Honors Analysis I

Andreas Tsantilas
Professor Jalal Shatah

Fall, 2020

Contents

Introduction

The books used here are Johnsonbaugh & Pfaffenberger, Rudin, Gleason, Abbott, as well as lectures given by Professor Jalal Shatah in the Fall of 2020.

1 Sets and Functions

1.1 Sets

Naively, we can understand a "set" to simply mean something which contains unique objects. For instance,

$$\{1, 3, 4, 8, 6\}$$

is a set. Note that the contents of this set do not have to be in any particular order, and can be rearranged without distrubing the uniqueness of the set. That is, the set $\{1, 2, 3, 4\}$ is equivalent to the set $\{4, 1, 2, 3, 3\}$. A set is considered a type of object, so it is meaningful to talk about sets within sets.

For finite sets, it may suffice to put each object down in writing. For infinite sets, we are allowed to define a set that contains everything that obeys a certain property P:

$$A = \{x : P(x)\}.$$

If P is the property that $x > 0$, we have the set of positive numbers (the colon is to be read as "such that"). If x is contained in the set A, we call x and *element* or *member* of A, and denote it by the expression

$$x \in A.$$

Similarly, if x is not in A, we write

$$x \notin A.$$

This leads rise to the notion of set equality:

Definition 1.1.1 (Set equality)**.** A set A is equal to a set B if they have the same elements. That is $A = B$ if and only if whenever $x \in A$, $x \in B$, and whenever $y \in B$, $y \in A$.

We also have notions of combining sets, and finding common elements.

Definition 1.1.2 (Union and Intersection)**.** For sets A and B, we define the *union* of A and B to be

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

and the *intersection* of A and B to be

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

This definition of sets extends easily into multiple, and perhaps infinitely many, sets. If $\mathcal{A} = \{A_1, A_2, \ldots\}$ is a set of sets, then we define

$$\cup\mathcal{A} = \bigcup_{n=1}^{\infty} A_n = \{x : x \in A_i \text{ for some } A_i \in \mathcal{A}\}$$

and

$$\cap\mathcal{A} = \bigcap_{n=1}^{\infty} A_n = \{x : x \in A_i \text{ for every } A_i \in \mathcal{A}\}$$

Sets do not necessarily have to contain elements. We are free to define the *empty set*:

Definition 1.1.3 (Empty Set)**.** There exists a set $\emptyset$ known as the *empty set.* This set contains no elements; that is, for every x we have $x \notin \emptyset$. Moreover, this set is unique; if we have $\emptyset$ and another empty set $\emptyset'$, it follows from the contrapositive of Definition 1.1 that $\emptyset = \emptyset'$.

Certain sets appear to be larger than other sets. From this, we can describe the notion of subsets.

Definition 1.1.4 (Subset)**.** A set A is a *subset* of a set B, denoted $A \subseteq B$, if for any object x,

$$x \in A \Rightarrow x \in B.$$

If $A \subseteq B$ but $A \neq B$, then we say that A is a *proper subset* of B and we denote it $A \subset B$.

Definition 1.1.5 (Set Difference)**.** If A and B are sets, the *difference* of A and B is

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}.$$

For instance, if $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$, then $A \setminus B = \{1\}$.

Sometimes, if we are working with sets which are all subsets of a larger set U, we typically call U the *universe* in which we are working. When we work with sets of integers, U could be the set of all integers, $\mathbb{Z}$. If we are working in a fixed universe U, then it makes sense to define complements.

Definition 1.1.6. We define the *complement* A' (sometimes A^c or $\bar{A}$) of A by

$$A' = U \setminus A$$

We can now state the basic properties of sets, all of which may be proven from the definitions given in this section:

Claim (Properties of Sets)**.** Let A, B, C be sets and let X be a set containing A, B, C as subsets. Then the following properties hold:

1. $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$.
2. $A \cup X = X$ and $A \cap X = A$.
3. $A \cup A = A$ and $A \cap A = A$.
4. $A \cup B = B \cup A$ and $A \cap B = B \cap A$.
5. $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$.
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (B \cup C)$.
7. $A \cup (X \setminus A) = X$ and $A \cap (X \setminus A) = \emptyset$.
8. $(A \cup B)' = A' \cap B'$, and $(A \cap B)' = A' \cup B'$.

Oftentimes in math, it is useful to consider the set of all possible subsets of a set X.

Definition 1.1.7 (Power Set)**.** The power set of X is the set of all possible subsets of X (including $\emptyset$ and X itself). It is denoted by

$$\mathcal{P}(A)$$

or sometimes by 2^X.

1.2 Relations

Foremost, we introduce the notion of an ordered pair, and equality between ordered pairs.

Definition 1.2.1 (Ordered Pair)**.** An *ordered pair* is an object of the form

$$(x, y)$$

which can equivalently be expressed in set-theoretic notation as

$$(x, y) \equiv \{\{x\}, \{x, y\}\}.$$

Two ordered pairs (x, y) and (x', y') are considered to be equal if and only if their components match. That is,

$$(x, y) = (x', y') \Leftrightarrow x = x' \text{ and } y = y'$$

Notice how the set theoretic expression has an element dedicated to defining the first element, and an element defining the second element in the pair.

Given two sets A and B, we can construct a new set consisting of ordered pairs of their elements.

Definition 1.2.2 (Cartesian Product)**.** Given two sets A and B, we define their Cartesian Product $A \times B$ to be:

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

We define the Cartesian Product of n sets to be

$$A_1 \times \cdots \times A_n = \{(a_1 \dots a_n) : a_i \in A_i\}.$$

The Cartesian Product of a set A with itself n times is often denoted

$$A^n = A \times \cdots \times A.$$

An ordered list of n objects is known as an n-tuple. This notion is consistent with the idea of vectors of the form $(x, y, z) \in \mathbb{R}^3$; each "slot" of the vector is an element of $\mathbb{R}$, and the order in which the components of a vector are listed matter.

Given this notion of Cartesian Products, we would like to be able to relate elements of a set X to one another.

Definition 1.2.3 (Relation)**.** A *relation* R on X is a subset of $X \times X$.

$$R \subseteq X \times X.$$

We say for $x, y \in X$ that xRy, or x is related to y, whenever

$$xRy \Leftrightarrow (x, y) \in R.$$

For instance, we can think of "greater than" to be a relation defined on $\mathbb{R} \times \mathbb{R}$. Thus, R will consist of ordered pairs (x, y) such that $x > y$. However, note that (y, x) and (x, x) are not an elements of R. Particularly interesting in math are relations which satisfy certain properties and can be thought of as a generalization of equality.

Definition 1.2.4 (Equivalence Relation)**.** The relation R on a set X is an equivalence relation on A if the following are satisfied:

1. $(x, x) \in R$ for all $x \in X$.
2. $(a, b) \in R \Leftrightarrow (b, a) \in R$.
3. $(a, b) \in R$ and $(b, c) \in R \Rightarrow (a, c) \in R$.

For a general binary relation $\sim$ on X, $\sim$ is an equivalence relation if for all $x, y, z \in X$:

1. $x \sim x$
2. $x \sim y \Rightarrow y \sim x$
3. $x \sim y$ and $y \sim z \Rightarrow x \sim z$.

These properties are called reflexivity, symmetry, and transitivity, respectively.

Definition 1.2.5 (Equivalence Class)**.** If X is a set and $\sim$ is an equivalence relation, then the *equivalence class* of x, denoted $[x]$, is

$$[x] := \{y \in X : x \sim y\}$$

Claim (Properties of Equivalence Classes)**.** The following are properties of equivalence classes. For a set X, a relation $\sim$, and $x_1, x_2 \in X$, the following are true:

1. $[x_1] = [x_2] \Leftrightarrow x_1 \sim x_2$.
2. $[x_1] \cap [x_2] \neq \emptyset \Leftrightarrow [x_1] = [x_2]$.

Proof. We will now prove both of these claims:

1. Since $x_1 \in [x_1]$, that implies $x_1 \in [x_2]$ since the sets are equal. Therefore $x_1 \in [x_2]$. For the backwards direction, we suppose $y \in [x_1]$. By transitivity, $y \sim x_1$ and $x_1 \sim x_2$ imply $y \sim x_2$, so $y \in [x_2]$. Since $y \in [x_1] \Rightarrow y \in [x_2]$, $[x_1] \subseteq [x_2]$. By symmetry, we can arrive at a similar conclusion for $[x_2] \subseteq [x_1]$, so $[x_1] = [x_2]$.
2. If we pick an $x_3 \in [x_1] \cap [x_2]$, that means $x_3 \sim x_1$ and $x_3 \sim x_2$, so $x_1 \sim x_2$ from transitivity, and their equivalence classes are equal. For the backwardsd direction, it is clear that equality between the calsses implies a nonempty intersection.

□

These equivalence classes form subsets of X, and seem to be either equal or completely disjoint. This leads to the notion of a partition.

Definition 1.2.6 (Partition)**.** A partition P on a set X is a subset of $\mathcal{P}(X)$ such that the following properties hold:

1. For all $A \in P$, $A \neq \emptyset$,
2. For all $A, B \in P$, $A \neq B \Rightarrow A \cap B = \emptyset$, and
3. $\bigcup\limits_{A \in P} A = X$.

In plain English, it is a collections of disjoint subsets of X whose union equals X.

Definition 1.2.7 (Quotient Set of a Relation)**.** The *quotient set of the set X by a relation $\sim$* is denoted $X/{\sim}$ and is given by:

$$X/{\sim} := \{[x] : x \in X\}.$$

Note that $X/{\sim} \subset \mathcal{P}(X)$.

We will now demonstrate that the set of all equivalence classes given by $\sim$ form a very natural way to partition the set X. Indeed, all an equivalece relation is is a partition on X, and all a partition is is an equivalence relation.

Theorem 1.2.1 (The Quotient Set is a Partition)**.** Partitions and equivalence relations are equivalent. That is, every partition denotes an equivalence relation, defined by

$$x \sim y := \exists A \in P : x, y \in A,$$

and every equivalence relation $\sim$ forms a set $X/{\sim}$ that is a partition on X.

Proof. We show that every Partition is an equivalence relation, as defined in the previous theorem.

1. $x \sim x$, since we know from condition 3 of Definition 1.2.6 that there is an $A : x \in A$.
2. $x \sim y \Rightarrow \exists A \in P : x, y \in A \Rightarrow y \sim x$.
3. $x \sim y \Rightarrow \exists A \in P : x, y \in A$. $y \sim z \Rightarrow \exists B \in P : y, z \in B$. Therefore, $A \cap B \neq \emptyset$ so from condition 2, $A = B$. Therefore $x, z \in A$ so $x \sim z$.

Now we will demonstrate how the quotient set $X/{\sim}$ satisfies all 3 parts of Definition 1.2.6.

1. Clearly, $[x] \in X/{\sim} \neq \emptyset$ for all x, since $x \in [x]$.
2. Condition 2's contrapositive is $A \cap B \neq \emptyset \Rightarrow A = B$, which was proved previously.
3. Suppose $S = \bigcup_{[x] \in X/\sim} [x] \neq X$. Then that implies the existence of an $x' \notin S$. But since $x' \sim x'$, $[x'] \in X/{\sim}$, so it is in S.

$\square$

1.3 Functions

Definition 1.3.1 (Function)**.** A *function* from a set X to a set Y is a subset f of $X \times Y$ such that

1. If $(x, y), (x, y') \in f$, then $y = y'$ and
2. If $x \in X$, then $(x, y) \in f$ for some $y \in Y$.

If $(x, y) \in f$, we define $f(x)$ to equal y. The first condition ensures that each element in x can be associated with a unique element in Y, and the second stipulates that f "captures" every element of X. Two functions $f : X \to Y$ and $g : X \to Y$ are said to be equal if and only if $f(x) = g(x)$ for every $x \in X$.

From the definition of a relation as a subset of $X \times Y$, functions are a special kind of relation.

Notation (Function)**.** If a function f is a subset of $X \times Y$, then we write

$$f : X \to Y$$

to denote the function f from X to Y. This notation can be used for any mapping from X to Y, but f almost always denotes a function.

Remark (Quotient Map)**.** Given a set X and a relation $\sim$, we call

$$\begin{aligned} \pi &: X \to X/{\sim} \\ &: x \mapsto [x] \end{aligned}$$

a quotient map.

Definition 1.3.2 (Image)**.** If $f : X \to Y$, and $S \subseteq X$, then we define the *image of S under f*, denoted $f(S)$, as

$$f(S) := \{f(x) : x \in S\}.$$

Definition 1.3.3 (Onto)**.** A function $f : X \to Y$ is said to be *onto* if

$$f(X) = Y.$$

That is, for every $y \in Y$, $y = f(x)$ for some $x \in X$. This condition is sometimes known as "surjectivity."

Definition 1.3.4 (One-to-one)**.** A function $f : X \to Y$ is said to be *one-to-one* if

$$x \neq x' \Rightarrow f(x) \neq f(x')$$

or, equivalently,

$$f(x) = f(x') \Rightarrow x = x'.$$

That is, different inputs have different outputs. This condition is sometimes known as "injectivity."

Definition 1.3.5 (Bijectivity)**.** A function $f : X \to Y$ is said to be *bijective* if it is both one-to-one and onto.

Definition 1.3.6 (Inverse of f)**.** If $f : X \to Y$ is a one-to-one function such that $f(X) = B \subseteq Y$, then we define the *inverse* of f denoted f^{-1}, where

$$f^{-1} : B \to X$$

such that

$$(y, x) \in f^{-1} \Leftrightarrow (x, y) \in f.$$

Definition 1.3.7 (Function Composition)**.** Given functions $f : X \to Y$ and $g : Y \to Z$, we define the *composition of g and f* to be

$$(g \circ f) : X \to Z$$

where

$$(g \circ f)(x) = g(f(x)).$$

2 The Real Numbers

While it is certainly possible for one to construct number systems such as the integers ($\mathbb{Z}$) and rationals ($\mathbb{Q}$) from the naturals ($\mathbb{N}$), we will take it on faith that these have been constructed and the familiar rules of algebra apply.

Therefore under these familiar rules, we can introduce the concept of a *field*.

Definition 2.0.1. A field is a set $\mathbb{F}$ equipped with:

1. Two distinct elements called 0 and 1.
2. A function called additive inversion from $\mathbb{F}$ to $\mathbb{F}$ and denoted by the minus symbol, "$-$," and a function from $\mathbb{F} \setminus \{0\}$ to $\mathbb{F} \setminus \{0\}$ called multiplicative inversion that takes an element x to an element called x^{-1}.
3. Two functions from $\mathbb{F} \times \mathbb{F} \to \mathbb{F}$ called addition and multiplication, denoted $x + y$ and $x \cdot y$ or xy.

These functions and elements must in turn satisfy the following nine conditions or axioms for $x, y, z \in \mathbb{F}$:

1. $x + 0 = 0 + x = x$
2. $x + (-x) = (-x) + x = 0$
3. $x + (y + z) = (x + y) + z$
4. $x + y = y + x$
5. $1 \cdot x = x \cdot 1 = x$
6. $xx^{-1} = x^{-1}x = 1$
7. $x(yz) = (xy)z$
8. $xy = yx$
9. $x(y + z) = (xy) + (xz)$, and $(x + y)z = (xz) + (yz)$

We can extend the concept of a field to the concept of an ordered field.

Definition 2.0.2. An ordered field $\mathbb{X}$ is a field with an order operation "¡" from $\mathbb{X} \times \mathbb{X} \to \mathbb{X}$ such that for all $x, y, z \in \mathbb{X}$,

1. If $x < y$, then $x + z < y + z$.
2. If $0 < x$ and $0 < y$, then $0 < xy$.

From these axioms, it is clear that the rationals $\mathbb{Q}$ form an ordered field.

Despite the concepts we have just described, we still have a certain incompleteness to our numerical systems. The rationals allow us to be infinitely granular in our arithmetic (between any two rational numbers, you can always squeeze one more in). However, this is a countable granularity. Between any two rationals, we can find a bijection between the rationals between them and $\mathbb{N}$. This has caveats, as we will demonstrate below.

Claim. Define $\sqrt{x}$ to be $y : y^2 = x$. Then there is no rational number equal to $\sqrt{2}$.

Proof. Suppose we could write the square root of 2 as an *irreducible* ratio of two integers p and q. Then we would have

$$\left(\frac{p}{q}\right)^2 = 2$$

From this, we have

$$p^2 = 2q^2$$

so p must be even. If it were odd, then it could be written in the form $p = 2k + 1$, and $p^2 = 2(2k^2 + k) + 1$ which is odd. Therefore, since p is even, it can be written in the form $p^2 = 4k^2$. Therefore

$$2k^2 = q^2$$

so q must also be even, so $q = 2m$ similarly to above. But that means the ratio can be reduced to just k/m, which contradicts the assumption that the ratio was irreducible. Therefore there is no rational number r such that $r^2 = 2$. $\square$

Despite this, we consider that $\sqrt{2}$ exists somehow, even if it is out of the realm of rationals. Indeed, the Ancient Greeks were able to construct this by creating a right triangle with side lengths 1 and 1.

GO BACK AND COVER ALL THE STUFF

3 Functions and Continuity

So far we have primarily talked about sequences, which can viewed as functions from $\mathbb{N} \to \mathbb{R}$. However, this was a discrete set of points, whereas functions from $\mathbb{R} \to \mathbb{R}$ can be thought of as functions on a continuum (i.e., a set which contains no "holes").

3.1 Functional Limits

Notation (Open and Closed Intervals)**.** Let $a, b \in \mathbb{R}$ and $a < b$.

1. (Closed Interval.) $x \in [a, b] \iff a \leq x \leq b$
2. (Open Interval.) $x \in (a, b) \iff a < x < b$
3. $x \in [a, b) \iff a \leq x < b$
4. $x \in (a, b] \iff a < x \leq b$

Definition 3.1.1 (Limit Point)**.** Let $X \subseteq \mathbb{R} : X \neq \emptyset$. Then we say an element a is an *limit point* or accumulation point of X if for every $\delta > 0$, there exists an $x \in X$ such that $0 < |a - x| < \delta$. A *left limit point* is one such that for every $\delta > 0$, $0 < a - x < \delta$, and a *right limit point* is one such that for every $\delta > 0, 0 < x - a < \delta$.

We can see that b is a left limit point of the interval (a, b), even though $b \notin (a, b)$. There is also an equivalent definition of limit points:

Theorem 3.1.1. A point a is a limit point if and only if

Definition 3.1.2 (Functional Limit)**.** Let $f : X \to \mathbb{R}$, and let a be an accumulation point of $X \subseteq \mathbb{R}$. Then we say that *the limit of* f *as* $x \to a$ *is* L if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\forall \varepsilon > 0, \exists \delta > 0 : |x - a| < \delta \Rightarrow |f(x) - L| < \varepsilon.$$

We denote this by

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

Moreover, from theorem [NAME THEOREM], this limit L is unique.

In plain English, this means that $\lim_{x \to a} f(x) = L$ if the distance from $f(x)$ and L can be made arbitrarily small if the difference between $x \neq a$ and a is arbitrarily small. We can also think of this in terms of "neighborhoods," or open intervals, since $0 < |x - a| < \delta \iff x \in (a - \delta, a + \delta)$. If x is in the δ-neighborhood of a, then we are guaranteed that $f(x)$ is in the ϵ-neighborhood of L.

As a matter of fact, this definition can be interpreted in many different ways.

Theorem 3.1.2 (Sequences and Functional Limits)**.** Let $X \subseteq \mathbb{R}$, $f : X \to \mathbb{R}$, and a be an accumulation point of X. Then the following are equivalent:

1. $\lim_{x \to a} f(x) = L$.
2. For every sequence $(x_n) : x_i \in X$, $x_n \neq a$, and $(x_n) \to a$, the sequence $(f(x_n)) \to L$.

Proof. We first show that $1 \Rightarrow 2$. Clearly, this is the case, since $(x_n) \to a$ so we can find an N_δ such that $\forall n \geq N_\delta, |x_n - a| < \delta$. But this implies that $\forall n \geq N_\delta, |f(x_n) - L| < \varepsilon$. Therefore $(f(x_n)) \to L$.

Now we show that $2 \Rightarrow 1$. Suppose for the sake of contradiction that $\lim_{x \to a} f(x) \neq L$. Then, this is the negation of the definition, so there exists an $\varepsilon > 0$ such that $\forall \delta > 0, |x' - a| < \delta$ and $|f(x') - L| \geq \varepsilon$ for some x'. Thus, since this is true for all δ, let $\delta_n = 1/n$. Then for each n, we can construct a sequence based on the x'_n such that $|f(x'_n) - L| \geq \varepsilon$. Since $\delta_n \to 0$, this means that $(x'_n) \to a$, but since $|f(x'_n) - L| \geq \varepsilon$ for all n, $(f(x'_n)) \not\to L$, contradicting 2. Therefore $\lim_{x \to a} f(x) = L$. □

Theorem 3.1.3 (Functional Limit Algebra)**.** Let $X \subseteq \mathbb{R}$, and let $f, g : X \to \mathbb{R}$. Assume $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, where a is an accumulation point of X. Then the following hold:

1. $\lim_{x\to a} cf(x) = cL$ for all $c \in \mathbb{R}$.
2. $\lim_{x\to a}[f(x) + g(x)] = L + M$.
3. $\lim_{x\to a} f(x)g(x) = LM$.
4. $\lim_{x\to a} f(x)/g(x) = L/M$ provided $M \neq 0$.

Proof. These hold because of the equivalence between sequences and limits, as proved in Theorem 3.1.1. We already have these same theorems for sequences, and can therefore easily extend them to functional limits. $\square$

If we consider the function
$$f(x) = \frac{|x|}{x},$$
it is easy to show that it does not have a limit as $x \to 0$, since $|f(x) - L| \geq \varepsilon$ for all L and a fixed ε. However, if we ignore all $x > 0$, then this function does have a limit, namely -1. This gives rise to the notion of a one-sided limit.

Definition 3.1.3 (Sided Limits)**.** Let $f : X \to \mathbb{R}$ and let a be a left (right) accumulation point of X. Then we say that the limit of $f(x)$ as x approaches a from the left (right) if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that
$$\forall\varepsilon > 0 \exists\delta > 0 : 0 < a - x < \delta (\text{right: } 0 < x - a < \delta) \Rightarrow |f(x) - L| < \varepsilon.$$
We denote this by
$$\lim_{x\to a^-} f(x) = L \text{ (right: } \lim_{x\to a^+} f(x) = L).$$

Definition 3.1.4 (Infinite Limits)**.** Let $X \subseteq \mathbb{R}$ and $f : X \to \mathbb{R}$. Then we say that the limit of $f(x)$ as $x \to \infty$ is L if for every $\varepsilon > 0$, there exists a number M such that if $x > M$, then $|f(x) - L| < \varepsilon$. We denote this by
$$\lim_{x\to\infty} f(x) = L.$$
If X is bounded above, then we say that this infinite limit is undefined. Moreover, we can extend this definition to $-\infty$.

3.2 Continuity

Definition 3.2.1 (Continuity)**.** Let $X \subseteq \mathbb{R}$ and $f : X \to R$. Let $c \in X$. Then we say that f is continuous at X if and only if
$$\lim_{x\to c} f(x) = f(c).$$
That is, f is continuous if and only if
$$\forall\varepsilon > 0, \exists\delta > 0 : |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon.$$
We say that f is continuous on X (or continuous) if for every $c \in X$ this holds. If $X = [a, b]$, then all we demand is that $\lim_{x\to a^+} f(x) = a$, and $\lim_{x\to b^-} f(x) = f(b)$.

This condition can be thought of as saying the graph (which we have not defned) of a continuous function does not suddenly "jump" from one value to the next, and $f(c)$ is always defined for $c \in X$. With this picture in mind, we can formulate many equivalent notions of continuity.

Theorem 3.2.1 (Sequences and Continuity)**.** Let $X \subseteq \mathbb{R}$ and let $f : X \to \mathbb{R}$, and $c \in X$. Then f is continuous at c if and only if:

1. For all $\varepsilon > 0$, there exists a $\delta > 0 : |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon$.
2. For all $(x_n) \to c$ such that $x_n \in X$, we have $(f(x_n)) \to f(c)$.

Proof. This is proved like theorem 3.1.1. □

Oftentimes, a way to prove (dis)continuity is by contradiction. This sometimes involves assuming continuity, then constructing a sequence such that there exists an ε such that there is a subsequence such that $x_{n_k} \to c$, but $|f(x_{n_k}) - f(c)| \geq \varepsilon$ for all k. Then we would have a sequence always larger than $f(c)$ which clearly cannot converge to $f(c)$.

Moreover, an $\varepsilon - \delta$ proof can be thought of as a game, where one "team" tries to outdo the other. The first team picks a small ε, and asks to see if you can provide (or simply show that there exists!) a suitable δ as a response, such that for all x within this δ range, $|f(x) - f(c)| < \varepsilon$. If you can, then you win and f is continuous at c.

Theorem 3.2.2 (Continuity Algebra)**.** Let $X \subseteq \mathbb{R}$ and let $f, g : X \to \mathbb{R}$, and f and g are continuous at a point $c \in X$. It follows that

1. $kf(x)$ is continuous at c for all $k \in \mathbb{R}$.
2. $f(x) + g(x)$ is continuous at c.
3. $f(x)g(x)$ is continuous at c.
4. $f(x)/g(x)$ is continuous at c if the quotient is defined.

Proof. These follow from theorem 3.1.2 and theorem 3.2.1. □

3.3 Bounded Functions and the Heine-Borel Theorem

We now turn to an interesting theorem that can better characterize our understanding of continuous functions.

Definition 3.3.1 (Bounded Function)**.** We say that a function $f : D \to \mathbb{R}$ is bounded on $X = [a, b]$ if there exists an M such that for every $c \in X \cap D$, $|f(c)| < M$. Note how f can be bounded on a set X which is not a subset of the domain D of f. This will be the definition for this section only.

Theorem 3.3.1. Let $f : X \to \mathbb{R}$ such that f is defined and continuous on $[a, b] \subseteq X$. Then f is bounded on the closed interval $[a, b]$.

Proof. For the sake of contradiction, for all $n \in \mathbb{N}$, let there exist a $c \in [a, b]$ such that $|f(c)| > n$ (this is the negation of boundedness). Let x_n be a sequence such that $\forall n \in \mathbb{N}, \exists x_n \in [a, b]$ such that $|f(x_n)| > M$. However, from the Bolzano-Weierstrass theorem, since $a \leq x_n \leq b$, there exists a convergent subsequence $x_{n_k} \to x^*$ for some $x^* \in [a, b]$. Therefore, $f(x^*)$ exists and is finite since x^* is in the domain; due to the sequential limit criterion, $(x_n) \to x^* \Rightarrow |f(x_n)| \to |f(x^*)|$. However, this is a contradiction since $|f(x_n)| \to \infty$. □

Warning. This proof cannot be used on an open interval (a, b), because the subsequence x_{n_k} might go to b or a, but these would be out of the domain. For instance, the function $f(x) = 1/x$ is continuous and defined on the closed interval $(0, 1)$, but it is unbounded since for all M, we have that $f(1/(M + 1)) > M$.

A natural consequence of a function having no "holes" is that if on a closed interval $[a, b]$ we have $f(a) \leq f(b)$, then it will attain every value in between the two. As an analogy, you cannot take the elevator from the first floor to the third floor without crossing the second floor at some point. We can codify this idea into an important theorem:

Theorem 3.3.2 (Intermediate Value)**.** Let $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Without loss of generality, assume $f(a) \leq f(b)$, and let $f(a) \leq c \leq f(b)$. Then there exists an $x_c \in [a, b]$ such that $f(x_c) = c$.

Proof. GIVE PROOF!!! □

Lemma 3.3.3. If f is bounded on $X_1, \ldots, X_n$, then f is bounded on $X_1 \cup \cdots \cup X_n$.

Proof. Our base case is $n = 2$. Let M_1 and M_2 be the respective bounds. Clearly, for $c \in X_1 \cup X_2 \cap D$, $|f(c)| < \max(M_1, M_2)$. Assming it holds for $n-1$, let M^* be the bound for $f(c)$ for $c \in X_1 \cup \cdots \cup X_{n-1}$. then we see that for $c \in X_1 \cup \cdots \cup X_{n-1} \cup X_n \cap D$, that $|f(c)| < \max(M^*, M_n)$. Therefore f is bounded $X_1 \cup \cdots \cup X_n$. □

Lemma 3.3.4. If f is continuous and $f(c)$ is bounded, then there exists a $\delta > 0$ such that for $x \in (c - \delta, c + \delta)$, $f(x)$ is bounded.

Proof. This follows from the definition of continuous. This means that we can choose $\varepsilon = 1$, and there exists a δ such that

$$x \in (c - \delta, c + \delta) \Rightarrow -1 < f(x) - f(c) < 1$$

or equivalently, $f(c) - 1 < f(x) < f(c) + 1$ or $|f(x)| < M + 1$. Therefore f is bounded on $(c - \delta, c + \delta)$. □

Now we know that if $c \in [a, b]$, there is an open interval I_c containing c such that f is bounded on I_c. However, for all c, there are infinitely many such intervals, so we cannot conclude that f is continuous on $\cup\{I_c : c \in [a, b]\}$. However, the following theorem demonstrates that we can find a finite union of intervals such that $[a, b] \subseteq I_1, \ldots, I_n$. This would mean that the funcition f would be bounded on all of $[a, b]$.

The importance of the Heine-Borel theorem may escape us for the time being, but it is of incredible importance in analysis. It's a theorem which states that something infinite can be reduced to something finite. This is welcome news; much of mathematical analysis is concerned with taming the infinite.

Theorem 3.3.5 (Heine-Borel)**.** Let $[a, b]$ be a bounded and closed interval, and let Λ be any collection of open intervals I_α such that

$$[a, b] \subseteq \bigcup_{\alpha \in \Lambda} I_\alpha$$

Then there exists a finite subset $\{I_1, \ldots, I_n\}$ of I_α such that

$$[a, b] \subseteq \bigcup_{k=1}^{n} I_k.$$

In other words, for every infinite family of open intervals that "covers" the interval $[a, b]$, then there exists a finite subset of this family that covers $[a, b]$, sometimes called a "subcover."

Proof. Foremost, the index α may not even be countable. However, we do know of one countably infinite set such that no interval can be in between two elements; that is the set $\mathbb{Q}$ of rational numbers, which is dense in $\mathbb{R}$. Therefore we know that there is at least one rational number q in every open interval $q \in I_\alpha$. Therefore, we can re-index the $I_\alpha = I_r$ so that r is a rational and $r \in I_\alpha$. However, this assignment is not unique, so we can say that if $r \in I_{\alpha_1}$ and $r \in I_{\alpha_2}$, then $I_r = I_{\alpha_1} \cup I_{\alpha_2}$. Now we can convert these rationals to integers, since there is a bijection between $\mathbb{N}$ and $\mathbb{Q}$ by the definition of countability. Thus,

$$[a,b] \subseteq \bigcup_{j=1}^{\infty} I_j.$$

For the sake of contradiction, assume the theorem is false and that there is no finite subcover. Then

$$[a,b] \not\subseteq \bigcup_{j=1}^{k} I_j \Rightarrow \exists x_k : x_k \in [a,b], x_k \notin \bigcup_{j=1}^{k} I_j$$

where we are justified in writing x_k since it depends on k. Now since x_k is contained in $[a,b]$ which is closed and bounded, then there is a convergent subsequence $x_{k_i} \to x^*, x^* \in [a,b]$. Therefore $x^* \in I^*$ for some I^* in the infinite collection. Since I^* is open, then x^* is not a boundary of the interval but inside the interval proper. Since $x_{k_i} \to x^*$, the x_{k_i} are eventually in I^*. We know that $I^* = I_\ell$ for some ℓ. Therefore, for $k > \ell$, the x_{k_i} are in I_ℓ, which is a contradiction since we assumed each x_k is not in any finite collection. This completes the proof. $\square$

Without the hypotheis that $[a,b]$ is closed and bounded, then this theorem fails. For instance, consider the unbounded interval $[1,\infty)$[1], and the collection of open intervals $I_n = (0,N)$. Clearly, any finite collection will not be able to cover it. Similarly, if we omitted the closed hypothesis, then we could consider the interval $(0,1)$ and the infinite collection $I_n = (1/n, 1)$. For any finite n, their union will fail to cover $(0,1)$.

Later on, we will translate this theorem into the language of compactness, and it will state that closed and bounded sets are compact.

One quick application of this theorem is that when determining if a functoin is continuous, we can choose a δ that does not depend on the limiting point c.

Claim. If $f : [a,b] \to \mathbb{R}$ and f is continuous, then our choice of δ does not depend on c.

Proof. We have that $|x - c| < \delta_c$ for some δ_c which implies that $|f(x) - f(c)| < \varepsilon$. This suggests an open interval $I_c = (c - \delta, c + \delta)$ which contains c. Since we can do this for every $c \in [a,b]$, then

$$[a,b] \subseteq \bigcup_{c \in [a,b]} I_c.$$

By the Heine-Borel theorem,

$$[a,b] \subseteq \bigcup_{i=1}^{k} I_i, \text{ where } I_i = (c_i - \delta_i, c_i + \delta_i)$$

for some k. Therefore, we can choose $\delta^* = \min(\delta_1, \ldots, \delta_k)$. [CLEAN UP] $\square$

We used the Bolzano-Weierstrass theorem to prove the Heine-Borel theorem; however, we can show that the Heine Borel-theorem implies the Bolzano-Weierstrass theorem.

[1] This is considered closed, since every convergent sequence in $[1,\infty)$ converges to a point in the interval.

Theorem 3.3.6 (Maximum Principle/Extremum Principle)**.** Let $f : [a, b] \to \mathbb{R}$. Then f reaches a maximum and a minimum on $[a, b]$.

Proof. We will only prove the case of the maximum; the minimum is worked out similarly. We know that since f is on a closed and bounded interval, it is bounded. Let

$$M = \sup\{f(x) : x \in [a, b]\}.$$

Now all we need to show is that $M = f(c)$ for some $c \in [a, b]$. Suppose this is not true. Then let

$$g(x) := \frac{1}{M - f(x)}\ x \in [a, b].$$

By the algebraic properties of continuity, we have that $g(x)$ is continuous. Therefore, g must be bounded by some N. This implies that

$$\frac{1}{M - f(x)} < N$$

Which implies that

$$f(x) < M - \frac{1}{N}\ \forall x \in [a, b].$$

This contradicts the fact that $M = \sup\{f(x) : x \in [a, b]\}$. Therefore, f attains its maximum on the interval $[a, b]$. □

4 Metric Spaces

In many respects, this section constitutes a small glimpse of higher mathematics. It is a vast generalization of many things in linear algebra, such as norms and the concept of distance.

4.1 Norms

Vector spaces sometimes come with a bilinear form called an *inner product*. That is, it is a function from $V \times V \to \mathbb{F}$, where V is a vector space over some field $\mathbb{F}$. In $\mathbb{R}^n$, it is simply

$$\langle x, y\rangle = x \cdot y = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

This inner product gives rise to a norm on $\mathbb{R}^n$, or a measure of a vector's "length:"

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2}.$$

This norm obeys the folliwing properties:

(i) $\|x\| = 0 \iff x = 0$.

(ii) $\|\lambda x\| = |\lambda|\|x\|$ for $\lambda \in \mathbb{R}$.

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

Thus, it makes sense that any generalization of a norm would satisfy the above properties. Therefore, we could think of a different norm on $\mathbb{R}^n$. Let

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|,$$

where $|\cdot|$ denotes the absolute value operation. You should verify that this is a norm. Moreover, for any $p \geq 1$,

$$\|x\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \cdots + |x_n|^p}$$

is a norm. If we take this limit to infinity, then we get an infinite norm,

$$\|x\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|).$$

Recall that a vector space V of dimension d is isomorphic to $\mathbb{R}^d$ (i.e., there exists a bijection between them).

Thus, we can think of an infinite-dimensional vector as living in $\mathbb{R}^\infty$. That is,

$$a \in \mathbb{R}^\infty,\ a = (a_1, a_2, a_3, \ldots),\ a_i \in \mathbb{R}.$$

This looks very much like a sequence in $\mathbb{R}$; indeed, $R\infty$ can be thought of as the space of sequences, since there is a very natural bijection between them. However, we may not be able to to meaningful algebra on certain vectors in $\mathbb{R}^\infty$, since certain series may diverge. Therefore, we can define the following sets:

Definition 4.1.1 (Sequence Spaces)**.** For an infinite dimensional vector space $\mathbb{R}^\infty$, we define the p-norm to be

$$\|x\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \ldots} = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}$$

and the sequence space ℓ^p:

$$\ell^p = \{a = (a_n)_{n=1}^\infty : \|a\|_p < \infty\}.$$

In the infinite case, we have

$$\ell^\infty = \{a = (a_n)_{n=1}^\infty : \sup(|a|) < \infty\}.$$

Notice how ℓ^1 is simply the set of all convergent series. Moreover,

$$\ell^1 \subset \ell^2 \subset \cdots \subset \ell^\infty.$$

4.2 Definitions and Theorems

In order to generalize the notion of length, we would like to create a function that obeys properties similar to the norm.

Definition 4.2.1 (Metric)**.** Let M be a set. A *metric* on M is a function $d : M \times M \to \mathbb{R}^+$ such that the following hold:

1. $d(x, y) = 0 \iff x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, z) \leq d(x, y) + d(y, z)$.

Definition 4.2.2 (Metric Space)**.** A *metric space* is an ordered pair (M, d) where M is a set and d is a metric on M.

For the standard 2-norm, the following result should be familiar.

Theorem 4.2.1 (Cauchy-Schwarz Inequality)**.** Let x, y be vectors in $\mathbb{R}^n$, $\cdot$ denote the dot product, and $\|\cdot\|$ be the 2-norm. Then

$$|x \cdot y|^2 \leq \|x\|\|y\|.$$

Or rather

$$\left|\sum_{k=1}^{n} x_k y_k\right| \leq \sqrt{\left(\sum_{k=1}^{n} x_k^2\right)\left(\sum_{k=1}^{n} y_k^2\right)}.$$

Proof. FLILL IN PROOF □

Corollary 4.2.2. The equation

$$d(x, y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

defines a metric on $\mathbb{R}^n$.

Proof. FILL □

Corollary 4.2.3. If $\{a_k\}, \{b_k\} \in \ell^2$, then the series

$$\sum_{k=1}^{\infty} a_k b_k$$

converges absolutely (that is, it is in ℓ^1).

Proof. FIL:: □

Theorem 4.2.4 (Cauchy-Schwarz for ℓ^2)**.** If $\{a_k\}, \{b_k\} \in \ell^2$, then

$$\left|\sum_{k=1}^{\infty} a_k b_k\right| \leq \sqrt{\left(\sum_{k=1}^{\infty} a_k^2\right)\left(\sum_{k=1}^{\infty} b_k^2\right)}.$$

Proof. Fill □

4.3 Sequences in Metric Spaces

A sequence in a metric space is simply a function from $\mathbb{N} \to M$, and we denote it $\{a_n\}$. We will soon find it convenient to also use the notation $\{a^{(n)}\}$ to refer to a sequence. Now, the definition for sequentiak convergence is similar to the regular one for sequences in $\mathbb{R}$:

Definition 4.3.1. Let $\{a_n\}$ be a sequence in M. We say that the sequence $\{a_n\}$ converges to $L \in M$ if $\forall \varepsilon > 0$, $\exists N : \forall n \geq N$,

$$d(L, a_n) < \varepsilon.$$

Now we can define convergence for points in $\mathbb{R}^n$. These are not vectors, since they do not necessarily have the structures of addition or multiplcation defined on them. They are vector-like insofar as they are ordered arrays of objects.

Theorem 4.3.1. Let $\{a^{(k)}\}$ be a sequence of points in $\mathbb{R}^n$. Let $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. Let $a^{(k)} = (a_1^{(k)}, \dots, a_n^{(k)})$, for k in $1, 2, \dots$. Then $\{a^{(k)}\}$ converges to a if and only if $a_k^{(k)} \to a_j$.

Proof. Suppose $\{a^{(k)}\}$ converges to a. Pick an $\varepsilon > 0$. Then we have that $d(a^{(k)}, a) < \varepsilon$ for all $k \geq K$. Moreover,

$$|a_j^{(k)} - a_j| \leq \sqrt{\sum_{i=1}^{n} (a_i^{(k)} - a_i)^2} < \varepsilon.$$

Therefore, $a_j^{(k)} \to a_j$.

Conversely, suppose that each $a_j^{(k)} \to a_j$. Then we know that there exists an N such that for all $k \geq K$,

$$|a_j^{(k)} - a_j| < \frac{\varepsilon}{\sqrt{n}} \quad \text{for} 1 \leq j \leq n.$$

Therefore,

$$d(a^{(k)}, a) = \sqrt{\sum_{i=1}^{n} (a_i^{(k)} - a_i)^2} < \sqrt{\frac{n\varepsilon^2}{n}} = \varepsilon,$$

so $a^{(k)} \to a$. □

Sometimes, our notation is ambiguous when we transition to $\mathbb{R}^\infty$.

Notation. When we say "$\{a_n\}$ is a sequence in ℓ^p", we mean that each term of $\{a_n\}$ is expressed as an entry in an element of ℓ^p. That is,

$$\{a_n\} = (a_1, \dots, a_n).$$

When we refer to a sequence such that $a_j \in \ell^p$, then we say that $\{a^{(n)}\}$ is a sequence of points in ℓ^p, and it is expressed as

$$\begin{aligned} a^(1) &= (a_1^{(1)}, a_2^{(1)}, \dots) \\ &\vdots \\ a^{(k)} &= (a_1^{(k)}, a_2^{(k)}, \dots) \\ &\vdots \end{aligned}$$

where each kth term is a point with infinite entries.

Now we show that our idea of convergence does not hold for sequences of points in ℓ^1, nor for ℓ^2, c_0, or ℓ^∞.

Lemma 4.3.2. There exist sequences of points in ℓ^1 such that $a^(k)_j \to a_j$, but $a^(k)$ does not converge.

Proof. Consider the sequence

$$\delta_n^{(k)} = \begin{cases} 1 & \text{if} n = k \\ 0 & \text{if} n \neq k \end{cases}.$$

Clearly, $\delta_n^{(k)} \to 0$ for every positive integer n. However, we also have that $d(\delta^{(k)}, 0) = 1$ for every positive integer k (here, 0 is the element of ℓ^1 with 0 in all the entries). Therefore, there is no ε such that $d(\delta^{(k)}, 0) < \varepsilon$. ☐

However, one direction of the previous theorem still holds.

Theorem 4.3.3. If $\{a^{(k)}\}$ is a sequence of points in ℓ^p which converges to $a \in \ell^p$, then for every positive integer n, $a_n^{(k)} \to a_n$.

Proof. Let $\varepsilon > 0$. Then there exists a positive integer K such that for all $k \geq K$, we have $d(a^{(k)}, a) < \varepsilon$. Therefore,

$$|a_n - a_n^{(k)}| \leq \left(\sum_{i=1}^{\infty} |a_i^{(k)} - ai|^p \right)^{1/p} = d(a^{(k)}, a) < \varepsilon.$$

Therefore, for every positive integer n, $a_n^{(k)} \to a_n$. ☐

4.4 Closed & Open Sets

4.4.1 Closed Sets

A closed interval $[a, b]$ of the real line $\mathbb{R}$ has the property that every convergent sequence completely contained in $[a, b]$ converges to a limit $L \in [a, b]$. We now generalize this property to a metric space M.

Definition 4.4.1 (Limit Point)**.** Let $X \subseteq M$, where M is a metric space. We say that $x \in X$ is a limit point of X if there exists a sequence $\{x_n\}$ where $x_n \in X$ for all n such that $x_n \to x$.

Definition 4.4.2 (Closed Set)**.** Let M be a metric space and $X \subseteq M$. We say that X is closed if every limit point of X is in X. That is, X contians all its limit points.

Definition 4.4.3. Let M be a metric space and let $X \subseteq M$. Then we denote the set of all limit points of X by $\overline{X}$.

It should be clear that $X \subseteq \overline{X}$; this is because if $x \in X$, then y is the limit of the sequence $x_n = y$ for all n. Therefore, y is a limit point of X, so $y \in \overline{X}$.

Lemma 4.4.1. A subset X of M is closed if and only if $X = \overline{X}$.

Proof. This follows from the definition. If X is closed, it contains all its limit points, then $\overline{X} \subseteq X$. But we already established that $X \subseteq \overline{X}$, so $X = \overline{X}$. ☐

Theorem 4.4.2 (Finite Union of Closed Sets)**.** Let $X_1, \ldots, X_n$ be cloesd subsets of a metric space M. Then

$$\bigcup_{i=1}^{n} X_i$$

is also closed.

Proof. The proof proceeds by induction. Let x be a limit point of $X_1 \cup X_2$. Then let $x_n \to x$ be an arbitrary sequence. Then either $x_n \in X_1$ for infinitely many n or $x_n \in X_2$. Without loss of generality, lets suppose there are infinitely many $x_n \in X_1$. Therefore, construct a subsequence x_{n_k} such that $x_{n_k} \in X_1$ for all k. Clearly, $x_{n_k} \to x$, and since X_1 is closed, $x \in X_1$. Therefore, $x \in X_1 \cup X_2$, so $X_1 \cup X_2$ is closed.

Now since $X_1 \cup \cdots \cup X_{n-1}$ is closed, and X_n is closed by assumption, their union is closed. This completes the proof. □

Theorem 4.4.3 (Intersection of Closed Sets)**.** Let Λ be an infinite family of closed sets. Then

$$S = \bigcap_{\alpha \in \Lambda} \alpha$$

is closed. Note that this family of sets may be uncountably infinite.

Proof. Let x be a limit point in S. Then there exists a sequence $x_n \in S$ such that $x_n \to x$. Since $x_n \in S$, then $x_n \in \alpha$ for all $\alpha \in \Lambda$. Because α is closed, then $x \in \alpha$ for all $\alpha \in \Lambda$. Therefore, $x \in S$. Since S contains all its limit points, S is closed. □

In sum, closed sets are sets in a metric space such that they contains their limit points. Now we turn to what is known as open sets.

4.4.2 Open Sets

Definition 4.4.4 (Open Ball)**.** Let (M, d) be a metric space, and let $x \in M$. Then for an $\varepsilon > 0$, we define the open ball of radius ε centered around x to be

$$B_\varepsilon(x) = \{c \in M : d(x, c) < \varepsilon\}.$$

In one dimension, this 1-ball is just an interval. In two dimensions it's a circle minus its boundary; in three it is a sphere minus its boundary. In math, when we say "ball," we really mean to say that we include the stuff inside the boundary too. Normally, a circle just refers to the boundary of a 2-ball and a sphere the boundary of a 3-ball. With this concept, we can now define an open set.

Definition 4.4.5 (Open Set)**.** Let M be a metric space, and $X \subseteq M$. Then we say that X is open if for all $x \in X$, there exists and ε such that

$$B_\varepsilon(x) \subseteq X.$$

Visually, this is a consequence of there being no boundary to the set; if there were, then we could center a ball on a point on the boundary, and there will always be a little bit hanging off the edge, no matter how small the ε.

Lemma 4.4.4. Let M be a metric space. Then M and $\emptyset$ are open.

Proof. If $\emptyset$ is not open, then there is an $x \in \emptyset$ such that $x \notin B_\varepsilon(y)$. But since $x \notin \emptyset$, this is false, so $\emptyset$ is closed. □

Theorem 4.4.5. Let M be a metric space, and $B_\varepsilon(x)$ be an open ball where $x \in M$. Then $B_\varepsilon(x)$ is open.

Proof. For a point $y \in B_\varepsilon(x)$, select a $\delta = \varepsilon - d(x, y)$. Then clearly,

$$B_\delta(y) \subseteq B_\varepsilon(x)$$

since y is distance $d(x, y)$ from x. Let $z \in B_\delta(y)$. Then we have that

$$d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \delta < \varepsilon$$

so $z \in B_\varepsilon(x)$. □

Now let us show a relationship between closed and open sets.

Theorem 4.4.6. Let X' denote the complement of X. Then X is open if and only if X' is closed.

Proof. Suppose X is open. For the sake of contradiction, assume X' is not closed. Let x be a limit point of X'. We must show that $x \in X'$. Suppose $x \notin X'$; then $x \in X$. Therefore, we can construct an open ball $B_\varepsilon(x) \subseteq X$. Now, construct a sequence $\{x_n\} \in X'$ such that $x_n \to x$. Since this sequence converges, then $\exists N : n \geq N \Rightarrow d(x_n - x) < \varepsilon$. However, this means that x_N is also in $B_\varepsilon(x)$, which implies that $x \in X \cap X'$, which is impossible.

Now assume X' is closed. Suppose for the sake of contradiction that X is not open. Then there exists an $x \in X$ such that for all $\varepsilon > 0$, $B_\varepsilon(x) \nsubseteq X$. This is equivalent to saying $\exists x_n \in B_\varepsilon(x) \cap X'$, since X' are the points not in X. Now consider the balls $B_{1/n}(x)$. There exists a point x_n in this ball. Therefore, construct a sequence such that

$$x_n \in B_{1/n}(x) \cap X'.$$

Clearly, since $d(x, x_n) < 1/n$, $x_n \to x$, so x is a limit point of X'. Since X' is closed, then $x \in X'$ since X' contains its limit points. However, we also had that $x \in X$, therefore leading to a contradiction. □

Theorem 4.4.7 (Intersection of Open Sets)**.** Let M be a metric space, and let $Y_1, dots, Y_n$ be open subsets of M. Then

$$Y_1 \cap Y_2 \cap \cdots \cap Y_n$$

is open.

Proof. Via deMorgan's Laws, we have that $(Y_1 \cap Y_2 \cap \cdots \cap Y_n)' = Y_1' \cup Y_2' \cup \cdots \cup Y_n'$. This is closed since Y_i is open, then Y_i' is closed and we showed this in theorem [THEOREM]. Therefore, since the complement is closed,

$$Y_1 \cap Y_2 \cap \cdots \cap Y_n$$

is open. □

Theorem 4.4.8. Let Γ be a family of open sets, not necessarily countable. Then

$$\bigcup_{\beta \in \Gamma} \beta$$

is open.

Proof. We have that $\forall \beta \in \Gamma,\ \beta'$ is closed. Therefore,

$$\bigcap_{\beta \in \Gamma} \beta'$$

is closed. This implies that the complement $\cup\Gamma$ is open, which completes the proof. □

Note that we can have that the intersection of an arbitrary family of sets is not open. For example,

$$\bigcap_{n=1}^{\infty} (-1/n, 1/n) = \{0\}$$

which is not open. It is also important to note that simply because something is not open doesn't mean it's cloesd. For instance,

$$[0, 1)$$

is neither open nor closed.

4.5 Continuous Functions on Metric Spaces

The following definition is motivated by the canonical definition of continuity.

Definition 4.5.1 (Continuity for Metric Spaces)**.** Let (M_1, d_1) and (M_2, d_2) be metric spaces. Let $f : M_1 \to M_2$. Then f is continuous at a point c if $\forall \varepsilon > 0$, there exists a $\delta > 0$ such that $d_1(x, c) < \delta \Rightarrow d_2(f(x), f(c)) < \varepsilon$. If f is continuous at every point in M_1, then we say f is continuous on M_1.

Now we can also generalize the equivalent sequential criterion for continuity.

Theorem 4.5.1 (Sequential Criterion for Continuity)**.** Let $f : M_1 \to M_2$. Then f is continuous at c if and only if for all sequences $x_n \to c$, we have that $f(x_n) \to f(c)$ in M_2.

Proof. Suppose that f is continuous. We want to show that $f(x_n) \to f(c)$. Then let $\varepsilon > 0$. We know that since f is continuous, we know that there exists a δ such that $d_1(x, c) < \delta \Rightarrow d_2(f(x), f(c)) < \varepsilon$. Moreover, since $x_n \to c$, we know that there exists a $N_\delta : n \geq N_\delta \Rightarrow d_1(x, c) < \delta$. Therefore, for $n \geq N_\delta$, $d_2(f(x), f(c)) < \varepsilon$.

Now suppose for all sequences such that $x_n \to c$ we have $f(x_n) \to f(c)$. Suppose f is not continuous. Then $\exists \varepsilon > 0$ such that for all $\delta > 0$, we have $d_1(x_n, c) < \delta$ and $d_2(f(x), f(c)) \geq \varepsilon$ for some $x \in M_1$. Therefore, we can a construct a sequence of these x_n's such that for all n,

$$d_1(x_n, c) < \frac{1}{n},\ d_2(f(x_n), f(c)) \geq \varepsilon.$$

Therefore, $x_n \to c$, but $f(x_n) \to L > f(c)$, which is a contradiction. Therefore, f must be continuous. ☐

Theorem 4.5.2 (Algebra of Metric Continuity)**.** If $f, g : M_1 \to M_2$ is continuous, then the following hold:

1. $|f|$ is continuous
2. $f + g$ is continuous
3. cf is continuous for some $c \in M_2$
4. fg is continuous
5. f/g is continuous if $g(x) \neq 0$ for all $x \in M_1$.

Theorem 4.5.3 (Equivalent Notions of Metric Continuity)**.** Let $f : M_1 \to M_2$. The following are equivalent :

1. f is continuous on M_1;
2. $f^{-1}(C)$ is closed whenever C is a closed subset of M_2;
3. $f^{-1}(U)$ is open whenever U is an open subset of M_2.

Corollary 4.5.4 (Composition of Functions)**.** Let $f : M_1 \to M_2$ and $g : M_2 \to M_3$, where f and g are continuous. Then the composition $g \circ f : M_1 \to M_3$ is continuous.

In many respects, this section is simply a generalization of what we already knew to be true for functions on the real line. We have, like in a few sections ago, that continuity and sequential continuity are really equivalent ideas, both of which will be useful in making proofs.

4.6 Compact Metric Spaces

What follows is a very important concept when it comes to anything related to higher mathematics. This is something to be studied very carefully.

Recall from the Heine-Borel theorem that a closed and bounded interval $[a, b]$ has the property that any infinite cover can in some sense be "reduced" to a finite sub-cover.

Definition 4.6.1. An open cover of a metric space M is a collection Γ of open subsets of M such that $M = \cup\Gamma$. A subcover of Γ is a subcollection Γ^* such that $M = \cup\Gamma^*$.

Now we can generalize this notion of reducability:

Definition 4.6.2. A metric space M is said to be compact if *every* open cover of M has a finite subcover.

That is, we can take any collection Γ and have the subcollection $\cup\Gamma^* = \beta_1 \cup \cdots \cup \beta_n$, where $M = \Gamma^*$. Crucially, we have to be able to do this for *every* open cover. It is also worth noting that M could be a smaller metric space contained within a larger one; for instance, we know from the Heine-Borel theorem that the metric space $[a, b]$ is compact, but $\mathbb{R}$ itself is not compact, since $\cup\{(-n, n) : n \in \mathbb{N}\}$ has no finite subcover. Similarly, we see that $\mathbb{R}^n$, ℓ^1, ℓ^2, ℓ^∞, and c_0 are not compact.

Definition 4.6.3 (Relative Metric)**.** Let M be a metric space, and $X \subset M$. Then we say that d' is the metric for X relative to M, defined by

$$d'(x, y) = d(x, y) \ \text{ for} x, y \in X.$$

This is important, since if X is a subset of a metric space, and we want to talk about X being a metric space, we must use the relative metric. This is because we can have odd conclusions if we don't. For instance, the set $[0, 1/2)$ is open in the metric space $[0, 1]$, but it is not open in $\mathbb{R}$. This is because open balls centered around 0 in the metric space $[0, 1]$, denoted

$$B_\varepsilon^{[0,1]}(0)$$

will be completely contained in $[0, 1/2)$; that is, we "cut off" the part of the ball that is outside of the metric space. However, if we are in the metric space of $\mathbb{R}$, then the "overhang" exists and therefore no ball around 0 can be completely contained in $[0, 1/2)$. Therefore, we are forced to consider more specific notions of closed and open in relative metric spaces.

Theorem 4.6.1. Let M be a metric space, and let X be a subset of M with the relative metric, and let $Y \subseteq X$. Then we have

1. Y is open in X if and only if $Y = U \cap X$ where U is open in M; that is, it is the intersectoin of X and an open set in M.
2. Y is closed X if and only if $Y = C \cap X$, where C is a closed set in M.

With this in mind, we can prove that the metric space $[a, b] \subset R$ is compact. Notice how this is slightly different than the Heine-Borel theorem, which only tells us that there are finite subcovers in for subsets of a metric space $\mathbb{R}$, but if we wish to treat $[a, b]$ as its own metric space (which is what the definition of compactness considers), then we need the relative metric.

Theorem 4.6.2 (Closed Sets are Compact)**.** Let $[a, b]$ be a closed interval in the metric space M. Then $[a, b]$ is compact.

Proof. Let Γ be an open cover for $[a,b]$. For every $x \in [a,b]$, there exists a $\beta_x \in \Gamma$ such that $x \in \beta_x$. Since β_x is open, there is an open ball in $[a,b]$ centered around x such that $B^{[a,b]}_{\varepsilon_x}(x) \subseteq \beta_x$. Now if we consider the collection of all these balls, we have

$$\mathcal{B} = \{B^{\mathbb{R}}_{\varepsilon_x}(x) : x \in [a,b]\}$$

so clearly, $[a,b] \subseteq \cup\mathcal{B}$. Now we can invoke the Heine-Borel theorem to reduce the elements of $\mathcal{B}$ into a finite number; that is, there exist $x_1, \ldots, x_n$ such that

$$[a,b] \subset \bigcup_{i=1}^{n} B^{\mathbb{R}}_{\varepsilon_{x_i}}(x_i)$$

Moreover,

$$[a,b] = \bigcup_{i=1}^{n} [a,b] \cap B^{\mathbb{R}}_{\varepsilon_{x_i}}(x_i).$$

But then since the balls are open in $\mathbb{R}$, relative to $[a,b]$ they must also be open. Therefore,

$$[a,b] = \bigcup_{i=1}^{n} B^{[a,b]}_{\varepsilon_{x_i}}(x_i) \subset \bigcup_{i=1}^{n} \beta_{x_i}.$$

Thus,

$$[a,b] = \bigcup_{i=1}^{n} \beta_{x_i},$$

which is a finite subcover, so $[a,b]$ is compact. □

The other theorems of continuity which follow from this can be generalized:

Definition 4.6.4. Let $f : X \to \mathbb{R}$. We say that f is bounded on X if there exists a real number M such that $|f(x)| \leq M$ for all $x \in X$.

Lemma 4.6.3. If $f : M \to \mathbb{R}$ is continuous at a, then there exists an open set U containing a such that f is is bounded on U.

Proof. Let (M,d) be a metric space. Take $\varepsilon = 1$. Since f is continuous, we know that there exists a $\delta > 0$ such that if $d(x,a) < \delta$, we have $|f(x) - f(a)| < \varepsilon$. Thus if $U = B_\delta(a)$,

$$|f(x)| \leq |f(x) - f(a)| + |f(a)| < 1 + |f(a)|.$$

□

Now we generalize the notion of boundedness on a closed interval. Recall that we have the theorem stating that if f is continuous on a closed and bounded interval, it must be bounded.

Theorem 4.6.4. If f is continuous on a compact metric space M, then f is bounded on M.

Proof. We have that for every $x \in M$, then there is an open set U_x such that $x \in U_x$ and f is bounded on U_x by A_x. Therefore, we can take the union of all such $U_x : x \in M$. Since M is compact, we can reduce it to

$$U_1 \cup \cdots \cup U_n.$$

Now let $M^* = \max(A_1, \ldots, A_n)$. Therefore, this is a bound on all of M. □

4.7 A Dual Version of Compactness

In the previous section, we were made aware of the fact that the Heine-Borel theorem has its analogues in an arbitrary metric space. Therefore, much like the Heine-Borel theorem is equivalent to the Bolzano-Weierstrass theorem, so too is this the case for compact spaces.

In this section, we show that a space M is compact if and only if it satisfies the Bolzano-Weierstrass property.

Theorem 4.7.1. If M is a compact metric space, then every sequence in M has a convergent subsequence.

Proof. Let x_n be a sequence in the compact metric space M. Then suppose there is no convergent subsequence of x_n. Then, for all $x \in N$, there exists an $\varepsilon > 0$ such that the ball $B_\varepsilon(x)$ contains only finitely many terms of the sequence x_n. If not, then x_n would get arbitrarily close to a particular value, and therefore a subsequence x_{n_k} would be convergent. Now, we take the union of all these open balls:

$$M \subseteq \mathcal{B} = \bigcup_{x \in M} B_\varepsilon(x)$$

By the definition of compactness, the set $\mathcal{B}$ can be reduced to a finite collection such that

$$M \subseteq \mathcal{B}' = \bigcup_{i=1}^{n} B_\varepsilon(x_i).$$

However, each $B_\varepsilon(x_i)$ has only finitely many of the terms of x_n; since M is a subset of $\mathcal{B}'$, then this implies there are only finitely many terms of the infinite sequence x_n, which is impossible. Therefore, x_n has a convergent subsequence in M. $\square$

This result can also be interpreted as saying the Heine-Borel theorem for $\mathbb{R}$ is equivalent to the Bolzano-Weierstrass theorem, since we used BW to prove HB, and now we just used HB to prove BW. Now we wish to show the reverse direction; that if every sequence in M has a convergent subsequence, then M is compact. For this, we'll have to split our work up across a few lemmae.

Lemma 4.7.2. Let M be a metric space where every sequence has a convergent subsequence. Then let $\varepsilon > 0$. Then there exist a finite $x_1, \ldots, x_n$ such that

$$M = \bigcup_{i=1}^{n} B_\varepsilon(x_i).$$

Lemma 4.7.3. Let M be a metric space where every subsequence converges. Then if Γ is an open cover, there exists an $\varepsilon > 0$ such that if $x \in M$, $B_\varepsilon(x) \subset U$ for some $U \in \Gamma$.

Theorem 4.7.4. Let M be a metric space where every subsequence converges. Then M is compact.

Now let's translate our definition of boundedness to compact metric spaces.

Definition 4.7.1. A subset S of a metric space is bounded if for all $x, y \in S$, $d(x, y) \leq A$ for some A.

Theorem 4.7.5. If C is a compact subset of M, then C is closed and bounded.

this always holds, except the converse is not necessarily true. Let $X = \{x \in \ell^1 : d(x, 0) = 1\}$ Considering $\mathbb{R}^\infty$, the Kronecker sequence $\{\delta^{(k)}\}$ has no convergent subsequence, yet X is still closed and bounded. However, it is possible to state a partial converse.

Theorem 4.7.6. Let C be a closed subset of a compact metric space M. Then C is compact.

Proof. Let x_n be a sequence in C. Since M is compact, x_n has a subsequence x_{n_k} such that x_{n_k} converges to some value L in M. However, C is closed; since L is a limit point of C, $L \in C$. Therefore, every subsequence of C has a subsequence which converges to a point in C. Therefore, C is compact. $\square$

Although the same cannot be said for $\mathbb{R}^\infty$, we can show an equivalence between these ideas.

Theorem 4.7.7. A subset $X \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.

4.8 Continuity on Compact Spaces

Theorem 4.8.1 (Preservation of Compactness)**.** If $f : M_1 \to M_2$ is continuous, and M_1 is compact, then $f(M_1)$ is also compact.

Proof. It should be shown that any subsequence $x_{n_k} \to c$, $f(x_n)$ has a subsequence which converges to $f(c)$, because of continuity. $\square$

Corollary 4.8.2. If $f : M_1 \to M_2$ is continuous, then $f(M_1)$ is closed and bounded.

Theorem 4.8.3. If f is a continuous function from a compact metric space M_1 to a metric space M_2, then f is uniformly continuous on M_1.

4.9 Complete Metric Spaces

The idea of completeness bears a lot of relevance to how we developed the idea of $\mathbb{R}$. The idea of a complete metric space can be seen as a space without having any holes. That is, a space is complete if every Cauchy sequence converges to a point in M.

Definition 4.9.1. A sequence $x_n \in M$ is said to be Cauchy if for every $\varepsilon > 0$, There exists an N such that for all$n, m \geq N$, $d(x_n, x_m) < \varepsilon$.

While it is true that every convergent sequence is Cauchy, it is not true that every Cauchy sequence is convergent. For that, we need the idea of completeness.

Definition 4.9.2. A metric space M is said to be complete if every Cauchy sequence is convergent.

A list of complete metric spaces:

1. $\mathbb{R}$
2. $\mathbb{R}^n$
3. ℓ^1
4. c_0
5. ℓ^2
6. ℓ^∞

5 The Riemann-Stieltjes Integral

5.1 Review of Riemann Integral

We now study the integral of a positive function, so we can later generalize to other functions. We can think of the integral as the area under the curve of a positive function. Foremost, we slice the domain into N equidistant pieces, where the length of $\Delta x = b - a/N$.

For any interval, let m_i be the minimum value attained by f on that interval, and let M_i be the maximum value. Therefore, the true area A, which is the sum of the true area of each interval A_i, is bounded by

$$\sum m_i \Delta x \leq \sum A_i \leq \sum M_i \Delta x.$$

This approximation will get better when Δx gets smaller. At any given interval j, we can split it into two pieces, in half. Now for each of the two new regions, say i and $i+1$, we have that $M_i \leq M_j$ and $m_j \leq m_i$, and the same for $i+1$; this is because the maximum must be the maximum for at most one of two of the intervals, and same for the minimum, which must be bigger than the minimum for both.

Therefore, as Δx gets smaller, the sequence $\sum M_i \Delta x$ gets smaller monotonically, and the minimum sum gets larger monotonically.

Definition 5.1.1. Let $[a, b]$ be an interval on $\mathbb{R}$. Then we call

Definition 5.1.2 (Upper sum, lower sum)**.** Define the upper sum to be

$$U_N = \sum_{i=1}^{N} M_i \Delta x_i, \ M_i = \sup_{x_{i-1} \leq x \leq x_{i+1}} f(x)$$

and the lower sum to be

$$L_N = \sum_{i=1}^{N} m_i \Delta x_i, \ m_i = \inf_{x_{i-1} \leq x \leq x_{i+1}} f(x)$$

Now for any partition, we can also denote the maximum length of any one interval in the partition to be h, so our upper sum only depends on h, written as U_h.

Now we want to show that

$$L_h \leq A \leq U_h.$$

Now a monotone decreasing sequence which is bounded below will always have an infimum. Likewise for an increasing sequence bounded above. Therefore, we can make the bound

$$\sup_h L_h \leq A \leq \inf_h U_h$$

and from this make the following definition.

Definition 5.1.3. If $\sup L_h = \inf U_h = A$, then we call A the area under the function f as $a \leq x \leq b$.

Our motiviation for making this not depend on the partitions we choose, but only the maximum length, means that we can rule out any "cheating" by judicious choice of partition. Now we ask, which functions do or don't have well-defined areas?

Theorem 5.1.1. If $f : [a, b] \to \mathbb{R}$ is continuous, then

$$\int_a^b f(x) dx := A$$

exists. The symbol for integration comes from Liebniz, and is an elongated "s" meaning sum.

Proof. A continuous function closed, bounded interval $[a, b]$ is uniformly continuous. Now we know there is a δ such that for any $x, y \in [a, b]$, $|f(x) - f(y)| < \frac{\varepsilon}{(b-a)}$. Therefore, let $\delta = h$. Thus, for any interval, the function differs by ε. Therefore,

$$0 \leq U_\delta - L_\delta = \sum_i (M_i - m_i)\delta_i \leq \frac{\varepsilon}{(b-a)} \sum \delta_i = \frac{\varepsilon}{(b-a)}(b-a) = \varepsilon.$$

Therefore, we see that the supremum of the lower sum is the infimum of the upper sum, meaning that the area for f is well-defined. □

Note that f does not have to be positive, since f can be written as the difference between two positive and negative functions. For a continuous function on a closed and bounded interval, we can simplify our life; our partition can be of equal size, and we can select any value of f in an interval, since the difference will always be less than an arbitrary ε. Therefore, any such sum will converge to the area.

The function need not be continuous for an area to exist underneath it. Imagine a jump discontinuity; we know that at the jump, there will be at most one interval within which the jump is contained. But since our $h \to 0$, this means that it will not contribute anything to the overall sum. This is true for finitely many such discontinuities.

But even one point of discontinuity can give you trouble. Consider

$$f(x) = \begin{cases} 0 & x = 0 \\ \frac{1}{x} & 0 < x \leq 1. \end{cases}$$

This has no integral; the lower sum at the point i/N will be N/i. The lower sum of the first rectangle, then, will be

$$L_1 = \frac{N}{h}\frac{1}{N} = \frac{1}{h}.$$

Therefore, since the lower sum is unbounded, the total area must be unbounded. In this case, we can say that the integral diverges. The Riemann integral for unbounded functions may not exist. However, they may. Modifying our first function,

$$f(x) = \begin{cases} 0 & x = 0 \\ \frac{1}{\sqrt{x}} & 0 < x \leq 1 \end{cases}.$$

This has an integral. Moreover, there are bounded functions who do not have integrals. For instance, they can have too many jumps. Consider

$$g(x) = \begin{cases} 1 & x \text{ is irrational} \\ 0 & x \text{ is rational} \end{cases}.$$

This has no Riemann integral, but it does have an *Lebesgue* integral of 1. These sorts of functions are the motivation for other types of integrals. The first general idea of a function having an integral is that the set of its discontinuities is measure zero. We will discuss what this means in later sections.

5.2 The Riemann Integral

The idea is now to see if instead of forming the upper and lower sums, we want to see if we can take *any* point ξ_i that's in the ith interval. Then we form the sum over all intervals:

$$\sum_i f(\xi_i)\Delta x_i.$$

Then the integral exists if this has a limit. But in what sense does this have a limit? We can say it has one if the size of the partition goes to zero. That is, $\forall \varepsilon > 0, \exists \delta$: for any partition $P, \|P\| < \delta \Rightarrow$

$$\left|\sum f(\xi_i)\Delta x_i - A\right| < \varepsilon \ \forall \xi_i \in [x_i, x_{i+1}].$$

But there is also another way of taking a limit of a partition. We can say that $\forall \varepsilon > 0$, $\exists P^*$: for all refinements P of P^*, then

$$\left|\sum f(\xi_i)\Delta x_i - A\right| < \varepsilon \ \forall \xi_i \in [x_i, x_{i+1}].$$

This is equivalent to the above one, but the difference will be clear when we consider the Riemann-Stieltjes integral.

According to these definitions,

1. Continuous functions are integrable.
2. Finitely discontinuous functions are integrable.

5.2.1 The Fundamental Theorem of Calculus

Students get upset that Calculus I is only a one-theorem class, which is just the fundamental theorem of calculus. We had better remember this theorem, then. Informally put, if $f'(x) = g(x)$, then

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

If we look at the summation

$$\sum g(\xi_i)\Delta x_i,$$

we know that for some ξ_i in the interval (which we get to choose),

$$\frac{f(x_{i+1}) - f(x_i)}{\Delta x_i} = f'(\xi_i).$$

We can use the mean value theorem to prove the fundamental theorem of calculus.

5.3 Oscillation of a Function

One goal we have in Analysis is to measure how continuous a given function is. Consider

$$f(x) = x.$$

It's continuous at 0. Now consider

$$f(x) = \sqrt{|x|}.$$

This is also continuous at 0, but our gut feeling should be that it's somehow less continuous. Now we introduce the oscillation of a function f on an interval $[a, b]$.

Definition 5.3.1 (Oscillation on an Interval)**.** Let $f : [a, b] \to \mathbb{R}$. We define the *oscillation* of a function to be

$$\begin{aligned}\Omega_f([a,b]) :&= \sup_{x,y\in[a,b]} |f(x) - f(y)| \\ &= \sup_{x\in[a,b]} f(x) - \inf_{x\in[a,b]} f(x).\end{aligned}$$

Simply put, this is a quantity which measures how much a function f changes on an interval.

Similarly, we can define oscillation for a point:

Definition 5.3.2 (Oscillation at a Point)**.** Let $f : [a, b] \to \mathbb{R}$. Then we say the oscillation of f at a point x is just

$$\omega_f(x) := \lim_{h \to 0} \Omega_f([x - h, x + h]).$$

This should fit with our intuition. It measures how flat a function is at a point x. Note that if $x = a$, then we can extend the domain of the function so $f(x) = a$ for $x < a$, and similarly for b.

Theorem 5.3.1. If f is continuous at x, then $\omega_f(x) = 0$.

Proof. This proof follows from the definition of a continuous function. □

This said, Ω_f can be thought of as measuring the continuity of f, and ω_f can be thought of as measuring the discontinuity.

Given a function f, we know that if $\omega_f(0) = 1$ is discontinuous. An example of such a function would be a piecewise function, or $\frac{1}{2}\sin(1/x)$. The overall goal of these measures, therefore, is to quantify how many discontinuities a function can have for it to still be Riemann-integrable. This ω_f will tell us how to measure the different jumps.

Example. Let $f : [a, b] \to \mathbb{R}$ such that f is monotone. We can prove that the set of discontinuities of f must indeed be countable; this is becasuse for every jump, we can find a rational between the two values. Therefore, we have an injection from the set of discontinuities to the rationals. Now we show how to get to a similar result by using ω_f. Let

$$E_n := \{\omega_f(x) \geq \frac{1}{2^n} : x \in [a, b]\}.$$

For a bounded function, we can only have finitely many such discontinuities on $[a, b]$. If we didn't, f would be unbounded since it is montone. Thus, E_n is finite for any n we choose. The set of discontinuities is therefore

$$D = \bigcup_{i=1}^{\infty} E_i.$$

But since each E_i is finite, then we know that D must be countable, since the infinite union of finite sets is at most countable.

5.4 Sets and Measure Zero

We have two ideas of sets.

Definition 5.4.1 (Content Zero)**.** We say a set A has content zero if

$$A \subset \bigcup_{i=1}^{N} I_i, \; I_i = (a_i, b_i).$$

and the length of this union can be made arbitrarily small:

$$\sum |I_i| < \varepsilon.$$

Immediately, we can show the following theorem:

Theorem 5.4.1. If f is continuous except on a set of content zero, then f is integrable.

Proof. We can cover all the discontinuities by N intervals whose length is arbitrarily small (say ε/M, where M is $sup_{x\in[a,b]}f(x)$.) Now if we look at the complement,

$$[a,b] \setminus \bigcup_{i=1}^{N} I_i,$$

then the integral exists since the function is continuous. Looking at the Riemann sum,

$$\sum f(\xi_i)\Delta x_i,$$

we see that for the points of discontinuity,

$$\left|\sum f(\xi_i)\Delta x_i\right| \leq M \sum |I_i| < M\varepsilon/M = \varepsilon.$$

□

This concept will be important to prove the ultimate condition for Riemann integrability; namely that a function is Riemann integrable if and only if the set of its discontinuity is measure zero.

5.4.1 Measure Zero and the Cantor Set

Definition 5.4.2. Let $E \subset \mathbb{R}$. We use the following notation and say $\mu(E) = 0$, or E is measure zero, if $\forall\varepsilon > 0$, if there exists an open cover such that

$$E \subset \bigcup_{i=1}^{\infty} I_i$$

such that

$$\sum_{i=1}^{\infty} |I_i| < \varepsilon.$$

For instance, any countable set E, $\mu(E) = 0$. This is beause if $E = \{x_1, x_2 \ldots,\}$ then let $x_1 \in I_1$ where $|I_1| < \varepsilon/2$, and let $x_2 \in I_2$ where $|I_2| < \varepsilon/4$, and so forth. Then the sum of all the lengths is less than ε. Since the rationals are countable, they are measure zero. However, we can have uncountable sets with measure zero. Consider the Cantor set, which can be constructed thusly.

Take the interval $[0, 1]$, and divide it into thirds, throwing the (open) middle third away. Then we have our original set $C = [0, 1/3] \cup [2/3, 1]$. Repeat the pattern with these two intervals as well, and repeating this process indefinitely. At the first step, we throw away a set of measure $1/3$. On the next one, we throw two sets away of measure $1/9$, for a combined $2/3^2$. On the nth step, we throw away a set of measure $2^n/3^{n+1}$. This is a geometric sum:

$$\frac{1}{3}\left(1 + \frac{2}{3} + \cdots + \frac{2^n}{3^n} + \ldots\right) = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1.$$

Therefore, the remainder of the Cantor set, $\mu(C) = 0$. There are elements in this set, since we keep $1/3$, $2/3$, etc. However, the Cantor set has no interval. It is also closed, since each step leaves you with a closed set. The Cantor set is a countable intersection of closed sets, so it's closed. Essentially, you're keeping numbers in base 3 who can be expressed as a decimal of 1's and 2's. This is a closed set with no interval, so it is nowhere-dense, but its cardinality is the same as the real line via the bijection

$$1 \mapsto 0$$
$$2 \mapsto 1$$

which gives us the same diagonalization argument we used to prove that the reals are not countable. Now we are in a position to prove a necessary and sufficient condition for Riemann-integrability.

Theorem 5.4.2. Let $f : [a, b] \to \mathbb{R}$, and let f be bounded. Then we know that f is Riemann integrable if and only if f is continuous except on a set of measure zero.

Proof. Recall that we defined the set of discontinuities as

$$E = \{x \in [a, b] : \omega_f(x) > 0\}$$

Which can be decomposed as the union of

$$E_n = \{x \in [a, b] : \omega_f(x) \geq \frac{1}{n}\}.$$

We know that E_n is closed, since its complement is open. Moreover, it's bounded since $a \leq x \leq b$. We know that on $\mathbb{R}$, closed and bounded sets are compact, so E_n is compact. By the Heine-borel theorem, we know that

$$E_n \subseteq \bigcup_{i=1}^{K} I_i, \ \sum_{i=1}^{K} |I_i| < \varepsilon.$$

Now let's look at the Riemann sum. [FINISH] □

5.5 The Riemann-Stieltjes Integral

In probability theory, we define something called a distribution function. Here, we refer to it as α, where α is monotone increasing. One question you ask is what the probability of an event happening between a and b. If α is differentiable, then we can say that it is

$$\int_a^b \alpha' dt.$$

In practice, however, α may not be differentiable. Therefore, we may sometimes write

$$\int_a^b d\alpha$$

which we will give rigorous meaning. The Riemann-Stieltjes integral is very apt to deal with these kinds of functions. There are two starting points for this integral.

Definition 5.5.1 (Riemann-Stieltjes Sum)**.** Let $f : [a, b] \to \mathbb{R}$ Then we define

$$S(f, P, T) = \sum f(t_i)\Delta\alpha_i$$

where $P = \{x_1, dots, x_k\}$ is a partition, $T = \{t_1, \ldots, t_k : t_i \in [x_{i-1}, x_i]\}$, and $\Delta\alpha = \alpha(x_i) - \alpha(x_{i-1})$.

Now we have a definition of convergence:

Definition 5.5.2. Let f be a bounded function on $[a, b]$. Then f is Stieltjes integrable on $[a, b]$ if and only if there exists a number I such that for every $\varepsilon > 0$, there exists a partition P such that for every P^* that is a parition of P ($P \subset P^*$), then

$$|S(f, P^*, T) - I| < \varepsilon$$

for any points T.

Notice in this definition, we don't require α to be increasing.
The next definition is similar to the upper and lower sum:

Definition 5.5.3. Let $P = \{x_1, \ldots, x_k\}$, let $M_i = \sup_{[x_{i-1},x_i]} f$, and $m_i = \inf_{[x_{i-1},x_i]} f$. Then define

$$U(f,P) = \sum M_i \Delta\alpha_i$$

and

$$L(f,P) = \sum m_i \Delta\alpha_i.$$

Then we let

$$\overline{\int_a^b} f d\alpha = \inf_P U, \ \underline{\int_a^b} f d\alpha = \sup_P L.$$

Then f is Stieltjes integrable with respect to α if

$$\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha.$$

Notice how in this definition, monotinicity of α is required, since otherwise we would not be able to assert that the upper sum is greater than the lower sum, and thus could not take the infimum.

Let's work with the second definition for a little bit, just to get a feeling of what we're defining.

Example. Let

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

We would like to compute

$$\int_{-1}^1 dH.$$

If we want to form the upper sum, let us first take a partition $\{-1, 0, 1\}$. On $[0, 1]$, the supremum of the function is $M = 1$. The infimum is $m = 1$. The supremum on $[-1, 0]$ is $M = 1$, and the lower $m = 1$. For the piece $[0, 1]$, $\delta\alpha_2 = 0$, and for the other, $\delta\alpha_1 = 1$. Therefore, the upper sum is just

$$U = M\Delta\alpha_1 + M\Delta\alpha_2 = 1$$

and the lower sum is

$$L = m\Delta\alpha_1 + m\Delta\alpha_2 = 1.$$

Thus we have the Inequality

$$L \leq \underline{\int_{-1}^1} dH \leq \overline{\int_{-1}^1} dH \leq U$$

Since $L = U = 1$, the upper and lower integrals are the same. Therefore it is Stieltjes integrable.

It should be verified that HdH is not Stieltjes integrable, since $L = 0$ and $U = 1$ for all partitions P. However, define

$$\widetilde{H}(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}.$$

Then we have that $\widetilde{H}dH$ is integrable on $[-1, 1]$, and the integral is always 0.

The purpose of the example is to show that there is a problem when there is discontinuity in the integrand and the α. When their points of discontinuity overlap, as in the HdH case, the definition becomes very delicate. Once we learn how to integrate a step function, we can calculate the Stieltjes integral to practically anything.

Theorem 5.5.1. If $\alpha \in C^1([a,b])$, then

$$\int_a^b f d\alpha = \int_a^b f(x)\alpha'(x)dx.$$

Proof. Let $S = \sum f(t_i)\alpha'(t_i)\Delta x_i$. Then we also have that $\sum f(t_i)\Delta\alpha_i = \sum f(t_i)\alpha'(\xi_i)\Delta x_i$ by the mean value theorem. But since α' is continuous, we know that $x_i \leq t_i, \xi_i \leq x_{i+1}$, which can be made arbitrarily small. So by the continuity of α', we have $|\alpha'(t_i) - \alpha'(\xi_i)| < \varepsilon$. Therefore, the sum is just

$$M\varepsilon(b-a)$$

where M is the upper bound of f on the interval. This is arbitarily small, so the two sums the same. $\square$

This is nothing new; the only novelty in the Stieltjes integral is when α is not necessarily differentiable. In many cases, however, integration by parts can save us.

Theorem 5.5.2. For α, f:

$$\int_a^b f d\alpha = -\int_a^b \alpha df + f(b)\alpha(b) - f(a)\alpha(a)$$

If f is itself C^1, then the left hand side can be thought of as a Riemann integral.

Proof. Prove by summation by parts. The summation by parts formula is

$$\sum a_n b_n = A_k b_{k+1} - \sum_n^k An(b_{n+1} - b_n)$$

Therefore, we take the sum

$$\sum_{i=1}^k \alpha(t_i)\Delta f_i = \sum_{i=1}^k \alpha(t_i)(f(x_i) - f(x_{i-1})).$$

The left hand side, via the summation by parts formula, becomes

$$[f(x_k) - f(x_0)]\alpha(t_{k+1}) - \sum_{i=1}^k (f(x_i) - f(x_0))(\alpha(t_{i+1}) - \alpha(t_i))$$

This becomes

$$[f(x_k) - f(x_0)]\alpha(t_{k+1}) + f(x_0)(\alpha(t_{k+1}) - \alpha(t_0)) - \sum_{i=1}^k f(x_i)(\alpha(t_{i+1}) - \alpha(t_i))$$

We cam do a cancellation, getting

$$f(x_k)\alpha(t_{k+1}) - f(x_0)\alpha(t_0) - \sum_{i=1}^k f(x_i)(\alpha(t_{i+1}) - \alpha(t_i))$$

Now you might notice that there is technically no t_{k+1} term in our partition T. Therefore, we are free to choose, as long as it obeys the partition behavtior. That is, we can elect $t_{k+1} = b$. Likewise, we don't have t_0, so we let $t_0 = a$. In order to resolve the admixture of t_i's and x_i's in our formula, $\square$

Example. We know that

$$\int_{-1}^{1} d\alpha = \alpha(1) - \alpha(-1).$$

This is because, in our Riemann-Stieltjes sum, all the terms cancel save for the first and last of our partition.

Now, consider a continuous f, and define $\alpha(x)$:

$$\alpha(x) = \begin{cases} A & x < x_0 \\ B & x = x_0 \\ C & x > x_0 \end{cases}.$$

Now create a partition $P = \{a, x_1, \ldots, b\}$ where $x_i < x_{i+1}$. When we take a partition, we have a neighborhood around x_0 that is as small as we would like. For this interval $[x_j, x_{j+1}]$, we have that the jump in α from x_0 to x_{j+1} is $C - B$. Then we have to multiply this by the value of f over this interval. Since f is continuous, we have

$$(\alpha(x_{j+1}) - \alpha(x_0))f(t) = (C - B)f(x_0) + \varepsilon/2.$$

Likewise, we have that

$$(\alpha(x_0) - \alpha(x_j))f(t) = (B - A)f(x_0) + \varepsilon/2.$$

Combining the equations, we get $(C - A)f(x_0) + \varepsilon$. We have that C is the limit from the right, and A is the limit from the left. Moreover, as long as f is continuous at x_0, the value of α at x_0 does not figure into the equation. Therefore, we have shown that for such a function,

$$\int_a^b f(x)d\alpha = f(x_0)(\alpha(x_0^+) - \alpha(x_0^-)).$$

Now we cannot do this if α and f are discontinuous at the same place, since we will not be able to leverage the continuity of either f or α in the same way. It can be done if α is continuous from the left and f from the right or vice versa.

Example. Consider the floor function $\lfloor x \rfloor$, and a differenetiable function f. We want to compute

$$\int_a^b f'(x)\lfloor x \rfloor dx = \int_a^b \lfloor x \rfloor df.$$

Integrating by parts, we see that this equals

$$-\int_a^b f(x)d\lfloor x \rfloor + f(b)\lfloor b \rfloor - f(a)\lfloor a \rfloor.$$

As we discussed, the integral is just the value of f at the discontinuities times the difference of the discontinuity (we know f is continuous since it is differentiable). Thus, we have

$$\int_a^b f(x)d\lfloor x \rfloor = -\sum_{a<n<b} f(n) + f(b)\lfloor b \rfloor - f(a)\lfloor a \rfloor$$

6 Sequences and Series of Functions

6.1 Sequences of Functions

Recall that we regard functions as points in a vector space, endowed with a metric. For instance, we have the space of continuous functions on $[a, b]$, denoted by $C([a, b])$. Let's suppose we have a sequence of functions such that $f_n \in C([a, b])$. What would it mean for such a sequence to converge?

Definition 6.1.1 (Uniform convergence)**.** Let f_n be a sequence of functions such that $f_n \in C([a, b])$. Then we say the sequence f_n converges if

$$\exists f \in C([a, b]) : \|f_n - f\|_C = \sup_{x \in [a,b]} |f_n(x) - f| \to 0$$

as $n \to \infty$.

It should be noted that this is a way to measure the distance between f_n and f independent of x. An alternative way might be

$$\|f_n - f\|_p = \left(\int_a^b |f_n(x) - f(x)|^p dx \right)^{1/p}.$$

It should be noted that $\|\cdot\|_C$ is the p-norm as $p \to \infty$.

Now consider

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx,$$

then we really mean ordinary convergence in $\mathbb{R}$, since that's all a definite integral is—a number. Is such a statement true?

Theorem 6.1.1. *Suppose $f_n \to f$ uniformly. Then*

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx.$$

Proof. Let $\varepsilon/(b-a) > 0$ such that $\|f_n - f\|_C < \varepsilon/(b-a)$. Then consider the difference

$$\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| = \left| \int_a^b f_n(x) - fxdx \right| \leq \left| \int_a^b \sup |f_n(x) - fx| dx \right| < \frac{\varepsilon}{b-a}(b-a) = \varepsilon.$$

Therefore the integral converges. □

However, we can imagine other useful notions of convegence which are possibly weaker.

Definition 6.1.2 (Pointwise Convergence)**.** We say a sequence f_n converges pointwise to f if

$$f_n(x) \to f(x) \forall x \in [a, b].$$

Note that x is fixed, so each $f_n(x)$ is a number. We look to see if the sequence of these numbers converges to the number $f(x)$. The main distinction is convergence between numbers versus convergence between functions.

Under this notion, we will show that the integrals may not end up being the same. Note that if a function converges uniformly it also converges pointwise, but the converse might not be true.

Example. Consider the sequence of functions $f_n(x) = x^n$ on $[0,1)$ This does not converge uniformly. This is because for any n, we can find $x \in [0,1)$ such that $x^n > 1/2$; indeed, $x > \left(\frac{1}{2}\right)^{1/n}$ does the job. This is always a number strictly less than 1. However, for this particular x, its pointwise limit is 0 since it is less than 1 and we take the exponent to infinity. Therefore,

$$\sup_{x\in[0,1)} |f_n(x) - f(x)| > 1/2$$

so we don't have uniform convergence.

Now let's go over three sequences that may prove useful when trying to find a counterexample.

1. Consider the following function:

$$f_n(x) = \begin{cases} n & 1/n \leq x \leq 2/n \\ 0 & \text{otherwise} \end{cases}.$$

   This is a function that converges pointwise to 0 for $x \in [0,1]$, but not uniformly, Moreover, its integral is always 1, but the integral of $f = 0$ is 0. This shows that the integrals of two functions may not be the same for pointwise convergence. Moreover, this means that it doesn't converge to 0 in L^1. In L^2, its integral is infinite.

2. Consider the function

$$f_n(x) = \begin{cases} 1 & n \leq x \leq n+1 \\ 0 & \text{otherwise} \end{cases}.$$

   This converges pointwise to 0 on $\mathbb{R}$. We do not have uniform convergence. Moreover, its norm in L^1 is 1, so it does not converge to 0 in L^1. Moreover, this is true for any $p \geq 1$.

3. The third is the function

$$f_n(x) = \begin{cases} 1/n & x \leq n \\ 0 & \text{otherwise} \end{cases}.$$

   This converges uniformly with the C-norm. The p-norm is just

$$\|f_n\|_p = \begin{cases} 1 & p = 1 \\ 1/n^{p-1} & p > 1 \end{cases}.$$

Recall the theorem that $C([a,b])$ is a complete normed vector space with the norm

Theorem 6.1.2. $C([a,b])$ is a complete normed vector space with the norm

$$\|f\|_C := \sup_{x\in[a,b]} |f(x)|.$$

Proof. We went over the fact that this indeed constitutes a norm. Now we just need to show that all Cauchy sequences converge in the sense of this norm. Suppose we have that the sequence f_n is Cauchy. Then we have that $\|f_m - f_k\|_C$ is Cauchy, so

$$|f_m(x) - f_k(x)| \leq \|f_m - f_k\|_C.$$

However, since the left-hand term is just a norm on $\mathbb{R}$, we now establisehd that there is pointwise convergence between these functions. However, we need to show uniform convergence, and that the limit f is continuous (i.e., in $C([a,b])$).

By Cauchy, $\forall \varepsilon > 0$, there exists an N such that for $m, k > N$, $\|f_m - f_k\|_C < \varepsilon$. This means that

$$|f_m(x) - f_k(x)| < \varepsilon.$$

As we send $m \to \infty$, we have that it converges pointwise to some $f(x)$. Thus,

$$|f(x) - f_k(x)| < \varepsilon.$$

But this choice of N was independent of x; therefore, we have that

$$\sup |f(x) - f_k(x)| < \varepsilon.$$

Now we have to show that this f is continuous. For all $x \in [a, b]$, and for $N > N^*$, we have

$$|f_n(x) - f(x)| < \varepsilon/3.$$

FINISH THIS PROOF PLEASE $\square$

6.2 Series of Functions

We consider functions of the form

$$\sum_{n=1}^{\infty} u_n(x).$$

For this class of functions, what do we mean by uniform convergence? Simply put, we just mean that the partial sums converge uniformly.

Definition 6.2.1 (Uniform Convergence for a Series of Functions)**.** Let $u_n(x)$ be a sequence of functions. We say the infinite sum $\sum u_(x)$ converges uniformly if for every $\varepsilon > 0$, there exists an N such that for all partial sums s_k of index greater than or equal to N,

$$\left| \sum_{n=1}^{k} u_n(x) - U(x) \right| < \varepsilon.$$

Theorem 6.2.1. If each u_n is continuous and the partial sums converge uniformly, then

$$\sum_{n=1}^{\infty} u_n(x)$$

is continuous.

This is an incredibly useful theorem to have, since we would like to know, for instance, when can an integral be interchanged with an infinite sum? That is, we know that if we have uniform convergence of a sequence, the limit of the integrals is equal to the integral of the limit:

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

Analogously, we want to know if

$$\int_a^b \sum_{n=1}^{\infty} u_n(x)dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)dx$$

makes sense, if ever. The above might not always be true, since we know that we cannot intergange limits; we take one for the sum and then another for the integral.

Theorem 6.2.2 (Weierstrass M-Test)**.** Let u_n be a sequence of functions. If we can bound each u_n by some bound M_n which depends only on n, and the sum

$$\sum_{n=1}^{\infty} M_n < \infty,$$

then the series $\sum u_n(x)$ converges uniformly.

What unoform convergence affords is again a way to freely interchange limits. Since integration and differentiation are limiting processes, we have that

$$\left(\sum_{i=1}^{\infty} u_n(x) \right)' = \sum_{i=1}^{\infty} u_n'(x)$$

only if the sum $\sum u_n'(x)$ converges uniformly. The reason we're doing all of these is for its applications to power series.

6.3 Power Series

The concept of a power series should be familiar from Calculus.

Definition 6.3.1. Given a sequence a_n, a power series is a series of the form

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Our central question will be when does this sum make sense? It can work trivially when we set $x = 0$; but we want to think of this object as a function of x.

Lemma 6.3.1. If each $|a_n| \leq C$, then $f(x)$ makes sense for $|x| < 1$. And everything converges uniformly, as long as x is strictly less than 1.

Proof. If a_n is bounded, then we can think of $f(x)$ as just

$$f(x) = C \sum_{n=0}^{\infty} x^n = C \cdot \frac{1}{1-x}.$$

□

In general, we can think of this power series converging for some radius ρ in the real line.

Definition 6.3.2 (Radius of Convergence)**.** We call the (conntected) interval R the *radius of convergence* for the power series $\sum a_n x^n$ such that for every $\rho \in I$, the sum

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

converges, and does not converge for elements outside of the interval.

Lemma 6.3.2. For $0 < \rho < R$, we have that the sum

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

converges uniformly on $[-\rho, \rho]$.

Proof. For $x \leq \rho < R$, we have that $|a_n x^n| \leq |a_n|\rho^n$, the latter of which converges. Therefore, by the Weierstrass M-test, the power series converges uniformly on $[-\rho, \rho]$. □

How do we go about finding if a function has a non-trivial radius of convergence?

Theorem 6.3.3 (Tests for Radii of Convergence)**.** We have two tests we can use:

1. (Root Test) In which we see if
$$\limsup (a_n x^n)^{1/n} = \limsup (a_n)^{1/n} |x| < 1$$
As long as $\limsup(a_n)^{1/n} = 1/\rho < \infty$, then the power series is defined for $x < \rho$.

2. (Ratio Test) Where we see if
$$\limsup \left| \frac{a_{n+1}x^{n+1}}{a_n x^n} \right| = \limsup \left| \frac{a_{n+1}}{a_n} \right| x| < 1$$
where the ratio a_{n+1}/a_n is $1/\rho$.

Example. Check by the above tests that
$$f(x) = \sum_{n=0}^{\infty} n^2 x^n$$
converges for $x < 1$.

Moreover,
$$\sum n! x^n$$
only converges for $x = 0$.

It should be noted that for the radius of convergence, we can always normalize it so that the radius is 1, and we can shift it so that the radius is centered around 0, without any loss of generality.

6.4 Abel Summability

From the above discussion, when does
$$\sum_{n=1}^{\infty} x^n \tag{6.4.1}$$
converge? It converges for $|x| < 1$, so that we can write that
$$\sum_{n=1}^{\infty} x^n = \frac{1}{1-x}. \tag{6.4.2}$$

Clearly, when $x = 1$, both sides blow up to infinity. However, as we take $x \to -1$, the left hand side alternates, and the right hand side is just $1/2$! We want to see when it is useful to consider the behavior of a power series at certain limits. Consider the derivative of the above,
$$\sum_{n=1}^{\infty} n x^{n-1},$$
where we can do term-by term differentiation since the derivatives converges uniformly on $(-1, 1)$ by the ratio or root tests. Therefore, we have that
$$\sum_{n=1}^{\infty} n x^{n-1} = \left(\frac{1}{1-x} \right)' = \frac{1}{(1-x)^2}.$$

In the same vein, let's consider the integral

$$\int_0^x \sum_{n=0}^{\infty} \xi^n d\xi = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}, \ |x| < 1.$$

Since we have uniform convergence of the original powe series (EQN), then the limit of the sum and the integral can be exchanged, and we have that it is equal to

$$\sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = \int_0^x \frac{1}{1-\xi} d\xi = \log \frac{1}{1-x}$$

Here, we applied both theorems to obtain Taylor's theorem for bot the natural log and the derivative of the "vanilla" power series. Therefore, we know that if we consider

$$\sum_{n=1}^{\infty} \frac{x^n}{n}, \ |x| < 1,$$

if we plug in $x = 1$, we get the notoriously divergent harmonic series. However, if we plug in -1, we know it converges by the alternating series test. We know that

$$\log \frac{1}{1-x} = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}, \ |x| < 1.$$

We can't plug 1 into the left hand side, but we can plug -1 in to get $-\log 2$. But we don't know if the two sides are *equal*, since the equality was only for $|x| < 1$. Moreover, since we do not have uniform convergence of the right hand side, so we can't plug -1 into the sum. There is no obvious reason why the two sides should be the same when $x = -1$, even though our intuition tells us so.

The reason this substitution can be justified is from a theorem of Abel.

Theorem 6.4.1 (Abel)**.** If

$$f(x) = \sum a_n x^n$$

is defined for $|x| < 1$, and $\sum a_n = L$, then

$$\lim_{x \to 1} f(x) = \sum a_n.$$

Proof. This theorem can be proven by summation by parts. Recall the formula

$$\sum a_n b_n = A_n b_{k+1} + \sum A_n (b_n - b_{n+1}).$$

Let our b_n be x^n. Thus,

$$\sum_{n=0}^{\infty} a_n x^n = L \lim_{n \to \infty} x^n + \sum_{n=0}^{\infty} A_n (x^n - x^{n+1})$$

But since $|x| < 1$, the boundary term is just $L \times 0 = 0$. Thus

$$f(x) = \sum_{n=0}^{\infty} A_n (x^n - x^{n+1}) = (1-x) \sum_{n=0}^{\infty} A_n x^n$$

As $A_n \to L$, we have that x^n sums to $1/(1-x)$ which would cancel the $(1-x)$ term outside the sum. This way, we relate the function f to the limit L. Now we show that $f(x) \to L$ as $x \to 1$. We have that

$$f(x) - L = (1-x) \sum_{n=0}^{\infty} A_n x^n - L(1-x) \sum_{n=0}^{\infty} x^n$$

Thus

$$f(x) - L = (1-x)\sum_{n=0}^{\infty}(A_n - L)x^n, \ |x| < 1$$

since we are in the regime where everything converges uniformly. Looking at this equation, we want to show that we can let $x \to 1$. Splitting this into two pieces, we have

$$f(x) - L = (1-x)\sum_{n=0}^{K}(A_n - L)x^n + (1-x)\sum_{K+1}^{\infty}(A_n - L)x^n$$

where since the first part is a finite sum, it goes to 0 as $x \to 1$. But then if we take K to be very large, we have that $|A_n - L| < \varepsilon$. Thus,

$$f(x) - L < 0 + (1-x)\varepsilon\frac{1}{1-x} = \varepsilon.$$

$\square$

6.5 Cesàro Summability

If a sequence $a_n \to L$, then it can be proven without much difficulty that

$$s_k = \frac{1}{k}\sum_{n=1}^{k} a_n \to L.$$

If $a_n \to L$, then intuitively,

$$(1-x)\sum_{n=0}^{\infty} a_n x^n \to L$$

as $x \to 1$ (we can do the same trick of splitting it into a finite sum and showing that it is within ε of L.). Moreover, if $a_n \to L$, then we claim that

$$(1-x)^2\sum_{n=0}^{\infty} n a_n x^n \to L$$

as $x \to 1$, which is kind of like the derivative of the vanilla power series in the earlier section. These three statements are used for the fact that if something is summable, then it is Abel summable.

Recall Abel's theorem, wherein we have the liberty to write for $\sum a_n = L$,

$$f(x) = \sum a_n x^n = (1-x)\sum s_n x^n = (1-x)^2\sum \sigma_n x^n \tag{6.5.1}$$

where $\sigma_n = s_1 + \cdots + s_n$, or the sum of the partial sums. We can introduce factors of $(1-x)$ by summing the coefficients. If $\sum a_n \to L$, then we have that s_n goes to L, so this shows that the equality is true. Therefore, if something is summable, then it is Abel-summable, since we can turn the power series into that form.

Now we are ready to introduce another form of summability, known as *Cesáro Summability.*

Definition 6.5.1. We say that a sum $\sum a_i$ is Cesàro summable, if

$$\frac{s_1 + \cdots + s_n}{n} \to L.$$

Note that $\sum a_i$ may not converge.

Theorem 6.5.1. If $\sum a_i$ is Cesàro summable, then $\sum a_i$ is Abel summable. That is,

$$\text{Cesàro} \Rightarrow \text{Abel}$$

Proof. In order to prove this, we can use the third term in the equality of equation (EQN). Then we immediately have the following situation:

$$f(x) = (1-x)^2 \sum \sigma_n x^n$$

And the Cesàro condition where $\sigma_n/n \to L$. Thus,

$$f(x) = (1-x)^2 \sum \frac{\sigma_n}{n} n x^n.$$

Now we can think of $c_n = \sigma_n/n \to L$. This is the third term in the equation (EQN), so we have that if $x \to 1$,

$$(1-x)^2 \sum \frac{\sigma_n}{n} n x^n \to \lim_{n\to\infty} \frac{\sigma_n}{n} = L.$$

$\square$

Remark. It should be noted that Abel summability $\not\Rightarrow$ Cesàro. This is because as $x \to -1$,

$$\sum n x^n = 1 - 2 + 3 - 5 + 5 - \dots$$

Is Abel-summable to 1/4 since we have

$$\frac{1}{(1-(-1))^2} = \frac{1}{4}.$$

This is not Cesàro summable.

In general, we have *Tauberian Theorems* which are of the form:

$$\star\text{-summable} + \text{ condition on } a_n \Rightarrow \sum a_n \text{ converges.}$$

The theorem Tauber actually proved is the following

Theorem 6.5.2 (Tauber's Theorem)**.** Abel Summability and $na_n \to 0$ implies that $\sum a_n$ converges.

Theorem 6.5.3. If $\sum a_n$ is Cesaro (and thus Abel summable) and $|na_n| \leq M$, then $\sum a_n$ converges.

7 Fourier Series

7.1 Baby Complex Analysis and Introduction

Recall that any complex number $z = x + iy$ can be written as

$$z = Re^{i\theta} = R(\cos\theta + i\sin\theta).$$

When we discussed convergent power series, we had that

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

conberges uniformly when $|x| < 1$. The idea is then to extend what we have on the real line to functions that are complex. That is, we want to define

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n e^{in\theta} r^n$$

which is a power series in r, which is real. Now the question we have is the same question we've had for the last couple sections; what happens when $r \to 1$?

If $\sum |a_n|$ converges as $r \to 1$, then $f(z) \to \sum_{n=0}^{\infty} a_n e^{in\theta}$. Since we have absolute convergence, which, by the M-test, gives us uniform convergence. The question is really, then, what happens when $\sum |a_n|$ does not converge? From the Tauberian theorems, we know that if $\sum a_n$ is Abel summable, and $na_n \to 0$, then $f(z) \to \sum a_n e^{in\theta}$. Given a function for $|r| < 1$ we can let $r \to 1$ if we have certain conditions. Now that it's a function of θ, then this is a periodic function.

What we're essentially trying to do is given a power series, we are going to the $r = 1$ region, which is well-defined if we have Abel summability and $na_n \to 0$.

A particularly interesting question is the reverse; given a function on the boundary, can we obtain a_n? That is, given $f(z)$, does

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n r^n e^{in\theta}.$$

And the answer is yes. Recall that

$$\int_0^{2\pi} e^{in\theta} d\theta = \begin{cases} 2\pi & n = 0 \\ 0 & n \neq 0 \end{cases}.$$

Thus, if $|r| < 1$, we have uniform convergence of the complex power series since [WHY]????? . Thus, since convergence is uniform,

$$\int_0^{2\pi} f(z) d\theta = \sum_{n=0}^{\infty} a_n r^n \int_0^{2\pi} e^{in\theta} d\theta = 2\pi a_0.$$

If we wish to isolate a_1, we just divide f by z, so that

$$\int_0^{2\pi} \frac{f(z)}{z} = \int_0^{2\pi} \frac{a_0}{z} + a_1 + a_2 z + \cdots = \int_0^{2\pi} \frac{a_0}{r} e^{-in\theta} + a_1 + \cdots = 2\pi a_1.$$

In general, we just divide $f(z)$ by z^n, or, taking $r = 1$,

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} d\theta.$$

In order to compute the a_n's, we only need the values of f on the unit circle, which are defined as a function of θ.

Given c_n such that $\sum |c_n| < \infty$, then

$$g(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}$$

is a 2π-periodic function of θ. Moreover, g is continuous.

Proof. We have that

$$g(\theta) - g(\theta_0) = \sum_{n=-\infty}^{\infty} c_n \left(e^{in\theta} - e^{in\theta_0}\right).$$

since the c_n are absolutely convergent. Thus,

$$|g(\theta) - g(\theta_0)| \leq |\sum_{n=-K}^{K} c_n(e^{in\theta} - e^{in\theta_0})| + |\sum_{n=K+1}^{\infty} c_n(1+1)| + |\sum_{n=K+1}^{\infty} c_n(1+1)|$$

since $|e^{i\phi}| = 1$ so we can bound it by 2. Therefore, for K large enough, we can make it such that the last two terms are less than $\varepsilon/2$, and thus, since the first term is a finite sum, we habe arbitrary closeness since the first term is continuous. □

Now the question of Fourier series boils down to two questions;

1. Given a 2π-periodic function, can I write it as

$$g(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{in\theta}?$$

   Does it converge? Does it Abel-converge, Cesàro converge?

2. What is the relationship between $g(\theta)$ and the sum?

7.2 Fourier Series

Recall that given a sequence $\{a_n\}$, we discussed when the series

$$\sum_{n=0}^{\infty} a_n x^n$$

converges; if it converges in the sense of Abel, Cesàro, etc. Recall that we normalized things so that this was convergent when $|x| < 1$. That is, so far we have

1. Uniform Convergence (**)
2. Abel Summability
3. Cesaro Summability

Next, we changed the x to be something like $z = re^{i\theta}$. Thus, we have

$$\sum a_n z^n = \sum a_n r^n e^{in\theta}$$

And, normalizing everything again, we have convergence when $|r| < 1$. We want to investigate the behavior as $r \to 1$. Note that we can consider $\overline{z}$, the complex conjugate. Recall the theorem,

Theorem 7.2.1. If $na_n \to 0$ as $n \to \pm\infty$, then

$$\text{Cesàro (or Abel) summable } \Rightarrow \sum_{n=-\infty}^{\infty} a_n r^n e^{in\theta} \to\Rightarrow \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$$

Example. Let

$$a_n = \frac{1}{(|n| + 1\log(|n|+1)}.$$

We have that $|na_n| \to 0$ as $n \to \pm\infty$. But this sum is not summable. If we do the Cauchy condensation test, and see that this doesn't converge. But when things are monotone, we can use the integral test:

$$\int_*^\infty \frac{1}{x\log x} dx$$

where the "*" was included to indicate that it doesn't matter where we start the integral. This gives us

$$\log\log(x)\Big|_*^\infty = \infty$$

This is a sequence which is not summable, but which does satisfy Abel's theorem; the sequence $na_n \to 0$, and thus we have that

$$f(\theta) = \sum_{n=-\infty}^{\infty} \frac{1}{(|n| + 1\log(|n|+1)} e^{in\theta}$$

which is the limit of the complex power series as $r \to 1$. This converges for every θ, but not uniformly. Now we want to ask the question; is $f(\theta)$ differentiable? In this case the answer is no; we can check this by differentiating, to get

$$f'(\theta) \stackrel{?}{=} \sum_{n=-\infty}^{\infty} \frac{n}{(|n| + 1\log(|n|+1)} e^{in\theta}$$

but the right hand side doesn't converge, not even in the Abel sense.

The general question we need to answer is: given a 2π-periodic function f, can we write it as

$$f(\theta) = \sum_{n=-\infty}^{\infty} c_n e^{-n\theta}?$$

Moreover, how do we relate the c_n to the function? We will use the following fact:

$$\int_0^{2\pi} e^{in\theta} e^{-ik\theta} d\theta = \begin{cases} 2\pi & n = k \\ 0 & n \neq k \end{cases} = 2\pi \cdot \delta_{nk}.$$

For the sake of simplicity, we assume all f's are 2π-periodic, then we have 2 big if's: if f is indeed equal to the summand, and if we can freely exchange the sum and the integral (i.e., uniform convergence), then we have that

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) e^{-ik\theta} d\theta$$

since we isolate the term when $n = k$. The obvious things we need for this to work are that f is Riemann-integrable. But if f is integrable, we don't have control over the size of the c_k. If we had $|kc_k| \to 0$, then that'd be great; if this is satisfied, then we have that the infinite sum exists in the sense of Abel.

Now we want to investigate under which conditions the c_k go to 0, and if they do that fast enough to be summed. Let's shift our focus to integrals of the form

$$\int_0^{2\pi} f(x)e^{i\lambda x}dx.$$

What happens when $\lambda \to \infty$? If we decompose our exponential into sines and cosines, we see that we get periodic behavior on an period of $2\pi/\lambda$. Thus as λ gets larger, these periods shrink. Thus, if we look at f over one period, it does not change very much if f is continuous since our interval is very small (think of $2\pi/\lambda = \delta$.) Then if f is essentially constant over that period, we can consider the integral over one period:

$$\int_{2\pi j/\lambda}^{2\pi(j+1)/\lambda} f(x)e^{i\lambda x} \approx f(x)\int_{2\pi j/\lambda}^{2\pi(j+1)/\lambda} e^{i\lambda x} = 0$$

as long as we can add up these periods, and the error doesn't accumulate, we can expect the entire integral to be 0.

We have other methods of introducing a λ into the denominator to manage the limit as $\lambda \to \infty$. For instance, we can integrate by parts. Thus

$$\int_0^{2\pi} f(x)\sin\lambda x dx = \frac{-f(2\pi)\cos(2\pi\lambda)}{\lambda} + \frac{f(0)}{\lambda} + \int_0^{2\pi} f'(x)\frac{cos\lambda}{\lambda}dx$$

where we consider only one part of the trigonometric form of our complex exponential. We can do the same for cos. The boundary terms go to 0, and if f is differentiable, and f' is differentiable, then we can bound the integrand by $\sup f$ and that also goes to 0. Therefore, we can arrive at the conclusion that If f' is Riemann integrable, then

$$\left|\int_0^{2\pi} f(x)e^{i\lambda x}\right| \leq \frac{C}{\lambda}.$$

Of course, there's nothing special about 2π; we can extend this to be true for any closed and bounded interval. Thus we have reduced the problem to

$$|nc_n| \leq C$$

which is a constant. It was the work of a mathematician named Littlewood that proved if we have $|nc_n|$ is bounded, and we have Cesàro summability, then we have summability. Let's codify our work into a lemma:

Lemma 7.2.2 (Riemann-Lebesgue)**.** If f is integrable, then

$$\int_0^{2\pi} f(x)e^{i\lambda x}dx \to 0 \text{ as} \lambda \to \infty.$$

Notice how this is far stronger than what we proved. We only considered functions whose derivatives exist and are integrable; The Riemann-Lebesgue lemma is saying for integrable functions, their integral goes to 0 as $\lambda \to \infty$. If we also assume that there is an integrable derivative, we can get a boundedness condition on nc_n. We are on the cusp of providing a strong sufficient condition on when we can write a function f as a sum of c_n's.

Why is this true for integrable functions? If you remember how we defined the integral as the division of intervals, we can turn to the first of our calculations in this chapter to see how this condition is satisfied. We will provide a sketch of this proof:

Proof. If f is integrable, then $\exists$ a step function S_ε such that

$$\int_a^b |f(x) - S_\varepsilon(x)| dx < \varepsilon$$

which is true because the step function is just the upper or lower integral. Therefore,

$$\int_a^b |f(x) - S_\varepsilon(x)| e^{i\lambda x} dx \leq \int_a^b |f(x) - S_\varepsilon(x)| dx < \varepsilon$$

which is true since the magnitude of $|e^{i\lambda x}| = 1$. [FINISH UP] □

This lemma is still not sufficient. We want to get a theorem that gives us the rate at which c_n goes to zero, so we can see if it's summable or not. If we integrate by parts a second time, we can get a λ^2 in the denominator.

Lemma 7.2.3. If $f \in C^2$ and periodic, then

$$\left| \int_0^{2\pi} f(x) e^{inx} dx \right| \leq \frac{C}{n^2}.$$

Technically, we can show the same result if f'' is merely integrable, not necessarily continuous.

Proof. We integrate by parts twice, as in our above calculations. □

This proves that if we have a function $f \in C^2$, then each of our $|c_n| \leq C/n^2$. Thus, by the Weierstrass M-test, then

$$\sum_{n=-\infty}^{\infty} c_n e^{in\theta}$$

converges uniformly; but to what? We still have to show that this is truly equal to $f(\theta)$.

7.3 Convergence Theorems

Recall that we motivated Fourier series when we wanted to go to Cesàro or Abel summability to summability. To each $f(\theta)$ we associate a fourier series

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx$$

And we made some observations:

1. If f is integrable, then $c_n \to 0$.
2. If $f \in C^1$, then $|c_n| \leq C/n$.
3. If $f \in C^2$, then $|c_n| \leq C/n^2$.

The last condition is useful, since now we have that the sum of c_n's makes perfectly good sense since we can sum it. Moreover, if $f \in C^2$, then the sum is also continuous, since we have uniform convergence and each $c_n e^{in\theta}$ is continuous. But we still haven't considered whether the sum is actually equal to f. We really have two functions:

$$f \in C^2,\ g(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}.$$

We can do some substitutions for c_n:

$$g(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} = \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi} f(\xi) e^{-in\xi} d\xi \right) e^{inx}.$$

But since we have uniform convergence (since f is C^2), then we can interchange the infinite sum and the integral. Thus

$$= \frac{1}{2\pi} f(\xi) \left(\sum_{n=-\infty}^{\infty} e^{inx-\xi} \right) d\xi.$$

Except there's a problem; the magnitude of the $e^{in(x-\xi)}$ is just 1. Therefore, we really ought to write the infinite sum as the limit of partial sums. Then if we think of $a = e^{i(x-\xi)}$. Using trigonometric identities, we can find that

$$\sum_{n=-K}^{K} a^n = \frac{\sin[(K+1/2)x]}{\sin(x/2)}$$

Therefore

$$g(x) = \lim_{k\to\infty} \frac{1}{2\pi} \int_0^{2\pi} f(\xi) \cdot \frac{\sin[(K+1/2)(x-\xi)]}{\sin((x-\xi)/2)} d\xi$$

Now let $x - \xi = z$. This integral becomes

$$g(x) = \lim_{k\to\infty} \frac{1}{2\pi} \int_0^{2\pi} f(x-z) \cdot \frac{\sin[(K+1/2)z]}{\sin(z/2)} dz$$

Where the bounds of integration are the same, since if f is periodic we can integrate over any interval of length 2π and still get the same quantity.

Now we want to see how close g gets to f.

$$g(x) - f(x) = \lim_{k\to\infty} \frac{1}{2\pi} \int_0^{2\pi} [f(x-z) \cdot \frac{\sin[(K+1/2)z]}{\sin(z/2)} - f(x)] dz$$

Since we don't want to keep writing the trigonometric expansion over and over again, we can assign a name to it:

$$D_K(z) := \sum_{n=-K}^{K} e^{inz} = \frac{\sin[(K+1/2)z]}{\sin(z/2)}.$$

It should be verified that

$$\frac{1}{2\pi} \int_0^{2\pi} D_K(z) dz = 1$$

which can be obtained from the sums. Therefore, the difference is

$$g(x) - f(x) = \lim_{K\to\infty} \frac{1}{2\pi} \int_0^{2\pi} [f(x-z) - f(x)] D_K(z) dz.$$

which we hope to be 0. This is very suggestive of the Riemann-Lebesgue lemma. The only term that has K in it is the sin in the numerator of $D_K(z)$. We can re-write the difference as

$$g(x) - f(x) = \lim_{K\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(x-z) - f(x)}{\sin(z/2)} \sin(z(K+1)/2)dz.$$

Again, we can freely change the bounds of integration since f is periodic. Notice that we're only in trouble when $z = 0$, since the denominator is 0. But then again, so is the numerator. We should split the integral into the following three integrals: one from $[-\pi, -\delta]$, then from $[\delta, \pi]$, and lastly one on the small interval of $[\delta, \delta]$. We only have to consider the last of these, since this is the only integral for which z assumes the value of 0. If z is small, then $\sin(z/2) \approx z/2$. The numerator also tells us that if the numerator is differentiable, then

$$f(x-z) - f(x) \approx f(x) - f'(x)z.$$

Thus we can cancel the z's in the numerator and the denominator. This takes care of the problem, since it's now a removable discontinuity. One we've done that problem, we can integrate by parts to obtain a K in the denominator by integrating by parts twice. Then all the integrals go to 0; and hence the difference between f and g goes to 0. This proves

Theorem 7.3.1. If $f \in C^2$, then

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

and the convergence is uniform.

Notice how we need the *hypotheis* that $f \in C^2$, even though we only integrated by parts once. This is becausae we also used taylor's theorem, and we need a second order derivative in order to quantify the remainder exactly.

What if the function is just C^1, or simply only continuous? If the function is just continuous, then we don't have uniform convergence, nor do we have convergence of the fourier series. However, we can state another theorem by expanding our definition of convergence:

Theorem 7.3.2. If $f \in C$, and f is 2π-periodic, then

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

converges in the sence of Cesàro.

And now for a quick diversion concerning non-periodic functions. If f is non-periodic, we may still be able to do analysis on it by treating it as if it were periodic. That is, if we wish to consider an interval $[a, b]$ of a function f, then we can chop it into repeating parts of length $b - a$. In general, there may be jumps when we proceed to the next period, so it may be that the fourier series does not converge at the point b.

So far we've talked about functions f which are 2π periodic. For each integrable function f, we associate it with a fourier series:

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where the symbol $\sim$ indicates that we do not have equality; for all we know, the right side may not even converge. All we know is that each of the c_n can be computed since f is integrable.

We proved in the last section that the sum on the right converges uniformly to f, so we have equality. This is because each integral can be bounded by a constant over n^2, which, by the Weierstrass M-test, we know implies uniform convergence. We established equality if $f \in C^2$, using the function $D_K(x)$, known as the Dirichlet kernel.

We will consider questions about the convergence of the partial sums of Fourier series. We can write the partial sums as

$$S_K(f)(x) = \int_0^\pi (f(x-t) + f(x+t))D_K(t)dt$$

which we can do by splitting the integral in half and changing the variable of one of them from t to $-t$. Now consider

$$S_K(f)(x) - f(x).$$

There are many senses in which the partial sums can converge to f (or this difference goes to 0).

1. Pointwise
2. Uniformly
3. Cesàro
4. In the sense of an integral.

It may be surprising that some functions, the partial sums of the Fourier series can converge in some of them, but not the others. For instance, it is possible to find a continuous function for which the difference does not go to 0 pointwise, for any point, but it does in the square-integral sense.

Theorem 7.3.3. If $f \in C^1$, then the Fourier series converges pointwise to f.

Proof. For this proof, we only really need that

$$\left| \frac{f(x-t) - f(x)}{t} \right| < C$$

which can be obtained by expanding the Dirichlet kernel and taylor expanding the denominator. Or, setting $t = x - y$ and rearranging the terms,

$$|f(x) - f(y)| < C|x-y|$$

which is known as the Lipschitz condition. If this is met, then we have that for every point, we have that the difference between the partial sums and f goes to 0 by integrating by parts. However, we don't have the uniformity condition. □

What if f is discontinuous? Fourier series were motivated by engineering, which often has discontinuities. If f is discontinuous, at every point of continuity we don't have a problem, but the real question is what happens at discontinuous points.

Theorem 7.3.4. If f has a jump at x_0, then

$$S_K(f)(x_0) \to \frac{f(x_0^+) + f(x_0^-)}{2}.$$

Proof. Let's make a simplification. We're trying to look at

$$\frac{1}{2\pi} \int_0^\pi [f(x+t) - f(x-t)] \frac{\sin(t(N+1)/2}{\sin(t/2)}$$

The simplification we make is that we only care about $t = 0$, so the denominator can be t treated like $t/2$ and we can let $\lambda = (N+1)/2$ which can be sent to infinity. Thus we can understand the problem like

$$\int_0^A h(t) \frac{\sin(\lambda t)}{t}$$

where $0 < A < 2\pi$. If $h(0) \neq 0$, then we have a problem. Changing the integral,

$$\int_0^A (h(t) - h(0^+) + h(0^+))\frac{\sin(\lambda t)}{t}.$$

This is because t approaches 0 from above. We have $h(t) - h(0^+)$ goes to 0 as $t \to 0$. Thus,

$$\int_0^A \frac{h(t) - h(0^+)}{t} \sin \lambda t + \int_0^A h(0^+)\frac{\sin \lambda t}{t}$$

The first integral goes to 0 by integrating by parts. Focusing our attention to the second integral, we can execute a change of variables given by

$$h(0^+) \int_0^{\lambda A} \frac{\sin x}{x}$$

But this is just our familiar improper integral; which is just $\pi/2$. Thus, for t close to 0,

$$\frac{1}{2\pi} \int_{-\delta}^{\delta} (f(x+t) + f(x-t))2 \cdot \frac{\pi}{2}$$

From which we get the desired result. □

There is still one more theorem we will consider before we move to the L^2 theory, or square-integral convergence.

Theorem 7.3.5. If f is integrable, then the partial sums $S_N(f)(t) \to f$ in the sense of Cesàro.

7.4 Convolution and Kernels

Recall that the partial sums of the Fourier series is given by the Dirichlet kernel:

$$S_n(f)(x) = \int_{-\pi}^{\pi} f(x-t)D_n(t)dt.$$

For all n, we have that the integral of the Dirichlet kernel is 2π; but it can be shown that the integral of the absolute value of $|D_n(x)|$ goes to infinity as $n \to \infty$.

If f is an integrable function, we express the convolution of f with an integrable function K to be

$$(f * K)(x) = \int f(x-t)K(t)dt.$$

This satisfies certain properties, such as commutativity, associativity, and linearity. This should look familiar; we were able to express the partial sums as a convolution with the Dirichlet kernel, $D_n(x)$. Near 0, this function peaks; we can use Taylor's theorem to show that it reaches a height of about n. Everywhere else, it oscillates around 0. The problem is, if we take the absolute value, the oscillations are strictly positive and so they accumulate, making the integral infinite. The heights of each of them decay like n/x.

For the sake of argument, we consider K to be a function whose integral is 1, and it is C^∞. The first observation is that

$$|f * K(x)| \leq |\sup_x f|.$$

We can use a change of variables to get

$$K * f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)K(x-t)dt$$

But since the convolution is uniformly bounded, we have that the convolution $f * K \in C^\infty$. Thus

$$(K * f(x)) = \int_{-\pi}^{\pi} f(t)K'(x-t)dt$$

whereby we have taken a function that was just integrable and have made it into a C^∞ function. Let $K_n(t) = K(nt)$. Given this, K_n is still infinitely differentiable. From changes of variables, we know that the integral of this is still 1. For a function that has a single bump (e.g., a Gaussian), K_n is a function that with a taller but narrower bump, of the scale n and $1/n$, respectively. Outside of this bump area, the functoin can be made arbitrarily small.

Now let's assume that f is continuous on a compact interval. This means that f is uniformly continuous. Let's look at the quantity

$$|K_n * f(x) - f(x)| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x-t) - f(x)|K_n(t)dt.$$

All the action of $K_n(t)$ is concentrated around 0. THis means we can Then we can split these integrals:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x-t)-f(x)|K_n(t)dt \leq \frac{1}{2\pi}\int_{-\delta}^{\delta} |f(x-t)-f(x)|K_n(t)dt + \int_{\delta\leq|t|\leq\pi} |f(x-t)-f(x)|K_n(t)dt$$

roughly speaking. Since f is continuous, and t is very small for the first integral, then the firsst integral is bounded by ε since the integral of K_n is bounded by 1. Moreover, we can bound

$$|f(x-t) - f(x)| < 2\sup_t f.$$

But away from the δ-interval, $K_n(t)$ is arbitrarily small. Therefore, we can bound

$$|K_n * f(x) - f(x)| < \varepsilon.$$

This is a very powerful statement, since we can take any continuous function, and by convolving it, we can get another function that is ε away from it, but C^∞.

Now we summarize the properties of what we call this special class of functions, known as "good kernels:"

Definition 7.4.1 (Good Kernels)**.** A family of functions $\{K_n(x)\}_{n=1}^{\infty}$ are said to be "good kernels" if they satisfy the following properties:

1. (Normalized.) For all $n \geq 1$,
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(x)dx = 1$$

2. (Absolutely Integrable.) There exists $M > 0$ such that for all $n \geq 1$,
$$\int_{-\pi}^{\pi} |K_n(x)dx \leq M$$

3. (Mass concentration.) For every $\delta > 0$,
$$\int_{\delta\leq|x|\leq\pi} |K_n(x)|dx \to 0, \quad \text{as } n \to \infty.$$

That is, the tail ends of the function are arbitrarily small.

What's more, a class of good kernels satisfy the "approximate identity" property.

Theorem 7.4.1. Let f be an integrable function on the unit circle. Then

$$\lim_{n\to\infty}(f * K_n)(x) = f(x)$$

whenever f is continuous at x. If f is continuous everywhere, then this convergence is uniform.

Proof. Since f is continuous at x, choose δ such that

$$|f(x-t) - f(x)| < \varepsilon/2.$$

By the first property of good kernels, then

$$(f * K_n)(x) - f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x-t) - f(x)]K_n(t)dt$$

We can split up this integral into two:

$$|(f * K_n)(x) - f(x)| \leq \frac{1}{2\pi}\int_{-\delta}^{\delta}|f(x-t) - f(x)||K_n(t)dt| + \int_{\delta\leq|t|\leq\pi}|f(x-t) - f(x)||K_n(t)|dt$$

We have that the quantity $|f(x-t) - f(x)| < \varepsilon/2$, and thus we can remove it from the integral. Moreover, from the third property of good kernels, the integral of $|K_n(t)|$ goes to 0; and since f is integrable, it is bounded by $B := \sup_x f$.Thus, we can select an n such that the integral of K_n is less than $\pi\varepsilon/2B$.Thus, we have that

$$\varepsilon/2\frac{1}{2\pi}\int_{-\pi}^{\pi}|K_n(t)|dt + 2B\frac{1}{2\pi}\int_{\delta\leq|t|\leq\pi}|K_n(t)|dt < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

If f is continuous on the entire interval, it is unifomrly continuous. Since we can select a δ^* independent of x, this convergence will be uniform. □

This has a very powerful application to Fourier series. The Dirichlet kernel satisfies one of the properties, since it is infinitely differentiable. However, it does not converge as $n \to \infty$ in absolute value. Let us define a new function, which is like the Dirichlet kernel:

Definition 7.4.2 (Fejér Kernel)**.** We define the Fejér Kernel to be:

$$F_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} D_i(x) = \frac{1}{2\pi n}\frac{\sin^2 nx}{\sin^2 x/2}$$

which is just the Cesàro sum of the Dirichlet kernels.

There is an immediate contrast between F_n and D_n; namely, F_n is positive, but it still has an integral of 1.

Now we are in position to prove a powerful theorem about continuous functions:

Theorem 7.4.2. If $f \in C[-\pi, \pi]$, then the Fourier series of f is Cesàro summable to f. Moreover, the convergence of the Cesàro sums is uniform.

Proof. We sketched out the proof in the above calculations. Split up the integral into two parts, and □

What we have shown is that given a continuous function, we can show that its fourier series is Cesaro summable to f. Both D_n and F_n are both C^∞ since they are the sums of both sines and cosines. If we use Taylor's theorem, then we can approximate F_n by a polynomial in x. Therefore, we have the following corollary:

Theorem 7.4.3 (Weierstrass Approximation Theorem)**.** If f is continuous on an interval, then we can construct a polynomial p_ε such that

$$|f(x) - p_\varepsilon(x)| < \varepsilon.$$

for all $x \in [a, b]$.

Now if we have any integrable function, we used step functions to approximate it in defining the integral. This step function can be further approximated by continuous functions, which can further be approximated by polynomials.