} \geq \epsilon$ implies that $g_n \geq \epsilon$. Since $g_n \to 0$ pointwise, given $x \in K$, there exsits N_x such that $x \notin K_n$ for all $n \geq N_x$ (as $g_n(x) < \epsilon$ for large enough n). We therefore have that $x \notin \bigcap_n K_n$ for all $x \in K$. Then, applying the corollary to Theorem 2.36, we obtain that K_N is empty. This means that for this N, $g_N^{-1}([\epsilon, \infty)) = \emptyset$, and hence $g_N^{-1}([0, \epsilon]) = K$, which is to say that $0 \leq g_n(x) < \epsilon$ for all $x \in K$. □

Definition 7.14: $\mathcal{C}(X)$ and the Supremum Norm

For a metric space X, define:

$$\mathcal{C}(X) = \{f : X \mapsto \mathbf{C} \text{ such that } f \text{ is bounded and continuous.}\}$$

The **supremum norm** of $f \in \mathcal{C}(X)$ is then defined as $\|f\| = \sup_{x \in X} |f(x)|$. We claim that $\|f - g\|$ defines a metric on $\mathcal{C}(X)$, and we prove this assertion below. Thus, we have that:

$$\begin{aligned} f_n \to f \text{ uniformly} &\iff \forall \epsilon > 0, \exists N \text{ such that } |f_n(x) - f(x)| < \epsilon \ \forall n \geq N \text{ and } \forall x \in X \\ &\iff \forall \epsilon > 0, \exists N \text{ such that } \|f_n(x) - f(x)\| < \epsilon \ \forall n \geq N \\ &\iff f_n \to f \text{ in the metric space } \mathcal{C}(X) \end{aligned}$$

We have hence "metrized" uniform convergence.

Theorem

$\|f-g\|$ defines a metric on $\mathcal{C}(X)$.

Proof

We recall the three properties of a metric as per Definition 2.15:

(a) $d(f,g)=0 \iff f=g$

(b) $d(f,g)=d(g,f)$

(c) $d(f,g)\leq d(f,h)+d(h,g)$

We now show that $\|f-g\|$ satisfies these three properties.

(a) $\|f-g\|=0$ means that $0=\sup_{x\in X}|f(x)-g(x)| \implies |f(x)-g(x)|=0$ for all x, hence $f(x)=g(x)$.

(b) $\|f-g\|=\sup_{x\in X}|f(x)-g(x)|=\sup_{x\in X}|g(x)-f(x)|=\|g-f\|$

(c) We have that $|f(x)-g(x)|\leq |f(x)-h(x)|+|h(x)-g(x)|$ for all $x\in X$, so $\|f-g\|\leq \|f-h\|+\|h-g\|$. □

Note that sometimes $\|f\|$ is written as $\|f\|_\infty$ as it is the $n\to\infty$ limit of the L_p norm. See HW3Q3 for the proof that the supremum norm is the limit of the L_p norm.

Theorem 7.15

$\mathcal{C}(X)$ is a complete metric space (every Cauchy sequence in $\mathcal{C}(X)$ has a limit in $\mathcal{C}(X)$).

Proof

Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{C}(X)$. Then, given $\epsilon>0$, $\exists N$ such that $m,n\geq N$ implies $\|f_m-f_n\|=\sup_{x\in X}|f_m(x)-f_n(x)|<\epsilon$. By the Cauchy criterion (Theorem 7.8), $f_n\to f$ for some f. What is left to show is that $f\in\mathcal{C}(X)$. f is continuous as it is the uniform limit of continuous functions (Theorem 7.12). Additionally, f is bounded as there exists N_0 such that $|f(x)-f_{N_0}(x)|<1$ for all x, and hence $|f(x)|\leq |f_{N_0}(x)|+|f(x)-f_{N_0}(x)|\leq M_0+1$ for all x where M_0 is the bound on $f_{N_0}(x)$ that exists as $f_{N_0}\in\mathcal{C}(X)$. As f is continuous and bounded, we conclude that $f\in\mathcal{C}(X)$. □

7.3 Uniform Convergence and Integration

Theorem 7.16

Suppose $f_n\in\mathcal{R}_\alpha[a,b]$ for all $n\in\mathbb{N}$ and that $f_n\to f$ uniformly on $[a,b]$. Then, $f\in\mathcal{R}_\alpha[a,b]$, and:

$$\lim_{n\to\infty}\int_a^b f_n d\alpha=\int_a^b f d\alpha$$

In other words, the above Theorem tells us that we can interchange the integral with the limit if the sequence is uniformly convergent. Compare this to our earlier example with pointwise convergence, where such an interchange was not possible (as it yielded different values).

Proof

First, we show that $f \in \mathcal{R}_\alpha[a,b]$. Let $\epsilon > 0$. Since $f_n \to f$ uniformly, there exists N such that $|f_n(X) - f(x)| < \epsilon$ if $n \geq N$ for all $x \in [a,b]$. So, $f_n(x) - \epsilon < f(x) < f_n(x) + \epsilon$. Hence,

$$\underline{\int_a^b}(f_n - \epsilon)d\alpha \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq \overline{\int_a^b}(f_n + \epsilon)d\alpha$$

Since $f_n \pm \epsilon \in \mathcal{R}_\alpha[a,b]$, we have that:

$$\int_a^b (f_n - \epsilon)d\alpha \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq \int_a^b (f_n + \epsilon)d\alpha$$

Therefore:

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq \int_a^b 2\epsilon d\alpha \implies \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq 2\epsilon(\alpha(b) - \alpha(a))$$

Since ϵ is arbitrary, we have that $\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$ and hence $f \in \mathcal{R}_\alpha[a,b]$.

Next, we show that $\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b f d\alpha$. To do this, we show that $\left|\int_a^b f d\alpha - \int_a^b f_n d\alpha\right|$ goes to 0 as $n \to \infty$. We have that:

$$\left|\int_a^b f d\alpha - \int_a^b f_n d\alpha\right| = \left|\int_a^b (f - f_n)d\alpha\right| \leq \int_a^b |f - f_n| d\alpha \leq \int_a^b \epsilon d\alpha = \epsilon(\alpha(b) - \alpha(a))$$

Where in the first equality we use Linearity (Theorem 6.12), the first inequality we apply Theorem 6.13, and in the second inequality, we use that for any $\epsilon > 0$, there exists N such that $|f - f_n| < \epsilon$ for $n \geq N$. Since ϵ is arbitrary, we conclude that $\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b f d\alpha$. □

Corollary

If $f_n \in \mathcal{R}_\alpha[a,b]$ and $f(x) = \sum_{n=1}^\infty f_n(x)$ converges uniformly on $[a,b]$, then $\int_a^b f d\alpha = \sum_{n=1}^\infty \int_a^b f_n d\alpha$. That is to say, the infinite series and the integral can be interchanged.

Proof

Let $S_n(x) = \sum_{i=1}^n f_i(x)$. Then, $S_n(x) \to f(x)$ uniformly by assumption, so:

$$\int_a^b f d\alpha = \lim_{n\to\infty} \int_a^b s_n d\alpha = \lim_{n\to\infty} \sum_{i=1}^n \int_a^b f_i d\alpha = \sum_{i=1}^\infty \int_a^b f_i d\alpha$$

Where the first equality follows from the previous theorem, and the second equality follows from the fact that a finite sum and integral can be interchanged by Linearity. □

7.4 Uniform Convergence and Differentiation

Recall Example 7.5, where we looked at the sequence of functions $f_n(x) = \frac{\sin nx}{\sqrt{n}}$. We showed that $f_n \to 0$ uniformly on $\mathbb{R}$, but we found in the example that $f_n'(x)$ does *not* converge. We are therefore motivated to find a condition that if a function converges and is differentiable, then f_n' converges.

As a point of notation, note that for $a < b$ we denote $\int_b^a f d\alpha = -\int_a^b f d\alpha$.

Theorem 7.17

Suppose:

(a) f_n is differentiable on $[a, b]$;

(b) $\exists x_0 \in [a, b]$ such that $f_n(x_0)$ converges as $n \to \infty$;

(c) f_n' converges uniformly on $[a, b]$.

Then, there exists f such that $f_n \to f$ uniformly on $[a, b]$, and:

$$\lim_{n\to\infty} f_n'(x) = f'(x) \ \forall x \in [a, b]$$

A couple remarks before we move to the proof. First, we note that hypothesis (b) seems strange; why would we require convergence f_n at a single point? This has to do with the fact that in differentiating, we lost our constants. For example, let $f_n(x) = n$ as the simplest example. In this case, we have that f_n is differentiable everywhere (with derivative zero everywhere on $[a, b]$) and that gfn' uniformly converges (it is just the sequence of the zero function). However, f_n does not even converge!

Note that we can and will assume that $f_n(x_0) \to 0$ at the specified x_0; if this is not true, we can simply replace $f_n(x)$ by $f_n(x) - f_n(x_0)$.

Proof

The proof of the above theorem is not so trivial. We will therefore prove a weaker theorem. Namely, we add a fourth hypothesis (d) that f_n' is continuous on $[a, b]$. The proof of the stronger/original theorem can be found in Rudin.
First, by (c) there exists a g such that $f_n' \to g$ uniformly on $[a, b]$ (and also on any subinterval of $[a, b]$). Furthermore, by (d) and Theorem 7.12, g is continuous.
Next, applying Theorem 7.16 (to either $[x_0, x]$ or $[x, x_0]$) we have that:

$$\int_{x_0}^{x} f_n'(t) \to \int_{x_0}^{x} g(t)dt = f(x)$$

Then applying the Fundamental theorem of calculus (Theorem 6.21), we have $f_n(x) - f_n(x_0) \to f(x)$ and $f'(x) = g(x)$. But we also assume that $f_n(x_0) \to 0$, so $f_n(x) \to f(x)$ and $f_n'(x) \to g(x) = f'(x)$. So, we have shown pointwise convergence of f_n to f! We have obtained that $\lim_{n\to\infty} f_n'(x) = f'(x)$ for all $x \in [a, b]$. Finally, we show $f_n \to f$ uniformly on $[a, b]$. We have that:

$$\begin{aligned} |f(x) - f_n(x)| = \left| \int_{x_0}^{x} g(t)dt - \int_{x_0}^{x} f_n'(t)dt + f_n(x_0) \right| &\leq \int_{x_0}^{x} \left| g(t) - f_n'(t) \right| dt + |f_n(x_0)| \\ &< \int_{x_0}^{x} \frac{\epsilon}{2(b-a)} dt + \frac{\epsilon}{2} \leq \frac{\epsilon}{2(b-a)}(b-a) + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Where we apply Theorem 6.13 for the first inequality, the fact that $f_n'(t) \to g$ and $f_n(x_0) \to 0$ in the second last inequality, and Theorem 6.12(d) in the last inequality. □

Theorem 7.18

There exists a continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ such that $f'(x)$ does not exist for any $x \in \mathbb{R}$.

The proof of the above theorem will follow by the construction of an "infinitely spiky" real function. Though this might seem like a very pathological counterexample, there are actually many examples of non-differentiable phenomena in mathematics. Looking at the field of probability, we find that brownian motion, brownian maps, and discrete exploration processes (to name a few) all have this property. A visualization of the brownian map, as well as other beautiful probability pictures can be found here `https://secure.math.ubc.ca/Links/Probability/pages/pic_gallery.html`.

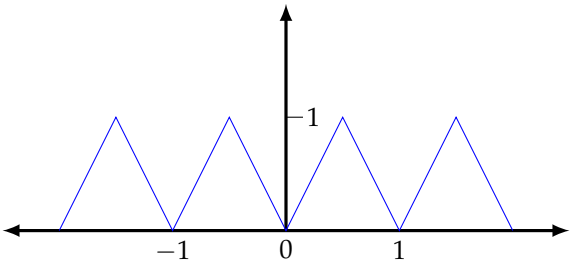

Figure 45: Plot of the ϕ function defined in the proof of Theorem 7.18.

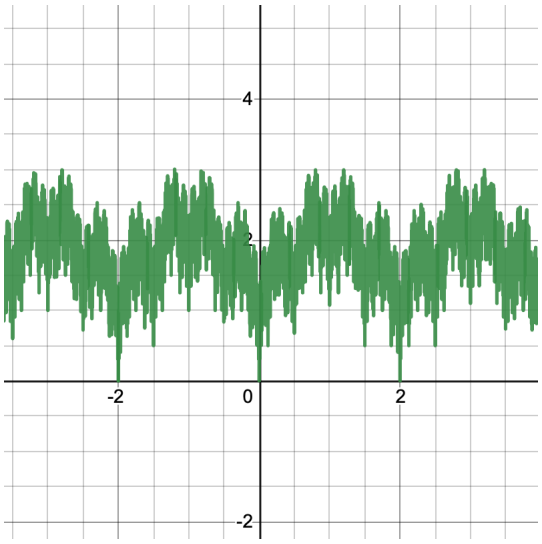

Figure 46: Desmos visualization of the nowhere-differentiable f constructed in the proof of Theorem 7.18. Note that a partial sum $(N = 10)$ of the series that f is defined to be is shown, as the infinite series is impossible to plot. Readers can play around the function with themselves at `https://www.desmos.com/calculator/onhkmblgo6`. There is a notion that f is "infinitely spiky", no matter how much one is to zoom into the above graph.

Proof

Define $\phi : \mathbb{R} \mapsto \mathbb{R}$ by $\phi(x) = |x|$ for $-1 \leq x \leq 1$ and $\phi(x+2) = \phi(x)$ for all $x \in \mathbb{R}$. (See figure 45 above). Then, ϕ is continuous; moreover, it is Lipschiz continuous, with $\left|\phi(s) - \phi(t)\right| \leq |s-t|$ for all $s, t \in \mathbb{R}$ (with equality where there are no integers between s, t). Define $f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \phi(4^n x)$. The series converges uniformly on $\mathbb{R}$ by Theorem 7.10, since $0 \leq \left(\frac{3}{4}\right)^n \phi(4^n x) \left(\frac{3}{4}\right)^n$ and $\sum_n \left(\frac{3}{4}\right)^n$ converges (it is a geometric series with $r < 1$). Hence, f is continuous as it is a uniform limit of a continuous function (Theorem 7.12). We now prove that $f'(x)$ does not exists for any $x \in \mathbb{R}$; let us then fix x. It suffices to find $\delta_m \to 0$ such that:

$$\left|\frac{f(x+\delta_m) - f(x)}{\delta_m}\right| \to \infty \text{ as } m \to \infty.$$

We then choose $\delta_m = \pm\frac{1}{2}\frac{1}{4^m}$. We choose the sign of δ_m depending on the choice of x as follows. At most one of $(4^m x - \frac{1}{2}, 4^m x)$ and $(4^m x, 4^m x + \frac{1}{2})$ contains an integer. We choose the sign such that no integer lies between $4^m x$ and $4^m(x+\delta_m)$. Note that we may choose a differnet sign for each m. Next, we make the observation that $\left|\phi(4^m(x+\delta_m)) - \phi(4^m x)\right| = 4^m x$; this holds as for the difference between two $\phi(x)$ values at two points without an integer between them is just the difference between the x values. Looking back at our definition of δ_m, we then see that $\left|\phi(4^m(x+\delta_m)) - \phi(4^m x)\right| = \frac{1}{2}$.
Furthermore, we see that if $n > m$, we have that $\phi(4^n(x+\delta_m)) - \phi(4^n x \pm \frac{1}{2}4^{n-m}) = \phi(4^n x)$ as $\frac{1}{2}4^{n-m}$ is an even integer and ϕ is 2-periodic. This leads us to conclude that $\phi(4^n(x+\delta_m)) - \delta(4^n x) = 0$ if $n > m$. Given m, then define:

$$\gamma_n = \frac{\phi(4^n(x+\delta_m)) - \delta(4^n x)}{\delta_m}$$

Then, $\gamma_n = 0$ if $n > m$, $|\gamma_m| = \left|\frac{4^m \delta_m}{\delta_m}\right| = |4^m| = 4^m$, and if $0 \leq n < m$, $|\gamma_n| \leq \frac{1}{\delta_m}\left|4^n(x+\delta_m) - 4^n x\right| = \frac{1}{|\delta_m|}|4^n \delta_m| = 4^n$. Finally, we have that:

$$\begin{aligned}
\left|\frac{f(x+\delta_m) - f(x)}{\delta_m}\right| = \left|\sum_{n=0}^{\infty}\left(\frac{3}{4}\right)^n \gamma_m\right| = \left|\sum_{n=0}^{m}\left(\frac{3}{4}\right)^n \gamma_m\right| &\geq \left(\frac{3}{4}\right)^m |\gamma_m| - \sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n |\gamma_n| \\
&\geq \left(\frac{3}{4}\right)^m 4^m - \sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n 4^n \\
&= 3^m - \sum_{n=1}^{m-1} 3^n \\
&= 3^m - \frac{3^m - 1}{3-1} \\
&= \frac{1}{2}(3^m + 1)
\end{aligned}$$

Where the second equality follows as all terms $n > m$ are zero, the first inequality follows by the reverse triangle inequality, the second inequality follows by the bounds on $|\gamma_n|$, and the third-to-last equality is the geometric sum formula. As $m \to \infty$, the difference quotient goes to infinity, and we therefore conclude that f is differentiable nowhere. □

7.5 Equicontinuous Families of Functions

Definition 7.22: Equicontinuity

A family $\mathcal{F}$ of functions on E (that is, a possibly finite, countable, or uncountable set of functions on E) is **equicontinuous** on E if for every $\epsilon > 0$, there exists $\delta > 0$ such that if $f \in \mathcal{F}$ and $x, y \in E$ with $d(x,y) < \delta$, then $\left|f(x) - f(y)\right| < \epsilon$. Note that the functions $f \in \mathcal{F}$ are either real or complex valued.

The above definition of equicontinuity is essentially an even stronger version of uniform continuity; the δ works not just for all x and y for a single f, but for all x and y for all of the fs in $\mathcal{F}$. If $\mathcal{F} = \{f\}$, then this is just uniform continuity.

Theorem

(a) If a family $\mathcal{F}$ of functions on E is equicontinuous, then every $f \in \mathcal{F}$ is uniformly continuous on E.

(b) Any finite family $\mathcal{F} = \{f_1, \dots f_n\}$ of uniformly continuous functions on E is equicontinuous.

Proof

(a) The claim follows immediately from the definition.

(b) Let $\epsilon > 0$. Then, by the uniform continuity of each $f_i \in \mathcal{F}$, there exists δ_i such that if $d(x,y) < \delta_i$, then $\left|f_i(x) - f_i(y)\right| < \epsilon$. Taking $\delta = \min \delta_1, \dots, \delta_n$, we have that for any $f \in \mathcal{F}$, if $d(x,y) < \delta$ then $\left|f(x) - f(y)\right| < \epsilon$. Hence $\mathcal{F}$ is equicontinuous. □

Example

Let $\mathcal{F} = \{f_1, f_2, \dots\}$ with $f_n(x) = \frac{\sin(nx)}{\sqrt{n}}$ for $x \in [0,1] \in E$. Then, $\mathcal{F}$ is equicontinuous.

Proof

We have that $\left|f_n(x) - f_n(y)\right| = \frac{1}{\sqrt{n}}\left|\sin(nx) - \sin(ny)\right| \leq \frac{2}{\sqrt{n}}$ for all $x, y \in E$. Let $\epsilon > 0$. Choose N such that $\frac{2}{\sqrt{n}} < \epsilon$ if $n > N$. Since the remaining f_n (i.e. $\{f_1, \dots, f_n\}$) are a finite collection of uniformly continuous functions (they are uniformly continuous by Theorem 4.19, as they continuous functions on a closed and bounded interval), by the above theorem, $\{f_1, \dots f_n\}$ is equicontinuous. So, there exists a δ such that for $n \geq N$ and $\left|x - y\right| < \delta$, $\left|f_n(x) - f_n(y)\right| < \epsilon$. We then have that for any $n \in \mathbb{N}$ and for any $x, y \in E$ with $\left|x - y\right| < \delta$, then $\left|f_n(x) - f_n(y)\right| < \epsilon$. We conclude that $\mathcal{F}$ is equicontinuous. □

Theorem (Problem 7.16)

Let $\{f_n\}$ be an equicontinuous sequnece of functions such that $f_n : K \mapsto \mathbb{C}$ with K compact. Suppose there is a pointwise limit $f(x) = \lim_{n\to\infty} f_n(x)$ that exists for all $x \in K$. Then, $f_n \to f$ uniformly on K.

Proof

We use an "$\frac{\epsilon}{3}$ argument". Let $\epsilon > 0$. Then, choose $\delta > 0$ such that $\left|f_n(x) - f_n(y)\right| < \frac{\epsilon}{3}$ for all n and for all x, y such that $d(x,y) < \delta$ (such a choice is possible by the equicontinuity of the sequence). Take an open cover of K by considering the set of neighbourhoods of radius δ around every point $x \in K$. Since K is compact, the open cover $\{N_\delta(x) : x \in K\}$ has a finite subcover $\{N_\delta(x_1), \ldots, N_\delta(x_k)\}$. Thus, given $x \in K$, there exists x_j such that $x \in N_\delta(x_j)$ and hence $d(x_j, x) < \delta$. Therefore by the triangle inequality:

$$\begin{aligned}\left|f_n(x) - f_m(x)\right| &\leq \left|f_n(x) - f_n(x_j)\right| + \left|f_n(x_j) - f_m(x_j)\right| + \left|f_m(x_j) - f_m(x)\right| \\ &< \frac{\epsilon}{3} + \left|f_n(x_j) - f_m(x_j)\right| + \frac{\epsilon}{3}\end{aligned}$$

where the last inequality follows from the equicontinuity. For each $i \in \{1, \ldots k\}$, we know that $\left\{f_n(x_j)\right\}$ is a convergent sequence as f_n converges pointwise by assumption. Hence, it is a Cauchy sequence. Therefore, there exists a N_i such that $m, n \geq N_i$ implies $\left|f_n(x_i) - f_m(x_i)\right| < \frac{\epsilon}{3}$. Take $N = \max N_1, \ldots, N_k$. Then, we have that $m, n \geq N$ implies:

$$\left|f_n(x) - f_m(x)\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

So, f_n satisfies the Cauchy criterion for uniform convergence, and hence $\{f_n\}$ converges uniformly on K. □

Theorem 7.24

If $f_n : K \mapsto \mathbb{C}$ is continuous, K is compact, and $f_n \mapsto f$ uniformly on K, then $\{f_n\}$ is equicontinuous.

Proof

We again use an "$\frac{\epsilon}{3}$ argument". Let $\epsilon > 0$. Since $f_n \to f$ uniformly, we have that there exists N such that $m, n \geq N$ implies $\left|f_n(x) - f_m(x)\right| < \frac{\epsilon}{3}$ for all $x \in K$. Also, since K is compact, each f_i is unformly continuous, so $\{f_1, \ldots f_N\}$ is equicontinuous for any $N \in \mathbb{N}$ (as it is a finite set of uniformly continuous functions). Hence, there exists $\delta > 0$ such that $\left|f_i(x) - f_i(y)\right| < \frac{\epsilon}{3}$ if $d(x,y) < \delta$ and $i \leq N$. Finally, for $n > N$, we habe that:

$$\left|f_n(x) - f_n(y)\right| \leq \left|f_n(x) - f_N(x)\right| + \left|f_N(x) - f_N(y)\right| + \left|f_N(y) - f_n(y)\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

where the first/third $\frac{\epsilon}{3}$s come from uniform convergence and the second from the equicontinuity. We conclude that $\{f_n\}$ is equicontinuous. □

Example

We here discuss a set of functions which is not equicontinuous, by returning to a prior example. Let $K = [0,1]$, and define:

$$f_n(x) = \begin{cases} 2nx & 0 \leq x < \frac{1}{2n} \\ 2 - 2nx & \frac{1}{n} \leq x \leq \frac{1}{n} \\ 0 & \frac{1}{n} < x \leq 1 \end{cases}.$$

See Figure 44 for a visualization. We have that $\{f_n\}$ obeys $|f_n(x)| \leq 1$ for all $x \in [0,1]$, but that $\{f_n\}$ is not equicontinuous, as $\left|f_n(\frac{1}{2^n} - f_n(0)\right| = 1 - 0 = 1$ for all n. We can get as close as we like to 0, but the difference will remain large. Also, there is no subsequence of $\{f_n\}$ that can be uniformly convergent on $[0,1]$, as $f_n(x) \to 0$ for all $x \in [0,1]$ pointwise but $f_n(\frac{1}{2n}) = 1$ for all n. $f_n(x)$ "stays far" from the limit. The takeaway message is that a sequence that converges pointwise but is not equicontinuous is not guaranteed to have a uniformly convergent subsequence. This is motivation for the later Theorem 7.25, which gives crtieria for a sequence of functions having a uniformly convergent subsequence.

Definition 7.19: Pointwise/Uniform Bounded Functions

$\{f_n\}$ is **pointwise bounded** on E if there exists a $\phi : E \mapsto \mathbb{R}$ such that $|f_n(x)| < \phi(x)$ for all $x \in E$ and for all $n \in \mathbb{N}$. $\{f_n\}$ is **uniformly bounded** on E if there exists M such that $|f_n(x)| \leq M$ for all $x \in E$ and all $n \in \mathbb{N}$.

Example

Let $f_n(x) = \frac{1}{x} + \frac{1}{n}$. Then, $\{f_n\}$ is pointwise bounded, by (for example) $\phi(x) = \frac{1}{x} + 2$. But, it is not uniformly bounded, as $\frac{1}{x}$ grows arbitrarily large as $x \to 0$.

Theorem 7.23: Selection Theorem

Suppose $f_n : E \mapsto \mathbb{C}$ is pointwise bounded on a countable set E. Then, some subsequence $\{f_{n_k}\}$ of $\{f_n\}$ is pointwise convergent on E; that is to say, $\lim_{k\to\infty} f_{n_k}(x)$ exists for all $x \in E$.

Note that the above theorem plays a large role in probability!

Proof

We invoke a "diagonal argument". Let $E = \{x_1, x_2, \ldots\}$. Consider the sequence $\{f_n(x_1)\}$. We have that it is pointwise bounded by hypothesis, so there exists a subsequence $\left\{f_{1_k}\right\}$ such that $\lim_{n\to\infty} f_{1_n}(x_1)$ converges. (Theorem 2.42). We can apply the same logic for $x_2, x_3, \ldots$ in term, such that the proceeding sequence is a subsequence of the former. In other words, we form the array:

$$\begin{array}{lllll} S_1: & f_{1_1} & f_{1_2} & f_{1_3} & \cdots \\ S_2: & f_{2_1} & f_{2_2} & f_{2_3} & \cdots \\ S_3: & f_{3_1} & f_{3_2} & f_{3_3} & \cdots \\ & \vdots & & & \end{array}$$

In doing so, we have that S_1 converges on x_1, S_2 is a subsequence of S_1 that converges on x_1 and x_2, S_3 is a subsequence of S_2 that converges on x_1 and x_2 and x_3 and so on. Then, we consider the sequence formed by the diagonal of the above array, with $S: f_{1_1}, f_{2_2}, f_{3_3}, \ldots$. This is a subsequence of our original sequence f_n. If we fix some N, then this subsequence is a subsequence of S_n for $n \geq N$ (as S is eventually a subsequence of each S_n), so it converges on the same points that S_n does, namely, $x_1, \ldots, x_n$. But this is true for every $n \in \mathbb{N}$, so the subsequence S converges for $x_i \in E$ for every $i \in \mathbb{N}$. □

Lemma (Problem 2.25)

If K is compact, then K has a countable dense subset $E \subset K$ (i.e. $\overline{E} = E \cup E' = K$). Alternatively, for all $x \in K$, there exists $r > 0$ such that there exists $p \in E$ such that $d(p, x) < r$. In other words, K is separable.

Proof

For $n \in \mathbb{N}$, $\left\{N_{1/n}(p)\right\}_{p\in K}$ is an open cover. So, it has a finite subcover $\left\{N_{1/n}(p)\right\}_{p\in E_n}$ where $E_n \subset K$ is finite. Let $E = \bigcup_{n=1}^{\infty} E_n$, then E is at most countable (Theorem 2.12). To see that it is dense, let $x \in K$, and $r > 0$. Then, choose n_0 such that $\frac{1}{n_0} < r$. We can then find $p_0 \in E_{n_0} \subset E$ such that $x \in N_{1/n_0}(p_0)$ as E_{n_0} is an open cover of K. Then, $d(x, p_0) < \frac{1}{n_0} < r$. Hence, E is dense. □

Theorem 7.25: Arzela-Ascoli

Suppose K is compact, and that $\mathcal{F} = \{f_n\} \subset \mathcal{C}(K)$ is equicontinuous and pointwise bounded. Then,

(a) $\{f_n\}$ is uniformly bounded.

(b) $\{f_n\}$ has a uniformly convergent subsequence (i.e. a subsequence that converges in $\mathcal{C}(K)$).

Proof

(a) The goal is to find M such that $\left|f_n(x)\right| \leq M$ for all $n \in \mathbb{N}$ and for all $x \in K$. Let $\epsilon > 0$ (though we can take $\epsilon = 0$ for this proof of part (a)). Since $\mathcal{F}$ is equicontinuous, we have that there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $\left|f_n(x) - f_n(y)\right| < \epsilon$ for all n. K is compact, we can cover K with balls of radius δ around each point in K and then take a finite subcover; i.e. there exists a finite set $\{p_1, \ldots, p_r\} \in K$ such that $\{N_\delta(p_i)\}_{i=1,\ldots,r}$ covers K. For each i, $\{f_n(p_i)\}_n$ is bounded, that is, $\left|f_n(p_i)\right| \leq M_i$ for all n. Let $M_0 = \max M_1, \ldots, M_n$. Given $x \in K$, choose p_i such that $x \in N_\delta(p_i)$. Then:

$$\left|f_n(x)\right| \leq \left|f_n(p_i)\right| + \left|f_n(x) - f_n(p_i)\right| < M_i + \epsilon \leq M_0 + \epsilon$$

letting $M = M_0 + \epsilon$, we see that $\{f_n\}$ is a uniformly bounded.

(b) Our goal is to construct a uniformly convergent subsequence. We do this in three steps. First, we construct a subsequence (we will show it is uniformly convergent afterwards!). By the above Lemma, K has a countable dense subset E. By Theorem 7.23, there exists a subsequence $\{f_{n_i}\}$ such that $\lim_{i\to\infty} f_{n_i}(x)$ exists for all $x \in E$. Write $g_i = f_{n_i}$.

Secondly, we set up the argument to show uniform convergence of the subsequence constructed in the first step. Let $\epsilon > 0$. By the equicontinuity assumption, there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $\left|g_i(x) - g_i(y)\right| < \frac{\epsilon}{3}$ for all i. Consider $\{N_\delta(p)\}_{p\in E}$, which covers K since E is dense. By the compactness of K, there exists a finite subset $\{N_\delta(x_1), \ldots, N_\delta(x_m)\}$ with $x_i \in E$. Hence, given $x \in K$, there exists x_s such that $d(x, x_s) < \delta$.

For the third step, we complete the proof with an "$\frac{\epsilon}{3}$ argument". Using the triangle inequality, we have that:

$$\begin{aligned}\left|g_i(x) - g_j(x)\right| &\leq \left|g_i(x) - g_i(x_s)\right| + \left|g_i(x_s) - g_j(x_s)\right| + \left|g_j(x_s) - g_j(x)\right| \\ &< \frac{\epsilon}{3} + \left|g_i(x_s) - g_j(x_s)\right| + \frac{\epsilon}{3}\end{aligned}$$

where the last inequality follows from the arguments in step 2. For the second term, we consider that for $s = 1, \ldots, m$, we can choose N_s such that $\left|g_j(x_s) - g_i(x_s)\right| < \frac{\epsilon}{3}$ for $i, j \geq N_s$ (this choice of N_s is possible as $\{g_n(x_s)\}$) converges. There are finitely many N_ss, so let $N = \max N_1, \ldots, N_m$. Then, we have that:

$$\left|g_i(x) - g_j(x)\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

for $i, j \geq N$ and for all $x \in K$. Hence, $\{g_i\}$ converges uniformly on K. □

7.6 The Stone-Weierstrass Theorem

Theorem 7.26: Weierstrass

Let $f : [a,b] \mapsto \mathbb{R}$ be continuous. Then, there exists polynomials P_n such that $P_n \to f$ uniformly on $[a,b]$.

Note that it will suffice to consider the case where $[a,b] = [0,1]$; we can get to arbitrary $[a,b]$ to $[0,1]$ by a change of variable, and the composition of apolynomial with a change of variable is still a polynomial.

Note that our proof will take a different angle from Rudin's proof of the theorem; we shall be exploring the proof by Bernstein. However, before we begin the proof, we will need to establish some basic background in probability.

Definition: Bernoulli Trials

A **Bernoulli trial** is a random experiment with a success outcome of probablity p and a failure outcome with probability $1-p$ (here, $p \in [0,1]$ is fixed). Consider n independent Bernoulli trials (where each experiment does not affect any of the others) and let S_n be the number of successes. Then, we have that:

$$p_m = P(S_n = m) = \binom{n}{m} p^m (1-p)^{n-m} \quad (m = 0, 1, \ldots, n).$$

Also note that:

$$\sum_{m=0}^{n} p_m = [p + (1-p)]^n = 1^n = 1$$

Definition: Random Variables

A **random variable** is a function $X : \{0, 1, \ldots n\} \mapsto \mathbb{R}$.

For example, S_n is the identity function, with $X(m) = S_n(m) = m$.

Definition: Expectation

The **expectation** of a random variable X, denoted EX, is defined as:

$$EX = \sum_{m=0}^{n} X(m) p_m$$

We can interpret the expectation of X as the sum over the values of X, weighted by the likelihoods. As an example, we have that:

$$ES_n = \sum_{m=0}^{n} S_n(m) p_m = \sum_{m=0}^{n} m p_m = \sum_{m=0}^{n} m \binom{n}{m} p^m (1-p)^{n-m} = \ldots = np$$

Definition: Variance & Standard Deviation

The **variance** of a random variable X, denoted $\mathrm{Var}(X)$, is defined as:

$$\mathrm{Var}(X) = E[(X - EX)^2] = E(X^2) - (EX)^2.$$

The **standard deviation**of X, denoted by σ_x, is then defined as $\sigma_x = \sqrt{\mathrm{Var}(X)}$.

For example, $\mathrm{Var}(S_n) = E(S_n^2) - (ES_n)^2 = np(1-p)$, and $\sigma_{S_n} = \sqrt{np(1-p)}$.

Definition: Proportion of Successes

Define X_n to be the **proportion of successes** in n independent Bernoulli trials, with $X_n = \frac{1}{n}S_n$. Then, we have that $EX_n = \frac{1}{n}np = p$, $\text{Var}(X_n) = \text{Var}(\frac{1}{n}S_n) = \frac{1}{n^2}\text{Var}(S_n) = \frac{1}{n}p(1-p)$. We then have that $\sigma_{X_n} = \sqrt{\frac{p(1-p)}{n}}$.

Analyzing the standard deviation σ_{X_n}, we see that as we do more trials (n increases), the fluctuation of the proportion of successes gets smaller. This phenomena is known as the *Law of large numbers*, which states that the proportion of successes should converge to the probability of success in a single trial. $\sigma_{X_n} \to 0$ as $n \to \infty$ tells us this fact.

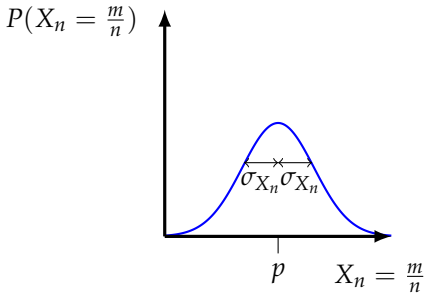

Figure 47: Visualization of how $P(X_n)$. Since σ_{X_n} (the width of the distribution) scales as $\frac{1}{\sqrt{n}}$, as n grows, the distribution becomes more sharply peaked around p.

Theorem: Chebychev's Inequality

$P(|X_n - p| > \delta) \leq \frac{1}{\sigma^2}p(1-p)\frac{1}{n}$. Note that this inequality can be generalized to random variables in general, but here it suffices to consider the inequality just for the case of X_n.

Proof

We have that:

$$P(|X_n - p| > \delta) = \sum_{m:|\frac{m}{n}-p|>\sigma} p_m \leq \sum_{m=0}^{n} \left|\frac{\frac{m}{n}-p}{\sigma}\right|^2 p_m = \frac{1}{\sigma^2}\sum_{m=0}^{n}\left(\frac{m}{n}-p\right)^2 p_m = \frac{1}{\sigma^2}\text{Var}(X_n)$$
$$= \frac{1}{\sigma^2}p(1-p)\frac{1}{n}.$$

where in the first inequality we use the fact that $\left|\frac{\frac{m}{n}-p}{\sigma}\right|^2 \geq 1$ and we hence add non-negative terms to the sum, and in the second to last equality we invoke the definition of the variance. □

With the machinery of basic probability established, we move to the proof of the Weierstrass theorem.

Proof

Take $p = x \in [0,1]$. We then have that $p_m = \binom{n}{m} x^m (1-x)^{n-m}$. Let $P_n(x) = Ef(x_n) = \sum_{m=0}^{n} f(\frac{m}{n}) p_m$ (why? as we take n large, we have that $x_n \to x$, so $f(x_n) \to f(x)$, showing that $P_n(x)$ approximates $f(x)$ well). We note that $\sum_{m=0}^{n} f(\frac{m}{n}) p_m$ is a polynomial in x of degree n. This is our candidate for a uniformly convergent polynomial. We then have that:

$$f(x) - P_n(x) = \sum_{n=0} \left(f(x) - f\left(\frac{m}{n}\right) \right) p_m$$

We will show that this is small by dividing it into two parts. For $\sigma > 0$, we have that:

$$\begin{aligned} \left| f(x) - P_n(x) \right| &\leq \sum_{m:\left|\frac{m}{n} - x\right| \leq \sigma} \left| f(x) - f\left(\frac{m}{n}\right) \right| p_m + \sum_{m:\left|\frac{m}{n} - x\right| > \sigma} \left| f(x) - f\left(\frac{m}{n}\right) \right| p_m \\ &\leq \sum_{m:\left|\frac{m}{n} - x\right| \leq \sigma} \left| f(x) - f\left(\frac{m}{n}\right) \right| p_m + \sum_{m:\left|\frac{m}{n} - x\right| > \sigma} 2M p_m \end{aligned}$$

where $M = \sup\left\{ f(x) : x \in [0,1] \right\}$. Let $\epsilon > 0$. we choose $\delta > 0$ such that $\left|x - y\right| < \delta$ implies $\left|f(x) - f(y)\right| < \frac{\epsilon}{2}$. This choice is possible by the uniform continuity of f (it is a continuous (polynomial) function on a compact set ($[0,1]$)). Then, for the first term above we have that:

$$\sum_{m:\left|\frac{m}{n} - x\right| \leq \sigma} \left| f(x) - f\left(\frac{m}{n}\right) \right| p_m \leq \frac{\epsilon}{2} \sum_{m=0}^{n} p_m = \frac{\epsilon}{2}.$$

For the second term, we apply Chebychev's inequality to get:

$$\sum_{m:\left|\frac{m}{n} - x\right| > \sigma} 2M p_m \leq 2M \frac{1}{\delta^2} \frac{x(1-x)}{n}.$$

Since $x(1-x) \leq \frac{1}{4}$ for $x \in [0,1]$ we have:

$$\sum_{m:\left|\frac{m}{n} - x\right| > \sigma} 2M p_m \leq 2M \frac{1}{\delta^2} \frac{x(1-x)}{n} \leq \frac{M}{2\delta^2} \frac{1}{n}.$$

Now, choose n such that $n > N \geq \frac{4\delta^2}{M\epsilon}$. We then have that:

$$\frac{M}{2\delta^2} \frac{1}{n} < \frac{\epsilon}{2}$$

Then, we have that:

$$\left|f(x) - P_n(x)\right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which proves the claim. □

Our conclusion is that $P_n(x)$ is very close to $f(x)$. In the above proof, we split up the sum into two parts, and used different methods to obtain nice estimates/bounds on each. The next topic we will look at is generalizing this theorem; we will be building up to Rudin 7.32 (Stone-Weierstrass).

Definition 7.28: Algebras

Let $\mathcal{A}$ be a set of functions $f : E \mapsto \mathbb{C}$ (or $\mathbb{R}$). Then, $\mathcal{A}$ is an **algebra** if for all $f, g \in \mathcal{A}$ and for all $c \in \mathbb{C}$, $f + g \in \mathcal{A}$, $fg \in \mathcal{A}$, and $cf \in \mathcal{A}$.

Example

Let $E = [0,1]$ and let $\mathcal{A} = \mathbb{P}$ be the set of polynomials on $[0,1]$. Then, $\mathcal{A}$ is an algebra as the sum and product of two polynomials is also a polynomial, and a constant times a polynomial is a polynomial.

Definition 7.28: Uniformly Closed Algebras and Uniform Closure

We say that an algebra $\mathcal{A}$ is **uniformly closed** if $f_n \in \mathcal{A}$ and if $f_n \to f$ uniformly on E, then $f \in A$. In other words, the uniform limit of sequences of functions in the algebra is contained in the algebra. The **uniform closure** of $\mathcal{A}$ is then defined as $\mathcal{B} = \{f : E \mapsto \mathbb{C} : \exists f_n \in \mathcal{A} \text{ such that } f_n \to f \text{ uniformly}\}$.

Example

$\mathbb{P}$ in the above example is not closed, as the uniform limit of a polynomial is not necessarily a polynomial. $\mathcal{C}([0,1])$ is uniformly closed as the limit of uniform and continuous functions are closed and bounded. $\mathcal{C}([0,1])$ is also the uniform closure of $\mathbb{P}$ by the Weierstrass theorem (Theorem 7.26).

Theorem 7.29

The uniform closure $\mathcal{B}$ (sometimes denoted $\overline{\mathcal{A}}$) of an algebra $\mathcal{A}$ of bounded functions is a uniformly closed algebra. Note that $\mathcal{A}$ has a metric $d(f,g) = \sup_{x \in E} |f(x) - g(x)| = \|f - g\|$ and uniform convergence is equivalent to convergence in this metric.

Proof

Suppose $f, g \in \mathcal{B}$ and $c \in \mathbb{C}$. Then, there exsits $\{f_n\}, \{g_n\} \subset \mathcal{A}$ such that $f_n \to f$ uniformly and $g_n \to g$ uniformly (that is, $\|f - f_n\| \to 0$ and $\|g - g_n\| \to 0$). We then have that:

$$\begin{aligned} f_n + g_n &\to f + g \text{ uniformly,} \\ f n g_n &\to fg \text{ uniformly,} \\ cf_n &\to cf \text{ uniformly.} \end{aligned}$$

Note that the first two lines above correspond to Rudin problems 7.2 and 7.3 respectively (these are left as exercises to the reader). We conclude that $f + g, fg, cg \in \mathcal{B}$ and hence $\mathcal{B}$ is an algebra. Furthermore, it is uniformly closed as it consists of $\mathcal{A}$ and all limit points of $\mathcal{A}$ (hence $\mathcal{B} = \overline{\mathcal{A}}$). $\square$

Definition 7.30: Separating Points and Vanishing at No Point

A set $\mathcal{A}$ consisting of functions $f : E \mapsto \mathbb{C}$ **separates points** on E if for all x_1, x_2 in E with $x_1 \neq x_2$, there exists $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$. In other words, there are enough functions in the set such that whatever pair of points we choose, we can always find distinct function values at these points. We say that $\mathcal{A}$ **vanishes at no point** in E if for all $x \in E$, there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$.

Example

(a) The set of polynomials $\mathbb{P}$ on $[-1, 1]$ separates points and vanishes at no point.

(b) The set of even polynomials on $[-1, 1]$ vanishes at no point but does not separate points (as for any $x \in (0, 1]$ and even polynomial f, $f(x) = f(-x)$).

(c) The set of odd polynomials on $[-1, 1]$ separates points, but all odd polynomials vanish at zero.

Theorem 7.32: Stone-Weierstrass

The uniform closure of any algebra $\mathcal{A}$ of real continuous functions on a compact set K which separates points and vanishes at no point is $\mathcal{C}(K)$ (i.e. the set of all continuous functions on K).

In other words, the above theorem tells us that given $\mathcal{A}$ that separates points and vanishes at no point on compact K, we can generate a sequence that uniformly converges to any continuous function on K. Note that the above theorem gives the Weierstrass theorem as a special case. Take $[a, b]$ and $\mathcal{A} = \mathbb{P}$ to be the polynomials on $[a, b]$. Then, $\mathbb{P}$ separates points and vanishes at no point, so according to the theorem, the uniform closure of $\mathbb{P}$ is all continuous functions on $[a, b]$.

Theorem 7.33: Complex Stone-Weierstrass

Let $\mathcal{A}$ be a set of real complex functions on a compact set K which separates points and vanishes at no point. Furthermore, suppose that $\mathcal{A}$ is self adjoint, that is, if $f \in \mathcal{A}$, then $\overline{f} \in \mathcal{A}$ (where $\overline{f}(x) = \overline{f(x)}$). Then, the uniform closure of $\mathcal{A}$ is $\mathcal{C}(K)$.

We establish three ingridients necessary for our proof of the Stone-Weierstrass theorem.

Lemma 1

Let $\mathcal{A}$ be an algebra of real, continuous functions on a compact set K. Then, if $f \in \overline{\mathcal{A}}$, then $|f| \in \overline{\mathcal{A}}$.

Proof

Let $f \in \overline{\mathcal{A}}$ and $M = \sup_{x \in K} |f(x)|$. This M is finite. Let $\epsilon > 0$. By Theorem 7.26, there exists a polynomial $\tilde{P}_n$ duch that:

$$\sup_{|y| \leq M} \left| \tilde{P}_n(y) - |y| \right| < \frac{\epsilon}{2}.$$

Then, let $P_n(y) = \tilde{P}_n(y) - \tilde{P}_n(0) = \sum_{j=1}^{n} c_j y^j$. We then have that:

$$\left| P_n(y) - |y| \right| \leq \left| \tilde{P}_n(0) - |0| \right| + \left| \tilde{P}_n(y) - |y| \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Also, $P_n(f) = \sum_{j=1}^{n} c_j f^j \in \overline{\mathcal{A}}$ as $f \in \overline{\mathcal{A}}$ and hence sums/products of f will be in the algebra. Note that the constant function may or may not be in the algebra, which is the reason why we define P_n with the constant term left out (by subtracting $\tilde{P}_n(0)$ from $\tilde{P}_n(y)$). Moreover, we have that $\sup_{x \in K} \left| P_n(f)(x) - |f(x)| \right| < \epsilon$ as $\left| P_n(y) - |y| \right| < \epsilon$ for any $|y| \leq M$ (since $|f(x)| \leq M$ for all $x \in K$). Hence, $|f| \in \overline{\mathcal{A}}$. □

Lemma 2

For $\overline{\mathcal{A}}$ as in the previous Lemma (where $\overline{\mathcal{A}}$ is the uniform closure of a set of real, continuous functions on compact K), if $f_1, \ldots, f_n \in \overline{\mathcal{A}}$, then $\max\{f_1, \ldots, f_n\} \in \overline{\mathcal{A}}$ and $\min\{f_1, \ldots, f_n\} \in \overline{\mathcal{A}}$.

Proof

It suffices to consider the case where $n = 2$. For this we use that:

$$\max\{f_1, f_2\} = \frac{1}{2}(f_1 + f_2) + \frac{1}{2}|f_1 - f_2|$$
$$\min\{f_1, f_2\} = \frac{1}{2}(f_1 + f_2) - \frac{1}{2}|f_1 - f_2|$$

Then, as $\overline{\mathcal{A}}$ is an algebra, we have that $\frac{1}{2}(f_1 + f_2) \pm \frac{1}{2}|f_1 - f_2| \in \overline{\mathcal{A}}$ by Lemma 1. □

As a quick verification of the above formulas for the max/min over $\{f_1, f_2\}$, we can WLOG consider the case where $f_1 \geq f_2$:

$$\frac{1}{2}(f_1 + f_2) + \frac{1}{2}|f_1 - f_2| = \frac{1}{2}(f_1 + f_2) + \frac{1}{2}(f_1 - f_2) = f_1 = \max\{f_1, f_2\}$$
$$\frac{1}{2}(f_1 + f_2) - \frac{1}{2}|f_1 - f_2| = \frac{1}{2}(f_1 + f_2) - \frac{1}{2}(f_1 - f_2) = f_2 = \min\{f_1, f_2\}$$

Theorem 7.31

If an algebra $\mathcal{A}$ of functions on E separates points and vanishes at no point, then given any $x_1, x_2 \in E$ with $x_1 \neq x_2$ and constants c_1, c_2, there exists $f \in \mathcal{A}$ such that $f(x_1) = c_1$ and $f(x_2) = c_2$.

Proof

By hypothesis, there exists $g \in A$ such that $g(x_1) \neq g(x_2)$ (as $\mathcal{A}$ separates points). Furthermore, there exists $h, k \in \mathcal{A}$ such that $h(x_1) \neq 0$ and $k(x_2) \neq 0$ as $\mathcal{A}$ vanishes at no point. Let:

$$u(x) = (g(x) - g(x_1))k(x)$$
$$v(x) = (g(x) - g(x_2))h(x)$$

Then, $u(x_1) = 0$ and $u(x_2) \neq 0$, and $v(x_1) \neq 0$ and $v(x_2) = 0$. Note that the k, hs are necessary to include in the above definitions of u, v to ensure that u, v lie in our algebra; $g(x) - g(x_i)$ may not be in $\mathcal{A}$ if the constant functions are not in $\mathcal{A}$. Now, let:

$$f(x) = c_1 \frac{v(x)}{v(x_1)} + c_2 \frac{u(x)}{u(x_2)}$$

which is a meaningful definition as $v(x_1) \neq 0$ and $u(x_2) \neq 0$. Since $\mathcal{A}$ is an algebra, $f \in \mathcal{A}$. Furthermore, we have that:

$$f(x_1) = c_1 \frac{v(x_1)}{v(x_1)} + c_2 \frac{u(x_1)}{v(x_2)} = c_1 + 0 = c_1$$

and identically $f(x_2) = c_2$, proving the claim. □

We now proceed into the proof of Theorem 7.32. As a brief refresher, we have an algebra $\mathcal{A}$ of continuous real functions on a compact set K, which vanishes points and separates at no point. We wish to show that for all continuous functions $f : K \mapsto \mathbb{R}$ and for all $\epsilon > 0$, there exists $h \in \mathcal{B} = \overline{\mathcal{A}}$ such that $\|h - f\| < \epsilon$.

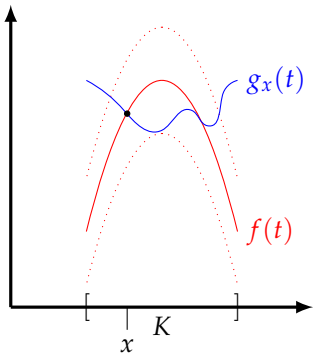

Figure 48: Visualization of Claim 1 in the below proof of the Stone-Weierstrass theorem. We have that $g_x(x) = f(x)$, and that $g_x(t)$ lies above the bottom of the ϵ-tube around f. As we will see in the proof, this g_x is obtained by considering h_ys that satisfy $h_y(t) - f(t) > -\epsilon$ for t in some neighbourhood of y. By the compactness of K, we can consider a finite subcover of these neighbourhoods, and defining g_x to be the maximum of some finite number of h_{y_i}s.

Proof

Claim 1: Let $f : K \mapsto \mathbb{R}$ be continuous, $x \in K$, and $\epsilon > 0$. Then, $\exists g_x \in \mathcal{B}$ such that $g_x(x) = f(x)$ and $g_x(t) - f(t) > -\epsilon$ for all $t \in K$.
We now prove the claim. Fix $x \in K$. Given $y \in K$ with $y \neq x$, by Theorem 7.31 there exists $h_y \in \mathcal{A}$ such that $h_y(x) = f(x)$ and $h_y(y) = f(y)$. $h_y - f$ is continuous, and $h_y(y) - f(y) = 0$, so by continuity there exists an open set $J_y \subset K$ such that $y \in J_y$ and $h_y(t) - f(t) > -\epsilon$ for all $t \in J_y$. We can form an open cover of K from considering the set of J_ys around each $y \in K$. By the compactness of K, there exists a finite subcover $\left\{J_{y_1}, \dots J_{y_n}\right\}$. Let $g_x = \max\left\{h_{y_1}, \dots h_{y_n}\right\}$ (a maximum can be taken over a finite set). Each of the h_{y_i}s are continuous and in $\mathcal{A}$, so $g_x \in \mathcal{B}$ by Lemma 2. Also, we have that:

$$g_x(x) = \max h_{y_1}(x), \dots, h_{y_n}(x) = \max f(x), \dots, f(x) = f(x)$$

as well as that $g_x(t) - f(t) \geq h_{y_i}(t) - f(t) > -\epsilon$ where we choose i such that $t \in J_{y_i}$. This proves the claim.
Claim 2: Let $f : K \mapsto \mathbb{R}$ be continuous, and let $\epsilon > 0$. Then, there exists $h \in \mathcal{B}$ such that $\sup_{x \in K} \left|h(x) - f(x)\right| < \epsilon$. This claim implies the Stone-Weierstrass theorem. We will be "fixing" the function from claim 1 such that the function does not lie above the ϵ tube.
Let us move onto the proof of the claim. Since $g_x - f$ is continuous and $g_x(x) - f(x) = 0$, we ahve that there exists an open set $V_x \in K$ such that $\left|g_x(t) - f(t)\right| < \epsilon$ for $t \in V_x$. Since K is compact, we have that $K \subset V_{x_1} \cup \ldots \cup V_{x_n}$ for some $x_1, \dots, x_n \in K$. Let $h = \min\left\{g_{x_1}, \dots, g_{x_n}\right\}$. By Lemma 2, $h \in \mathcal{B}$. Then, take any $t \in K$. We then have that:

$$h(t) - f(t) = g_{x_i} - f(t) > -\epsilon$$

where we pick i to give the minimum, and the lower bound of $-\epsilon$ follows by Claim 1. We also have that:

$$h(t) - f(t) \leq g_{x_{i'}}(t) - f(t) < \epsilon$$

where we pick a new i' such that $t \in V_{x_{i'}}$. We conclude that $\left\|h - f\right\| < \epsilon$, proving the claim. □

Having proven the real case of the Stone-Weierstrass theorem, we now move to the proof of the complex generalization (Theorem 7.33). Recall that we add the hypothesis that $\mathcal{A}$ is self-adjoint in this generalization (if $f \in \mathcal{A}$, then $\overline{f} \in \mathcal{A}$). As a recap of the statement of the theorem, we suppose that $\mathcal{A}$ is a self-adjoint algebra of complex continuous functions on a compact set K that separates points and vanishes at no point. Then, $\mathcal{B} = \overline{\mathcal{A}} = \mathcal{C}(K)$. In other words, for any $f \in \mathcal{C}(K)$ we can find an element in $\overline{\mathcal{A}}$ such that the difference is arbitrarily small.

Proof

Let $\mathcal{A}_{\mathbb{R}}$ be the algebra of real-valued continuous functions contained in $\mathcal{A}$. If $f = u + iv \in \mathcal{A}$, then $u = \frac{1}{2}(f + \overline{f}) \in \mathcal{A}_{\mathbb{R}}$ and $v = \frac{1}{2}(f - \overline{f}) \in \mathcal{A}_{\mathbb{R}}$. We have that $\mathcal{A}_{\mathbb{R}}$ separates points; to see this, let $x_1, x_2 \in K$ with $x_1 \neq x_2$. By Theorem 7.31 (applied to $\mathcal{A}$) there exists $f \in \mathcal{A}$ such that $f(x_1) = 1$ and $f(x_2) = 0$. Writing $f = u + iv$, we have that $u(x_1) = 1 \neq u(x_2) = 0$ and $u \in \mathcal{A}_{\mathbb{R}}$. Furthermore, $\mathcal{A}_{\mathbb{R}}$ vanishes at no point. For all $x \in K$, by assumption there exists $f \in \mathcal{A}$ such that $f(x) \neq 0$, and hence $u(x) \neq 0$ (or $v(x) \neq 0$). Hence by Theorem 7.32, we have that $\mathcal{A}_{\mathbb{R}} = \mathcal{C}_{\mathbb{R}}(K)$ (that is, the real values continuous functions on K). Let $f = u + iv \in \mathcal{C}_{\mathbb{C}}(K)$ and let $\epsilon > 0$. There exist $\tilde{u}, \tilde{v}$ such that $\|u - \tilde{u}\| < \frac{\epsilon}{2}$ and $\|v - \tilde{v}\| < \frac{\epsilon}{2}$, so it follows that $\left\|f - (\tilde{u} + i\tilde{v})\right\| \leq \|u - \tilde{u}\| + \|v - \tilde{v}\| < \epsilon$. □

8 Some Special Functions

8.1 Power Series, Revisited

Recall our definition of power series (Definition 3.38), functions of the form:

$$f(x) = \sum_{n=0}^{\infty} c_n x^n.$$

Also recall the radius of convergence (Theorem 3.39) of such power series, defined as:

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}}.$$

Note that if $\limsup_{n\to\infty} \sqrt[n]{|c_n|} = \infty$, then $R = 0$, and if $\limsup_{n\to\infty} \sqrt[n]{|c_n|} = 0$, then $R = \infty$. The series converges absolutely for $|x| < R$ and diverges for $|x| > R$.

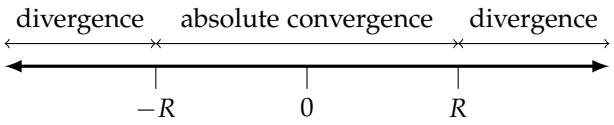

Figure 49: Visualization of the radius of convergence for $f(x) = \sum_{n=0}^{\infty} c_n x^n$, $x \in \mathbb{R}$.

Theorem 8.1

If $\sum_{n=0}^{\infty} c_n x^n$ converges for $|x| < R$, let $f(x) = \sum_{n=0}^{\infty} c_n x^n$ for $|x| < R$. Then, the series converges uniformly on $[-R+\epsilon, R-\epsilon]$ for any $\epsilon > 0$, f is differentiable (and hence continuous) on $(-R, R)$ and $f'(x) = \sum_{n=0}^{\infty} n c_n x^{n-1}$.

Proof

We first show the uniform convergence on $[-R+\epsilon, R-\epsilon]$. For $|x| \leq R-\epsilon$, we have that $|c_n x^n| \leq |x_n|(R-\epsilon)^n$. Since $\sum_n |c_n|(R-\epsilon)^n < \infty$ (by the assumed absolute convergence on $(-R, R)$), we have that $\sum_n c_n x^n$ converges uniformly in $|x| \leq R-\epsilon$ by the M-test (Theorem 7.10).
We next prove the claim about the differentiability/derivative of f. The radius of convergence of $\sum_n n c_n x^{n-1}$ is:

$$\frac{1}{\limsup_{n\to\infty} \sqrt[n]{|nc_n|}} = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}} = R$$

so since f converges in $(-R, R)$, so does $\sum_n n c_n x^{n-1}$. Now, let $s_n(x) = \sum_{m=0}^{n} c_m x^m$. Then, by the linearity of the derivative we have that $s_n'(x) = \sum_{m=1}^{n} m c_m x^{m-1}$. By the first part of the proof, we have that $s_n'(x) \to \sum_{m=1}^{\infty} m c_m x^{m-1}$ uniformly on $[-R+\epsilon, R-\epsilon]$. Since $s_n(x) \to f(x)$ uniformly, by Theorem 7.17, we have that f' exists on $[-R+\epsilon, R-\epsilon]$ and $f'(x) = \sum_{m=1}^{\infty} m c_m x^{m-1}$. Since ϵ is arbitrary, f' exists and is equal to $\sum_{m=1}^{\infty} m c_m x^{m-1}$ for all $x \in (-R, R)$. □

As a remark, note that we can (interestingly) prove the differentiability of f on $(-R, R)$ from the uniform convergence on $[-R+\epsilon, R-\epsilon]$.

Corollary

If $f(x) = \sum_{n=0}^{\infty} c_n x^n$ converges for $|x| < R$, then $f^{(k)}(x)$ exists for all $k \in \mathbb{N}$ and for all $x \in (-R, R)$. It is given by:

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\dots(n-k+1)c_n x^{n-k} \quad (*)$$

and consequently, we have that $c_k = \frac{f^{(k)}(0)}{k!}$, so $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$.

Compare the above Corollary to Taylor's theorem (Theorem 5.15). Here, we take our taylor polynomial and extend it to an infinite series (the limit of polynomials).

Proof

By Theorem 8.1, we have that $f'(x) = \sum_{n=1}^{\infty} nc_n x^{n-1}$ and $f''(x) = \sum_{n=2}^{\infty} n(n-1)c_n x^{n-2}$ and so on. Setting $x = 0$ in $(*)$, we have that $f^{(k)}(0) = n(n-1)\dots 1c_n = n!c_k$, so $c_n = \frac{f^{(k)}(0)}{n!}$. □

Recall the definition of *analytic functions*, which are infinite differentiable and can be represented as sums or series of derivatives evaluated at zero.

Example

As we discussed in Chapter 5, there are infinitely differentiable functions that are not analytic. Let:

$$f(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

By Theorem 8.1, f is infinitely differentiable, and $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. But, $f(x) \neq 0$ except at $x = 0$. Hence, $f(x) \neq \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$ except at $x = 0$. This is true despite the fact that the RHS converges to zero for all $x \in \mathbb{R}$.

Example

Bump functions are continuous, infinitely differentiable functions of compact support (it is zero outside of a compact set). For example,

$$f(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right) & x \in (-1, 1) \\ 0 & |x| \geq 1 \end{cases}$$

is an example of a bump function. Such functions are very useful in the study of functional analysis and PDEs.

Before we continue, some remarks on the radius of convergence are in order. Of course, the definition of $R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{|c_n|}}$ always holds. But in practice, the ratio test is much nicer to use (though it does not

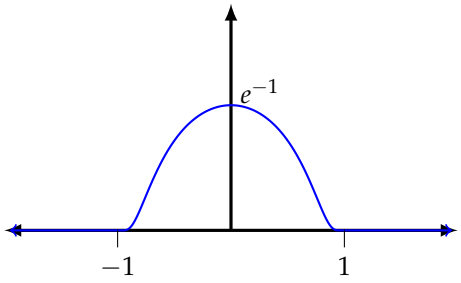

Figure 50: Plot of the bump function f from the above example.

always work). We can make the observation that:

$$\left|\frac{c_{n+1}x^{n+1}}{c_n x^n}\right| = |x|\left|\frac{c_{n+1}}{c_n}\right| \to |x|L$$

Where we take the $n \to \infty$ limit in the final expression (assuming the limit exists). We have that the power series converges if $|x| < \frac{1}{L}$ and diverges if $|x| > \frac{1}{L}$. So, when the limit exists, we can write:

$$R = \frac{1}{\lim_{n\to\infty}\left|\frac{c_{n+1}}{c_n}\right|}.$$

In general, by Theorem 3.37 (not covered in lecture in 320, see Rudin), we have that:

$$\frac{1}{\limsup_{n\to\infty}\left|\frac{c_{n+1}}{c_n}\right|} \leq R \leq \frac{1}{\liminf_{n\to\infty}\left|\frac{c_{n+1}}{c_n}\right|}$$

Theorem 8.2: Abel's Theorem

Suppose $\sum_{n=0}^{\infty} c_n$ converges (perhaps conditionally). Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$. Then, $f(x)$ converges if $|x| < 1$, and $\lim_{x\to 1^-} f(x) = \sum_{n=0}^{\infty} c_n$.

Proof

For the first claim, we have that $\limsup_{n\to\infty} \sqrt[n]{|c_n|} \leq 1$ so by the root test, $\sum_n c_n x^n$ has $R \geq 1$. The interesting case is when $R = 1$ (as if $R > 1$, then f is continuous on $(-R, R)$ and the result follows immediately). In this case, $f(x)$ converges if $|x| < 1$. Let $s_n = \sum_{m=0}^{n} c_m$ and $s = \sum_{m=0}^{\infty} c_m = \lim_{n\to\infty} s_n$. Let $s_{-1} = 0$. Then, we have that $s_n - s_{n-1} = c_n$ for $n \geq 0$.
Let $\epsilon > 0$. We wish to show that there exists $\delta > 0$ such that for $1 - \delta < x < 1$ we have that $\left|f(x) - s\right| < \epsilon$. We start with the partial sum of $f(x)$. For $|x| < 1$, we have:

$$\begin{aligned}\sum_{m=0}^{n} c_m x^m &= \sum_{m=0}^{n} (s_m - s_{m-1}) x^m \\ &= \sum_{m=0}^{n} s_m x^m - x \sum_{m=0}^{n-1} s_m x^m \\ &= (1-x) \sum_{m=0}^{n} s_m x^m + s_n x^{n+1}.\end{aligned}$$

Now, let $n \to \infty$. We then have that $s_n x^{n+1} \to 0$ as $s_n \to s$ and $x^{n+1} \to 0$. We then have that $f(x) = (1-x)\sum_{m=0}^{n} s_m x^m + 0$, and using that $\sum_{m=0}^{\infty} x^m = \frac{1}{1-x}$, we obtain that:

$$\left|f(x) - s\right| = \left|(1-x) \sum_{m=0}^{\infty} (s_m - s) x^m\right| \leq |1-x| \sum_{m=0}^{\infty} |s_m - s||x|^m.$$

Now, choose N such that $m \geq N$ implies $|s_m - s| < \frac{\epsilon}{2}$. Let $x \in (0, 1)$, and then:

$$\left|f(x) - s\right| \leq (1-x) \sum_{m=0}^{N} |s_m - s| x^n + (1-x)\frac{\epsilon}{2}\frac{1}{1-x}$$

The second term on the RHS is bounded using the geometric series. The first term is a polyonimal in x and hence continuous everywhere (including at $x = 1$). It equals 0 at $x = 1$, so it has absolute value less than $\frac{\epsilon}{2}$ if $x \in (1-\delta, 1)$ for some δ. Hence:

$$\left|f(x) - s\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

See Rudin page 175 for an application of this Theorem to prove Theorem 3.51 in a different way. Not that for the case where $\sum_n c_n = \infty$, the theorem still holds, with $\lim_{x\to 1^-} \sum_{n=0}^{\infty} c_n x^n = \infty$. □

Theorem 8.3

Suppose $\sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} \left|a_{ij}\right|\right) < \infty$. Then, $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij}$ (and both converge).

Proof

Rudin uses an overly clever proof. See HW7Q2 for a more natural one. □

Theorem 8.4

Suppose $f(x) = \sum_{n=0}^{\infty} c_n x^n$ has radius of convergence R (Taylor series of f at 0/Maclaurin series). Let $|a| < R$. Then, $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}(x-a)^n$ for (at least) $|x-a| < R - |a|$.

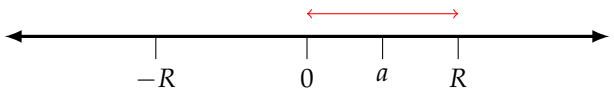

Figure 51: Visualization of Theorem 8.4. The new Taylor series around $x = a$ converges in the region up to the radius of convergence of the original series.

Proof

We have that $f(x) = \sum_{n=0}^{\infty} c_n \left[(x-a)+a\right]^n = \sum_{n=0}^{\infty} c_n \sum_{m=0}^{n} \binom{n}{m}(x-a)^m a^{n-m}$, and we want to find a way to interchange the order of summation. By Theorem 8.3, the interchange is permitted if:

$$\sum_{n=0}^{\infty}\sum_{m=0}^{n} |c_n| \binom{n}{m} |x-a|^m |a|^{n-m} < \infty$$

but the above is equaivalent to:

$$\sum_{n=0}^{\infty} |c_n| \left(|x-a| + |a|\right)^n$$

which converges if $|x-a| + |a| < R$. Therefore:

$$f(x) = \sum_{m=0}^{\infty}\sum_{n=m}^{\infty} c_n \binom{n}{m} a^{n-m}(x-a)^m$$

over $n \geq m$. We need to show that these new coefficients $\left(\sum_{n=m}^{\infty} c_n \binom{n}{m} a^{n-m}\right)$ are equal to $\frac{f^{(m)}(a)}{m!}$. Expanding out, we have that:

$$\sum_{n=m}^{\infty} c_n \binom{n}{m} a^{n-m} = \frac{1}{m!} n(n-1)\ldots(n-m+1) c_n a^{n-m} = \frac{1}{m!} f^{(m)}(a)$$

where in the last equality we use Theorem 8.1. We conclude that:

$$f(x) = \sum_{m=0} \frac{f^{(m)}(a)}{m!}(x-a)^m \text{ for } |x-a| < R - |a|$$

Example

Let $f(x) = \sum_{n=0}^{\infty} x^n$. If $|x| < 1$, then the series converges and $f(x) = \frac{1}{1-x}$ ($R = 1$). Choose $a = -\frac{1}{2}$ (we look at the Taylor serires around $x = -\frac{1}{2}$). For any $|x| < 1$, we have that $f^{(n)}(x) = \frac{n!}{(1-x)^{n+1}}$, so therefore:

$$f^{(n)}(-\frac{1}{2}) = \frac{n!}{(1+\frac{1}{2})^{n+1}} = \left(\frac{2}{3}\right)^{n+1} n!.$$

By Theorem 8.4, we then have that:

$$f(x) = \sum_{n=0}^{\infty} \frac{\left(\frac{2}{3}\right)^{n+1} n!}{n!}(x-a)^n = \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{n+1} (x+\frac{1}{2})^n$$

which is valid for $\left|x-(-\frac{1}{2})\right| < 1 - \left|-\frac{1}{2}\right|$ and hence if $\left|x+\frac{1}{2}\right| < \frac{1}{2}$. Note that in fact, $\sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{n+1} (x-a)^n$ converges whenever $\left|\frac{2}{3}(x+\frac{1}{2})\right| < 1$, in other words, whenever $\left|x+\frac{1}{2}\right| < \frac{3}{2}$, so the series converges in $(-\frac{3}{2}, \frac{1}{2})$.

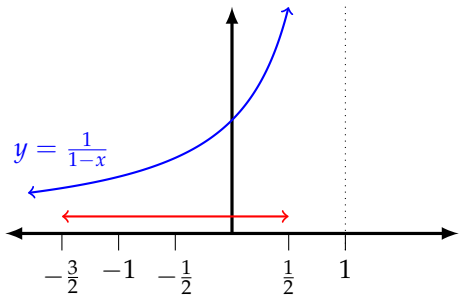

Figure 52: Plot of $y = \frac{1}{1-x}$ and region for which the Taylor series around $a = -\frac{1}{2}$ is valid (red).

As a remark, note that although Theorem 8.4 only guarantees convergence up to the previous boundary, in general the Taylor series will converge up to the nearest singularity. The above theorem is a good example of *analytic continuation*, where a representation of a function converges in a larger interval than the original series (notice that the above taylor series for f around $a = -\frac{1}{2}$ converges up to $-\frac{3}{2}$, when the original series only converged up to -1). We extend the function to a larger interval.

Note that there is another way to obtain the Taylor series around $a = -\frac{1}{2}$ (in a technique reminiscent of that used in MATH 300). We can cleverly manipulate the original expression and the geometric series formula to observe that:

$$f(x) = \frac{1}{1-x} = \frac{1}{1-x+\frac{1}{2}-\frac{1}{2}} = \frac{2}{3}\frac{1}{1-\frac{2}{3}(x+\frac{1}{2})} = \frac{2}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n (x+\frac{1}{2})^n.$$

Theorem 8.5: Principle of Permanence of Form

Suppose $\sum_n a_n x^n$ and $\sum_n b_n x^n$ each have radius of convergence larger or equal to R. Suppose $D \subset (-R, R)$ has a limit point in $(-R, R)$ (for example, $D = \left\{\frac{R}{n} : n = 2, 3, \ldots\right\}$ with limit point 0). If $\sum_n a_n x^n = \sum_n b_n x^n$ in all $x \in D$, then $a_n = b_n$ for all $n \in \mathbb{N}$ and $\sum_n a_n x^n$ and $\sum_n b_n x^n$ for all $x \in (-R, R)$.

Note that this also holds for complex variables, so it can be a way of taking things we know in a real context and promoting it to a complex context. We will first prove a lemma.

Lemma (Problem 2.6)

Let $E \subset X$ for a metric space X. Then the set E' of limit points of E is closed.

Proof

Let x be a limit point of E'. We wish to show that $x \in E'$. Let $\delta > 0$. Since x is a limit point of E', for any $r > 0$ there exists some $y \in E'$, $y \neq x$ such that $y \in N_r(x)$. Since $y \in E'$, for any $\delta - r > 0$ we have that $N_{\delta - r}(y)$ contains a point z of E. Therefore, we have that:

$$d(x, z) \leq d(x, y) + d(y, z) < r + \delta - r = \delta$$

so any neighbourhood $N_\delta(x)$ of x contains some point of E. Hence, x is a limit point of E and $x \in E'$. Hence E' is closed. □

We now move to the proof of the theorem.

Proof

Let $c_n = a_n - b_n$ and $f(x) = \sum_n c_n x^n$. Then, $f(x) = 0$ for all $x \in D$. Let $E = \{x \in (-R, R) : f(x) = 0\}$, so $D \subset E$. We want to show that $E = (-R, R)$. Let $A = E' \cap (-R, R)$. By hypothesis, $A \neq \emptyset$ as D has a limit point in $(-R, R)$. Let $B = (-R, R) \setminus A$. Then, $(-R, R) = A \cup B$ and $A \cap B = \emptyset$. By the above Lemma, the set of limit points E' is closed. Hence, A is closed and B is open. We claim that A is also open.

To see that this is the case, we show that E' is open (then $E' \cap (-R, R)$ is open as a finite intersection of open sets is open). Let $x_0 \in A$. Then, there exists d_n such that $f(x) = \sum_n d_n (x - x_0)^n$ for $|x - x_0| < R - |x_0|$ by Theorem 8.4. We will show that $d_n = 0$ for all n, showing that $f(x) = 0$ for all $x \in I_0 = N_{R - |x_0|}(x_0)$, and therefore that A is open. Suppose for the sake of contradiction that this is false. Then, there exists k such that $d_k \neq 0$, i.e. $f(x) = \sum_{n=k}^{\infty} d_n (x - x_0)^n$ with $d_k \neq 0$. So, $f(x) = (x - x_0)^k \sum_{n=0}^{\infty} d_{n+k} (x - x_0)^n = g(x)$. Then, $g(x_0) = d_k \neq 0$. By the continuity of g, there exists $\delta > 0$ such that $g(x) \neq 0$ for $|x - x_0| < \delta$. But then, $f(x) = (x - x_0)^k g(x) \neq 0$ if $0 < |x - x_0| < \delta$. But then we have that there exists a neighbourhood of x_0 such that $f(x) \neq 0$, but then x_0 cannot be a limit point of $\{x : f(x) = 0\}$. So, $x_0 \notin A$, which is a contradiction. Hence $d_n = 0$ for all $n \in \mathbb{N}$, and A is open.

Given the claim, we have that $A \cap \overline{B} = \overline{A} \cap B = \emptyset$ and hence A and B are separated sets. Since $(-R, R)$ is connected and equals $A \cup B$, one of A, B must be empty. It cannot be A (as $E' \neq \emptyset$ by assumption) so it must be B. Hence $B = \emptyset$ and $A = (-R, R) = E'$. This means that any $x \in (-R, R)$ is a limit point of E, so there exists $\{x_n\} \subset E$ such that $x_n \to x$. Since f is continuous on $(-R, R)$, we have that $f(x) = \lim_{n \to \infty} f(x_n) = 0$. So, $x \in E$ and $E = (-R, R)$. □

Having proven some more properties of power series, we now move to formal definitions of the exponential/logarithmic/trigonometric functions and a from-first-principles proof of their properties.

8.2 The Exponential Function

Recall Definition 3.30, where we defined $e = \sum_{n=0}^{\infty} \frac{1}{n!}$. We also showed in Theorem 3.31 that $e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$. We also know that $2 < e < 3$, and by Theorem 3.32 that e is irrational. We now define a power series based on e.

Definition: The Exponential Function

For $z \in \mathbb{C}$, we define:

$$E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

In particular, we note that $E(1) = e$. Also note that (from Example 3.40) that the above power series has radius of convergence $R = \infty$. We will show with the next sequence of theorems that $E(z)$ coincides with the exponential function $\exp(z) = e^z$ that we are familiar with.

Theorem: Addition Formula

For $z, w \in \mathbb{C}$, we have that $E(z + w) = E(z)E(w)$.

Proof

From the defintion, we have that:

$$\begin{aligned} E(z)E(w) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k}{k!} \frac{w^{n-k}}{(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^k w^{n-k} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} (z + w)^n \\ &= E(z + w) \end{aligned}$$

where in the second equality we use Theorem 3.50 for the multiplication of series. □

We can now use this addition formula to show how $E(p)$ coincides with $\exp(p)$:

Theorem

For $p \in \mathbb{Q}$, $E(p) = e^p$.

Proof

Setting $z = w = 1$ in the addition formula above, we note that:

$$E(2) = E(1+1) = E(1)^2 = e^2, \quad E(3) = E(2+1) = E(2)E(1) = e^2 e^1 = e^3$$

so by induction, it follows that for $n \in \mathbb{N}$:

$$E(n) = e^n$$

Furthermore, we observe that:

$$E(z)E(-z) = E(z-z) = E(0) = 1$$

so it follows that $E(z) = \frac{1}{E(z)}$. Hence, $E(-N) = \frac{1}{e^N} = e^{-N}$ for $N \in \mathbb{N}$. Finally, let $p = \frac{m}{n}$ with $m, n \in \mathbb{N}$. Then, we have that:

$$e^m = E(m) = E(np) = E(p + \ldots + p) = E(p)^n$$

so therefore:

$$E(\frac{m}{n}) = E(p) = e^{\frac{m}{n}}$$

Additionally, we have that:

$$1 = E(p-p) = E(p)E(-p) = e^p E(-p) \implies E(-p) = \frac{1}{e^p}.$$

We have hence successfully shown that $E(p) = e^p$ for $p \in \mathbb{Q}$. □

Next, we will extend the theorem for real exponents and study properties. We will have to come up with a different definition for an exponential of irrational powers; note that so far, our definition for exponentials a^x only could accomodate rational x.

Definition: Real Exponentials

For $x \in \mathbb{R}$, let:

$$e^x = E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

The above defintion defines $x \mapsto e^x$ as an analytic function on $\mathbb{R}$, which is therefore infinitely differentiable on $\mathbb{R}$. Futhermore, the derivatives have a certain property.

Theorem

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

Proof

By the defintion of e^x we have that:

$$\frac{\mathrm{d}}{\mathrm{d}x} e^x = \sum_{n=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{x^n}{n!} \right) = \sum_{n=0}^{\infty} \left(\frac{n x^{n-1}}{(n-1)!} \right) = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

□

Theorem

$e^x > 0$ for all $x \in \mathbb{R}$.

Proof

Since $e^{x+y} = e^x e^y$ by the addition formula, we have for any $x \in \mathbb{R}$ that $e^x e^{-x} = 1$. Since $e^x > 0$ if $x \geq 0$ by definition, it follows that $e^{-x} > 0$ as well if $e^x e^{-x} = 1 > 0$ is to be satisfied. □

Theorem

e^x is strictly increasing and strictly convex.

Proof

Since $\frac{\mathrm{d}}{\mathrm{d}x} e^x = e^x > 0$ and $\frac{\mathrm{d}^2}{\mathrm{d}x^2} e^x = e^x > 0$ by the two previous theorems, we conclude that e^x is strictly increasing and convex by Theorems 5.11 (and its extension). □

Theorem: Superpolynomial Asymptotic Growth

$\lim_{x \to \infty} \frac{e^x}{x^n} = \infty$.

Proof

For $n \geq 0$, the defintion shows that:

$$e^x > \frac{x^{n+1}}{(n+1)!}$$

if $x > 0$, as on the RHS we drop all but the $k = n+1$ term in the series (and all of the terms are positive for $x > 0$). Rearranging, we have that:

$$\frac{e^x}{x^n} > \frac{x}{(n+1)!}$$

and the claim follows by taking $x \to \infty$. □

The above theorem shows that e^x grows faster than any power of x. We also obtain the Corollary:

Corollary

$\lim_{x\to\infty} e^x = \infty$, $\lim_{x\to\infty} e^{-x} = 0$.

Proof

The first claim follows by setting $n = 0$ in the previous Theorem, and the second by using that $e^{-x} = \frac{1}{e^x}$. □

Since e^x is strictly increasing, we may now define an inverse of e^x at each $y > 0$.

8.3 The Logarithm

Definition: The Logarithm

For $y > 0$, define $L(y) = \log(y)$ by $E(L(y)) = y$ for $y > 0$, $y \in \mathbb{R}$. Equivalently, we can define the logarithm as $L(E(x))) = x$ for $x \in \mathbb{R}$ or $e^{\log y} = y$ or $\log e^x = x$.

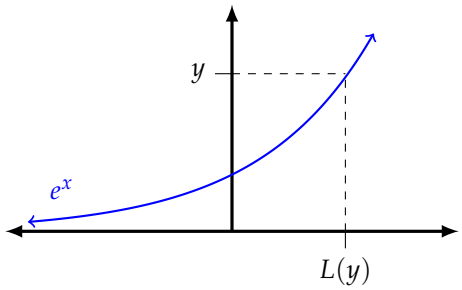

Figure 53: Plot of e^x and extraction of its inverse $L(y)$.

Theorem: Derivative and Addition Formula

$L'(y) = \frac{1}{y}$ for $y > 0$ and $L(uv) = L(u) + L(v)$

Proof

Since $L(E(x)) = x$ by definition, by the chain rule (Theorem 5.5) we have that:

$$L'(E(x))E'(x) = 1 \implies L'(E(x)) = \frac{1}{E(x)}.$$

Where we use that $E'(x) = E(x)$ in the above implication. Letting $E(x) \mapsto y$, we have that $L'(y) = \frac{1}{y}$ as claimed. For the second claim, consider that for all $u, v > 0$, there exists unique $x, y \in \mathbb{R}$ (which can be denoted as $\exists! x, y \in \mathbb{R}$) such that $E(x) = u, E(y) = v$. Hence:

$$L(uv) = L(E(x)E(y)) = L(E(x+y)) = x + y = L(u) + L(v).$$

□

Corollary

For $y > 0$, $L(y) = \int_1^y \frac{1}{t} dt$.

Proof

By the Fundamental Theorem of Calculus (Theorem 6.21) we have that $\int_1^y \frac{1}{t} dt = L(y) - L(1) = L(y)$. □

Theorem

For $p \in \mathbb{Q}$, $L(x^p) = pL(x)$.

Proof

By the addition formula, we have that $L(\frac{1}{x}) + L(x) = L(\frac{1}{x}x) = L(1) = 0$. Hence, $L(\frac{1}{x}) = -L(x)$. Furthermore, by the addition formula, we have that $L(x^n) = nL(x)$ by induction. Furthermore, $L(x^0) = L(1) = 0 = 0L(x)$ which shows that the formula holds for $\mathbb{N} \cup \{0\}$. Combining the previous facts, we have that:

$$nL(x^{\frac{1}{n}}) = L((x^{\frac{1}{n}})^n) = L(x^1) \implies L(x^{\frac{1}{n}}) = \frac{1}{n}L(x)$$

We therefore have that for $p = \frac{m}{n}$ with $m, n \in \mathbb{N}$ that:

$$L(x^{\frac{m}{n}}) = mL(x^{\frac{1}{n}}) = \frac{m}{n}L(x)$$

Furthermore,

$$L(x^{-\frac{m}{n}}) = L\left(\frac{1}{x^{\frac{m}{n}}}\right) = -L(x^{\frac{m}{n}}) = -\frac{m}{n}L(x)$$

which shows that the proposed identity holds for all $p \in \mathbb{Q}$. □

Definition: Real Exponentials with Arbitrary Base

For $\alpha \in \mathbb{R}$, we define $x^\alpha = e^{\alpha \log \alpha}$.

Note that the above definition is equivalen to the definition made in MATH 320 HW3Q1, but it is a much cleaner definition (as we will soon see)!

Theorem: Generalized Power Rule

$\frac{d}{dx}x^\alpha = \alpha x^{\alpha-1}$ and hence x^α has antiderivative:

$$\begin{cases} \frac{1}{\alpha+1}x^{\alpha+1} & \text{if } \alpha \neq -1 \\ \log x & \text{if } \alpha = 1 \end{cases}.$$

Proof

From the definition of x^α, we have that:

$$\frac{\mathrm{d}}{\mathrm{d}x}\, x^\alpha = \frac{\mathrm{d}}{\mathrm{d}x}\left(e^{\alpha \log x}\right) = e^{\alpha \log x}\alpha\frac{1}{x} = \alpha\frac{x^\alpha}{x} = \alpha x^{\alpha-1}$$

where in the second equality we use the chain rule and the fact that $L'(y) = \frac{1}{y}$. □

Theorem: Subpolynomial Asymptotic Growth

$\lim_{x\to\infty} \log x = \infty$, $\lim_{x\to 0^+} = -\infty$, and $\lim_{x\to\infty} \frac{\log x}{x^\alpha} = 0$ if $\alpha > 0$.

Proof

To realize the first two equalities, we first make the observation that $\log(x)$ is (Strictly) monotonically increasing. To see this, let $x_1, x_2 > 0$ and $x_1 < x_2$ and then we have that:

$$\log(x_2) - \log(x_1) = \int_0^{x_2} \frac{1}{t}dt - \int_0^{x_1} \frac{1}{t}dt = \int_{x_1}^{x_2} \frac{1}{t}dt > 0$$

where the bound follows from the fact that $\frac{1}{t} > 0$ for all $t > 0$ and $x_2 - x_1 > 0$. Hence, $\log(x)$ is monotonicaly increasing, and to compute the limits it suffices to compute the limit along a specific choice of sequence that tends to ∞ or 0 respectively. Using that $\lim_{n\to\infty} e^n = \infty$ and $\lim_{n\to\infty} e^{-n} = 0$, we have that:

$$\lim_{n\to\infty} \log\left(e^n\right) = \lim_{n\to\infty} n\log(e) = \lim_{n\to\infty} n = \infty$$

$$\lim_{n\to\infty} \log\left(e^{-n}\right) = \lim_{n\to\infty} -n\log(e) = \lim_{n\to\infty} -n = -\infty$$

so we conclude that $\lim_{x\to\infty} \log(x) = \infty$ and $\lim_{x\to 0} \log(x) = -\infty$.
For the third claim, we let $a > 0$ and $x > 1$. Then:

$$\log(x) = \int_1^x \frac{1}{t}dt < \int_1^x t^a\frac{1}{t}dt = \left.\frac{t^a}{a}\right|_1^x = \frac{x^a}{a} - \frac{1}{a} < \frac{x^a}{a}$$

so choosing $a \in (0, \alpha)$ we have that:

$$\frac{1}{x^\alpha}\log x < \frac{1}{a}\frac{1}{x^{\alpha-a}} \to 0 \text{ as } x \to \infty.$$

□

8.4 Cosine and Sine

We have seen that $E(z+w) = E(z)E(w)$ for all $z, w \in \mathbb{C}$. A natural definition for the complex exponential follows.

Definition: Complex Exponentials

Given $z \in \mathbb{C}$, we define:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = E(z)$$

As a remark, note in taking the complex conjugate of $\exp(z)$, we can absorb the conjugation into the argument:

$$\overline{\exp(z)} = \overline{\sum_{n=0}^{\infty} \frac{z^n}{n!}} = \sum_{n=0}^{\infty} \frac{\overline{z}^n}{n!} = \exp(\overline{z})$$

We will now define the trigonometric functions using the complex exponential, and prove the properties that we would expect them to have from our prior geometric notions.

Definition: Cosine and Sine

Let $x \in \mathcal{R}$. We then define:

$$C(x) = \operatorname{Re} E(ix) = \frac{1}{2}\left[e^{ix} + e^{-ix}\right]$$
$$S(x) = \operatorname{Im} E(ix) = \frac{1}{2i}\left[e^{ix} - e^{-ix}\right]$$

Theorem: Euler's Formula

$E(ix) = C(x) + iS(x)$.

Proof

The formula is an immediate consequence of the definitions of $C(x), S(x)$. □

Note that $C(x), S(x)$ can alternatively (equivalently) be defined as power series:

$$C(x) = \frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{(ix)^n}{n!} + \frac{(-ix)^n}{n!}\right) = \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}$$
$$S(x) = \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n+1)!}$$

Theorem

Let $x \in \mathbb{R}$. We then have that:

(a) $C(x) = C(-x)$, $C(0) = 1$ and $S(x) = -S(-x)$, $S(0) = 0$.

(b) $C^2(x) + S^2(x) = 1$.

(c) $C'(x) = -S(x)$ and $S'(x) = C(x)$.

Proof

(a) From the definitions of C and S, we have:

$$\begin{aligned} C(-x) &= \frac{1}{2}\left[e^{-ix} + e^{ix}\right] = C(x) \\ C(0) &= \frac{1}{2}\left[e^{i(0)} + e^{-i(0)}\right] = \frac{1}{2}\left[1+1\right] = 1 \\ -S(-x) &= -\frac{1}{2}\left[e^{-ix} - e^{ix}\right] = \frac{1}{2}\left[e^{ix} - e^{-ix}\right] = S(x) \\ S(0) &= \frac{1}{2}\left[e^{i(0)} - e^{-i(0)}\right] = \frac{1}{2}\left[1-1\right] = 0 \end{aligned}$$

(b) Expanding out the expression, we have:

$$\begin{aligned} C^2(x) + S^2(x) &= \frac{1}{4}\left[e^{ix}e^{ix} + 2e^{ix}e^{-ix} + e^{-ix}e^{-ix}\right] - \frac{1}{4}\left[e^{ix}e^{ix} - 2e^{ix}e^{-ix} + e^{-ix}e^{-ix}\right] \\ &= e^{ix}e^{-ix} \\ &= e^{ix-ix} \\ &= e^{0} \\ &= 1 \end{aligned}$$

(c) Using the linearity of the derivative, and the known result for the derivative of exponentials we have:

$$\begin{aligned} C'(x) &= \frac{1}{2}\left[ie^{ix} - ie^{-ix}\right] = -\frac{1}{2i}\left[e^{ix} - e^{-ix}\right] = -S(x) \\ S'(x) &= \frac{1}{2i}\left[ie^{ix} + ie^{-ix}\right] = \frac{1}{2}\left[e^{ix} + e^{-ix}\right] = C(x) \end{aligned}$$

□

Lemma

There exists $x > 0$ such that $C(x) = 0$.

Proof

Suppose for the sake of contradiction that $C(x) > 0$ for all $x > 0$. Then, S is strictly increasing as $S' = C$. So, for all $y > x$, we then have that:

$$S(x)(y-x) < \int_x^y S(t)dt = -C(y) + C(x) \leq 2$$

but letting $y \to \infty$, we get that $\infty \leq 2$ which is a contradiciton. □

Definition: π

Given the above Lemma, $x_0 = \inf\{x > 0 : C(x) = 0\}$ exists. In particular, $x_0 > 0$ since $C(0) = 1$ and C is continuous. Then, we define $\pi = 2x_0$.

Theorem

$S(\frac{\pi}{2}) = 1$.

Proof

By the definition of π, we have that $C(x) > 0$ for $x \in [0, \frac{\pi}{2})$ and $C(\frac{\pi}{2}) = 0$. From the previous theorem we have that $C^2(\frac{\pi}{2}) + S^2(\frac{\pi}{2}) = 1$ so we obtain that $S(\frac{\pi}{2}) = \pm 1$. Since $S(0) = 0$ and $S'(x) = C(x) > 0$ for $x \in [0, \frac{\pi}{2})$, we conclude that $S(\frac{\pi}{2}) = 1$. □

Note the implication this result has for e^{ix}; using Euler's Formula, we find that:

$$e^{i\frac{\pi}{2}} = \cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right) = 0 + i1 = i$$

Therefore:

$$e^{i\pi} = \left(e^{i\frac{\pi}{2}}\right)^2 = i^2 = -1$$
$$e^{i\frac{3\pi}{2}} = \left(e^{i\frac{\pi}{2}}\right)^3 = i^3 = -i$$
$$e^{i2\pi} = \left(e^{i\frac{\pi}{2}}\right)^4 = i^4 = 1.$$

From this we notice the periodicity of e^{ix}.

Theorem: Periodicity of Trigonometric Functions

(a) $e^{x+2\pi i} = e^x$.

(b) $C(x + 2\pi) = C(x)$.

(c) $S(x + 2\pi) = S(x)$.

Proof

(a) $e^{x+2\pi i} = e^x e^{2\pi i}$ by the addition formula. Then, $e^{2\pi i} = 1$ by the argument above, proving the identity.

(b) Using the above periodicity of e^{ix}, we have that $C(x + 2\pi) = \text{Re}(e^{i(x+2\pi)}) = \text{Re}(e^{ix}) = C(x)$.

(c) $S(x + 2\pi) = \text{Im}(e^{i(x+2\pi)}) = \text{Im}(e^{ix}) = C(x)$. □

Note that there is a way to relate S and C via a phase shift. We observe that:

$$S(x + \frac{\pi}{2}) = \text{Im}(e^{i(x+\frac{\pi}{2})}) = \text{Im}(e^{ix}e^{i\frac{\pi}{2}}) = \text{Im}(e^{ix}i) = \text{Im}((C(x) + iS(x))i) = \text{Im}(iC(x) - S(x)) = C(x)$$

We can also generalize this formula to get the familiar trigonometric sum identity.

Theorem

$S(x + y) = C(x)S(y) + S(x)C(y)$.

Proof

Using the definition of S and Euler's Formula, we observe that:

$$\begin{aligned} S(x+y) = \text{Im}(e^{i(x+y)}) &= \text{Im}(e^{ix}e^{iy}) \\ &= \text{Im}((C(x)+iS(x))(C(y)+iS(y))) \\ &= \text{Im}(C(x)C(y)+iS(x)C(y)+iC(x)S(y)-S(x)S(y)) \\ &= C(x)C(y)-S(x)S(y) \end{aligned}$$

□

As a final remark before moving onto the next section, we observe that $x \mapsto e^{ix}$ is a bijection from $[0, 2\pi)$ onto the unit circle (points $z \in \mathbb{C}$ with $|z| = 1$). See Rudin for more details.

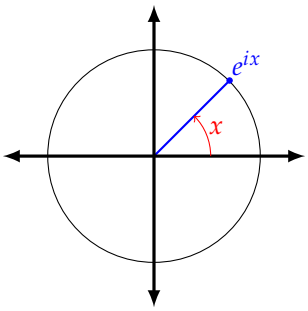

Figure 54: Visualization of the map $x \mapsto e^{ix}$ from $[0, 2\pi)$ onto the unit circle in the complex plane.

8.5 The Algebraic Completeness of the Complex Field

Theorem 8.8: The Fundamental Theorem of Algebra

Let $n \in \mathbb{N}$, and $a_0, a_1, \ldots, a_n \in \mathbb{C}$ such that $a_n \neq 0$. Define the polynomial:

$$P(z) = a_0 + a_1 z + \ldots + a_n z^n$$

with $z \in \mathbb{C}$. Then, there exists $z_0 \in \mathbb{C}$ such that $P(z_0) = 0$.

Corollary

By division, $P_n(z)$ has n roots.

Proof

Assume WLOG that $a_n = 1$ (this can be realized by dividing out by the original nonzero a_n). Let $\mu = \inf_{z\in\mathbb{C}} \left|P(z)\right|$. We wish to show that $\mu = 0$, and that the inf is attained at some $z_0 \in \mathbb{C}$.
We first show that the infimum is attained. The idea of the argument is that for large z, z^n grows the most rapidly and hence it dominates. Hence, $\left|P(z)\right|$ is large. Hence, the inf is attained in some compact disk, but since P is continuous, the inf is therefore obtained somewhere on this compact set. Formally, for $|z| = R$, we have that:

$$\begin{aligned}\left|P(z)\right| = \left|z^n\right|\left|\frac{a_0}{z^n} + \frac{a_1}{z^{n-1}} + \ldots + \frac{a_{n-1}}{z} + 1\right| &\geq |z|^n \left(1 - \frac{|a_0|}{|z|^n} - \frac{|a_1|}{|z|^{n-1}} - \ldots - \frac{|a_{n-1}|}{|z|}\right) \\ &= R^n \left(1 - \frac{|a_0|}{|z|^n} - \frac{|a_1|}{|z|^{n-1}} - \ldots - \frac{|a_{n-1}|}{|z|}\right)\end{aligned}$$

We see that this expression goes to infinity as $R \to \infty$. Hence, $\left|P(z)\right| \geq \mu + 1$ if $|z| \geq R_0$ for some $R_0 > 0$. As $|P|$ is continuous, and $|z| \leq R_0$ is a compact subset of $\mathbb{C}$, μ is attained somewhere on the subset by the Extreme Value Theorem (Theorem 4.16). Therefore, $\mu = \left|P(z_0)\right|$ for some z_0 with $|z_0| \leq R_0$.
Next, we show that $\mu = 0$. Suppose (for the sake of contradiction) that $\mu > 0$ and hence $P(z_0) \neq 0$. Let $Q(z) = \frac{P(z_0+z)}{P(z_0)}$. Then $Q(0) = 1$, so:

$$Q(z) = 1 + b_k z^k + \ldots + b_n z^n$$

where b_k is the first nonzero coeffient. Furthermore, we see that $\left|Q(z)\right| = \frac{\left|P(z_0+z)\right|}{\mu} \geq 1$ for all $z \in \mathbb{C}$ (as μ is the infimum). We will now derive a contradiction by looking at small z (where $z^k > z^{k+1} > \ldots > z^n$). We consider that:

$$\left|Q(z)\right| \leq \left|1 + b_k z^k\right| + \sum_{m=k+1}^{n} |b_m||z^m|$$

We want to choose z small enough such that the first term in the above expression is less than 1 and the others are negligeble. To this end, let us write $bk = \frac{b_k}{|b_k|}|b_k| = e^{it_1}|b_k|$ for some $t_1 \in \mathbb{R}$. Then, let $t = \frac{-t_1+\pi}{k}$ so $t_1 = -kt + \pi$. We then have that $b_k = -e^{-itk}|b_k|$ (where the minus sign comes from $e^{i\pi}$). Choose $z = re^{it}$ with $r > 0$. Then, $z^k = r^k e^{itk}$ and $b_k z^k = -e^{-itk}|b_k|r^k e^{itk} = -r^k|b_k| < 0$. Let us choose r small enough so that $|b_k|r^k < 1$ is satisfied. We then have that:

$$\left|1 + b_k z^k\right| = \left|1 - |b_k|r^k\right| = 1 - |b_k|r^k$$

so therefore:

$$\left|Q(z)\right| \leq 1 - |b_k|r^k + \sum_{m=k+1}^{n} |b_m|r^k = 1 - r^k\left(|b_k| - r|b_{k+1}| - \ldots - r^{n-k}|b_n|\right)$$

where $\left(|b_k| - r|b_{k+1}| - \ldots - r^{n-k}|b_n|\right) > 0$. Hence $\left|Q(z)\right| < 1$, which is a contradiction. We conclude that $\left|P(z_0)\right| = 0$. □

8.6 Fourier Series

We now begin our discussion on Fourier Series. Note that the theory of Fourier Series (and more generally, Harmonic Analysis) is rich enough for it to be its own course, but we will here give an introduction to the topic. Fourier Series show up everywhere, such as (for example) in partial differential equations, or in signal processing. It is an interesting topic of study that combines analysis and linear algebra.

For some additional references on the topic, Katznelson's "Harmonic Analysis" (`https://www.cambridge.org/core/books/an-introduction-to-harmonic-analysis/67C4CE356E7420BA17F3F1337291EF82`) and Grafakos' "Classical Fourier Analysis" (`https://link.springer.com/book/10.1007/978-1-4939-1194-3`) are good texts.

Definition: Inner Product/Norm of Functions

Given $a, b \in \mathbb{R}$ with $a < b$ and integrable $f, g : [a, b] \mapsto \mathbb{C}$ we write:

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx \ \left(= \overline{\langle g, f \rangle}\right)$$

to denote the inner product on the space of complex functions. Note that this inner product induces the norm $\|f\|_2 = \sqrt{\langle f, f \rangle} = \left(\int_a^b |f(x)|^2 dx\right)^{1/2}$.
Note that we will often take $[a, b] = [0, 2\pi]$ and in this case we may prefer:

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x)\overline{g(x)}dx.$$

As a remark, we have that $d(f, g) = \|f - g\|_2$ induces a metric, but we have to be careful; it defines a metric on the metric space of continuous functions (that is, $\mathcal{C}[a, b]$) but *not* as on the space of Riemann-integrable functions (i.e. $\mathcal{R}[a, b]$). To see this, consider that we can have $f \in \mathcal{R}[a, b]$ with $\int_a^b |f(x)|dx = 0$ but $f \neq 0$ (consider any function f which is zero everywhere but is nonzero for a finite number of points).

Definition: Orthogonal/Orthonormal Families of Functions

We say that a family of functions $\phi_n[a, b] \mapsto \mathbb{C}$ are mutually **orthogonal** if:

$$\langle \phi_n, \phi_m \rangle = 0 \text{ if } n \neq m.$$

We say that a family of functions is **orthonormal** if:

$$\langle \phi_n, \phi_m \rangle = \delta_{nm} = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}.$$

Example

(a) Let $[a,b] = [-\pi, \pi]$. Then, $\phi_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$, $n \in \mathbb{Z}$ obey:

$$\langle \phi_n, \phi_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)x} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos\big((m-n)x\big) + i \sin\big((m-n)x\big) dx = \delta_{nm}.$$

(b) Take $\phi_1, \phi_2, \ldots$ to be $\frac{1}{\sqrt{2\pi}}$, $\frac{1}{\sqrt{2\pi}} \cos(x)$, $\frac{1}{\sqrt{2\pi}} \sin(x)$, $\frac{1}{\sqrt{2\pi}} \cos(2x), \ldots$. We then have that this family is orthonormal on $[-\pi, \pi]$.

(c) The Legendre polynomials, defined by $P_n(x) = \sum_{k=0}^{n} \binom{n}{k}\binom{n+k}{n} \left(\frac{x-1}{2}\right)^k$ for $n \in \mathbb{N} \cup \{0\}$ are orthogonal on $[-1, 1]$. Note that $\|P_n\|_2 = \sqrt{\frac{2}{2n+1}}$.

With this definition defined, we will want to use sets of orthonormal functions as a *basis* of L^2 space. We want to be able to write arbitrary $f \in L^2$ as a linear combination of these functions.

Example

As motivation for the next definition, suppose $f(x) = \sum_{m=0}^{N} c_n \phi_n$ with $\{\phi_n\}$ orthonormal on $[a,b]$. Then, we can write:

$$\langle f, \phi_n \rangle = \sum_{m=0}^{N} c_m \langle \phi_m, \phi_n \rangle = \sum_{m=0}^{N} N c_m \delta_{mn} = c_n.$$

That is, we can write $c_n = \langle f, \phi_n \rangle = \int_a^b f(x)\overline{\phi_n(x)} dx$.
As a specific example, suppose $f(x) = \sum_{n=-N}^{N} c_n e^{inx} = \sum_{n=-N} c_n \sqrt{2\pi} \frac{e^{inx}}{\sqrt{2\pi}}$. From the previous example, we have that $\frac{e^{inx}}{\sqrt{2\pi}}$ is orthonormal on $[-\pi, \pi]$. Hence:

$$\sqrt{2\pi} c_n = \langle f, \phi_n \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) e^{inx}.$$

That is, $c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$.

Definition: Fourier Coefficients/Series

If $f \in \mathcal{R}[a,b]$ and $\{\phi_n\}$ is an orthonormal family of functions, then $c_n = \langle f, \phi_n \rangle$ is called a **Fourier coefficient of** f, and the **Fourier series of** f is $\sum_n c_n \phi_n$.

A natural question after making this definition is "when does the Fourier series of f converge?" A follow-up question is "when it converges, is it equal to f?" We will investigate these questions with the next sequence of theorems.

Theorem 8.11

Suppose $f \in \mathcal{R}[a,b]$ and $\{\phi_n\}$ is an orthonormal on $[a,b]$. Let $c_n = \langle f, \phi_n \rangle$ and $s_n = \sum_{m=1}^{n} c_m \phi_m$. Let $t_n = \sum_{m=1}^{n} a_m \phi_m$ for some $a_m \in \mathbb{C}$. Then, we have that:

$$\|f - s_n\|_2^2 \leq \|f - t_n\|_2^2$$

with equality if and only if $c_m = a_m$ for each m.

The moral of the theorem is that if we are to use a linear combination of the first n functions out of an orthonormal set of functions to best approximate f in the L^2 norm, the best way to do so is by using Fourier coefficients.

Proof

By the definition of the norm, we have that:

$$\|f - t\|_2^2 = \langle f - t_n, f - t_n \rangle^2 = \|f\|_2^2 - \langle t_n, f \rangle - \langle f, t_n \rangle + \|t_n\|_2^2$$

Looking at the terms, we have that:

$$\|t_n\|_2^2 = \langle t_n, t_n \rangle = \sum_{k,m=1}^{n} a_k \overline{a_m} \langle \phi_k, \phi_m \rangle = \sum_{k,m=1}^{n} a_k \overline{a_m} \delta_{km} = \sum_{m=1}^{n} a_m \overline{a_m} \quad (1)$$

$$\langle f, t_n \rangle = \sum_{m=1}^{n} \overline{a_m} \langle f, \phi_n \rangle = \sum_{m=1}^{n} c_m \overline{a_m}$$

Therefore:

$$\begin{aligned} \|f - t_n\|_2^2 &= \|f\|_2^2 + \sum_{m=1}^{n} a_m \overline{a_m} - c_m \overline{a_m} - \overline{c_m} a_m \\ &= \|f\|_2^2 + \sum_{m=1}^{n} (a_m \overline{a_m} - c_m \overline{a_m} - \overline{c_m} a_m + c_m \overline{c_m}) - \sum_{m=1}^{n} c_m \overline{c_m} \\ &= \|f\|_2^2 + \sum_{m=1}^{n} |a_m - c_m|^2 - \sum_{m=1}^{n} |c_m|^2 \end{aligned}$$

Putting $a_m = c_m$, we obtain:

$$\|f - s_n\|_2^2 = \|f\|_2^2 - \sum_{m=1}^{n} |c_m|^2. \quad (2)$$

Hence, we have that:

$$\|f - t_n\|_2^2 = \|f - s_n\|_2^2 + \sum_{m=1}^{n} |a_m - c_m|^2$$

And $\sum_{m=1}^{n} |a_m - c_m|^2 \geq 0$ and is 0 if and only if $a_m = c_m$ for all m. This proves the claim. □

Note that putting (1) and (2) together in the above proof give the identity that $\|f\|_2^2 = \|f - s_n\|_2^2 + \|s_n\|_2^2$. There is a geometric interpretation to this identity, which we picture below:

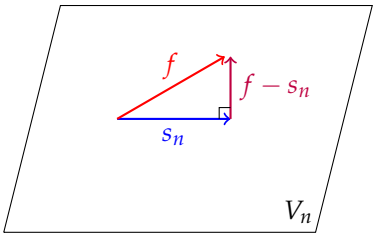

Figure 55: Visualization of the identity $\|f\|_2^2 = \|f - s_n\|_2^2 + \|s_n\|_2^2$. If we let V_n be the set of all linear combinations of $\phi_1, \ldots, \phi_n$ (i.e. all sums of the form $\sum_{m=1}^{n} a_m \phi_m$), then s_n is the orthogonal projection of f onto V_n.

Theorem 8.12: Bessel's Inequality

Suppose $\{\phi_n\}$ is orthonormal on $[a, b]$ and is an infinite family. If $f(x) = \sum_{n=1}^{\infty} c_n \phi_n$, then:

$$\sum_{n=1}^{\infty} |c_n|^2 \leq \int_a^b |f(x)|^2 dx.$$

We call this the Bessel Inequality. Note that this implies:

$$\lim_{n \to \infty} c_n = 0$$

Proof

We start with the identity that $\|f\|_2^2 = \|f - s_n\|_2^2 + \|s_n\|_2^2$. It then follows that $\sum_{m=1}^{n} |c_m|^2 = \|s_m\|_2^2 \leq \|f\|_2^2$. We then take $n \to \infty$. Since $\sum_{m=1}^{n} |c_m|$ is bounded above (by $\|f\|_2^2$) for any n and is monotonically increasing, we conclude that the infinite series converges and hence:

$$\sum_{n=1}^{\infty} |c_n|^2 \leq \int_a^b |f(x)|^2 dx.$$

By the divegence test (Theorem 3.23) we obtain that $\lim_{n \to \infty} c_n = 0$. □

Example

Suppose $f \in \mathcal{R}[-\pi, \pi]$. Then, $\lim_{n \to \infty} \int_{-\pi}^{\pi} f(x) \cos(nx) dx = 0$ by the above Theorem.

The above example is a version of the Riemann-Lebesgue Lemma. The intuitive interpretation of the above example is that $\cos(nx)$ oscillates wildly as $n \to \infty$, and hence under the integral, the peaks/valleys cancel.

Definition: Inner Product/Norm of Functions (Revisited)

From now on, we will restrict ourselves to $f : [-\pi, \pi] \mapsto \mathbb{C}$ with $f \in \mathcal{R}[-\pi, \pi]$. Take $\phi_n(x) = e^{inx}$ for $n \in \mathbb{Z}$. We extend f to all of $\mathbb{R}$ by $f(x + 2\pi) = f(x)$. To make the ϕ_ns orthonormal over our interval, we change our definition of our inner product and norm:

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx, \|f\|_2^2 = \langle f, f \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx$$

We then have that $\langle \phi_n, \phi_m \rangle = \delta_{nm}$. Theorem 8.11 and its consequences hold with this new definition.

Definition: Fourier Series (Revisited)

We define the **Fourier series** of f by:

$$\sum_{n=-\infty}^{\infty} c_n \phi_n(x)$$

where $c_n = \langle f, \phi_n \rangle$ and $\phi_n(x) = e^{inx}$.

Note that at this point, we do note claim that the Fourier series converges, nor that it equals f.

Definition: Partial Fourier Series

The Nth partial Fourier seris of f is defined as $s_N(f; x) = \sum_{n=-N}^{N} c_n e^{inx}$. Where f is clear from context, we sometimes will write $s_N(x)$.

Note that Theorem 8.12 implies that:

$$\sum_{n=-N}^{N} |c_n|^2 = \|s_N\|_2^2 \leq \|f\|_2^2.$$

Definition: The Dirchlet Kernel

The **Dirchlet Kernel** is defined as:

$$D_n(t) = \sum_{n=-N}^{N} e^{int}.$$

Lemma 1

$$D_n(t) = \frac{\sin\left((N+\frac{1}{2})t\right)}{\sin\left(\frac{1}{2}t\right)}.$$

Proof

By definition, we have that $D_N(t) = \sum_{n=-N}^{N} e^{int}$. We then have that:

$$\begin{aligned}
D_N(t) &= \sum_{n=-N}^{N} e^{int} \\
&= e^{-iNt} \sum_{k=0}^{2N} (e^{it})^k && \text{(Common factor)} \\
&= e^{-iNt} \left[\frac{e^{i(2N+1)} - 1}{e^{it} - 1} \right] && \text{(Geometric sum)} \\
&= e^{-iNt} \left[\frac{e^{i(2N+1)} - 1}{e^{it} - 1} \right] \frac{e^{-it/2}}{e^{-it/2}} \\
&= \frac{e^{i(N+1/2)t} - e^{-i(N+1/2)t}}{e^{it/2} - e^{-it/2}} \\
&= \frac{\sin\left((N + \frac{1}{2})t\right)}{\sin\left(\frac{1}{2}t\right)}
\end{aligned}$$

□

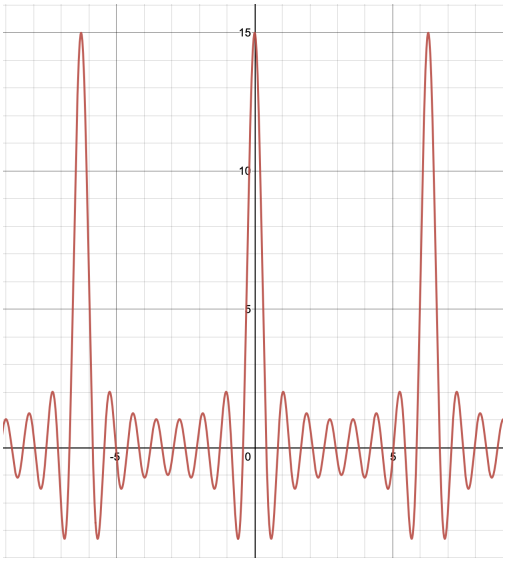

Figure 56: Desmos Visualization of the Dirchlet Kernel $D_N(t)$ for $N = 7$. $D_N(t)$ is 2π periodic, and becomes more sharply peaked at $t = \pi n$, $n \in \mathbb{Z}$ as we increase N. It can be understood as an oscillating function that approaches a δ "function". Readers can play around with the function at `https://www.desmos.com/calculator/satukmy8kj`.

Lemma 2

For $x \in \mathbb{R}$, $s_N(f; x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x - t) D_N(t) dt$.

What we would like to see is that $s_N(f;x) \to f$ as $N \to \infty$. Given the above Lemma, this can be realized if $D_N(t) \to \delta(t)$ (where $\delta(t)$ is the Dirac delta "function"; of course this is not actually a function, but the intuition is that $D_N(t)$ becomes sharply peaked around t) and then $f(x-t) = f(x)$. Note that the integral formula above is of a *convolution integral*. We will discuss properties of it and the Dirchlet kernel, and use it to show that $s_N \to f$ if f is Lipschitz continuous.

Proof

By calculation and algebraic manipulation, we have that:

$$\begin{aligned}
s_N(x) = \sum_{n=-N}^{N} c_n e^{inx} &= \sum_{n=-N}^{N} \langle f, \phi_n \rangle e^{inx} \\
&= \sum_{n=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt e^{inx} \\
&= \sum_{n=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{in(x-t)} dt \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \sum_{n=-N}^{N} e^{in(x-t)} dt \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) D_N(x-t) dt \\
&= \frac{1}{2\pi} \int_{x-\pi}^{x+\pi} f(x-s) D_N(s) ds && \text{(Substitute } s = x-t\text{)} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-s) D_N(s) ds && \text{(Periodicity of } f, D_N\text{)}
\end{aligned}$$

□

Lemma 3

(a) $\frac{1}{2\pi} \int_{-N}^{N} D_N(t) dt = 1$

(b) $D_N(t) = \cos(nt) + \cot\left(\frac{1}{2}t\right) \sin(nt)$.

Proof

(a) $\frac{1}{2\pi} \int_{-N}^{N} D_N(t) dt = \langle D_N, 1 \rangle = \sum_{n=-N}^{N} \langle \phi_n, \phi_0 \rangle = \sum_{n=-N}^{N} \delta_{N_0} = 1.$

(b) $D_N(t) = \frac{1}{\sin\left(\frac{1}{2}t\right)} \left(\sin\left(\frac{1}{2}t\right) \cos(Nt) + \cos\left(\frac{1}{2}t\right) \sin(Nt) \right) = \cos(nt) + \cot\left(\frac{1}{2}t\right) \sin(nt).$ □

Theorem 8.14

Let $x \in \mathbb{R}$, $f \in \mathbb{R}$ on $[-\pi, \pi]$. Suppose there exist $\delta > 0$, $M < \infty$ such that $|f(x+t) - f(x)| \leq M|t|$ for all $|t| < \delta$. Then, we have that $\lim_{N \to \infty} s_N(x) = f(x)$.

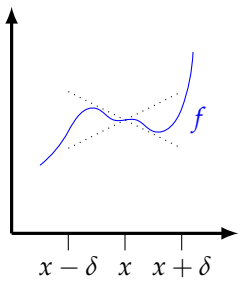

Figure 57: Visualization of the Lipschitz continuity condition on f in Theorem 8.14. f is bounded in between lines of slope $\pm M$ for a neighbourhood $N_\delta(x)$ around x.

Proof

We show that $f(x) - s_N(x)$ goes to zero. The intuition for the proof we will use is that $D_N(t)$ will behave like a delta function, picking out a specific value of x. Using the result of the previous Lemma, we have:

$$\begin{aligned} f(x) - s_N(x) &= f(x)\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(t)dt - \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)D_N(t)dt && \text{Lemma 3(a)/2} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} (f(x) - f(x-t))D_N(t)dt && \text{Theorem 6.12} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} (f(x) - f(x-t))\cos(Nt)dt && \text{Lemma 2(a)} \\ &+ \frac{1}{2\pi}\int_{-\pi}^{\pi} (f(x) - f(x-t))\cot\left(\frac{1}{2}t\right)\sin(Nt)dt \end{aligned}$$

The first term is easy; by the Riemann-Lebesgue Lemma (see the example after Theorem 8.12), since $f(x) - f(x-t)$ is Riemann integrable, we have that the first term goes to 0 as $N \to \infty$. For the second term, we show that $(f(x) - f(x-t))\cot\left(\frac{1}{2}t\right)$ is Riemann-integrable and then use the Riemann-Lebesegue Lemma again. We have that:

$$(f(x) - f(x-t))\cot\left(\frac{1}{2}t\right) = \frac{f(x) - f(x-t)}{t}\frac{\frac{t}{2}}{\sin\left(\frac{t}{2}\right)}2\cos\left(\frac{t}{2}\right)$$

where $\frac{f(x)-f(x-t)}{t}$ is bounded by M by the Lipschitz continuity assumption at x, $\frac{\frac{t}{2}}{\sin\left(\frac{t}{2}\right)} \to 1$ as $t \to 0$ and is bounded and continuous, and $2\cos\left(\frac{t}{2}\right)$ is of course continuous. Hence, $(f(x) - f(x-t))\cot\left(\frac{1}{2}t\right)$ is Riemann Integrable, and we conclude that the second term also goes to 0 as $N \to \infty$. Hence $\lim_{N\to\infty} f(x) - s_N(x) = 0$. □

Theorem

The partial sum s_N is linear in f.

Proof

Recall the definition that $s_N(f;x) = \sum_{n=-N}^{N} \langle f, \phi_n \rangle \phi_n(x)$. Using the linearity of the inner product in the first argument, we then have that:

$$\begin{aligned} s_N(af+bg;x) &= \sum_{n=-N}^{N} \langle af+bg, \phi_n \rangle \phi_n(x) \\ &= a \sum_{n=-N}^{N} \langle f, \phi_n \rangle + b \sum_{n=-N}^{N} \langle g, \phi_n \rangle \phi_n(x) \\ &= a s_N(f;x) + b s_N(g;x) \end{aligned}$$

so s_N is linear in f as claimed. □

With linearity established, we give a corollary of Theorem 8.14.

Corollary

(a) If $f(x) = 0$ for all $x \in (x_0 - \epsilon, x_0 + \epsilon)$ then $s_N(f;x) \to 0$ for all such x (the Theorem can be immediately applied as $|f(x+t) - f(x)| < M|t|$ is satisfied for all constant x).

(b) If $f(x) = g(x)$ for all $x \in (x_0 - \epsilon, x_0 + \epsilon)$ then by the linearity of s_N in f we have that $s_N(f;x) - s_N(g;x) \to 0$ for all such x.

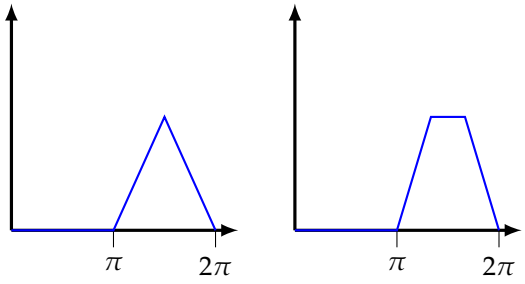

Figure 58: Two functions to which we can apply the above corollary. The two functions have different Fourier series, but both series converge to zero on $[0, \pi]$. This is very different behaviour from power series, see for example Theorem 8.5. This corollary/example shows off the "localization principle" of Fourier series.

Theorem 8.15

If f is continuous and has period 2π, then for all $\epsilon > 0$ there exists a trigonometric polynomial $P(x) = \sum_{n=-N}^{N} a_n e^{inx}$ such that $\sup_{x \in \mathbb{R}} |f(x) - P(x)| < \epsilon$.

Note that the above theorem does *not* imply that Fourier series converge uniformly.

Proof

Let $T = \{z \in \mathbb{C} : |z| = 1\}$ (i.e. the unit circle in the complex plane). T is compact. Define $F : T \mapsto \mathbb{C}$ by $F(e^{ix}) = F(x)$. By the periodicity of f, F is well defined. Let $\mathcal{A}$ be the set of trig polynomials $\sum_{n=-N}^{N} a_n z^n$ with $z \in T$ and $a_n \in \mathbb{C}$. We show that $\mathcal{A}$ satisfies the conditions for the (complex) Stone-Weierstrass theorem to be applied. $\mathcal{A}$ is closed under addition and multiplication (as is easily verified by condidering the sum/products of finite sums). $\mathcal{A}$ vanishes at no point in T as $1 \in \mathcal{A}$. $\mathcal{A}$ separates points in T as $f(x) = x \in \mathcal{A}$. $\mathcal{A}$ is self-adjoint as if $\sum_{n=-N}^{N} a_n z_n \in \mathcal{A}$:

$$\overline{\sum_{n=-N}^{N} a_n z_n,} = \sum_{n=-N}^{N} \overline{a_n} \overline{z}^n = \sum_{n=-N}^{N} \overline{a_n} z^{-n} = \sum_{m=-N}^{N} \overline{a_m} z^m \in \mathcal{A}$$

where we let $m = -n$. Hence, by the Stone-Weierstrass theorem (Theorem 7.33) there exists $P \in \mathcal{A}$ such that $\left|F(z) - P(z)\right| < \epsilon$ for all $z \in T$. Write $P(z) = \sum_{n=-N}^{N} a_n z^n$ and set $P(x) = P(e^{ix}) = \sum_{n=-N}^{N} a_n e^{inx}$. Then, we have that $\left|f(x) - p(x)\right| = \left|F(e^{ix}) - P(e^{ix})\right| < \epsilon$. for all $x \in \mathbb{R}$. □

Note that problem 8.15 gives an explicit sequence of trigonometric polynomials that converge uniformly to f.

As a point of notation for the next theorem, for $\{c_n\}_{n\in\mathbb{Z}}$ and $\{\gamma_n\}_{n\in\mathbb{Z}}$, let $(c, \gamma) = \sum_{n=-N}^{N} c_n \overline{\gamma_n}$. Note that there is no guarantee that this sum converges. Furthermore, let us review the notation that we have already established. $\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx$, $\|f\|_2 = \sqrt{\langle f, f\rangle}$ (when the integral converges), $\phi_n(x) = e^{inx}$. $s_N(f; x) = \sum_{n=-N}^{N} \langle f, \phi_n\rangle \phi_n$.

Theorem: Cauchy Shwartz Inequality for Norms (Problem 6.10)

$|\langle f, g\rangle| \leq \|f\|_2 \|g\|_2$.

Theorem: Minkowski Inequality (Problem 6.11)

$\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$, and $\|f - g\|_2 \leq \|f - h\|_2 + \|h - g\|_2$.

Theorem (Problem 6.12)

For $f \in \mathcal{R}[-\pi, \pi]$ and $\epsilon > 0$, there exists a continuous (in fact, piecewise linear) function h such that $\|f - h\|_2 < \epsilon$.

Proof

Left as an exercise. Solutions to the problems can be found at `https://minds.wisconsin.edu/bitstream/handle/1793/67009/rudin%20ch%206.pdf?sequence=6&isAllowed=y`. □

Theorem 8.16: Parseval's Relation & The Bessel Equality

For $f, g \in \mathcal{R}[-\pi, \pi]$, let $c_n = \langle f, \phi_n\rangle$ and $\gamma_n = \langle g, \phi_n\rangle$. Then $\lim_{N\to\infty} \|f - s_N(t)\| = 0$ (Convergence of partial Fourier series to F in L_2). Furthermore, $\langle f, g\rangle = (c, \gamma)$ (Parseval's Relation) and in particular $\|f\|_2^2 = (c, c)$. I.e. $\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |c_n|^2$ (Bessel Equality).

Proof

Let $\epsilon > 0$. Choose a continuous h such that $\|f - h\|_2 < \frac{\epsilon}{3}$ (Problem 6.12). Then:

$$\|s_N(f;x) - f\|_2 \leq \|s_N(f;x) - s_N(h;x)\|_2 + \|s_N(h;x) - h\|_2 + \|h - f\|_2.$$

We have that the third term is less than $\frac{\epsilon}{3}$ by assumption. For the first term, by the linearity of s_N we have that $\|s_N(f;x) - s_N(h;x)\| = \|s_N(f-h)\|_2 \leq \|f-h\|_2 < \frac{\epsilon}{3}$ (using Bessel's Inequality). For the second term, by Theorem 8.15 we have that there exists a trigonometric polynomial P such that $\|h - P\|_\infty < \frac{\epsilon}{3}$. But, $\|h-P\|_2 \leq \|h-P\|_\infty < \frac{\epsilon}{3}$. Say $\deg P = N_0$. By Theorem 8.11, if $N \geq N_0$ then:

$$\|s_N(h;x) - h\|_2 \leq \|p - h\|_2 < \frac{\epsilon}{3}$$

as "the best L_2 approximation of f by $\sum_{n=-N}^{N} a_n \phi_n$ is $s_N(f;x)$". But then we have that:

$$\|s_N(f;x) - s_N(h;x)\| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

if $N \geq N_0$ so $\lim_{N\to\infty} \|s_N(f;x) - f\|_2 = 0$. This proves the first claim.
For Parseval's Relation, we observe that $\langle s_N(f;x), g\rangle = \sum_{n=-N}^{N} c_n \langle \phi_n, g\rangle = \sum_{n=-N}^{N} c_n \overline{\gamma_n}$. So:

$$\left| \langle f, g\rangle - \sum_{n=-N}^{N} c_n \overline{\gamma_n} \right| = |\langle f - s_N(f;x), g\rangle| \leq \|f - s_N(f;x)\|_2 \|g\|_2$$

where we use the Cauchy-Shwartz inequality for the last inequality. We then have that $\lim_{N\to\infty} \|s_N(f;x) - f\|_2 = 0$ by the previous part of the theorem, proving the Parseval relation. To obtain the Bessel equality, let $g = f$ in Parseval's relation.
We expand on a detail in the proof; we show that $\sum_{n=-\infty}^{\infty} |c_n \overline{\gamma_n}|$ converges, as if we can show absolute convergence then taking the $n \to \infty$ limit of $\sum_{n=-N}^{N} c_n \overline{\gamma_n}$ is justified. To see that this is the case, observe that:

$$\sum_{n=-\infty}^{\infty} |c_n \overline{\gamma_n}| = (|c_n|, |\gamma_n|) \leq \sqrt{(c,c)}\sqrt{(\gamma,\gamma)} \leq \|f\|_2 \|g\|_2.$$

Hence, since $\sum_{n=-N}^{N} |c_n \overline{\gamma_n}|$ is bounded and monotonic in N, the series converges and the infinite sum is equal to the symmetric limit; that is, the absolute convergenceof the sum implies that $\sum_{n=0}^{\infty} c_n \overline{\gamma_n}$ and $\sum_{n=0}^{-\infty} c_n \overline{\gamma_n}$ converge individually. □

9 Functions of Several Variables

9.1 Banach Fixed Point Theorem

Our goal in this chapter will be to work up to the Inverse Function Theorem. This chapter in Rudin begins by covering the necessary results in linear algebra; we will assume this has been covered in a prior course, so we will omit discussion of items 9.1-9.9. However, these can be read for a refresher of the material.

We will start off this chapter with discussion of the Banach fixed point theorem (also known as the Contraction principle), as it is independent of the rest of the chapter's content. It will be used in the proof of the inverse function theorem, but also applies in a more general setting.

Definition 9.22: Contractions

Let (X, d) be a metric space. Suppose there exists $c < 1$ such that the map $\phi : X \mapsto X$ satisfies $d(\phi(x), \phi(y)) \leq cd(x, y)$ for all $x, y \in X$ (that is, the images of x, y are closer by a factor c compared to the original x, y). Then, we call ϕ a **contraction** of X into X.

Lemma

Every contraction $\phi : X \mapsto X$ is uniformly continuous.

Proof

Take $\delta = \frac{\epsilon}{c}$ in the definition of uniform continuity. □

Theorem 9.23: Banach Fixed Point Theorem/Contraction Mapping Theorem

Let (X, d) be a complete metric space, and suppose $\phi : X \mapsto X$ is a contraction. Then, there exists a unique $x \in X$ such that $\phi(x) = x$. We call this x a *fixed point*.

The proof of the above theorem gives an algorithm to find x which converges exponentially fast.

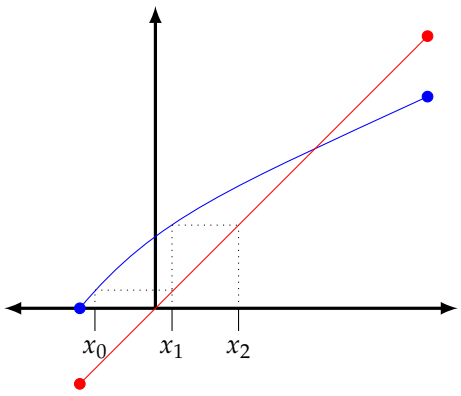

Figure 59: Visualization of the algorithm for finding the fixed point x in Theorem 9.23 for the case where $X = \mathbb{R}$. The fact that ϕ is a contraction makes it such that it has slope less than 1. The fixed point is the point of intersection between $y = x$ and $y = \phi(x)$. The iterative algorithm sketched above gives an exponentially fast way of finding this point of intersection, by iteratively applying ϕ to the initial guess x_0.

Proof

We first show uniqueness. If $\phi(x) = x$ and $\phi(y) = y$, then $d(\phi(x), \phi(y)) \leq cd(x, y)$. But since $c < 1$, $d(x, y) \leq cd(x, y)$ is only satisfied if $d(x, y) = 0$. Hence, $x = y$ and the fixed point is unique.
We next show existence. Given $x_0 \in X$, let $x_1 = \phi(x_0)$, $x_2 = \phi(x_1) = \phi \circ \phi(x_0)$, $x_3 = \phi(x_2) = \phi \circ \phi \circ \phi(x_0)$ and so on, with $x_{n+1} = \phi(x_n) = \phi \circ \ldots \circ \phi(x_0)$ with the composition carried out $n+1$ times. The goal is to show this sequence is Cauchy, and has a limit which is a fixed point. We then have that $d(x_{n+1}, x_{n+2}) = d(\phi(x_n), \phi(x_{n+1})) \leq cd(x_{n+1}, x_n)$, so by induction, it follows that $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$. Hence, for $n > m$ we have that:

$$\begin{aligned} d(x_n, x_m) &\leq \sum_{i=m+1}^{n} d(x_i, x_{i-1}) && \text{(Triangle Inequality)} \\ &\leq \sum_{i=m+1}^{n} c^{i-1} d(x_1, x_0) \\ &\leq \sum_{i=m+1}^{\infty} c^{i-1} d(x_1, x_0) \\ &= \frac{c^m}{1-c} d(x_1, x_0) && \text{(Convergent Geometric Series)} \end{aligned}$$

c^m can be as small as we like by taking sufficiently large m, so $\{x_n\}$ is Cauchy, and by the completeness of X, $x_n \to x$ for some x. Since ϕ is a contraction, it is (uniformly) continuous by the above Lemma, and since $x_n \to x$, we have that $\phi(x_n) \to \phi(x)$, and hence:

$$\phi(x) = \lim_{n \to \infty} \phi(x_n) = \lim_{n \to \infty} x_{n+1} = x.$$

This shows that x is the desired fixed point. □

9.2 Differentiation of Functions of Several Variables

In this section we discuss the differentiation of functions $f : \mathbb{R}^n \mapsto \mathbb{R}^m$. It will be instructive to remind ourselves of the familiar case of $n = m = 1$. In this case, we defined the derivative as:

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

where we would say f was differentiable at x if the above limit existed. We want to now try to generalize this notion to higher dimensions. It obviously does not apply directly, as in this general setting, $\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x})$ is a vector in $\mathbb{R}^m$ and $\mathbf{h}$ is a vector in $\mathbb{R}^n$, and the division of such vectors is not well defined. Let us try to recast the $n = m = 1$ derivative into a form that lends itself better to generalization. We could equivalently write the above derivative expression as:

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

and hence:

$$\lim_{h \to 0} \left| \frac{f(x+h) - f(x) - f'(x)h}{h} \right| = 0.$$

Equivalently, we can say that $f(x+h) = f(x) + f'(x)h + r(h)$ with $\lim_{h \to \infty} \frac{r(h)}{h} = 0$; $r(h)$ is then the "remainder term". With this interpretation, $f'(x)h$ is the best linear approximation to $f(x+h) - f(x)$.

This last characterization makes sense for general vectors in $\mathbb{R}^k$, so let us define derivatives with this notion!

Definition 9.11: Derivatives of $f : \mathbb{R}^n \mapsto \mathbb{R}^n$

Let $m, n \in \mathbb{N}$ and $E \subset \mathbb{R}^n$ be open. Define a function $\mathbf{f}$ such that $\mathbf{f} : E \mapsto \mathbb{R}^m$, and let $\mathbf{x} \in E$. We say that f is **differentiable** at $\mathbf{x}$ and has **derivative** $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ ($\mathbf{f}'(\mathbf{x}) = A$) if:

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0$$

Note that the $||$ in the above definition refer to the L_2/Euclidean norm of $\mathbb{R}^m$ (in the numerator) and $\mathbb{R}^m$ (in the denominator) respectively. Furthermore, observe that A is not just a number (as it is in the linear case) but an *linear transformation* such that $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, where $L(\mathbb{R}^n, \mathbb{R}^m)$ is the vector space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. Equivalently, the above definition can be phrased as:

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + A\mathbf{h} + \mathbf{r}(\mathbf{h}) \text{ with } \lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0$$

where again, $A\mathbf{h}$ is the "best linear approximation" to $\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})$.

Theorem 9.12: Uniqueness of Higher Dimensional Derivatives

The derivative A defined above is unique.

Proof

Suppose $\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + A_1\mathbf{h} + \mathbf{r}_1(\mathbf{h})$ and $\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + A_2\mathbf{h} + \mathbf{r}_2(\mathbf{h})$ with $\mathbf{r}_1(\mathbf{h}) \in O(\mathbf{h})$ and $\mathbf{r}_2(\mathbf{h}) \in O(\mathbf{h})$. Let $B = A_1 - A_2$. Then, $0 = (A_1 - A_2)\mathbf{h} + (\mathbf{r}_1 - \mathbf{r}_2)$. Therefore, $B\mathbf{h} = \mathbf{r}_2 - \mathbf{r}_1$, so for $\mathbf{h} \neq \mathbf{0}$ and scalar $t \neq 0$ we have that:

$$\frac{|B\mathbf{h}|}{|\mathbf{h}|} = \frac{|B(t\mathbf{h})|}{|t\mathbf{h}|} \leq \frac{|\mathbf{r}_1(t\mathbf{h})|}{|t\mathbf{h}|} + \frac{|\mathbf{r}_2(t\mathbf{h})|}{|t\mathbf{h}|} \quad \text{(Triangle Inequality)}$$

Letting t go to zero, we have that the RHS goes to zero. Hence, $\frac{|B\mathbf{h}|}{|\mathbf{h}|} \to 0$. But $\frac{|B\mathbf{h}|}{|\mathbf{h}|}$ is independent of t, so it must be zero. Hence, B is the zero map and hence $A_1 = A_2$. We conclude that the derivative A is unique. □

Note that if $f'(x)$ exists for all x in E, then we can regard $\mathbf{f}'$ as a function $\mathbf{f}' : E \mapsto L(\mathbb{R}^n, \mathbb{R}^m)$. Let us clarify that $\mathbf{f}'$ is *not* a vector valued function, but a linear map; we use the notation $\mathbf{f}'$ to denote it as the derivative of a vector valued function.

Example 9.14

Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is linear. Then:

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{h}) + 0$$

for all $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$. Hence, $\mathbf{f}'(x) = \mathbf{f}$ for all $\mathbf{x}, \mathbf{h} \in \mathbb{R}^n$. The derivative is the function itself (*not* the value $\mathbf{f}(\mathbf{x})$) when $\mathbf{f}$ is linear.
Note that for $n = m = 1$, we can identify a linear map $f : \mathbb{R} \mapsto \mathbb{R}$ as $f(x) = \alpha x$, with $\alpha \in \mathbb{R}$. Then, $f'(x) = \alpha$ is consistent with the above general result.

Theorem

If $\mathbf{f}'(\mathbf{x})$ exists, then $\mathbf{f}$ is continuous at $\mathbf{x}$.

Proof

$\lim_{\mathbf{h}\to 0} \mathbf{f}(\mathbf{x}+\mathbf{h}) = \lim_{\mathbf{h}\to 0} \left(\mathbf{f}(x) + \mathbf{f}'(x)\mathbf{h} + O(\mathbf{h})\right) = \mathbf{f}(x)$.

Definition: Norm of a linear map

Let $A : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear map. We then define:

$$\|A\| = \sup\left\{|A\mathbf{x}| : |\mathbf{x}| \leq 1\right\}.$$

In other words, we have that $|A\mathbf{x}| = \left|A\frac{\mathbf{x}}{|\mathbf{x}|}\right||\mathbf{x}| \leq \|A\|\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. $\|A\|$ is the best constant bound. Before moving onto the next theorem, we note a couple facts about $\|A\|$:

(a) $\|A\| < \infty$ for any linear map A.

(b) $\|AB\| \leq \|A\|\|B\|$.

(c) $d(A, B) = \|A - B\|$, which defines a metric on the space of linear maps.

Theorem 9.15: Chain Rule

Suppose that $E \subset \mathbb{R}^n$, E is open, and $\mathbf{f} : E \mapsto \mathbb{R}^m$. Furthermore, suppose $V \subset \mathbb{R}^m$ is open and $\mathbf{f}(E) \subset V$. Define $\mathbf{g} : V \mapsto \mathbb{R}^k$, and suppose $\mathbf{f}'(\mathbf{x}_0)$ exists and $\mathbf{g}'(\mathbf{f}'(\mathbf{x}_0))$ exists for some $\mathbf{x}_0 \in E$. Let $\mathbf{F} = \mathbf{g} \circ \mathbf{f} : E \mapsto \mathbb{R}^k$. Then, we have that $\mathbf{F}'(\mathbf{x}_0)$ exists, and $\mathbf{F}'(\mathbf{x}_0) = \mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0) \in L(\mathbb{R}^n, \mathbb{R}^k)$.

As a point of clarification, note that $\mathbf{g}'(f(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0)$ denotes a composition of linear operators, where $\mathbf{f}' :\in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbf{g}' \in L(\mathbb{R}^m, \mathbb{R}^k)$.

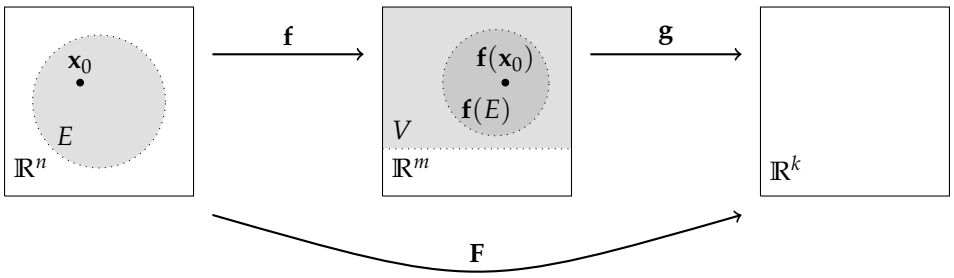

Figure 60: Visualization of the sets and functions in Theorem 9.15.

Proof

The proof is very similar to the one dimensional case. We observe that:

$$\mathbf{F}(\mathbf{x}_0+\mathbf{h})-\mathbf{F}(\mathbf{x}_0)=\mathbf{g}(\mathbf{f}(\mathbf{x}_0+\mathbf{h}))-\mathbf{g}(\mathbf{f}(\mathbf{x}_0)).$$

We can then write $\mathbf{f}(\mathbf{x}_0+\mathbf{h})$ as $\mathbf{f}(\mathbf{x}_0)+\mathbf{K}$ where $\mathbf{K}=\mathbf{f}'(\mathbf{x}_0)\mathbf{h}+\mathbf{u}(\mathbf{h})$ with $\mathbf{u}(\mathbf{h})\in O(\mathbf{h})$. We can then write:

$$\begin{aligned}\mathbf{F}(\mathbf{x}_0+\mathbf{h})-\mathbf{F}(\mathbf{x}_0)&=\mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{K}+\mathbf{V}(\mathbf{K})\\&=\mathbf{g}'(\mathbf{f}(x_0))\left[\mathbf{f}'(x_0)\mathbf{h}+\mathbf{u}(\mathbf{h})\right]+\mathbf{V}(\mathbf{K})\\&=\mathbf{g}'(\mathbf{f}(x_0))\mathbf{f}'(x_0)\mathbf{h}+\mathbf{g}'(\mathbf{f}(x_0))\mathbf{u}(\mathbf{h})+\mathbf{V}(\mathbf{K})\end{aligned}$$

where $\mathbf{g}'(\mathbf{f}(x_0))\mathbf{u}(\mathbf{h})+\mathbf{V}(\mathbf{K})=\mathbf{r}(\mathbf{h})$ is our remainder term. We want to prove that $\mathbf{r}(\mathbf{h})\in O(\mathbf{h})$, i.e. that $\frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|}\to 0$ as $\mathbf{h}\to\mathbf{0}$. To this end, we observe that:

$$\frac{\left|\mathbf{g}'(\mathbf{f}(x_0))\mathbf{u}(\mathbf{h})\right|}{\mathbf{h}}\leq\left\|\mathbf{g}'(\mathbf{f}(\mathbf{x_0}))\right\|\frac{|\mathbf{u}(\mathbf{h})|}{|\mathbf{h}|}$$

where we have that $\left\|\mathbf{g}'(\mathbf{f}(\mathbf{x_0}))\right\|<\infty$ and $\frac{|\mathbf{u}(\mathbf{h})|}{|\mathbf{h}|}\to 0$ as $\mathbf{h}\to\mathbf{0}$ as $\mathbf{u}(\mathbf{h})\in O(\mathbf{h})$. Furthermore, let $\boldsymbol{\eta}(\mathbf{K})=\frac{|\mathbf{V}(\mathbf{K})|}{|\mathbf{K}|}$ and then we have that:

$$\frac{|\mathbf{V}(\mathbf{K})|}{\mathbf{h}}\frac{|\mathbf{K}|}{|\mathbf{K}|}=\boldsymbol{\eta}(\mathbf{K})\frac{|\mathbf{K}|}{\mathbf{h}}\leq\boldsymbol{\eta}(\mathbf{K})\left(\frac{|\mathbf{f}'(\mathbf{x_0})\mathbf{h}|}{|\mathbf{h}|}+\frac{|\mathbf{u}(\mathbf{h})|}{|\mathbf{h}|}\right)\leq\boldsymbol{\eta}(\mathbf{K})\left(\left\|\mathbf{f}'(\mathbf{x}_0)\right\|\frac{|\mathbf{h}|}{|\mathbf{h}|}+\frac{|\mathbf{u}(\mathbf{h})|}{|\mathbf{h}|}\right)$$

where we have that $\left\|\mathbf{f}'(\mathbf{x}_0)\right\|<\infty$ and $\frac{|\mathbf{u}(\mathbf{h})|}{|\mathbf{h}|}\to 0$ as $\mathbf{h}\to\mathbf{0}$. Furthermore, Since $\mathbf{K}\to\mathbf{0}$ as $\mathbf{h}\to\mathbf{0}$ (since $\mathbf{K}=\mathbf{f}'(\mathbf{x}_0)+\mathbf{u}(\mathbf{h})$ and $\mathbf{f}'(\mathbf{x}_0)\to 0$ as $\mathbf{f}'(\mathbf{x}_0)$ is continuous and $\mathbf{u}(\mathbf{h})\in O(\mathbf{h})$, we have that $\boldsymbol{\eta}(\mathbf{K})\to\mathbf{0}$. Hence, $\mathbf{r}(\mathbf{h})\in O(\mathbf{h})$, so we conclude that:

$$\lim_{h\to\mathbf{0}}\frac{\left|\mathbf{F}(\mathbf{x_0}-\mathbf{h})-\mathbf{F}(\mathbf{x}_0)-\mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0)\mathbf{h}\right|}{|\mathbf{h}|}=0$$

□

Next, we recall the Jacobian matrix/determinant that was introduced in second year multivariable calculus. Is it equivalent to what we have defined here? Well, not quite; $\mathbf{F}'(\mathbf{x})$ is not a matrix, but a linear operator. However, linear operators do have a matrix representation. Up until now, we have done things basis-agnostically, but we will now start to work in particular bases to probe this question further.

Definition: Canonical Basis

The **canonical bases** (also known as the standard basis) of $\mathbb{R}^n$ and $\mathbb{R}^m$ are defined to be $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$ and $\{\mathbf{u}_1,\dots\mathbf{u}_m\}$ where $\mathbf{e}_1=(1,0,\dots 0)$ (length n) and $\mathbf{u}_1=(1,0,\dots,0)$ (length m).

Definition: Function Components

Let $\mathbf{f}:\mathbb{R}^n\mapsto\mathbb{R}^m$. We can then write $\mathbf{f}(\mathbf{x})=s\sum_{i=1}^m f_i(\mathbf{x})\mathbf{u}_i$. So, $\mathbf{f}=(f_1,\dots,f_n)$ where f_i is the ith **component** of $\mathbf{f}$. Hence, $f_i:\mathbb{R}^n\mapsto\mathbb{R}$.

Definition 9.16: Partial Derivatives

We define the **partial derivative** as

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) = (D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t}$$

where $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

We observe that $D_j f_i$ is just the derivative of f_i with respect to x_j.

Theorem 9.17

If $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^m$ is differentiable at $\mathbf{x} \in \mathbb{R}^n$, then $(D_j f_i)(\mathbf{x})$ exists for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, and:

$$[\mathbf{f}'(\mathbf{x})] = \begin{bmatrix} (D_1 f_1)(\mathbf{x}) & \cdots & (D_n f_1)(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ (D_1 f_m)(\mathbf{x}) & \cdots & (D_n f_m)(\mathbf{x}) \end{bmatrix}.$$

In other words, we have $\mathbf{f}'(\mathbf{x})\mathbf{e}_j = \sum_{i=1} (D_i f_i)(\mathbf{x})\mathbf{u}_i$, and hence $[\mathbf{f}'(\mathbf{x})]_{ij} = (D_j f_i)(\mathbf{x})$.

Proof

We want to show that the linear operator $A = \mathbf{f}'(\mathbf{x})$ has matrix elements:

$$\mathbf{u}_i A \mathbf{e}_j = A_{ij} = \frac{\partial f_i}{\partial x_j}.$$

In other words, that $f_i(\mathbf{x} + t\mathbf{e}_j) = f_i(\mathbf{x}) + A_{ij}t + O(t)$, which identifies A_{ij} as $\frac{\partial f_i}{\partial x_j}$. To this end, let $\epsilon(t) = f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x}) - A_{ij}t$. We then have that:

$$\begin{aligned}
|\epsilon(t)| &= \left|\mathbf{u}_i(\mathbf{f}(x + t\mathbf{e}_j) - \mathbf{f}(x) - A(t\mathbf{e}_j))\right| && \text{(Linearity to say } A_{ij}t = A(t\mathbf{e}_j)) \\
&\leq \left|\mathbf{u}_j\right|\left|\mathbf{f}(x + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x}) - A(t\mathbf{e}_j)\right| && \text{(Cauchy-Shwartz inequality on } \mathbb{R}^n) \\
&= \left|\mathbf{f}(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x}) - A(t\mathbf{e}_j)\right| && \mathbf{u}_i \text{ is of unit length} \\
&\in O(t\mathbf{e}_j) \text{ and therefore } O(t) && \text{(Definition of derivative)}
\end{aligned}$$

which proves the claim. □

As a remark, note that if $\mathbf{f}$ is differentiable at $\mathbf{x}$, then the above theorem implies that all partial derivatives exist at that point. The converse is *not* true (see the discussion on page 215). But, the converse is true with an extra additional condition that the partial derivatives are continuous.

Definition: Gradient

Let $E \subset \mathbb{R}^n$ be open and $f : E \mapsto \mathbb{R}$. Suppose all partial derivatives of f exist. Then, we define the **gradient** of f to be $\nabla f : E \mapsto \mathbb{R}^n$, where:

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x})\mathbf{e}_i = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right).$$

Note that this is the matrix representation of $f'(\mathbf{x})$ for real valued f. In particular, for $\mathbf{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$, we have that:

$$f'(\mathbf{x})\mathbf{h} = \nabla f \cdot \mathbf{h} = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

By the chain rule, for $\mathbf{u} \in \mathbb{R}^n$ with $|\mathbf{u}| = 1$, we have that:

$$\lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t} = \frac{\mathrm{d}}{\mathrm{d}t}_{t=0} f(\mathbf{x} + t\mathbf{u}) = f'(\mathbf{x})\mathbf{u} = \nabla f \cdot \mathbf{u}$$

We use this to define the directional derivative:

Definition: Directional Derivatives

The **directional derivative** of f in direction $\mathbf{u}$ (where $\mathbf{u}$ is a unit length vector in $\mathbb{R}^n$) at $\mathbf{x}$, denoted as $(D_{\mathbf{u}} f)(\mathbf{x})$, is defined as:

$$(D_{\mathbf{u}} f)(\mathbf{x}) = \nabla f \cdot \mathbf{u}$$

We note that $(D_{\mathbf{u}} f)(\mathbf{x})$ is maximal when $\mathbf{u} \parallel \nabla f(\mathbf{x})$; hence, $\nabla f(\mathbf{x})$ points in the direction of maximal increase of f at $\mathbf{x}$. This illustrated in the example below.

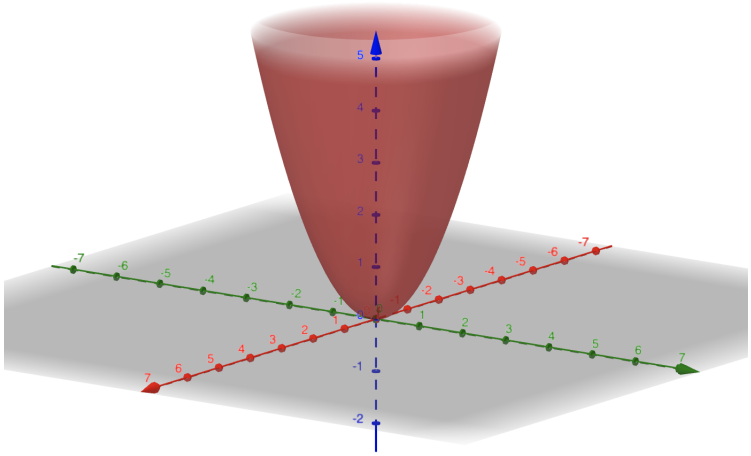

Figure 61: Visualized is the surface $z = f(x, y) = x^2 + y^2$. At any given (x, y), we have that $\nabla f = (2x, 2y)$. The direction of ∇f gives the direction for which f has a maximal rate of increase at (x, y).

Definition: Convex Sets

Let $E \subset \mathbb{R}^n$. We say that E is **convex** if for all $\mathbf{x}, \mathbf{y} \in E$ and for all $\lambda \in [0,1]$, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in E$.

Geometrically, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ represents the points on the line segment that joins $\mathbf{x}, \mathbf{y}$. In this picture, convexity means that for any two points in a set, all points in the line segment that joins them is also in the set. This turns out to be a useful notion.

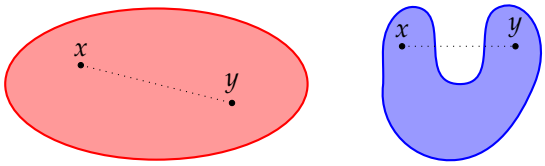

Figure 62: A picture of a convex set (left) and non-convex set (right) in $\mathbb{R}^2$.

Theorem 9.19

Let $E \subset \mathbb{R}^n$ be convex and $\mathbf{f} : E \mapsto \mathbb{R}^m$. Suppose $\mathbf{f}$ is differentiable on E and $\|\mathbf{f}'(\mathbf{x})\| \leq M$ for all $x \in E$. Then, for all $\mathbf{a}, \mathbf{b} \in E$, we have that $|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$.

Proof

By convexity, component-wise integration (Rudin 6.23) and the FTC applied component-wise (Rudin 6.24) we have that:

$$\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a}) = \int_0^1 \mathbf{f}((1-t)\mathbf{a} + t\mathbf{b})dt.$$

By the chain rule (Theorem 9.15), we then have that:

$$\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a}) = \int_0^1 \mathbf{f}'((1-t)\mathbf{a} + t\mathbf{b})(\mathbf{b} - \mathbf{a})dt.$$

Applying the multidimensional ML bound (Rudin 6.25), we have that:

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq \int_0^1 \left|\mathbf{f}'((1-t)\mathbf{a} + t\mathbf{b})(\mathbf{b} - \mathbf{a})\right|dt$$

but then using the bound on $\|\mathbf{f}'(\mathbf{x})\|$ we have that:

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$$

which is exactly what we wanted to show. □

Corollary

If $\mathbf{f}'(\mathbf{x}) = 0$ for all $\mathbf{x} \in E$, then $\mathbf{f}$ is constant on E.

Note that convexity of E is not actually needed for this Corollary; we can instead consider a finite sequence of closed disks and apply Theorem 9.19 to each of them.

Definition 9.20: : Continuous Differentiability

Let $E \subset \mathbb{R}^n$ be open. $\mathbf{f} : E \mapsto \mathbb{R}^m$ is **continuously differentiable** on E (denoted $\mathbf{f} \in C^1(E)$) if $\mathbf{f}'(\mathbf{x})$ exists for every $\mathbf{x} \in E$ and $\mathbf{f}'$ is continuous on E.

Note that (as always), $\mathbf{f}' : E \mapsto L(\mathbb{R}^n, \mathbb{R}^m)$. To phrase the above definition another way, for all $\mathbf{x} \in E$ and for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{y} \in E$ that satisfy $|\mathbf{x} - \mathbf{y}| < \delta$, we have that $\left\|\mathbf{f}'(\mathbf{x}) - \mathbf{f}'(\mathbf{y})\right\| < \epsilon$.

Theorem 9.21

Suppose $\mathbf{f} : E \mapsto \mathbb{R}^m$ where $E \subset \mathbb{R}^n$ is open. Then, $\mathbf{f} \in C^1(E)$ if and only if $\frac{\partial f_i}{\partial x_j}$ exist for all $i \in \{1, \dots m\}, j \in \{1, \dots, n\}$ and all are continuous.

Proof

Not covered in lecture, see Rudin. □

Example

Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ where:

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}.$$

Calculating the partial derivatives of f at $(0,0)$ with respect to x and y, we have:

$$\left.\frac{\partial f}{\partial x}\right|_{(0,0)} = \lim_{t\to 0} \frac{f(0+t,0) - f(0,0)}{t} = \lim_{t\to 0} \frac{0-0}{t} = 0$$
$$\left.\frac{\partial f}{\partial y}\right|_{(0,0)} = \lim_{t\to 0} \frac{f(0,0+t) - f(0,0)}{t} = \lim_{t\to 0} \frac{0-0}{t} = 0.$$

However, f is not even continuous at $(0,0)$; to see this, approach $(0,0)$ by the line (t, at):

$$\lim_{t\to 0} f(t,at) = \lim_{t\to 0} \frac{at^2}{t^2 + a^2t^2} = \lim_{t\to 0} \frac{a}{1+a^2} = \frac{a}{1+a^2} \neq 0 = f(0,0).$$

As f is not continuous at $(0,0)$, it is therefore not differentiable at $(0,0)$. Note that f *is* continuous differentiable on $\mathbb{R}^2 \setminus \{0\}$. Also note that if we try to compute the derivative along the ray (t, at), we get ∞:

$$\lim_{t\to 0} \frac{f(t,at) - f(0,0)}{t} = \lim_{t\to 0} \frac{1}{t} \frac{a}{1+a^2} = \infty.$$

This example shows the partial derivatives are not continuous everywhere as f is not continuous differentiable. $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist everywhere, but this itself does not imply much.

9.3 The Inverse Function Theorem

Before we give the statement and proof of the Inverse function theorem, it is instructive to revist the one-dimensional case.

Theorem (Problem 5.2)

Let $f : [a, b] \mapsto \mathbb{R}$. Suppose f' exists for all (a, b), and suppose $f'(x) > 0$ (or alternatively $f'(x) < 0$) for all $x \in (a, b)$. Then, f is strictly increasing, and has an inverse function g. This g is differentiable and has derivative $g'(f(x)) = \frac{1}{f'(x)}$.

Proof

Left as an exercise. Solution can be found at `https://minds.wisconsin.edu/bitstream/handle/1793/67009/rudin%20ch%205.pdf?sequence=7&isAllowed=y`.

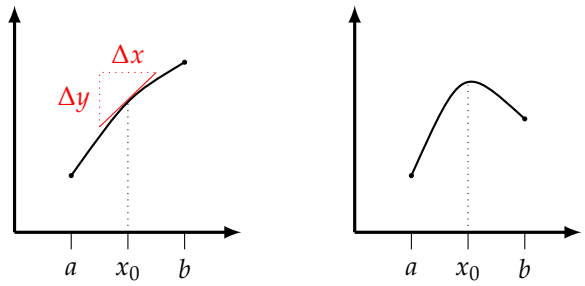

Figure 63: Example demonstrating the above theorem. For the left function, we have that $g = f^{-1}$ exists, and $g'(f(x_0)) = \frac{1}{f'(x_0)} = \frac{1}{\frac{\Delta y}{\Delta x}} = \frac{\Delta x}{\Delta y}$. For the left function, we have that $f'(x_0) = 0$, so no inverse exists near this point (the inverse does not exist on $(x_0 - \epsilon, x_0 + \epsilon)$ for some ϵ).

Theorem 9.24: The Inverse Function Theorem

Suppose $\mathbf{f} : E \mapsto \mathbb{R}^n$ where $E \subset \mathbb{R}^n$ is open. Suppose $\mathbf{f} \in C^1(E)$, $\mathbf{a} \in E$, and $\mathbf{f}'(\mathbf{a})$ is invertible (c.f. $f'(a) \neq 0$ for the $n = 1$ case). Let $\mathbf{b} = \mathbf{f}(\mathbf{a})$. Then:

(1) There exists an open set U, V with $\mathbf{a} \in U, \mathbf{b} \in V$ such that $\mathbf{f} : U \mapsto V$ is a bijection. Hence, the inverse function $\mathbf{g} = \mathbf{f}^{-1}$ exists with $\mathbf{g} : V \mapsto U$ and $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in U$.

(2) $\mathbf{g} \in C^1(V)$.

Note in the above theorem that both the domain and codomain are n-dimensional; the dimensions must match for an inverse to exist (there would be no inverse for $f : \mathbb{R}^2 \mapsto \mathbb{R}$, for example). Now, a couple more remarks before we move onto the proof.

a) By the chain rule, we have that $\mathbf{g}'(\mathbf{f}(\mathbf{x}))\mathbf{f}'(x) = I$, so $\mathbf{g}'(\mathbf{f}(\mathbf{x})) = \left(\mathbf{f}'(\mathbf{x})\right)^{-1}$ for $\mathbf{x} \in U$.

b) Let us write $\mathbf{y} = \mathbf{f}(\mathbf{x})$ as $\mathbf{y} = (y_1, \ldots, y_n)$ where $y_1 = f_1(x_1, \ldots, x_n)$, $y_2 = f_2(x_1, \ldots, x_n)$ and so on (up to $y_n = f_n(x_1, \ldots x_n)$). Note that $f_1, \ldots f_n$ could (in general) be horrible nonlinear functions. Writing $\mathbf{y}$ in this way, we generate a system of equations. In math, we are often interested in whether a system of equations is solvable. If $\mathbf{f}'(\mathbf{a})$ is indeed invertible, then for $\mathbf{y}$ near $\mathbf{b} = \mathbf{f}'(\mathbf{a})$, there exists a unique C^1 solution $x_1 = g_1(y_1, \ldots y_n), x_2 = g_2(y_1, \ldots, y_n), \ldots, x_n = g_n(y_1, \ldots, y_n)$. $\mathbf{f}$ is a bijection

from $U \mapsto V$, and hence there is an inverse function that gives a unique solution $\mathbf{x}$ to any $\mathbf{y}$. And this function is continuously differentiable!

c) $\mathbf{f} \in C^1$ is assumed. So, the invertibility of $\mathbf{f}'(\mathbf{a})$ is equivalent to the determinant of the matrix representation of the linear operator being non-zero:

$$J\mathbf{f}(\mathbf{a}) = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_1}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix} \neq 0$$

this is because $\mathbf{f}'(\mathbf{a})$ being invertible if and only if the matrix representation is invertible if and only if the matrix representation has nonzero determiant. This determinant is known as the *Jacobian determinant* of $\mathbf{f}$ at $\mathbf{a}$.

Proof of (1)

For the first step, we show that there exists open $U \ni \mathbf{a}$ such that $\mathbf{f}$ is one-to-one on U. Write $A = \mathbf{f}'(\mathbf{a})$. We will now used the Banach fixed point theorem/Theorem 9.23 (Note: This requires some ingenuity). Given $\mathbf{y} \in \mathbb{R}^n$, let $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}))$ (note that $\boldsymbol{\phi}$ depends on $\mathbf{y}$). Then, $\boldsymbol{\phi} = \mathbf{x}$ if and only if $\mathbf{y} = \mathbf{f}(\mathbf{x})$ (as $A^{-1}\mathbf{z} = 0$ if and only if $\mathbf{z} = \mathbf{0}$). Hence, the uniqueness of the fixed point of $\boldsymbol{\phi}$ implies a fixed point of $\mathbf{x}$ such that $\mathbf{y} = \mathbf{f}(\mathbf{x})$, which is the claim. It therefore suffices to show that $\boldsymbol{\phi}$ is a contraction. We will demonstrate this via the derivative. We have that $\boldsymbol{\phi}'(\mathbf{x}) = I - A^{-1}\mathbf{f}'(\mathbf{x})$. We can factor out A^{-1} to write $\boldsymbol{\phi}'(\mathbf{x}) = A^{-1}(\mathbf{f}'(\mathbf{a}) - \mathbf{f}'(\mathbf{x}))$. We therefore have that $\|\boldsymbol{\phi}'(\mathbf{x})\| \leq \left\|A^{-1}\right\| \|\mathbf{f}'(\mathbf{a}) - \mathbf{f}'(\mathbf{x})\|$. Since $\mathbf{f}'$ is continuous, there exists $\delta > 0$ such that $\|\mathbf{f}'(\mathbf{a}) - \mathbf{f}'(\mathbf{x})\| \leq \frac{1}{2\|A^{-1}\|}$ if $|x - a| < \delta$. Hence, $\|\boldsymbol{\phi}'(\mathbf{x})\| \leq \frac{1}{2}$ if $\mathbf{x} \in N(\mathbf{a})$ which identifies the desired set U. For $\mathbf{x}_1, \mathbf{x}_2 \in U$, we have that $|\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_2)| \leq \frac{1}{2}|\mathbf{x}_1 - \mathbf{x}_2|$ by Theorem 9.19. So, $\boldsymbol{\phi}$ is a contraction of U, and hence $\boldsymbol{\phi}$ has at most one fixed point. Therefore, $\mathbf{f}$ is one-to-one on U.
Fo the second step, let $V = \mathbf{f}(U)$. We show that V is open (as this proves (1)). Let $\mathbf{y}_0 \in V$, say, $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ ($\mathbf{x}_0 \in U$). We need to find $\epsilon > 0$ such that $N_\epsilon(\mathbf{y}_0) \subset V$, i.e. for every $\mathbf{y} \in N_\epsilon(\mathbf{y}_0)$, there exists $\mathbf{x} \in U$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. Given $\mathbf{y} \in \mathbb{R}^n$, as before, define $\boldsymbol{\phi}_\mathbf{y} : \mathbb{R}^n \mapsto \mathbb{R}^n$ by $\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}))$ (with $\mathbf{x} \in U$). Note that this neighbourhood of U is *not* complete (unless it happens to be all of $\mathbb{R}^n$). Next, choose $r > 0$ such that $\overline{B} = N_r(\mathbf{x}_0) \subset U$ which is possible as U is open. For $\mathbf{x} \in \overline{B}$, we show that $\left|\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) - \mathbf{x}_0\right| < r$. For this, we observe that:

$$\begin{aligned} \left|\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) - \mathbf{x}_0\right| \leq \left|\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) - \boldsymbol{\phi}_\mathbf{y}(\mathbf{x}_0)\right| + \left|\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}_0) - \mathbf{x}_0\right| &\leq \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \left|\mathbf{x}_0 + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}_0)) - \mathbf{x}_0\right| \\ &= \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \left|A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}_0))\right| \\ &\leq \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \left\|A^{-1}\right\| |\mathbf{y} - \mathbf{f}(\mathbf{x}_0)| \\ &= \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \left\|A^{-1}\right\| |\mathbf{y} - \mathbf{y}_0| \end{aligned}$$

We choose $\epsilon = \frac{r}{2\|A^{-1}\|} < \frac{1}{2}r + \left\|A^{-1}\right\| \frac{r}{2\|A^{-1}\|}$. Therefore, $\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) \in \overline{B}$, that is, $\boldsymbol{\phi}_\mathbf{y} : \overline{B} \mapsto \overline{B}$. We know from step 1 that $\boldsymbol{\phi}_\mathbf{y}$ is a contraction on $\overline{B}$, and $\overline{B}$ is complete, so by the fixed point theorem/Theorem 9.23, there exist a unique fixed point $\mathbf{x} \in B \subset U$ such that $\boldsymbol{\phi}_\mathbf{y}(\mathbf{x}) = \mathbf{x}$. This $\mathbf{x}$ obeys $\mathbf{y} = \mathbf{f}(\mathbf{x})$ so $\mathbf{y} \in V$. □

As a remark, note that coming up with $\boldsymbol{\phi}$ in the above first part of the proof is quite difficult/inspired; this is definitely not a simple proof to come up with! After this first part, we have shown the existence of an open $U \ni \mathbf{a}$ such that $\mathbf{f} : U \mapsto V$ is a bijection. Note that we need V to be open for the proof of the second part of the theorem, as we need the derivative (which we only defined on open sets) to make sense.

Proof of (2)

We show that $\mathbf{g} = \mathbf{f}^{-1} : V \mapsto U$ is in $C^1(V)$. Let $\mathbf{y} \in V$. Then, $\mathbf{y} = \mathbf{f}(\mathbf{x})$ sor some unique $\mathbf{x}$. Since U is open, take $\mathbf{k}$ small enough such that $\mathbf{y} + \mathbf{k} \in V$. Say, $\mathbf{y} + \mathbf{k} = f(\mathbf{x_k})$ where $\mathbf{x_K} \in U$. Let $S = \mathbf{f}'(\mathbf{x})$ and $T = S^{-1}$. Note that $(\mathbf{f}'(\mathbf{x}))^{-1}$ exists by a continuity argument ($\mathbf{f}'(\mathbf{x})$ is close to $\mathbf{f}'(\mathbf{a})$ and $\mathbf{f}'(\mathbf{a})$ is invertible; see Rudin 9.8). Consider the expression:

$$\frac{\left|\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - T(\mathbf{k})\right|}{|\mathbf{k}|}.$$

We want to show that this goes to zero as $\mathbf{k} \to \mathbf{0}$. The idea is that somehow, we will turn this ratio into the derivative of $\mathbf{f}$. Doing some algebra, we have that:

$$\begin{aligned}\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - T(\mathbf{k}) = \mathbf{g}(\mathbf{f}(\mathbf{x_k})) - \mathbf{g}(\mathbf{f}(\mathbf{y})) - T\mathbf{k} &= \mathbf{x_k} - \mathbf{x} - T\mathbf{k} \\ &= -T\left((\mathbf{f}(\mathbf{x_k}) - \mathbf{f}(\mathbf{x})) - S(\mathbf{x_k} - \mathbf{x})\right)\end{aligned}$$

Taking the norm of both sides, we have:

$$\left|\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - T(\mathbf{k})\right| \leq \|T\|\left|\mathbf{f}(\mathbf{x_k}) - \mathbf{f}(\mathbf{x}) - S(\mathbf{x_k} - \mathbf{x})\right|$$

Note that $\mathbf{f}(\mathbf{x_k}) - \mathbf{f}(\mathbf{x}) = \mathbf{k}$. Also, we want to prove $|\mathbf{x_k} - \mathbf{x}| \leq 2\left\|A^{-1}\right\||\mathbf{k}|$. Next, we have that:

$$\boldsymbol{\phi}_{\mathbf{y}}(\mathbf{x_k}) - \boldsymbol{\phi}_{\mathbf{y}}(\mathbf{x}) = \mathbf{x_k} - \mathbf{x} + A^{-1}(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x_k}))$$

So therefore:

$$\begin{aligned}\mathbf{x_k} - \mathbf{x} = \boldsymbol{\phi}_{\mathbf{y}} - \boldsymbol{\phi}_{\mathbf{y}} + A^{-1}\mathbf{k} \implies |\mathbf{x_k} - \mathbf{x}| &\leq \left|\boldsymbol{\phi}_{\mathbf{y}} - \boldsymbol{\phi}_{\mathbf{y}}\right| + \left\|A^{-1}\right\||\mathbf{k}| \\ &\leq \frac{1}{2}|\mathbf{x_k} - \mathbf{x}| + \left\|A^{-1}\right\||\mathbf{k}|\end{aligned}$$

where in the last inequality we use the result from step 1 of the proof ($\boldsymbol{\phi}$ is a contraction with constant $\frac{1}{2}$). From this, we obtain that $|\mathbf{x_k} - \mathbf{x}| \leq 2\left\|A^{-1}\right\||\mathbf{k}|$ (as we wanted!) and hence $\frac{1}{|\mathbf{k}|} \leq \frac{2\left\|A^{-1}\right\|}{|\mathbf{x_k} - \mathbf{x}|}$.We therefore have that:

$$\begin{aligned}\left|\mathbf{g}(\mathbf{y}+\mathbf{k}) - \mathbf{g}(\mathbf{y}) - T\mathbf{k}\right| &\leq \frac{2\left\|A^{-1}\right\|}{|\mathbf{x_k} - \mathbf{x}|}\|T\|\left|\mathbf{f}(\mathbf{x_k}) - \mathbf{f}(\mathbf{x}) - S(\mathbf{x_k} - \mathbf{x})\right| \\ &= 2\left\|A^{-1}\right\|\|T\|\frac{\left|\mathbf{f}(\mathbf{x_k}) - \mathbf{f}(\mathbf{x}) - S(\mathbf{x_k} - \mathbf{x})\right|}{|\mathbf{x_k} - \mathbf{x}|}\end{aligned}$$

Now, let $\mathbf{k} \to \mathbf{0}$. Then, $\mathbf{x_k} - \mathbf{x} \to \mathbf{0}$, so the RHS goes to 0 as S is $\mathbf{f}'$ and hence this is just the definition of the derivative. Continuity follows from the fact that when you move an operator in a continuous way, so too does its image. □

THE END