Analysis

Luke Baker

May 2022

Contents

Chapter 1

Real Numbers

This section is currently a major work in progress and excluded for now!

Chapter 2

An Introduction to Topology

The goal of this chapter is to study abstract Metric and Topological Spaces and prove some very useful and important theorems regarding these objects and their uses in Analysis. In most Analysis texts General Topological Spaces usually aren't studied, but often focusing too much on the metric rather than the topological properties in a Metric Space can over complicate things for the student and further there are many situations where a metric does not apply very well, we will motivate Topology through exploring a more basic example of these spaces. This chapter acts somewhat like a boomerang as we start with metric spaces, move to more abstract topological spaces, and then gradually work our way back to metric spaces having explored many of the important topological properties in generality.

2.1 More on Metric Spaces

Having finished chapter one we recall that a Metric Space is simply a pair of a set and a distance function (X, d). So far we have given examples of Metric Spaces, but we have not studied them or given any explanation for why they are useful. This section will begin to rectify this issue. To begin we will present many new definitions, it is important the reader try their best to commit these definitions to memory and try to gain some intuitive understanding of the concepts.

Definition 2.1. Given a metric space X and a point $p \in X$ we define **The Open Ball** of Radius r around p as the collection of all points x with $d(x, p) < r$, denoted as $B_r(p) := \{x \in X | d(p, x) < r\}$. Similarly, the Closed Ball of Radius r is defined $\overline{B}_r(p) := \{x \in X | d(p, x) \leq r\}$.

In the above definitions the open ball does not contain points at a distance of radius r, while the closed ball does contain these points. The open ball is very useful for measuring "closeness" to a point. For examples in $\mathbb{R}$ an open ball is just an open interval we have previously discussed! In $\mathbb{R}^2$ an open ball is a disc of radius r, and in $\mathbb{R}^3$ it is an open sphere. Similarly, closed balls given closed intervals, as well as discs and spheres this time with the boundary points included. Having introduced the idea of an open ball we can now present the rest of the important definitions.

Definition 2.2. For the following definitions assume X is a Metric Space, $p \in X$, and $E \subseteq X$.

(a) We say that p is an **interior point** of E if there exists an open ball $B_r(p)$ such that $B_r(p) \subseteq E$. Intuitively an interior point of a set is a point that is fully surrounded by points of a set.

(b) We say that $E \subseteq X$ is **open** (relative to X) if every point of E is an interior point of E, in open sets every point is fully surrounded by other points of the set.

(c) We say that p is a **limit point** of E if *every* open ball of p contains a point of E distinct from the point p itself. Intuitively, limit points of a set are the points that are "very close" or "touching" the set. Notice the definition does not require that $p \in E$.

(d) Finally, we say that $E \subseteq X$ is **closed** (relative to X) if every limit point of E is contained in E. Intuitively, closed sets contain all points that are "close" to the set.

Some examples of open and closed sets can help with intuition. We first want to show that Open Balls are open as their name suggests.

Theorem 2.3. Let X be a Metric Space, $p \in X$, and $r > 0$. Then $B_r(p)$ is open.

Proof. We need to show that every point of $B_r(p)$ is an interior point. So first we choose an arbitrary point $q \in B_r(p)$. Now we need to show that there exists some open ball $B_{r_2}(q) \subseteq B_r(p)$. Often times drawing a picture can aid in the proof and give a visual understanding of the situation (refer to figure). Let $m = d(p, q)$. Then since $q \in B_r(p)$ by definition we have that $0 \leq m < r$ and so $r - m > 0$. Let $r_2 = r - m$. Now let x be an arbitrary point in $B_{r_2}(q)$, it follows by definition that $d(q, x) < r_2$ and now using the triangle inequality we get that $d(p, x) \leq d(p, q) + d(q, x) < m + r - m = r$ and so we see that $x \in B_r(p)$. But since x was an arbitrary point of $B_{r_2}(q)$ we see that $B_{r_2}(q) \subseteq B_r(p)$ so that q is interior to $B_r(p)$. Finally, because q was arbitrary we see that every point of $B_r(p)$ is an interior point and so the set is open! $\square$

This gives us a great deal of open sets to work with in the metric spaces we are familiar with! For example open intervals (a, b) are open as a subset of $\mathbb{R}$, again showing that the names are consistent!

Given a metric space (X, d) the set X is always an open subset of itself, this follows because for any point $p \in X$ and any distance r we must have that $B_r(p) \subseteq X$ as there are no other points in the space to work with! The empty set is also always an open subset of X as vacuously every point in the empty set is interior! Finally, for our last group of open sets the reader should verify that for any distance r and point $p \in X$ the set $\{x \in X | d(p, x) > r\}$ is open.

We now prove an important theorem relating open and closed sets.

Theorem 2.4. In a metric space (X, d) a set G is closed if and only if G^c is open.

Proof. Firstly, assume G is closed and let $x \in G^c$. Then x is not a limit point of G, since G contains all its limit points, so that there is some radius r such that $B_r(x)$ contains no points of G (if there was no radius, then x would be a limit point). But since $B_r(x)$ contains no point of G, then $B_r(x) \subseteq G^c$ so that x is interior to G^c, and since x was arbitrary we see that G^c is open! On the other hand if we assume G^c is open and we let p be a limit point of G, we see it cannot possibly be in G^c as that would imply there was an open ball surrounding it entirely in G^c and so it would not be a limit point of G. □

This proof is also very useful for discovering many closed sets. For example since the reader verified earlier that $\{x \in X | d(x,p) > r\}$ is open, we have that $\{x \in X | d(x,p) \leq r\}$ is closed in any metric space! This implies the closed interval $[a,b]$ is closed, as are closed discs any many other sets! One strange thing readers may notice is that given a metric space (X,d) since both $\emptyset$ and X were open this implies that we also have both $\emptyset$ and X are closed! We see that sets can be closed, open, or both! We can also see that sets can be neither for consider the set $(0,1]$! It does not contain all its limit points, so it cannot be closed, but it is certainly not open for no open ball of 1 will be in the set!

We end this section with one last theorem and corollary concerning open and closed sets.

Theorem 2.5. Given an arbitrary sized collection of open sets $\{V_a\}_{a \in A}$ we have that the union of all V_a is also open, and if the collection of open sets is finite $V_1, ..., V_n$, then the intersection of all V_i is also open.

Proof. Let $\{V_a\}$ be the arbitrary collection of open sets as above and let

$$p \in \bigcup_{a \in A} V_a.$$

Since p is in the union there must be some index $\alpha \in A$ so that $p \in V_\alpha$, but since V_α is open it follows that p is interior to V_α so that there is some open ball $B_r(p) \subseteq V_\alpha$, but then by definition we have

$$B_r(p) \subseteq \bigcup V_a$$

so that p is interior to the union of all $\{V_a\}$ and since p was arbitrary we see the union is open.

For the case of finite intersections suppose we have a finite collection of open sets $V_1, ..., V_n$ as above, and suppose

$$p \in \bigcap_{i=1}^{n} V_i.$$

Then by definition $p \in V_i$ for all $i = 1, ..., n$ and since each V_i is open it follows that p is interior to each V_i, so that for each i there exists an open ball of radius r_i so that $B_{r_i}(p) \subseteq V_i$. Since the set of all radii $\{r_1, ..., r_n\}$ is finite it has a minimum so set $r' = \min\{r_1, ..., r_n\}$. Then it follows that for all $i = 1, ..., n$ we have that $B_{r'}(p) \subseteq B_{r_i}(p) \subseteq V_i$ and so we see that

$$B_{r'}(p) \subseteq \bigcap V_i$$

so that p is interior to the intersection and so the intersection is open. □

Corollary 2.5.1. From DeMorgan's Laws and the fact that a set is closed if and only if its compliment is open we have that arbitrary intersections of closed sets are closed and finite unions of closed sets are closed.

The last theorem of this section will deal with metric spaces contained inside other metric spaces. Given a metric space (X, d) and a subset $Y \subseteq X$ it makes sense to consider (Y, d) as its own metric space. This can be seen with $\mathbb{R}^2$ and $\mathbb{R}$ for example. We mentioned earlier that every metric space is an open subset of itself so that on its own $\mathbb{R}$ is open, but as a subset of $\mathbb{R}^2$ it is easy to show that $\mathbb{R}$ is NOT open. The following theorem helps resolve this conflict and will guide an important definition in the next section!

Theorem 2.6. Given a metric space (X, d) and a sub-metric space (Y, d) we have that a subset $E \subseteq Y$ is an open subset of Y if and only if there is an open subset $V \subseteq X$ so that $E = V \cap Y$.

Proof. For the first direction assume that E is an open subset of Y. Since E is open we know that for every point p of E there exists a corresponding radius r_p so that $B_{r_p}(p) \subseteq Y$.

Let $B'_{r_p}(p)$ be the same open balls but now regarded as subsets of X. Since each of these $B'_{r_p}(p)$ are open, by the previous theorem it follows that if we set

$$V = \bigcup_{p \in E} B'_{r_p}(p),$$

then V is an open subset of X and we see that $E = V \cap Y$ completing the first direction. Now assume that $E = V \cap Y$ where V is an open subset of X and let $p \in E$. Since V is open there is an open ball $B_r(p) \subseteq V$ and so we get that $B_r(p) \cap Y \subseteq V \cap Y = E$ so that p is interior to E. $\square$

2.2 Topological Spaces

So far we have provided a host of definitions regarding metric spaces and proved various facts regarding these definitions. We also in Chapter One discussed a couple different examples of metric spaces and showed that most of the spaces we constructed in Chapter One were indeed examples of metric spaces. However, one space we constructed was missing from the list of metric spaces: The Extended Real Numbers. Indeed the Extended Real Numbers are *metrizable*, meaning they can be given a metric, but we should think about this for a moment. Imagine we gave $\overline{\mathbb{R}}$ a metric. What would the following expression equal: $d(0, \infty)$? Recall ∞ is not a real number and a distance MUST be a real number, so to assign any distance between 0 and ∞ would not truly capture how we would expect the distance between these two points to be. Thinking about this for a while should give the reader the idea that no metric space truly fits how we expect $\overline{\mathbb{R}}$ to behave and so we must move to a more general type of space. This more general type of space is called a **Topological Space**! We will now introduce the notion of a topological space and show that topological spaces generalize the properties we have just proven about metric spaces, and we will give a few examples of these new spaces!

Definition 2.7. Given a set X and a collection of subsets of X we will label τ the pair (X, τ) is called a **Topological Space** if it satisfies the following three axioms:

1. $X \in \tau$ and $\emptyset \in \tau$.
2. The union of an arbitrary collection of members of τ is again a member of τ.
3. The intersection of a finite collection of members of τ is a member of τ.

The members of τ are called **Open Sets** or **Open Subsets of** X and τ is called a **Topology**.

In light of the Theorem's in the last chapter we see that Metric Spaces are a specific type of Topological Space as they obey all of the above axioms! We now must reintroduce some of the definitions from Metric Spaces into these new and more abstract spaces!

Definition 2.8. Given a Topological Space (X, τ) and a point $p \in X$ we define a **Neighborhood of** p to be an arbitrary open set containing p and we will often denote a Neighborhood of p with $N(p)$. The primary reason for neighborhoods is that for metric spaces we could often find a "nice" open set containing a point, being an open ball around the point, but without a metric we can only talk about arbitrary neighborhoods of a point. With the neighborhood definition we can now give a topological definition of an interior point.

Given a subset $E \subseteq X$ we say that a point $p \in E$ is an **Interior Point** of E if there is a neighborhood of p, $N(p)$ such that $N(p) \subseteq E$.

We can also redefine limit points as folllows: Given a subset $E \subseteq X$ we say that p is a **Limit Point** of E if every neighborhood of p contains a point of E distinct from the point p itself.

Finally we say a subset $C \subseteq X$ is **Closed** if it contains all its limit points. This definition is the same as in a Metric Space, but we are using the topological definition of a limit point.

It is not too difficult to show that if our Topological space (X, τ) is also a Metric Space then all of the above definitions coincide so that in a Metric Space we can continue to use our old definitions. Further we can show that Theorem 2.4 also applies to Topological Spaces!

That is in these new spaces a set is closed if and only if its compliment is open. This follows essentially be repeating Theorem 2.4 but replacing open balls with neighborhoods.

Definition 2.9. Given a Topological space (X, τ) and a subset E let E' be the set of all limit points of E and let E° be the set of all interior points of E. We call E° **The Interior of** E and we call $\overline{E} = E \cup E'$ **The Closure of** E. We have the following facts.

Theorem 2.10. If E is a subset of X, then

1. The Interior of E is open.
2. If O is an open set with $O \subseteq E$, then $O \subseteq E^\circ$.
3. $E = E^\circ$ if and only if E is open.
4. The Closure of E is closed.
5. If F is a closed set with $E \subseteq F$, then $\overline{E} \subseteq F$.
6. $E = \overline{E}$ if and only if E is closed.

Proof. To begin, for (1) we see that if $p \in E^\circ$, then there is a neighborhood $N(p) \subseteq E$. Let

$$V = \bigcup_{p \in E^\circ} N(p),$$

then V is open as each $N(p)$ is open and $E^\circ \subseteq V$. But now if $q \in V$, then it is contained in some $N(p)$ so that $q \in N(p) \subseteq E$ showing q is also an interior point of E so $V \subseteq E^\circ$. We now see $E^\circ = V$ and (1) follows from V being open. For (2) if $x \in O$, then since O is an open set containing x it is a neighborhood of x and since $O \subseteq E$ it follows that x is an interior point of E so that $O \subseteq E^\circ$. For (3) we see if $E = E^\circ$ E must be open as E° is always open by (1). On the other hand we always have that $E^\circ \subseteq E$ no matter what but if E is open by (2) we have that since $E \subseteq E$ and E is open then $E \subseteq E^\circ$ showing that $E = E^\circ$.

Now we will deal with the closed set properties. For (4) we have if $x \in \overline{E}^c$ then $x \notin E$ and $x \notin E'$. Since x isn't in E and isn't a limit point of E we have there is a neighborhood $N(x)$ that contains no points of E. Further, $N(x)$ has no limit points of E, for if $p \in N(x)$

was a limit point of E, then it would need to contain a point of E but we just showed this is false. Since $N(x) \cap \overline{E} = \emptyset$ we see that $N(x) \subseteq \overline{E}^c$ so that x is interior to $\overline{E}^c$. Since x was arbitrary every point of $\overline{E}^c$ is interior and so by (3) we see $\overline{E}^c$ is open and so by definition we have $\overline{E}$ is closed. To show (5) we see that every limit point of E is also a limit point of F so that $E' \subseteq F$ and so the union $E \cup E' \subseteq F$. Finally, for (6) if $E = \overline{E}$ it is closed by (4), and we always have $E \subseteq \overline{E}$ but if E is closed then (5) and the fact that $E \subseteq E$ give us that $\overline{E} \subseteq E$ so that $E = \overline{E}$. □

The above theorem gives a lot of results but (3) and (6) are particularly nice as they tell us the old definitions of open and closed sets from the section on Metric Spaces agree with the new topological definitions! That is a set is open if and only if every point is interior to the set, and a set is closed if and only if it contains all its limit points. Another brief note is that (4) and (6) tell us since $\overline{E}$ is closed then $\overline{\overline{E}} = \overline{E}$. Intuitively, one should think of the interior of a set as the largest open set contained in the set, and the closure as the smallest closed set containing our given set.

Having discussed the closure we can now introduce a new and very important definition!

Definition 2.11. In a Topological space (X, τ) a subset $E \subseteq X$ is **Dense** in X if $\overline{E} = X$. Intuitively, a subset is dense if every point of the topological space is close to the subset. Using the Metric Topology on $\mathbb{R}$ we can see from Theorem $1.XXX$ that the rational numbers are an example of a dense subset.

Often times rather than give the entire collection,τ, of open sets we give a smaller collection that can generate the collection τ through unions and intersections.

Definition 2.12. More formally, we call a collection of open sets $\mathcal{B} \subseteq \tau$ a **Basis** for the Topology τ if given a point $x \in X$ and an open set O with $x \in O$, then there is a basis set $V \in \mathcal{B}$ such that $x \in V \subseteq O$. We then have the following theorem helping show why the collection $\mathcal{B}$ is called a basis.

Theorem 2.13. Given a Topological space (X, τ) and a basis $\mathcal{B}$ for τ, we have that any open set O is the union of elements of $\mathcal{B}$.

Proof. For each $x \in O$, let V_x be a basis element satisfying the above definition. Then certainly we have $x \in V_x \subseteq O$ for all $x \in O$ so that

$$\bigcup_{x \in O} V_x \subseteq O,$$

but since every $x \in O \in V_x$ it also immediately follows that

$$O \subseteq \bigcup_{x \in O} V_x.$$

□

Having discussed most of the important topological definitions we can now move on to some examples! The first and easiest examples are Metric Spaces! As previously mentioned every Metric Space is a Topological Space giving a collection familiar spaces to work with. The Metric Space of Real Numbers $(\mathbb{R}, |\cdot|)$ when viewed as a Topological space will be called the **Standard Topology of** $\mathbb{R}$. Similarly, the Standard Topology of $\mathbb{R}^n$ will be the topology that coincides with the metric induced topology. Given any set X we can make two more very simple topologies. The **Trivial Topology** for a set X is $\tau = \{\emptyset, X\}$ and the **Discrete Topology** for a set X is $\tau = \mathcal{P}(x)$. At the start of the section we discussed the Extended Real Numbers $\overline{\mathbb{R}}$. The basis of the Topology for $\overline{\mathbb{R}}$ is all sets of the form $[-\infty, a)$, (a, b), and $(b, \infty]$ for $a, b \in \mathbb{R}$, the reader should verify this is indeed a topology. In the section on Metric Spaces we briefly discussed the idea of a subspace and we can extend this idea to Topological Spaces as a result of Theorem 2.XXX as follows.

Definition 2.14. Given a Topological Space (X, τ) and a subset $Y \subseteq X$ we define the **Subspace Topology on Y** as $\tau_Y = \{Y \cap T | T \in \tau\}$. It is clear how Theorem 2.XXX influences this definition.

Definition 2.15. Given two Topological Spaces (X, τ_X) and (Y, τ_Y) we can form the **Product Topology** of X and Y as the topology on the cartesian product $X \times Y$ with basis sets $\mathcal{B} = \{U \times V | U \in \tau_X \quad V \in \tau_Y\}$. This definition is the most natural way to extend a topology to cartesian products. It also turns out that taking the Product Topology of $\mathbb{R} \times \mathbb{R}$ (both with the standard topology) agrees with the standard topology of $\mathbb{R}^2$ and in general this holds for $\mathbb{R}^n$ with $n-$products of $\mathbb{R}$ using the product topology.

2.3 Sequences in Topological Spaces

The concept of a sequence should be familiar to readers, it was introduced in Chapter Zero, but to recall we define a sequence as a function $a : \mathbb{N} \to X$ where we denote the n-th element as a_n and the sequence as a whole as $\{a_n\}_{n=1}^{\infty}$. In general a sequence can map to any set, but now that we have built up the notion of a Topological Space we can discuss convergence of a sequence. Since topology broadly gives us a notion of closeness, via open sets, we can say that informally a sequence converges if the terms get arbitrarily close and stay arbitrarily close to some limit value of the sequence.

Definition 2.16. Formally, we say a sequence $\{a_n\}_{n=1}^{\infty}$ **Converges to a point** p if for every neighborhood $N(p)$ there exists a natural number $N \in \mathbb{N}$ so that for all $n \geq N$ we have $a_n \in N(p)$. In this case we say that p is a limit of the sequence a_n and denote this by $\lim_{n\to\infty} a_n = p$ or $a_n \to p$ as $n \to \infty$. This must hold for EVERY neighborhood of the point p although the number N need not be the same for every neighborhood. If a sequence does not converge it is said to **Diverge**. Similarly, we can regard a subsequence as a sequence in its own right and discuss convergence or divergence of said subsequence.

From Calculus and previous experiences with sequences we have a lot of preexisting notions of how sequences should behave. For example we intuitively should think that limits of sequences need to be unique as converging to two points seems ludicrous. Similarly, we described earlier that limit points were in a sense "close" to a set and so one might think that if we have a limit point we might be able to construct a sequence of elements in the set converging to said limit point since it is so close to the set. It turns out that in general topology neither of these hold!

Consider the set $\mathbb{R}$ where we specify a set X is closed if $X = \mathbb{R}$ or if X is countable. It is clear that any finite union of countable sets is countable as is any arbitrary intersection of countable sets, so our closed sets follow the necessary rules. Since open sets are just the compliments of closed sets we have defined a topology through the compliments of each of the closed sets given. This topology is called **The Co-Countable Topology**. Now consider the set $A = [0, 1]$ with this topology. The closure $\overline{A}$ is the smallest closed set containing A, but the only closed set containing A is $\mathbb{R}$ itself! So that for example $-1 \in \overline{A}$ and since $\overline{A} = A \cup A'$ we have -1 is a limit point of $[0, 1]$ with this bizarre topology. However, there is no sequence in A that converges to -1 as given any sequence $a_n \in A$ the set of all sequence members is a countable set not containing -1 and therefore the complement is an open set containing -1, so that it is a neighborhood of -1 containing no members of the sequence. This shows that -1 is a limit point of A with no sequence in A converging to it.

For an easy example of a sequence having two limits consider the Topological Space where $X = \{1, 2, 3\}$ and $\tau = \{\emptyset, \{1, 2\}, \{1, 2, 3\}\}$. Now consider the sequence given by $a_n = 2$ for all n. The reader should guess that the sequence a_n converges to the point 2 and indeed it does! But notice that by our definition the sequence also converges to the point 1!

Throughout the next two sections we will see that with some additional properties these problems of sequences are fixed!

2.4 Topological Properties

Often times in math we want to generalize things as much as possible and through the notion of a Topological Space we have very successfully generalized many different types of spaces. Now however we must work backwards and re-instill our Topological Spaces with a collection of properties that are very useful for the purposes of Analysis! The main properties we will discuss here are: Countability Axioms, Connectedness, and Compactness. The Hausdorff property will be discussed in its own section!

We now want to discuss the Countability Axioms starting with the Axiom of First Countability.

Definition 2.17. We have already discussed the concept of a basis for a Topological Space, but now we want to discuss a basis of some point x of a Topological Space. Given a Topological Space (X, τ) and a point $x \in X$ we say that a collection of open sets containing the point x, $\mathcal{B}_x$, is a **Local Basis** for the point x if for every neighborhood of x, $N(x)$, there contains an element $B \in \mathcal{B}_x$ so that $x \in B \subseteq N(x)$. If every element of X has a countable local basis then we say that (X, τ) is a **First Countable Space**.

The next theorem shows that First Countable spaces fix one of our issues with sequences!

Theorem 2.18. Given a first countable space (X, τ) and a subset $S \subseteq X$ a point x is in $\overline{S}$ if and only there is a sequence of elements in S that converges to x.

Proof. First assume that there is a sequence of elements in S that converge to x. That is there is a sequence x_n that converges to x. Then by our definition of convergent sequence we know that every neighborhood of x contains all but finitely many points of x_n which are all points of S. If $x \in S$, then trivially it is in $\overline{S}$, but if $x \notin S$ then every neighborhood contains a point of S, being one of the points in the sequence x_n, so that by definition x is a limit point of S and so $x \in \overline{S}$. Note this first direction did not require first countability and is actually true in every Topological Space! For the second direction we assume that $x \in \overline{S}$. If $x \in S$ then the constant sequence $x_n = x$ for all $n = 1, 2, ...$ satisfies the criterion. If $x \notin S$ then we can now use the first countability of X. Let $\mathcal{B}_x$ be a countable local basis of x and then label the elements of $\mathcal{B}_x$ as $B_1, B_2, ...$ We know x must be a limit point of S so every neighborhood around x contains a point of S. Let x_1 be a point of S in the set B_1. We will use induction for the rest of the elements. Assume we have x_N chosen, then we will take the intersection of all $B_1 \cap B_2 \cap ... \cap B_N = \mathcal{O}_N$, since $\mathcal{O}_N$ is the union of a finite number of open sets it is open and it certainly contains x so there is a new local basis element B_m so that $x \in B_m \subseteq \mathcal{O}_N$ and now we choose x_{n+1} to be a point of S lying in B_m, we again know this exists as EVERY neighborhood of x must contain some point of S. This sequence of x_n converges to x as every neighborhood eventually contains all of the remaining terms in the sequence and so our work is done. □

This theorem is very nice as it allows us to regard limits of sequences and limit points as very similar objects, and we shall see in Chapter 4 that more nice properties follow. For the mean time however we will continue to the next Countability Axiom!

Definition 2.19. We say that a Topological Space (X, τ) is a **Second Countable Space** if there exists a countable basis of X. Note, this is not a local basis, but a full basis for the Topological Space as discusses in the previous section. Second Countablity also instantly guarantees the space is also First Countable! Another useful property comes in the form of the following Theorem.

Theorem 2.20. A Second Countable Space (X, τ) contains a countably dense subset.

Proof. Since X has a countable basis $\mathcal{B}$ first index the elements of $\mathcal{B}$ as $B_1, B_2, ...$ and then for each nonempty B_n choose an element $x_n \in B_n$. Then the claim is that the set $S = \{x_n | n \in \mathbb{N}\}$ is dense in X. To see this let $x \in X$ we need to show that $x \in \overline{S}$, but given any neighborhood $N(x)$ we see that there is a basis element B_i so that $x \in B_i \subseteq N(x)$ and since $x_i \in B_i \subseteq N(x)$ we see that the neighborhood of x contains an element of S. So either $x \in S$ or x is a limit point of S and in either case we get $x \in \overline{S}$. □

While there are more applications of The Countability Axioms we will need continuous functions to explore the rest. This will be a common theme througout this Chapter as the strong connection between Topology and Continuous Functions prevents us from fully exploring each topic we introduce. Having said that we can now discuss the next topic being Connectedness. Informally, most people have a sense of what it means to be connected. A space is "connected" if it consists of one piece and is not split into several distinct objects. We will now make this informal definition more concrete.

Definition 2.21. Given a Topological Space (X, τ) we say two sets $A, B \subseteq X$ are **Separated** if $\overline{A} \cap B = A \cap \overline{B} = \emptyset$. We say that the Topological Space X is **Connected** if it cannot be expressed as the union of two nonempty separated sets. Essentially, a space is connected if it cannot be split into to separated pieces.

Connected sets do not fully show their importance until the section on continuous functions, but we will give a few minor theorems now that can be nice at times.

Theorem 2.22. Given a Topological Space (X, τ) and two connected subsets $A, B \subseteq X$. If there is a point p in both A and B, then the union $A \cup B$ is connected.

Proof. For contradiction assume that $A \cup B$ is not connected, then there are nonempty separated sets C, D so that $A \cup B = C \cup D$. We know that the point p must be exclusively in either C or D (it cannot be both as the sets are separated) so without loss of generality we will say that $p \in C$. Now, since D is nonempty it contains either a point of A or B, if it contains a point of A, then we have that $A = A \cap (A \cup B) = (A \cap C) \cup (A \cap D)$ and since $A \cap C \subseteq C$ and $A \cap D \subseteq D$ and both are nonempty we see that $(A \cap C)$ and $(A \cap D)$ are two nonempty separated sets whose union gives A contradicting the fact that A was connected! The same contradiction arises if D contained instead a point of B! $\square$

Theorem 2.23. If A, B are disjoint closed sets then they are separated. Same applies if they are disjoint open sets.

Theorem 2.24. A space X is connected if and only if the only sets that are both open and closed are: $\emptyset$ and X.

Proof. For the first direction assume X is connected and for contradiction assume that there is some other set $A \subseteq X$ that is both open and closed. Then it immediately follows that A^c is also open and closed and since both A and A^c are closed we have $\overline{A} = A$ and $\overline{A^c} = A^c$. It then follows that since $A \cap A^c = \emptyset$ we immediately have that A and A^c are two nonempty separated sets whose union gives X so that X is not connected, a contradiction. Now for the other direction assume that X and $\emptyset$ are the only subsets that are both open and closed and assume for contradiction that X is not connected. Then there are nonempty separated sets A, B so that $X = A \cup B$. Let $x \in A$ since $A \cap \overline{B} = \emptyset$ we see that x is not a limit point of B so that there is some neighborhood $N(x) \subseteq A$ and so x is interior to A. Since x was arbitrary we see that A is open. Further the same argument shows that B is open. For the last step, we see that if $p \notin A$ since $X = A \cup B$ it must be the case that $p \in B$ so that $A^c = B$ and since $B = A^c$ is open, we have that A is closed so that A is a open and closed subset of X, a contradiction. $\square$

Since this last theorem was a bidirectional it could be taken as the definition of Connectedness if so desired and characterizes Connectedness completely!

The last and most important Topological Property of this section is Compactness. Compactness is perhaps the most important topic in all of Analysis and has incredibly powerful and useful consequences. Unfortunately, it has a quite confusing definition for newcomers and can be quite hard to use and appreciate until a great deal of familiarity is developed. Intuitively, compactness means a set or space is "Topologically Finite" in a certain sense and with this we will see that some very useful facts about finite sets can be extended to compact sets! But first we must give a precise definition.

Definition 2.25. Given a Topological Space (X, τ) a subset $K \subseteq X$ an **Open Cover** of K is a set of open sets $\{\mathcal{O}_\alpha | \alpha \in A\}$ such that $K \subseteq \bigcup_{\alpha \in A} \mathcal{O}_\alpha$, ie an open cover is a collection of open sets that "covers" K. We say that K is **Compact** if for EVERY open cover $\{\mathcal{O}_\alpha\}$ there exists a finite collection of cover elements $\mathcal{O}_1, \mathcal{O}_2, ..., \mathcal{O}_n$ so that $K \subseteq \mathcal{O}_1 \cup ... \cup \mathcal{O}_n$. In other words K is Compact if EVERY open cover has a finite subset of open sets that also covers K, a finite sub-cover.

The very important thing is that this must hold for EVERY open cover, not just any in particular cover. This means to show that some set K is compact you must start with an arbitrary open cover and find some way of showing that only finitely many are needed. We claimed that Compactness is the topological version of finiteness and the reader should prove that finite sets are indeed compact. To prove in general that a set is compact can be quite difficult, but in the next theorems and in later sections we will build a few tools that can help!

Theorem 2.26. Given a Topological Space (X, τ) and a compact subset $K \subseteq X$ we have that if $D \subseteq K$ is closed then D is also compact.

Proof. To begin we start with an arbitrary open cover of D so let $\{\mathcal{O}_\alpha\}$ be an open cover of D. Now since D is closed we have that D^c is open, and since $D \subseteq K$ we have that $\{\mathcal{O}_\alpha\} \cup \{D^c\}$ is an open cover for K. Since K is compact we know that it has a finite subcover. Let $\{\mathcal{O}_1, ..., \mathcal{O}_n\}$ be the open subcover possibly minus the set D^c if that was part of $K's$ subcover. Then all $\mathcal{O}_i$ $i = 1, ..., n$ were part of the original cover of D and also form a subcover of D so that D has a finite subcover. Since the open cover we chose for D was arbitrary we see any open cover has a finite subcover and so D is compact. $\square$

Theorem 2.27. If (X, τ) is a Topological Space and $K \subseteq X$ then K is a compact subset of X if and only if K is a compact subset of the subspace (K, τ_K).

The proof of the above is omitted since it is quite easy, but it is important as it implies we can regard compact sets as subsets of a larger Topological Space or we can focus on the subspace on its own if we like! However, note if we are regarding the subspace (K, τ_K) as a compact space, then an open cover $\{O_i | i \in I\}$ must satisfy the fact that $\bigcup_{i \in I} O_i = K$ since every set in the subspace is a subset of K.

Definition 2.28. A collection of sets A has **The Finite Intersection Property** if every finite subset $a_1, a_2, ..., a_n \in A$ has a nonempty intersection, $a_1 \cap a_2 \cap ... \cap a_n \neq \emptyset$.

Theorem 2.29. A space (K, τ) is compact if and only if every collection of closed subsets of K, we will call $\mathcal{A}$, guarantees that the intersection of all subsets in $\mathcal{A}$ is nonempty. Note here we are regarding the whole space as a compact space.

Proof. If K is compact and the intersection is empty, let $A^* \in \mathcal{A}$ be any member. Then every point of A^* cannot be in all other $A \in \mathcal{A}$ since the intersection is empty and since A^* is a closed subset of K it is also compact. But then we see that the sets $\{A^c | A \in \mathcal{A}\}$ form an open cover of A^* since each point of A^* is in some A^c. But then there is a finite subcover so that $A^* \subseteq A_1^c \cup ... \cup A_n^c$, but then DeMorgan's Laws tell us that $(A_1^c \cup ... \cup A_n^c)^c = A_1 \cap ... \cap A_n$ so that $A^* \cap A^1 \cap ... \cap A^n$ must be empty contradicting the finite intersection property.

Now assume that every collection of closed subsets of K having the finite intersection property has a nonempty intersection and let $\mathcal{O}$ be an open cover of K. For contradiction we assume K is NOT compact, then it follows for every finite collection of $\{O_i \in \mathcal{O}\}_{i=1,...,n}$ we have that $\bigcup_{i=1}^n O_i \neq K$, but then complimentation and DeMorgan's Laws give that $\bigcap_{i=1}^n O_i^c \neq \emptyset$ since there was a point in K not in the union. So the collection $\{O^c | O \in \mathcal{O}\}$ has the finite intersection property so by assumption we have that $\bigcap_{O \in \mathcal{O}} O^c \neq \emptyset$, but this implies again from complementation and DeMorgan's Laws that $\bigcup_{O \in \mathcal{O}} O \neq K$ contradicting the fact that $\mathcal{O}$ was an open cover of K! So we see that K must be compact. $\square$

Corollary 2.29.1. As an easy corollary, we have that if K is a compact set and A_n is a collection of nonempty nested closed subsets of K, ie $A_{n+1} \subseteq A_n \subseteq K$, then the intersection of all A_n is nonempty.

Theorem 2.30. If X and Y are two compact spaces, then the product space $X \times Y$ is compact.

To prove this theorem we will need to build up some machinery. More specifically, we will need the notion of projection and a few Lemmas involving projection.

Definition 2.31. The function $\pi_1 : X \times Y \to X$ by $\pi_1(x, y) = x$ is the **Projection** of $X \times Y$ onto its first factor. We will prove a small Lemma about projection mappings before Proving Theorem 2.29

Lemma 2.32. Given a product Topology $X \times Y$ the function π_1 given above is an open mapping, meaning that if S is an open subset of $X \times Y$, then $\pi_1(S)$ is an open subset of X.

Proof. Let $x \in \pi_1(S)$. It follows that for some $y \in Y$ that $(x, y) \in S$. Since S is open there is a basis element of $X \times Y$ we shall call $\mathcal{B}$ so that $(x, y) \in \mathcal{B} \subseteq S$. But recalling how we defined the basis for a product topology we see that $\mathcal{B} = O_1 \times O_2$ where O_1 is an open subset of X. It follows immediately that $x = \pi_1(x, y) \in O_1 \subseteq \pi_1(S)$ so that x is interior to $\pi_1(S)$. Since x was arbitrary we see that $\pi_1(S)$ is open. □

Lemma 2.33. Given a product Topology $X \times Y$ if Y is compact then the projection π_1 given above is a closed mapping, meaning that if S is a closed subset of $X \times Y$, then $\pi_1(S)$ is a closed subset of X.

Proof. Let S be a closed subset of $X \times Y$. We want to prove that $\pi_1(S)$ is closed. Let $p \notin \pi_1(S)$, then it follows for all $y \in Y$ that $(p, y) \notin S$ and since S was closed we see that for each $y \in Y$ there is a neighborhood $N(p, y)$ disjoint from S. We then have that the collection $\{\pi_2(N(p, y))|y \in Y\}$, using the previously found neighborhoods, form an open cover for Y and since Y is compact there is a finite subcover so that $Y = \pi_2(N(p, y_1)) \cup ... \cup \pi_2(N(p, y_n))$ for $i = 1, ..., n$, so now the intersection of $N(p, y_1) \cap ... \cap N(p, y_n) = U$ is disjoint from S,is open (finite intersection of open sets), and for all $y \in Y$ we have $(p, y) \in U$. It is immediate that $\pi_1(U)$ is open by previous lemma and we claim that $\pi_1(U)$ is disjoint from $\pi_1(S)$. For if there was a point q in both then there must be some $y \in Y$ so that $(q, y) \in A$, but $(q, y) \in S$ contradicting the fact S and A were disjoint. It follows that $\pi_1(U)$ is an open set containing p disjoint from $\pi_1(S)$, so that p is not a limit point of $\pi_1(S)$. This means $\pi_1(S)$ contains all its limit points and is therefore closed. □

Our last Lemma here is commonly called the Tube Lemma.

Lemma 2.34 (The Tube Lemma)**.** Given a product Topology $X \times Y$ with Y compact. Let $p \in X$ and let S be open in $X \times Y$. If S contains the "slice" $\{p\} \times Y$, then there an open subset $O \subseteq X$ so that $O \times Y \subseteq S$. The set $O \times Y$ is the so called "Tube" containing the slice $\{p\} \times Y$.

Proof. Consider $S^c \subseteq X \times Y$. This set is closed and so $\pi_1(S^c)$ is as well by above theorem. Since the slice $\{p\} \times Y \subseteq S$ it follows that $p \notin \pi_1(S^c)$ and since $\pi_1(S^c)$ is closed there is a neighborhood $N(p) \subseteq X$ disjoint from $\pi_1(S^c)$ (since p is not a limit point). But then the set $N(p) \times Y$ is our set! We see that if $x \in N(p)$ then $x \notin \pi_1(S^c)$ so that for all $y \in Y$ $(x, y) \notin S^c$ giving $(x, y) \in S$ and since x was an arbitrary point of $N(p)$ we see that $N(p) \times Y \subseteq S$. □

We can now finally prove theorem 2.29.

Proof. Let $\mathcal{O}$ be an open cover of $X \times Y$. Since Y is compact, we see the subset $\{x\} \times Y \subseteq X \times Y$ is compact for any $x \in X$ (this follows easily from the Lemma that projections are open mappings). Let $x \in X$, by previous comment since $\{x\} \times Y \subseteq X \times Y$ there exists a finite collection $O_1, ..., O_n \in \mathcal{O}$ that covers $\{x\} \times Y$, denote this collection $\mathcal{O}_x$. The union $S_x = O_1 \cup ... \cup O_n$ is an open set containing $\{x\} \times Y$ and so by the Tube Lemma there is an open subset $O_x \subseteq X$ so that $O_x \times Y \subseteq S \times Y$. The collection $\{O_x | x \in X\}$ forms an open cover of X and by compactness there is a finite subcover $O_{x_1}, ..., O_{x_n}$ that covers X. Since for each O_{x_i} we have $O_{x_i} \times Y \subseteq S_{x_i}$ and each $S_{x_i} = \bigcup \mathcal{O}_{x_i}$ with each $\mathcal{O}_{x_i}$ finite, it follows that the collections of all $\mathcal{O}_{x_i}$ $i = 1, ..., n$ is a finite collection of $\mathcal{O}$ that covers $X \times Y$ completing the proof. □

Corollary 2.34.1. As an easy corollary to the difficult theorem above, we see that the product of any finite number of compact sets is again compact. This follows by repeating the above theorem as needed.

We will not prove it here, but it is infact true that the product of any arbitrary collection of compact sets is again compact. This very important result is known as Tychonoff's Theorem! This theorem requires the axiom of choice and will not be proven here!

The last topics of this section are two related notions of compactness.

Definition 2.35. Given a Topological Space (X, τ) a subset K is said to be **Limit Point Compact** if every infinite subset of K has a limit point in K. We say that K is **Sequentially Compact** if every sequence in K has a convergent subsequence in K.

Theorem 2.36. A compact space (X, τ) is always limit point compact.

Proof. Suppose X is compact but not limit point compact. Then there is an infinite subset ,S, of X containing no limit point in X. It follows then that for each $x \in X$ there is a neighborhood $N(x)$ containing no other point of X, but this implies that $N(x) = \{x\}$, so that $\{x\}$ is open. Clearly then the collection $\{\{x\}|x \in X\}$ is an open cover of X with no finite subcover, contradicting the assumption that X was compact. $\square$

So far we see that Compactness always implies limit point compactness. To conclude this section we will present one more theorem relating sequential compactness to compactness, and in the next section we will show that in certain situations, the three are equivalent!

Theorem 2.37. If (K, τ) is a second countable and sequentially compact space, then it is compact.

Proof. Let $\mathcal{A}$ be an open cover for K. Since K is second countable, we can find a countable subcover we will call $\mathcal{A}'$ by choosing one open set per basis element in $\mathcal{A}$. Now for contradiction we assume $\mathcal{A}'$ has not finite subcover. We will make a sequence in K with no convergent subsequence. Since $\mathcal{A}'$ is a countable collection of sets they can be indexed by the natural numbers and so for each $n = 1, 2, ...$ let $A_n \in \mathcal{A}'$ be the nth set, by the given index. Furhter, since K has no finite subcover we know for all n the set $K - (A_1 \cup ... \cup A_n)$ is nonempty, otherwise $A_1, ..., A_n$ would cover K. So we form a sequence $\{a_n\}_{n=1}^{\infty}$ by letting

a_n be any element in the set $K - (A_1 \cup ... \cup A_n)$. Since K is sequentially compact it has a convergent subsequence, $a_{n_k \to a}$ as $k \to \infty$. But this means that for every neighborhood, $N(a)$, of a there exists an $N \in \mathbb{N}$ so that for all $k \geq N$ $a_{n_k} \in N(a)$. But since $\mathcal{A}'$ was an open cover of K there is some set A_i so that $a \in A_i$. From here we see the contradiction as for all $k \geq i$ we see that $a_{n_k} \notin A_i$ so that we have reached a contradiction! $\square$

2.5 Hausdorff Spaces

We have covered many Topological Properties so far, but several issues still remain. Most notably at the end of the section on Topological Spaces we showed that sequences need not have unique limits, a bizarre fact we would not typically expect. The following section is dedicated to the property that resolves this. It turns out that the reason our sequence was able to converge to more than one point is because these points were topologically indistinguisable, there was no open sets splitting them apart. It turns out that there are many so called **Separation Axioms** we can impose on our Topological Spaces to distinguish points and sets, however for this book we will focus on the most important one: being Hausdorff. This specific axiom will give rise to a great deal of important and useful results specifically for Analysis, but first we must define this property.

Definition 2.38. We say that a Topological Space (X, τ) is **Hausdorff** if for each pair of points $x, y \in X$ there exist neighborhoods $N(x)$ and $N(y)$ of x and y respectively so that $N(x) \cap N(y) = \emptyset$. We may also call the space X a **Hausdorff Space**.

We will now deliver on the promise that limits are unique in Hausdorff Spaces.

Theorem 2.39. Let (X, τ) be a Hausdorff Space, and let $\{a_n\}_{n=1}^{\infty}$ be a sequence in X, then a_n has at most one limit.

Proof. Suppose for contradiction that a_n has two limits x and y. Since X is Hausdorff there exist neighborhoods $N(x)$ and $N(y)$ with $N(x)$ and $N(y)$ disjoint. Since $a_n \to x$ there exists an $N_0 \in \mathbb{N}$ so that for all $n \geq N_0$ $a_n \in N(x)$ and similarly there exists an N_1 so that for $n \geq N_1$ $a_n \in N(y)$. But from here we see the contradiction as setting $N = \max\{N_0, N_1\}$ we see that for all $n \geq N$ that $a_n \in N(x)$ and $a_n \in N(y)$ contradicting the fact they were disjoint sets! □

Another nice fact about Hausdorff Spaces is that it guarantees that comapct sets are closed!

Theorem 2.40. Suppose (X, τ) is a Hausdorff space and $K \subseteq X$ is compact, then K is closed.

Proof. We will prove this by showing that K^c is open. Let $p \in K^c$ and for each $x \in K$ let $N(x)$ and $N(p)_x$ be the disjoint sets separating p and x. Clearly the collection of sets $\{N(x)|x \in K\}$ form an open cover of K and so by compactness there is a finite subcollection $N(x_1), N(x_2), ..., N(x_n)$ that covers K. But to each $N(x_i)$, $i = 1, ..., n$ we have the corresponding sets around p being $N(p)_{x_i}$. Now let $U = \bigcap_{i=1}^n N(p)_{x_i}$, then U is an open set containing p as it is the intersection of a finite number of open sets containing p. Further, we see that $U \cap K = \emptyset$ because if there was a point $q \in U \cap K$, then $q \in N(x_i)$ for some i but $q \in U \subseteq N(p)_{x_i}$ which is disjoint from $N(x_i)$, so we see there cannot be a point in the intersection of U and K. So now we see that $U \subseteq K^c$ so that p is interior to K^c. Since p was arbitrary we see K^c is open and therefore K is closed. □

Corollary 2.40.1. As a nice corollary, since finite sets are always compact, in a Hausdorff space finite sets, and even more specifically singletons $\{x\}$, are always closed.

Theorem 2.41. If (X, τ) is Hausdorff, $S \subseteq X$, and p is a limit point of S, then every neighborhood of p contains infinitely many points of S, so that finite sets have no limit points.

Proof. Suppose for contradiction there was a neighborhood of p containing only finitely many points of S (distinct from p itself) labelled $x_1, ..., x_n$. Then for each x_i there are disjoint neighborhoods $N(x_i)$ and $N(p)_{x_i}$. The intersection $U = N(p)_{x_1} \cap ... \cap N(p)_{x_n}$ is an open set containing p and since $U \subseteq N(p)_{x_i}$ we have $x_i \notin U$ for all $i = 1, ..., n$ so the only point of S that can possibly be in U is perhaps p itself. It follows that p cannot possibly be a limit point of S contradicting the assumption. $\square$

To conclude this short section we want to finish our discussion on the various forms of compactness. So far we have that compactness implies limit point compactness and that in a second countable space sequential compactness implies compactness. If we can bridge the gap from limit point compactness to sequential compactness we can show that in certain cases all three are equivalent. Our next theorem addresses this!

Theorem 2.42. Suppose (K, τ) is a first countable, Hausdorff, limit point compact space. Then K is sequentially compact. For those with a stronger Topological background, the Hausdorff requirement can be restricted to a space being T_1, but this is not discussed further here.

Proof. To begin let $\{a_n\}_{n=1}^{\infty}$ be a sequence of elements in K. If there are only finitely many different terms in the sequence, then it is obvious that some subsequence must converge as at least one term must repeat infinitely many times. Now we can focus on the case that there are infinitely many distinct terms of the sequence. Let $A = \{a_n | n = 1, 2, 3, ...\}$ be the set of all terms in the sequence. Clearly A is an infinite subset of K so by limit point compactness, A has a limit point, p, in K. Since K is Hasdorff, this also implies that every neighborhood of p contains infinitely many points of A. Now we use first countability. Let $\mathcal{B}$ be the countable local basis for p, indexed as $B_1, B_2, ... \in \mathcal{B}$. Let a_{n_1} be any element of $A \cap B_1$, there are infinitely many points satisfying this so we can certainly pick one. Now we inductively pick the remaining elements of the subsequence. Suppose we have picked a_{n_k} and we want to pick $a_{n_{k+1}}$. We see that $O = B_1 \cap ... \cap B_{k+1}$ is an open set containing p, so there is a basis element B_m so that $p \in B_m \subseteq O$. Since a_{n_k} was chosen to be an element of A we know there is an index i so that $a_{n_k} = a_i$ in our original sequence. Since B_m contains an infinite number of points of A we can certainly choose an element in A and B_m with a higher value index that a_i and so let this point be our next term in the subsequence $a_{n_{k+1}}$. By induction we will construct a subsequence of our original sequence and it is easy to see that this subsequence will converge to p. Since our sequence was arbitrary, we see every sequence has a convergent subsequence and so K is sequentially compact. $\square$

Theorem 2.43. If (K, τ) is a second countable Hausdorff space, then the following are equivalent.

1. K is compact.
2. K is sequentially compact.
3. K is limit point compact.

This Theorem follows from our previous theorems relating versions of compactness.

2.6 Return to Metric Spaces

Having discussed Topological Spaces in some detail we can now come back to our more familiar Metric Spaces with a deeper understanding and appreciation. Our next theorem will instantly show us that many of the useful theorems we've proved can transfer to metric spaces.

Theorem 2.44. Every metric space (X, d) is first countable and Hausdorff.

Proof. Let $p \in X$. The sets $B_{\frac{1}{n}}(p)$ for $n = 1, 2, 3, ...$ form a countable local basis for p. We see this as if $N(p)$ is an arbitrary neighborhood of p then p is interior to $N(p)$ so there is some $B_r(p) \subseteq N(p)$ and by the Archimedean Property of $\mathbb{R}$ we can choose $N \in \mathbb{N}$ so that $\frac{1}{r} < N$ giving $B_{\frac{1}{N}}(p) \subseteq B_r(p) \subseteq N(p)$. To see that X is Hausdorff let $x, y \in X$ and let $r = \frac{1}{2}d(x, y)$. Then it follows that $B_r(x)$ and $B_r(y)$ are disjoint open sets containing x and y respectively. $\square$

Another incredible fact about metric spaces is that in a metric space all forms of compactness we've discussed are equivalent!

Theorem 2.45. If (X, d) is a sequentially compact metric space, it is second countable.

To prove this theorem we will need a useful lemma.

Lemma 2.46. If (X, d) is a metric space with a countable dense subset, then it is second countable.

Proof. Let A be the countable dense subset. For each $a \in A$ consider the collection of open balls $B_q(a)$ where $q \in \mathbb{Q}$. The collection of all these open balls, we will call $\mathbb{B}$, is certainly countable as it is a countable union of countable sets. Now we show it forms a basis for X. Let p be an arbitrary point of X and let O be an open set containing p. Then there is some open ball $B_r(p) \subseteq O$. If $p \in A$ then choosing a value $q \in \mathbb{Q}$ with $0 < q < r$ gives $B_q(p) \subseteq O$ is a member of of $\mathbb{B}$. If $p \notin A$, then since $\overline{A} = X$ we see p is a limit point of A. So there is some point $w \in B_r(p)$ so that $w \in A$j and $d(w, p) < r$. But now choosing $q \in \mathbb{Q}$ so that $q < r - d(w, p)$ tells us that $p \in B_q(w)$ and if $x \in B_q(w)$ then $d(p, x) \leq d(p, w) + d(w, x) < d(w, p) + r - d(w, p) = r$ so that $p \in B_q(w) \subseteq B_r(p) \subseteq O$ and $B_q(w) \in \mathcal{B}$. So we see that $\mathcal{B}$ is a countable basis for X. □

Now we prove the original Theorem.

Proof. Let X be our sequentially compact metric space. Let $\epsilon > 0$. We will show that X can be covered by finitely many open balls of radius ϵ. Assume this is not the case. We construct a sequence as follows. Let $x_1 \in X$. Inductively assume x_n has been chosen and choose x_{n+1} so that $d(x_i, x_{n+1}) \geq \epsilon$ for all $i = 1, ..., n$, this can be done as no finite collection of balls of radius ϵ will cover X. The sequence $\{x_n\}_{n=1}^{\infty}$ has no convergent subsequence, however as if $p \in X$ then by construction we see that $B_\epsilon(p)$ is a neighborhood of p that cannot contain infinitely many terms of x_n, so we see that it must be the case that X can be covered by finitely many open balls of radius ϵ. Since ϵ was arbitrary we can do this for any $\epsilon > 0$ and so certainly for $\epsilon = \frac{1}{n}$ for any $n \in \mathbb{N}$. Now for each $n \in \mathbb{N}$ we know $X = \bigcup_{i=1}^{k} B_{\frac{1}{n}}(x_{n_i})$ so let $X_n = \{x_{n_i}\}$. Each X_n is finite so the union $X^* = \bigcup_{n=1}^{\infty} X_n$ is a countable union of finite sets and so is countable. Finally, we claim that X^* is a dense subset of X. If $t \in X$ is not in X^* then consider some $B_r(t)$ we can choose $n \in \mathbb{N}$ so that $\frac{1}{n} <$ and it follows then since the collection $B_{\frac{1}{n}}(x_{n_i})$ covers X that $t \in B_{\frac{1}{n}}(x_{n_m})$ for some m so that $d(x_{n_m}, t) < \frac{1}{n} < r$ and so x_{n_m} is a point of $X^* \in B_r(t)$. We now see that t is a limit point of X^* and since t was an arbitrary point of X it follows X^* is a countable dense subset of X and by the lemma we see that X is second countable. □

Corollary 2.46.1. The above theorem plus Theorem 2.43 tells us that in any metric space all three forms of compactness are equivalent! Since compactness implies limit point compactness which in a Hausdorff first countable space, of which all metric spaces are, implies sequential compactness which with the above theorem implies regular compactness!

Definition 2.47. We want to end our discussion on Compactness by discussing boundedness. We say a set A in a Metric Space X is **Bounded** if there exists some ope nball $B_r(p)$ so that $A \subseteq B_r(p)$. Essentially, a set is bounded if it can be fit inside some open ball. It is easy to see that compact sets are bounded, for consider the open cover by open balls of radius 1.

Definition 2.48. We now want to talk more about Sequences in Metric Spaces as there is a major property of sequences that cannot be explored in the more general topological space. Since we have access to a distance function our definition of convergence can be restated as follows: We say that a sequence a_n **Converges** to a point p if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for for all $n \geq N$ we have $d(a_n, p) < \epsilon$. It can easily be seen that this definition coincides with the more general version. If a_{n_k} is a subsequence that converges to a point q then we say that q is a **Subsequential Limit**. We say that a sequence is bounded if the set of elements in the sequence $\{a_n | n \in \mathbb{N}\}$ is a bounded set. It is not too difficult to see that convergent sequences are bounded.

Definition 2.49. We now introduce a very important definition regarding sequences. We say a sequence a_n is **Cauchy** if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for all $m, n \geq N$ we have $d(a_n, a_m) < \epsilon$. Explicitly, rather than the terms of a sequence getting close to some point p the terms are getting arbitrarily close to each other. It is easy to see that every convergent sequence is Cauchy, but the converse need not be true, consider the metric space $(0, 1]$ with the standard metric and the sequence $a_n = \frac{1}{n}$. This sequence is Cauchy, but does not converge. What is true is that every Cauchy sequence is also bounded. If X is a metric space where every Cauchy sequence converges, it is called a **Complete Metric Space**. Complete Metric Spaces are incredibly important as you can prove convergence just by showing that a sequence is Cauchy.

We now want to prove that Compact Metric Spaces are Complete.

Lemma 2.50. Let a_n be a sequence in some Metric Space X and let A be the set of all subsequential limits. If a_n is Cauchy then A has at most one element.

Proof. Suppose for contradiction we have $p, q \in A$ where $p \neq q$. Let $\epsilon > 0$. Since there is a subsequence a_{n_k} that converges to p we can choose a $K \in \mathbb{N}$ so that for all $k \geq K$ we have $d(a_{n_k}, p) < \frac{\epsilon}{3}$. Similarly, we can find a subsequence a_{n_m} and an M so that $d(a_{n_m}, q) < \frac{\epsilon}{3}$ for $m \geq M$. Finally, since the sequence is Cauchy we can find an N so that for all $n, w \geq N$ $d(a_n, a_w) < \frac{\epsilon}{3}$. But now if we set $N' = \max(n_K, n_M, N)$ then the triangle inequality gives for all $n, m \geq N'$ $d(p, q) \leq d(p, a_n) + d(a_n, a_m) + d(a_m, q) \leq \epsilon$. Since ϵ was arbitrary we see that $d(p, q) = 0$ so that $p = q$ a contradiction. $\square$

Theorem 2.51. Let X be a Compact metric Space. Then X is also Complete.

Proof. We need to show that every Cauchy sequence converges. Let a_n be Cauchy, then by above the set of all subsequential limits A has at most one element, but since X is compact it is also sequentially compact so that a_n has a convergent subsequence, and so A has at least one element. We see A has exactly one element we will call x. We now show that a_n converges to x. Given $\epsilon > 0$ since there is a subsequence a_{n_k} converging to x there is an M so that for all $m \geq M$ $d(a_{n_m}, x) < \frac{\epsilon}{2}$ and by Cauchyness there is an N so that for all $n, k \geq N$ $d(a_n, a_k) < \frac{\epsilon}{2}$. But now for all $n \geq N' = \max(n_M, N)$ we have $d(a_n, x) \leq d(a_n, a_{n_N}) + d(a_{n_N}, x) \leq \epsilon$. □

2.7 Return to $\mathbb{R}^n$

Finally we return to $\mathbb{R}^n$. Having developed a host of Topological Properties we want to apply them to this more familiar space and see how they can be helpful. We will first cover compactness.

Theorem 2.52. Let A be a compact subset of $\mathbb{R}$. Then $\alpha = \sup A \in A$. The same applies to $\inf A$. In other words, a Compact set of real numbers has a maximum and a minimum.

Proof. Since compact sets are bounded A is a bounded subset of R so α exists and because A is compact it is closed. We will assume that $\alpha \notin A$ for a contradiction. Given $\epsilon > 0$ since α is the least upper bound we know that $\alpha - \epsilon$ is not an upper bound so that there exists some $x \in A$ with $\alpha - \epsilon \leq x < \alpha$, but this implies that $x \in B_\epsilon(\alpha)$. Since ϵ was arbitrary we see that α is a limit point of A, but since A is closed it contains all its limit points and therefore $\alpha \in A$ a contradiction! □

We have discussed compactness a great deal, we now want to know which subsets of $\mathbb{R}^n$ are compact, the very important Heine-Borel Theorem will answer this question!

Theorem 2.53. For $a < b$ the interval $[a, b]$ is compact.

Proof. Assume $[a, b]$ is not compact and let $\mathcal{O}$ be an open cover. Since $[a, b]$ is not compact at least one of $[a, \frac{b-a}{2}]$ or $[\frac{b-a}{2}, b]$ does not have a finite subcover, if both did then $[a, b]$ would be compact. Similarly, we can split whichever interval we chose in half and at least one of those halves will not have a finite subcover. We can continue this process on and on and we will label the nth interval $[a_n, b_n]$. The length of the nth interval is $\frac{1}{2^n}(b-a)$ as at each step the length is halved. It is clear that $a_n \leq a_{n+1} < b_{n+1} \leq b_n \leq b$ at every step, so that the set $A = \{a_n | n \in \mathbb{N}\}$ is a bounded set of real numbers so let $\alpha = \sup A$. Then we claim that $\alpha \in [a_n, b_n]$ for all n. It is clear that $a_n \leq \alpha$ for all n. Let b_n be fixed, and let a_m be arbitrary, then we have that $a_m \leq a_{m+n} \leq b_{m+n} \leq b_n$ so that b_n is an upper bound for the set of a_n and so $\alpha \leq b_n$ and since this holds for all b_n we see that for all $n \in \mathbb{N}$ $a_n \leq \alpha \leq b_n$ and so $\alpha \in [a_n, b_n]$ for all n. But then the intersection $\bigcap_{n=1}^{\infty} [a_n, b_n]$ is nonempty. Since $\mathcal{O}$ is an open cover of $[a, b]$, we see that there is some open set O containing α. Since O is open there is some $r > 0$ with $B_r(\alpha) \subseteq O$, but then if we can choose an $n \in \mathbb{N}$ so that $\frac{b-a}{2^n} < r$ and we see that for this n interval $[a_n, b_n] \subseteq B_r(\alpha) \subseteq O$. But this is a contradiction as we claimed this interval had no finite subcover! So we see that $[a, b]$ must be compact! □

Corollary 2.53.1. Every n-box is compact. This follows from above theorem and the theorem that the product of compact sets is compact as an n-box equals $[a_1, b_1] \times ... \times [a_n, b_n]$.

Theorem 2.54 (The Heine-Borel Theorem)**.** In $\mathbb{R}^n$ a set is compact if and only if it is closed and bounded.

Proof. We have already seen that in any metric space compactness implies being closed and bounded. Now assume that A is a closed and bounded subset of $\mathbb{R}^n$. Then there is some n-box K containing A. But then since K is compact, and A is a closed subset of a compact set it follows that A is also compact! □

We now give a useful theorem describing which subsets of $\mathbb{R}$ are connected.

Theorem 2.55. We have that a set A is a connected subset of $\mathbb{R}$ if and only if $a \in A$, $b \in A$, and $a < c < b$ implies that $c \in A$.

Proof. Suppose A is connected and $c \notin A$, then consider the sets $S = (-\infty, c) \cap A$ and $W = (c, \infty) \cap A$. We see $A = S \cup W$ and clearly S is not empty as $a \in S$ and likewise $b \in W$. We want to show that $\overline{S} \cap W = \emptyset$. It isn't difficult to see that $\overline{S} \subseteq (-\infty, c]$ and since $(-\infty, c] \cap W = \emptyset$ so does $\overline{S} \cap W$. Similarly we see that $S \cap \overline{W}$ is empty. Since S and W are separated sets whose union equals A, A is not connected contradicting our assumption! This proves the first direction.

Now assume we have a set A with the property that if $a, b \in A$ with $a < c < b$, then $c \in A$ and for contradiction assume that A is not connected. Then there exist nonempty separated sets M, N with $M \cup N = A$. Let $m \in M$ and $n \in N$ and without loss of generality assume that $m < n$ (if not rename). Let $\alpha = \sup M \cap [m, n]$ we know this set is bounded above by n so α exists. Then $\alpha \in \overline{M}$ as it is a closed bounded (therefore compact) set by Theorem 2.XXX, and since M and N are separated $\alpha \notin N$, but by our property $m < \alpha < n$ so $\alpha \in A$ and therefore must be in M. But then $\alpha \notin \overline{N}$ so $\alpha \neq \inf N \cap [\alpha, n] = \beta$ (if it was the inf it would need to be in $\overline{N}$) so we can choose some p with $\alpha < p < \beta$ with $p \notin M$ and $p \notin N$ so $p \notin A$, but $m < p < n$ contradicting the fact that $p \in A$ by our assumption. $\square$

Corollary 2.55.1. As a corollary to this we see that the only connected subsets of $\mathbb{R}$ are intervals as they are the sets satisfying the property given above!

We end this section by talking about sequences as well as touching on the extended real numbers. We mention the extended real numbers as sequences can "diverge" to infinity, in which case if we look at the sequence from the perspective of the extended reals they would indeed converge in this larger picture. We will be careful to state when one should focus on the extended reals. We first show an unexpected result concerning $\overline{\mathbb{R}}$.

Theorem 2.56. The extended real numbers $\overline{\mathbb{R}}$ is compact with the topology given earlier in the chapter.

Proof. Let $\mathcal{O}$ be an open cover of $\overline{\mathbb{R}}$. Then there is some set of the form $[-\infty, a)$ and some set of the form $(b, \infty]$ (these could be closed interval's it does not change the proof). Then choosing these sets we see that the interval $[a-1, b+1]$ is compact and $\mathcal{O}$ is an open cover for this set as well so that there is a finite collection $O_1, ..., O_n$ that covers $[a-1, b-]$. But then our collection $O_1, ..., O_n, [-\infty, a), (b, \infty]$ is a finite subcover of $\overline{\mathbb{R}}$ completing the proof. $\square$

Theorem 2.57 (Bolzano-Weierstrass Theorem)**.** Let a_n be a sequence of real numbers. Then a_n has a convergent subsequence in $\overline{\mathbb{R}}$. If the sequence is bounded it has a convergent subsequence in $\mathbb{R}$.

Proof. The theorem follows almost immediately from compactness. Since $\overline{\mathbb{R}}$ is compact it is limit point compact giving the first result. If the sequence is bounded then there is some interval $[a, b]$ containing the sequence. Since $[a, b]$ is compact the full theorem follows. $\square$

Definition 2.58. We say that a sequence of real numbers a_n is **Monotonically Increasing** if $a_n \leq a_{n+1}$ for all n. Similarly, a sequence is **Monotonically Decreasing** if $a_n \geq a_{n+1}$ for all n. If the sequence is either monotonically increasing or decreasing we say the sequence is **Montonic**.

Theorem 2.59 (Monotone Convergence Theorem)**.** If a_n is a monotone sequence of real numbers then it converges in $\overline{\mathbb{R}}$. The sequence converges in $\mathbb{R}$ if and only if the sequence is bounded.

Proof. Assume the sequence a_n is montonically increasing, the case for decreasing is the same proof with swapped inequalities. If the sequence is unbounded then choose any neighborhood of infinity $N(\infty) = (b, \infty]$. Since the sequence is unbounded there is an $n \in \mathbb{N}$ so that $a_n > b$ and sicne the sequence is montonically increasing it follows that for all $m \geq n$ $a_m \geq a_n > b$ so that for all $m \geq n$ $a_m \in (b, \infty]$. Since this was an arbitrary neighborhood it follows that the sequence converges to ∞ (in $\overline{\mathbb{R}}$), this limit is not in $\mathbb{R}$ so clearly the sequence diverges if looking in $\mathbb{R}$. If instead the sequence is bounded, then let $\alpha = \sup\{a_n | n \in \mathbb{N}\}$ which is in $\mathbb{R}$. Then for any $\epsilon > 0$ since $\alpha - \epsilon$ is not an upper bound for the terms in the sequence it follows there is some $N \in \mathbb{N}$ so that $\alpha - \epsilon < a_N \leq \alpha$ and so for all $n \geq N$ we have that $|a_n - \alpha| < \epsilon$ so that the sequence converges to $\alpha \in \mathbb{R}$. $\square$

Chapter 3

Continuous Functions

Having spent a great deal of time on topology we now turn towards our first familiar topic: Continuous functions! However, unlike in Calculus we will be looking at these functions from the very general topological point of view as it turns out that continuity is inherently a topological concept. To see this consider what it means for a function to be continuous. When you first learn about these types of functions they are described as being drawn without picking up a pencil. In Calculus you learn instead that a function is continuous if the limit of the function at a point equals the value at said point. What this means is that as you move the input closer to some point, the values of the function also get closer to said point. But measuring closeness is something topology does very well! We now move on to definitions and explain how this agrees with this intuition we've painted.

3.1 Limits

Definition 3.1. Let X, Y be topological spaces with $A \subseteq X$. Let $f : A \to Y$ and let p be a limit point of X and let $L \in Y$. Then we say that

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

if for every neighborhood $N_Y(L) \subseteq Y$ there exists a neighborhood $N_X(p) \subseteq X$ so that $f(N_X(p) - \{p\}) \subseteq N_Y(L)$.

Essentially, we are saying if we measure closeness with open sets of L that we can find a close set around p that stays close to the limit after going through the function. Note the value p is excluded from the functional values as the actual value of the function at that point plays no role in determining the limit.

If we instead have metric spaces (X, d_X) and (Y, d_Y) with $A \subseteq X$, p a limit point of A, and $L \in Y$ then we have the definition that

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

if for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $0 < d_X(x, p) < \delta$ then $d_Y(f(x), L) < \epsilon$.

So far the functional limit has nothing to do with the limit of sequences, but the next theorem will relate these two ideas for first countable spaces.

Theorem 3.2. Suppose X, Y are first countable spaces, $A \subseteq X$, and $f : A \to Y$. Then we have that $\lim_{x \to p} f(x) = L$ if and only if for every sequence a_n where $\lim_{n \to \infty} a_n = p$ and $a_n \neq p$ for all $n = 1, 2, ...$, we have that $\lim_{n \to \infty} f(a_n) = L$.

Proof. First suppose that $\lim_{x \to p} f(x) = L$ and let a_n be a sequence satisfying the above. We need to show that $\lim_{n \to \infty} f(a_n) = L$. Let $N_Y(L)$ be any neighborhood of L, then since $\lim_{x \to p} f(x) = L$ there exists some neighborhood $N_X(p)$ so that $f(N_X(p) - \{p\}) \subseteq N_Y(L)$. But since $a_n \to p$ as $n \to \infty$ there is some $N \in \mathbb{N}$ so that for all $n \geq N$ we have $a_n \in N_X(p)$ and since $a_n \neq p$ for all n, this above with the previous comment tells us that $f(a_n) \in f(N_X(p) - \{p\}) \subseteq N_Y(L)$. Since $N_Y(L)$ was arbitrary this holds for any neighborhood and so we see that $\lim_{n \to \infty} f(a_n) = L$.

Now to prove the other direction we prove the converse, that is assume that $\lim_{x \to p} f(x) \neq L$, and we must show that there is some sequence a_n with $a_n \to p$ as $n \to \infty$ ($a_n \neq p$), but where

$\lim_{n\to\infty} f(a_n) \neq L$. By first countability p has a countable local basis $\mathcal{B}$. Index the sets in order of inclusion $B_1, B_2, ...$ with $B_{n+1} \subseteq B_n$ as was done in the proof of theorem 2.XXX. Since $\lim_{x\to p} f(x) \neq L$ there exists some neighborhood $N_Y(L)$ so that for every neighborhood $N_X(p)$ we have $f(N_X(p) - \{p\}) \nsubseteq N_Y(L)$, if no such $N_Y(L)$ existed then L would be the functional limit. Certainly B_1 is a neighborhood of p and by the previous comment there is some element $x \in B_1$ so that $f(x) \notin N_Y(L)$ and $x \neq p$. We will create a new sequence by first setting $a_1 = x$. Inductively continue this process setting a_n as an element of $x' \in B_n$ satisfying that $f(x') \notin N_Y(L)$ and $x' \neq p$. Then the sequence a_n converges to p, $a_n \neq p$ for all n, and the sequence $f(a_n)$ certainly does not converge to L as $N_Y(L)$ contains no terms of the sequence. This completes the proof. $\square$

Theorem 3.3. If $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}$ where X is first countable and $\lim_{x\to p} f(x) = A$ and $\lim_{x\to p} g(x) = B$, then

(a) $\lim_{x\to p}(f + g)(x) = A + B$.

(b) $\lim_{x\to p}(f \cdot g)(x) = A \cdot B$,

(c) $\lim_{x\to p}(\frac{f}{g})(x) = \frac{A}{B}$, provided $B \neq 0$.

This theorem follows immediately from the preceding theorem as well as the corresponding theorem on limits of sequences.

3.2 Continuous Functions

We can now discuss continuous functions having proven a few useful facts about limits.

Definition 3.4. Let X, Y be topological spaces and $f : X \to Y$ with $p \in X$. Then we say that f is **Continuous at the point p** if for every neighborhood $N_Y(f(p))$ there exists a neighborhood $N_X(p)$ such that $f(N_X(p)) \subseteq N_Y(f(p))$.

For metric spaces this definition can be stated as f is continuous at p if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $d_X(x, p) < \delta$ implies that $d_Y(f(x), f(p)) < \epsilon$. If f is continuous at every point of X we say that f is **Continuous on X** or more simply f is **Continuous**. If p is a limit point, then we immediately have that f is continuous at p if and only if $\lim_{x \to p} f(x) = f(p)$. We will now handle the case p is not a limit point.

Definition 3.5. Let X be a topological space and let $p \in X$. If p is not a limit point of X, then we say that p is **An Isolated Point** of X. This implies that there is a neighborhood $N(p)$ containing no other points of X, excluding the point p itself, but this immediately tells us that p is an isolated point of X if and only if $\{p\}$ is an open set.

Theorem 3.6. If $f : X \to Y$ and p is an isolated point of X, then f is always continuous at p.

Proof. Let $N(f(p))$ be any neighborhood of $f(p)$. Since p is an isolated point $\{p\}$ is a neighborhood of p and so clearly $f(\{p\}) = \{f(p)\} \subseteq N(f(p))$. $\square$

As a result of this theorem as well as our comments about limit points we have that Theorem 3.2 applies to continuous functions ie a function is continuous at a point p if and only if for sequences that converge to p, a_n have the property that $f(a_n) \to f(p)$. Similarly, Theorem 3.3 also applies, so that all functions involved are continuous if f, g are continuous.
Our next theorem gives a very useful criterion for when a function is continuous based on open sets.

Theorem 3.66 (Topological Continuity) Let X, Y topological spaces. Then f is continuous (everywhere on X) if and only if given an open subset $V \subseteq Y$ we have that the inverse image $f^{-1}(V) \subseteq X$ is an open subset of X.

Proof. Assume f is continuous and let V be an open set in Y. Now let $x \in f^{-1}(V)$. Since f is continuous at x there is some neighborhood $N(X)$ so that $f(N(X)) \subseteq V$, since V is a neighborhood of $f(x)$. But then x is interior to $f^{-1}(V)$ and since x was arbitrary we see that $f^{-1}(V)$ is open.
Now assume that for every open set V in Y we have that $f^{-1}(V)$ is open in X. Now let $x \in X$,

we need to show that f is continuous at x. So let $f(x) = L$ and let $N(L)$ be a neighborhood of $f(x) = L$. Then we know that $f^{-1}(N(L))$ is an open subset of X containing x and it is therefore a neighborhood of x. Since $f(f^{-1}(N(L))) \subseteq N(L)$ we are done. $\square$

We now see that continuity is very closely related to topology. We have shown that a function is continuous if the inverse image of an open set is an open set and so we have a way of discussing continuity using only open sets. It is also easy to show the above theorem holds if we use closed sets instead of open sets. To end this section we give a nice theorem that is proven very easily as a result of the Topological Continuity Theorem.

Theorem 3.8. If X, Y, Z are topological spaces, $f : X \to Y$ is continuous, and $g : Y \to Z$ is continuous, then $(g \circ f) : X \to Z$ is continuous.

Proof. Let $V \subseteq Z$ be open. Then $g^{-1}(V)$ is an open subset of Y by continuity of g, and $f^{-1}(g^{-1}(V))$ is an open subset of X by continuity of f, this completes the theorem as $(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V))$. $\square$

3.3 Continuity and Topology

This section will show that continuity interacts very nicely with several of the topological properties we previously defined in chapter two.

Definition 3.9. Let X, Y be topological spaces. Then if $f : X \to Y$ is a bijective continuous function and f^{-1} is also continuous, then we say that f is a **Homeomorphism**, and say X and Y are **Homeomorphic**. As a result of the last theorem on continuous functions we see that a set $V \subseteq X$ is open in X if and only if $f(V)$ is open in Y. In other words homeomorphic spaces look the same in terms of open sets, or are topologically the same! This is why mathematicians say that if a shape can be continuously morphed into another and then continuously morphed back, they are topologically the same. See the classic example of a coffee cup and a donut! We could also say that f is a homeomorphism if it is a bijective continuous open mapping. To continue along with this properties that are preserved under homeomorphism are called **Topological Properties**. In this section we will prove that the properties we discussed in Chapter 2 are indeed Topological Properties and we will find that a couple familiar results follow!

The first properties we will discuss are the Countability and Hausdorff Properties. The property of Second Countability is quite easy to see and so no proof is given

Theorem 3.10. If X, Y are homeomorphic and X is first countable, then Y is first countable.

Proof. We need to show that every point of Y has a countable local basis. Let $p \in Y$ and let $f : X \to Y$ our homeomorphism. Let $x \in X$ be such that $f(x) = y$ and let $\mathcal{B}_x$ be a countable local basis for x. Then the claim is that $f(\mathcal{B}_x)$ is a local basis for y. For if V is any neighborhood of y, then $f^{-1}(V)$ is an open neighborhood of x, so there is a basis element B_i with $x \in B_i \subseteq f^{-1}(V)$, but then $y \in f(B_i) \subseteq V$ so that our collection is indeed a countable local basis! □

Theorem 3.11. If X, Y are homeomorphic and X is second countable, then Y is second countable.

Theorem 3.12. If X, Y are homeomorphic and X is Hausdorff, then Y is also Hausdorff.

Proof. Let $p, q \in Y$ with $p \neq q$. Then choose $x, y \in X$ so that $f(x) = p$ and $f(y) = q$. Since X is Hausdorff there exists neighborhoods $N(x), N(y)$ with $N(x) \cap N(y) = \emptyset$. But then if f is our homeomorphism, since f is bijective we have that $f(N(x)) \cap f(N(y)) = \emptyset$ and $f(N(x))$ is a neighborhood of $f(x) = p$ and $f(N(y))$ is a neighborhood of $f(y) = q$, so Y is Hausdorff. □

We now turn towards the two more important properties: Compactness and Connectedness. We will see that we only need a continuous function in these cases and in fact these properties will give easy proofs of some very famous results!

Theorem 3.72 (Compactness and Continuity) Let X, Y be topological spaces. If X is compact and $f : X \to Y$ is continuous, then $f(X) \subseteq Y$ is compact.

Proof. Let $\{O_\alpha | \alpha \in A\}$ be an open cover of $f(X)$. Then it's easy to see the sets $\{f^{-1}(O_\alpha) | \alpha \in A\}$ is an open cover of X, so by compactness a finite subcover exists. Let $f^{-1}(O_1), ..., f^{-1}(O_n)$ be our finite subcover. Then it follows that since $f(f^{-1}(O_i)) \subseteq O_i$, that the collection $O_1, ..., O_n$ is a finite subcover of $\{O_\alpha\}$ and so $f(X)$ is compact. $\square$

Theorem 3.73 (Extreme Value Theorem) Let $f : [a, b] \to \mathbb{R}$, then there exists $x, y \in [a, b]$ so that for all $c \in [a, b]$ $f(y) \leq f(c) \leq f(x)$. In short f attains a maximum and a minimum.

Proof. This famous theorem from Calculus is essentially a corollary of the Compactness and Continuity theorem. Because $[a, b]$ is compact it follows that $f([a, b])$ is a compact subset of $\mathbb{R}$ and so by Theorem 2.XXX it achieves a maximum and a minimum! $\square$

Theorem 3.15. Let X, Y be Topological Spaces with X compact and Y Hausdorff. Let $f : X \to Y$ be a bijective continuous function. Then f is a homeomorphism.

Proof. We essentially need to prove that f is also an open mapping. Let $U \subseteq X$ be open. Then U^c is compact as it is a closed subset of X. It follows that $f(U^c)$ is compact by Compactness and Continuity Theorem, but since Y is Hausdorff this implies that $f(U^c)$ is closed. But by bijectivity we see that $f(U^c)^c = f(U)$ so that $f(U)$ is open and so f is an open mapping. $\square$

Now we discuss connectedness.

Theorem 3.75 (Conectedness and Continuity) Let X, Y be topological spaces, with X connected. Then $f(X) \subseteq Y$ is connected.

Proof. As is the case with many connectivity proofs we assume that $f(X)$ is not connected. Then there are nonempty separated sets A, B with $f(X) = A \cup B$. Let $M = f^{-1}(A)$ and

$N = f^{-1}(B)$, these are both nonempty sets and $M \cup N = X$. Since $\overline{A}$ is closed by continuity of f we see that $f^{-1}(\overline{A})$ is closed and we see that $M \subseteq f^{-1}(\overline{A})$. Since the closure of M is the smallest closed set containing M and $f^{-1}(\overline{A})$ is a closed set containing M we get that $\overline{M} \subseteq f^{-1}(\overline{A})$. But from here we can see that $\overline{M} \cap N = \emptyset$ as if there was some point p in both it would follow that $f(p) \in \overline{A}$ and B. A similar argument shows that $M \cap \overline{N} = \emptyset$ so we see that M, N are separated sets whose union is X so that X is not connected, a contradiction. $\square$

Theorem 3.76 (Intermediate Value Theorem) If $f : [a, b] \to \mathbb{R}$ is continuous and if $f(a) \leq c \leq f(b)$, then there exists a $d \in [a, b]$ so that $f(d) = c$.

Proof. Another famous theorem from Calculus follows very easily from our labors in Topology! Since $[a, b]$ is connected by theorem 2.XXX, it follows $f([a, b])$ is a connected subset of $\mathbb{R}$ so by theorem 2.XXX since $f(a) \leq c \leq f(b)$ it follows $c \in f([a, b])$ so that there is some $d \in [a, b]$ with $f(d) = c$. We can further strengthen this theorem. By the extreme value theorem f attains a maximum and a minimum M and m respectively. By the exact same logic we just used if $m \leq c \leq M$ there is some $d \in [a, b]$ so that $f(d) = c$. $\square$

To end this section we introduce a stronger notion of connectedness that will be useful in showing that many sets we expect to be connected are so.

Definition 3.18. Let X be a topological space. We say that a **Path** is a continuous function $f : [0, 1] \to X$. We can easily replace $[0, 1]$ with a general $[a, b]$ as there is a continuous function from $[0, 1] \to [a, b]$ namely $f(x) = a + bx$ and the composition of continuous functions is continuous. We say that X is **Path Connected** if given any two points $x, y \in X$ there is a path f with $f(0) = x$, $f(1) = y$, and $f([0, 1]) \subseteq X$. A space is path connected if we can find a path between two points that stays in that space. We easily see from how Convex sets were defined that they are path connected. We now want to show that path connected spaces are connected.

Theorem 3.19. Let X be a path connected space. Then X is connected.

Proof. We as usual assume X is not connected. Then there is a separation $A \cup B = X$. Let $x \in A$ and $y \in B$. Since X is path connected there exists a path $f : [0,1] \to X$ with $f(0) = x$ and $f(1) = y$. Let $M = f^{-1}(A)$ and $N = f^{-1}(B)$. Since $f([0,1]) \subseteq X$ and $X = A \cup B$ we see that $M \cup N = [0,1]$. Using the same argument as in the proof of The Connectedness and Continuity Theorem we see that $\overline{M} \subseteq f^{-1}(A)$ and see that $\overline{M} \cap N = \emptyset$ and we see $M \cap \overline{N} = \emptyset$, so that M, N are nonempty separated sets whose union gives $[0,1]$. However, since we already know that $[0,1]$ is connected we reach a contradiction and we see that X must be connected. $\square$

3.4 Uniform Continuity

Having shown that our Topological Properties were indeed what we said they were, and showing how they nicely interact with Continuity, we now restrict ourselves to a topic exclusive to Metric Spaces. That topic is Uniform Continuity.

Definition 3.20. Let X, Y be metric spaces and let $f : X \to Y$. We say that f is **Uniformly Continuous** if for all $\epsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in X$ if $d_X(x, y) < \delta$ then $d_Y(f(x), f(y)) < \epsilon$. This should look very similar to the definition of continuity. The key difference is that for continuity the value of δ depended on ϵ and the point we were checking for continuity. This definition means for any given ϵ the same δ will work for all points in the space! It is immediately obvious that all Uniformly Continuous functions are continuous, but the converse is not true. However, our next theorem will show that this is true if we assume X is compact!

Theorem 3.21. Let X, Y be metric spaces with X compact. If $f : X \to Y$ is continuous, then f is uniformly continuous.

Proof. Let $\epsilon > 0$, we must find a δ that works for all $x, y \in X$. For each $x \in X$ we know there is some $\delta > 0$ so that for all $y \in X$ we have $d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \frac{\epsilon}{2}$. Denote for each x its corresponding δ as δ_x and then $\rho_x = \frac{\delta_x}{2}$. Consider the set of $B_{\rho_x}(x)$. This set is clearly an open cover for X and so by continuity there is a finite subcover: $B_{\rho_{x_1}}(x_1), ..., B_{\rho_{x_n}}(x_n)$. Consider the set of $\{\rho_{x_1}, ..., \rho_{x_n}\}$, this is a finite set so a minimum ρ_{x_i} exists. Then the claim is that $\delta' = \frac{\rho_{x_i}}{2}$ will do the trick. To see this let $x, y \in X$ with $d_X(x, y) < \delta'$. Then $x \in B_{\rho_{x_i}}(x_i)$ for some $1 \leq i \leq n$ so that $d_X(x, x_i) < \rho_{x_i}$ and so $d_X(x_i, y) \leq d_X(x_i, x) + d_X(x, y) < \rho_{x_i} + \delta' < 2\rho_{x_i} = \delta_{x_i}$. It then follows that $d_Y(f(x), f(y)) \leq d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ completing the proof. $\square$

The last proof was quite obnoxious as it required dividing delta's and epsilons by $\frac{1}{2}$ constantly. The reader may wonder how to figure this out on their own and there are two answers. Firstly, whenever you get a result with something like 2ϵ on one side that indicates that going and dividing by two earlier would likely fix the problem, in the last problems case this might have had to happen several times! The other solution is to say any multiple of ϵ is acceptable! As long as the multiple of ϵ has no dependency on δ this is equivalent and will frequently be used later to save time on problems!

Another comment about the last problem is that in practice we will often take uniform continuity for granted as we very frequently will consider continuous functions from the interval $[a, b]$, which will guarantee uniform continuity. That is why uniform continuity is so nice is guarantees $\delta' s$ that work regardless of what points we pick in a given interval and there will be a few times that this property is critical!

3.5 Discontinuous Functions

The last section of this chapter is on discontinuous functions. We will be focusing solely on real valued functions here as these are the most interesting types of discontinuities. We introduce some tools to help us explore these issues.

Definition 3.22. Let $f : (a, b) \to \mathbb{R}$, and let $p \in (a, b)$. We define the **Right Hand Limit** at p as

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

means that for all $\epsilon > 0$ there exists a $\delta > 0$ such that $0 < x - p < \delta$ implies that $|L - f(p)| < \epsilon$. Similarly, the **Left Hand Limit** $\lim_{x \to p^-} f(x) = L$ means that for all $\epsilon > 0$ there is a $\delta > 0$ such that $0 < p - x < \delta$ implies that $|L - f(p)| < \epsilon$. Intuitively, the right hand limit is whatever value the function gets close to as it takes values greater than p (or approaches from the right) and similarly for the left hand limit. We see that the regular limit exists if and only if both the left hand and right hand limits exist and are equal from a comparison of definitions.

We can now classify types of discontinuity based on whether or not the right and left hand limits exist.

Definition 3.23. Let $f : (a, b) \to \mathbb{R}$ and let $p \in (a, b)$. Suppose f is discontinuous at p. If $\lim_{x \to p^+} f(x)$ and $\lim_{x \to p^-} f(x)$ exist, then we say that f has a **Simple Discontinuity** at p. Otherwise, we say f has an **Essential Discontinuity** at p. Simple discontinuities are like jump discontinuities or removable discontinuities. Essentially, it is discontinuous because it either approaches two different values from the left and right, or they approach the same value but this value is not the value of the function at said point. Essential discontinuities are more complex, like asymptotic behavior or oscillatory discontinuities. The following theorem is quite interesting!

Theorem 3.24. Let $f : (a, b) \to \mathbb{R}$. Then f has at most a countable number of simple discontinuities.

Proof. For the first case consider if $\alpha = \lim_{x \to p^-} f(x) < \lim_{x \to p^+} f(x) = \beta$. Let D be the set of all points in this category and let $p \in D$. In this case we will assign an ordered triple of rationals to this point p. For the first value choose some rational number q so that $\alpha < q < \beta$, this exists as the rationals are dense in $\mathbb{R}$. Next choose a rational number r such that $a < r < p$ so that if $r < t < p$ we have that $f(t) < q$. This can be done as by definition of left limit taking $\epsilon = q - \alpha$ we can pick δ so that $p - r < \delta$ gives the result. Similarly, we choose our third rational number s so that $p < s < b$ and for $p < t < s$ we have $f(t) > q$. We thus have an ordered triple (q, r, s) of rational numbers. We now must show two distinct elements of D cannot have the same ordered triple. So let $w, v \in D$ with $w < v$. Let (q, r, s) be the ordered triple for v. If $\lim_{x \to w^-} f(x) < q < \lim_{x \to w^+} f(x)$ is not true clearly w has a different ordered triple, so assume that the previous statement is true. But now they cannot have the same second and third elements of the triple. For if $r < w < v$ we can choose a t with $r < w < t < v < s$ so that $f(t) < q$ because $r < t < v$, but at the same time $f(t) > q$ as $w < t < s$. Since each ordered triple can correspond to at most one point of D and each point of D corresponds to an ordered triple, and the set of ordered triple's of rationals are countable, we see that D is countable! The exact same argument handles the case where D is the set of p with $\lim_{x \to p^-} f(x) > \lim_{x \to p^+} f(x)$.

Our last case for simple discontinuities is if $\lim_{x \to p^-} f(x) = \lim_{x \to p^+} f(x)$. In this case we see that $\lim_{x \to p} f(x)$ exists, but is distinct from the function value $f(p)$. So in this case we assume $\lim_{x \to p} f(x) < f(p)$ and choose q rational so that $\lim_{x \to p} f(x) < q < f(p)$ and choose r, s as before getting (q, r, s). Then again the same argument gives our result and the case where $\lim_{x \to p} f(x) > f(p)$ works the same. Since all four of these cases we found the resulting sets were countable, the union of all four countable sets is again countable and so we have our result! $\square$

Essential discontinuities have no such restrictions! For consider the function

$$f(x) = \{0, x \in \mathbb{Q}, 1, x \in \mathbb{I}\}$$

then it can be seen that x is discontinuous at every point and all of these discontinuities are essential!

We end this section and this chapter with the concept of monotonic functions and a neat fact about these functions. We've already discussed monotonic series and the functions are very similar in concept.

Definition 3.25. Let $f : \mathbb{R} \to \mathbb{R}$. We say that f is **Monotonically Increasing** if $x < y$ implies that $f(x) \leq f(y)$. Similarly f is **Monotonically Decreasing** if $x < y$ implies that $f(x) \geq f(y)$. If f is either monotonically increasing or decreasing we say f is **Monotonic**. The next theorem and its corollary highlight some of the nice qualities of monotonic functions.

Theorem 3.26. Let $f : (a, b) \to \mathbb{R}$ be monotonically increasing. Then for all $p \in (a, b)$ we have that $\lim_{x\to p^-} f(x)$ and $\lim_{x\to p^+} f(x)$ exist. The same holds for monotonically decreasing functions.

Proof. We claim that for $p \in (a, b)$ that $\lim_{x\to p^-} f(x) = \sup\{f(x)|x < p\}$ and that $\lim_{x\to p^+} f(x) = \inf\{f(x)|x > p\}$. We prove the first one equality, the second is the same argument. Let $\alpha = \sup\{f(x)|x < p\}$. Let $\epsilon > 0$. By definition of sup we can find some $y < p$ so that $\alpha - \epsilon < f(y) < \alpha$. It follows then that if we take $\delta = p - y$ we have that for all t satisfying $0 < p - t < \delta$ we have that $y < t$ so that $\alpha - \epsilon < f(y) \leq f(t) < \alpha$ so that $0 < p - t < \delta$ implies that $|f(t) - \alpha| < \epsilon$ and so by definition we see that $\lim_{x\to p^-} f(x) = \alpha$. $\square$

Corollary 3.26.1. It is clear that a monotonic function can only have simple discontinuities as the left and right hand limits exist at every point. And from this we see that a monotonic function can have at most a countable number of discontinuities.

Chapter 4

Series of Complex Numbers

This text can be seen as consisting of four parts. Part One deals with Topology and the structure of the Real Numbers. In Part Two we will develop the familiar concepts from Calculus: Series, Derivatives, and Integrals. Part Three will deal with Sequences and Series of Functions and finally Part Four briefly covers several further branches of Analysis. With the completion of Chapter Three we have finished part one and now turn to more familiar topics! The concept of Series are very familiar as any Calculus Two course likely spent a great deal of time on these topics! It is quite annoying trying to find a good place to fit this topic in however. We will see that our work on sequences will be used constantly to prove the familiar results about series and so putting this chapter right after chapter two would be valid. However, Continuity is essentially a continuation of all the ideas of Topology found in chapter two so putting that as chapter three seems more fitting, and we postponed our discussion of series until now. With that being said a brief review of sequences and the topology of $\mathbf{R}$ before preceeding is advised!

4.1 Introduction to Series

Intuitively, we think of a series as a sum of an infinite number of terms. This is clearly impossible! Addition is a binary operation and we can never sum an infinite collection of numbers as we can only add two numbers at a time. However, what we can do is use the idea of sequences and convergence to estimate sums for as many of the numbers as we'd like. Using this we can make a sequence of "Partial Sums" and ask what this sequence converges to! We make this more formal as follows.

Definition 4.1. Given a sequence $\{a_n\}$ of complex numbers, we define the **N-th Partial Sum** of the sequence $\{a_n\}$ as

$$S_n = \sum_{k=1}^{N} a_k = a_1 + ... + a_N.$$

We now have two sequences to work with the original sequence $\{a_n\}$ and the sequence of partial sums $\{S_n\}$. For our original sequence $\{a_n\}$ we define **The Series** of the terms $\{a_n\}$ with the following:

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k = \lim_{n\to\infty} S_n.$$

If the limit exists (in $\mathbb{R}$) we say the series **Converges** and otherwise it **Diverges**. Note we could chose to start at any index not just $n = 1$ and frequently we will choose $n = 0$.
If $\sum_{n=1}^{\infty} |a_n|$ converges we say the series **Converges Absolutely**. If $\sum_{n=1}^{\infty} a_n$ converges, but not absolutely we say it **Conditionally Converges**.

We now use the fact that $\mathbb{C}$ is Complete to get a very important theorem.

Theorem 4.87 (Cauchy Criterion) A series $\sum_{n=1}^{\infty} a_n$ converges if and only if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $m, n \geq N$ with $m > n$ we have $|\sum_{k=n}^{m} a_k| < \epsilon$.

Proof. If the series converges, then by definition the sequence $\{S_n\}$ is Cauchy so that for all $\epsilon > 0$ there exists an N so that for all $m \geq n \geq N$ we have $|S_m - S_n| < \epsilon$, but this is the same as $|\sum_{k=n}^{m} a_k| < \epsilon$.

The other direction is easy as well as the condition implies that S_n is Cauchy and the completeness of $\mathbb{C}$ means that S_n converges. $\square$

Theorem 4.88 (Divergence Test) If $\lim_{n\to\infty} a_n \neq 0$, then $\sum_{n=1}^{\infty} a_n$ diverges.

Proof. We prove contra-positive that is if $\sum_{n=1}^{\infty} a_n$ converges then $a_n \to 0$. Suppose the series converges. Then by Cauchy Criterion we have for all $\epsilon > 0$ there exists N so that $m \geq n \geq N$ implies $|\sum_{k=n}^{m} a_K| < \epsilon$. Let $\epsilon > 0$ and choose N using previous. Then for all $n \geq N$ we have (choosing $m = n+1$) $|\sum_{k=n}^{n+1} a_k| < \epsilon$ but this means $a_{n+1} < \epsilon$ so that the sequence converges to 0. $\square$

4.2 The Comparison Test

We will see that many times to analyze the convergence of complex series, we need only look at the series of absolute values. For we would hope that if the absolutely converges then it converges in the normal sense too. The next two theorems discuss this.

Theorem 4.4. Let a_n be a sequence of Complex Numbers. If a_n converges to L then $|a_n|$ converges to $|L|$.

Proof. Let $\epsilon > 0$. Then since $a_n \to L$ there exists an N so that for all $n \geq N$ we have $|a_n - L| < \epsilon$. But by the reverse triangle inequality we have then that $||a_n| - |L|| < |a_n - L| < \epsilon$ completing the short proof. $\square$

Theorem 4.5. If $\sum |a_n|$ conveges, then so does $\sum a_n$.

Proof. This follows from the Cauchy Criterion and the triangle inequality for $|\sum_{k=n}^{m} a_k| \leq |\sum_{k=n}^{m} |a_k|| < \epsilon$. $\square$

We now give a very familiar and useful test: The Comparison Test.

Theorem 4.91 (The Comparison Test) Suppose a_n is a sequence of complex numbers and b_n, c_n are sequences of nonnegative real numbers.
(a) If there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ $|a_n| \leq b_n$ and $\sum b_n$ converges, then $\sum a_n$ converges absolutely.
(b) If there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ $b_n \leq c_n$ and $\sum b_n$ diverges, then $\sum c_n$ diverges.

Proof. We first prove (a) using Cauchy Criterion. Given $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for all $m \geq n \geq N$ we have $|\sum_{k=n}^{n} c_k| < \epsilon$. But as long as N is large enough to satisfy part (a) we have then that $|\sum_{k=n}^{m} |a_k|| \leq |\sum_{k=n}^{m} c_n| < \epsilon$ so that $\sum |a_n|$ converges by the Cauchy Criterion.
For part (c) it follows easily from part (a) for if $\sum b_n$ converged, part (a) would imply $\sum b_n$ converges contradicting the fact it converges. $\square$

The Comparison Test is our most useful test for analyzing the Converge of Complex Series as it essentially turns it into analyzing the convergence of a series of nonnegative numbers for which we will develop many familiar tests in the next few sections. We also introduce the famous Limit Comparison Test.

Theorem 4.92 (Limit Comparison Test) Suppose a_n is a sequence of nonnegative real numbers, and b_n is a sequence of positive real numbers ($b_n > 0$). If

$$\lim_{n \to \infty} \frac{a_n}{b_n} = c,$$

where $0 < c < \infty$, then either both series converge or diverge.

Proof. By definition of the sequential limit we have given $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $c - \epsilon < \frac{a_n}{b_n} < c + \epsilon$, and multiplying by b_n we have $(c - \epsilon)b_n < a_n < (c + \epsilon)b_n$. We see that the convergence of each series hence depends on the convergence of the other because multiplying by $(c - \epsilon)$ or $(c + \epsilon)$ has no impact on convergence. $\square$

4.3 Geometric Series

We move on to one very important type of series: Geometric Series. A geometric series is one where the terms a_n are of the form $a_n = a_0 \cdot r^n$ for some a_0 and some $r \in \mathbb{C}$. What makes these series so easy to analyze is the fact we have an exact formula for their partial sums!

Theorem 4.8. If $a_n = a_0 \cdot r^n$ for some $a_0, r \in \mathbb{C}$, then we have the following formula for the partial sums

$$S_n = \sum_{k=0}^{n} = \frac{a_0(1 - r^{n+1})}{1 - r}.$$

Proof. We see by definition $S_n = a_0 + a_0 r + ... + a_0 r^n = a_0(1 + ... + r^n)$. Multiplying by $(1-r)$ we get $S_n(1-r) = a_0(1+...+r^n)(1-r) = a_0((1-r)+(r-r^2)+...+(r^n-r^{n+1})) = a_0(1-r^{n+1})$. We then divide everything by $(1 - r)$ to get $S_n = \frac{a_0(1-r^{n+1})}{1-r}$ as desired. □

Theorem 4.94 (Geometric Series Test) Suppose $\sum_{n=0}^{\infty} a_0 r^n$ is a geometric series. Then if $0 \leq r < 1$ we have the series converges and $\sum_{n=0}^{\infty} a_0 r^n = \frac{a_0}{1-r}$. If $r \geq 1$ the series diverges.

Proof. We know by the preceding theorem the partial sums are

$$S_n = \frac{a_0(1 - r^{n+1})}{1 - r} = \frac{a_0}{1 - r} - \frac{r^{n+1}}{1 - r}.$$

To prove the theorem we just need to analyze $\lim_{n\to\infty} r^n$. If $0 \leq r < 1$ then we want to show $\lim_{n\to\infty} r^n = 0$. We see that r^n is a decreasing sequence because $r^{n+1} \leq r^n$ and the sequence is bounded below by 0 as all terms are positive. Therefore the sequence does converge by the monotone convergence theorem. Let $lim_{n\to\infty} r^n = L$, then by our algebraic theorem of sequences we have $r \cdot L = r \cdot \lim_{n\to\infty} r^n = \lim_{n\to\infty} r^{n+1} = L$ so that $r \cdot L = L$ but this is certainly only true if $L = 0$. Since $\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{a_0(1-r^{n+1}}{(1-r)} = \frac{a_0}{1-r}$ we have finished the first case.

If $r \geq 1$ it is clear that the series diverges by the Divergence Test as $a_n = a_0 r^n$ which clearly does not converge to zero. $\square$

4.4 The Cauchy Condensation Test

One famous test that we haven't seen yet, and won't see for quite some time is the Integral Test. The reason we can't prove this test is that we haven't even defined what we mean by an Integral yet and we still are two chapters away from this task! Instead of waiting two chapters to prove the Integral Test we instead prove the Cauchy Condensation Test! This test was likely not seen in a Calculus Two course as the integral test has similar applications, but is easier to use. We now state and prove the test.

Theorem 4.95 (The Cauchy Condensation Test) Let a_n be a monotonically decreasing sequence of nonnegative real numbers. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series $\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges.

Proof. Let S_n be the n-th partial sums of the series $\sum_{n=1}^{\infty} a_n$. Let C_n be the n-th partial sums of the series $\sum_{k=0}^{\infty} 2^k a_{2^k}$. Then $S_n = a_1 + a_2 + ... + a_n$ and $C_n = a_1 + 2a_2 + 4a_4 + ... + 2^n a_{2^n}$. It follows then that if $n = 2^k$ for some k we have

$$S_n = a_1+a_2+...+a_n = a_1+(a_2+a_3)+(a_4+a_5+a_6+a_7)+...+a_n \leq a_1+2a_2+4a_4...2^n(a_n) = C_n$$

so that $S_n \leq C_n$. But at the same time

$$S_N = a_1 + a_2 + ... + a_n = a + 1 + (a_2 + a_3) + ... + a_n \geq \frac{a_1}{2} + a_2 + ...2^{n-1}a_{2^n} = \frac{C_n}{2}$$

so that $S_n \geq \frac{1}{2}C_n$ so that the convergence and divergence of the series are linked. $\square$

Theorem 4.96 (The P-Series Test) Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ for some nonnegative number p. If $p > 1$ the series converges otherwise it diverges.

Proof. The series converges if and only if the series $\sum_{k=0}^{\infty} 2^k \cdot \frac{1}{2^{kp}}$ converges. This is the same as $\sum_{k=0}^{\infty} 2^{(1-p)k}$ or $\sum_{k=0}^{\infty} (\frac{1}{2^{p-1}})^k$. This converges if and only if $\frac{1}{2^{p-1}} < 1$ or $p > 1$ completing the proof. $\square$

4.5 Limit Supremum and Infimum

We now take a short break from proving series convergence tests to discuss the topic of the limit supremum and infimum, often also called the upper and lower limits!

Definition 4.12. Let $\{a_n\}$ be a sequence of real numbers. We define the limit supremum, upper limit, or $\limsup$ of a_n as

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} (\sup\{a_m | m \geq n\}),$$

likewise we define the limit infimum as

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} (\inf\{a_m | m \geq n\}).$$

The limit on the right hand side of the above is taken to be the limit in $\overline{\mathbb{R}}$. It is easy to see that both $\limsup a_n$ and $\liminf a_n$ are real numbers if and only if the sequence is bounded in $\mathbb{R}$. Also if we define a sequence by $s_n = \sup\{a_m | m \geq n\}$ we see that s_n is a monotonically decreasing sequence so that $\limsup a_n = \lim_{n\to\infty} s_n$ always exists as an extended real number, and the same applies to the $\liminf a_n$. This follows from the monotone convergence theorem.

Our next theorem gives another classification for the $\limsup$ and $\liminf$.

Theorem 4.13. Let a_n be a sequence of real numbers and let A be the set of all extended real subsequential limits of $\{a_n\}$. Then we have that $\limsup a_n = \sup A$ and $\liminf a_n = \inf A$.

Proof. Firstly since $\overline{\mathbb{R}}$ is compact it is sequentially compact so A is nonempty and the sup and inf exist in $\overline{\mathbb{R}}$. Let $\alpha = \sup A$ we prove that $\limsup a_n = \sup A$, the case of the $\liminf$

is similar. In the case where $\alpha = \infty$ it is clear that $\limsup a_n$ must also be ∞ so they are equal. Now we consider the case α is finite. In this case we first construct a subsequence that converges to α. We will denote our subsequence by a_{m_k} to distinguish it from the other subsequences we will be using. By definition $\alpha - 1$ is not an upper bound for A so there is some subsequence of a_n, a_{n_k} such that $\alpha - 1 < \lim a_{n_k}$, and so there is some $N \in \mathbb{N}$ so that $\alpha - 1 < a_{n_N} < \alpha$, make the first term of our subsequence $a_{m_1} = a_{n_N}$. Now we inducitvely assume our subsequence is defined up to $a_{m_{k-1}}$. Since $\alpha - \frac{1}{k}$ is not an upper bound for A we find some subsequence a_{n_k} whose limit L satisfies $\alpha - \frac{1}{k} < L < \alpha$ and then we choose some N_0 so that $\alpha - \frac{1}{k} < a_{n_{N_0}} < \alpha$ and where $N_0 > m_{k-1}$. We then set $a_{n_k} = a_{n_{N_0}}$. The sequence we have constructed a_{n_k} converges to α. Since there is some subsequence that converges to α it is easy to see that $\lim_{n\to\infty}(\sup\{a_m | m \geq n\}) = \limsup a_n \geq \alpha$. Now we want to show there is some subsequence that converges to $\limsup a_n$. This is much easier however. Set $a_{n_1} = a_1$. Inductively assume we have chosen $a_{n_{k-1}}$, then consider $\beta = \sup\{a_m | m \geq n_{k-1}\}$. Since $\beta - \frac{1}{k}$ is not an upper bound for the previous set we select some $N \in \mathbb{N}$ $\beta - \frac{1}{k} < a_N$ and where $N > n_{k-1}$. If we set $a_{n_k} = a_N$ then our subsequence a_{n_k} converges to $\limsup a_n$ and so $\limsup a_n$ is a subsequential limit and since α is the supremum of all subsequential limits it follows that $\limsup a_n \leq \alpha$ completing the proof. □

The last theorem while quite ugly due to managing all the subsequences is very useful in giving another way of thinking about the lim sup and lim inf. The next section will put these concepts to good use in the ratio and root tests! We end this section with a quick coupletheorem about these concepts followed by a very easy proof of the squeeze theorem!

Theorem 4.14. If a_n and b_n are sequences of real numbers and there exists some N such that for all $n \geq N$ $a_n \leq b_n$ then we have that $\liminf a_n \leq \liminf b_n$ and $\limsup a_n \leq \limsup b_n$.

Theorem 4.15. If a_n is a sequence of real numbers then $\limsup a_n = \liminf a_n$ if and only if $\lim a_n$ exists and in this case all three are equal.

Theorem 4.101 (The Squeeze Theorem) If a_n, b_n, and c_n are sequences of real numbers such that there exists an N such that for all $n \geq N$ $a_n \leq b_n \leq c_n$ and $\lim a_n = \lim c_n = L$, then $\lim b_n = L$. Essentially, the outside sequences "squeeze" the middle sequence to converge with them.

Proof. We have that from the previous theorem that $\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup c_n$. Since $\liminf a_n = \lim a_n = \lim c_n = \limsup c_n$ we then see that $L \leq \liminf b_n \leq \limsup b_n \leq L$ so that they are equal and so $\lim b_n = L$. □

4.6 Ratio and Root Test's

We return to proving our familiar series tests from Calc 2 with the Ratio and Root tests! We first prove a useful theorem relating these two different tests.

Theorem 4.17. If a_n is a sequence of positive real numbers we have that

$$\liminf \left|\frac{c_{n+1}}{c_n}\right| \leq \liminf \sqrt[n]{c_n} \leq \limsup \sqrt[n]{c_n} \leq \limsup \left|\frac{c_{n+1}}{c_n}\right|.$$

We prove a quick lemma to aid in this proof.

Lemma 4.18. If $p > 0$ then $\lim_{n\to\infty} \sqrt[n]{p} = 1$.

Proof. If $p = 1$, this is easy to see. If $p > 1$ we have that the sequence is monotonically decreasing and bounded below by 1, so that it converges. If $\lim_{n\to\infty} p^{\frac{1}{n}} = L$ we have that $L \cdot p = \lim_{n\to\infty} p^{\frac{n+1}{n}} = p$ so that $L = 1$. The case $p < 1$ follows similarly.

□

We now prove the original theorem.

Proof. We prove the second half as the first is similar. Put $\alpha = \limsup \frac{c_{n+1}}{c_n}$ and $\beta = \limsup \sqrt[n]{c_n}$. We need only consider the case $\alpha \neq \infty$ for everything is less than ∞! Now assume α is finite and choose M to be some number greater than α. It follows that there exists an $N \in \mathbb{N}$ so that for all $n \geq N$, $\frac{c_{n+1}}{c_n} \leq M$, if this was not so α would not be the lim sup. It follows that for all $n \geq N$ we have $c_{n+1} \leq c_N M$. This is the same as saying $\sqrt[n]{c_{n+1}} \leq \sqrt[n]{c_N M^{-N}} M$. Notice that the term inside the root on the right is a constant, so by our lemma we have $\limsup \sqrt[n]{c_n} = \beta \leq M$. Since M was an arbitrary number greater than α we see that $\beta \leq \alpha$. $\square$

We now handle the Ratio Test.

Theorem 4.104 Let a_n be a sequence of complex numbers. If $\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$ then the series converges absolutely, if $\liminf \left|\frac{a_{n+1}}{a_n}\right| > 1$ then the series diverges.

Proof. For the first case if $\alpha = \limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$ choose $\alpha < \beta < 1$. It follows that there exists an N such that for all $n \geq N$ we have $\left|\frac{a_{n+1}}{a_n}\right| \leq \beta$. This gives that $|a_{N+1}| \leq \beta |a_N|$. Similarly, $|a_{N+2}| \leq \beta |a_{N+1}| \leq \beta^2 |a_N|$. In general $|a_{N+k}| < \beta^K |a_N|$. It follows that since $\beta < 1$ the series $|a_N| \sum \beta^n$ converges that by the comparison test $\sum |a_n|$ converges giving the result. In the case the $\liminf > 1$ the series must diverge as this implies that there is some point where $|a_{n+1}| > |a_n|$ so that a_n doesn't approach zero and diverges by the Divergence Test. $\square$

Theorem 4.105 (The Root Test) Let a_n be a sequence of complex numbers. If $\limsup \sqrt[n]{|a_n|} < 1$ the series $\sum a_n$ converges absolutely, if $\liminf \sqrt[n]{|a_n|} > 1$ the series diverges.

Proof. This follows from the ratio test and the theorem relating these concepts. $\square$

A nifty tool for using the Root Test is the fact that $\lim_{n\to\infty} \sqrt[n]{n} = 1$. To see this use The Binomial Theorem to have that $\sqrt[n]{n} = 1 + a_n$. Then $n = (1 + a_n)^n \geq \frac{n(n-1)}{2} a_n^2$ so that $0 < a_n \leq \sqrt{\frac{2}{n-1}}$ so that $a_n \to 0$ or $\sqrt[n]{n} \to 1$.

4.7 Alternating Series Test

We are finally approaching the end of the convergence tests and now we give the last familiar one, the Alternating Series Test.

Theorem 4.106 (The Alternating Series Test) Let a_n be a sequence of real numbers. If $|a_n| \geq |a_{n+1}|$, $a_{2m} < 0$, $a_{2m+1} > 0$, and $\lim_{n\to\infty} a_n = 0$, then the series $\sum a_n$ converges.

Proof. We claim the sequence of partial sums is bounded. We have $S_1 = a_1$. We see that for any $n \in \mathbb{N}$ we have $S_{2n+1} \leq S_{2n-1}$ as $S_{2n+1} = S_{2n-1} + (a_{2n+1} + a_{2n}) \leq S_{2n-1}$ since $|a_{2n}| \geq |a_{2n+1}|$. The same argument shows that $S_{2(n+1)} \geq S_{2n}$. Since each subsequence of the previous subsequences are monotonic and bounded, they are convergent. Now we just need to show they converge to the same thing. So let $\epsilon > 0$. Let $L_1 = \lim_{n\to\infty} S_{2n+1}$ and let $L_2 = \lim_{n\to\infty} S_{2n}$. Choose N_1 large enough that for all $n \geq N_1$ we have $|S_{2n+1} - L_1| < \frac{\epsilon}{3}$, N_2 large enough that for all $n \geq N_2$, $|S_{2n} - L_2| < \frac{\epsilon}{3}$, and finally choose N_3 large enough that for all $n \geq N_3$ we have $|a_n| < \frac{\epsilon}{3}$. We then have for all $n \geq \max\{N_1, N_2, N_3\}$ that $|L_1 - L_2| \leq |L_1 - S_{2n+1}| + |S_{2n+1} - S_{2n}| + |S_{2n} - L_2| < \epsilon$. Since ϵ was arbitrary we see that $L_1 = L_2$ so that the series converges! $\square$

4.8 The Cauchy Product Series

The last three sections of this chapter will cover material that should be new and can be quite interesting. The first of these is the Cauchy Product.

Given two series $\sum a_n$ and $\sum b_n$, it is clear that addition and multiplication by some constant is a very simple to understand concept, the next theorem makes this precise.

Theorem 4.22. Let $\sum a_n = A$ and $\sum b_n = B$ be convergent series. Then $\sum(a_n + b_n) = \sum a_n + \sum b_n = A + B$ and $c\sum a_n = \sum c_n = cA$.

One harder question is how to define the product of two series. You might be tempted to define $\sum a_n \cdot \sum b_n = \sum a_n b_n$, but this isn't a very useful definition, and it is not true that

$\sum a_n b_n = AB$.

Instead we should try multiplying the partial sums as generalize this into a formula.

It turns out that the correct formula is as follows:

$$\left(\sum_{k=0}^{n} a_k\right)\left(\sum_{k=0}^{n} b_k\right) = \sum_{m=1}^{k} a_m b_{k-m}.$$

We can now generalize this from the Partial Sums to the full series'.

Definition 4.23. Given two series $\sum a_n$ and $\sum b_n$, we define the **Cauchy Product** $\sum c_n$ as follows

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k}.$$

We might hope that if $\sum a_n = A$ and $\sum b_n = B$, then the product $\sum c_n = AB$, but this is not necessarily the case. The next theorem however gives a situation in which it is true!

Theorem 4.109 (Mertens' Theorem) Suppose $\sum a_n = A$, $\sum b_n = B$, and $\sum c_n$ is their Cauchy Product. If $\sum a_n$ converges absolutely, then $\sum c_n = AB$.

Proof. Let A_n, B_n, and C_n be the parital sums of the respective series. We have that

$$C_N = \sum_{n=0}^{N}\sum_{k=0}^{n} a_k b_{n-k} = \sum_{n=0}^{N} a_n B_{N-n} = \sum_{n=0}^{N} B_n a_{N-n} = \sum_{n=0}^{\infty}(B - B + B_n)a_{N-n}$$

$$= \sum_{n=0}^{N} B a_{N_n} + \sum_{n=0}^{N}(B_n - B)a_{N-} = A_N B + \sum_{n=0}^{N}(B_n - B)a_{N-n}.$$

If we can show that $\sum(B_n - B)a_{N-n} \to 0$ we are done as $\lim_{n\to\infty} A_N B = AB$.

Let $\epsilon > 0$ and let $A' = \sum_{n=0}^{\infty}|a_n|$. If we choose $N \in \mathbb{N}$ so that $|B_n - B| < \epsilon$ we have that

$$\left|\sum_{n=0}^{N}(B_n - B)a_{N-n}\right| \leq \sum_{n=0}^{N}|(B_n - B)||a_{N-n}| \leq \epsilon\sum_{n=0}^{N} a_{N-n} < \epsilon A'.$$

Since ϵ is arbitrary we see we can make the sum as small as desired and so we have that

$$\lim_{n\to\infty} C_n = \lim_{n\to\infty} A_n B + \sum_{k=0}^{n}(B_k - B)a_{n-k} = AB.$$

This completes the proof. $\square$

4.9 Rearrangments

We now come to the topic of rearrangements. It is very easy to informally describe a rearrangement of a series. We think of a rearrangement as just changing the order in which we add the terms in the series. One would expect this would have no impact on the result, but it turns out that without the stronger assumption of absolute convergence bizzare things can happen! We make this rigorous before giving two theorems on this topic.

Definition 4.25. Let $f : \mathbb{N} \to \mathbb{N}$ be a bijection. We will denote the $f(n)$ by f_n instead. Then the sum $\sum_{n=0}^{\infty} a_{f_n}$ called a **Rearrangement** of the series $\sum a_n$. The following theorem shows that rearrangments need not converge to the same value as the original series.

Theorem 4.111 (The Riemann Series Theorem) If $\sum a_n$ is a conditionally convergent series of real numbers. Then if $A \in \mathbb{R}$ there exists a rearrangement with $\sum a_{f_n} = A$.

Proof. Let $a_n^+ = \frac{|a_n|+a_n}{2}$ and $a_n^- = \frac{|a_n|-a_n}{2}$. Essentially, a_n^+ is the sequence of positive terms where negative terms are replaced by zeros, and a_n^- is the sequence of all negative terms multiplied to be positive and where the original positive terms are replaced by zeroes. It is clear that both $\sum a_n^+$ and $\sum a_n^-$ cannot converge, for if both did then their sum would be $\sum |a_n|$ which diverges, and infact neither can converge for if $\sum a_n^+$ converges then $\sum(a_n^+ - a_n) = \sum a_n^-$. We construct our rearrangement as follows take the first n terms in a_n^+ so that $S_n^+ \leq A < S_{n+1}^+$. $\square$

We now show that this problem is fixed by Absolute Convergence.

Theorem 4.27. Suppose $\sum a_n$ converges absolutely. Then if $\sum a_{f_n}$ is a rearrangement, then $\sum a_n = \sum a_{f_n}$.

Proof. Let $\epsilon > 0$. Then by the Cauchy Criterion there exists an N such that for all $m \geq n \geq N_1$ we have $\sum_{k=n}^{n} |a_k| < \epsilon$. Now consider $\sum a_n$. Let $M \in \mathbb{N}$ be chosen so that for all $i = 1, ..., N$ we have for some $w \leq M$ that $f(w) = i$. Essentially we have gone out far enough so that all the first n terms in the original series have been covered by our rearrangement. Then it follows that $\sum a_n - \sum a_{f_n}$ will consist only of terms greater than N and so by the Cauchy Criterion we have $|\sum a_n - \sum a_{f_n}| < \epsilon$. It follows that taking the limit they will get arbitrarily close and so the rearrangements are equal. □

4.10 The Contraction Fixed Point Theorem

To end this Chapter we prove a very famous theorem that requires both the tools of series as well as continuity: The Contraction Fixed Point Theorem. It was proven by Stefan Banach whose name has already appeared before.

Definition 4.28. Let (X, d) be a metric space. We say that a function $f : X \to X$ is a **Contraction** if there exists some $0 \leq c < 1$ such that for all $x, y \in X$ we have $d(x, y) \leq cd(f(x), f(y))$. It is immediate that a contraction is a continuous function.

Theorem 4.114 (Contraction Fixed Point Theorem) Let (X, d) be a complete nonempty metric space and let f be a contraction on X. Then f has a unique fixed point.

Proof. Uniqueness is immediate as if x, y were two distinct fixed points then $d(x, y) = d(f(x), f(y)) \leq cd(x, y) < d(x, y)$ shows a contradiction.

We now prove existence, and the method of our proof gives a method of finding the fixed point!

Since X is nonempty let $x \in X$ be any arbitrary point. We define a sequence $\{x_n\}$ by setting $x_1 = x$ and then letting $x_{n+1} = f(x_n)$. Let $K = d(x_1, x_2)$, then we see that

$d(x_2, x_3) \leq cK$ and by induction we have that $d(x_n, x_{n+1}) \leq c^{n-1}K$. We see then if $n < m$ then $d(x_n, x_m) \leq d(x_n, x_{n+1}) + ... + d(x_{m-1}, x_m) = \sum_{i=n+1}^{m} d(x_{i-1}, x_i) \leq \sum_{i=n+1}^{m} c^i K \leq K \sum_{i=n+1}^{\infty} c^i = K \frac{c^n}{(1-c)}$. It follows since $c < 1$ we can choose n large enough so that $d(x_n, x_m)$ is as small as we need and so the sequence is Cauchy. By completeness it follows the sequence converges so let $x' = \lim_{n\to\infty} x_n$. Since x is continuous at x' and $x_n \to x$ we have that $f(x') = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x$. $\square$

Note we needed series as we needed to rely on the limit of a geometric series to prove Cauchyness and we needed Continuity to justify that x' was a fixed point!

Chapter 5

Derivatives of Real Functions

We now turn to a very familiar topic from Calculus: The Derivative! The motivation and several uses of the derivative should be familiar to all students who've taken introductory calculus and most are not talked about here. This section is significantly easier and more familiar than almost all of the previous sections and the reader should treat this as a nice reward after some significant work! This chapter will primarily deal with derivatives of real functions, but the last section will briefly mention a few facts about derivatives of complex functions.

5.1 The Derivative

Definition 5.1. Let $f : [a, b] \to \mathbb{R}$ and let $x \in [a, b]$. We say that f is **Differentiable** at x if the following limit exists:

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

The value $f'(x)$ is called the **Derivative** of f at x. If f is differentiable on all of $[a, b]$ (or some substantial subset) then we can form a new function $f'(x) : [a, b] \to \mathbb{R}$ called

The Derivative Function of f.

If f is differentiable at a point x, then there exists some number $f'(x)$ and some function $r(h)$ so that $r(h) \to 0$ as $h \to 0$ and

$$f(x+h) = f(x) + f'(x) \cdot h + r(h).$$

Or essentially

$$f(x+h) \approx f(x) + f'(x) \cdot h.$$

The above statements follow immediately from the definition of the derivative and can be interpreted as saying that if f is differentiable at some point x, then closely around that point the function appears to be a line (i.e. of the form $f(x) = mx + b$). This outlook will be very important when considering derivatives for multivariate functions.

Theorem 5.2. Let $f : [a, b] \to \mathbb{R}$ be differentiable at p. Then f is continuous at p.

Proof. We need to show that $\lim_{x \to p} f(x) = f(p)$ (as p is certainly a limit point of $[a, b)$. We have that

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

so that multiplying everything by h

$$\lim_{h \to 0} f(p+h) - f(p) = \lim_{h \to 0} f'(p)h = 0$$

but this implies $\lim_{h \to 0} f(p+h) = f(p)$ and substituting $x = p + h$ gives the result. □

5.2 Basic Derivative Rules

We now give proofs of some of the very famous Derivative Rules!

Theorem 5.3. The function $f(x) = c$ for some $c \in \mathbb{R}$ is differentiable with derivative $f'(x) = 0$.

Proof. This follows as

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

$\square$

Theorem 5.4. The function $f(x) = x$ is differentiable with derivative $f'(x) = 1$.

Proof. This follows as

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{x + h - x}{h} = \lim_{h \to 0} \frac{h}{h} = \lim_{h \to 0} 1 = 1.$$

$\square$

Theorem 5.5. The function $f(x) = \frac{1}{x}$ is differentiable, when $x \neq 0$, with derivative $f'(x) = \frac{-1}{x^2}$.

Proof. This follows as

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h \to 0} \frac{-1}{x^2 + xh} = \frac{-1}{x^2}.$$

$\square$

Theorem 5.6. Let $f, g : [a, b] \to \mathbb{R}$. If f and g are differentiable at x, then the function $H(x) = af(x) + bg(x)$ is differentiable at x and $H'(x) = af'(x) + bg'(x)$.

Proof. We have

$$\lim_{h \to 0} \frac{H(x+h) - H(x)}{h} = \lim_{h \to 0} \frac{af(x+h) + bg(x+h) - af(x) - bg(x)}{h}$$

$$= a \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + b \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = af'(x) + bg'(x).$$

□

Theorem 5.121 (The Product Rule) Let $f, g : [a, b] \to \mathbb{R}$ both be differentiable at x, then the function $H(x) = (f \cdot g)(x)$ is differentiable at x and $H'(x) = f'(x)g(x) + f(x)g'(x)$.

Proof. We have

$$\lim_{h\to 0} \frac{H(x+h) - H(x)}{h} = \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}$$

$$= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h}$$

$$= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} + \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h}$$

$$= \lim_{h\to 0} f(x+h) \cdot \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} + g(x) \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = f(x)g'(x) + g(x)f'(x).$$

□

Theorem 5.122 (The Power Rule) Let $f(x) = x^n$ for $n \in \mathbb{Z}$, then $f'(x) = nx^{n-1}$.

Proof. We prove by induction. The case of $n = 0$ follows from Theorem XXX above. Now assume it holds for $g(x) = x^n$ and consider $f(x) = x^{n+1}$. Then $f(x) = x \cdot g(x)$ and by the product rule we have $f'(x) = g(x) + xg'(x) = x^n + x \cdot nx^{n-1} = (n+1)x^n$. By induction the result follows. The cases where $n < 0$ use the exact same method and is no more difficult. □

Corollary 5.8.1. We see from the above theorem and the theorem on linearity that every polynomial is differentiable.

Theorem 5.123 (The Chain Rule) Let $f : [a, b] \to \mathbb{R}$ be differentiable at x, and let $f(x) = y$ and let $g : f([a, b]) \to \mathbb{R}$ be differentiable at y. Then $h(x) = g(f(x))$ is differentiable at x and $h'(x) = g'(y) \cdot f'(x) = g'(f(x)) \cdot f'(x)$.

Proof. We need to show that

$$\lim_{h\to 0} \frac{H(x+h) - H(x)}{h} = \lim_{h\to 0} \frac{g(f(x+h)) - g(f(x))}{h} = g'(f(x)) \cdot f'(x).$$

If we multiplied the top and bottom by $f(x+h) - f(x)$ it would seemingly do the trick, but the expression would then be undefined if $f(x+h) = f(x)$ for values of $h \neq 0$, so we need some other trick. This trick is to use a function the mimics multiplying the top by the previous expression, but when the expression is equal it is instead replaced. Formally consider the function:

$$R(h) = \begin{cases} \frac{g(f(x+h) - g(y)}{f(x+h)-y} & f(x+h) \neq y \\ g'(y) & f(x+h) = y \end{cases}$$

Then we see that for all h we have

$$\lim_{h \to 0} \frac{g(f(x+h)) - y}{h} = \lim_{h \to 0} R(h) \frac{f(x+h) - y}{h},$$

this is because when $f(x+h) \neq y$, the expressions cancel, and for $f(x+h) = y$ the expression equals zero as $g(f(x+h)) = g(y)$. Now we can take the limit and see that

$$\lim_{h \to 0} R(h) \frac{f(x+h) - y}{h} = \lim_{h \to 0} R(h) \cdot \lim_{h \to 0} \frac{f(x+h) - y}{h} = g'(y) f'(x),$$

and this completes the proof. $\square$

Theorem 5.124 (The Quotient Rule) Let $f, g : [a, b] \to \mathbb{R}$ be differentiable at x, with $g(x) \neq 0$. Then the function $H(x) = \frac{f(x)}{g(x)}$ is differentiable at x with derivative

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

Proof. This essentially follows from the power rule and the chain rule. If we rewrite H as $H(x) = f(x) \cdot \frac{1}{g(x)}$, then the power rule gives

$$H'(x) = \frac{1}{g(x)} f'(x) + \left(\frac{1}{g(x)} \right)' f(x).$$

But the chain rule and our example on the function $\frac{1}{x}$ tells us that

$$\left(\frac{1}{g(x)}\right)' = \frac{-1}{g(x)^2} \cdot g'(x).$$

The result follows from these facts. $\square$

5.3 The Mean Value Theorem

We have now proven almost all of the very familiar derivative rules one encounters in an introductory Calculus course. This section will prove a few of the common theorems involving derivatives that one likely saw with the most important theorem being The Mean Value Theorem. Our first theorem should be very familiar as it is the basis for many optimization aspects encountered in Calculus, but first we must define some familiar terminology.

Definition 5.11. Let $f : X \to \mathbb{R}$, where X is some topological space and let $x \in X$. We say that x is a **Local Maximum** of the function f if there exists some neighborhood $N(x) \subseteq X$ so that for all $p \in N(x)$ we have that $f(p) \leq f(x)$. A **Local Minimum** is defined similarly.

Theorem 5.12. Let $f : (a, b) \to \mathbb{R}$ be differentiable. If $x \in (a, b)$ is a local maximum, then $f'(x) = 0$. The same holds for local minimum's.

Proof. Since x is a local maximum we can find some neighborhood of x, $B_r(x)$ satisfying the condition, ie for all $p \in (x - r, x + r)$ we have $f(p) \leq f(x)$. We have that $f'(x)$ exists, and so both the following expressions exist and are equal:

$$A = \lim_{h \to 0+} \frac{f(x+h) - f(x)}{h} \qquad \text{and} \qquad \lim_{h \to 0-} \frac{f(x+h) - f(x)}{h} = B.$$

For small enough values of h we see that $f(x + h) \in B_r(x)$. If we look at the terms in A we see that since $f(x)$ is the local max that for small h we have $f(x + h) < f(x)$ so that $A \leq 0$.

When we take the limit in B we will be approaching through negative h values so that we can rewrite b as

$$\lim_{h\to 0+} \frac{f(x)-f(x+h)}{h}.$$

Using a similar argument to what we used for A we see that $B \geq 0$ and since $A = B$ we see that the limit must be 0 or $f'(x) = 0$. $\square$

Theorem 5.127 (Rolle's Theorem) Let $f : [a,b] \to \mathbb{R}$ be continuous and differentiable on (a,b). If $f(a) = f(b) = 0$, then there exists some point $c \in (a,b)$ such that $f'(c) = 0$.

Proof. Since f is continuous and since $[a,b]$ is a compact set the Extreme Value Theorem guarantees a maximum $p \in [a,b]$ and a minimum $q \in [a,b]$ for f. It's easy to see that a maximum (sometimes called a global maximum) is also a local maximum, same applies to minimums. We may assume at least one p or q is not equal to a or b for if they both equalled one of these, then the function would be constant and the theorem follows. Without loss of generality assume $p \neq a$ and $p \neq b$. Then clearly $p \in (a,b)$ and from our above theorem since $f'(p)$ exists and p is a local maximum we have that $f'(p) = 0$, so that the point $p \in (a,b)$ satisfies the theorem. $\square$

We can now use Rolle's Theorem to prove the more general Mean Value Theorem!

Theorem 5.128 (The Mean Value Theorem) Let $f : [a,b] \to \mathbb{R}$ be continuous and differentiable on (a,b). Then there exists some point $c \in (a,b)$ such that $f(b) - f(a) = f'(c)(b-a)$.

Proof. Consider the function $g(x) = f(x) - f(a) - \frac{(f(b)-f(a))(x-a)}{b-a}$. Then g is continuous on $[a,b]$ and differentiable on (a,b) and $g(a) = g(b) = 0$, so by Rolle's Theorem there is some point $c \in (a,b)$ so that $g'(c) = 0$. But we can differentiate g as see that in general

$$g'(x) = f'(x) + \frac{f(b)-f(a)}{b-a},$$

and plugging in c we get

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

and the theorem follows. $\square$

The Mean Value Theorem is so important because it gives us a way of relating values of the function, which we may know about, with values of the derivative of the function, which we generally will know less about! We now prove an even more general version known as The Generalized Mean Value Theorem. This theorem includes the original MVT as a special case.

Theorem 5.129 (The Generalized Mean Value Theorem) Let $f, g : [a, b] \to \mathbb{R}$ be continuous and differentiable on (a, b). Then there exists some point $c \in (a, b)$ such that

$$f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)].$$

The case where $g(x) = x$ is the ordinary Mean Value Theorem!

Proof. Let $H(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)]$. Then we see that H is differentiable on the open interval (a, b), continuous on $[a, b]$, and $H(b) - H(a) = 0$. The theorem follows from applying the regular MVT. $\square$

A nice way to interpret this result is for parameteric functions. In this scenario we can interpret this result as saying that the MVT holds for parametric equations $(x(t), y(t))$ as the average slope $\frac{y(b)-y(a)}{x(b)-x(a)}$ will equal the instantaneous slope at some point c. One very important and familiar conseqeunce of the MVT is a famous tool to help with evaluating limits!

Theorem 5.130 (L'Hopital's Rule) Let $f, g : [a, b] \to \mathbb{R}$ be continuous and differentiable,

$p \in (a, b)$ with $\lim_{x\to p} f(x) = \lim_{x\to p} g(x) = 0$ and

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

Then

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

The same result holds if $\lim_{x\to p} f(x) = \lim_{x\to p} g(x) = \pm\infty$, the reader should attempt this version.

Proof. Let $\epsilon > 0$. From continuity we have that $f(p) = g(p) = 0$ and so

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

Applying the Generalized MVT on the interval $[p - \frac{\epsilon}{2}, p + \frac{\epsilon}{2}]$ gives there is some c so that

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

but now we are done as

$$\left| \frac{f(x) - f(p)}{g(x) - g(p)} - L \right| = \left| \frac{f'(c)}{g'(c)} - L \right| < \epsilon.$$

$\square$

5.4 Derivatives Classes

We showed earlier, that if a function is differentiable it must be continuous. We have not yet shown the converse is false but consider the function $f(x) = |x|$ on $[-a, a]$. This function is continuous, but fails to be differentiable at $x = 0$ as the left and right limits approach differing values! This generally shows that being differentiable is a quite nice property, even stronger than being continuous. One may wonder if being differentiable is nice enough to ensure that the derivative function is continuous as well. The following example shows this is false.

Theorem 5.17. The function

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases},$$

is differentiable, but its derivative $f'(x)$ is not continuous at $x = 0$.

Proof. We will assume knowledge of the sin function for this theorem, all properties of sin will be developed later. The function f is differentiable as the product of two differentiable functions for $x \neq 0$. At $x = 0$, we know since $-1 \leq \sin(p) \leq 1$ then the limit

$$\lim_{h \to 0} \frac{h^2 \sin\left(\frac{1}{h}\right)}{h} = 0,$$

so that f is differentiable everywhere and

$$f'(x) = \begin{cases} 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

We see that $f'(0) = 0$, but the $\lim_{x \to 0} f(x)$ does not exist and so f' is certainly not continuous at $x = 0$. $\square$

This theorem shows that in general for a function f to be differentiable it must be a "nice" function, but that the derivative function does not keep these "nice" properties. If we measure smoothness by how many times a function is differentiable, we see that a taking the derivative makes a function less smooth. This will be important when we discuss sequences of functions, as generally it is harder to talk about derivatives as they will be behave more bizarre than the original function, whereas we will see integral's of functions will behave more nicely!

Definition 5.18. Let $f : [a, b] \to \mathbb{R}$. We say that $f \in \mathcal{C}^0([a, b], \mathbb{R})$ if f is continuous. We will often denote this more simply by dropping the 0 and the $\mathbb{R}$, so that the sets $\mathcal{C}^0([a, b], \mathbb{R}) = \mathcal{C}([a, b])$. If the domain of the function is also obvious we may drop that as well. We now recursively define the sets $\mathcal{C}^k([a, b], \mathbb{R}) = \mathcal{C}^k([a, b])$ by saying $f \in \mathcal{C}^k([a, b])$ if f is differentiable and its derivative is in $\mathcal{C}^{k-1}([a, b])$. If $f \in \mathcal{C}^k([a, b])$ for all $k \in \mathbb{N}$ we say that $f \in \mathcal{C}^\infty([a, b])$ or that f is **Smooth** or **Infinitely Differentiable** on $[a, b]$.

As an easy example polynomial's are smooth! We will later see that so are the exponential and trig functions!A natural question one might have is about what types of functions are derivatives. If we have a function $f(x)$, how can we know if f is the derivative of some function? More precisely, is there some function $F(x)$ so that $F'(x) = f(x)$?

Definition 5.19. If $f : [a, b] \to \mathbb{R}$ and $F : [a, b] \to \mathbb{R}$ is a continuous and differentiable function, such that $F'(x) = f(x)$, then we say that F is an **Antiderivative** of the function $f(x)$. If F, G are two antiderivatives of f it is easy to see that $F - G = c$ for some $c \in \mathbb{R}$ using the mean value theorem. We can now reframe our posed question as asking when can we know if f has an antiderivative. It is clear that f can be discontinuous and still have an antiderivative, but our next two theorems will give tools to show when an antiderivative doesn't exist. In particular we will see that while derivatives need not be continuous, they strangely must still obey an intermediate value theorem!

Theorem 5.20. Let $f : [a, b] \to \mathbb{R}$ be differentiable. Then if $f'(a) \leq d \leq f'(b)$, then there is some $c \in [a, b]$ such that $f'(c) = d$. The same result holds if $f'(a) \geq d \geq f'(b)$.

Proof. The case where $d = f'(a)$ or $d = f'(b)$ is immediate so we may assume that $f'(a) < d < f'(b)$. Consider the function $g(x) = f(x) - d \cdot x$. Then g is differentiable on $[a, b]$ with derivative $g'(x) = f'(x) - d$. Since g is continuous on $[a, b]$ it attains a maximum at some point $c \in [a, b]$. It follows from Theorem XXX that $g'(c) = 0$, but this gives that $f'(c) - d = 0$ or $f'(c) = d$. $\square$

This shows us that functions that don't satisfy the IVT cannot have an antiderivative! More importantly it will be used in the next theorem to give an even stronger criterion. Namely, that if f has any simple discontinuities, then f has no antiderivative!

Theorem 5.21. Let $f : [a, b] \to \mathbb{R}$ be differentiable. Then $f'(x)$ has no simple discontinuities.

Proof. Suppose there is some point $c \in [a, b]$ such that f' has a simple discontinuity at c. Then we have that $M = \lim_{x \to c^+} f'(x)$ and $N = \lim_{x \to c^-} f'(x)$ both exist.
If $M \neq N$ and without loss of generality assume $M < N$ (the other case is similar). Then there is some number $q \neq f'(c)$ so that $M < q < N$. If we choose $\epsilon > 0$ so that $M + \epsilon < c < N - \epsilon$ we see we there exists some $\delta > 0$ so that $f'((c - \epsilon, c)) \subseteq B_\epsilon(M)$ and $f'((c, c + \epsilon)) \subseteq B_\epsilon(N)$ so that for all $x \in (c - \epsilon, c + \epsilon)$ we have $f'(x) \neq q$ contradicting the previous theorem. The case where $M = N$ is also handled similarly. $\square$

We have just given a few basic tools concerning the discussion of Antiderivatives. Next chapter we will see that any continuous function has an antiderivative, but for this we will need Integration! So for the meantime we postpone this discussion and move to the last section of the chapter.

5.5 Complex Differentiability

To end this chapter we will briefly discuss Complex Differentiability!

Definition 5.22. Let $U \subseteq \mathbb{C}$ be open and let $f : U \to \mathbb{C}$. If $z \in U$ we say that f is **Differentiable** at z if

$$\lim_{h \to 0} \frac{f(z + h) - f(z)}{h}$$

exists and denote this limit $f'(z)$.

This definition is almost exactly the same as the one we gave for real functions. The only difference is of course the domain we chose, but obviously that would change if we were discussing complex functions. One might then think that differentiability in $\mathbb{R}$ and $\mathbb{C}$ behave similarly and share similar properties. The answer is no. Differentiability in $\mathbb{C}$ is quite different and is far more restrictive than in $\mathbb{R}$ (and even functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ which we will explore much later). Most of the theorems concerning differentiability rules such as the power, product, quotient, and chain rule's still hold and the proofs go without change essentially. On the other hand the MVT and L'Hopital's Rule do not hold. The main difference is that in $\mathbb{R}$ there are two possible paths you can approach the point z in the above limit: the left and right. However, in $\mathbb{C}$ there are an infinite number of paths you can take to approach the point z and the limit must be the same along all possible paths! It turns out there is even more restrictions at play, but for the meantime it is important that when complex differentiability comes up again in Chapter 8 the reader remembers that it is indeed far more restrictive.

Chapter 6

Riemann Integration

6.1 The Riemann Integral

We now turn to the second pillar of introductory calculus: Integration! To clear up confusion, in Calculus you likely dealt with two things called integrals. You likely saw countless examples of the so called "Indefinite Integral" of a function and you probably saw many "Definite Integrals". For the purposes of Analysis, it turns out the Definite Integral is the more important one, where as the indefinite integral is really just an antiderivative of a function. With that out of the way we will now develop what we mean by a definite integral, hence on only called the integral.

Definition 6.1. Consider an interval $[a, b]$. A **Riemann Partition** of $[a, b]$ or simply a **Partition** is a finite set of ordered points in the interval $[a, b]$ including a and b. We will typical label and order the points so that for the partition $P = \{a, x_2, ..., x_{n-1}, b\}$ corresponds to the points $a = x_0 < x_1 < ... < x_n = b$. For a given partition P the $i - th$ subinterval is $[x_{i-1}, x_i]$.

Definition 6.2. Given an interval $[a, b]$, a partition P, and a bounded function $f : [a, b] \to \mathbb{R}$ we write $M_i = \sup f(\Delta_i)$, $m_i = \sup f(\Delta_i)$, and the length of the $i - th$ subinterval is

$\Delta_i = x_i - x_{i-1}$. We now define the **Upper and Lower Riemann Sums** of f over $[a,b]$ respectively as:

$$U(P,f) = \sum_{i=1}^{n} M_i\Delta_i \qquad \text{and} \qquad L(P,f) = \sum_{i=1}^{n} m_i\Delta_i.$$

Definition 6.3. Given a bounded function $f : [a,b] \to \mathbb{R}$ we define the **Upper and Lower Integral's** of f over $[a,b]$ respectively as

$$\overline{\int_a^b} f dx = \inf\{U(P,f) | P \text{ is a partition of } [a,b]\},$$

$$\underline{\int_a^b} f dx = \sup\{L(P,f) | P \text{ is a partition of } [a,b]\}.$$

If both the upper and lower integral's of f exist and are equal then we say that f is **Riemann Integrable** or just **Integrable** over $[a,b]$ and denote the **Riemann Integral** as

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

If f is integrable over $[a,b]$ we say that $f \in \mathcal{R}$ on $[a,b]$, if it's obvious we will simply say $f \in \mathbb{R}$.

The definition above can be quite a lot but it is probably the easiest possible method of defining the integral! It is clear that for any partition P we have $L(P,f) \leq U(P,f)$. Since this holds for every partition we can also see that

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

To end this section we introduce a few facts about the integral and partitions to aid us in the next section in which we discuss what function are integrable.

Definition 6.4. Given a partition P of the interval $[a,b]$, we define a **Refinement** of P to be another partition P' so that $P \subseteq P'$. Similarly, if P_1 and P_2 are both partitions we define the **Common Refinement** to be $P_1 \cup P_2$. Our next theorem will show why considering refinements is useful.

Theorem 6.5. If $f : [a, b] \to \mathbb{R}$ is bounded and P is a partition of $[a, b]$, with a refinment P'. Then

$$L(P, f) \leq L(P', f) \leq U(P', f) \leq U(P, f).$$

Proof. We prove the last inequality, the rest are all similar. Let P be a partition and let P' be a partition with one more point. Then there is some subinterval $[x_{i-1}, x_i]$ so that P' has a point $q \in [x_{i-1}, x_i]$ which splits this interval into two $[x_{i-1}, q], [q, x_i]$. Clearly $M_i \geq \sup[x_{i-1}, q]$ and $M_i \geq \sup[q, x_i]$. This case then follows immediately from that observation. The general case follows from induction. □

6.2 Integrable Functions

We now turn our attention to discussing exactly which functions are Riemann Integrable. The first question that one should have is something along the lines of: is every function integrable? Trivially, unbounded functions aren't integrable as we've only defined integrals for bounded function. But besides these silly cases are there any examples? The next theorem will answer that question.

Theorem 6.6. Let $f : [a, b] \to \mathbb{R}$ be the Dirichlet function

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

Then f is NOT integrable.

Proof. This proof also gives an example of how to work with partitions! Let P be an arbitrary partition of $[a, b]$. Given any subinterval $[x_{i-1}, x_i]$ it is clear that $M_i = 1$ and $m_i = 0$. So that $U(P, f) = \sum_{i=1}^n \Delta_i = b - a$ and $L(P, f) = \sum_{i=1}^n 0 = 0$. Since, this is independent of partition we see that the lower integral must be 0 and the upper integral must be 1. Since these are not equal we see f is not integrable! $\square$

With this example out of the way we now see the need of determining what functions are integrable. It turns out that there is infact an easy criterion for determining which functions are integrable due to Henri Lebesgue, whose name will become very familiar in the later chapter on Real Analysis. The goal of this section is to prove this criterion. For now we prove a much simpler and harder to use criterion.

Theorem 6.7. Let $f : [a, b] \to \mathbb{R}$ be bounded. Then f is integrable if and only if for all $\epsilon > 0$ there exists a partition P so that $U(P, f) - L(P, f) < \epsilon$.

Proof. First assume f is integrable. Then $\underline{\int}_a^b f dx = \overline{\int}_a^b f dx$. By definition there exists a partition P_1 so that $\underline{\int}_a^b f dx - L(P_1, f) < \frac{\epsilon}{2}$ and similarly there exists a partition P_2 so that $U(P_2, f) - \overline{\int}_a^b f dx < \frac{\epsilon}{2}$. It follows from taking P to be the common refinement of P_1 and P_2 we have $U(P, f) - L(P, f) < \epsilon$.

Now for the other direction suppose for all $\epsilon > 0$ we can find a partition satisfying our requirement. Then given $\epsilon > 0$ we can choose a partition P so that $U(P, f) - L(P, f) < \epsilon$. Then $\overline{\int}_a^b f dx - \underline{\int}_a^b f dx \leq U(P, f) - L(P, f) < \epsilon$. Since this holds for all $\epsilon > 0$ we see they must be equal completing the proof. $\square$

We will need this criterion in the proof of Lebesgue's Criterion later on as well as a great deal of extra machinery we will now build up. The most important of which is the notion of measure, specifically a zero measure set. Informally, for real sets we think of measure as how much space on the real line a set takes up. So that means informally a set of measure zero takes up no space on the real line. We now formalize this.

Definition 6.8. Given an open interval (a, b) we define the **Length** of (a, b) as $L((a, b)) = b - a$ and $L(\emptyset) = 0$. Given an arbitrary subset $S \subseteq \mathbb{R}$ we say that S has **Zero Measure** or is **Measure Zero** if for all $\epsilon > 0$ there exists a countable collection of open intervals $\{I_n | n \in \mathbb{N}\}$ so that $S \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} L(I_n) < \epsilon$.

The next theorem shows a few very useful facts about sets of measure zero.

Theorem 6.9.
(a)Let $S_1, S_2, ...$ be a countable collection of sets of measure zero. Then $\bigcup_{n=1}^{\infty} S_n$ is a set of measure zero.
(b) A singelton is a set of measure zero.
(c) A countable set has measure zero.
(d) If $A \subseteq B$ and B has measure zero, then A has measure zero.

Proof. First will prove (a). Let $\epsilon > 0$ By definition for each i we can find a countable collection of open intervals $\{I_{i_n} | n \in \mathbb{N}\}$ so that $\sum_{n=1}^{\infty} L(I_{i_n}) < \frac{\epsilon}{2^i}$ and so that $S_i \subseteq \bigcup_{n=1}^{\infty} I_{i_n}$. It follows that $\bigcup_{i=1}^{\infty} S_i \subseteq \bigcup_{i=1}^{\infty}(\bigcup_{n=1}^{\infty} I_{i_n})$ and $\sum_{i=1}^{\infty}\sum_{n=1}^{\infty} L(I_{i_n}) < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon$. Since each $\{I_{i_n}\}$ is countable and we have a countable collection of these it follows that the collection of all $\{I_{i_n} | i, n \in \mathbb{N}\}$ is countable completing (a).
For (b), given $\epsilon > 0$ and a point $p \in \mathbb{R}$ then the interval $I_1 = (p - \frac{\epsilon}{4}, p + \frac{\epsilon}{4})$ has $L(I_1) = \frac{\epsilon}{2} < \epsilon$. It then follows that taking $I_n = \emptyset$ for $n > 1$ gives that $\{p\} \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} L(I_n) < \epsilon$.
Part (c) follows from (a) and (b).
Part (d) follows as any covering of B also covers A. $\square$

Definition 6.10. Let $S \subseteq \mathbb{R}$. Let P be a property. If there exists some set B with zero measure and the property P holds on $S - B$ we say that P holds **Almost Everywhere** on S. Simply put a property holds almost everywhere if the set it fails to hold on has measure zero.

The last set of tools we will need is the concept of Oscillation.

Definition 6.11. Let $f : X \to \mathbb{R}$ with X some metric space. Given $\epsilon > 0$ and some subset $S \subseteq X$ we define the **Oscillation** of f over S as $\omega(f, S) = \sup_{x \in S} f(x) - \inf_{x \in S} f(x)$. Given some point $x_0 \in X$ we define the **Oscillation at** x_0 as $\omega(f, x_0) = \lim_{r \to 0} \omega(f, B_r(x_0))$. The oscillation measures how variant the function is over some set. The oscillation at a point measures this limiting value as the sets close around some point. It is easy to see that $\omega(f, x_0) \geq 0$. The next theorem relates oscillation to continuity.

Theorem 6.12. Let $f : X \to \mathbb{R}$ where X is some metric space. For $x_0 \in X$ we have $\omega(f, x_0) = 0$ if and only if f is continuous at x_0.

Proof. Suppose $\omega(f, x_0) = 0$ and let $\epsilon > 0$. We need to find a $\delta > 0$ so that $d_X(x, x_0) < \delta$ implies $|f(x) - f(x_0)| < \epsilon$. But since $\lim_{r \to 0} \omega(f, B_r(x_0)) = 0$ we see there exists some $r > 0$ so that $\omega(f, B_r(x_0)) < \epsilon$. It follows that taking $\delta = r$ we have that if $d_X(x, x_0) < \delta$, then $x \in B_r(x_0)$ so that $|f(x) - f(x_0)| \leq \sup_{p \in B_r(x_0)} f(p) - \inf_{p \in B_r(x_0)} f(p) = \omega(f, B_r(x_0)) < \epsilon$, so that f is continuous at x_0.

For the other direction suppose f is continuous at x_0. Let $\epsilon > 0$ and choose r so that $d_X(x, x_0) < r \implies |f(x) - f(x_0)| < \frac{\epsilon}{2}$ and consider $B_r(x_0)$. Then we can find some points $p, q \in B_r(x_0)$ so that $\sup_{x \in B_r(x_0)} f(x) - f(p) < \frac{\epsilon}{4}$ and $f(q) - \inf_{x \in B_r(x_0)} f(x) < \frac{\epsilon}{4}$. But then it follows that

$$\omega(f, B_r(x_0)) = \sup f(x) - \inf f(x) \leq |f(p) - f(q)| + \frac{\epsilon}{2} < \epsilon$$

. □

Since $\omega(f, x_0)$ is always positive it follows that f is discontinuous at x_0 if and only if $\omega(f, x_0) > 0$. We need one last tool.

Theorem 6.13. Let $f : [a, b] \to \mathbb{R}$. Let $v > 0$, then the set $S = \{x \in [a, b] | \omega(f, x) < v\}$ is open so that the set $C = \{x \in [a, b] | \omega(f, x) \geq v\}$ is closed.

Proof. Let $x \in S$. Then $\lim_{r\to 0} \omega(f, B_r(x)) = p < v$. Now choose r_0 so that $\omega(f, B_{r_0}(x)) < v$. It follows that for all $x' \in B_{r_0}(x)$ we have $\omega(f, x') < v$ so that $B_{r_0}(x) \subseteq S$ so S is open. But then it follows that $C = [a, b] \cap S^c$ as the union of two closed sets is closed. □

We finally have enough tools to state and prove Lebesgue's Criterion.

Theorem 6.150 (Lebesgue's Criterion for Riemann Integration) Let $f : [a, b] \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}$ on $[a, b]$ if and only if f is continuous almost everywhere on $[a, b]$.

This theorem is perfect as it tells us exactly which functions are Riemann integrable and shows the close connection between Riemann's definition of the integral and continuity.

Proof. For the first direction suppose $f \in \mathcal{R}$. We need to show the set of discontinuities of f, we will call D, is a set of measure zero. Let $D_n = \{x \in [a, b] | \omega(f, x) > \frac{1}{n}\}$, then $\bigcup_{n=1}^{\infty} D_n = D$ and by Theorem XXX (a) it suffices to show that each D_n has measure zero. Fix D_N and let $\epsilon > 0$. Then there is some partition P of $[a, b]$ so that

$$U(P, f) - L(P, f) = \sum_{i=1}^{n} (M_i - m_I)\Delta_i = \sum_{i=1}^{n} \omega(f, [x_{i-1}, x_i])L([x_{i-1}, x_i] < \frac{\epsilon}{N}.$$

This follows from Theorem 6.7 and comparing definitions of M_i, m_i and $\omega(f, [x_{i-1}, x_i])$. Since P is a collection of finite points P has measure zero. Now let I_k be the collection of all open sub-intervals (x_{k-1}, x_k) so that $D_N \cap I_k \neq \emptyset$. Then clearly we have for these sub-intervals that $\omega(f, I_k) > N$. It follows then that

$$N \cdot \sum_{i=k_0}^{k} L(I_i) \le \sum_{i=k_0}^{k} \omega(f, I_i)L(I_i) \le \sum_{i=1}^{n} \omega(f, [x_{i-1}, x_i])\Delta_i < \frac{\epsilon}{N},$$

so that $\sum L(I_i) < \epsilon$. Or the collection of sub-intervals intersecting D_N is a set of measure zero. Since every point of D_N is either in the collection of sub-intervals or is a member of P, both of which are measure zero, we see that D_N is a set of measure zero, and therefore so is D.

Now suppose the set of discontinuities of f, we call D, is a set of measure zero. Let $\epsilon > 0$ and consider the set $S = \{x \in [a,b] | \omega(f,x) \geq \epsilon\}$. We have that $S \subseteq D$ so that it has measure zero so that there is a countable collection of open intervals I_n that cover S where $\sum_{n=1}^{\infty} L(I_n) < \epsilon$. Since S is a closed subset of $[a,b]$ it is compact and so we actually need only finitely many $I_1, ..., I_n$ to cover S and clearly we have $\sum_{i=1}^{n} L(I_i) < \epsilon$. Now consider the set $M = [a,b] - (I_1, ..., I_n) = [a,b] \cap I_1^c \cap ... \cap I_n^c$. M is closed as it is the intersection of closed sets and since $M \subseteq [a,b]$ it is compact! Further, for all $x \in M$ we have that $\omega(f,x) < \epsilon$, but this means that for each x there is some $r_x > 0$ so that $\omega(f, B_{r_x}(x)) < \epsilon$. The collection of sets $B_{r_x}(x)$ is clearly an open cover for M so that there exists a finite subcover $B_{r_{x_1}}(x_1), ..., B_{r_{x_k}}(x_k)$. We now simplify our open balls as follows: we let $I_1' = \overline{B_{r_{x_1}}(x_1)} \cap [a,b]$, and we set $I_i' = \overline{B_{r_{x_i}}(x_i)} \cap [a,b] - I_1' - ... - I_{i-1}'$. It is clear that the sets I_i $i = 1, ..., k$ still cover M and $\sum_{i=1}^{k} L(I_i'^{\circ}) < (b-a)$. Now we form a partition P of $[a,b]$ by taking the end points of all the following sets $\overline{I_1}, ..., \overline{I_n}, I_1', ...I_k'$. Then we have that

$$U(P,f) - L(P,f) = \sum_{i=1}^{m} \omega(f, [x_{i-1}, x_i])\Delta_i = \sum_{i=1}^{n} \omega(f, I_i) L(I_i) + \sum_{i=1}^{k} \omega(f, I_i') L(I_i'^{\circ})$$

$$\leq \omega(f, [a,b]) \sum_{i=1}^{n} L(I_i) + \epsilon \sum_{i=1}^{k} L(I_i'^{\circ})$$
$$\leq \omega(f, [a,b])\epsilon + \epsilon(b-a) = (\omega(f, [a,b]) + (b-a))\,\epsilon.$$

Since the ending inequality can be made as small as we like we are done. □

Corollary 6.14.1. The following are all true for bounded functions $f, g : [a,b] \to \mathbb{R}$

(a) If f is continuous then $f \in \mathcal{R}$.

(b) If f is in $\mathcal{R}$ then $|f| \in \mathcal{R}$.

(c) If $f, g \in \mathcal{R}$, then $f \cdot g \in \mathcal{R}$.

(d) If $f \in \mathcal{R}$ and g is continuous then $g \circ f \in \mathcal{R}$.

(e) If f has only simple discontinuities $f \in \mathcal{R}$. In particular monotonic functions are integrable.

Proof. All of these results follow from Lebesgue's Criterion along with the continuity properties of the functions involved! $\square$

6.3 Properties of the Integral

After the lengthy proof of Lebesgue's Criterion, its corollary has given us a great group of functions we can integrate! We now would like to investigate the properties the Riemann Integral satisfies! The first theorem deals with linearity!

Theorem 6.15. If $f, g \in \mathcal{R}$ on $[a, b]$ then the function $f(x) + g(x) \in \mathcal{R}$ and if $c \in \mathbb{R}$ is some constant then $cf(x) \in \mathcal{R}$ and

$$\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx,$$

and

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx.$$

Proof. Let P be a partition of $[a, b]$. Then $L(P, f) + L(P, g) \le L(P, f + g) \le U(P, f + g) \le U(P, f) + U(P, g)$, so that the first claim follows. The second claim follows as $cL(P, f) = L(P, cf)$ and $cU(P, f) = U(P, cf)$. $\square$

We now present a group of other properties as one large theorem.

Theorem 6.16. Let $f, g \in \mathcal{R}$ on $[a, b]$. Then

(a) If $c \in (a, b)$, then $f \in \mathcal{R}$ on $[a, c]$ and $[c, b]$ and $\int_a^b f dx = \int_a^c f dx + \int_c^b f dx$.

(b) If $m < f < M$ for all $x \in [a, b]$, then $m(b - a) \le \int_a^b f dx \le M(b - a)$.

(c) If $f \le g$ for all $x \in [a, b]$ then $\int_a^b f dx \le \int_a^b g dx$.

(d) We have $\left|\int_a^b f dx\right| \le \int_a^b |f| dx$.

Proof. The proof of (a) is the most difficult. Let P be a partition of $[a, b]$ and let P' be the refinement including c (if c was not in the original partition. Now let $P_1 = P' \cap [a, c]$ and $P_2 = P' \cap [c, b]$. Then $L(P, f) \leq L(P', f) = L(P_1, f) + L(P_2, f) \leq U(P_1, f) + U(P_2, f) = U(P', f) \leq U(P, f)$. From this part (a) can be seen.

Part (b) follows almost immediately from definition. As for all partitions we have $m(b-a) \leq L(P, f) \leq U(P, f) \leq M(b-a)$.

For part (c) the function $h = g - f$ is integrable and always positive so that (b) plus linearity tells us $0 \leq \int_a^b h dx = \int_a^b g dx - \int_a^b f dx$.

Finally, for (d) we have that for all x, $f(x) \leq |f(x)|$ so that (c) gives $\int_a^b f dx \leq \int_a^b |f| dx$. We also see that for all x that $-|f(x)| \leq f(x)$ so that we have $-\int_a^b |f| dx = \int_a^b -|f| dx \leq \int_a^b f dx$ so the result follows. □

Our final property will show that we need not use M_i or m_i, but can approximate the integral for sufficient partitions.

Theorem 6.17. Let $f \in \mathcal{R}$ on $[a, b]$ and for $\epsilon > 0$ let P be a partition satisfying $U(P, f) - L(P, f) < \epsilon$. Then if we have $p_i \in [x_{i-1}, x_i]$ for all $i = 1, ..., n$ we have

$$\left| \sum_{i=1}^{n} f(p_i) \Delta_i - \int_a^b f dx \right| < \epsilon.$$

Proof. Such a partition clearly exists. It follows that for such a partition we have that $L(P, f) \leq \sum_{i=1}^{n} f(p_i) \Delta_i \leq U(P, f)$ and $L(P, f) \leq \int_a^b f dx \leq U(P, f)$. The result follows. □

6.4 The Fundamental Theorem of Calculus

So far we've discussed what functions are integrable and the properties the integral satisfies. We now turn our attention to evaluating integrals. It turns out that the answer to this question is related to our questions about antiderivatives in the last section! The famous

Fundamental Theorem of Calculus presented below links to the two pillars of calculus: Differentiation and Integration by showing that they are inverses! We will also present a few easy, and likely familiar, corollaries!

Theorem 6.154 (The Fundamental Theorem of Calculus Part One) Let $f \in \mathcal{R}$ on $[a,b]$ and let $F : [a,b] \to \mathbb{R}$ be the function defined by $F(x) = \int_a^x f(t)dt$. Then F is continuous on $[a,b]$ and if f is continuous at $x \in [a,b]$, then F is differentiable at x with $F'(x) = f(x)$.

The above theorem tells us that the integral function given above is always continuous! Further, it tells us that if the original function was continuous, then $F(x)$ is actually an antiderivative!

Proof. Let $\epsilon > 0$. Since f is bounded we have there is some M so that $|f| \leq M$ for all $x \in [a,b]$, assume this $M > 0$. Choose $\delta = \frac{\epsilon}{M}$. It follows then that for all $x, y \in [a,b]$ if $|x-y| < \delta$ (assume WLOG $x < y$) then $|F(X) - F(Y)| = |\int_a^x f dx - \int_a^y f dx| \leq \int_x^y |f| dx \leq M(y-x) < M\delta = \epsilon$. This shows continuity, and in fact uniform continuity.
Now assume that f is continuous at a point $x \in [a,b]$ and let $\epsilon > 0$. Then there exists some $\delta > 0$ so that for all $y \in [a,b]$ if $|x-y| < \delta$ then $|f(x) - f(y)| < \epsilon$. It follows then that

$$\frac{F(x+\delta) - F(x)}{h} - f(x) = \left(\frac{1}{\delta}\int_x^{x+\delta} f(t)dt\right) - f(x),$$

$$= \frac{1}{\delta}\int_x^{x+\delta} f(t)dt - \frac{1}{\delta}\int_x^{x+\delta} f(x)dt = \frac{1}{\delta}\int_x^{x+\delta} f(t) - f(x)dt,$$

$$\leq \frac{1}{\delta}\int_x^{x+\delta} |f(t) - f(x)|dt < \frac{1}{\delta}\int_x^{x+\delta} \epsilon \; dt = \epsilon.$$

This shows that $\lim_{h\to 0} \frac{F(x+h)-F(x)}{h} = f(x)$ and so we are done. $\square$

The above proof was quite simple and the only trick was rewriting $f(x)$ as $\frac{1}{\delta}\int_x^{x+\delta} f(x)dt$. Make sure to notice the variable of integration was t so that the integral was of just a constant! The First Fundamental Theorem of Calculus tells us that essentially the derivative of the integral of a function gives the original. The Second Fundamental Theorem will tell us that the integral of the derivative will give us the original, so that they truly are inverses! It also gives us a clever tool for evaluating integrals.

Theorem 6.155 (The Fundamental Theorem of Calculus Part Two) If $f \in \mathcal{R}$ on $[a, b]$ and $F : [a, b] \to \mathbb{R}$ is differentiable with $F'(x) = f(x)$ for all $x \in [a, b]$. Then

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

Proof. One might hope we can apply the FTC Part One and get our result quickly. One would be wrong! FTC Part One requires continuity of $f(x)$ in order to generate an antiderivative and in this problem we have only assumed that f is integrable! So we must use a different method!

Let P be a partition of $[a, b]$ so that $U(P, f) - L(P, f) < \epsilon$. It follows that since F is differentiable on each sub-interval $[x_{i-1}, x_i]$ that by the mean value theorem there is some point $c_i \in [x_{i-1}, x_i]$ so that

$$\frac{F(x_i) - F(x_{i-1})}{\Delta_i} = F'(c_i) = f(c_i).$$

But then for these c_i we have then that

$$\sum_{i=1}^n f(c_i)\Delta_i = \sum_{i=1}^n F(x_i) - F(x_{i-1}) = F(b) - F(a).$$

But now Theorem 6.17 shows that

$$\left|F(b) - F(a) - \int_a^b f dx\right| < \epsilon.$$

Since this holds for all $\epsilon > 0$ we see that

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

and so we are done! $\square$

We have now proven the the most important theorem in Calculus! The reader should be quite happy and the remaining few theorems of this chapter all follow quite easily and should be familiar results from elementary calculus.

Theorem 6.156 (Substitution Rule) Let $f : [a, b] \to \mathbb{R}$ be continuous and let $u : [a, b] \to \mathbb{R}$ be continuously differentiable. Then

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

Proof. We have both sides are integrable as f is continuous and so is $f(u(x)) \cdot u'(x)$ it is just a question of equality. Then FTC1 tells us that f has an antiderivative $F(x)$ and FTC2 tells us that

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

Next we have if $h(x) = F(u(x))$ then $h'(x) = F'(u(x)) \cdot u'(x) = f(u(x))u'(x)$ so that FTC2 again gives

$$\int_a^b f(u(x))u'(x)dx = \int_a^b h'(x)dx = h(b) - h(a) = F(u(b)) - F(u(a)),$$

so that both integrals are equal. $\square$

Theorem 6.157 (Integration by Parts) Suppose F, G are differentiable functions so that $F'(x) = f(x)$, $G'(x) = g(x)$ and $f, g \in \mathcal{R}$, then

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

Proof. This follows from the Product Rule of derivatives $(F \cdot G)' = F(x)g(x) + f(x)G(x)$ plus the Fundamental Theorem of Calculus. □

Definition 6.22. We define the **Average Value** of an integrable function over an interval $[a, b]$ as

$$\text{Avg}(f) = \frac{1}{b-a}\int_a^b f\,dx.$$

Theorem 6.159 (Average Value Theorem) If $f : [a, b] \to \mathbb{R}$ is continuous then there exists some point $c \in (a, b)$ so that $\text{Avg}(f) = f(c)$.

Proof. Since f is continuous FTC1 gives us an antiderivative $F(x) = \int_a^x f\,dx$. Clearly F is differentiable so that it satisfies the Mean Value Theorem so that there is some point $c \in (a, b)$ where $\frac{F(b)-F(a)}{b-a} = F'(c)$. Rewriting we have

$$\frac{F(b) - F(a)}{b - a} = \frac{1}{b-a}\int_a^b f\,dx = F'(c) = f(c),$$

which completes the proof. □

6.5 Complex Integration

We now briefly discuss how to integrate some complex functions, namely functions from a real interval $[a, b]$ to the complex plane. We will need this for some results in the next few chapters.

Definition 6.24. Let $f : [a, b] \to \mathbb{C}$. Then we can write f in its real and imaginary components as $f(x) = u(x) + iv(x)$. We then define the **Riemann Integral** of f over $[a, b]$ as:

$$\int_a^b f(x)\,dx = \int_a^b u(x)\,dx + i\int_a^b v(x)\,dx.$$

The integrals on the right exists if and only if $u(x)$ and $v(x)$ are continuous almost everywhere on $[a, b]$ which happens if and only if f is continuous almost everywhere on $[a, b]$.

Chapter 7

Sequences and Series of Functions

We have officially finished Parts One and Two of the book. We have proven almost all results one would find in an elementary Calculus. We now turn our attention to some of the remaining results in Calculus 2 we have yet to prove and move to new results. Part Three of this book discusses sequences and series of functions! This chapter will deal with the general theory discussing what it means for a sequence of functiosn to converge, the next chapter will deal with the special class of "Analyitic" Functions, and the last chapter in this part will deal with how we can represent functions as a series of sines and cosines via Fourier Series! These chapters, this one in particular, will reembrance Topology using it much more than Part Two. We now begin our discussion of the present chapter.

7.1 Pointwise Convergence

Back in Chapters 2 and 4 we talked a great deal about sequences and series. Chapters 3,5, and 6 talked a great deal about functions. It is not time to combine these two areas and discuss how sequences and series of functions behave! This topic may seem very foreign or unfamiliar to you, but this could not be further from the truth. For all the time in Calculus Two students are concerned with Taylor Series and Power series for functions. Further students often integrated a Taylor Series by performing something such as

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

But whose to say these two values are actually equal? We have never proved that the integral of the sum is the sum of the integral and students likely took this for granted without ever wondering about the validity. This chapter will begin to address these claims. We start with what we mean for a sequence of functions to converge.

Definition 7.1. Let X be a topological space and for each $n \in \mathbb{N}$ let $f_n : X \to \mathbb{C}$. We say that the sequence of functions f_n **Converges** or **Converges Pointwise** to the limit function f if for all $x \in X$ we have $\lim_{n\to\infty} f_n(x) = f(x)$. We denote this by saying

$$\lim_{n\to\infty} f_n = f,$$

or saying $f_n \to f$ as $n \to \infty$. Put simply a sequence of functions converges pointwise to f if it converges at every point.

With this new definition in place we can begin to ask and answer questions about convergence!
For example is it true that if f_n is a sequence of continuous functions that converges to f must f be continuous? Or is it true the sequence of integrals equals the integral of the sequence? What about derivatives? It turns out that with pointwise convergence all of these can be false.

Example 7.2. Consider the sequence of continuous functions $f_n : [0,1] \to \mathbb{R}$ defined by

$$f_n(x) = \begin{cases} 0 & x = 0 \\ nx & 0 < x < \frac{1}{n} \\ 1 & x \geq \frac{1}{n} \end{cases}.$$

Then the sequences of functions converges pointwise to the limit function

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

So that the sequence of continuous functions need not converge to a continuous limit function.

Example 7.3. Let $\{q_n\}$ be an enumeration of the rational numbers in $[0,1]$. Then define a sequence of functions $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 & x \in \{q_1, ..., q_n\} \\ 0 & \text{else} \end{cases}.$$

Then each f_n is integrable as it is discontinuous at only finitely many points and $\int_0^1 f_n dx = 0$ or all n so that $\lim_{n\to\infty} \int_0^1 f_n dx = \lim_{n\to\infty} 0 = 0$. However the limit function $f : [0,1] \to \mathbb{R}$ is Dirichlet's function! We know this function is not even integrable and so clearly the limit of the integrals is not the integral of the limit!

Example 7.4. Finally, we show derivatives fail. Let $f_n : [0,1] \to \mathbb{R}$ be defined by $f_n = \frac{x^n}{n}$. Then each f_n is differentiable with derivative $f_n' = x^{n-1}$, so that at $x = 1$ we have

$\lim_{n\to\infty} f_n'(1) = \lim_{n\to\infty} 1^{n-1} = 1$. The original sequence converges to the function

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

So that f is not even continuous at the point $x = 1$ and so clearly we do not have

$$\lim_{n\to\infty} f_n' = f'.$$

The reader should wonder why they were allowed to interchange limits, derivatives, etc in Calculus 2 but now everything is failing. Was all the work performed in such as class invalid? At this point we have two options: we can give up the concept of sequences of functions, or we can try to see what is needed to make the above work! Obviously, we choose the latter and it turns out that for most results what we need is the stronger notion of convergence, namely Uniform Convergence!

7.2 Uniform Convergence

The name Uniform Convergence should sound familiar to you, for it has a similar name to uniform continuity! Recalling uniform continuity the different between that and stronger continuity is that the same δ worked for all points in our domain. But now we can apply this idea to our new problem. In the scenario of pointwise convergence the sequence of functions must converge at every point, but some points may converge very quickly, while others quite slowly. If we state the exact definitions we would say that: a sequence of functions $f_n : X \to \mathbb{C}$ converges pointwise to a limit function f if for all $x \in X$, for all $\epsilon > 0$ there exists an N so that for all $n \geq N$ we have $|f_n(x) - f(x)| < \epsilon$. The key point is that the value of N required to make the functions close to the limit function depend on both the value of ϵ and the value of x! But now we see how giving this scenario uniform treatment might work.

Definition 7.5. Let X be a topological space and let $f_n : X \to \mathbb{C}$ be a sequence of functions. We say that f_n **Uniformly Converges** to a limit function f if for all $\epsilon > 0$ there exists some $N \in \mathbb{N}$ so that for all $x \in X$ and for all $n \geq N$ we have that

$$|f_n(x) - f(x)| < \epsilon.$$

The important distinction here is that the value of N depends only on ϵ and must work for all x in the domain, similarly to how for uniform continuity the same δ needed to work for all x in the domain! This important improvement will allow us to prove very powerful results. It is also clear that if a sequence f_n converges uniformly to a limit f that it also converges pointwise! Our first proof involving uniform continuity will be another so called Cauchy Criterion!

Theorem 7.166 (Cauchy Criterion for Sequences of Functions) Let X be a topological space and let $f_n : X \to \mathbb{C}$ be a sequence of functions. Then f_n converges uniformly to a limit function if and only if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for all $m, n \geq N$ and for all $x \in X$ we have $|f_n(x) - f_m(x)| < \epsilon$.

Proof. If f_n converges uniformly Cauchyness is obvious.
The proof of the other direction essentially follows from the Cauchy Criterion of $\mathbb{C}$. Assume the Cauchy Criterion and fix some $x \in X$. It is clear then that the sequence of points $f_n(x)$ is a Cauchy sequence and so it converges to some point x'. If we just set $f(x) = x'$ and its clear that $f_n \to f$. Now we need to show this is uniform. It's clear that for a fixed n that $\lim_{m\to\infty} |f_n(x) - f_m(x)| = |f_n(x) - f(x)|$ for all $x \in X$, so that given $\epsilon > 0$ we can choose N so that for all $x \in X$ we have $|f_n(x) - f(x)| < |f_n(x) - f_m(x)| + \frac{\epsilon}{2}$ and so that $|f_n(x) - f_m(x)| < \frac{\epsilon}{2}$ completing the proof. $\square$

For series of functions this allows for a very easy and useful uniform convergence proof!

Theorem 7.167 (Weierstrass M-Test) Let $f_n : X \to \mathbb{C}$ be a sequence of functions so that there exists a sequence of positive real numbers M_n satisfying $|f_n| \leq M_n$ for all n and for all $x \in X$. If $\sum M_n$ converges then $\sum f_n$ converges uniformly.

Proof. Let $\epsilon > 0$ and choose N so that for all $m > n \geq N$ we have $\sum_{k=n}^{n} M_k < \epsilon$. Then if $S_n(x) = \sum_{k=1}^{n} f_k(x)$ we have for all $m > n \geq N$ and for all $x \in X$ that $|S_n(x) - S_m(x)| \leq \sum_{k=n}^{n} |f_n(x)| \leq \sum_{k=n}^{m} M_k < \epsilon$ so that the result follows from the Cauchy Criterion. $\square$

7.3 Properties of Uniform Convergence

We now show that uniform convergence can indeed preserve the important properties we would hope! It is also worth noting that while all of the following proofs are for sequences, they will generalize to series by considering a series as a limit of partial sums!

Theorem 7.168 (Uniform Continuity and Limits) Let $f_n : X \to \mathbb{C}$ that converges uniformly to a limit function f. If $\lim_{x\to a} f_n(x) = a_n$ exists for all n and $\lim_{n\to\infty} a_n = a'$ also exist, then

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

We can rewrite the above to get that

$$\lim_{x\to a} \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \lim_{x\to a} f_n(x),$$

or in words the order we evaluate the limits does not matter as long as we have uniform convergence!
Note we are implicitly requiring that a be a limit point of X.

Proof. Let $\epsilon > 0$ and let $N(a) \subseteq X$ be some neighborhood of a. Choose $N \in \mathbb{N}$ so that for all x and for all $n \geq N$ we have
(1) $|f_n(x) - f(x)| < \epsilon$,

(2) $|a_n - a'| < \epsilon$, and

(3) $x \in N(a) \implies |f_n(x) - a_n| < \epsilon$.

We can satisfy part (1) by uniform convergence, (2) by the fact $a_n \to a$, and (3) since $\lim_{x \to a} f_n(x) = a_n$. Then it follows that if $x \in N(a)$ then the triangle inequality gives

$$|f(x) - a'| \leq |f(x) - f_n(x)| + |f_n(x) - a_n| + |a_n - a'| < 3\epsilon.$$

This shows that $f(N(a)) \subseteq B_{3\epsilon}(a')$. Since $N(a)$ was arbitrary we see that

$$\lim_{x\ toa} f(x) = a',$$

and so we are done. □

We have a few notes about the last proof. Firstly, it's important to brush up on topological definitions of limits to make sure the neighborhood argument made sense. Secondly, normally in a proof such as the above we would find three distinct N_1, N_2, and N_3 satisfying (1), (2), and (3) and then make N the largest. The reader has gotten to an advanced enough point this can be done implicitly and does not need to be directly stated, but the reader should be aware why this is possible. Finally, some may be confused why we allowed $|f(x) - a'| < 3\epsilon$ instead of ϵ. But this is allowed as any multiple of epsilon can work and again if we wanted we could go in and demand each thing be less than $\frac{\epsilon}{3}$. But again as the reader has reached a higher maturity these tiny details can be more or less ignored and understand that this is allowed. As an easy corollary we have the following very famous and important theorem!

Theorem 7.169 (Uniform Limit Theorem) Let $f_n : X \to \mathbb{C}$ converge uniformly to f. If each f_n is continuous at $x \in X$, then f is continuous at x. In particular if each f_n is continuous, then so is the limit function f.

Proof. Suppose for all n that f_n is continuous at the point x. If x is not a limit point there is nothing to prove. If x is a limit point then $\lim_{t\to x} f_n(t) = f_n(x)$. Since $\lim_{n\to\infty} f_n(x) = f(x)$, the above theorem gives that $\lim_{t\to x} f(t) = f(x)$ and so we are done. $\square$

We now cover integration!

Theorem 7.170 (Uniform Convergence and Integration) Let $f_n : [a,b] \to \mathbb{C}$ converge uniformly to f. If each f_n is integrable, then so is f and

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

This shows we can change the order of limits and integration if convergence is uniform!

Proof. We prove this for real integrals, the general case follows as complex integrals are defined in terms of real integrals. Let $D(f_n)$ be the set of all discontinuities for the function f_n. For each n since f_n is integrable this is a set of measure zero and since there are a countable collection of f_n it follows that $D = \bigcup_{n=1}^{\infty} D(f_n)$ is also a set of measure zero. On $[a,b] - D$ we know that each f_n is continuous and so by the Uniform Limit Theorem we have f is continuous on $[a,b] - D$. Therefore, f is continuous almost everywhere on $[a,b]$ and so it is integrable! Now let $\epsilon > 0$, by uniform convergence choose $N \in \mathbb{N}$ so that for all $n \geq N$ and for all $x \in [a,b]$ $|f_n(x) - f(x)| < \epsilon$. Then we have that for all $n \geq N$

$$\left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right| = \left|\int_a^b (f_n(x) - f(x))\, dx\right| \leq \int_a^b |f_n(x) - f(x)|\, dx < \epsilon(b-a),$$

and so we can conclude that

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

and so we are done. $\square$

At this point the reader might expect a theorem essentially stating the same as the above for differentiation. It turns out that in general, not even uniform convergence preserves differentiation. In general integrability is a rather weak property and we will see when discussing Lebesgue integration later on that it can be preserved even with pointwise convergence in some cases! Continuity is a stronger property than integrability so that we need something stronger than pointwise convergence to preserve it and uniform convergence does the trick. But for differentiation, this is such a strong and demanding property that not even uniform convergence can preserve it in every case. The next theorem is really the best we can do, it won't guarantee limit functions be differentiable, but it will give us a scenario where interchanging limits and derivatives is allowed.

Theorem 7.171 (Uniform Convergence and Differentiation) Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of differentiable functions. If there exists some $p \in [a, b]$ so that $f_n(p)$ converges and the sequence of derivatives f_n' converges uniformly to some function g, then f_n converges uniformly to a differentiable function f and

$$f' = g = \lim_{n\to\infty} f_n'.$$

Proof. Let $\epsilon > 0$. We must first show uniform convergence of f_n. We know there exists some $N \in \mathbb{N}$ so that for all $m, n \geq N$ we have that $|f_n(p) - f_m(p)| < \epsilon$ and so that for all $x \in [a, b]$ $|f_n'(x) - f_m'(x)| < \epsilon$, by convergence of $f_n(p)$ and uniform convergence of the derivatives. Now consider the function $f_n - f_m$. This is differentiable on $[a, b]$ so that the mean value theorem tells us that for all $x_1, x_2 \in [a, b]$ we have $|f_n(x_1) - f_m(x_1) - f_n(x_2) + f_m(x_2)| \leq \epsilon(b - a)$. But now we have that for all $x \in [a, b]$ that

$$|f_n(x) - f_m(x)| \leq |f_n(x) - f_m(x) - f_n(p) + f_m(p)| + |f_n(p) - f_m(p)| < \epsilon(b - a + 1).$$

This shows that the sequence f_n converges uniformly by the Cauchy Criterion.

Now we must show f is differentiable and $f'(x) = g(x)$. Let $x \in [a, b]$. We then have that for $n \in \mathbb{N}$ that

$$\left|\frac{f(x+h)-f(x)}{h}-g(x)\right| \leq$$

$$\left|\frac{f(x+h)-f(x)-f_n(x+h)+f_n(x)}{h}\right| + \left|\frac{f_n(x+h)-f_n(x)}{h}-f_n'(x)\right| + |f_n'(x)-g(x)|\,.$$

Now we must show that we can force all of the above to be as small as needed. Clearly, we can choose N so that $n \geq$ gives $|f_n'(x)-g(x)|$ is small. Similarly, we can choose h so that $\left|\frac{f_n(x+h)-f_n(x)}{h}-f_n'(x)\right|$ is small by definition of the derivative. So we have one thing left to show. Given $\epsilon > 0$ we choose $m, n \geq N$ so that for all x we have $|f_n'(x)-f_m'(x)| < \epsilon$. Now we have the function $f_n(x+h)-f_m(x+h)$ is differentiable on $[x, x+h]$ so that the MVT gives a point $c \in (x, x+h)$ so that $|\frac{f_n(x+h)-f_m(x+h)-f_n(x)-f_m(x)}{h} = |f_n'(c)-f_m'(c)| < \epsilon$. But now we see all three conditions can be satisfied at the same time and so we are done. $\square$

The two conditions we had above were that $f_n(x)$ needed to converge for some point and that the sequence of derivatives converge uniformly. These are indeed necessary as consider the following two examples.

Example 7.12. Consider the sequence of functions $f_n : [a,b] \to \mathbb{R}$ by $f_n(x) = n$. Then each f_n is differentaible with derivative $f_n'(x) = 0$ so that $f_n' \to g(x) = 0$ uniformly, but the original functions f_n converge at not point and so certainly do not converge to a differentiable function!

Example 7.13. Assuming some knowledge of the sin and cos functions, consider the series of functions

$$f(x) = \sum_{n=1}^{\infty} a^n \cos(b^n \pi x),$$

where $0 < a < 1$ and b is an odd integer where $ab > 1 + \frac{3\pi}{2}$. This function is called the Weierstrass Function. It converges uniformly by the M-Test and is continuous as a uniform limit of continuous functions. But the series of derivatives is of the form

$$\sum_{n=1}^{\infty} -a^n b^n \pi \sin(b^n \pi x) = \sum_{n=1}^{\infty} -(ab)^n \pi \sin(b^n \pi x),$$

which diverges for say $x = 1$ and so clearly the sequence of derivatives do not converge uniformly. Whats more interesting however is that as the following theorem will show the Weierstrass Function is not differentiable at any point! The proof of this claim is a little complicated and involves a great deal of trig identities, while not being very important to our overall progression, and so the proof is excluded. Intuitively, this claim can be seen as the "derivative" given above wildly oscillates and increases in size. For a nice proof see this article:xxx.

7.4 The $\mathcal{C}$(X) Banach Space

We now turn our attention away from sequences of functions to discuss more general collections of functions. We will need to introduce some new terminology that will be used throughout the rest of the chapter.

Definition 7.14. Let $\mathcal{F}$ be a set (or commonly called a collection or family) of functions from a domain X to $\mathbb{C}$. Then we say that $\mathcal{F}$ is **Pointwise Bounded** if there exists some function $g : X \to \mathbb{R}$ so that for all $x \in X$ and all $f \in F$ we have $|f(x)| < g(x)$. We say that $\mathcal{F}$ is **Uniformly Bounded** if there exists some number $M \in \mathbb{R}$ so that for all $x \in X$ and for all $f \in F$ $|f(x)| < M$.

Example 7.15. The sequence of functions $f_n(x) = \cos(nx)$ is uniformly bounded as
$|\cos(nx)| < 1$.
The sequence of functions $g_n(x) = \frac{1}{n}x$ is pointwise bounded as we have $|\frac{1}{n}x| < x + 1$.
The sequence of functions $h_n(x) = n$ is neither uniformly bounded or pointwise bounded.

These different notions of being bounded have different uses but it is clear that being uniformly bounded is a stronger condition than pointwise bounded. One nice application is the following, a sort of monotone convergence theorem for sequences of functions.

Theorem 7.16. Suppose $f_n : X \to \mathbb{R}$ is a sequence of functions that is pointwise bounded, and monotonic in the sense that for all $x \in X$ we have $f_n(x) \leq f_{n+1}(x)$. Then f_n converges pointwise to some function f.

Proof. This follows by applying the monotone convergence theorem of real sequences at each point of x with the sequence $f_n(x)$. □

In Chapter 5 on Differentiation we discussed the space $\mathcal{C}([a, b], \mathbb{R})$ of continuous functions. We expand this slightly below.

Definition 7.17. If X, Y are topological space and $f : X \to Y$, then we say that $f \in \mathcal{C}(X, Y)$ if f is continuous. If $Y = \mathbb{C}$ or $Y = \mathbb{R}$ we simplify the notation so that $\mathcal{C}(X, Y) = \mathcal{C}(X)$ and require $f \in \mathcal{C}(X)$ to be bounded. This agrees with our definition in Chapter 5 as $[a, b]$ was compact. Note context will tell us if $Y = \mathbb{C}$ or $\mathbb{R}$.

Currently, the space $\mathcal{C}(X)$ lacks structure, but we will soon endow it with a norm turning it into a Normed Space, and therefore a Metric Space as well!

Definition 7.18. Let $f : X \to \mathbb{C}$, be bounded. We define the **Supremum Norm** of f on X as

$$||f||_\infty = \sup_{x \in X} |f(x)|.$$

It is clear that for all $f, g \in \mathcal{C}(X)$ we have

$||f||_\infty \geq 0$,

$||f||_\infty = 0$ if and only if $f(x) = 0$,

If $a \in \mathbb{C}$ then $||af||_\infty = |a| ||f||_\infty$, and the triangle inequality of real numbers gives that $||f+g||_\infty \leq ||f||_\infty + ||g||_\infty$, so that the supremum norm is indeed a norm for $\mathcal{C}(X)$, turning the pair $(\mathcal{C}, ||\cdot||_\infty)$ into a normed space, and in particular also into a metric space! Note that being bounded in $\mathcal{C}(X)$ is equivalent to being uniformly bounded and if we simply say bounded this is what is implied.

To distinguish notation if $f_n \to f$ with respect to the sup norm we will denote this by $f_n \to_\infty f$. We can now view the set of continuous functions as a metric space and ask questions about dense subsets, compact subsets, and more. One very important tool is the following: if $f_n \in \mathcal{C}(X)$ then $f_n \to_\infty f$if and only if $f_n \to f$ uniformly. This is easy to see just by comparing definitions. But this observation gives rise to the following very important theorem.

Theorem 7.179 ($\mathcal{C}(X)$ is complete) Let $f_n \in \mathbb{C}(X)$ be a Cauchy sequence (w.r.t $\mathcal{C}(X)$). Then f_n converges to a limit function $f \in \mathcal{C}$. In particular the normed space $\mathcal{C}(X)$ is a Banach Space.

Proof. If f_n is a Cauchy sequence (w.r.t the sup norm), then clearly it is also Cauchy in the sense of definition 7.XXX above. So that the Cauchy Criterion Theorem 7.XX tells us that f_n converges uniformly to some function f. But then by the Uniform Limit Theorem we have that f is continuous so that $f \in \mathcal{C}(X)$, and finally since $f_n \to f$ uniformly, we have that $f_n \to_\infty f$ and so we are done. $\square$

7.5 Arzela-Ascoli Theorem

Moving along as promised we now want to discuss what subsets of $\mathcal{C}(X)$ are compact! We now that if a subset $S \subseteq \mathcal{C}(X)$ is compact then it is closed and bounded, but is the converse true here as it is in $\mathbb{R}^n$? No!

Example 7.20. Consider the open ball $B_1(0) \subseteq \mathcal{C}(X)$, this set is closed and bounded. Consider the sequence of functions $f_n : [0,1] \to \mathbb{R}$ given in example 7.XXX above. Each

$f_n \in B_1(0)$ and it converges pointwise to a limit function f. Clearly, this convergence is not uniform as the limit function is not continuous, and so the sequence $f_n \not\to_\infty f$! But we also have every subsequence $f_{n_k} \to f$ pointwise, but not uniform, so that $f_{n_k} \not\to_\infty f$. Since no subsequence converges $B_1(0)$ is NOT sequentially compact and therefore not compact!

One might think that there is in general to characterization of compactness in $\mathcal{C}(X)$ like the Heine-Borel theorem gave us for $\mathbb{R}^n$. But infact this is not the case! It turns out we can completely characterize compactness and this is exactly what the Arzela-Ascoli theorem this section is named after will do! We will just need a few new tools, namely Equicontinuity, and a few more restrictions in place!

Definition 7.21. Let $\mathcal{F}$ be a collection of functions from a Topological space X to $\mathbb{C}$. Given $x \in X$ we say that $\mathcal{F}$ is **Equicontinuous at** x if for all $\epsilon > 0$ there exists a neighborhood $N(x) \subseteq X$ so that for all $y \in N(x)$ and for all $f \in \mathcal{F}$ we have $|f(x) - f(y)| < \epsilon$. If $\mathcal{F}$ is equicontinuous at every $x \in X$ we say that $\mathcal{F}$ is **Equicontinuous**. It is clear that this implies that every $f \in \mathcal{F}$ is continuous and so we have that $\mathcal{F} \subseteq \mathcal{C}(X)$.

The next theorem gives a hint of the relationship between convergence in $\mathcal{C}(X)$ and equicontinuity.

Theorem 7.22. If $f_n \subseteq \mathcal{C}(X)$ is a sequence of functions and $f_n \to_\infty f$ and K is compact, then the collection $\{f_n | n \in \mathbb{N}\}$ is equicontinuous.

Proof. Let $x \in X$ and let $\epsilon > 0$. Since $f_n \to_\infty f$ there exists an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $||f_n - f||_\infty < \epsilon$. The function f is continuous by Uniform Limit Theorem, so let $N(x)$ be a neighborhood of x so that $y \in N(x) \implies |f(x) - f(y)| < \epsilon$. Then for all $y \in N(x)$ and for all $n \geq N$ we have

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f(x)| + |f(x) - f(y)| + |f(y) - f_n(y)|$$

$$\leq 2||f_n - f||_\infty + |f(x) - f(y)| < 3\epsilon.$$

Now if $k < N$ we know that f_k is continuous so that there is some neighborhood $N(x))_k$ so that for all $y \in N(x)_k$ we have $|f_k(x) - f_k(y)| < \epsilon$. The intersection of all $N(x)_1 \cap ... \cap N(x)_{N-1} \cap N(x) = U$ is a finite intersection of open sets, and so it is a neighborhood of x and for all $y \in U$ and for all $n \in \mathbb{N}$ we see that $|f_n(x) - f_n(y)|$ is small, and so $\{f_n\}$ is equicontinuous at x. Since x was arbitrary we see that $\{f_n\}$ is equicontinuous. □

We now want to tackle the Arzela-Ascoli Theorem. The next few lemmas will be related to equicontinuity and will be used in the proof of the Arzela-Ascoli Theorem.

Lemma 7.23. If $S \subseteq \mathcal{C}(X)$ is compact, then S is equicontinuous.

Proof. Let $x_0 \in X$ and let $\epsilon > 0$. For each $f \in S$ consider the open ball $B_\epsilon(f)$. Clearly the collection $\{B_\epsilon(f) | f \in S\}$ is an open cover for S so that there is some finite subcover, $B_\epsilon(f_1), ..., B_\epsilon(f_n)$. For a fixed $1 \leq i \leq n$ since f_i is continuous at x_0 there is some neighborhood $N_i(x_0)$ so that for all $x \in N_i(x_0)$ we have that $|f_i(x_0) - f_i(x)| < \epsilon$. But now given any function $f \in S$ we have that $f \in B_\epsilon(f_k)$ for some k so that

$$|f(x_0) - f(x)| \leq |f(x_0) - f_k(x_0)| + |f_k(x_0) - f_k(x)| + |f_k(x) - f(x)| < 3\epsilon,$$

and so we see that S is equicontinuous at x_0, and therefore equicontinuous. □

Lemma 7.24. If K is a compact space and $S \subseteq \mathcal{C}(K)$ is equicontinuous and pointwise bounded, then S is uniformly bounded, or bounded w.r.t the sup-norm.

Proof. For each $x \in K$ by equicontinuity there exists some neighborhood $N(x)$ so that for all $y \in N(x)$ and for all $f \in S$ we have that $|f(x) - f(y)| < 1$. The collection $\{N(x)\}$ is an open cover for K and so by compactness we have a finite subcover $N(x_1), ..., N(x_n)$. Since S is pointwise bounded there exists some function g so that for all $f \in S$ we have $|f(x)| < g(x)$. In particular this holds at each point x_i, $i = 1, ..., n$. Take $M = \max\{g(x_1), ..., g(x_n)\} + 1$, then for all $f \in S$ and for all $x \in X$ we have that $x \in N(x_i)$ for some i so that

$$|f(x)| \leq |f(x_i)| + 1 < M,$$

so that S is uniformly bounded. $\square$

The last lemma gives a result similar to the Arzela-Ascoli thereom but in a very peculiar setting.

Lemma 7.25. Suppose E is a countable set and f_n is a sequence of functions in $\mathcal{C}(E)$. Then there exists some subsequence f_{n_k} that converges pointwise on E.

Proof. Let x_n be the sequence of all points of E and let g be the pointwise bound for $\{f_n\}$. Then we know that for all n we have $|f_n(x_1)| < g(x_1)$ so that the sequence $f_n(x_1)$ is a bounded sequence of complex numbers, and therefore has a convergent subsequence we will call $f_{1,n}(x_1)$ that converges. But now the collection of points $\{f_{1,n}(x_2)\}$ is bounded, so that there is some subsequence $f_{2,n}(x_2)$ that converges. Inducting we assume we have $f_{k,n}$ chosen so that $f_{k,n}(x_k)$ converges, then again we know that $f_{k,n}(x_{k+1})$ is a bounded sequence, so that there is some subseqeuence that converges we call $f_{k+1,n}$. Lets arrange our sequences in a grid as follows:

$$\begin{array}{cccc} f_{1,1} & f_{1,2} & f_{1,3} & \cdots \\ f_{2,1} & f_{2,2} & f_{2,3} & \cdots \\ f_{3,1} & f_{3,2} & f_{3,3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array}$$

Now we form a new subsequence of our original sequence by letting $f_{n_k} = f_{k,k}$. Then for each $x_i \in E$ we know that for all $k \geq i$ we have that f_{n_k} is a subsequence of $f_{i,n}$ so that $f_{n_k}(x_i)$ converges. So that as $n \to \infty$ f_{n_k} converges for every point in E. $\square$

We are now ready to prove the very important Arzela-Ascoli Theorem!

Theorem 7.26 (The Arzela-Ascoli Theorem)**.** Let K be a Compact Space with a countable dense set. Then a subset $S \subseteq \mathcal{C}(K)$ is compact if and only if S is equicontinuous, closed, and pointwise bounded. Since every second countable space has a countable dense set this holds if K is a compact second countable spaces, and in particular if K is a compact metric space!

Proof. For the first direction assume S is compact. Then S is closed and bounded and so clearly it is also pointwise bounded. Equicontinuity follows from Lemma 7.XXX above completing the first direction!

For the second direction we assume that S is closed, pointwise bounded, and equicontinuous. We prove that S is sequentially compact. Let f_n be a sequence of functions in S and let E be a countable dense subset of K. We can find a subsequence f_{n_k} that converges for every point in E, by Lemma 7.XX above. We now want to show this subsequence actually converges for every point in K and it converges uniformly. Let $\epsilon > 0$, then for each $x \in K$ by equicontinuity there is some neighborhood $N(x)$ so that for all $f \in S$ and all $y \in N(x)$ we have $|f(x) - f(y)| < \epsilon$. By compactness of K we can choose a finite subcover $N(x_1), ..., N(x_j)$. Since E is dense for each $i = 1, ..., j$ there exists a point e_i so that $e_i \in N(x_i)$, and now we can choose an $N \in \mathbb{N}$ so that for all $i = 1, ..., j$ and for all $m, n \geq N$ we have $|f_{n_n}(e_i) - f_{n_m}(e_i)| < \epsilon$. But now given any $p \in K$ we know that $p \in N(x_i)$ for some i, so that we have for all $m, n \geq N$

$$|f_m(p) - f_n(p)| \leq |f_m(p) - f_m(e_i)| + |f_m(e_i) - f_n(e_i)| + |f_n(e_i) - f_n(p)| < 3\epsilon.$$

Since p was arbitrary this holds for all $p \in K$ so that f_{n_k} is a Cauchy sequence and therefore converges uniformly, so that the sequence $f_{n_k} \to_\infty f$, or f_n has a uniformly convergent subsequence (or just convergent w.r.t. the sup metric). Therefore, S is sequentially compact, and therefore since $\mathcal{C}(K)$ is a metric space, S is compact. $\square$

7.6 Stone Weierstrass Theorem

Having discussed Compact subsets of $\mathcal{C}(X)$ we want to turn our attention to dense subsets. The next very easy theorem shows why this is an important discussion.

Theorem 7.27. If $S \subseteq \mathcal{C}(X)$ is dense, then given any $f \in \mathcal{C}(X)$ there is a sequence f_n of functions in S so that $f_n \to_\infty f$.

Proof. This almost follows immediately from definitions. If $f \in \mathcal{C}(X)$ then since S is dense we have that f is a limit point of S so that there is some sequence in S converging to f. $\square$

In other words, this theorem is useful as we can arbitrarily approximate any continuous function if we have a dense subset! It will turn out that the polynomials are dense in $\mathcal{C}([a,b])$! This is especially nice as polynomials are the nicest functions we can work with and this result is called The Weierstrass Aprproximation Theorem and was proven by Weierstrass. The famous Stone Weierstrass Theorem, that this section is dedicated to proving will be more a general and powerful theorem that gives us a nice criterion for which sets are dense in $\mathcal{C}(X)$! It was proven by Stone some time after Weierstrass's original theorem and it turns out that to prove Stone-Weierstrass we will need can use Weierstrass's original result to aid us!

Definition 7.28. If f, g are integrable functions the **Convolution at x** of f and g is $(f * g)(x) = \int_{-\infty}^{\infty} f(x-t)g(t)\,dt$, provided the indefinite integral exists, for our use we will often restrict this so that $(f * g)(x) = \int_{-1}^{1} f(x-t)g(t)\,dt$. The **Dirac Delta "Function"** $\delta(x)$ is defined so that if $x \neq 0$ $\delta(x) = 0$ and if $x = 0$ then $\delta(x) = \infty$, and to also satisfty the requirement that $\int_{-\infty}^{\infty} \delta(x)\,dx = 1$. Of course no function satisfies these conditions hence why it was in quotations. The Dirac Function as a concept however, is incredibly useful for convolution as given any function $f(x)$ then

$$(f * \delta)(x) = \int_{-\infty}^{\infty} f(x-t)\delta(t)\,dt = f(x)\int_{-\infty}^{\infty} \delta(t)\,dt = f(x)!$$

The reason we care about any of this is that given any continuous function, what we plan to do is Convolve it with a sequence of functions that behave similarly to the Dirac Function! This convolution will be a polynomial function, so that as our sequence goes to infinity the convolution looks much more like the original function and uniformly approximates the function!

We define the **Landau Kernel** is defined as $K_n(x) = c_n(1-x^2)^n$, where c_n normalizes the Kernel in the sense that $c_n = \int_{-1}^{1}(1-x^2)^n\,dx$. As $n \to \infty$ $K_n(x)$ will look more and more like the Delta function and behave more like it. We have that

$$\frac{1}{c_n} = \int_{-1}^{1}(1-x^2)^n\,dx = 2\int_0^1(1-x)^n(1+x)^n\,dx \geq 2\int_0^1(1-x)^n\,dx = \frac{2}{n+1},$$

so that $c_n \leq \frac{n+1}{2} \leq n$.

It is easy to see that $K_n(x)$ is always a polynomial, so that we also have for a continuous function $f : [-1,1] \to \mathbb{R}$ that

$$(K_n * f)(x) = (a_0 + a_1 x + ... + a_n x^n * f)(x) = \int_{-1}^{1} a_0 + a_1(x-t)f(t) + ...a_n(x-t)^n f(t)\,dt.$$

So that the above is a polynomial function for the variable x.

We now use all of theses tools to prove the following important theorem!

Theorem 7.189 (The Weierstrass Approximation Theorem) If $f : [a,b] \to \mathbb{R}$ is continuous, then there exists a sequence of polynomials P_n so that $P_n \to f$ uniformly on $[a,b]$.

Proof. We prove the case where $f : [-1,1] \to \mathbb{R}$. We claim that the sequence of polynomials $(K_n * f)(x)$ will do the trick! Let $\epsilon > 0$. Choose δ so that for all $x, y \in [-1,1]$ if $|x-y| < \delta$, then $|f(x) - f(y)| < \epsilon$, by uniform continuity of f. Let $M = \max|f(x)|$, which exists as f is continuous. Let $x \in [-1,1]$. Since $\int_{-1}^{1} K_n(t)\,dt = 1$ we immediately have that

$f(x) = f(x)\int_{-1}^{1} K_n(t)\,dt = \int_{-1}^{1} f(x)K_n(t)\,dt$. So now we have that

$$|f(x) - (K_n * f)(x)| = \left|\int_{-1}^{1} f(x)K_n(t) - f(x-t)K_n(t)\,dt\right| \leq \int_{-1}^{1} |f(x) - f(x-t)|K_n(t)\,dt.$$

We can split this integral up to equal

$$\int_{-1}^{-\delta} |f(x) - f(x-t)|K_n(t)\,dt + \int_{-\delta}^{\delta} |f(x) - f(x-t)|K_n(t)\,dt + \int_{\delta}^{1} |f(x) - f(x-t)|K_n(t)\,dt. \tag{7.1}$$

For the outside integrals we see that

$$\int_{\delta}^{1} |f(x)-f(x-t)|K_n(t)\,dt \leq 2M\int_{\delta}^{1} K_n(t)\,dt \leq 2Mn\int_{1}^{\delta} (1-x^2)^n\,dt \leq 2Mn(1-\delta^2)^n(1-\delta).$$

The last inequality follows from the fact that $K_n(t)$ is decreasing on $[\delta, 1]$ so that $(1-\delta^2)^n$ is an upper bound for the function on this interval. Since $(1-\delta^2) < 1$ we have that $\lim_{n\to\infty} n(1-\delta^2)^n = 0$ so that the outer integrals in (X) can be made as small as desired. Now for the inner integral we have

$$\int_{-\delta}^{\delta} |f(x) - f(x-t)K_n(t)\,dt \leq \epsilon\int_{-\delta}^{\delta} K_n(t) \leq \epsilon,$$

so that the inner integral can be made small!

Since all three integrals in (X) can be made as small as desired and x was an arbitrary point of $[-1, 1]$ we see that $(K_n * f)(x) \to f$ uniformly. □

The case where $[a, b] \neq [-1, 1]$ follows easily from just changing our bounds of integration and using a modified Kernel, one that is positive on $[a, b]$, but the numbers will not be as nice! We can reinterpret the above theorem by seeing that the set of Polynomials $P \subseteq \mathcal{C}([a, b])$ is dense! While a great result again we would like to expand this to more general settings. The Stone-Weierstrass will extend this result to more general domains, specifically Compact

Metric Spaces, and to broader collections of functions. More specifically we will consider a special classes of functions called Algebras of Functions.

Definition 7.30. Let $\mathcal{A} \subseteq \mathcal{C}(X)$ be a collection of functions. We say that $\mathcal{A}$ is a **Algebra** of functions if for all $f, g \in \mathcal{A}$ and for all scalars c, we have that $f + g \in \mathcal{A}$, $f \cdot g \in \mathcal{A}$, and $cf \in \mathcal{A}$. In other words an Algebra of functions is closed under addition, multiplication, and scalar multiplication. If our Algebra is of real functions we will of course require the scalars be real. The set of polynomials on $[a, b]$ clearly form an Algebra. In fact there is a close relationship between polynomials and Algebras. For if $f \in \mathcal{A}$, then the new function $a_1 f(x) + a_2 f^2(x) + ... + a_n f^n(x) \in \mathcal{A}$, where the new function looks like a polynomial of f functions! This it turns out is why we needed the Weierstrass Approximation Theorem as we will soon see!

Definition 7.31. Before turning to Stone-Weierstrass, we need to discuss the important properties that an will enable an Algebra to be dense in $\mathcal{C}(X)$. Let X be a space and $\mathcal{A} \subseteq \mathcal{C}(X)$ an Algebra. Then $\mathcal{A}$ **Vanishes at No Point of X** if for all $x \in X$, there exists an $f \in A$ so that $f(x) \neq 0$. If $\mathcal{A}$ did vanish at some point then it would certainly be impossible to approximate every possible continuous function! The other big property is this: we say $\mathcal{A}$ **Separates Points** if given $x, y \in X$ then there exists a function $f \in \mathcal{A}$ so that $f(x) \neq f(y)$. If $\mathcal{A}$ did not separate points, then it certainly could not approximate every function as there would be some pair of points that always got mapped to the same value. It turns out that while these properties are necessary, they are also sufficient! We will see that this is exactly what we need for Stone-Weierstrass!

We now prove some Lemmas to assist us with Stone-Weierstrass!

Lemma 7.32. Let K be a compact metric space and let $\mathcal{A} \subseteq \mathcal{C}(K)$ be an Algebra of real functions that separates points and vanishes at no point and let $f \in \mathcal{C}(K)$. If f_n is a sequence of functions in $\mathcal{A}$ so that $f_n \to_\infty f$, then there is a sequence of functions $g_n \in \mathcal{A}$ so that $g_n \to_\infty |f|$.

Proof. Since K is compact it follows that since f is continuous then $|f|$ has a maximum, say M. Let f_n be a sequence in $\mathcal{A}$ so that $f_n \to_\infty f$. By Weierstrass Approx. Theorem there is a sequence of polynomials P_n so that $P_n \to |x|$ on $[-M, M]$. We may assume that $P_n(0) = 0$ for all n, for if it doesn't then the new function $P_n(x) - P_n(0)$ is still a polynomial, and it still uniformly converges to $|x|$. We then have that given $\epsilon > 0$ choose N so that for all $y \in [-M, M]$ and for all $x \in K$ that $|P_n(y) - |y|| < \epsilon$ and $|f_n(x) - f(x)| < \epsilon$ when $n \geq N$. It follows that if we set $g_n(x) = P_n(f_n(x))$, then $g_n \in \mathcal{A}$ as it is a polynomial of f (with no constant term) and that $|g_n(x) - |f(x)|| < \epsilon$. So that $g_n \to_\infty |f|$. $\square$

Corollary 7.32.1. If f, g are continuous real functions an there are sequences f_n and g_n in $\mathcal{A}$ that uniformly converge to f and g, then there are sequences that converge to $\max(f, g) = \frac{|f-g|+f+g}{2}$ and $\min(f, g) = \frac{f+g-|f-g|}{2}$.

Lemma 7.33. Let $\mathcal{A} \subseteq \mathcal{C}(X)$ be an algebra of functions, that vanishes at no point and separates points. Then if $x_1, x_2 \in X$ and c_1, c_2 scalars then there exists a function $f \in \mathcal{A}$ so that $f(x) = c_1$ and $f(y) = c_2$. If $\mathcal{A}$ is a collection of real functions the scalars of course are real.

Proof. By separating points there is some function $h \in \mathcal{A}$ so that $h(x_1) \neq h(x_2)$. WLOG assume that $h(x_1) \neq 0$ (if it does relabel x_1 and x_2 as in this case $h(x_2) \neq 0$). If $h(x_2) \neq 0$ let $j(x) = h(x)$. If $h(x_2) = 0$, then by vanishing at no point there is some function $g \in \mathcal{A}$ so that $g(x_2) \neq 0$ in which case we set $j(x) = h(x) + kg(x)$ where k is any number guaranteeing $g(x_1)k \neq h(x_1)$. Then in either case we see that $j \in \mathcal{A}$ and j satisfies that $j(x_1) \neq j(x_2)$ and $j(x_1) \neq 0$ and $j(x_2) \neq 0$. Finally, we let

$$f(x) = \frac{c_1(j(x) - j(x_2))}{(j(x_1) - j(x_2))} + \frac{c_2(j(x) - j(x_1))}{(j(x_2) - j(x_1))}.$$

Then $f \in \mathcal{A}$ and satisfies all the requirements. $\square$

We are now ready for Stone-Weierstrass, at least in the case of real functions!

Theorem 7.194 (The Stone Weierstrass Theorem) Let K be a compact metric space. Let $\mathcal{A} \subseteq \mathcal{C}(K, \mathbb{R})$. Then $\mathcal{A}$ is dense in $\mathcal{C}(K)$ if and only if $\mathcal{A}$ separates points and vanishes at no point.

Proof. For the first direction assume that $\mathcal{A}$ is dense in $\mathcal{C}(K)$. Then there is some sequence converging to the function $f(x) = 1$ so that $\mathcal{A}$ vanishes at no point, and similarly for any point $p \in K$ there is some sequence converging to the function $g(x) = d(p, x)$ so that $\mathcal{A}$ must separate points.

Now for the more difficult direction. Let $f \in \mathcal{C}(K)$, we must show that f is a limit point of $\mathcal{A}$, or find some sequence in $\mathcal{A}$ that converges to f. Let $\epsilon > 0$ and let $x \in K$ we some arbitrary point. For every other $y \in K$ by Lemma 7.XX above we can find some function f_y so that $f_y(x) = f(x)$ and $f_y(y) = f(y)$. Since these f_y are continuous we can choose a neighborhood $N(y)$ so that for all $z \in N(y)$ we have $f_y(z) > f(z) - \epsilon$. Since the neighborhoods $N(y)$ are an open cover of K by compactness we can choose a finite subcover $N(y_1), ..., N(y_n)$. Now let $g_x = \max(f_{y_1}, ..., f_{y_n})$. Then there is a sequence of functions $g_{n,x}$ in $\mathcal{A}$ that converge uniformly to g_x by Lemma 7.XX and for all $y \in K$ we have that $g_x(y) \geq f(y) - \epsilon$. We can find such a g_x for each $x \in K$. Now however, each g_x is continuous and $g_x(x) = f(x)$ so that there is some $N(x)$ satisfying for all $y \in N(x)$ we have $g_x(y) \leq f(x) + \epsilon$. The collection of $N(x)$ form an open cover so by compactness we have a finite subcover $N(x_1), ..., N(x_n)$. Then again finally setting $h = \min(g_{x_1}, ..., g_{x_n})$ we see that for all $x \in K$ we have $f(x) - \epsilon \leq h(x) \leq f(x) + \epsilon$, and there is a sequence $h_n \in \mathcal{A}$ that converges uniformly to h. Since ϵ was arbitrary we see we can find a sequence that arbitrary approaches f so that f and so we see that $\mathcal{A}$ is dense in $\mathcal{C}(K)$! □

Finally, to end this section we extend the above result to $\mathcal{C}(K, \mathbb{C})$.

Theorem 7.195 (Complex Stone-Weierstrass Theorem) Let K be a compact metric space. Let $\mathcal{A} \subseteq \mathcal{C}(K)$. Then $\mathcal{A}$ is dense in $\mathcal{C}(K)$ if $\mathcal{A}$ separates points, vanishes at no point, and if $f \in \mathcal{A}$ then $\overline{f} \in \mathcal{A}$.

Proof. For any $f \in \mathcal{C}(K)$ we can write $f = u + iv$ where u, v take only real values. We need to show we can approximate these functions. The problem we don't know what purely real valued functions we have available to us! Let $\mathcal{B}$ be the subset of $\mathcal{A}$ of functions that only take real values. For $g \in \mathcal{A}$ we also have $\overline{g} \in \mathcal{A}$ so that $\frac{g+\overline{g}}{2} = \Re g \in \mathcal{A}$ and similarly $\Im g \in \mathcal{A}$. Since both $\Re g$ and $\Im g$ take only real values we see $\mathcal{B}$ is nonempty. Given a point $p \in K$ there is some function $h \in \mathcal{A}$ where $h(p) \neq 0$, so that at least one of $\Re g$ or $\Im g$ is nonzero and in $\mathcal{B}$ so that $\mathcal{B}$ vanishes at no point. Similarly, we show that $\mathcal{B}$ separates points. Now we show that $\mathcal{B}$ is an algebra of functions. If $f, g \in \mathcal{B}$, then $f + g$ and fg are in $\mathcal{A}$ and take only real values so that they are in $\mathcal{B}$, and same holds for real scalars. Therefore, we see that $\mathcal{B}$ is an algebra of real functions that separates points and vanishes at no points. Therefore, there are sequences $u_n \to_\infty u$ and $v_n \to_\infty v \in \mathcal{B}$ that converge to u and v respectively, so that the sequence $(u + iv)_n \in \mathcal{A}$ converges to f uniformly. $\square$

Chapter 8

Analytic Functions

We now turn our attention to a very important class of functions that naturally arise after studying sequences and series of functions. Namely, the Analytic Functions! These are a subclass of functions that have incredibly nice properties making them very nice functions to use. Along with this many of these functions should be very familiar to the reader. We start with a discussion of Power Series.

8.1 Power Series

Definition 8.1. We define a **Power Series** as a function of the form

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

For a complex power series, we allow complex a_n and z, where as for real power series we require these all be real numbers.

One may ask when a Power Series converges and when it diverges. The next theorem gives an incredibly nice result.

Theorem 8.197 (The Cauchy-Hadamard Theorem) Let $\sum a_n(z-z_0)^n$ be a power series, and let $R = \frac{1}{\limsup \sqrt[n]{|a_n|}}$. Note in the case $R = \frac{1}{0}$ set $R = \infty$ and if $R = \frac{1}{\infty}$ set $R = 0$. The value R is called the **Radius of Convergence** for the Power Series. Then we have if $0 < R < \infty$ that $\sum a_n(z-z_0)^n$ absolutely converges for all z with $|z - z_0| < R$, diverges for all z where $|z - z_0| > R$, and if $|z - z_0| = R$ then either can happen. In the case $R = 0$ the series converges only at z_0, and in the case $R = \infty$ the series converges on all of $\mathbb{R}$ or $\mathbb{C}$.

Proof. This is essentially an application of the Root Test! Assume $0 < R < \infty$, the special cases use the exact same argument. Then if $z \in B_R(z_0)$ that
$\limsup_{n\to\infty} \sqrt[n]{|a_n(z-z_0)^n|} = \limsup_{n\to\infty} \sqrt[n]{|a_n|} \cdot \sqrt[n]{|z-z_0|^n} < \frac{1}{R}R = 1$. By the Root Test the series converges.
Now suppose that $|z - z_0| > R$. Then $\limsup_{n\to\infty} \sqrt[n]{|a_n| \cdot |z-z_0|^n} > \frac{1}{R}R = 1$. But this means that $\lim_{n\to\infty} a_n(z-z_0)^n \neq 0$ so that the series must diverge. □

The Radius of Convergence is an incredibly useful tool. It guarantees that convergence happens in one specific region and that outside this region the series will never converge again giving us a nice set to work on! The next question one might have is whether or not this convergence is point wise or uniform? The next theorem answers this.

Theorem 8.3. Consider the power series $\sum a_n(z-z_0)^n$ with radius of convergence $0 < R < \infty$. Then for all $0 < \alpha < R$ we have that $\sum a_n(z-z_0)^n$ converges uniformly on the set $\overline{B_\alpha(z_0)}$.

Proof. We have that $|\sum a_n(z-z_0)^n| \leq \sum |a_n(z-z_0)^n| \leq \sum |a_n|\alpha$. Since $\sum |a_n|\alpha$ converges absolutely, then by the M-Test $\sum a_n(z-z_0)^n$ converges uniformly. □

Corollary 8.3.1. Power Series are Continuous in this region as well. This follows as a power series is the uniform limit of continuous functions.

The next theorem will expand on this Corollary Significantly.

Theorem 8.4. Suppose we have a power series $f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$, with Radius of Convergence R and $0 < \alpha < R$. Then on $\overline{B_\alpha(z_0)}$ we also have that f is differentiable with $f'(z) = \sum_{n=1}^{\infty} na_n(z - z_0)^{n-1}$.

Proof. We know the series converges uniformly on this region.

The function $g(z) = \sum_{n=1}^{\infty} na_n(z - z_0)^{n-1}$ is also a power series and in fact also converges uniformly on this region as $\limsup \sqrt[n]{|na_n|} = \limsup \sqrt[n]{|a_n|}\sqrt[n]{n} = \limsup \sqrt[n]{|a_n|}$. If f and g are real power series then the fact that $g(x) = f'(x)$ follows from Theorem 7.XX. The case where these are complex series must be handled differently. Assume f, g are complex power series and let z be an arbitrary point in $\overline{B_\alpha(z_0)} \subseteq \mathbb{C}$. To save space $f_n(z) = \sum_{k=0}^{n} a_k(z - z_0)^k$ and $Rf_n = \sum_{k=n}^{\infty} a_k(z - z_0)^k$. It's easy to see that each f_n is differentiable with $f_n'(z) = \sum_{k=1}^{n} ka_k(z - z_0)^{k-1}$ (this hints at the approach!). Then we have

$$\left|\frac{f(z+h) - f(z)}{h} - g(z)\right| \leq$$

$$\left|\frac{f_N(z+h) - f_N(z)}{h} - f_N'(z)\right| + |f_N'(z) - g(z)| + \left|\frac{Rf_N(z+h) - Rf_N(z)}{h}\right|.$$

Let the terms in the last inequality be (1), (2), and (3) respectively. The reader should verify the equation $\frac{a^n - b^n}{a-b} = \sum_{k=1}^{n-1} a^k b^{n-k-1}$ holds for $n \geq 2$ (hint use induction). Then we have that we can rewrite (3) as

$$\left|\sum_{n=N}^{\infty} a_n \sum_{k=1}^{N-1} (z+h)^k z^{N-k-1}\right| \leq \sum_{n=N}^{\infty} n|a_n|\alpha^{k-1}.$$

We know the above series converges and so choosing N large enough we can make (3) as small as desired. Similarly, by comparing terms we see that f_n' is the partial sums of g_n so that if we choose N large enough we can also make (2) small. Finally, from definition if we choose h close enough to z we can make (1) small.Therefore we conclude that f is complex differentiable or Holomorphic and $f'(z) = g(z)$. $\square$

Corollary 8.4.1. If f is a power series, then by the above so is the function f', so that f' is differentiable, and so on. This shows that if f is a power series it is infinitely differentiable or $f \in \mathcal{C}^{\infty}$.

8.2 Taylor Series

We now turn to the main topic of this chapter, Analytic Functions. These are very closely related to Power Series as the following definition will show.

Definition 8.5. Let $S \subseteq \mathbb{R}$ be open and let $f : S \to \mathbb{R}$. We say that f is **Real Analytic** if for all $x_0 \in S$ there is some radius r so that on $B_r(x_0)$ $f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n$, or in other words if f is locally representable as a power series at every point of the domain. Similarly, if O is an open subset of $\mathbb{C}$ we say that a function $g : O \to \mathbb{C}$ is **Complex Analytic** if for all $z_0 \in 0$ on some $B_p(z_0)$ we have $f(z) = \sum_{n=0}^{\infty} c_n(z - z_0)^n$. If a function is real or complex analytic we will simply call it **Analytic**. We will also sometimes use the notation that if f is analytic then $f \in \mathcal{C}^{\omega}(D)$, where D is the domain of the function.

By our theorems on Power Series we immediately have that both real and complex analytic functions are infinitely differentiable on their domains. The next important theorem gives a formula for calculating the coefficients of the local power series of an analytic function.

Theorem 8.201 (Taylor Series) Suppose that f is analytic (real or complex) and let z_0 be a point of the domain. Then for the power series expanded at z_0, $f(z) = \sum a_n(z - z_0)^n$ we have that

$$a_n = \frac{f^{(n)}(z_0)}{n!},$$

where $f^{(n)}$ is the n-th derivative of f. When expressed in this form the series above is often called the **Taylor Series** of f about z_0.

Proof. This essentially follows from taking successive derivatives of the power series. For

$$f^{(n)}(z) = \sum_{k=n}^{\infty} k(k-1)(k-2)...(k-n+1)a_k(z-z_0)^{k-n},$$

so that

$$f^{(n)}(z_0) = a_n n(n-1)(n-2)...(n-n+1) = a_n n!,$$

which completes the proof. $\square$

The converse of the above theorem holds too, that is that if f can be represented as $\sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z-z_0)^n$ at every point of the domain, then f is analytic! This gives the primary tool in determining if a function is anaylytic! For to show this we must show that its Taylor Series converges to the function! In general this is quite difficult, it is in general quite tough to decide whether a real function is Analytic or not. For complex functions in the section on Complex Analysis we will see that there is an incredibly useful tool for determining this, but for real functions we are not so lucky. The next familiar theorem gives a tool for showing a Taylor Series converges, though it may not converge to the original function.

Theorem 8.202 (Taylor's Theorem) Suppose $f \in \mathcal{C}^{\infty}([a,b])$ and consider the function $f_n(x) = \sum_{k=0}^{n} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$, where $x_0 \in (a,b)$. Then for all $x \in (a,b)$ where $x \neq x_0$ there exists some point p so that $|p - x_0| < |x - x_0|$ and

$$f(x) = f_n(x) + \frac{f^{(n+1)}(p)}{(n+1)!}(x-x_0)^{n+1},$$

in particular if M is an upper bound for $f^{(n+1)}$ on (a,b), then

$$|f(x) - f_n(x)| \leq \frac{M}{(n+1)!}(b-a)^{n+1}.$$

Proof. Without loss of generality assume $x > x_0$ and consider the function

$R_n = f - f_n - M(x-x_0)^{n+1}$, where M is chosen so that $R_n(x) = 0$. We see for all $0 \le k \le n$ that $R_n^{(k)}$ exists and further that $R_n^{(k)}(x_0) = 0$. Then by Rolle's Theorem is some point $c_1 \in (x_0, x)$ so that $R_n'(c) = 0$. Another application of Rolle's Theorem gives another point $c_2 \in (x_0, c_1)$ so that $R_n''(c_2) = 0$. Continuing this along we get points $x_0 < c_n < c_{n-1} < ... < c_1$ so that we have $R_n^{(n)}(c_n) = 0$. But now by a final application of Rolle's Theorem we have that there is some point p so that $R_n^{(n+1)}(p) = 0$, but $R_n^{(n+1)} = f^{(n+1)}(p) - (n+1)!M$, so that $M = \frac{f^{(n+1)}(p)}{(n+1)!}$ completing the proof. $\square$

The bound the theorem gives is often called the **Lagrange Error Bound**, and can be used to show that a Taylor Series converges to the original function on some suitable neighborhood. For if M is small, $b-a$ is small, then as $n \to \infty$ we will see that the difference between f and f_n is bounded by something that approaches zero showing uniform convergence on this range.

8.3 Properties of Analytic Functions

We now want to discuss some of the properties analytic functions satisfy! The next theorem is one of the most important results concerning Analytic Functions as it gives conditions on their behavior.

Theorem 8.203 (The Identity Theorem) Suppose f is an analytic function on some open connected set O. Then $L = \{z \in O | f(z) = 0\}$ has a limit point in O if and only if $f(z) = 0$ for all $z \in O$. This applies whether $O \subseteq \mathbb{R}$ or $\mathbb{C}$.

Proof. The case where $f(z) = 0$ for all $z \in O$ implies that L has a limit point in O immediately, it is the other direction which is quite interesting.
Suppose L has some limit point in O. We would like to show that the set of limit points of L, L', is both open and closed, and since the only open and closed subset of a connected set is the empty set and itself this will complete the proof since L' is nonempty. If p is a limit

point of L', then we can easily show that p is a limit point of L so that $p \in L'$ and so L' is closed. So now we just need to show that L' is open.

Let p be in L', we must find an open ball containing p in L'.

Let $\sum a_n(z-p)^n$ be the power series of f about p with radius of convergence $r > 0$. If $a_n = 0$ for all n then clearly on $B_r(p) \subseteq L'$, as $f(z) = 0$ on all of $B_r(p)$. Suppose this is not the case and let k be the smallest number so that $a_k \neq 0$. Then we can rewrite f as $f(z) = (z-p)^k \cdot g(z)$ where $g(z) = \sum_{n=0}^{\infty} a_{n+k}(z-p)^n$. We have that $a_k \neq 0$ so that $g(p) = a_k \neq 0$. Since g is continuous at p there is some disc $B_\rho(p)$ so that on this disc $g(z) \neq 0$. It follows then that on $B_\rho(p) - \{p\}$ we have that $f(z) \neq 0$ so that p is not a limit point of L, a contradiction. Therefore, we see that $B_r(p) \subseteq L'$ completing the proof. □

Corollary 8.8.1. If two functions f, g are analytic on some open connected set O and the set of points where $f = g$ has a limit point in O, then $f = g$ everywhere on O. This follows by applying The Identity Theorem to the analytic function $h = f - g$.

The Identity Theorem leads to many useful consequences, one important one is the idea of Analytic Continuation.

Definition 8.9. Suppose $V \subseteq \mathbb{C}$ is some open set and f is analytic on V. If $V \subseteq U$ and there is an analytic function g on U so that for all $z \in V$ we have $g(z) = f(z)$, then g is called the **Analytic Continuation** of f on U. Such a continuation is unique by the Corollary to The Identity Theorem. Analytic continuation is a very important topic and the next theorem will touch on this a bit. Note we may also require that V be only open in $\mathbb{R}$ if so desired, just so long as V has a limit point in V we can guarantee that any Analytic Continuation is unique!

Theorem 8.10. Let $V \subseteq R$ be a nonempty connected open set and let $O \subseteq \mathbb{C}$ be a nonempty connected set containing V. Then a function $f : V \to \mathbb{R}$ is real analytic if and only if it has an unique complex analytic continuation $g : O \to \mathbb{C}$ in the sense that $g(x) = f(x)$ for all $x \in V$.

Proof. Assume f is real analytic. Then for each $x \in V$ there exists a set $B_r(x) \subseteq \mathbb{R}$ so that f is locally representable as a power series in this region, expanded about x. Now consider the same collection of all $B_r(x)$ but this time taken as subsets of $\mathbb{C}$. Then the union of this region $O = \bigcup B_r(x)$ is an open connected subset of $\mathbb{C}$ and for any point $z \in O$, we have $z \in B_r(x_0)$ for some x_0 so that there is a power series expansion for $f(x) = \sum a_n(x - x_0)^n$. Take in this region $g(z) = \sum a_n(z - x_0)^n$. Then g is complex analytic and uniqueness follows from The Identity Theorem as the set $f(z) = g(z)$ holds on all of z.
On the other hand if g is an analytic continuation on O that agrees with f on V, then since g is locally representable as a power series so is f completing the proof. □

We would now like to discuss which functions are analytic. We know that to be analytic, a function must be smooth: ie $\mathcal{C}^\omega(V) \subseteq \mathcal{C}^\infty(V)$ for any open set V. The next theorem tells us how we can combine analytic functions to get new ones.

Theorem 8.11. If f, g are analytic on some open set V, then $f + g$, $f \cdot g$, and $f \circ g$ are all analytic on V. If $f(z) \neq 0$ for all $z \in V$, then $\frac{1}{f}$ is also analytic on V. If $f'(z) \neq 0$ for all $z \in V$ and f^{-1} exists then f^{-1} is analytic on V.

Proof. The case of $f + g$ is trivially easy. The case of $f \cdot g$ holds by taking the Cauchy Product of power series, which are absolutely convergent inside their radius of convergence. The last three cases are delayed until the section on Complex Analysis. They will rely on a very important result concerning Holomorphic Functions. One might be unhappy to see these proofs delayed, but they will not be used again except for a brief mention of Elementary Functions, and these results are not used in the proof so that no circular logic is applied. □

It is obvious that polynomials are all Analytic, simply take all $a_n = 0$ past a given n! By the above theorem, this gives that rational functions are Analytic. As are any rational powers such as $\sqrt{x}$ and so on. The next section will discuss the famous Exponential and Trig functions and we will see that by construction these functions are also Analytic!

At this point one might be asking if every smooth function is analytic. For every smooth function, we can find its Taylor Series and so does the Taylor Series always converge? Another way to reword this question is to ask if $\mathcal{C}^\omega(D) = \mathcal{C}^\infty(D)$. We've already shown that $\mathcal{C}^\omega(D) \subseteq \mathcal{C}^\infty(D)$, so we would just have to show the other direction holds. It turns out that this is false! There are smooth functions that are not analytic! The next example will show this.

Example 8.12. Assuming some knowledge of the exponential function, consider the function

$$f(x) = \begin{cases} e^{\frac{-1}{x^2}} & x > 0 \\ \\ 0 & x \leq 0 \end{cases},$$

then the reader can verify that f is smooth on $\mathbb{R}$. However, f is NOT analytic. As $f(x) = 0$ for all $x \leq 0$ so that by the Identity Theorem if f were analytic it would need to equal 0 on all of $\mathbb{R}$, which it clearly does not.

8.4 Exponential and Trigonometric Functions

To end this section we will introduce a handful of very familiar and useful functions: The Exponential and Trig Functions! These functions should be incredibly familiar to any reader, but likely not at this level of rigor. So we will formally introduce them and prove all of their familiar properties to close and mention the class of Elementary Functions to end this chapter.

Definition 8.13. We define the **Exponential Function** as

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

The above series converges absolutely on all of $\mathbb{C}$ as an easy application of the Ratio Test. Since $\exp(z)$ is given by a Power Series it is automatically Analytic and therefore differentiable with derivative

$$\exp'(z) = \sum_{n=1}^{\infty} \frac{nz^{n-1}}{n!} = \sum_{n=0}^{\infty} \frac{z^n}{n!} = \exp(z). \tag{8.1}$$

So that the exponential function is its own derivative! Since $\exp(z)$ converges absolutely on all of $\mathbb{C}$ we can use the Cauchy Product and the Binomial Theorem to find that

$$\exp(z) \cdot \exp(w) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k}{k!} \cdot \frac{w^{n-k}}{(n-k)!} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} = \exp(z+w), \tag{8.2}$$

proving another very important property. Our last observation is the fact that clearly if $x \in \mathbb{R}$, then $\exp(x)$ is also real. This can be seen directly from the formula.

We now use these base properties to prove some more familiar properties.

Theorem 8.14. The following are all true:

(a) For all $z \in \mathbb{C}$, we have $\exp(z) \neq 0$.

(b) $\overline{\exp(z)} = \exp(\overline{z})$.

(c) $|\exp(ix)| = 1$ for all $x \in \mathbb{R}$.

(d) If we define $e = \sum_{n=0}^{\infty} \frac{1}{n!}$, then $\exp(n) = e^n$.

Proof. We have that $\exp(0) = 1$, so that for all $z \in \mathbb{C}$ we have $\exp(z) \cdot \exp(-z) = \exp(0) = 1$ so that clearly $\exp(z) \neq 0$, proving (a). We see that $\overline{\exp(z)} = \overline{\sum_{n=0}^{\infty} \frac{z^n}{n!}} = \sum_{n=0}^{\infty} \frac{\overline{z^n}}{n!}$, as $\overline{z^n} = (\overline{z})^n$ proving (b). We have that (c) essentially follows from (b) as $|\exp(ix)| = \sqrt{\exp(ix) \cdot \overline{\exp(ix)}} = \sqrt{\exp(0)} = 1$. Finally, (d) follows from the multiplication property above. $\square$

Because of property (d) we will often instead refer to $\exp(z)$ as e^z, especially in the case of real numbers. The next theorem also deals with e^x, but is specifically properties for the real case.

Theorem 8.15. The following are all true:

(a) $\lim_{x\to\infty} e^x = \infty$ and $\lim_{x\to\infty} e^x = 0$.

(b) e^x grows faster than any polynomial in the sense that $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$, for any $n \in \mathbb{N}$.

(c) $e^x > 0$ for all $x \in \mathbb{R}$.

(d) e^x is strictly increasing on $\mathbb{R}$.

Proof. The first part of (a) follows immediately from the definition. To see the other we have that $e^{-x} = \frac{1}{e^x}$ so that the first part of (a) gives the second. For (b) this is easy as $e^x > \frac{x^{n+1}}{(n+1)!}$. To see (c) we have that for all $x \geq 0$ this follows from the definition, and Property (8.2) gives for all $x < 0$. Finally, for (d) this again follows straight from the definition, and we also get from this that e^x is bijective when the codomain is $(0, \infty)$. □

Definition 8.16. Since e^x is a bijection from $\mathbb{R}$ to $(0, \infty)$ it has an inverse function here. We define the function $\log : (0, \infty) \to \mathbb{R}$ by $\log(x) = (e^x)^{-1}$, the inverse of the exponential function. By Theorem 8.xx we have that the $\log(x)$ function is Analytic, and therefore differentiable. We can calculate its derivative using the chain rule by

$$1 = (x)' = (\exp(\log(x))' = \exp(\log(x)) \cdot \log'(x) = x \log'(x) \implies \log'(x) = \frac{1}{x}. \tag{8.3}$$

Further, property 8.2 for exponentials can be reworked. For any $x, y \in \mathbb{R}$ we can rewrite these as $x = e^w$ and $y = e^v$ to get that

$$\log(xy) = \log(e^w e^v) = \log(e^{w+v}) = w + v = \log(x) + \log(y). \tag{8.4}$$

The next theorem includes several facts concerning the log function. Most proofs follow from corresponding facts about e^x and the above properties, so proofs are omitted

Theorem 8.17. The following are true:

(a) $\log(x) = \int_1^x \frac{1}{x} dx$.

(b) $\log(x)$ grows slower than any polynomail, ie $\lim_{x\to\infty} \frac{\log(x)}{x^n} = 0$ for $n \geq 1$.

(c) $\lim_{x\to 0} \log(x) = -\infty$, $\lim_{x\to\infty} \log(x) = \infty$, and log is strictly increasing on $(0, \infty)$.

Definition 8.18. In algebra/Calculus you probably worked with expressions like x^a for some $a \in \mathbb{R}$, and were allowed to rewrite this as $e^{a\log(x)}$, we will allow this by making this a definition! For any $x \in \mathbb{R}$ and for any $a \in \mathbb{R}$, we define $x^a = e^{a\log(x)}$, as long at least one of x or a is nonzero. This holds for integer powers as well and agrees with the old definition.

Theorem 8.214 (Power Rule) Let $f(x) = x^a$ for some $a \neq 1$. Then f is differentiable and $f'(x) = ax^{a-1}$.

Proof. We have $f(x) = x^a = e^{a\log(x)}$, giving differentiability at $x \neq 0$, and the chain rule gives $f'(x) = e^{a\log(x)} \cdot \frac{a}{x} = ae^{a\log(x)} \cdot e^{-\log(x)} = ax^{a-1}$. □

We now turn to the Trigonometric Functions!

Definition 8.20. We define the function $\cos(z)$ by

$$\cos(z) = \frac{\exp(ix) + \exp(-ix)}{2},$$

and the function $\sin(z)$ by

$$\sin(z) = \frac{\exp(ix) - \exp(-ix)}{2i}.$$

These functions are analytic and infact the following power series converge to the functions on all of $\mathbb{C}$

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} \qquad \sin(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}. \tag{8.5}$$

We immediately have from the definitions that

$$\exp(iz) = \cos(z) + i\sin(z). \tag{8.6}$$

It's easy to see from the Power Series that for real x both sin and cos are real, so that for real x we have $\cos(x) = \Re(\exp(ix))$ and $\sin(x) = \Im(\exp(ix))$ giving that for all real x we have

$$\cos^2(x) + \sin^2(x) = 1. \tag{8.7}$$

We also have since $\cos'(z) + i\sin'(z) = (\exp(iz))' = i\exp(ix) = -\sin(z) + i\cos(z)$, we get that

$$\sin'(z) = \cos(z) \qquad \cos'(z) = -\sin(z). \tag{8.8}$$

The next theorem captures a few more of the important properties of the sin and cos functions.

Theorem 8.21. The following are true:

(a) There exists a real number π so that $\cos(\frac{\pi}{2}) = 0$ and for all $x < \frac{\pi}{2}$, $\cos(x) > 0$.

(b) $\exp(z)$ is periodic with period $2\pi i$.

(c) sin and cos are periodic with period 2π.

(d) For all $z \in \mathbb{C}$, $z \neq 0$, there exists a number $r \in (0, \infty)$ and a number $\theta \in [0, 2\pi)$ so that $z = r\exp(i\theta)$.

Proof. For (a) we assume that $\cos(x) \neq 0$ for all $x \in \mathbb{R}$. Then we know $\cos(0) = 1$, so that by continuity for all $x \in \mathbb{R}$, $\cos(x) > 0$. But then $\sin(x)$ is strictly increasing, so that $x < y$ implies $\sin(x) < \sin(y)$ giving

$$\sin(x)(y - x) \leq \int_x^y \sin(t)\,dt = \cos(x) - \cos(y) \leq 2.$$

But choosing $x > 0$ and y large enough makes the above false, so that there is some real number t where $\cos(t) = 0$. The set $[0, t] \cap \cos^{-1}(\{0\})$ is compact so that a minimum exists. Simply define π to be twice this minimum and (a) follows!

We then immediately have from 8.x that $\sin(\frac{\pi}{2}) = 1$, giving $\exp(\frac{\pi i}{2}) = i$, so that $\exp(\pi i) = \exp(\frac{\pi i}{2})^2 = -1$ or $\exp(2\pi i) = 1$. Then (b) follows as $\exp(z+2\pi i) = \exp(z)\exp(2\pi i) = \exp(z)$.

We have that (c) follows almost immediately from (b).

Finally, to prove (d) suppose $z \neq 0$. Then take $r = |z|$ and let $x = \Re\left(\frac{z}{|z|}\right)$ and $y = \Im\left(\frac{z}{|z|}\right)$. If $x \geq 0$, then $x \in [0, 1]$ so that for some $\theta \in [0, \frac{\pi}{2}]$ we have $\cos(\theta) = x$, and it follows that $\sin(\theta) = y$, so that $\exp(i\theta) = \frac{z}{|z|}$. The rest of the cases are not much more complicated. For $x < 0$ apply previous to $\frac{-iz}{|z|}$ and for $y < 0$ apply previous to $\frac{-z}{|z|}$. $\square$

The remaining properties of the trig functions can be developed in similar manner, but we leave that to the reader. To end this discussion we briefly mention the Hyperbolic functions and the Elementary functions.

Definition 8.22. We define

$$\cosh(z) = \frac{e^z + e^{-z}}{2} \qquad \text{and} \qquad \sinh(z) = \frac{e^z - e^{-z}}{2}.$$

The reader can verify that $\sinh'(z) = \cosh'(z)$ and $\cosh'(z) = \sinh(z)$. They have more uses, but are not as important as the trig functions.

Definition 8.23. We define the collection of **Elementary Functions** as all functions of the form of finite sums, products, and compositions of: Polynomial, Rational, Exponential, Trig, and Hyperbolic functions as well as their inverses when applicable. We see that Elementary Functions are all Analytic due to Theorem 8.XX. The Elementary Functions are incredibly nice and special as they are the functions most commonly used functions to model the real world. In fact, at one point in time Elementary Functions were the only functions mathematicians considered! We do not talk further on the topic, but feel it is nice to introduce this familiar class of functions.

Chapter 9

Fourier Series

In the last Chapter we discussed the very important class of Analytic Functions. We can represent these functions quite nicely with power series, using The Taylor Formula giving a great tool to approximate Analytic functions. Unfortunately, Analytic Functions are also quite rare, they are more special than most functions we work with in Pure Math. For example, smooth functions are already quite rare requiring derivatives of all orders, and Analytic functions are an even smaller subclass of this group! Fourier Series give a different tool to approximate functions, that have a larger scope. We will see that the conditions for Fourier Series to converge, can be much more relaxed. However, we will also see that the way in which we measure convergence plays a large role as well. We start by examining abstract Hilbert Spaces, which were mentioned previously in chapter 2. We will build up some abstract theory, and then employ this theory through Fourier Series.

9.1 Inner Product Spaces

We briefly recall some facts from Chapter 2.

Definition 9.1. Give a vector space V over some field $\mathbb{F}$ (either $\mathbb{R}$ or $\mathbb{C}$ in our case), then an **Inner Product** is a binary operation $< .,. >: V^2 \to \mathbb{F}$, satisfying

1. Conjugate Symmetry: $\langle x, y\rangle = \overline{\langle y, x\rangle}$.
2. Linearity: $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$.
3. Positive Definite: $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

The pair $< V, \langle .,. \rangle >$ is called an **Inner Product Space**, and gives rise to a norm on V by $||x|| = \sqrt{\langle x, x\rangle}$. If the normed space $< V, || \cdot || >$ is a Banach Space, then $< V, \langle .,. \rangle >$ is called a **Hilbert Space**.

In the case that $\mathbb{F} = \mathbb{R}$, then clearly (1) gives $\langle x, y\rangle = \langle y, x\rangle$.

The reason Inner Product Spaces are so important is the fact that the Inner Product can measure "angles" or more generally orthogonality.

Definition 9.2. Let $x, y \in V$ where V is an Inner Product Space. Then we say that x is **Orthogonal** to y and denote this by $x \perp y$ if $\langle x, y\rangle = 0$. Given a collection of vectors $O = \{v_i\}$ we say that O is an **Orthogonal Set** if every pair of vectors in O are orthogonal. If an Orthogonal Set O also satisfies that for any $x \in O$ we have $||x|| = 1$, then O is also an **Orthonormal Set**.

We now recall a few definitions from Linear Algebra before continuing with the new material.

Definition 9.3. Recall as collection of vectors $\{v_i\} \subseteq V$ is called **Linearly Independent** if the equation $v_1x_1 + ...v_nx_n = 0$ has only the trivial solution $x_1 = ... = x_n = 0$, and the **Span** of the collection $\{v_i\}$ is the set of all $\{v_1x_1 + ...v_nx_n | x_1, ..., x_n \in \mathbb{F}\}$. If a collection of vectors $\mathcal{B} = \{b_1, ..., b\}$ is Linearly Independent and Spans V, then it is called a **Basis** for V. It is easy to see that any Orthogonal Set is Linearly Independent.

We proved in Chapter 2 that in an Inner Product space we have the Cauchy-Schwarz inequality

$$|\langle x, y\rangle| < ||x|| \, ||y||,$$

the parallelogram law

$$||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2,$$

and the Polarization Identity

$$\Re\langle x, y\rangle = \frac{1}{2}\left(||x+y||^2 - ||x||^2 - ||y||^2\right).$$

We now want to prove a few more Properties that follow from the properties of the Inner Product, we will prove a more general Pythagorean Theorem and a weak version of Parseval's Theorem.

Theorem 9.222 (The Pythagorean Theorem) If x and y are Orthogonal, then

$$||x||^2 + ||y||^2 = ||x+y||^2.$$

Proof. This follows as $||x+y||^2 = \langle x+y, x+y\rangle = \langle x, x+y\rangle + \langle y, x+y\rangle = \overline{\langle x+y, x\rangle} + \overline{\langle x+y, y\rangle} = \overline{\langle x, x\rangle} + 2\overline{\langle x, y\rangle} + \overline{\langle y, y\rangle} = \langle x, x\rangle + \langle y, y\rangle = ||x||^2 + ||y||^2$, we have $\langle x, y\rangle = 0$ by Orthogonality. □

Theorem 9.223 (Finite Parseval's Identity) If $\{x_1, ..., x_n\}$ is an orthogonal set, then

$$\left|\left|\sum_{i=1}^{n} x_i\right|\right| = \sum_{i=1}^{n} ||x_i||.$$

This can be seen by Inducting on the Pythagorean Theorem.

9.2 Hilbert Spaces

We have shown a few nice facts concerning Orthogonality using the Inner Product, but to go farther we will need the completeness offered by Hilbert Spaces.

Definition 9.6. Let H be a Hilbert Space. If $O = \{e_a | a \in A\}$ is an infinite Orthonormal Set, we will sometimes refer to the collection O as an **Orthonormal System** to indicate that this is infinite. For now we will assume that O is countable to simplify what it would mean to take the sum of an uncountable collection, spolier for next chapter this topic is quite uninteresting! Since we now have an infinite collection of vectors, we will call the **Span of The Orthonormal System** as the set of all $x \in H$ such that

$$x = \sum_{n=1}^{\infty} x_n e_n,$$

where each $x_n \in \mathbb{F}$. If the Span of an Orthonormal System is all of H we will say that the collection O is an **Orthonormal Basis**, or that the Orthonormal System is **Complete**. Given an Orthonormal System $O = \{e_1, e_2, ...\}$ and a vector $x \in H$ we define the n-th Fourier Coefficient of x (w.r.t. O) as $c_n = \langle x, e_n \rangle$. This can be thought of as projection of the vector x on to the vector e_n. If we think of the full projection of a vector x onto the entire Orthonormal Subspace, we get the **Generalized Fourier Series** of x (w.r.t. O) given by

$$\sum_{n=1}^{\infty} c_n e_n.$$

Fourier Series and projection are very useful as the next theorem will show.

Theorem 9.225 Let H be a Hilbert Space, let $O_n = \{e_1, e_2, ..., e_n\}$ be a finite subset of an orthonormal system, and let $x \in H$ be some vector. The span of O_n is some subspace of H we will call H_n. If we define the **Projection** of x onto the subspace H_n as $s_n = \sum_{i=1}^{n} c_i e_i$, where c_n is the n-th Fourier Coefficient of x, then s_n is the best approximation for x in H_O, in the sense that if $d_i \in \mathbb{F}$ is some other finite list of scalars then

$$||x - s_n|| = \left|\left| x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i \right|\right| \leq \left|\left| x - \sum_{i=1}^{n} d_i e_i \right|\right|.$$

This projection x' is often also called the n-th Partial Fourier Sum of x.

Proof. Let d_i be as above. We have that

$$\left\langle \sum d_i e_i, x \right\rangle = \langle d_1 e_1, x\rangle + ... + \langle d_n e_n, x\rangle$$

$$= d_1\langle e_1, x\rangle + ... + d_n\langle e_n, x\rangle = d_1\overline{c_1} + ... + d_n\overline{c_n} = \sum d_i\overline{c_i}.$$

Further we have from the Finite Parseval's Identity that $||\sum d_i e_i||^2 = \sum |d_i|^2$ and likewise $||\sum c_i e_i||^2 = \sum |c_i|^2$. Finally, we can also compute that $\langle \sum d_i e_i, \sum c_i e_i\rangle = \sum d_i\overline{c_i}$, showing that $\langle \sum d_i e_i, \sum c_i e_i\rangle = \langle \sum d_i e_i, f\rangle$.

Then altogether we have using the Polarization Identity.

$$\left|\left|x - \sum d_i e_i\right|\right|^2 = ||x||^2 + \left|\left|\sum d_i e_i\right|\right|^2 - 2\Re\left\langle x, \sum d_i e_i\right\rangle$$

$$= ||x||^2 + \left|\left|\sum d_i e_i\right|\right|^2 - 2\Re\left\langle \sum d_i e_i, \sum c_i e_i\right\rangle + \left|\left|\sum c_i e_i\right|\right|^2 - \left|\left|\sum c_i e_i\right|\right|^2$$
$$= ||x||^2 + \left|\left|\sum (d_i - c_i)e_i\right|\right|^2 - \left|\left|\sum c_i e_i\right|\right|^2.$$

Since all of the above are positive terms, and only the middle depends on d_i, it is clear the above expression is minimized when $d_i = c_i$. □

Corollary 9.7.1. As a nice corollary, we see the computation used above generates the following formula

$$\left|\left|x - \sum_{i=1}^n c_i e_i\right|\right|^2 = ||x||^2 + \left|\left|\sum_{i=1}^n c_i e_i\right|\right|^2 = ||x||^2 + \sum_{i=1}^n |c_i|^2.$$

We still have yet to use the completion offered by the Hilbert Space, but the next theorem will!

Theorem 9.226 (Bessel's Inequality) For a Hilbert Space H and a vector x if $\{e_1, e_2, ...\}$ is an orthonormal system and c_i are the Fourier Coefficients for x, then

$$\sum_{n=1}^{\infty} |c_n|^2 \leq ||x||^2.$$

In particular, $\sum_{n=1}^{\infty} |c_n|^2 < \infty$. Further since $||\sum^n c_i e_i||^2 = \sum^n |c_i|^2$ for all n, it follows by completeness the former series also converges and so there is some $x' \in H$ such that $x' = \sum^{\infty} c_n e_n$, though it may be the case that $x' \neq x$ if the orthonormal system is not a basis!

Proof. By the last corollary we have that for any $n \in \mathbb{N}$ that $\sum_{i=1}^{n} |c_i|^2 \leq ||x||^2$. Since this is independent of our choice of n it follows that the general case follows or
$\sum_{n=1}^{\infty} |c_n|^2 \leq ||x||^2$. □

Theorem 9.227 (The Riemann-Lebesgue Lemma) In the above situation where c_n are the Fourier Coefficients for x, then $c_n \to 0$ as $n \to \infty$. This is an obvious corollary of Bessel's Inequality.

While the x' above need not equal x, it is the case that it must equal x if the orthonormal system is complete, this is quite easy to see but we state it in the form of the following theorem.

Theorem 9.228 (Parseval's Identity) If H is a Hilbert Space, $x \in H$, and $\{e_1, e_2, ...\}$ is orthonormal basis, then The Generalized Fourier Series of x does converge to f and Bessel's Inequality becomes an equality, ie

$$\sum_{n=1}^{\infty} |c_n|^2 = ||x||^2.$$

9.3 The Trigonometric System

We now want to put all of this abstract theory to work! While it may not be immediately obvious, you can think of functions as vectors, uncountably long vectors but vectors nonetheless! And now we should decide how to turn the space of functions into an Inner Product Space! For the case of $\mathbb{R}^n$ our inner product was the dot product where $(x_1, ..., x_n) \cdot (y_1, ..., y_n) = x_1y_1 + ...x_ny_n$. Fortunately, we have a notion very similar to a sum for functions: The Integral. This is how we will give an inner product to functions! One last comment before getting to work, we will obviouly need to restrict ourselves to integrable functions to make this work.

Definition 9.11. Let $f, g \in \mathcal{R}$ on $[a, b]$. Then we can turn $\mathcal{R}$ on $[a, b]$ into an Inner Product Space with the Inner Product given by

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,dx.$$

This integral exists by the fact f, g are integrable and conjugation is continuous and one can verify that the Inner Product properties are satisfied. This also gives us a Norm we will call the L^2 Norm denoted

$$||f||_2 = \sqrt{\int_a^b |f|^2\,dx}.$$

At this point the reader might hope that the space $\mathcal{R}$ with the above inner product is a Hilbert Space, which would immediately give us access to all the tools we just proved. Unfortunately, $(\mathcal{R}, ||.||_2)$ is NOT complete! This is one of the major downsides of the Riemann Integral, but for now we must adapt to this situation and proceed the best we can. In the next couple chapters we will develop the Lebesgue Integral, a more powerful and general method of integration, and it turns out that the space of square Lebesgue Integrable functions will be complete!

With that out of the way we can apply some of the results proved in the previous sections, and will largely use these results for motivation of what follows. We may start to wonder if we have access to an Orthonormal System on the space $\mathcal{R}$, and it turns out the the Trigonometric Functions provide exactly this.

Theorem 9.12. The collection of functions $\{1, \cos(x), \sin(x), \cos(2x), \sin(2x), ...\}$ form an Orthogonal System on $[0, 2\pi]$, and when modified to be $\{1, \frac{\cos(x)}{\sqrt{\pi}}, \frac{\sin(x)}{\sqrt{\pi}}, ...\}$ they form an Orthonormal System! Similarly the collection of functions

$$\left\{\frac{e^{inx}}{\sqrt{2\pi}} | n \in \mathbb{Z}\right\}$$

also forms an Orthonormal System. We see that all of the above functions are integrable, and further have period 2π, so for that reason we will usually restrict ourselves to consider the interval $[0, 2\pi]$ rather than the more general $[a, b]$.

Since the above give orthonormal systems we can construct Fourier Series (this is originally how Fourier Series arose) using these functions. That is given a function f if we set $a_n = \langle f, \frac{\cos(nx)}{\sqrt{\pi}}\rangle = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x)\cos(nx)\, dx$, and similarly for $b_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x)\cos(nx)\, dx$ and $c_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)e^{-inx}\, dx$, then we can represent it as a Fourier Series of these functions! We write using the above coefficients for sin and cos

$$f \sim \sum_{n=0}^{\infty} a_n \frac{\cos(nx)}{\sqrt{\pi}} + b_n \frac{\sin(nx)}{\sqrt{\pi}}.$$

For the exponentials it is a little different we write $f \sim \sum_{-\infty}^{\infty} c_n \frac{e^{inx}}{\sqrt{2\pi}}$, as all integers are used. It is quite annoying to remember to manage all the roots of π, so to make matters easier one may replace c_n as follows

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

giving the nicer expression

$$f \sim \sum_{-\infty}^{\infty} c_n e^{inx}.$$

The same of course also holds for a_n,b_n and the respresentations with $\sin' s$ and $\cos' s$.

It is easy to guess what is meant by a partial sum for the sin and cos representation, but it turns out that using the exponential is much nicer so we will use the exponential orthonormal system, and what we mean by an n-th partial sum is

$$S_n(f,x) = \sum_{-n}^{n} c_k e^{ikx}.$$

9.4 L^2 Convergence of Fourier Series

Having introduced the important trigonometric/exponential systems we now wish to discuss convergence. It turns out that the most appropriate way to discuss how Fourier Series converge is in the L^2 Norm that our integral inner product on $\mathcal{R}$ provided! It turns out that these systems are in fact complete systems, though our space is not a Hilbert Space, so that our trig/exponential Fourier Series will converge in the L^2 Norm! This is made precise in the Riesz-Fischer Theorem! First, however, we must use an application of the Stone-Weierstrass Theorem.

Theorem 9.13. Consider the set $K = [0, 2\pi]$. Then for any complex function $f \in \mathcal{C}(K)$ for all $\epsilon > 0$ there exists an exponential polynomial $P(x) = \sum_{-n}^{n} d_k e^{ikx}$ so that $|P - f|_\infty < \epsilon$ on $[0, 2\pi]$. It is worth noting that if we instead assume that f is periodic with period 2π, the above result holds on all of $\mathbb{R}$ rather than just $[0, 2\pi]$!

Proof. As mentioned above this is an application of Stone-Weierstrass. The collection of all exponential polynomials on $[0, 2\pi]$ form an algebra of functions that separates points and vanishes at no point of $[0, 2\pi]$. Further, it is self-adjoined so that by The Complex Stone Weierstrass Theorem the collection of linear combinations of the exponential above are dense in $\mathcal{C}(K)$. The result follows. □

This theorem might trick the reader into assuming that the Fourier Series of a continuous function converges uniformly. This is necessarily the case! The polynomials chosen above may not agree with the partial sums, and may behave strangely. Further, the Partial Sums of the Fourier Series were only the best approximation in the L^2 Norm sense, not in the Sup-Norm sense! But we will use this as a stepping stone to show that The Fourier Series does converge in the sense of the L^2 Norm! This result is a weak version of the Riesz-Fischer Theorem which we now prove!

Theorem 9.232 (Riesz-Fischer Theorem) Let $f \in \mathcal{R}$ be 2π periodic and let $S_n(f)$ be the n-th partial sum of f as above. Then

$$\lim_{n\to\infty} ||f - S_n(f)||_2 = \lim_{n\to\infty} \frac{1}{2\pi}\int_0^{2\pi} |f(x) - S_n(f,x)|^2 \, dx = 0.$$

In other words the Fourier Series converges in the L^2 Norm!

We need the following Lemma

Lemma 9.15. Let $f \in \mathcal{R}$ on $[a,b]$. Then for all $\epsilon > 0$ there is a function $g \in \mathcal{C}([a,b])$ so that

$$||f - g||_2 < \epsilon.$$

In other words, the continuous functions are dense in $\mathcal{R}$ with the L^2 Norm.

Proof. Let $\epsilon > 0$ and let $P = \{a = x_0, x_1, ..., x_n = b\}$ be a partition of $[a,b]$ so that $U(P,f) - L(P,f) < \frac{\epsilon}{2M}$ where M is the maximum of f on $[a,b]$. Then define $g : [a,b] \to \mathbb{R}$ by the function

$$g(x) = \frac{x_i - x}{\Delta_i} f(x_{i-1}) + \frac{t - x_{i-1}}{\Delta_i} f(x_i), \quad x \in [x_{i-1}, x_i].$$

Then g is continuous on $[a,b]$ (this is obvious for $x \in (x_{i-1}, x_i)$ and at the boundary points g will agree regardless of which way you approach) and so is $|f-g|$. Since $|f-g| > 0$ it suffices to show that $U(P, |f-g|^2) < \epsilon$. Let M_i and m_i be as normal for the function f on

$[x_{i-1}, x_i]$. Then since $m_i \leq g(x) \leq M_i$ for all $x \in [x_{i-1}, x_i]$ we find

$$U(P, |f-g|^2) \leq \sum (M_i - m_i)^2 \Delta_i \leq 2M \sum (M_i - m_i)\Delta_i < 2M \frac{\epsilon}{2M} = \epsilon,$$

so that $\int_a^b |f-g|^2 \, dx \leq U(P, |f-g|^2) < \epsilon$. $\square$

We now prove Theorem 9.XXX

Proof. We need only show that $\lim_{n\to\infty} ||f - S_n(f)||_2 = 0$. So let $\epsilon > 0$. Then by the above Lemma there is some continuous function g so that $||f-g||_2 < \epsilon$. Then by Theorem 9.XX there is some Exponential Polynomial so that for all $x \in \mathbb{R}$ we have $|g(x) - P(x)| < \epsilon$. But this implies that $\int_0^{2\pi} |g(x) - P(x)|^2 \, dx \leq 2\pi\epsilon^2$, so that $||g - P||_2 < \sqrt{2\pi}\epsilon$. The L^2 Norm Triangle Inequality gives that $||f - P||_2 < (\sqrt{2\pi} + 1)\epsilon$. Let N be the degree of P, then we have that by Theorem 9.XX that

$$||f - S_N(f)||_2 \leq ||f - P||_2 < (\sqrt{2\pi} + 1)\epsilon.$$

Which completes the proof. $\square$

The above proof essentially followed by showing that S_N is the best L^2 approximation, that exponential polynomials are dense in continuous functions, and that continuous functions are dense in integrable functions (with L^2 Norm). To end this section we prove one more nice fact called Parseval's Theorem, which is closely related with Parseval's Identity but more specifically for functions!

Theorem 9.234 (Parseval's Theorem) If $f, g \in \mathcal{R}$ and are 2π periodic, with Fourier Series $f \sim \sum_{-\infty}^{\infty} c_n e^{inx}$ and $g \sim \sum_{-\infty}^{\infty} d_n e^{inx}$, then

$$\frac{1}{2\pi} \int_0^{2\pi} f(x)\overline{g(x)} \, dx = \sum_{-\infty}^{\infty} c_n \overline{d_n}.$$

In particular $\frac{1}{2\pi}\int_0^{2\pi}|f(x)|^2\,dx = \sum_{-\infty}^{\infty}|c_n|^2$.

Essentially this is saying that $\langle f, g\rangle = \langle \sum c_n e^{inx}, \sum d_n e^{inx}\rangle$.

Proof. If $S_n(f)$ is the partial sum for f, then

$$\frac{1}{2\pi}\int_0^{2\pi} S_n(f,x)\overline{g(x)}\,dx = \sum_{-n}^{n}\frac{c_n}{2\pi}\int_0^{2\pi} e^{inx}\overline{g(x)}\,dx = \sum_{-n}^{n} c_n\overline{d_n}.$$

It follows by the Cauchy Schwarz Inequality that

$$\left|\frac{1}{2\pi}\int_0^{2\pi} f\overline{g}dx - \sum_{-n}^{n} c_n\overline{d_n}\right| = \left|\frac{1}{2\pi}\int_0^{2\pi} f\overline{g}dx - \frac{1}{2\pi}\int_0^{2\pi} S_n(f)\overline{g}dx\right| = \frac{1}{2\pi}\left|\langle f - S_n(f), g\rangle\right|$$

$$\leq \frac{1}{2\pi}||g||_2||f - S_n(f)||_2,$$

and by The Riesz-Fischer Theorem we see that $\lim_{n\to\infty}\left|\frac{1}{2\pi}\int_0^{2\pi} f\overline{g}dx - \sum_{-n}^{n} c_n\overline{d_n}\right| = 0$, completing the proof. □

To end this section we give a modified version of the Riemann-Lebesgue Lemma, which will be useful for examining other methods of convergence.

Theorem 9.235 (Riemann-Lebesgue Lemma) Suppose $f \in \mathbb{R}$ and is 2π periodic. Then

$$\lim_{n\to\infty}\frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}\,dx = \lim_{n\to\infty}\frac{1}{\pi}\int_0^{2\pi} f(x)\cos(Nx)\,dx = \lim_{n\to\infty}\frac{1}{\pi}\int_0^{2\pi} f(x)\sin(Nx)\,dx = 0.$$

Proof. The case of the exponential follows as that integral gives the Fourier Coefficients c_n and by Parseval's Theorem $c_n \to 0$. The sin and cos integrals follow from the exponential one. □

9.5 Pointwise Convergence

We've proven some nice facts about Fourier Series, and importantly, that for any integrable function the Fourier Series will converge in the L^2 Norm! A great way to end this chapter would be to prove that L^2 convergence implies pointwise, or even uniform convergence showing that the Fourier Series converges pointwise to the function f, so that we can definitively say that $f(x) = \sum_{-\infty}^{\infty} c_n e^{inx}$. But, alas, this is not the case. It turns out that L^2 convergence does not guarantee uniform, or even pointwise convergence. Further, not only are there integrable functions whose Fourier Series does not pointwise converge, but there are even continuous functions whose Fourier Series does not converge pointwise! Rather than end on a sour note however, we will show a few examples of when we can guarantee pointwise, or even uniform convergence, by strengthening the requirements our functions must satisfy! We will give two different approaches: firstly with Lipschitz continuous functions and second with piecewise continuous functions! Our central tool in these proofs will be the same trick we used with the proof of the Weierstrass Approximation Theorem, we will use a special Kernel and use Convolution!

Definition 9.18. Just as we used the Landau Kernel for Convolution in the proof of the approximation theorem, we would like a new Kernel to relate our function and its Fourier Series. We define the **Dirichlet Kernel** as $D_N(x) = \frac{1}{2\pi}\sum_{-N}^{N} e^{inx}$. We can rearrange as follows to get another expression

$$D_N(x) = \frac{1}{2\pi}\sum_{-N}^{N} e^{inx} = \frac{1}{2\pi}\sum_{n=0}^{2N}(e^{ix})^{n-N} = \frac{1}{2\pi}\frac{e^{-iNx} - e^{i(N+1)x}}{1 - e^{ix}}$$

$$= \frac{1}{2\pi}\left(\frac{e^{i\left(N+\frac{1}{2}\right)x} - e^{-i\left(N+\frac{1}{2}\right)x}}{e^{i\frac{1}{2}x} - e^{-i\frac{1}{2}x}}\right) = \frac{\sin((N+\frac{1}{2})x)}{2\pi\sin\left(\frac{1}{2}x\right)}.$$

We also have that taking the Convolution (on $[0, 2\pi]$)

$$(D_N(x) * f) = \int_0^{2\pi} D_N(x-t) f(t)\, dt = \int_0^{2\pi} \frac{1}{2\pi} \sum_{-N}^{N} e^{in(x-t)} f(t)\, dt$$

$$= \sum_{-N}^{N} e^{inx} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int}\, dt = \sum_{-N}^{N} c_n e^{inx} = S_N(f).$$

Finally our last observation concerning the Kernel is the following

$$\int_0^{2\pi} D_N(x)\, dx = \sum_{-N}^{N} \frac{1}{2\pi} \int_0^{2\pi} e^{inx}\, dx = 1.$$

We can use the Dirichlet Kernel to prove that for a special class of continuous functions, the Fourier Series will converge pointwise!

Definition 9.19. Let $f : X \to Y$, where X, Y are metric spaces. We say that f is **Lipschitz Continuous** if there exists a $K \in \mathbb{R}$ such that for all $x, y \in X$ we have that $d_Y(f(x), f(y)) < K d_X(x, y)$. It's easy to see that every Lipschitz Continuous function, is also uniformly continuous, and any $\mathcal{C}^1$ function is Lipschitz Continuous.

Theorem 9.20. Let $f : \mathbb{R} \to \mathbb{C}$ be Lipschitz Continuous and 2π periodic. Then $S_n(f) \to f$ pointwise on $\mathbb{R}$.

Proof. We need to prove that for any fixed value x_0 that $\lim_{n\to\infty} Sn(f, x_0) = f(x_0)$. By assumption of 2π perodicity it does not matter which interval we integrate over whether it be $[0, 2\pi]$ or $[-\pi, \pi]$, but it will be more conveient to integrate over $[-\pi, \pi]$ for this problem! We have that

$$f(x_0) = f(x_0) \int_0^{2\pi} D_N(t)\, dt = \int_{-\pi}^{\pi} f(x_0) D_N(t)\, dt.$$

Let

$$g(t) = \frac{f(x_0 - t) - f(x_0)}{\sin\left(\frac{t}{2}\right)}.$$

Then since $\lim_{t\to 0} g(t) < M$, and g is continuous for all $x \neq 0$, it is bounded on $[-\pi, \pi]$ and therefore integrable on this region (despite g not being defined at 0, excluding a point has no impact on the integral). Next we compute that

$$|S_N(f, x_0) - f(x_0)| = \left| \int_{-\pi}^{\pi} f(x_0 - t) D_N(t) - f_(x_0) D_N(t)\, dt \right|$$

$$= \left| \int_{-\pi}^{\pi} D_N(t)(f(x_0 - t) - f(x_0)\, dt \right| = \left| \int_{-\pi}^{\pi} g(x) \sin\left(\left(N + \frac{1}{2} \right) t \right)\, dt \right|.$$

$$= \left| \int_{-\pi}^{\pi} g(x) \sin\left(\frac{t}{2}\right) \cos(Nt)\, dt + \int_{-\pi}^{\pi} g(x) \cos\left(\frac{t}{2}\right) \sin(Nt)\, dt \right|.$$

Since g is integrable so are $g(t)\sin(t/2)$ and $g(t)\cos(t/2)$. Then the Riemann-Lebesgue Lemma says that both of the above integrals tend to zero as $N \to \infty$. □

We now give a different approach for a different class of functions.

Definition 9.21. A function f is **Piecewise Continuous** if it is continuous at all but finitely many points, and the right and left hand limits exist at every point.

We now prove another pointwise result, due to Dirichlet.

Theorem 9.240 (Dirichlet's Theorem) Let f be a 2π periodic piecewise continuous real function, that is differentiable at all points of continuity. Then for any $x_0 \in \mathbb{R}$ we have that

$$\lim_{n\to\infty} S_N(f, x_0) = \frac{\lim_{t\to x_0+} f(t) + \lim_{t\to x_0-} f(t)}{2}.$$

In other words, $S_N(f)$ converges pointwise at all continuous points, and at discontinuities it converges to the average of the right and left hand limits.

Proof. Let x_0 be a point where f is continuous. Then there exists some $r > 0$ so that on $\overline{B_r(x_0)}$, f is continuous and differentiable. The derivative on this region is bounded by say M, so that for all $x, y \in B_r(x_0)$ we have that $|f(x) - f(y)| < M|x - y|$ by the Mean Value

Theorem, so that f is Lipschitz continuous here and so the result follows from the above case.

Now we investigate a point of discontinuity. Let x_0 be a point of discontinuity and let $a = \lim_{t\to x_0+} f(t)$ and $b = \lim_{t\to x_0-} f(t)$. Now let $L = \frac{a-b}{2}$. Then we define a new function

$$g(x) = \begin{cases} -L & x < x_0 \\ L & x > x_0 \\ f(x_0) + \frac{a+b}{2} & x = x_0 \end{cases}.$$

Then for all $x \neq x_0$ we have $S_N(g, x) \to g(x)$. For $x = x_0$ the Fourier Series converges to 0, this is because g is odd about x_0. Then the function $f - g$ is continuous about x_0 and $(f-g)(x_0) = \frac{a+b}{2} = b + L = a - L$ so that the left and right limits agree with $(f-g)(x_0)$ so that $f - g$ is also continuous at x_0. Further, $f - g$ is differentiable about x_0 so that it isn't hard to see that the Lipschitz Criterion holds at x_0. Then we see that $S_N(f - g, x_0) \to \frac{a+b}{2}$, so that finally we conclude that

$$\lim_{N\to\infty} S_N(f, x_0) = \lim_{N\to\infty} S_N(f - g, x_0) + S_N(g, x_0) = \frac{a+b}{2}.$$

This completes the proof as this can be performed at every point of discontinuity. $\square$

We end this chapter showing that if the above theorem is slightly strengthened, then we can instead guarantee uniform convergence!

Theorem 9.23. Let f be a 2π periodic continuous real function, whose derivative function is piecewise continuous. Then $S_N(f) \to f$ uniformly on all of $\mathbb{R}$.

Proof. By Dirichlet's Theorem we immediately have that $S_N(f) \to f$ pointwise on $[-\pi, \pi]$, so we must show this convergence is uniform. Since f' is integrable, we can calculate its

Fourier Coefficients, denoted c_n'. Using integration by parts we have that

$$|c_n'| = \left|\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx\right| = \left|n\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx\right| = n|c_n|.$$

Parseval's Theorem tells us that

$$\sum_{-\infty}^{\infty} |c_n'|^2 = \sum_{-\infty}^{\infty} n^2|c_n|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f'(x)|^2\,dx.$$

Then an infinite series version of the Cauchy Schwarz Inequality gives

$$\left|\sum_{n-\infty}^{\infty} |c_n|\right| \leq \left|\sum_{-\infty}^{\infty} \frac{1}{n^2}\right|^{\frac{1}{2}} \cdot \left|\sum_{-\infty}^{\infty} n^2|c_n|^2\right|^{\frac{1}{2}} < \infty.$$

Since $\sum |c_n|$ converges The Weierstrasss M-Test guarantees uniform convergence of $\sum c_n e^{inx}$!

$\square$

Chapter 10

Lebesgue Measure

10.1 Motivation and Sigma Algebras

We have completed what is largely considered the "core" of an undergraduate Analysis sequence.

The use of Fourier Series for problems like Partial Differential Equations, and other areas of research is incredibly important and common. It turns out that some time ago, as people were solving problems related to Differential Equations and Fourier Series, they found that they were quite limited. The major limitation is the weakness of the Riemann Integral. In the section of Sequences of functions we almost saw that the nicer properties our functions satisfied, the harder it was to guarantee taking limits of functions preserved these properties. We saw for Differentiation that not even Uniform Convergence was enough and for continuity that Uniform Convergence did do the trick. One would hope that since being integrable is significantly weaker than being continuous, that we might not even need Uniform Convergence, but this turned out to be false. Further, the issue was typically not that the limits didn't agree, but that the limit function could not be guaranteed to be integrable! Along with this, last chapter we developed some beautiful theory concerning abstract Hilbert Spaces hoping to use it with Fourier Series, only for the L^2 Norm to not be complete on the set of Riemann Integrable functions. At this point, if we want to push any further it is time to consider a new method of integration, and it turned out that Lebesgue was the one to provide the solution!

Lebesgue's solution was to change how we integrate. Instead of partitioning the domain $[a, b]$ into subintervals and approximating the function over this range, we instead work the other way! If the range is a subset of $\mathbb{R}$, then we can partition the range into subsets $[x_0, x_1], ..., [x_{n-1}, x_n]$ and then take the inverse images of these subranges $f^{-1}([x_{i-1}, x_i])$ and if we can "measure" the length/area/volume of this inverse image in a way that makes sense then we can approximate the integral as

$$\sum_{i=0}^{n} m(f^{-1}([x_{i-1}, x_i]) \cdot x_i.$$

The following picture elucidates this.

It might not be immediately obvious how this improves integration, but it turns out that with this sort of idea we can integrate much more abstract functions and on types of domains that are not simple intervals! Consider for example the Dirichlet Function, $D(x)$, on $[0, 1]$. If we decide that the measure of the rationals in $[0, 1]$ is zero, then we can actually integrate this function as $m(D^{-1}(\{1\})) = 0$ and $m(D^{-1}(\{0\})) = 1$ so that

$$\int_0^1 D(x)\, dx = 0 \cdot m(D^{-1}(\{0\})) + 1 \cdot m(D^{-1}(\{1\})) = 0 + 0 = 0.$$

It follows that we can integrate any function where inverse images are "measurable", which we will see is quite a large class of functions!

At this point we should begin discussing what a "Measure" is. We will formally define a measure in the section on Measure Spaces, so for now we will be quite informal. A measure is a generalization of the idea of length, area, volume etc. It essentially assigns a size to subsets of some space. For this chapter since we are doing a deeper dive into Real Analysis

we will typically be working with subsets of $\mathbb{R}$ so that you can think of Measure as the total length of a set. That being said we will give abstract definitions in these first few sections, so that when we generalize from $\mathbb{R}$ to $\mathbb{R}^n$ or other space we need not redefine things!
We now want to develop how we "measure" subsets of $\mathbb{R}$. For this reason, we should introduce the idea of an uncountable sum! For if we can measure the size of each individual point in $\mathbb{R}$, then the measure of a set would just be the sum of the measures of all the points! It will be shown soon that this nice idea does not quite workout.

Definition 10.1. Let A be any set of positive real numbers. Then we define $\sum_{a\in A} a$ as $\sup\{\sum_{i=1}^{n} a_n\}$ where this sup ranges over all finite subsets of A. The next theorem shows why this isn't a useful concept.

Theorem 10.2. Suppose A is a set of positive real numbers when $\sum_{a\in A} a < \infty$. Then there are at most countably many $a \in A$ such that $a > 0$. In essence, countable sums are as good as it gets!

Proof. Suppose $\sum_{a\in A} a = M < \infty$. Let $n \in \mathbb{N}$ and consider the set of $a \in A$ such that $a > \frac{1}{n}$. Let $K = \lceil \left(\frac{M}{n}\right) \rceil$. We see then that

$$\sum_{i=0}^{K} \frac{1}{n} > M = \sum_{a\in A} a \geq \sum_{a\geq \frac{1}{n}} a \geq \sum_{a\geq \frac{1}{n}} \frac{1}{n}.$$

So that for each $n \in \mathbb{N}$ there are finitely many $a \in A$ so that $a \geq \frac{1}{n}$. Taking the union over all $n \in \mathbb{N}$, we see there are at most countably many $a > 0$. $\square$

The above theorem shows that we cannot break our sets into uncountably many pieces and measure them up, so we will do the next best thing. We will require whatever measure we build to be able to be split up into countably many pieces, and then the measure of the whole should equal the sum of the measures of the countably many sub-pieces. This property is called **Countable Additivity**. Whatever measure we develop we might require that it agrees with the lengths of intervals ie $m([a, b]) = b - a$, it is invariant under translation so that sliding the measure to a new place does not change it, and that every subset of $\mathbb{R}$ is measurable! We make the following definition

Just as the case of uncountable sums however, this is impossible.

It turns out that if you have a countable additive, translation invariant measure where every subset of $\mathbb{R}$ is measurable, then $m([0, 1])$ is either 0 or infinite. Neither of these are useful, so instead we will restrict ourselves to a suitable class of so called "measurable" sets, that do satisfy the above rules.

Definition 10.3. Let X be any set. Then a collection of subsets $\mathcal{A} \subseteq \mathcal{P}(X)$ is called a $\sigma-$**Algebra** if

(1) $X \in \mathcal{A}$,

(2) $a \in \mathcal{A}$ implies that $a^c \in \mathcal{A}$,

(3) If $a_n \in \mathcal{A}$ is a countable collection of sets in $\mathcal{A}$, then

$$\bigcup_{n=1}^{\infty} a_n \in \mathcal{A}.$$

The above definition with DeMorgan's Laws also implies that $\mathcal{O}$ is closed under countable intersections.

We also see that (1) and (2) imply that $\emptyset \in \mathcal{A}$, and so that (3) implies that $\mathcal{A}$ is closed under finite unions and intersections, taking all $a_n = \emptyset$ for all $n \geq N$, for some N.

$\sigma-$Algebra's are perfect for Measure Theory as if our all our measurable sets lie in some σ-Algebra, then we can union countably many sets and the resulting set will still be measurable! It also makes sense that if we can measure a set we can measure its complement.

10.2 Borel Sets

It's a nice fact, that we will use later, that given any collection of sets we can always find a $\sigma-$algebra containing these sets that is minimal. The next theorem shows this.

Theorem 10.4. Let X be any set. Then given a collection of subsets $\mathcal{O}$, there exists a smallest $\sigma-$algebra, $\mathcal{A}$, containing $\mathcal{O}$.

Proof. Let $S = \{O_1, O_2, ...\}$ be the collection of all $\sigma-$algebras containing $\mathcal{O}$. S is nonempty, as $\mathcal{P}(X)$ is a $\sigma-$algebra containing $\mathcal{O}$. Then we can define

$$\mathcal{A} = \bigcap_{O \in S} O.$$

Clearly $\mathcal{O} \subseteq \mathcal{A}$. Then we claim that $\mathcal{A}$ is a $\sigma-$algebra. If $A \in \mathcal{A}$, then $A \in \mathcal{O}_\rangle$ for all $O_i \in S$. Since each O_i is a $\sigma-$algebra, it follows that $A^c \in O_i$ for all i, so that $A^c \in \mathcal{A}$.
The same idea shows that $\mathcal{A}$ is closed under countable unions.
The fact that $\mathcal{A}$ is the smallest $\sigma-$algebra containing $\mathcal{O}$ follows immediately from how we defined it, completing the proof. $\square$

The previous theorem gives a great tool of constructing a nice collection of sets to measure for any Topological Space. Because open sets are the key ingredients of any topological space, if we were to define a Measure on some Topological Space we'd certainly hope that all the open sets are measurable. With that said we introduce the Borel Sets to deal exactly that.

Definition 10.5. Let (X, τ) be a Topological Space. We define the **Borel Sets**, denoted $\mathcal{B}$(X), of (X, τ) to be the smallest $\sigma-$algebra containing τ. Such a σ-algebra exists by Theorem 10.XX above.
By the $\sigma-$algebra properties we also see that $\mathcal{B}(X)$ is the smallest $\sigma-$algebra containing all closed sets. There are many other sets besides open and closed sets, as though a countable

union of closed sets need not be open, it must be in $\mathcal{B}(X)$. Sets that are countable unions of closed sets are often called F_σ sets, and countable intersections of open sets are often called G_δ sets. It follows that every F_σ and G_δ set is also in $\mathcal{B}(X)$.
One note about notation, the same set X may have many different topologies so one should specify the topology, unless it is obvious, before discussing its Borel Sets.

Theorem 10.6. Consider the set $\mathcal{B}(\mathbb{R})$ (with the standard topology). Then every countable set, finite set, and every half-open interval $[a,b),(a,b]$ is in $\mathcal{B}(\mathbb{R})$. This follows as all of the previous sets are F_σ sets. The first two are obvious but to see that $[a,b)$ is an F_σ set consider

$$[a,b)=\bigcup_{n=1}^{\infty}\left[a,b-\frac{1}{n}\right].$$

10.3 Measure Spaces

We have now successfully discussed $\sigma-$algebra's and are ready to discuss what we exactly mean by a Measure! But first we will start with a Measurable Space.

Definition 10.7. Let X be a set and let $\mathcal{A}$ be a $\sigma-$algebra on X. Then the pair $(X,\mathcal{A})$ is called a **Measurable Space**. This is because there is no measure assigned to the space.

With a Measurable space, we are now ready to define exactly what we mean by a Measure and turn the Measurable Space into a Measure Space!

Definition 10.8. Let $(X,\mathcal{A})$ be a measurable space. Then we define a **Measure** on $(X,\mathcal{A})$ to a be a function $m:\mathcal{A}\to[0,\infty]$, satisfying
(1) $m(\emptyset)=0$ and
(2) If $E_1,E_2,...$ is a countable collection of disjoint sets in $\mathcal{A}$, then

$$m\left(\bigcup_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}m(E_n).$$

If we assign a measure m to a measurable space $(X, \mathcal{A})$, then the ordered triple $(X, \mathcal{A}, m)$ is called a **Measure Space**! Just as with topology and metric spaces, if the σ-algebra and measure on the space are obvious we will likely refer to X as the measure space!

It is also clear that (1) and (2) imply disjoint finite additivity of the measure!

Theorem 10.9. Let $(X, \mathcal{A}, m)$ be a measure space and let $E_1, E_2, ...$ be a countable collection of sets in $\mathcal{A}$. Then

(1) If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $m(A) \leq m(B)$.

(2)

$$m\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m(E_n),$$

(3) If E_n is an increasing sequence of sets, ie $E_n \subseteq E_{n+1}$ then

$$m\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} m(E_n),$$

(4) If E_n is a decreasing sequence $(E_{n+1} \subseteq E_n)$ and $M(E_1) < \infty$, then

$$m\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} m(E_n),$$

These three properties are sometimes called monoticity of measure, countable subadditivity, continuity from below, and continuity from above respectively, respectively.

Proof. For (1) this is easy as $B - A = B \cap A^c \in \mathcal{A}$ so by finite additivity we have $m(B) = m(B - A) + m(A) \geq m(A)$.

For (2) note we are not requiring the sets be disjoint this time. So now we define a new sequence of sets D_n by taking $D_1 = E_1$ and $D_n = E_n - \bigcup_{i=1}^{n-1} D_i$. Then D_n is disjoint sequence of sets in $\mathcal{A}$ and for each $n \in \mathbb{N}$ we have $\bigcup_{i=1}^{n} D_i = \bigcup_{i=1}^{n} E_i$, so that we can conclude that

$\bigcup_{n=1}^{\infty} D_n = \bigcup_{n=1}^{\infty} E_n$. Further, since each $D_n \subseteq E_n$ by (1) we have $m(D_n) \leq m(E_n)$. So that

$$m\left(\bigcup_{n=1}^{\infty} E_n\right) = m\left(\bigcup_{n=1}^{\infty} D_n\right) = \sum_{n=1}^{\infty} m(D_n) \leq \sum_{n=1}^{\infty} m(E_n).$$

For (3) we do a similar trick as above, set $O_1 = A_1$ and $O_n = A_n - O_{n-1}$. Then $\bigcup_{n=1}^{\infty} O_n = \bigcup_{n=1}^{\infty} E_n$ so that

$$m\left(\bigcup_{n=1}^{\infty} E_n\right) = m\left(\bigcup_{n=1}^{\infty} O_n\right) = \sum_{n=1}^{\infty} m(O_n) = \sum_{n=1}^{\infty} m(E_n) - m(O_{n-1}) = \lim_{n\to\infty} m(E_n).$$

Finally (4) follows from applying (3) to the sets $Q_n = A_1 - A_n$. □

10.4 Measure Completion

We now come to our last tool abstract tool well need, the concept of a complete measure space!

Definition 10.10. Suppose $(X, \mathcal{A}, m)$ is a measure space. A set $O \in \mathcal{A}$ is called a **Null Set** if $m(O) = 0$. If every subset of a null set is also a null set then we say that $(X, \mathcal{A}, m)$ is a **Complete Measure Space**.

This property is nice to have as we would expect that any subset of a zero measure set is also zero measure. However, this is not always the case. So we sometimes need to complete our measure space, similarly to how we can complete a metric space. The next theorem shows this!

Theorem 10.11. Let $(X, \mathcal{A}, m)$ be a measure space, and let N be the set of all subsets of all the null sets for this measure space. Then let $\mathcal{A}' = \{S \subseteq X | S = A \cup B$ where $A \in \mathcal{A}$ and $B \in N\}$, next define the set function $m' : \mathcal{A}' \to [0, \infty]$ by

$$m'(D) = m(A),$$

when $D = A \cup B$ with $A \in \mathcal{A}$ and $B \in N$.

Then $\mathcal{A}'$ is a $\sigma-$algebra and m' is a measure on $(X, \mathcal{A}')$, so that $(X, \mathcal{A}', m')$ is a complete measure space. This space is called the **Completion** of the original measure space.

Proof. First, we need to show that $\mathcal{A}'$ is a $\sigma-$algebra. Clearly, $X \in \mathcal{A}'$. If $S \in \mathcal{A}'$, then write $S = A \cup B$ for $A \in \mathcal{A}$ and $B \in N$. Then $S^c = (A \cup B)^c = A^c \cap B^c$. Let M be a null set containing B. Then we can rewrite B as $B = (F \cap (F^c \cup B)$ or $B^c = (F \cap (F^c \cup B))^c = (F^c \cup (F^c \cup B)^c)$ so that $(A^c \cap B^c) = (A^c \cap (F^c \cup (F^c \cup B)^c)) = ((A^c \cap F^c) \cup (A^c \cap (F^c \cup B)^c)) = ((A^c \cap F^c) \cup (A^c \cap (F \cap B^c))$. It's easy to see that $(A^c \cap F^c) \in \mathcal{A}$ and since $(A^c \cap (F \cap B^c)) \subseteq F$ it is in N so that $S^c \in \mathcal{A}'$.

Now let $E_1, E_2, ...$ be a countable sequence of sets in $\mathcal{A}'$, where each $E_n = A_N \cup B_n$, with $A_n \in \mathcal{A}$ and $B_n \in N$. Then

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} (A_n \cup B_n) = \bigcup_{n=1}^{\infty} A_n \cup \bigcup_{n=1}^{\infty} B_n.$$

We have $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ and since a countable union of null sets is a null set it follows that $\bigcup_{n=1}^{\infty} B_n \in N$, so that $\mathcal{A}'$ is a $\sigma-$algebra.

Now we must show that m' is a measure on $(X, \mathcal{A}')$. To see that m' is well defined consider we write $A \in \mathcal{A}'$ as $A = B \cup C$ and $A = F \cup G$ for $B, C \in \mathcal{A}$ and $C, G \in N$. Then let W, V be null sets containing C and G respectively. Then we see that $B \cup W \cup V = F \cup W \cup V$ so that $m(B) \leq m(B \cup C \cup G) = m(F \cup C \cup G) = m(F)$ and the other direction shows $m(F) \leq m(B)$, showing that m' is well defined. It is clear that $m'(\emptyset) = 0$ and that $m'(A) = m(A)$ if

$A \in \mathcal{A}$. So now let $E_1, E_2, ...$ be a countable collection of sets in $\mathcal{A}'$, and rewrite each E_n as $E_n = A_n \cup B_n$ for $A_n \in \mathcal{A}$ and $B_n \in N$. Then

$$m'\left(\bigcup_{n=1}^{\infty} E_n\right) = m\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n) = \sum_{n=1}^{\infty} m'(E_n).$$

This completes the proof. □

10.5 Lebesgue Outer Measure

After developing some abstract theory concerning measure, we now want to actually construct a useful measure on $\mathbb{R}$. The measure we will construct will be called the Lebesgue Measure, and it will satisfy everything we would hope about measure on $\mathbb{R}$ on the $\sigma-$algebra it is defined on.

To start we will however first define the Lebesgue Outer Measure. We will define this so called outer measure for every subset of $\mathbb{R}$, and then we will restrict it to a suitable σ-algebra!

Definition 10.12. Let $A \subseteq \mathbb{R}$. If $I = (a, b)$ we define the **Length** of I as $L(I) = b - a$, this was done already in chapter 6, but we review here. We define the **Lebesgue Outer Measure** of A as

$$m^*(A) = \inf\left\{\sum_{n=1}^{\infty} L(I_n) \Big| \{I_n\} \text{ countable collection of open intervals with } A \subseteq \bigcup_{n=1}^{\infty} I_n\right\}.$$

We have a few facts immediately. It is easy to see from the definition that if $A \subseteq B$ then $m^*(A) \leq m^*(B)$ so that outer measure is monotone. Further, $m^*(\{x\}) = 0$. It's easy to extend this to any countable set, for if $E = \{x_1, x_2, ...\}$ is countable consider the intervals $I_n = (x_n - \frac{\epsilon}{2^{n+1}}, x_n + \frac{\epsilon}{2^{n+1}})$ (this is essentially the same proof as in chapter 6).

Theorem 10.13. The function m^* is countably subadditive.

Proof. This uses the same trick as above. Let $E_1, E_2, ...$ be a countable collection of sets. We need to show that

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m^*(E_n).$$

In the case that one $m^*(E_i) = \infty$, this is obvious so assume each has finite outer measure and let $\epsilon > 0$. Then for each $n \in \mathbb{N}$ there exists a countable collection of open intervals $I_{k,n}$ so that $\sum_{k=1}^{\infty} L(I_{k,n}) < m^*(E_n) + \frac{\epsilon}{2^n}$ and where $E_n \subseteq \bigcup_{k=1}^{\infty} I_{k,n}$. We have that

$$\bigcup_{n=1}^{\infty} E_n \subseteq \bigcup_{n=1}^{\infty}\left(\bigcup_{k=1}^{\infty} I_{k,n}\right).$$

So then it follows that

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty}\left[\sum_{k=1}^{\infty} L(I_{k,n})\right] \leq \epsilon + \sum_{n=1}^{\infty} m^*(E_n).$$

Since this holds for all $\epsilon > 0$, the result follows. □

The above proof also tells us that m^* is finitely subadditive.

10.6 Lebesgue Measure

Having defined what the Lebesgue Outer Measure is, we would now like to turn this into a true measure. An incredibly easy method of doing this is given by Carathéodory.

Definition 10.14. Let $E \subseteq \mathbb{R}$. We say that E is **Measurable** if for all $A \subseteq \mathbb{R}$, that

$$m^*(A) = m^*(A \cup E) + m^*(A \cup E^c).$$

Informally, this says that a set is measurable if it nicely splits all other sets with respect to the Outer Measure.

We denote the set of all measurable subsets of $\mathbb{R}$ by $\mathcal{M}$.

We then define the **Lebesgue Measure** $\lambda : \mathcal{M} \to [0, \infty]$ by $\lambda(E) = m^*(E)$ for measurable sets E.

Theorem 10.256 (Lebesgue Measure Properties) The collection $(\mathbb{R}, \mathcal{M}, \lambda)$ is a complete measure space.

Further, every Borel Set is measurable, in parituclar interval's are measurable with $\lambda([a,b]) = \lambda([a,b)) = \lambda((a,b]) = \lambda((a,b)) = b - a$. If E is measurable, then the translate $E + y = \{x + y | x \in E\}$ is measurable and $\lambda(E + y) = \lambda(E)$.

This theorem asserts that the Lebesgue Measure is essentially our ideal measure for $\mathbb{R}$. The proof is quite long so we break it up into several steps.

Step 1: $\mathcal{M}$ is closed under complements and finite unions.

Proof. We immediately have from the definition that if E is measurable, that E^c is measurable. Now suppose E_1 and E_2 are measurable and let A be any subset of $\mathbb{R}$. By finite subadditivty we have that $m^*(A) \leq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c)$. For the other direction we have that

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c)$$

$$= m^*(A \cap E_1 \cap E_2) + m^*(A \cap E_1 \cap E_2^c) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c).$$

Since $A \cap E_1 \cap E_2 \subseteq A \cap (E_1 \cup E_2)$ and all of the above terms are positive we conclude that

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c).$$

This completes the proof of Step 1. $\square$

Step 2: $\mathcal{M}$ is a $\sigma-$algebra.

Proof. We need to prove that $\mathcal{M}$ is closed under countable unions. Assume $E_1, E_2, ...$ is a countable collection of sets in $\mathcal{M}$, with union E. We may assume that the $E_n's$ are disjoint, for if they are not the following are disjoint sets in $\mathcal{M}$ whose union is E:

$G_N = E_N - (E_{N-1} \cup ... \cup E_1)$.

To save space we will write $E_1 \cup ... \cup E_N = F_N$. We claim that by induction we have that

$$m^*(A \cap F_N) = \sum_{i=1}^{N} m^*(A \cap E_i).$$

The case $N = 1$ is obvious. Assume the $N - 1$ case holds. Then since $F_N = F_{N-1} \cup E_N$ we have that with E_N measurable that

$$m^*(A \cap F_N) = m^*(A \cap F_N \cap E_N) + m^*(A \cap F_N \cap E_N^c)$$

$$= m^*(A \cap E_N) + m^*(A \cap F_{N-1}) = \sum_{i=1}^{N} m^*(A \cap E_i).$$

The middle equality follows by the disjointness of the $E_n's$.

Now fix some N. Then since F_N is measurable, $F_N \subseteq E$ so that $E^c \subseteq F_N^c$, and the previous statement we get that

$$m^*(A) = m^*(A \cap F_N) + m^*(A \cap F_N^c) \geq m^*(A \cap F_N) + m^*(A \cap E^c) = \sum_{i=1}^{N} m^*(A \cap E_N) + m^*(A \cap E^c).$$

Since this holds for any N the above plus subaddivity gives that

$$m^*(A) \geq \sum_{n=1}^{\infty} m^*(A \cap E_N) + m^*(A \cap E^c) \geq m^*(A \cap E) + m^*(A \cap E^c).$$

So that E is measurable and $\mathcal{M}$ is a $\sigma-$algebra. $\square$

Step 3: $(\mathbb{R}, \mathcal{M}, \lambda)$ is a complete measure space.

Proof. It is easy to see that $\lambda(\emptyset) = m^*(\emptyset) = 0$. Now suppose $E_1, E_2, ...$ are a countable collection of sets in $\mathcal{M}$. By subaddivity of outer measure we already have that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \lambda(E_n).$$

Using essentially the same inductive proof used in Step 2 above, we get that for each N that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \geq \lambda\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} \lambda(E_i).$$

Since the expression above is independent of N we conclude that

$$\lambda\left(\bigcup_{n=1}^{\infty} E_n\right) \geq \sum_{n=1}^{\infty} \lambda(E_n).$$

This shows that λ is a measure on $\mathcal{M}$.

To see that the space is complete we have that if A is any set and N is a null set with $S \subseteq N$, then $m^*(S) \leq m^*(N) = \lambda(N) = 0$. Then $m^*(A \cap S) \leq m^*(S) = 0$ and $m^*(A \cap S^c) \leq m^*(A)$ giving $m^*(A) \geq m^*(A \cap S) + m^*(A \cap S^c)$, so that S is measurable. This completes Step 3. $\square$

Step 4: Every interval is measurable. Since every open set is can be expressed as a countable union of intervals, it follows every open set is measurable and (since $\mathcal{M}$ is a $\sigma-$algebra containing every open set) that every Borel Set is measurable.

Proof. It suffices to prove that intervals (a, ∞) are measurable. Let A be any set. If $A \subseteq (a, \infty)$ or $A \subseteq (-\infty, a]$ the result is obvious, so assume otherwise. Let $\{I_n\}$ be a countable collection of open intervals covering A. We have that $m^*(A) \leq \sum_{n=1}^{\infty} L(I_n)$. Now let $I_{n,l} =$

$\{I_n \cap (-infty, a)\}$ and $I_{n,r} = \{I_n \cap (a, \infty)\}$. Then $I_{n,l}$ and $I_{n,r}$ are both countable collections of open intervals covering $A \cap (-\infty, a)$ and $A \cap (a, \infty)$. Then

$$m^*(A \cap (-\infty, a)) + m^*(\{a\}) + m^*(A \cap (a, \infty) \leq \sum_{n=1}^{\infty} L(I_{n,l}) + \sum_{n=1}^{\infty} L(I_{n,r}) = \sum_{n=1}^{\infty} L(I_n).$$

Since this holds for any cover of A, we see that $m^*(A) \geq m^*(A \cap (a, \infty) + m^*(A \cap (-\infty, a]))$, so that (a, ∞) is measurable. □

Step 5: The measure of an interval is its length.

Proof. It suffices to prove this for intervals of the form $[a, b]$. For the other cases we can approximate (a, b) by $[a + \epsilon, b - \epsilon]$ arbitrarily close, and the half intervals follow. It also suffices to prove this in the case neither $a, b = \pm\infty$.

Now suppose we have an interval $[a, b]$. It's easy to see that $m^*([a, b]) \leq b - a$ as for every $\epsilon > 0$ the sets $(a - \epsilon, a + \epsilon), (a, b), (b - \epsilon, b + \epsilon)$, cover $[a, b]$ and their lengths are $b - a + 4\epsilon$. The other direction is trickier. Suppose we have any countable collection of open intervals covering $[a, b]$, say $\{I_n\}$. The Heine-Borel Theorem guarantees a finite subcover $I_1 = (a_1, b_1), ..., I_N = (a_n, b_n)$. We can assume that $a \in I_1$, for $a \in I_k$ for some k, and we can rename to have our result. But then proceeding we assume that $b_1 \in I_2$, for if it is not we again rename and proceed. We repeat this process until we arrive at a set I_N containing b. In all we have a collection of sets $a \in I_1$, $b_1 \in I_2$, $b_k \in I_{k+1}$, and $b \in I_N$. It follows then that

$$\sum_{n=1}^{\infty} L(I_n) \geq \sum_{k=1}^{N} L(I_k) = (b_N - a_N) + ...(b_1 - a_1) \geq b_N - a_1 \geq b - a.$$

We conclude then that $m^*([a, b]) \geq b - a$, and so by definition we have $\lambda([a, b]) = b - a$ completing the proof. □

Step 6: The translate of a measurable set, and the measures agree. This step completes the proof of Theorem 10.15

Proof. Let E be any set and let $\{I_n\}$ be a countable collection of open intervals covering E. If $y \in \mathbb{R}$ then it follows the sets $\{I_n + y\}$ cover $E + y$ and their lengths agree, so that $m^*(E) = m^*(E + y)$. Now suppose E is measurable and A is any set. Then $m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) = m^*(A \cap (E + y)) + m^*(A \cap (E + y)^c)$, so that $E + y$ is measurable, and the fact that their measures agree follows from the fact m^* is translation invariant. $\square$

To end this chapter we introduce the concept of Regular and Radon measures and prove that the Lebesgue Measure is a Radon Measure.

Definition 10.16. Let (X, τ) be a topological space. A measure on some measure space $(X, \mathcal{A}, m)$ is **Outer Regular** if for all $a \in \mathcal{A}$ we have

$$m(A) = \inf\{m(O)|A \subseteq O, \quad O \text{ is Open and Measurable}\}.$$

Similarly, m is **Inner Regular** if

$$m(A) = \sup\{m(C)|C \subseteq A, \quad C \text{ is Compact and Measurable}\}.$$

If m is both Inner and Outer Regular we say m is **Regular**.

A **Borel Measure** is a measure that is defined on all Borel Sets.

If a measure m is a Regular Borel measure such that the measure of every compact set is finite, then m is simply called a **Radon Measure**. Radon Measures are measures that in some sense agree with what we expect for Topological Spaces.

Theorem 10.258 (Regularity of Lebesgue Measure) For all $\epsilon > 0$ and for all measurable sets E there exists an open set O and a closed set C, so that $C \subseteq E \subseteq O$, where $\lambda(O - E) < \epsilon$ and $\lambda(E - C) < \epsilon$. Further, the measure of any compact set is finite. This proves that the Lebesgue Measure is a Radon Measure.

Proof. Let $\epsilon > 0$ and let $E \in \mathcal{M}$. By definition we have that there is a countable collection of open intervals $\{I_n\}$ satisfying $\sum_{n=1}^{\infty} L(I_n) < \lambda(E) + \epsilon$. It follows that setting $O = \bigcup_{n=1}^{\infty} I_n$, we have that O is open (and measurable) with $\lambda(O) < \sum_{n=1}^{\infty} L(I_n)$. Then we have
$\lambda(O) - \lambda(E) < \epsilon$, and since $E \subseteq O$, it follows that $\lambda(O - E) = \lambda(O) - \lambda(E) < \epsilon$.
By this we know that since E^c is measurable there is some open set O' with $\lambda(O' - E^c) < \epsilon$.
But now taking $C = (O')^c$, we see that $C \subseteq E$ and $\lambda(E - C) = \lambda(O' - E^c) < \epsilon$.
Finally, the fact that compact sets have finite measure is easy as for every compact set K we have $K \subseteq [-R, R]$ for some R and $\lambda([-R, R]) = 2R$. □

Chapter 11

Lebesgue Integration

Having spent a great deal of time developing the Lebesgue Measure and its useful properties, we now turn towards integration. The motivation behind Integration and general concept was outlined in the last chapter, so we will great straight to work. Throughout the Chapter we will be working with the Measure Space: $(\mathbb{R}, \mathcal{M}, \lambda)$, though we will often state things in terms of general Measure Spaces. We will start by discussing what we mean by Measurable Functions, then gradually we will build up integration, and finally we will end by discussing the advantages over the Riemann Integral!

11.1 Measurable Functions

We briefly motivated how we would define Measurable Functions in the last chapter. The idea was that preimages of interval's would be measurable sets allowing easy integration. Our definition will differ from this slightly, but has the same general idea.

Definition 11.1. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **Lebesgue Measurable** or simply **Measurable** if for all $a \in \mathbb{R}$, we have that $f^{-1}((a, \infty)) \in \mathcal{M}$. This definition extends to extended real functions by saying that $f^{-1}((a, \infty]) \in \mathcal{M}$.

Theorem 11.2. A function $f : \mathbb{R} \to \mathbb{R}$ is measurable if and only if any of the following holds for all $a \in \mathbb{R}$:

(1) $f^{-1}([a,\infty)) \in \mathcal{M}$,

(2) $f^{-1}((-\infty,a)) \in \mathcal{M}$,

(3) $f^{-1}(-\infty,a]) \in \mathcal{M}$.

Proof. The proof essentially says any infinite interval works. It is clear that measurability implies (2) and that (1) and (3) are equivalent. So we need just prove measurability implies (1) and vise versa.

Suppose f is measurable, then for all $a \in \mathbb{R}$ we have $f^{-1}(a,\infty)) \in \mathcal{M}$. In particular this holds for $(a - \frac{1}{n}, \infty)$, for $n \in \mathbb{N}$. But then since each of these are measurable we have

$$\bigcup_{n=1}^{\infty} f^{-1}((a - \frac{1}{n}, \infty)) = f^{-1}([a,\infty)) \in \mathcal{M},$$

so that measurability implies (1).

Using the same trick if (1) holds, then for all $n \in \mathbb{N}$ we have $f^{-1}([a + \frac{1}{n}, \infty)) \in \mathcal{M}$, so that intersecting gives measurability of f. $\square$

The next theorem gives a more useful, and more general criterion for measurability of a function. We will use this to make a definition of measurability that holds for more general functions from any measure space to $\mathbb{C}$!

Theorem 11.3. A function $f : \mathbb{R} \to \mathbb{R}$ is measurable if and only if for every open set $O \subseteq \mathbb{R}$, we have that $f^{-1}(O)$ is measurable.

Proof. For the first direction assume that $f^{-1}(O) \in \mathcal{M}$ for every open set O. Then in particular this holds for all sets of the form (a,∞), proving measurability.

Now for the other direction assume that f is measurable. We have that $f^{-1}((a,b)) = f^{-1}(a,\infty) - f^{-1}(b,\infty)$. Since both of the latter are measurable we have that $f^{-1}((a,b))$ is measurable. Since every open subset of $\mathbb{R}$ can be represented as a countable union of interval's we are done. $\square$

Definition 11.4. Let $(X, \mathcal{A}, m)$ be a measure space and (Y, τ) a topological space, and let $f : X \to \tau$. We say that f is **Measurable**, if for every open set $O \in \tau$, we have that $f^{-1}(O) \in \mathcal{A}$. By the preceeding theorem for functions $f : \mathbb{R} \to \mathbb{R}$, this agrees with the old definition, another proof show that this also works for extended real functions. The most important cases will be $\mathbb{C}, \mathbb{R}$, and $\overline{\mathbb{R}}$.

This new definition gives rise to a very useful class of measurable functions, namely continuous functions!

Theorem 11.5. Suppose $(X, \mathcal{A}, m)$ is a measure space, (Y, τ) a topological space, and m is a Borel Measure. Then every continuous function $f : X \to Y$ is measurable. This follows from the above definition and the definition of Borel Measure's immediately.
We also have that if f measurable and g is continuous, then $g \circ f$ is measurable.
Finally, if f is a measurable complex function, with $f = u + iv$, then both u and v are measurable. This is because the projection functions are continuous.

We want to make the class of measurable functions as large as possible, as these are the functions we can integrate. The fact that continuous functions are measurable is a good step in the right direction! It turns out that the most useful fact about measurable functions, is there interactions with limiting processes. The next theorem will explore this concept!

Theorem 11.6. The Pointwise Supremum Function is defined for a sequence of functions f_n as

$$(\sup_n f_n)(x) = \sup_n \{f_n(x)\}.$$

The same idea gives rise to the function $(\limsup_{n\to\infty} f_n)(x)$.
With these functions defined, let $(X, \mathcal{A}, m)$ be a measure space and let $f_n : X \to \overline{\mathbb{R}}$ be a sequence of measurable functions. Then $(\sup_n f_n)(x), (\inf_n f_n)(x), (\limsup_{n\to\infty} f_n)(x)$, $(\liminf_{n\to\infty} f_n)(x)$, and if it exists $\lim_{n\to\infty} f_n(x)$ are all measurable.

We also have that $\max(f_1, f_2, ..., f_n)$ and $\min(f_1, f_2, ..., f_n)$ are measurable, this follows from the sup and inf cases.

Proof. We first prove that $(\sup_n f_n)(x)$ is measurable, the case of inf is the same idea. Let $a \in \overline{\mathbb{R}}$. Let $h(x) = (\sup_n f_n)(x)$. Then $h^{-1}((a,\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}(a,\infty])$, so that it is a countable union of measurable sets, proving measurability of h. The fact that $(\limsup f_n)(x)$ and $(\liminf f_n)$ are measurable follows from the facts that

$$\limsup a_n = \inf_n \{\sup_{k\geq n} a_k\},$$

and likewise for $\liminf$. Finally, the fact that $\lim f_n$ is measurable follows from the $\limsup$ and $\liminf$ cases. □

We see that many of the important limit processes we are familiar with preserve measurability, which is a massive improvement over Riemann Integrable functions.

To end this section we prove one more incredibly nice fact about measurable functions, namely that changes on null sets do not affect measurability! First recall that we say a property P holds **Almost Everywhere** (abbreviated to a.e.) if the set where P fails, say E, is a null set. If we need to be specific of the measure we might say m-a.e.

Theorem 11.7. Let $(X, \mathcal{A}, m)$ be a complete measure space and (Y, τ) a topological space. Suppose $f : X \to Y$ is measurable on X and $g : X \to Y$ is a function such that $f = g$ a.e. on X. Then g is measurable.

Proof. Suppose O is an open subset of Y. Then we know that $A = f^{-1}(O) \in \mathcal{A}$. Let $N \subseteq X$ where $f \neq g$, then by assumption $m(N) = 0$. Let $B = \{x \in X | g(x) \notin A \text{ and } g(x) \in O\}$. Then $B \subseteq N$, so B is measurable with $m(B) = 0$. Similarly, let $C = \{x \in A | g(x) \notin O\}$. Then $C \subseteq A \cap N \subseteq N$, so that C is measurable with $m(C) = 0$, also implying that $(A - C)$ is measurable. Finally, we have that $g^{-1}(O) = (A - C) \cup B$, which is the union of measurable sets so that it is measurable! □

11.2 Simple Functions

The following definition will be useful for representing simple functions.

Definition 11.8. Let E be some set. Then the **Indicator Function** of E is

$$\mathbb{1}_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

It follows that $\mathbb{1}_E$ is a measurable function if and only if E is a measurable set.

While we haven't defined integration yet, we should hope that however we would define it that we would have that $\int \mathbb{1}_E = m(E)$.

The next simplest type of function we would hope to integrate are the so called Simple Functions, which we now define.

Definition 11.9. Let $(X, \mathcal{A}, m)$ be a measure space and Y a topological space. Then a function $s : X \to Y$ is **Simple** if the range of f consists of finitely many points and f is measurable. For every simple function s there is a **Canonical Representation** given by

$$s(x) = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}(x),$$

where $A_k = f^{-1}(\{a_k\})$ and $\{a_1, ..., a_N\}$ are the distinct values s takes. It is clear that each A_k must be measurable!

It's an easy exercise that sums of simple functions are simple and that scalar multiples of simple functions are simple!

Simple functions are incredibly useful tools. Firstly, they are easy functions to define integrals for. This will be seen in the next section where we define the integral of a simple function, but the reader can likely already guess how the process will go! Another reason is the next important theorem!

Theorem 11.268 (The First Simple Approximation Theorem) Let $(X, \mathcal{A}, m)$ be a measure space and let $f : X \to [0, \infty]$. Then there exists a sequence of simple functions s_n so that $0 \leq s_n \leq s_{n+1} \leq f$, and $s_n \to f$ pointwise on X.

Proof. Let N be arbitrary. We will divide the interval $[0, N]$ into 2^N subdivisions and measure how much of the range of f is in each of these subdivisions. This is similar to what we did for Lower Riemann Sums when we partitioned the domain, but now we are partitioning the range of f.
For each $k = 0, 1, ..., 2^N - 1$, set $A_k = f^{-1}\left(\left[\frac{Nk}{2^N}, \frac{N(k+1)}{2^N}\right)\right)$ and $a_k = \frac{Nk}{2^N}$. Finally, set $a_{2^N} = N$ and $A_{2^N} = f^{-1}([n, \infty))$. Then let

$$s_N(x) = \sum_{k=0}^{2^N} a_k \mathbb{1}_{A_k}(x).$$

Since each A_k is measurable, the above is a simple function for each $N \in \mathbb{N}$. It isn't too difficult to see that the $s_n' s$ are monotone increasing. If $f(x) < \infty$, then there is an interval $[0, N]$ containing $f(x)$ and so $s_n(x) \to f(x)$. For $f(x) = \infty$ the sequence $s_n(x) = n$ so that $s_n \to \infty$. $\square$

Definition 11.11. Let $f : E \to \overline{\mathbb{R}}$, where E is any set. Then we define

$$f^+(x) = \max(f(x), 0),$$

and

$$f^-(x) = \max(-f(x), 0),$$

as the **Positive and Negative Parts** of f respectively. These are both functions from $E \to [0, \infty]$ with $f = f^+ - f^-$. Further, if we replace E with a measure space $(X, \mathcal{A}, m)$ and f is measurable, then both f^+ and f^- are measurable.

We now prove another result very similar to the First Simple Approximation Theorem, hence the naming.

Theorem 11.270 (The Second Simple Approximation Theorem) Let $(X, \mathcal{A}, m)$ be a measure space and let $f : X \to \mathbb{C}$. Then f is measurable if and only if there exists a sequence of simple functions s_n that converges to f. The same result holds for functions to $\overline{\mathbb{R}}$, and clearly to $\mathbb{R}$.

Proof. If there is a sequence of simple functions that converges to f, the result follows as the limit of a sequence of measurable functions is measurable.
Now suppose that $f : X \to \mathbb{C}$ is measurable, with $f = u+iv$. Then by above we have that u, v are both measurable real functions, so that u^+, u^-, v^+, and v^- are all measurable functions from $[0, \infty)$. By the First Simple Approximation Theorem, we can find four sequences of simple functions $s_{n,u^+}, s_{n,u^-}, s_{n,v^+}, s_{n,v^-}$ that converge to the respective functions. Then set $s_n = (s_{n,u^+} - s_{n,u^-}) + i(s_{n,v^+} - s_{n,v^-})$. Then $s_n \to f$ as $n \to \infty$ completing the result.
For the case of an extended real function f we need only consider sequences for f^+ and f^-. $\square$

Corollary 11.12.1. If $f : X \to \mathbb{C}$ with $f = u + iv$ and u, v are measurable functions, then f is measurable.
Sums of (real/extended real/complex) measurable functions are measurable, as are products of measurable functions, and scalar products of measurable functions.
This means that the collection of measurable functions is an Algebra of Functions (this definition is in chapter 7).
These facts follow from the fact that sums of simple functions are simple and for products we have that

$$fg = \frac{(f+g)^2 - (f-g)^2}{4}.$$

11.3 The Integral of Positive Functions

Having discussed Measurable Functions, we now want to be able to integrate them! To define our new integral we use a multi step approach: First we define integration for simple functions, then for positive functions, and next section we will define the general integral! For the rest of this chapter we will work with real/extended real functions. We will discuss integration of Complex Functions in the chapters on Complex Analysis!

Definition 11.13. Let $(X, \mathcal{A}, m)$ be a measure space. Suppose s is a non-negative simple function with Cannonical Representation

$$s = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}.$$

If $E \in \mathcal{A}$ we define the **Simple Lebesgue Integral of s** (w.r.t. m) as

$$\int_E s \, dm = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k \cap E}.$$

Let $f : X \to [0, \infty]$ be measurable, and let S be the set of all simple functions satisfying $s \leq f$ on E. Then we define the **Lebesgue Integral of** f as

$$\int_E f \, dm = \sup_{s \in S} \int_E s \, dm.$$

If f is simple, these definitions agree!

It's not too difficult to see that

$$\int_E f \, dm = \int_X f \mathbb{1}_E \, dm,$$

so that we can regard integrals as being over the entire space when it is convenient.

We now want to discuss some of the properties this integral satisfies.

Theorem 11.14. Suppose $(X, \mathcal{A}, m)$ is a measure space and $f, g : X \to [0, \infty]$ are measurable. Then

(a) $0 \leq \int_X f\, dm \leq \infty$,

(b) If $f \leq g$, then $\int_X f\, dm \leq \int_X g\, dm$,

(c) If $n \leq f \leq N$ and $m(X) < \infty$, then $n \cdot m(X) \leq \int_X f\, dm \leq N \cdot m(X)$,

(d) If $c \in [0, \infty]$, then $c \int_X f\, dm = \int_X cf\, dm$.

(e) If $m(E) = 0$, then $\int_E f\, dm = 0$.

Proof. Part (a) is obvious as the integral of every non-negative simple function will be positive.

Part (b) follows as every simple function $s \leq f$ also satisfies $s \leq g$.

Part (c) follows from applying part (b).

For part (d) we see that for any simple function s on X that $c \int_X s\, dm = c \sum_{k=1}^N a_k m(A_k) = \sum_{k=1}^N c a_k m(A_k) = \int_X cs\, dm$, and the result follows by applying this to every simple function less than f.

Fianlly, to see (e) we see that for any simple function s on E, that

$\int_E s\, dm = \sum_{k=1}^N a_k m(A_K \cap E) = 0$, and since this holds for any simple function on E (e) follows. □

We needed a measure on our space to define the integral of functions, but now that we have an integral it turns out that we can use this integral to define new measures on our space! The next theorem makes this precise in the situation of simple functions and it will be used in the proof of the incredibly important Lebesgue Monotone Convergence Theorem!

Theorem 11.15. Suppose $(X, \mathcal{A}, m)$ is a measure space and s is a non-negative simple function. Then define a function $\gamma : \mathcal{A} \to [0, \infty]$ by

$$\gamma(E) = \int_E s\, dm.$$

Then $(X, \mathcal{A}, \gamma)$ is a measure space. Further if $(X, \mathcal{A}, m)$ is complete, then so is $(X, \mathcal{A}, \gamma)$.

Proof. By Theorem 11.XX (e) above it follows that $\gamma(\emptyset) = 0$. Now suppose $E_1, E_2, \ldots$ is a countable collection of sets in $\mathcal{A}$. Let $E = \bigcup_{n=1}^{\infty} E_n$. We have that

$$\int_E s\, dm = \sum_{k=1}^{M} a_k m(A_k \cap E) = \sum_{k=1}^{M} a_k \sum_{n=1}^{\infty} m(A_k \cap E_n)$$

$$= \sum_{n=1}^{\infty} \sum_{k=1}^{M} a_k m(A_k \cap E_n) = \sum_{n=1}^{\infty} \int_{E_n} s\, dm.$$

We conclude that

$$\gamma\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \gamma(E_n),$$

which completes the proof. □

Our next theorem is one of the most important theorems regarding Lebesgue Integration.

Theorem 11.274 (Lebesgue Monotone Convergence Theorem) Suppose $(X, \mathcal{A}, m)$ is a measure space and $\{f_n\}$ is a sequence of measurable functions from $X \to [0, \infty]$ that is monotone in the sense that $f_n \leq f_{n+1}$ and $f_n \to f$ pointwise on X. Then

$$\lim_{n \to \infty} \int_X f_n\, dm = \int_X f\, dm.$$

Proof. The sequence f_n must converge to some function f by monotonicity, and the limit function is measurable by Theorem 11.XXX. Since each $f_n \leq f$ we have immediately from Theorem 11.XX (b) that $\int_X f_n\, dm \leq \int_X f\, dm$ for all n, so that $\lim_{n\to\infty} \int_X f_n\, dm \leq \int_X f\, dm$.

We now need the other inequality. We will show that every simple function less than f must also be less than $\lim_{n\to\infty}\int_X f_n\,dm$, which will give the result. First choose $a \in (0,1)$ and let s be a non-negative simple function with $s \le f$. We cleverly choose a set to integrate over. Let $E_n = \{x \in X | f_n(x) \ge as(x)\}$. Since $a \le 1$ we know that $as(x) < f(x)$ (unless f is zero at a given point but in this case everything is obvious) so that for each $x \in X$ there is an N so that $x \in E_N$ and so

$$X = \bigcup_{n=1}^{\infty} E_n.$$

Further we have that for a fixed N that

$$\int_{E_N} as\,dm = a\int_{E_N} s\,dm \le \int_{E_N} f_N\,dm \le \int_X f_N\,dm. \tag{11.1}$$

By Theorem 11.XX above and Theorem 10.XX (x) continuity of measure we have that

$$\lim_{n\to\infty}\int_{E_n} as\,dm = \int_E as\,dm.$$

Now taking the limit as $n \to \infty$ on both sides of (X) gives that

$$\int_X as\,dm = a\int_X s\,dm \le \lim_{n\to\infty}\int_X f_n\,dm,$$

and since this holds for all $a \in (0,1)$ we see that

$$\int_X s\,dm \le \lim_{n\to\infty}\int_X f_n\,dm,$$

and the result follows. □

We've just proved one of the major theorems regarding Lebesgue Integration! The theorem clearly does not hold for Riemann Integration, Dirichlet's Function shows this. This major result will have many uses including even stronger convergence theorems, but it really highlights the power of Lebesgue integration! We will now use it to prove some more properties of the Lebesgue Integral starting with Linearity!

Theorem 11.17. Suppose $(X, \mathcal{A}, m)$ is a measure space, $f, g : X \to [0, \infty]$ measurable, and $a, b \in [0, \infty]$. Then

$$\int_X (af + bg)\, dm = a \int_X f\, dm + b \int_X g\, dm.$$

Proof. We already showed in Theorem 11.XX (x) that $a \int_X f\, dm = \int_X af\, dm$, so we need to show additivity. For simple functions addivitity is obvious so we will assume this. By the Simple Approximation Theorem there are sequences of simple functions with $s_{1,n} \to f$ and $s_{2,n} \to g$, which also shows that $s_1 + s_2 \to f + g$. By the Monotone Convergence theorem plus linearity of simple functions we have that

$$\int_X f{+}g\, dm = \lim_{n\to\infty} \int_X s_{1,n}{+}s_{2,n}\, dm = \lim_{n\to\infty} \int_X s_{1,n}\, dm{+} \lim_{n\to\infty} \int_X s_{2,n}\, dm = \int_X f\, dm{+}\int_X g\, dm,$$

completing the proof. □

Corollary 11.17.1. If $f_1, f_2, ...$ is a sequence of non-negative measurable functions then

$$\int_X \sum_{n=1}^{\infty} f_n\, dm = \sum_{n=1}^{\infty} \int_X f_n\, dm.$$

This follows from finite additivity above and the Monotone Convergence Theorem.

We can now extend Theorem 11.XX to the case of all non-negative functions!

Theorem 11.18. Suppose $(X, \mathcal{A}, m)$ is a measure space and $f : X \to [0, \infty]$. Set

$$\gamma(E) = \int_E f\, dm.$$

Then $(X, \mathcal{A}, \gamma)$ is a measure space, and if $(X, \mathcal{A}, m)$ is complete, then so is $(X, \mathcal{A}, \gamma)$.

Proof. It is again obvious that $\gamma(\emptyset) = 0$, so let $E_1, E_2, ...$ be countable collection of sets with union E. Then applying the preceding corollary gives

$$\int_E f\,dm = \int_E \sum_{n=1}^{\infty} f\mathbb{1}_{E_k}\,dm = \sum_{n=1}^{\infty} \int_E f\mathbb{1}_{E_k}\,dm = \sum_{n=1}^{\infty} \int_{E_k} f\,dm.$$

will give the desired result. $\square$

11.4 The Integral of General Functions

Having proved a few nice facts and properties of the Lebesgue Integral, we want to extend this to all real-valued functions! We do this as follows.

Definition 11.19. Suppose $(X, \mathcal{A}, m)$ is a measure space and $f : X \to \mathbb{R}$. We define the **Lebesgue Integral** of f over a set $E \in \mathcal{A}$ as

$$\int_E f\,dm = \int_E f^+\,dm - \int_E f^-\,dm,$$

provided this is not of the form $\infty - \infty$. If f is a non-negative function, this definition agrees with the old one. If both above integrals are finite, we say that f is **Integrable** over E. We denote the set of all integrable functions with respect to measure m with $L^1(m)$. One may wonder why the "1" is present here, and this will be explored in the Chapter on L^p spaces!

Just as before we want to establish some basic properties of the full Lebesgue Integral. We start this time with linearity.

Theorem 11.278 (Linearity of Lebesgue Integral) Suppose $(X, \mathcal{A}, m)$ is a measure space, $f, g \in L^1(m)$, and $a, b \in \mathbb{R}$. Then $af + bg \in L^1(m)$

$$\int_X (af + bg)\,dm = a\int_X f\,dm + b\int_X g\,dm.$$

Proof. The fact that $af+bg \in L^1(X)$ is trivial as $\int_X |(af+bg)|\, dm \leq |a| \int_X |f| + |b| \int_X |g|\, dm$ by the Triangle Inequality and interchanging constants for non-negative functions. Linearity holds for non-negative functions and constants so we must extend to the general case. We from the non-negative case that

$$\int_X (f+g)\, dm = \int_X f^+ + g^+\, dm - \int_X (f^- + g^-)\, dm$$

$$= \int_X f^+\, dm - \int_X f^-\, dm + \int_X g^+\, dm - \int_X g^-\, dm = \int_X dm + \int_X g\, dm.$$

For non-negative a we have that $\int_X af\, dm = a \int_X f\, dm$. In the case $a = -1$ we have that

$$-\int_X f\, dm = -\int_X f^+\, dm + \int_X f^-\, dm = \int_X (-f)^+\, dm - \int_x (-f)^-\, dm = \int_X (-f)\, dm,$$

the general case follows. $\square$

The next theorem gives another list of properties.

Theorem 11.21. Suppose $(X, \mathcal{A}, m)$ is a measure space and $f, g \in L^1(m)$. Then
(a) If $f \leq g$, then $\int_X f\, dm \leq \int_X g\, dm$,
(b) If $n \leq f \leq N$ for all x and $m(X) < \infty$, then $n \cdot m(X) \leq \int_X f\, dm \leq N \cdot m(X)$,
(c) If $m(E) = 0$, then $\int_E f\, dm = 0$,
(d) We have $|\int_X f\, dm| \leq \int_X |f|\, dm$.

Proof. It is obvious a-c hold from the respective positive cases. To see (d) the triangle inequality gives

$$\left|\int_X f\, dm\right| = \left|\int_X f^+\, dm - \int_X f^-\, dm\right| \leq \int_X f^+\, dm + \int_X f^-\, dm = \int_X |f|\, dm.$$

$\square$

11.5 Advantages over Riemann Integral

We've now defined the General Lebesgue Integral for functions from any measure space $X \to \mathbb{R}$ and some of the basic properties, as well as proven Lebesgue's Monotone Convergence Theorem. We now want to highlight more of the advantages this integral has over the Riemann Integral. The first major advantage is that integration is unaffected by null sets!

Theorem 11.280 (Almost Everywhere Integration) Suppose $(X, \mathcal{A}, m)$ is a complete measure space and $f, g :\in L^1(m)$.
(a) If $f \leq g$ a.e., then $\int_X f\, dm \leq \int_X g\, dm$
(b) If $f = g$ a.e., then $\int_X f\, dm = \int_X g\, dm$.

Proof. It suffices to prove this for non-negative functions f, g, the general case will follow. For (a) let A be the set where $f > g$. Then $X = (X - A) \cup A$ and so the above theorem plus Theorem 11.XX gives

$$\int_X f\, dm = \int_{X-A} f\, dm + \int_A f\, dm = \int_{X-A} f\, dm \leq \int_{X-A} g\, dm \leq \int_X g\, dm.$$

To see (b) apply (a) as $f \leq g$ a.e. and $g \leq f$ a.e. so the result follows. □

We can use this to weaken The Monotone Convergence Theorem so that $f_n \to f$ a.e..
The next theorem is another important convergence theorem that is not satisfied by the Riemann Integral.

Theorem 11.281 (Fatou's Lemma) Suppose $(X, \mathcal{A}, m)$ is a measure space with a sequence of non-negative measurable functions f_n. Then

$$\int_X (\liminf f_n)\, dm \leq \liminf \int_X f_n\, dm.$$

Proof. The proof is essentially an application of the Monontone Convergence Theorem! We create a new sequence of functions $g_n(x) = \inf_{k \geq n} f_k(x)$. We have that

$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} g_n$ and clearly that $g_n \le f_n$. Then g_n is a monotone sequence of non-negative measurable functions, so that by the Monotone Convergence Theorem we have

$$\int_X (\limsup f_n)\, dm = \int_X \lim g_n\, dm = \lim \int_X g_n\, dm \le \liminf \int_X f_n\, dm,$$

which completes the proof. $\square$

Theorem 11.282 (Reverse Fatou's Lemma) Suppose $(X, \mathcal{A}, m)$ is a measure space with a sequence of non-negative measurable functions f_n. If there exists a function $g \in L^1(m)$ with $f_n < g$ for all n, then

$$\limsup \int_X f_n\, dm \le \int_X (\limsup f_n)\, dm.$$

Proof. We can essentially apply the regular Fatou's Lemma. We make a new sequence of functions $g_n = g - f_n$. Then

$$\liminf g_n = g - \limsup f_n$$

so that

$$\int_X g\, dm - \int_X (\limsup f_n)\, dm = \int_X (g - \limsup f_n)\, dm$$

$$= \int_X (\liminf g_n)\, dm \le \liminf \int_X g_n\, dm = \int_X g\, dm - \limsup \int_X f_n\, dm),$$

and the result follows.

Note in this case we needed a dominating function g and the integral needed to be finite so that we could cancel it from both sides! $\square$

Combining both the above gives the more general following Theorem

Theorem 11.283 (Fatou-Lebesgue Theorem) Suppose $(X, \mathcal{A}, m)$ is a measure space and $\{f_n\}$ is a sequence of measurable functions dominated by a function $g \in L^1(m)$ in the sense that for all n we have $|f_n| < g$. Then each f_n is integrable, as well as the functions

$\liminf f_n$ and $\limsup f_n$ and

$$\int_X (\liminf f_n)\, dm \leq \liminf \int_X f_n\, dm \leq \limsup \int_X f_n\, dm \leq \int_X (\limsup f_n)\, dm.$$

Proof. At this point this is easy, apply Fatou's Lemma to $f_n + g$ for the first inequality, and the Reverse Fatou Lemma for the last inequality. The middle inequality is obvious. $\square$

As an easy Corollary we get one of the most important and powerful convergence theorems regarding Lebesgue Integration.

Theorem 11.284 (Lebesgue Dominated Convergence Theorem) Suppose $(X, \mathcal{A}, m)$ is a measure space and $\{f_n\}$ is a sequence of measurable functions that converges pointwise to a limit function f a.e. on X, and is dominated by a function $g \in L^1(m)$ in the sense that $|f_n| < g$ a.e., then $f \in L^1(m)$ and

$$\lim_{n\to\infty} \int_X |f - f_n|\, dm = 0,$$

which also implies that

$$\lim_{n\to\infty} \int_X f_n\, dm = \int_X f\, dm.$$

Proof. Let A be the set where some $f_n \geq g$ or where $f_n \not\to f$. Then A has measure zero so let $E = X - A$. We have that on E the sequence of functions $|f_n - f| \to 0$, so that $\limsup |f_n - f| = \liminf |f_n - f| = \lim |f_n - f| = 0$ on E, so that the integrals equal zero. Then the Fatou-Lebesgue Theorem gives convergence of the integrals and

$$\lim_{n\to\infty} \int_E |f_n - f|\, dm = 0.$$

We get the other case as $|\lim_{n\to\infty} \int_E f_n - f\, dm| \leq \lim_{n\to\infty} \int_E |f_n - f|\, dm = 0$. Finally, the integrals over X must all equal the integrals over E as $X - E$ is a null set. $\square$

The convergence theorems we've just proved are one of the Lebesgue Integral's massive advantages. We can now in many cases not require uniform convergence like was required for the Riemann Integral. One issue still lingers however: is the Lebesgue Integral truly a generalization of the Riemann Integral? Are there Riemann Integrable functions that fail to be Lebesgue Integrable with respect to the Lebesgue Measure developed last chapter? It turns out that as we would hope, the Lebesgue Integral truly does generalize the Riemann Integral.

Definition 11.27. Given an interval $[a,b]$ we define a **Step Function** as a function of the form

$$H(x) = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k},$$

where the $A_k's$ are disjoint subintervals partitioning $[a,b]$.
It is clear that each Step function is also a simple function.
We can reinterpret Riemann Integration, as each Riemann Partition gives two corresponding Step function whose Lebesgue Integrals are the Upper and Lower Sums respectively, with respect to the Lebesgue Measure on $\mathbb{R}$, λ.
This gives that $\underline{\int_a^b} f\,dx = \sup \int_{[a,b]} H(x)\,d\lambda$ over all step functions less than f, and similarly $\overline{\int_a^b} f\,dx = \inf \int_{[a,b]} H(x)\,d\lambda$ over all step functions greater than f.

Theorem 11.28. Recall that λ is the Lebesgue Measure on $\mathbb{R}$. Suppose f is a bounded function on $[a,b]$. Then if $f \in \mathcal{R}$ on $[a,b]$ we have that $L^1(\lambda)$ on $[a,b]$ and

$$\int_a^b f\,dx = \int_{[a,b]} f\,d\lambda.$$

As a result of this theorem we will typically use the original Riemann Notation and refer to the Lebesgue Measure on $\mathbb{R}$ as dx, but it should be understood that at this point we will only be using Lebesgue Integration going forward.

Proof. Suppose s is a simple function greater than f, and t is a simple function less than f.

Then $s - t \geq 0$ so that $\int_{[a,b]} s\,d\lambda \geq \int_{[a,b]} t\,d\lambda$. Since t was arbitrary we see that

$$\sup \int_{[a,b]} t\,d\lambda \leq \int_{[a,b]} s\,d\lambda,$$

where the sup is taken over all simple functions less than f. Now since s was arbitrary we see that

$$\sup \int_{[a,b]} t\,d\lambda \leq \inf \int_{[a,b]} s\,d\lambda,$$

where the sup ranges over all simple functions less than f and the inf ranges over all simple functions greater than f.

Since step functions are simple the previous statement plus our comments on step functions above gives the following inequalities.

$$\underline{\int_a^b} f\,dx \leq \sup \int_{[a,b]} t\,d\lambda \leq \inf \int_{[a,b]} s\,d\lambda \leq \overline{\int_a^b} f\,dx,$$

where s, t are simple with $t \leq f \leq s$. But if $f \in \mathcal{R}$, then the above inequalities all become equality's. It's not too difficult to choose a sequence of step functions that converges to the lower integral, which is also equal to the upper integral and therefore converges to f. Since, f is a limit of step functions it is measurable, and its Lebesgue Integral equals the Riemann Integral! $\square$

Chapter 12

Multivariable Differentiation

We have sucessfully covered Part Three of the Book! This last part focuses on a few individual topics: Multivariable Analysis, Further Topology, Complex Analysis, and Functional Analysis. This chapter and the following two will discuss Multivariable Analysis. The work we have done so far has been very useful, but a notably missing topic is the entirety of a standard Calculus Three course and a focus on multivariable functions. This first chapter will review some of the important linear algebra concepts needed to work in $\mathbb{R}^n$ and then we will discuss the derivative for multivariable functions.

12.1 Linear Algebra Overview

We recall that $\mathbb{R}^n$ is a Vector Space with **Dot Product** $(x_1, x_2, ..., x_n) \cdot (y_1, y_2, ..., y_n) = x_1y_1 + ... + x_ny_n$. For a vector $\mathbf{v} \in \mathbb{R}^n$ the **Norm** is $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.

Definition 12.1. Given a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ we define the **Operator Norm** of A by

$$\|A\|_{op} = \sup\{\|Av\| | v \in \mathbb{R}^n \text{ and } \|v\| = 1\}.$$

12.2 The Derivative

We recall that back in Chapter 5 we defined the integral of a real function as

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

This definition does not translate well to the multivariable case at all, but we also made the observation that f was differentiable if

$$f(x+h) \approx f(x) + f'(x)h.$$

This definition said that f was differentiable at x if the function was "linear" about $f(x)$. Having discussed Linear Algebra and linear transformations we can now rework this definition into one that will hold for general multivariable functions!

Definition 12.2. Suppose O is an open subset of $\mathbb{R}^n$ and $f : O \to \mathbb{R}^m$. Then we say that f is **Differentiable** at $\mathbf{x} \in O$ if there exists some linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\lim_{\mathbf{h} \to 0} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A\mathbf{h}}{\|\mathbf{h}\|} = 0.$$

In this case we say that A is the **Derivative** of f at $\mathbf{x}$ and we denote this by $Df(\mathbf{x})$. To distinguish this from the single-variable case or the other variations of the term we will see, this is also sometimes called the **Total Derivative** of f at $\mathbf{x}$.
It is easy to see that the derivative is unique.

This form of the derivative is most useful, but there are some other notions of derivatives for the higher dimensional case we outline now.

Definition 12.3. Suppose O is an open subset of $\mathbb{R}^n$ and $f : O \to \mathbb{R}^m$. If $\mathbf{x} \in O$ and $\mathbf{u} \in \mathbb{R}^n$ is a unit vector ($\|\mathbf{u}\| = 1$). Then we define the **Directional Derivative** of f at $\mathbf{x}$ in the direction of $\mathbf{u}$ as

$$Df(\mathbf{x}, \mathbf{u}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t},$$

if the above limit exists.

12.3 Basic Theorems

We've claimed that the Total Derivative is the best definition for multivariable functions. The next theorem gives a sense of why this is the case.

Theorem 12.4. Suppose O is an open subset of $\mathbb{R}^n$, $f : O \to \mathbb{R}^m$, and $\mathbf{x} \in O$. Suppose f is differentiable at $\mathbf{x}$. We then have that

(a) f is continuous at $\mathbf{x}$,

(b) If $\mathbf{u} \in \mathbb{R}^n$ is a unit vector then $Df(\mathbf{x}, \mathbf{u})$ exists and equals $Df(\mathbf{x})\mathbf{u}$.

Proof. To see (a) we have that since the derivative exists at $\mathbf{x}$ that

$$\lim_{\mathbf{h} \to 0} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - Df(\mathbf{x})\mathbf{h}}{\|\mathbf{h}\|} = 0.$$

Rearranging we get that

$$\lim_{\mathbf{h} \to 0} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \lim_{\mathbf{h} \to 0} \|\mathbf{h}\| Df(\mathbf{x})\mathbf{h} = 0,$$

so that f is continuous at $\mathbf{x}$.

For (b) we see that if we take the limit for $Df(a)$ in the direction of $\mathbf{u}$ we get that

$$\lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x}) - tDf(\mathbf{x})\mathbf{u}}{\|t\mathbf{u}\|} = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{|t|} - \frac{tDf(\mathbf{x})\mathbf{u}}{|t|} = 0.$$

So that

$$\lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{u})-f(\mathbf{x})}{t} = Df(\mathbf{x})\mathbf{u}.$$

This completes the proof. $\square$

The converse of (b) is not true, if all directional derivatives exist we cannot say that f is totally differentiable, nor can we say if they all exist that f is continuous at $\mathbf{x}$ highlighting why our definition of the Total Derivative is truly the correct one.

The Chain Rule also holds for the Total Derivative!

Theorem 12.291 (Multivariable Chain Rule) Suppose O is an open subset of $\mathbb{R}^n$ with $\mathbf{x} \in O$, and $f : O \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$. If there is some open set V with $f(\mathbf{x}) \in V$ and a function $g : V \to \mathbb{R}^w$ that is differentiable at $f(\mathbf{x})$. Then the composition function $(g \circ f)$ is differentiable at $\mathbf{x}$ and

$$Dg \circ f(\mathbf{x}) = Dg(f(\mathbf{x})) \cdot Df(\mathbf{x}).$$

The multiplication above is matrix multiplication

Proof. Let $\mathbf{y} = f(\mathbf{x})$ and let $A = Df(\mathbf{x})$ and $B = Dg(\mathbf{y})$. Then we certainly have that $B \cdot A$ is a Linear Transformation from $\mathbb{R}^m \to \mathbb{R}^w$, so it is a potential candidate. Let $\mathbf{h} \in \mathbb{R}^n$ and let $\mathbf{v} \in \mathbb{R}^m$.

We define $\Delta_f(\mathbf{h}) = f(\mathbf{x}+\mathbf{h}) - y, \quad \Delta_g(\mathbf{v}) = g(\mathbf{y}+\mathbf{v}) - g(\mathbf{y})$,

and $\quad \Delta_{g\circ f}(\mathbf{h}) = g(f(\mathbf{x}+\mathbf{h})) - g(\mathbf{y})$.

Then by differentiability we can write $\Delta_f(\mathbf{h}) = A\mathbf{h} + r_f(\mathbf{h})$ and $\Delta_g(\mathbf{v}) = B\mathbf{v} + r_g(\mathbf{v})$, where r_f and r_g are remainder functions satisfying $\frac{r_f(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ as $\mathbf{h} \to 0$, and likewise for $\frac{r_g(\mathbf{v})}{\|\mathbf{v}\|}$.

We now set $\mathbf{k} = \Delta_f(\mathbf{h})$. By our previous comments we have that

$$\|\mathbf{k}\| = \|A\mathbf{h} + r_f(\mathbf{h})\| = \left\| A\frac{\mathbf{h}}{\|\mathbf{h}\|} + \frac{r_f(\mathbf{h})}{\|\mathbf{h}\|} \right\| |\mathbf{h}\| \leq \left\| \|A\|_{op} + \frac{r_f(\mathbf{h})}{\|\mathbf{h}\|} \right\| \|\mathbf{h}\|,$$

which tends to 0 as $\mathbf{h} \to 0$, so that $\Delta_g(\mathbf{k})$ and $r_g(\mathbf{k})$ are defined for small $\mathbf{h}$.

We also have that $\Delta_{g\circ f}(\mathbf{h}) = \Delta_g(\mathbf{k})$ so that now we have if $\mathbf{k} \neq 0$

$$\frac{\Delta_{g\circ f}(\mathbf{h}) - BA\mathbf{h}}{\|\mathbf{h}\|} = \frac{\Delta_g(\mathbf{k}) - BA\mathbf{h}}{\|\mathbf{h}\|} = \frac{\Delta_g(\mathbf{k}) - B\mathbf{k} - Br_f(\mathbf{h})}{\|\mathbf{h}\|} = \frac{r_g(\mathbf{k}) - Br_f(\mathbf{h})}{\|\mathbf{h}\|}.$$

Then taking magnitudes we see that

$$\left\|\frac{r_g(\mathbf{k}) - Br_f(\mathbf{h})}{\|\mathbf{h}\|}\right\| \leq \frac{\|r_g(\mathbf{k})\|}{\|\mathbf{h}\|} + \left\|B\frac{r_f(\mathbf{h})}{\|\mathbf{h}\|}\right\| = \frac{\|r_g(\mathbf{k})\|}{\|\mathbf{k}\|} \cdot \frac{\|\mathbf{k}\|}{\|\mathbf{h}\|} + \left\|B\frac{r_f(\mathbf{h})}{\|\mathbf{h}\|}\right\|.$$

We know that as $\mathbf{h} \to 0$, that $\frac{r_f(\mathbf{h})}{\|\mathbf{h}\|} \to 0$, so that by continuity $B\left(\frac{r_f(\mathbf{h})}{\|\mathbf{h}\|}\right) \to 0$.
Since $\|\mathbf{k}\| \leq \left\|\|A\|_{op} + \frac{r_f(\mathbf{h})}{\|\mathbf{h}\|}\right\| \|\mathbf{h}\|$, we get that $\frac{\|\mathbf{k}\|}{\|\mathbf{h}\|}$ is finite, and since $\mathbf{k} \to 0$ as $\mathbf{h} \to 0$ we see $\frac{r_g(\mathbf{k})}{\|\mathbf{k}\|} \to 0$ as $\mathbf{h} \to 0$, so that the above tends to 0
If $\mathbf{k} = 0$, then

$$\frac{\Delta_g(\mathbf{k}) - BA\mathbf{h}}{\|\mathbf{h}\|} = \frac{-BA\mathbf{h}}{\|\mathbf{h}\|} = \frac{-B\mathbf{k} - Br_f(\mathbf{h})}{\|\mathbf{h}\|} = \frac{Br_f(\mathbf{h})}{\|\mathbf{h}\|},$$

which tends to zero by the same justification as above.

We see the original limit always tends to zero, so that $Dg \circ f(\mathbf{x}) = BA$. □

One nice corollary of the multivariable chain rule is a higher dimensional version of the mean value theorem, which we now prove!

Theorem 12.6. Suppose $O \subseteq \mathbb{R}^n$ is a convex open set and $f : O \to \mathbb{R}^m$ is differentiable on O. If there exists an M so that

$$\|Df(\mathbf{x})\|_{op} \leq M,$$

for all $\mathbf{x} \in O$, then for all $\mathbf{a}, \mathbf{b} \in O$ we have

$$\|f(\mathbf{b}) - f(\mathbf{a})\| \leq M\|b - a\|.$$

Proof. We prove the special case where $n = 1$ and $O = (0, 1)$ first, while adding the assumptions f be continuous on $[0,1]$. We will use this special case along with convexity to give the result. Then we define a new function $g:[0,1]\to\mathbb{R}$, by $g(x)=((f(1)-f(0))\cdot(f(x))$, where the product here is the dot product. Then the mean value theorem gives for some $c\in(0,1)$ that

$$g(1)-g(0)=((f(1)-f(0))\cdot(f(1)-f(0))=\|f(1)-f(0)\|^2=g'(c)=(f(1)-f(0))\cdot Df(c),$$

so that by the Schwarz Inequliaty we get

$$\|f(1)-f(0)\|\leq M.$$

Now we return to the general case and we similarly define a new function $g:[0,1]\to\mathbb{R}^m$ by $g(t)=f((1-t)\mathbf{a}+t\mathbf{b})$, essentially g gives the function values on the straight line connecting $\mathbf{a}$ to $\mathbf{b}$, which is contained in O by convexity. Then by the Chain Rule we have g is differentiable with $g'(t)=Df((1-t)\mathbf{a}+t\mathbf{b})(\mathbf{b}-\mathbf{a})$, and $\|g'(t)\|\leq M\|\mathbf{b}-\mathbf{a}\|$. Our special case above now gives the result. □

12.4 Partial Derivatives

We now turn our attention to a familiar topic from Calculus Three, and one that will ultimately give us the tools needed for computing the Total Derivative. So far we've treated the derivative as a linear transformation, but the fact that every linear transformation between $\mathbb{R}^n$ and $\mathbb{R}^m$ can be represented by a Matrix tells will be the more important perspective for actually computing and using the derivative. We now turn towards Partial Derivatives. For the rest of the chapter we will use $\{\mathbf{e_1},...,\mathbf{e_n}\}$ as the standard basis for the domain (typically $\mathbb{R}^n$) and $\{\mathbf{u_1},...,\mathbf{u_m}\}$ as the standard basis for the codomain (typically $\mathbb{R}^m$).

Definition 12.7. Let $O \subset \mathbb{R}^n$ be open and let $f : O \to \mathbb{R}$. Then the **j-th Partial Derivative** of f at $\mathbf{x} \in O$ is defined as the directional derivative of f in the direction of $\mathbf{e_j}$, $Df(\mathbf{x}, \mathbf{e_j})$, provided this exists. We will use the new notation $D_j f(\mathbf{x})$. If we look at the limit formula we see that

$$D_j f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e_j}) - f(\mathbf{x})}{h} = \lim_{h \to 0} \frac{f(x_1, ..., x_j + h, ..., x_n) - f(x_1, ..., x_j, ..., x_n)}{h},$$

so that to compute the Partial Derivative, you simply take the derivative of f with respect to the j-th variable, treating the others as constants! The notation here is different one likely saw the notation

$$\frac{\partial f}{\partial x_j}$$

in a typical Calculus coures, but this new notation is much more in like with our new notation and it will prove to be more convenient.

To extend this to more general functions $f : O \to \mathbb{R}^m$, we recall that in the single variable case we could write functions $g : \mathbb{R} \to \mathbb{R}^n$ as $g(x) = (g_1(x), ..., g_n(x))$, and we can do the same for multivariable case, however we will use our newer notation.

Given a function $f : O \to \mathbb{R}^m$ we define the **i-th Component** of f as $f_i(\mathbf{x}) = \mathbf{u_i}^T f(\mathbf{x})$ for all $\mathbf{x} \in O$. It is clear then that we can represent f as

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}.$$

We can then define the **ij-th Partial Derivative** of f as the j-th Partial Derivative of the i-th component of f.

The next big theorem tells us how we can represent the Total Derivative.

Theorem 12.294 (Jacobian Matrix) If $O \subseteq \mathbb{R}^n$ is open and $f : O \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$. Then

(a) Each component function f_i is differentiable at $\mathbf{x}$ and $Df_i(\mathbf{x}) = \mathbf{u_i}^T Df(\mathbf{x})$,

(b) Each partial derivative of f exists, and

(c) We have that

$$Df(\mathbf{x}) = \begin{bmatrix} D_1 f_1(\mathbf{x}) & D_2 f_1(\mathbf{x}) & \dots & D_n f_1(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathbf{x}) & D_2 f_m(\mathbf{x}) & \dots & D_n f_m(\mathbf{x}) \end{bmatrix}.$$

The above matrix is called the **Jacobian Matrix** of f at $\mathbf{x}$ and the determinant $\det Df(\mathbf{x})$ is called the **Jacobian** of f at $\mathbf{x}$.

Proof. Given $\mathbf{h}$ define a function F as

$$F(\mathbf{h}) = \frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - Df(\mathbf{x})\mathbf{h}}{\|\mathbf{h}\|}$$

and set $F_i(\mathbf{h}) = \mathbf{u_i}^T F(\mathbf{h})$ so that

$$F_i(\mathbf{h}) = \frac{f_i(\mathbf{x}+\mathbf{h}) - f_i(\mathbf{x}) - \mathbf{u_i}^T Df(\mathbf{x})\mathbf{h}}{\|\mathbf{h}\|}.$$

Then we see that since $F(\mathbf{h}) \to 0$, that $F_i(\mathbf{h}) \to 0$ as $\mathbf{h} \to 0$, so that f_i is differentiable at $\mathbf{x}$, and in fact we see that $Df_i(\mathbf{x}) = \mathbf{u_i}^T Df(\mathbf{x})$, completing (a).

Part (b) follows from (a) as since each f_i is differentiable we know all directional derivatives for each f_i exist, in paritcular all partial derivatives of f_i. Since this works for each f_i we see all partial derivatives of f exist.

For part (c) we know that $Df_i(\mathbf{x})$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}$, so that it is a

$1 \times n$ matrix. Since we have that $D_j f_i(\mathbf{x}) = Df_i(\mathbf{x})\mathbf{e_j}$, we see that

$$Df_i(\mathbf{x}) = \begin{bmatrix} D_1 f_i(\mathbf{x}) & \dots & D_n f_i(\mathbf{x}) \end{bmatrix}.$$

This together with (a) gives $Df(\mathbf{x})_{ij} = \mathbf{u_i}^T Df(\mathbf{x})\mathbf{e_j} = D_j f_i(\mathbf{x})$, giving (c). $\square$

Corollary 12.8.1. We see that if each f_i is differentiable at $\mathbf{x}$, then we see each $F_i(\mathbf{h})$ defined above tends to zero, so that $F(\mathbf{h}) \to 0$, and so the function f is also differentiable at $\mathbf{x}$.

Definition 12.9. In the special case $f : O \to \mathbb{R}$, where $O \subseteq \mathbb{R}^n$, is differentiable, we define the **Gradient** of f as the Jacobian Marix which in this case looks like

$$\nabla f = \begin{bmatrix} D_1 f(x) & \dots & D_n f(x) \end{bmatrix} = D_1 f(x)\mathbf{e_1} + ... + D_n f(x)\mathbf{e_n}.$$

For functions like f above and a unit vector $\mathbf{u} \in \mathbb{R}^n$ we also see that $Df(\mathbf{x}, \mathbf{u}) = \nabla f \mathbf{u}$.

While the existence of partial derivatives does not imply differentiability, the next theorem shows that if the partial derivatives are continuous we do guarantee differentiability!

Theorem 12.10. Suppose $O \subseteq \mathbb{R}^n$ is open and $f : O \to \mathbb{R}^m$. Then all partial derivatives exist and are continuous on O if and only if f is differentiable on O, and further the transformation $Df : O \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.

Proof. First assume all partial derivatives exist and are continuous. By the above corollary, it sufficies to prove this for the case where $m = 1$, or $f : O \to \mathbb{R}$. Let $\mathbf{x} \in O$, and given $\epsilon > 0$ choose $r > 0$ so that for all $y \in B_r(\mathbf{x})$ we have

$$|D_j f(\mathbf{x}) - D_j(\mathbf{y})| < \epsilon,$$

for all $j = 1, ..., n$. Now let $\mathbf{h} \in \mathbb{R}^n$ with $\|\mathbf{h}\| < r$, with $\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$.

We then define $\mathbf{v}_k = \begin{bmatrix} h_1 \\ \vdots \\ h_k \end{bmatrix}$, with the special case $\mathbf{v}_0 = \mathbf{0}$. At this point we now have that

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \sum_{k=1}^{n} f(\mathbf{x}+\mathbf{v}_k) - f(\mathbf{x}+\mathbf{v}_{k-1}).$$

We know that $\mathbf{x}+\mathbf{v}_k \in B_r(\mathbf{x})$ for all $k = 1,..,n$ and since $B_r(\mathbf{x})$ is convex, the line segment joining v_{k-1} to v_k is also contained in $B_r(\mathbf{x})$, and $\mathbf{v}_k = \mathbf{v}_{k-1} + h_k\mathbf{e}_k$. Now then if we make a new function $g : [0,1] \to \mathbb{R}$ $g(t) = f(\mathbf{x}+\mathbf{v}_{k-1} + th_k\mathbf{e}_k)$, then by the chain rule g is differentiable with derivative $Df(\mathbf{x}+\mathbf{v}_{k-1}+t\cdot h_k\cdot \mathbf{e}_k)\cdot h_k\mathbf{e}_k = D_kf(\mathbf{x}+\mathbf{v}_{k-1}+t\cdot h_k\cdot\mathbf{e}_k)h_k$. Finally, g satisfies the criterion of the Mean Value Theorem, so that

$$f(\mathbf{x}+\mathbf{v}_k) - f(\mathbf{x}+\mathbf{v}_{k-1}) = D_kf(\mathbf{x}+\mathbf{v}_{k-1}+c_kh_k\mathbf{e}_k),$$

for some $c_k \in (0,1)$. We also have that $|D_kf(\mathbf{x}+\mathbf{v}_{k-1}+c_kh_k\mathbf{e}_k) - D_kf(\mathbf{x})| < \epsilon$, by above. Substituting this into our original equation gives

$$\|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \begin{bmatrix} D_1f(\mathbf{x}) & \dots & D_nf(\mathbf{x}) \end{bmatrix}\mathbf{h}\| = \left\| f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \sum_{k=1}^{n} D_kf(\mathbf{x})h_k \right\|$$

$$\le \epsilon \sum_{k=1}^{n} |h_k| = n\epsilon\|\mathbf{h}\|.$$

Since ϵ was arbitrary, we see that f is differentiable at $\mathbf{x}$.

The other direction is easier for if Df is continuous and $\epsilon > 0$, then the ij-th Partial Derivative exists and

$$|D_jf_i(\mathbf{x}) - D_jf_i(\mathbf{y})| = |\mathbf{u}_i \cdot ((D_jf(\mathbf{x}) - D_jf(\mathbf{y}))\mathbf{e}_j)| \le \|Df(\mathbf{x}) - Df(\mathbf{y})\|_{op},$$

so that by continuity of Df we see that each D_jf_i must be continuous. $\square$

We can use this theorem to motivate a new definition.

Definition 12.11. For function of a single variable we have already defined what we mean by $\mathcal{C}^k(X)$. Based on the above theorem we can extend this to multivariable functions. Suppose $O \subseteq \mathbb{R}^n$ is open, and let $f : O \to \mathbb{R}^m$. Then we say that $f \in \mathcal{C}^1(O)$ if all of its partial derivatives exist, and are continuous. By the above theorem this is equivalent to Df being continuous on O, but our definition generalizes easier to higher dimensions. A partial derivative of f say $D_j f_i$ is a function from $O \to \mathbb{R}^m$, so we can ask about its partial derivatives. A partial derivative of $D_j f_i$ say $D_k(D_j f_i)$ is called a **Second-Order Partial Derivative** of f. If all second-order partial derivatives exist and are continuous we say that $f \in \mathcal{C}^2(O)$. Similarly, we can define third, fourth, etc... partial derivarives. If all k-order partial derivatives of f exist and are continuous we say that $f \in \mathcal{C}^k(O)$. Just as in the single variable case if $f \in \mathcal{C}^k(O)$ for all $k = 1, 2, ...$ we say that $f \in \mathcal{C}^\infty(O)$ or are smooth functions.

A nice fact is that composing $\mathcal{C}^k$ functions results in another $\mathcal{C}^k$ function.

Theorem 12.12. Suppose $O \subseteq \mathbb{R}^n$ is open, $V \subseteq \mathbb{R}^m$ is open, $f : O \to \mathbb{R}^m$, $f(O) \subseteq V$, $g : B \to \mathbb{R}^w$. If $f, g \in \mathcal{C}^k$ on their respective sets, then $g \circ f \in \mathcal{C}^k(O)$.

Proof. For the case of $\mathcal{C}^1$ we see $Dg \circ f(\mathbf{x}) = Dg(f(\mathbf{x})) \cdot Df(\mathbf{x})$. Since f is continuous (differentiable) and Dg is a continuous transformation then $Dg(f(\mathbf{x}))$ is continuous, so that its product with Df (another continuous transformation) is continuous, so that $g \circ f \in \mathcal{C}^1(O)$ Now assume it holds for the case of $k-1$ and that $f, g \in \mathcal{C}^k$. Then every element of the matrices Dg, Df, and $Dg \circ f$ are all $\mathcal{C}^{k-1}$ functions. By the $k-1$ case then we see that every element of the matrix $Dg \circ f$ is in $\mathcal{C}^{k-1}$, but then the composition function $g \circ f \in \mathcal{C}^k$ completing the proof.

If f, g are smooth then the composition is smooth as the above holds for all $k = 1, 2, ...$ $\square$

Another perhaps surprising fact is the following theorem.

Theorem 12.13. Suppose A is an $n \times n$ matrix of functions $f_{ij} : O \to \mathbb{R}$ where O is an open subset of $\mathbb{R}^n$, and A is invertible for all $\mathbf{x} \in O$. If each function in A is of class $\mathcal{C}^k$, then each function in A^{-1} is of class $\mathcal{C}^k$. This is because each element of A^{-1} by Theorem XXX can be written as a rational function of entries the A. Since rational functions are $\mathcal{C}^\infty$ the result follows from the above theorem.

To end this section we now prove a familiar result from Calculus Three.

Theorem 12.300 (Clairaut's Theorem) Let $f : O \to \mathbb{R}$ where $O \subseteq \mathbb{R}^n$ is open. Then if $f \in \mathcal{C}^2(O)$, then for all $j, k = 1, ..., n$ and for all $\mathbf{x} \in O$ we have

$$D_j(D_k f(\mathbf{x})) = D_k(D_j f(\mathbf{x})).$$

Essentially, we can swap the order of partial differentiation.

Proof. We can assume that $m = 2$ and show this for $i = 1, j = 2$. Since in the general case all other variables are regarded as constants, we are justified in doing this. Now for every $\mathbf{x} \in O$ we can write it as (x, y). Fix $(x, y) \in O$, and consider a small rectangle $[x, x+h] \times [y, y+k] \subseteq O$. Now consider the function

$$\Delta(t, s) = f(x+t, y+s) + f(x, y) - f(x+t, y) - f(x, y+s).$$

For a fixed s, $\Delta(t)$ is a function from $[0, h] \to \mathbb{R}$ so an application of the mean value theorem gives for some $c \in (x, x+h)$ that

$$\Delta(h, s) - \Delta(0, s) = \Delta(h, s) = D_1 f(x+c, y+s)h - D_1 f(x+c, y)h.$$

Since this is now a function of s another application of the MVT gives for some $d \in (y, y+k)$ that

$$\Delta(h, k) - \Delta(h, 0) = \Delta(h, k) = D_2(D_1 f(x+c, y+d))hk.$$

A similar trick says we can find points $p \in (x, x+h)$ and $q \in (y, y+k)$ so that

$$\Delta(h,k) = D_1(D_2 f(x+p, y+q))hk.$$

By continuity of $D_2 D_1 f$ and similarly $D_1 D_2 f$ given $\epsilon > 0$ if we choose δ small enough and make h, k small enough to satisfy $[x, x+h] \times [y, y+k] \subseteq B_\delta((x,y))$, then we guarantee that $\left|\frac{\Delta(h,k)}{hk} - D_1(D_2 f(x,y))\right| < \epsilon$ and $\left|\frac{\Delta(h,k)}{hk} - D_2(D_1 f(x,y))\right| < \epsilon$ giving the result. □

12.5 Inverse Function Theorem

We arrive at the first of two major theorems that will close out this chapter. The Inverse Function theorem roughly says that if $Df(\mathbf{x})$ is invertible and $f \in \mathcal{C}^1$, then f is invertible and its inverse function is differentiable at $f(\mathbf{x})$. This all happens only locally about $\mathbf{x}$, however, as globally it is certainly not true that being differentiable at a point guarantees that the function be bijective (and therefore invertible) on its entire domain. The theorem guarantees a small enough neighborhood where all these niceties can happen! The proof of the Inverse Function theorem will use the Contraction Fixed Point Theorem proven back in Chapter 4, so the reader should perhaps refresh on this theorem. With that said, we can now jump right into the theorem!

Theorem 12.301 (The Inverse Function Theorem) Suppose $O \subseteq \mathbb{R}^n$ is open, $f : O \to \mathbb{R}^n$, and $\mathbf{x}_0 \in O$. If $f \in \mathcal{C}^1(O)$ and $Df(\mathbf{x}_0)$ is invertible, then

(a) There exist open sets U containing $\mathbf{x}_0$ and V containing $f(\mathbf{x}_0)$ so that the restriction $f : U \to V$ is bijective.

(b) Since f is bijective on U it has an inverse function $f^{-1} : V \to U$ and $f^{-1} \in \mathcal{C}^1(V)$.

Proof. We first prove part (a) using the Contraction Fixed Point Theorem. For simplicity let $A = Df(\mathbf{x}_0)$, we have by assumption that A^{-1} exists. Now for each $\mathbf{y} \in \mathbb{R}^n$ consider the function $g_{\mathbf{y}} : O \to \mathbb{R}^n$ by

$$g_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - f(\mathbf{x})).$$

Then g is differentiable on O with

$$Dg_{\mathbf{y}}(\mathbf{x}) = I_n - A^{-1}Df(\mathbf{x}_0) = A^{-1}(A - Df(\mathbf{x}_0)),$$

so that we see

$$\|Dg_{\mathbf{y}}(\mathbf{x})\|_{op} \leq \|A^{-1}\|_{op}\|A - Df(\mathbf{x}_0)\|_{op}.$$

Since $f \in \mathcal{C}^1(O)$ we can find a neighborhood U of $\mathbf{x}_0$ so that for all $\mathbf{x} \in U$ we have $\|A - Df(\mathbf{x})\|_{op} \leq \frac{1}{2\|A^{-1}\|_{op}}$. It follows then on U we have for all $\mathbf{x}$ that

$$\|Dg_{\mathbf{y}}(\mathbf{x})\|_{op} \leq \frac{1}{2},$$

so that by Theorem 12.XX the multivariable case of the MVT we have that for all $\mathbf{a}, \mathbf{b} \in U$ that

$$\|g_{\mathbf{y}}(a) - g_{\mathbf{y}}(b)\| \leq \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|. \tag{12.1}$$

It follows that $g_{\mathbf{y}}$ is a contraction mapping on U, so that $g_{\mathbf{y}}$ has at most one fixed point (we do not yet know if the range of g is complete so we cannot guarantee a fixed point). But that means there is at most one $\mathbf{x} \in U$ so that $g_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$ which happens when $\mathbf{y} = f(\mathbf{x})$. This holds for each $\mathbf{y}$, so for each $\mathbf{y} \in \mathbb{R}^n$ there is AT MOST one $\mathbf{x} \in U$ so that $f(\mathbf{x}) = \mathbf{y}$, or f is injective on U, we cannot guarantee a solution yet as we don't know if the range of f is complete yet!

Now let $V = f(U)$, we want to show that V is open which will complete (a) so let $\mathbf{y}_0 \in V$. Then there is some $\mathbf{x} \in U$ so that $f(\mathbf{x}) = y_0$. Since U is open, there is some $r > 0$ so that $B_r(\mathbf{x}) \subseteq U$ giving $\overline{B}_{\frac{r}{2}}(\mathbf{x}) \subseteq U$, let $K = \overline{B}_{\frac{r}{2}}(\mathbf{x})$.

Now let $\mathbf{y} \in \mathbb{R}^n$ be chosen so that $\|\mathbf{y} - \mathbf{y}_0\| < \frac{r}{4\|A^{-1}\|_{op}}$. We then have that for this $\mathbf{y}$ and for any $\mathbf{v} \in K$ that

$$\|g_{\mathbf{y}}(\mathbf{v}) - \mathbf{x}\| \leq \|g_{\mathbf{y}}(\mathbf{v}) - g_{\mathbf{y}}(\mathbf{x})\| + \|g_{\mathbf{y}}(\mathbf{x}) - \mathbf{x}\| \leq \frac{1}{2}\|\mathbf{v} - \mathbf{x}\| + \|A^{-1}(\mathbf{y} - \mathbf{y}_0)\|_{op} \leq \frac{r}{2},$$

by (12.1) and the definition of $g_{\mathbf{y}}$. This shows that $g_{\mathbf{y}}(\mathbf{v}) \in K$ or $g_{\mathbf{y}}$ is a contraction mapping of K to K! Since K is compact (and therefore complete) it follows now by the Contraction Fixed Point Theorem, that there is a unique fixed point $\mathbf{v}'$ where $g_{\mathbf{y}}(\mathbf{v}') = \mathbf{v}'$, which further implies that $f(\mathbf{v}') = \mathbf{y}$ or $\mathbf{y} \in V$!

Since this holds for any $\mathbf{y}$ within a ball of $\mathbf{y}_0$, we see that V is open completing part (a)!

We now handle (b). Let $f^{-1} : V \to U$ be the inverse of f. Let $\mathbf{y} \in V$ and let $\mathbf{x}$ be so that $f(\mathbf{x}) = \mathbf{y}$. It follows if $\mathbf{k}$ is small enough that $\mathbf{y} + \mathbf{k} \in V$ so that there is some $\mathbf{h}$ so that $\mathbf{x} + \mathbf{h} \in U$ with $f(\mathbf{x} + \mathbf{h}) = \mathbf{y} + \mathbf{k}$. Then for this $\mathbf{y}$ we can again use $g_{\mathbf{y}}$ to see that

$$g_{\mathbf{y}}(\mathbf{x} + \mathbf{h}) - g_{\mathbf{y}}(\mathbf{x}) = \mathbf{h} - A^{-1}(\mathbf{y} - (\mathbf{y} + \mathbf{k})) - A^{-1}(\mathbf{y} - \mathbf{y}) = \mathbf{h} - A^{-1}k,$$

so that (12.1) gives $\|\mathbf{h} + A^{-1}\mathbf{k}\| \leq \frac{1}{2}\|\mathbf{h}\|$, and by the reverse triangle inequality we have $\|\mathbf{h}\| \leq 2\|A^{-1}\|_{op}\|\mathbf{k}\|$, so that $\mathbf{h} \to 0$ as $\mathbf{k} \to 0$. It follows by Theorem 12.XX that since $\|Df(\mathbf{x}) - A\|_{op} \leq \frac{1}{\|A^{-1}\|_{op}}$ that $(Df(\mathbf{x}))^{-1}$ exists, label it B. Now finally we have that

$$\frac{\|f^{-1}(\mathbf{y} + \mathbf{k}) - f^{-1}(\mathbf{y}) - B\mathbf{k}\|}{\|\mathbf{k}\|} = \frac{\|\mathbf{h} - B(f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}))\|}{\|\mathbf{k}\|}$$

$$= \frac{\|B(Df(\mathbf{x})\mathbf{h} - f(\mathbf{x} + \mathbf{h}) + f(\mathbf{x}))\|}{\|\mathbf{k}\|} \leq 2\|B\|_{op}\|A^{-1}\|_{op}\frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - Df(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|}.$$

This proves that f^{-1} is differentiable at $\mathbf{y}$ and therefore on all of V.

Finally, we see that since f^{-1} is continuous, inversion of linear transformations is continuous, $Df(\mathbf{x})$ is continuous, and $Df^{-1}(\mathbf{y}) = (Df(f^{-1}(\mathbf{y}))^{-1}$ a composition of continuous functions, then Df^{-1} is continuous proving that $f^{-1} \in \mathcal{C}^1(V)$. $\square$

We can in fact extend the result using Theorem's 12.XX and 12.XX to say that if $f \in \mathcal{C}^k$

then $f^{-1} \in \mathcal{C}^k$ on the set V whose existence is guaranteed in part (a). Before moving on to the Implicit Function Theorem, we briefly discuss diffeomorphisms and restate the Inverse Function Theorem.

Definition 12.16. Given an open subset $O \subseteq \mathbb{R}^n$ and a function f, we say that f is a **Diffeomorphism** if $f \in \mathcal{C}^1$, is bijective on O, and its inverse function $f^{-1} \in \mathcal{C}^1$. If we have that f and f^{-1} are infact $\mathcal{C}^k$ then they are $\mathcal{C}^k$ Diffeomorphisms. This definition is similar to that of homeomorphisms discussed back in Chapter 3, and every diffeomorphism is also a homomorphism.

We can restate the Inverse Function Theorem as follows.

Theorem 12.303 (Inverse Function Theorem Restated) If $O \subseteq \mathbb{R}^n$ is open, $f : O \to \mathbb{R}^n$ is of class $\mathcal{C}^k$, and $Df(\mathbf{x})$ is invertible at $\mathbf{x}$, then there exists open sets U containing $\mathbf{x}$ and V containing $f(\mathbf{x})$ and f is a $\mathcal{C}^k$ diffeomorphism from U to V.

12.6 Implicit Function Theorem

We now turn to the second major theorem of this chapter The Implicit Function Theorem. So far we've worked exclusively with functions and their derivatives, but we are not always presented with a function. For example the equation for a circle $x^2 + y^2 = r^2$, is clearly not a function, but we may still want to get information about this equation. For example in Calculus we can "Implicitly Differentiate" this equation to get a derivative $2x + 2y\frac{dy}{dx} = 0$ giving $\frac{dy}{dx} = \frac{-x}{y}$. Even though, y is not a function of x we have implicitly expressed it in this form and used it to get a useful formula for the derivative of this so called "implicit function".

The Implicit Function Theorem makes this rigorous. Essentially, if we have a function f with $n + k$ variables, an equation $f(x_1, ..., x_{n+k}) = 0$, and some solution to this equation then under a few additional conditions, if we are given k of these variables we can implicitly find a function of these k variables for the other n variables. Further, this implicit function will be guaranteed to be differentiable and we can calculate its derivative! Of course, as in the case of the inverse function theorem, this is only locally true and does not extend to the entire domain of our function f. We now introduce some notation to help us with the theorem.

It will be convenient to treat the set of n variables and the other k variables as separate entities. For $\mathbf{x} \in \mathbb{R}^n$ we can write this of course as $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, and for $\mathbf{y} \in \mathbb{R}^k$ we can of course write this as $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$. Then we can write the vector $\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_k \end{bmatrix} \in \mathbb{R}^{n+k}$ instead as $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+k}$.

Similarly we can write a function $f : \mathbb{R}^{n+k} \to \mathbb{R}^k$ as $f(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}, \mathbf{y}$ are as above.

If we now consider the derivative of f, Df it will be of the form

$$Df(\mathbf{x},\mathbf{y}) = \left[\begin{array}{ccc|ccc} D_{x_1}f_1(\mathbf{x},\mathbf{y}) & \dots & D_{x_n}f_1(\mathbf{x},\mathbf{y}) & D_{y_1}f_1(\mathbf{x},\mathbf{y}) & \dots & D_{y_k}f_1(\mathbf{x},\mathbf{y}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ D_{x_1}f_k(\mathbf{x},\mathbf{y}) & \dots & D_{x_n}f_k(\mathbf{x},\mathbf{y}) & D_{y_1}f_k(\mathbf{x},\mathbf{y}) & \dots & D_{y_k}f_k(\mathbf{x},\mathbf{y}) \end{array}\right]$$

To refer to the left side of this matrix we will use the notation $D_{\mathbf{x}}f$, which is a $k \times n$ matrix. Similarly for the right side we use $D_{\mathbf{y}}f$, which will be a $k \times k$ matrix. In fact the condition we were missing above was that this matrix $D_{\mathbf{y}}f$ be invertible.

We are now ready for the theorem.

Theorem 12.304 (The Implicit Function Theorem) Suppose $O \subseteq \mathbb{R}^{n+k}$ is open and $f : O \to \mathbb{R}^k$, with $f \in \mathcal{C}^M(O)$. We will use the notation $f(\mathbf{x}, \mathbf{y})$ as explained above. Suppose there is a point $(\mathbf{a}, \mathbf{b})$ so that $f(\mathbf{a}, \mathbf{b}) = 0$ and $D_{\mathbf{y}}f(\mathbf{a}, \mathbf{b})$ is invertible. Then there exists an open set $W \subseteq \mathbb{R}^n$ containing $\mathbf{a}$ and a function $g : W \to \mathbb{R}^k$ satisfying

(a) $g(\mathbf{a}) = \mathbf{b}$,

(b) $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in W$,

(c) $g \in \mathcal{C}^M(V)$, and

(d) $Dg(\mathbf{a}, g(\mathbf{a})) = -(D_{\mathbf{y}} f(\mathbf{a}, g(\mathbf{a})))^{-1} \cdot D_{\mathbf{x}} f(\mathbf{a}, g(\mathbf{a}))$.

Put simply there is an implicit function of $\mathbf{x}$ about $(\mathbf{a}, \mathbf{b})$ and this function is of class $\mathcal{C}^M$. Some of the implications of this theorem may look very similar to that of the Inverse Function Theorem. This is no coincidence, we will be using the Inverse Function Theorem as a major part of our proof!

Proof. We define a new function $F : \mathbb{R}^{n+k} \to \mathbb{R}^{n+k}$ by $F(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, f(\mathbf{x}, \mathbf{y})))$. Since $f \in \mathcal{C}^M$ and the identity function is smooth, we have that $F \in \mathcal{C}^M$, and we see the derivative is

$$DF(\mathbf{x}, \mathbf{y}) = \left[\begin{array}{ccc|ccc} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline D_{x_1} f_1(\mathbf{x}, \mathbf{y}) & \cdots & D_{x_n} f_1(\mathbf{x}, \mathbf{y}) & D_{y_1} f_1(\mathbf{x}, \mathbf{y}) & \cdots & D_{y_k} f_1(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ D_{x_1} f_k(\mathbf{x}, \mathbf{y}) & \cdots & D_{x_n} f_k(\mathbf{x}, \mathbf{y}) & D_{y_1} f_k(\mathbf{x}, \mathbf{y}) & \cdots & D_{y_k} f_k(\mathbf{x}, \mathbf{y}) \end{array}\right]$$

$$= \left[\begin{array}{c|c} I_n & 0 \\ \hline D_{\mathbf{x}} f & D_{\mathbf{y}} f \end{array}\right]$$

We can calculate the determinant of $DF(\mathbf{a}, \mathbf{b})$ and we see that
$\det DF(\mathbf{a}, \mathbf{b}) = \det D_{\mathbf{y}} f(\mathbf{a}, \mathbf{b}) \neq 0$, so that $DF(\mathbf{x}, \mathbf{y})$ is invertible!

Therefore, we can use the Inverse Function Theorem with F and we are guaranteed a neighborhood $U \subseteq \mathbb{R}^{n+k}$ of $(\mathbf{a}, \mathbf{b})$ and $V \subseteq \mathbb{R}^{n+k}$ of $(\mathbf{a}, \mathbf{0})$ so that F restricted to U has a $\mathcal{C}^M$ inverse function, in particular F is injective on U.

Let us define $W \subseteq \mathbb{R}^n$ to be the set of $\mathbf{x}$ so that $(\mathbf{x}, \mathbf{0}) \in V$. We know $\mathbf{a} \in W$ so that W is nonempty. If $\mathbf{x}' \in W$, then we have that $(\mathbf{x}', \mathbf{0}) \in V$ and since V is open there is some $r > 0$, so that $B_r(\mathbf{x}') \subseteq V$. If $\|\mathbf{x}' - \mathbf{x}\| < r$, then $\|(\mathbf{x}', \mathbf{0}) - (\mathbf{x}, \mathbf{0})\| < r$, so that $(\mathbf{x}, \mathbf{0}) \in B_r(\mathbf{x}')$, showing that $x \in W$, so that W is open.

For every $\mathbf{x} \in W$, we have $(\mathbf{x}, \mathbf{0}) \in V$, so that $F(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{0})$ for some $\mathbf{y}$, and this $\mathbf{y}$ is unique by the injectivity of F.

If we now define $g : W \to \mathbb{R}^k$ by $g(\mathbf{x})$ is the unique $\mathbf{y}$ satisfying the above, then we have that $g(\mathbf{a}) = \mathbf{b}$ by uniqueness of $\mathbf{y}$ above and $F(\mathbf{x}, g(\mathbf{x})) = (\mathbf{x}, 0) = (\mathbf{x}, f(\mathbf{x}, g(\mathbf{x})))$, so that $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in W$, giving (a) and (b).

For part (c) we consider $F^{-1} : V \to U$, guaranteed above. Then we have that

$F^{-1}(\mathbf{x}, 0) = (\mathbf{x}, g(\mathbf{x}))$, and again since $F^{-1} \in \mathcal{C}^M$ and the identity mapping is smooth, we get that $g \in \mathcal{C}^M(W)$, giving (c).

Finally, we compute Dg for $\mathbf{t} \in W$ with the composition $H(\mathbf{t}) = f(\mathbf{t}, g(\mathbf{t})) = 0$, giving

$DH(\mathbf{t}) = Df(\mathbf{t}, g(\mathbf{t})) \cdot \left[\begin{smallmatrix} I_n \\ Dg(\mathbf{t}) \end{smallmatrix}\right] = 0$ or

$\left[\begin{smallmatrix} D_{\mathbf{x}}(\mathbf{t}, g(\mathbf{t})) & D_{\mathbf{y}}(\mathbf{t}, g(\mathbf{t})) \end{smallmatrix}\right] \cdot \left[\begin{smallmatrix} I_n \\ Dg(\mathbf{t}) \end{smallmatrix}\right] = D_{\mathbf{x}}(\mathbf{t}, g(\mathbf{t})) + D_{\mathbf{y}}(\mathbf{t}, g(\mathbf{t})) \cdot Dg(\mathbf{t}) = 0$, which gives (d) and complete the proof. $\square$

Chapter 13

Multivariable Integration

Having now discussed differentiation in multiple dimensions, we turn back to the topic of integration. We spent multiple chapters developing the Lebesgue Theory of integration and this is a situation in where that work pays off. Having discussed integration on general measure spaces, if we can develop a suitable measure for $\mathbb{R}^n$, we will immediately be able to integrate functions $f : \mathbb{R}^n \to \mathbb{R}$. We now begin the process of constructing the higher dimensional Lebesgue Measure, we will startw with a discussion of Product Measures!

13.1 Product Measures

At this point in time, we have a measure for $\mathbb{R}$, the Lebesgue Measure λ. We know that $\mathbb{R}^n$ is the Cartesian Product of copies of $\mathbb{R}$, and so a natural question is if we can make a measure in a similar fashion. This is in fact true, and given any two measures we can make a Product Measure, that fits naturally! This section is dedicated to this topic.

Definition 13.1. Suppose we have two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, then a set of the form $A \times B$, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$ is called a **Measurable Rectangle**. We define the **Product σ-Algebra** of the two measurable spaces as the smallest σ-Algebra generated by the collection of Measurable Rectangles, and denote this by $\mathcal{A} \times \mathcal{B}$.

The pair $(X \times Y, \mathcal{A} \times \mathcal{B})$ is a measurable space.

Suppose $E \subseteq X \times Y$, $x \in X$, and $y \in Y$. Then we define the **x-Slice** and **y-Slice** of E respectively as $E_x = \{y \in Y | (x, y) \in E\}$ and $E^y = \{x \in X | (x, y) \in E\}$.

This can be extended to functions as well. If f is a function on $(X \times Y)$, then the **x-Slice** and **y-Slice** of f are defined respectively as f_x on Y by $f_x(y) = f(x, y)$ and f^y on X by $f^y(x) = f(x, y)$.

The next theorem gives some information regarding this.

Theorem 13.2. Suppose we have measurable spaces $(X, \mathcal{A})$, $(Y, \mathcal{B})$, and $(X \times Y, \mathcal{A} \times \mathcal{B})$, $E \in \mathcal{A} \times \mathcal{B}$, and $f : X \times Y \to \mathbb{R}$ is a measurable function. Then

(a) For all $x \in X$ and $y \in Y$ we have $E_x \in \mathcal{A}$ and $E^y \in \mathcal{B}$,

(b) $f_x : Y \to \mathbb{R}$ and $f^y : X \to \mathbb{R}$ are both measurable functions.

Proof. For (a) let $D = \{S \in \mathcal{A} \times \mathcal{B} | S_x \in \mathcal{A} \text{ for all } x \in X\}$, we would like to show that $D = \mathcal{A} \times \mathcal{B}$. It isn't hard to see that because both $\mathcal{A}, \mathcal{B}$ are σ-algebra's that D is a σ-algebra containing every measurable rectangle $A \times B$. But since $\mathcal{A} \times \mathcal{B}$ is the σ-algebra generated by measurable rectangles we see that $D = \mathcal{A} \times \mathcal{B}$. The case of S^y is of course the same completing (a).

For (b) if $O \subseteq \mathbb{R}$, is open then $f^{-1}(O) \in \mathcal{A} \times \mathcal{B}$, so that by (a) we have $f_x^{-1}(O) = f^{-1}(O)_x \in \mathcal{A}$ and of course the same for f^y completing (b). $\square$

We still have not defined a measure on the product measurable space, but we soon will. However, we will need to put some mild conditions on the underlying spaces, namely being σ-finite, and we will need the concept of a Monotone Class. We define these, and give an important theorem before moving onto defining the product measure.

Definition 13.3. Let $(X, \mathcal{A}, m)$ be a measure space. We say the space is a **σ-Finite Measure Space** if there is a countable collection of disjoint sets $X_1, X_2, ... \in \mathcal{A}$ with $m(X_n) < \infty$ for all $n = 1, 2, ..$ and

$$X = \bigcup_{n=1}^{\infty} X_n.$$

If $m(X) < \infty$ the space is instead just called **Finite Measure Space**, which is of course σ-finite.

Definition 13.4. A **Monotone Class** is a collection of sets $\mathcal{X}$ so that if $A_1, A_2, ... \in \mathcal{X}$ and $B_1, B_2, ... \in \mathcal{X}$ are countable collections of sets so that $A_i \subseteq A_{i+1}$ and $B_{i+1} \subseteq B_i$ and if

$$A = \bigcup_{n=1}^{\infty} A_n \qquad B = \bigcap_{n=1}^{\infty} B_n,$$

then $A, B \in \mathcal{X}$.

An algebra of sets is any collection of sets $\mathcal{O}$ containing $\emptyset, X$ and is closed under complements and finite unions. Given any algebra of sets $\mathcal{O}$, there is a smallest Monotone Class containing $\mathcal{O}$, this proof is just like the proof of generating a σ-algebra! The next theorem relates these ideas!

Theorem 13.309 (The Monotone Class Lemma) Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. The collection of all unions of measurable rectangles is an algebra, denote this by $\mathcal{O}$. The smallest monotone class containing these sets is equal to $\mathcal{A} \times \mathcal{B}$.

Proof. Let M be the smallest monotone class described above. Since every σ-algebra is a monotone class we automatically have that $\mathcal{A} \times \mathcal{B} \subseteq M$. To finish the proof we must show that M is a σ-algebra giving $M \subseteq \mathcal{A} \times \mathcal{B}$.

We have that $\emptyset, X \times Y \in M$. To aid us for every $E \in M$ let

$M(E) = \{A \in M | A - E \in M, E - A \in M, E \cap A \in M\}$.

Each $M(E)$ is a monotone class, this is easy to verify by performing the operations, and we have that for sets C, D that $D \in M(C)$ if and only if $C \in M(D)$. Further, since $\mathcal{O}$ is an algebra we have that for all $E, F \in A$ that $F \in M(E)$ and $E \in M(F)$. But now we clearly see that for any set $E \in A$ that $M(E)$ is a monotone class containing A, so that $M(E) = M$. In other words, for every two sets $C, D \in M$ we have that $C - D$, $D - C$, and $C \cap D$ are all in M. Since $D^c = X - D$, M is closed under complements. If $A_1, A_2, ...$ is a countable collection of sets in M, then we have

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left(\bigcup_{k=1}^{n} A_k \right) \in M,$$

so that M is a σ-algebra completing the proof. □

We are now ready to define the Product Measure!

Theorem 13.310 (The Product Measure) Suppose $(X, \mathcal{A}, m_X)$ and $(Y, \mathcal{B}, m_Y)$ are σ-finite measure spaces and $E \in \mathcal{A} \times \mathcal{B}$. Then for all $x \in X$ and $y, \in Y$ we have the functions $m_Y(E_x)$ are measurable, $m_x(E^y)$ are measurable, and

$$\int_X m_Y(E_x)\, dm_X = \int_Y m_X(E^y)\, dm_Y. \tag{13.1}$$

In other words we can integrate the slices of a $\mathcal{A} \times \mathcal{B}$ set and the order in which we integrate doesn't matter. In this situation we define the **Product Measure** of m_X and m_Y as $m_X \otimes m_Y : \mathcal{A} \times \mathcal{B} \to [0, \infty]$ as the common value of the integrals above. This is a measure by Theorem 11.XX.

The triple $(X \times Y, \mathcal{A} \times \mathcal{B}, m_X \otimes m_Y)$ is the **Product Measure Space**, and we have for all measurable rectangles $A \times B$ that $m_X \otimes m_Y(A \times B) = m_X(A) \times m_Y(B)$.

Proof. We work similarly to Theorem 13.XX, but this time we will use the Monotone Class Lemma. Let $D = \{E \in \mathcal{A} \times \mathcal{B} |\ (13.1) \text{ holds for } E\}$. For a measurable rectangle $A \times B$, we see that for all $x \in A$ we have $m_Y((A\times B)_x) = m_Y(B)$ and for $x \notin A$ we have $m_Y((A\times B)_x)) = 0$ so that $\int_X m_Y((A \times B)_x)\, dm_X = \int_A m_Y(B)\, dm_X = m_X(A) \times m_Y(B)$, and the same works in the other direction, so that (13.1) holds for every measurable rectangle. We can in fact extend this to unions of measurable rectangles by taking sums.

Suppose $A_1, A_2, ...$ is a countable collection of sets in D with $A_n \subseteq A_{n+1}$ and

$$A = \bigcup_{n=1}^{\infty} A_n.$$

Then setting $f_n = m_Y((A_n)_x)$ and $g_n = m_X((A_n)^y)$, we see that $f_n \leq f_{n+1}$, $g_n \leq g_{n+1}$, $f_n \to m_Y(A_x)$, and $g_n \to m_X(A^y)$ so that by the Monotone Convergence Theorem we have

$$\int_X m_Y(A_x)\, dm_X = \lim_{n\to\infty} \int_X f_n((A_n)_x)\, dm_X = \lim_{n\to\infty} \int_Y g_n((A_n)^y)\, dm_Y = \int_Y m_X(A^y)\, dm_Y,$$

so that $A \in D$.

If $B_1, B_2, ...$ is a countable collection of sets in D with $B_{n+1} \subseteq B_n$, and we add the assumption that B_1 is a measurable rectangle where (13.1) holds and is finite, then the Dominated Convergence Theorem shows that $B = \bigcap_{n=1}^{\infty} B_n \in D$.

To extend this to all decreasing sequences of sets (without restrictions on B_1), since X and Y are σ-finite, let X_n and Y_n be countable collections of sets as in the definition of σ-finite spaces, and let $E_1, E_2, ...$ be a decreasing sequence of sets in D, with $E = \bigcap_{n=1}^{\infty} E_n$. Then by the previous argument we see setting

$$E'_n = E \cap \bigcup_{k=1}^{n} (X_k \times Y_k)$$

that $E'_n \in D$ for all $n = 1, 2, 3...$ and $E'_n \subseteq E'_{n+1}$ with

$$E = \bigcup_{n=1}^{\infty} E'_n,$$

by the case on increasing sequences we see $E \in D$. Since D is now a monotone class containing all unions of measurable rectangles, the Monotone Class Lemma tells us that $D = \mathcal{A} \times \mathcal{B}$, completing the proof. $\square$

13.2 The N-Dimensional Lebesgue Measure

Recall that we defined the Lebesgue Measure space on $\mathbb{R}$ as $(\mathbb{R}, \mathcal{M}, \lambda)$. Having defined Product Measures we can turn towards defining the N-Dimensional Lebesgue Measure. It may seem like the easy thing to do is to simply define the N-Dimensional Lebesgue Measure as the Product Measure of $\lambda \otimes \lambda ... \otimes \lambda$. It is not quite this simple, though the general result is not much more difficult.

Consider the case of $\mathbb{R}^2$. Its easy to see that every "line" in $\mathbb{R}$ should have zero Area. For example if we've defined a measure space by $(\mathbb{R}^2, \mathcal{M} \times \mathcal{M}, \lambda \otimes \lambda)$ and $A = [0, 1] \subseteq \mathbb{R}$, then the set $A \times \{0\} \subseteq \mathbb{R}^2$ satisfies that $\lambda \otimes \lambda(A \times \{0\}) = 0$ as we would hope! But now consider a nonmeasurable subset say $B \subseteq A$. Then we have that $B \times \{0\} \notin \mathcal{A} \times \mathcal{A}$, so that our new measure space is not complete! This tells us exactly what problem we need to fix, and since we've discussed completions of measure spaces we can proceed!

Definition 13.7. We define the **N-Dimensional Lebesgue Measure Space** on $\mathbb{R}^n$ as the completion of the measure space $(\mathbb{R}^n, \mathcal{A} \times ... \times \mathcal{M}, \lambda \otimes ... \otimes \lambda)$, which we denote by $(\mathbb{R}^n, \mathcal{M}^n, \lambda^n)$.

We would now like to verify some properties concerning this measure space just as we did in the one dimensional case.

Theorem 13.312 (Properties of N-Dimensional Lebesgue Measure) The measure space $(\mathbb{R}^n, \mathcal{M}^n, \lambda^n)$ is a complete Radon Measure. Further, for an n-Box $A = [a_1, b_1] \times ... \times [a_n, b_n]$ we have $\lambda^n(A) = \Pi_{k=1}^n (b_k - a_k)$ (the same holds for open, half open, and any combination of these cases) and that λ^n is translation invariant.

Proof. Most of these follow from the one dimensional case and our definition. The two most interesting results are the regularity and translation invariance. For regularity let $\epsilon > 0$ and suppose $A \in \mathcal{M}^n$. Then $A = B \cup C$, where $B \in \mathcal{M} \times ... \times \mathcal{M}$ and $C \subseteq D \in \mathcal{M} \times ... \times \mathcal{M}$, with $\lambda \otimes ... \otimes \lambda(D) = 0$. It's not hard to see we can cover C by a collection of open rectangles with measure as small as we'd like, so we need only consider B. If we consider the projections of B onto each copy of $\mathbb{R}$ we can find open intervals covering each projection and the difference in measure can small as we like by the regularity of the one dimensional Lebesgue Measure. Taking measurable rectangles from these open intervals will give us a collection of open measurable rectangles covering B as arbitrarily closely as we'd like. The proof of inner regularity is the same idea.

Now for translation invariance. Suppose $E \in \mathcal{M}^n$ and $\mathbf{x} \in \mathbb{R}^n$. Then consider the set $E + \mathbf{x}$. As above we can write $E = B \cup C$, with B in the product measure and C a null set. By applying the one dimensional case to every side of B we get that $B + \mathbf{x}$ is measurable, and it's easy to see that $C + \mathbf{x}$ is a null set so that $E + \mathbf{x}$ is measurable. The fact its measure is equivalent follows from applying the regularity above (open set $O + \mathbf{x}$ is still open!). $\square$

The next theorem gives one more nice fact about Lebesgue Measure that we will need later on!

Theorem 13.9. We have $E \in \mathcal{M}^n$ if and only if $E = B \cup C$, where B is a F_σ set (a countable union of closed sets) and $\lambda^n(c) = 0$.

Proof. We will essentially use the regularity properties to prove this the first direction with $E \in \mathcal{M}^n$. For each $k \in \mathbb{N}$ we inductively choose a compact (therefore closed) set $C_k \subseteq E$ so that $C_k \subseteq E$, $C_{k-1} \subseteq C_k$, and $\lambda^n(E - C_k) < \frac{1}{k}$, or equivalently $\lambda^n(E) < \lambda^n(C_k) + \frac{1}{k}$.

If we set $B = \bigcup_{k=1}^{\infty} C_k$, then B is certainly a F_σ set with $B \subseteq E$ and

$\lambda^n(B) = \lim_{k\to\infty} \lambda^n(C_k) \geq \lim_{k\to\infty} \lambda^n(E) - \frac{1}{k} = \lambda^n(E)$, or $\lambda^n(B) = \lambda^n(E)$. Setting

$C = E - B$ gives the result.

The converse is immediate as F_σ sets and null sets are measurable.

It's easy to see that a similar result holds for G_δ sets, we have $E = D - F$ where D is a G_δ set and $\lambda^n(F) = 0$. $\square$

Having defined the N-Dimensional Lebesgue Measure, we immediately can discuss integration!

Definition 13.10. We define the **Integral** of a function $f : E \to \mathbb{R}$, where $E \in \mathcal{M}^n$, as the Lebesgue integral of f over E with respect to the N-Dimensional Lebesgue Measure:

$$\int_E f \, d\lambda^n.$$

We will more commonly use the following notations:

$$\int_E f \, d\mathbf{x} \quad \text{or} \quad \int_E f \, dV_n \quad \text{or} \quad \int_E f \, dx_1 dx_2 ... dx_n.$$

For the special cases of $\mathbb{R}^2$ and $\mathbb{R}^3$ we may use the notation used back in Calculus Three. Finally, to integrate functions from $f : E \to \mathbb{R}^m$, where $E \in \mathcal{M}^n$, we define this as

$$\int_E f \, d\mathbf{x} = \begin{bmatrix} \int_E f_1 \, d\mathbf{x} \\ \vdots \\ \int_E f_m \, d\mathbf{x} \end{bmatrix}.$$

We've successfully developed the integral for multivariable functions, but we have no tools for computing this integral. The next two sections will provide the two major tools that will assist us in this task.

13.3 Fubini's Theorem

In Caclculus Three we often computed integrals via iterated integration. For example if $E = [0,1] \times [0,1]$ and $f(x) = x + y$ we can do the following:

$$\int_E f\,dA = \int_0^1 \left[\int_0^1 (x+y)\,dy\right] dx = \int_0^1 \left(x + \frac{1}{2}\right) dx = 1.$$

Essentially, we fixed x and integrated first with respect to y, then integrated the resulting function with respect to x. Is this result justified? Not yet, but Fubini's Theorem will tell us that this technique is justified for general Product Measures! We first prove Tonelli's Theorem which is a related result, but in a more specialized situation.

Theorem 13.315 (Tonelli's Theorem) Suppose $(X, \mathcal{A}, m_X)$ and $(Y, \mathcal{B}, m_Y)$ are σ-finite measure spaces, and $f : X \times Y \to [0, \infty]$ is $(\mathcal{A} \times \mathcal{B})$ measurable. Then the function $g(x) = \int_Y f_x\,dm_Y$ is $\mathcal{A}$ measurable, and the function $h(y) = \int_X f^y\,dm_X$ is $\mathcal{B}$ measurable, and

$$\int_{X\times Y} f\,d(m_X \otimes m_Y) = \int_X g(x)\,dm_X = \int_Y h(y)\,dm_Y.$$

If we plug in the definitions for g and h we see that

$$\int_{X\times Y} f\,dm_X \otimes m_Y = \int_X \left[\int_Y f_x\,dm_Y\right] dm_X = \int_Y \left[\int_X f^y\,dm_X\right] dm_Y,$$

which means we can integrate functions by fixing one variable, integrating with respect to the other, and then integrate again!

Proof. From the definition of the product measure, this is automatically true for Indicator functions, so by linearity of the integral we have the result holds for all nonnegative simple functions. By the First Simple Approximation Theorem there is a monotone sequence of simple functions $s_1, s_2, \ldots$ converging to f on $X \times Y$. Suppose we define $g_n(x) = \int_Y (s_n)_x\,dm_Y$.

Then one application of the monotone convergence theorem for g_n gives that

$$\lim_{n\to\infty}\int_Y (s_n)_x\,dm_Y = \int_Y f_x\,dm_Y \qquad \text{or} \qquad g_n \to g.$$

Now applying the monotone convergence theorem two more times as well as from above we know the result holds for simple functions gives

$$\int_{X\times Y} f\,d(m_X\otimes m_Y) = \lim_{n\to\infty}\int_{X\times Y} s_n\,d(m_X\times m_Y) = \lim_{n\to\infty}\int_X g_n\,dm_X = \int_X g\,dm_X.$$

This completes the proof, as the same arguments gives the result for the function h. $\square$

Now essentially as a Corollary we can prove Fubini's Theorem!

Theorem 13.316 (Fubini's Theorem) Suppose $(X,\mathcal{A},m_X)$ and $(Y,\mathcal{B},m_Y)$ are σ-finite measure spaces, and $f : X\times Y\to\mathbb{R}$. If $f\in L^1(m_X\otimes m_Y)$, then $f_x\in L^1(m_Y)$ a.e. on X and $f^y\in L^1(m_X)$ a.e. on Y, and setting $g(x)=\int_Y f_x\,dm_Y$ and $h(y)=\int_X f^y\,dm_X$ we have

$$\int_{X\times Y} f\,d(m_X\otimes m_Y) = \int_X g(x)\,dm_X = \int_Y h(y)\,dm_Y.$$

Proof. If $f\in L^1(m_X\otimes m_Y)$, then $\int_{X\times Y} f\,d(m_X\otimes m_Y)<\infty$ and so by Tonelli's Theorem we see that $|g|<\infty$ a.e. on Y, giving that $f_x\in L^1(m_Y)$ a.e. on Y. If we now apply Tonelli's Theorem to the positive and negative parts of f we get Fubini's Theorem for g, and the same argument gives h.

One note, we must verify that f is both $(\mathcal{A}\times\mathcal{B})$ measurable and that it has a finite integral to apply Fubini's Theorem. But if we have measurability, then we can apply Tonelli's Theorem to $|f|$ and if the iterated integrals are finite so is the integral of $|f|$, giving that $f\in L^1(m_X\otimes m_Y)$. $\square$

Great the issue of iterated integration is solved! Except its not! The theorems above only apply to product measures, and we have defined the N-Dimensional Lebesgue Measure as a completion of product measures! Therefore, we cannot directly apply Fubini's or Tonelli's Theorems. To remedy this, we prove a small theorem and then give the hypotheses to ensure the theorems previously given hold!

Theorem 13.13. Suppose $(X, \mathcal{A}, m)$ is a measure space, and $(X, \mathcal{A}', m')$ is its completion. If f is $\mathcal{A}'$ measurable function, then there exists a $\mathcal{A}$ measurable function, g, so that $f = g$ a.e. (w.r.t. m).

Proof. Let $s_1, s_2, ...$ be a sequence of simple functions converging to f, which exist by the Second Simple Approximation Theorem. Then each $s_n = \sum_{i=1}^{k_n} a_{i,n} \mathbb{1}_{A_{i,n}}$. Since each $A_{i,n} \in \mathcal{A}'$ we can write is as $B_{i,n} \cup C_{i,n}$ with $B_{i,n} \in \mathcal{A}$ and $C_{i,n}$ a null set. Now if we define $t_n = \sum_{i=1}^{k_n} a_{i,n} \mathbb{1}_{B_{i,n}}$, then g is $\mathcal{A}$ measurable and equals f a.e. completing the proof. □

Theorem 13.318 (Fubini and Tonelli Theorem for Complete Spaces)
Suppose $(X, \mathcal{A}, m_X)$ and $(Y, \mathcal{B}, m_Y)$ are complete measure spaces. Denote the completion of their product measure by $(Z, \mathcal{C}, m_Z)$. If f is $\mathcal{C}$ measurable, then f_x is measurable for almost all x, f^y is measurable for almost all y, and Fubini's and Tonelli's Theorems hold with their various assumptions.

Proof. Let g be the function that is $\mathcal{A} \times \mathcal{B}$ measurable and is equal to f a.e.. Then set $h = f - g$, we see that $h = 0$ a.e. (w.r.t. m_Z).
Let $D = \{(x, y) \in X \times Y | h(x, y) \neq 0\}$. By definition of complete measure, there is some set $E \in \mathcal{A} \times \mathcal{B}$ so that $D \subseteq E$ and $m_X \otimes m_Y(E) = 0$. But from the definition of product measure we have then that $\int_X m_y(E_x)\, dm_X = \int_Y m_x(E^y)\, dm_Y = 0$ so that $E_x = 0$ a.e. on X and $E^y = 0$ a.e. on Y. Since $(X, \mathcal{A}, m_X)$ and $(Y, \mathcal{A}, m_Y)$ are complete, $D_x \subseteq E_x$, and $D^y \subseteq E^y$ we see that $D_x = 0$ a.e. on X and $D^y = 0$ a.e. on X. It follows then that $f_x = g_x$ a.e. on X and $f^y = g^y$ a.e. on Y. It follows then that if we apply Fubini's or Tonelli's theorem's to g, that all integrals involved will match with the corresponding integrals of f giving the result! □

It follows we can now integrate functions $f : \mathbb{R}^n \to \mathbb{R}$ as

$$\int_{\mathbb{R}^n} f\, d\mathbf{x} = \int_{\mathbb{R}} \ldots \left[\int_{\mathbb{R}} f\, dx_1 \right] \ldots dx_n,$$

where we can change which order we integrate each dx_i, and at each step we integrate only with respect to the x_ith variable

13.4 The Change of Variables Theorem

The other major technique used to evaluate higher dimensional integrals is found in the Change of Variables Theorem. It is essentially a generalization if the single variable integration by substitution formula. As an exmaple, in Caluclus Three you might encounter the integral of the function $f(x) = 1$ over, the unit sphere S. We could try to integrate using the x, y, z basis as

$$\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} 1\, dzdydx,$$

but this is hideous. We likely instead used a "change of variables" by substituting to spherical coordinates with $x = \rho \cos\theta \sin\phi$, $y = \rho \sin\theta \sin\phi$, and $z = \rho\cos\phi$, giving the simpler integral

$$\int_0^{\pi} \int_0^{2\pi} \int_0^1 \rho^2 \sin\phi\, drd\theta d\phi.$$

This integral is much easier to compute and gives the expected value of $\frac{4\pi}{3}$. But, how is this justified and why did a $\rho^2 \sin\phi$ term appear in the integral? The answer comes from the change of variables theorem, which this section is dedicated to proving. Essentially, we can think of a "change of variables" as a smooth ($\mathcal{C}^1$) mapping from one set to another. In this case we can integrate in the new space, but we must change our function by how this mapping scales the mapped space, which in this sense means multiplying by the Jacobian Determinant! The next theorem gives a special case of the change of variables theorem.

Theorem 13.15. Suppose $T \in GL(n,\mathbb{R})$, $E \subseteq \mathcal{M}^n$, $f : E \to \mathbb{R}$, and $f \in L^1(\lambda^n)$ on $\mathbb{R}^n$. Then $(f \circ T) \in L^1(\lambda^n)$ on $\mathbb{R}^n$ and

$$\int_{\mathbb{R}^n} f \, d\mathbf{x} = \int_{\mathbb{R}^n} (f \circ T)|\det T| \, d\mathbf{x}.$$

Proof. We start by assuming we are in the special case where we can assume that $(f \circ T)$ is measurable, the general case will follow from this.

If the theorem is true then for transformations A, B we have

$$\int_{\mathbb{R}^n} f \, d\mathbf{x} = |\det A| \int_{\mathbb{R}^n} (f \circ A) \, d\mathbf{x} = |\det A||\det B| \int_{\mathbb{R}^n} (f \circ A \circ B) \, d\mathbf{x},$$

so that it holds for compositions. That means we need only prove the theorem for the special transformations T_1, T_2, and T_3 as given in Chapter 12 in the Section on ... because every invertible transformation can be expressed as a finite composition of these three types. For T_1 we use Fubini's/Tonelli's Theorems. Isolate the variable that is being changed, say x_i. Then we have $|\det T_1| = c$ so that

$$\int_{\mathbb{R}} f(x_1, ..., x_i, ..., x_n) \, dx_i = |c| \int_{\mathbb{R}} f(x_1, ..., cx_i, ...x_n) \, dx_i,$$

giving

$$\int_{\mathbb{R}^n} f \, d\mathbf{x} = |\det T_1| \int_{\mathbb{R}^n} (f \circ T_1) \, d\mathbf{x},$$

completing the case of T_1. For T_2 we have $\det T_2 = 1$, and since $\int_{\mathbb{R}} f(x_1, ..., x_i + cx_j, ..., x_n) \, dx_i = \int_{\mathbb{R}} f(x_1, ..., x_i, ..., x_n) \, dx_i$, the case follows immediately. Finally, T_3 is just as we need only swap the order of integration and $\det T_3 = -1$ $|\det T_3| = 1$, which agrees with the fact that changing the order of integration doesn't change the result.

We now finish the case for a general function f. Since T is a homeomorphism, if E is some Borel set then $T(E)$ is also Borel, so that setting $f = \mathbb{1}_E$ the composition $(f \circ T) = \mathbb{1}_{T(E)}$ is measurable, and the above gives $\lambda^n(T(E)) = |\det T|\lambda^n(E)$.

Now we prove that in general $(f \circ T)$ is measurable, let O be open. Then $f^{-1}(O) = B \cup C$, where B is a F_σ set, so that $T^{-1}(B)$ is also a F_σ set. On the other hand by Theorem XXX we can find a G_δ set, D, containing C with measure zero. It follows that $T^{-1}(D)$ is another G_δ set and $\lambda^n(T^{-1}(D)) = 0$, and since $T^{-1}(C) \subseteq T^{-1}(D)$ we get that $\lambda^n(T^{-1}(C)) = 0$. In other words $(f \circ T)^{-1}(O) = T^{-1}(B) \cup T^{-1}(C)$, which is the union of a F_σ set and a set of measure zero and is therefore measurable! Therefore $(f \circ T)$ is a measurable function completing the proof. $\square$

Corollary 13.15.1. Setting $f(x) = \mathbb{1}_E$ for a measurable set E, we arrive at the formula $\lambda^n(T(E)) = |\det T|\lambda^n(E)$, so that the determinant gives the scaling factor of a transformation!

We are now ready for the full Change of Variables Theorem to end this chapter!

Theorem 13.320 (Change of Variables Theorem) Suppose that $O \subseteq \mathbb{R}^n$ is open, $g \in \mathcal{C}^1$ on O and $f \in L^1(\lambda^n)$ on $g(O)$. Then $(f \circ g) \in L^1(\lambda^n)$ on O and

$$\int_{g(O)} f(\mathbf{x})\, d\mathbf{x} = \int_O (f \circ g)(\mathbf{x})|\det Dg(\mathbf{x})|\, d\mathbf{x}.$$

Proof. Since O is open and g is a homeomorphism $g(O)$ is open and therefore measurable. The fact that $(f \circ g)$ is a measurable function follows just as in Theorem XXX above, we prove the special case where it is assumed measurable and then use a F_σ and null set argument for the general result, so we assume that $(f \circ g)$ is measurable. It is now a question of integral equality (which of course gives $(f \circ g) \in L^1$).

We will adopt a new norm for this problem as well, called the **Maximum Norm**. For a vector $\mathbf{x} \in \mathbb{R}^n$ we define $\|\mathbf{x}\|_m = \max\{|x_1|, ..., |x_n|\}$. This norm is useful, as closed/open balls in this norm are cubes with side lengths $2r$, about the center. It's also easy to verify that for transformations T we have $\|T\mathbf{x}\|_m \leq \|T\|_{op}\|\mathbf{x}\|_m$.

We start by assuming that $C = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{x}_0\|_m \leq r\}$ is a closed cube contained in O with side lengths $2r$. We see that $\lambda^n(C) = (2r)^n$, and since C is compact and Dg is continuous, $\|Dg\|_{op}$ attains a maximum, say M. A slight variation of Theorem XXX gives that for all $\mathbf{x} \in C$ we have

$$\|g(\mathbf{x}) - g(\mathbf{x}_0)\|_m \leq M\|\mathbf{x} - \mathbf{x}_0\|_m,$$

so that $g(C)$ is contained in a cube of side lengths $2Mr$, let C' be this cube, we immediately have that $\lambda^n(C') = (2Mr)^n$, so that

$$\lambda^n(g(C)) \leq M^n \lambda^n(C).$$

Now suppose $T \in GL(n, \mathbb{R})$ is arbitrary. Applying the above corollary we get

$$\lambda^n(g(C)) = |\det T| \lambda^n(T^{-1}(g(C)) = |\det T| \lambda^n(T^{-1} \circ g(C)).$$

Now performing the same argument we just used with $T^{-1} \circ g$ in place of g, and using K to correspond with M we get

$$\lambda^n(g(C)) \leq |\det T| K^n \lambda^n(C).$$

Let $\epsilon > 0$, by continuity of Dg we choose $\delta > 0$ so that $\|\mathbf{y} - \mathbf{x}\|_m < \delta$ implies

$$\|Dg^{-1}(\mathbf{y}) Dg(\mathbf{x})\|_{op} < 1 + \epsilon.$$

It follows if we divide C into smaller cubes of side length less that δ with disjoint interiors, say $C_1, ..., C_N$ with centers $x_1, ..., x_N$, then for each of these cubes we have

$$\lambda^n(g(C_i)) \leq |\det Dg(x_i)| \lambda^n(C_i)(1 + \epsilon),$$

so that

$$\lambda^n(g(C)) \leq \sum_{i=1}^{N} \lambda^n(g(C_i)) \leq (1+\epsilon) \sum_{i=1}^{N} |\det Dg(x_i)| \lambda^n(C_i)$$

$$= (1+\epsilon) \int_{C_i} \sum_{i=1}^{N} |\det Dg(x_i)| \, d\mathbf{x}.$$

Taking limits gives that $\lambda^n(g(C)) \leq \int_C |\det Dg(x)| \, d\mathbf{x}$. Via coverings with cubes, we can extend this formula to hold for all Borel Sets, and therefore for null sets as well contained in O. Thus, for all positive simple functions, and via the Monotone Convergence Theorem for all positive functions f we have

$$\int_{g(O)} f \, d\mathbf{x} \leq \int_O (f \circ g) |\det Dg| \, d\mathbf{x}.$$

The function $(f \circ g)|\det Dg|$ on $O = g^{-1}(g(O))$ is positive and since g^{-1} satisfies all the criterion's we apply the above with g^{-1} getting

$$\int_{g^{-1}(g(O))} (f \circ g) |\det Dg| \, d\mathbf{x} \leq \int_{g(O)} (f \circ g \circ g^{-1})(x) |\det Dg(g^{-1}(x))| |\det Dg^{-1}(x)| \, d\mathbf{x}$$

$$= \int_{g(O)} f \, d\mathbf{x},$$

which completes the proof. $\square$

Corollary 13.16.1. In the above situation we also have that for any subset $S \subseteq O$ that $g(S)$ is measurable and $\lambda^n(g(S)) = \int_S |\det Dg| \, d\mathbf{x}$, by applying the above to the function $f = \mathbb{1}_S$.

Our one issue remaining is that this doesn't quite work for Polar or Rectangular regions. We can compute the Jacobian of the Transformation $F(r, \theta)$ with $x = r\cos(\theta)$, $y = r\sin(\theta)$ to get $|\det DF| = r$, which is not invertible when $r = 0$, so that our theorem doesn't hold in any region with $r = 0$. Similar issues occur with cylindrical and spherical coordinate transformations. The next theorem says that a differentiable transformation g maps sets of measure zero to measure zero. This remedies this problem and will conclude the chapter! This proof is adapted from the article XXX by D. E. Varberg

Theorem 13.17. Suppose $\lambda^n(E) = 0$ and f is differentiable on E. Then $\lambda^n(f(E)) = 0$.

Proof. Fix $m, n \in \mathbb{N}$ and let E_{ij} be the set of $\mathbf{x} \in N$ so that $\|f(\mathbf{x}+\mathbf{t}) - f(\mathbf{x})\| \leq m\|\mathbf{t}\|$ for all $\mathbf{t}$ with $\|\mathbf{t}\| < \frac{1}{n}$. Since $E = \bigcup E_{ij}$ we need to show each E_{ij} has finite measure. Let $\epsilon > 0$. Next let D be a G_δ set of measure zero containing E_{ij}. Then $f(D)$ is also a G_δ set and therefore measurable. Since D has zero measure, we can find a countable collection of cubes $C_i = \|x - x_i\|_m < r$, where $\sum_{i=1}^\infty \lambda^n(C_i) < \epsilon$. It follows via the mean value theorem each $f(C_i)$ is contained in a cube with sidelengths $m(2r)$, called C_i', and so $\sum_{i=1}^\infty \lambda^n(C_i') = \sum_{i=1}^\infty m^n \lambda^n(C_i) < m^n\epsilon$. The sets C_i' cover $f(D)$ and this holds for all ϵ so $\lambda^n(f(D)) = 0$ and therefore $\lambda^n(f(E_{ij})) = 0$ completing the proof. $\square$

Chapter 14

Multivariable Chains and Forms

We've successfully covered a great deal of a typical Calculus Three course with multivariable differentiation and integration. We come to the final topic of such a course, typically referred to as Vector Calculus. This concerns integration of scalar and vector valued functions over lower dimensional subsets of a space. Things one likely saw were Line and Surface integrals, Green's Theorem, Stokes' Theorem, and more. We briefly discussed this last section, but our current setup for integration does not cover integration of this type. For example consider the path $\gamma : [0,1] \to \mathbb{R}^2$ by $f(x) = (x, 0)$. If we try to integrate any function, f, along this path with respect to the Lebesgue Measure we will get

$$\int_{\gamma([0,1])} f \, dxdy = 0,$$

because $\lambda^2(\gamma([0,1])) = 0$. In general, integrating over a lower dimensional subset of $\mathbb{R}^n$ will always give 0 if the integral is with respect to the Lebesgue Measure. There are several remedies to this problem, notably the Hausdorff Measure, but we will take a simpler approach here, opting to develop a very specialized measure for such purposes. But first, we want to discuss what regions we would like to integrate over. Again, there are several paths we could take, the most notable being integrating over **Manifolds**. These are incredibly general regions that locally resemble real space and are incredibly useful. This book won't explore this option (FOR NOW MAYBE LATER), instead opting for **K-Cells**. This chapter will be quite full of definitions, as the end goal is The Generalized Stokes' Theorem which is simple to prove, but requires an incredible amount of machinery! One last comment before we begin, is the difficulty in this chapter is less the conceptual difficulty of the topics, but rather the intense amount of definitions and nested levels of indices. It can be overwhelming, so concrete examples in low dimensions will be very helpful and should be considered to gain intuition. We now begin!

14.1 Cells and Chains

Recall, that a **k-Box** was a set of the form $[a_1, b_1] \times ... \times [a_k, b_k] \subseteq \mathbb{R}^k$, and in particular the **Unit k-Box** is $[0,1]^k \subseteq \mathbb{R}^k$. Throughout the chapter we will assume that $n \geq k$, when comparing spaces $\mathbb{R}^k$ and $\mathbb{R}^n$.

Definition 14.1. We define a **k-Cell** as a continuous function $c : [0,1]^k \to \mathbb{R}^n$. We will typically add the assumption that $c \in \mathcal{C}^1$, but this is not strictly part of the definition. Note that c maps the unit box into a space that may be of higher dimension. The **Standard k-Cell** is the mapping $I^k : [0,1]^k \to \mathbb{R}^k$ by $I^k(\mathbf{x}) = \mathbf{x}$.
We could have assumed c mapped any k-Box rather than the unit box, as these regions are homeomorphic, but this definition is a bit simpler. Another note of interest is that the cell is NOT the image of the function, but rather the function itself. This is in agreement with how we've defined paths (which are k-Cells!). That being said we typically think of taking an integral over the image of the k-Cell.

Definition 14.2. Typically a single Cell will not be sufficient, so we introduce the notion of a Chain! A **k-Chain**, C, is a "formal sum" of k-Cells, ie if $c_1, ..., c_m$ are all k-Cells, then a k-Chain is an expression of the form $C = \sum_{i=1}^{m} a_i c_i$, where each $a_i \in \mathbb{Z}$. Intuitively, think of these coefficients as being a multiplicity, as in this region shows up a_i times, and negative multiplicities mean taking the region in the opposite orientation. Treating these as sums is useful for adding chains, and being able to multiply by scalars! In particular, if we have a cell c and add it with $-c$, we will get 0. This agrees with how if we integrate along one orientation, then the opposite they will cancel! This notion will also be useful for the next definition.

Definition 14.3. Having discussed chains, we can now define the boundary of a cell, and therefore of a chain as well! It is easy enough to intuitively picture what we mean by the boundary of a k-cell, it should be the region mapped to by the boundary of the unit box, but we clearly need more than one cell for this, hence one example of the need for chains! Now for the definition, which is split into several parts. For k dimensions, $[0,1]^k$, will have $2k$ sides. For example the unit square has four sides, the unit cube six, and at this point we are limited by our own three dimensional space, but the idea extends. For $1 \leq i \leq k$, there are two sides, the one where the i-th coordinate is fixed at zero, and where it is fixed at one. The **i-0 Side** is defined as the (k-1)-Cell $I^k_{i,0} : [0,1]^{k-1} \to \mathbb{R}^k$ by $I^k_{i,0}(x_1, ..., x_{k-1}) = (x_1, ..., x_{i-1}, 0, x_i, ..., x_{k-1})$ and similarly the **i-1 Side** is defined as $I^k_{i,1}(x_1, ..., x_{k-1}) = (x_1, ..., x_{i-1}, 1, x_i, ..., x_{k-1})$.
Now we define the **Boundary of** I^k as the $(k-1)$-Chain

$$\partial I^k = \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} I^k_{i,a}.$$

The presence of the $(-1)^{i+a}$ should feel strange. This is designed this way so that when two different sides have a common side (a side of a side) that it will cancel. This will lead to Theorem XXX below, saying that the boundary of a boundary is zero, a very nice fact! It is a little annoying, but it is quite useful.

Now we define the **Boundary of a k-Cell** c as

$$\partial c = \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} c(I_{i,a}^{k}),$$

and we extend via linearity the **Boundary of a k-Chain** as

$$\partial \left(\sum_{i=1}^{m} a_i c_i \right) = \sum_{i=1}^{m} a_i \partial c_i.$$

Having thrown a great deal of definitions at the reader, lets finally prove something!

Theorem 14.4. As stated earlier, suppose $C = \sum_{i=1}^{m} a_i c_i$ is a k-chain, then $\partial(\partial C) = 0$.

Proof. From our definitions, we need only prove this for the Unit Box I^k, all other cases follow. Then we have that

$$\partial(\partial I^k) = \partial \left(\sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} I_{i,a}^{k} \right) = \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} \partial(I_{i,a}^{k})$$

$$= \sum_{i=1}^{k} \sum_{a=0}^{1} (-1)^{i+a} \sum_{j=1}^{k-1} \sum_{b=0}^{1} (-1)^{j+1} I_{i,a}^{k}(I_{j,b}^{k-1}) = \sum_{i=1}^{k} \sum_{a=0}^{1} \sum_{j=1}^{k-1} \sum_{b=0}^{1} (-1)^{i+j+a+b} I_{i,a}^{k}(I_{j,b}^{k-1}).$$

From plugging into the definitions however, we will get that $I_{i,a}^{k}(I_{j,b}^{k-1}) = I_{j+1,b}^{k}(I_{i,a}^{k-1})$, but the (-1) terms corresponding to these values will have different signs. Therefore we will have expressions in the sum of the form $I_{i,a}^{k}(I_{j,b}^{k-1}) - I_{j+1,b}^{k}(I_{i,a}^{k-1}) = 0$, so that the whole sum will equal zero. Again, a specific example may make this more clear. □

14.2 Functions on Chains

We now come to our first topic of integration, namely scalar function integration along Chains. This will generalize the basic line and surface integrals encountered in Calculus Three! Unfortunately, it is not immediately obvious how we should define this integral, as

mentioned integrating over the image of a lower dimensional region will always be zero. Our remedy is to find a suitable concept of an integral that applies to these situations, and then we will simply define the integral to agree with this!
Before discussing integration, we start with a brief discussion of an k-parallelepipeds, the k-dimensional equivalent to a parallelepiped. A more detailed discussion of this topic can be found in Munkres' Analysis on Manifolds XXX.

Definition 14.5. A **k-parallelepiped** is a xxx.

It turns out that a suitable definition of the "Volume" of a k-parallelepiped, P, is as follows: Let $v_1, ..., v_k$ be the side vectors of the parallelepiped, then if we make a transformation $A = \left[\begin{smallmatrix} v_1 & \cdots & v_k \end{smallmatrix} \right]$, then the correct definition of the **Volume of P** is $V(p) = \sqrt{|\det A^T A|}$. There are two main reasons for this:
1. If $k = n$ then this definition agrees with the Lebesgue Measure for P,
2. If $P \subseteq \mathbb{R}^k \times \mathbf{0}$ this value agrees with if we had just computed the measure in $\mathbb{R}^k$.
The first of these claims is easy to see, while the second can be proven via an induction argument.
We will now use this definition, along with the change of variables theorem, to motivate our definition of the integral of a scalar function over a chain!

Definition 14.6. For an arbitrary function $g \in \mathcal{C}^1$ we will define $D^T Dg(\mathbf{x}) = Dg(\mathbf{x})^T Dg(\mathbf{x})$, where this is matrix multiplication, to save time.
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous c is a $\mathcal{C}^1$ k-cell mapping into $\mathbb{R}^n$. Then we define the **Integral of f over c** as

$$\int_c f = \int_{I^k} f(c(\mathbf{x}))\sqrt{\det D^T Dg(\mathbf{x})}\, d\mathbf{x}.$$

Note there is no "dx" symbol, this is to signify this integral is NOT with respect to any measure, but rather just a definition.

Another comment is the restriction to continuous functions. This ensures the integral on the right exists and will be suitable for most purpose.

We give one familiar example of this formula now!

Example 14.7. Suppose $c : [0, 1] \to \mathbb{R}^3$ is a $\mathcal{C}^1$ k-cell. Let us name the components as follows

$$c(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}.$$

We will also rename the partial derivatives as $Dc_1(t) = x'(t)$, $Dc_2(t) = y'(t)$, and $Dc_3(t) = z'(t)$. Then we can evalute $D^T Dc(t) = x'(t)^2 + y'(t)^2 + z'(t)^2$ so that if f is a continuous function containing the image of $c([0, 1])$ then our definition gives

$$\int_c f = \int_0^1 f(c(t))\sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\, dt.$$

This is exactly the formula for the line integral of a scalar function you likely saw back in Calculus Three!
A similar computation for k-cells from $[0, 1]^2 \to \mathbb{R}^3$ gives the formula for surface integrals!

We have successfully re-derived some of the formulas familiar to us from back in Calculus Three! The next section will introduce the tool needed to re-derive the rest, namely Differential Forms!

14.3 Tangents and 1-Forms

As promised we now move onto Differential Forms. The rest of the chapter will be dedicated to these objects and the numerous definitions that go along with them. Our first objective will be to define what we mean by 1-forms. The next section will introduce the Wedge Product which will allow us to combine 1-forms and create more general k-forms.
We start by discussing Tangent Vectors and Spaces.

Definition 14.8. Given a point $p \in \mathbb{R}^n$ we define a **Tangent Vector of** p to be the ordered pair (p, v) where v is another vector in $\mathbb{R}^n$. Think of this object as the vector v, but starting from the point p, rather than starting from the origin. This idea seems silly, but it is very useful to think in this mindset.

We define the **Tangent Space of** $\mathbb{R}^n$ **at** p as the set $\mathcal{T}_p(\mathbb{R}^n) = \{(p, v) | v \in \mathbb{R}^n\}$, the set of all tangent vectors of p.

For each $p \in \mathbb{R}^n$ it is easy to see that $\mathcal{T}_p(\mathbb{R}^n)$ is a vector space if we define

$$(p, v) + (p, w) = (p, v + w),$$

and

$$c(p, v) = (p, cv).$$

Finally, we define the **Tangent Bundle** as $\mathcal{T}(\mathbb{R}^n) = \{\mathcal{T}_p(\mathbb{R}^n) | p \in \mathbb{R}^n\}$, the set of all Tangent Spaces. If it is obvious, the $\mathbb{R}^n$ may be excluded.

We can now define a **Vector Field**, sometimes referred to as a **Tangent Vector Field**, is a function $F : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ satisfying $F(\mathbf{x}) = (\mathbf{x}, G(\mathbf{x}))$, where $G : \mathbb{R}^n \to \mathbb{R}^n$ is some function. Essentially, a vector field assigns, to each point in $\mathbb{R}^n$, a tangent vector from that point, ie $F(\mathbf{x}) \in \mathcal{T}_{\mathbf{x}}(\mathbb{R}^n)$.

If we have a vector space V over a field $\mathbb{F}$, then a **Linear Functional**, T, or simply a **Functional** is a linear transformation from V to its scalar field, ie $T \in \mathcal{L}(V, \mathbb{F})$. The set of functionals on V is called the **Dual Space of** V, denoted by V^*.

Finally, we define a **Differential 1-Form**, or for short a 1-Form, as a function ω on $\mathbb{R}^n$ (or some subset) that assigns to each $\mathbf{x}$ a functional $\omega(\mathbf{x}) \in \mathcal{T}_{\mathbf{x}}(\mathbb{R}^n)^*$. So a differential 1-form takes in a point in space, and assigns a functional on the space of tangent vectors to that point, that's a mouthful!

Given a 1-Form ω, each $\omega(\mathbf{x})$ is a function that takes in a tangent vector and assigns a real number, so we have given a vector $\mathbf{v} \in \mathbb{R}^n$ we have $\omega(\mathbf{x})(\mathbf{x}, \mathbf{v}) \in \mathbb{R}$.
It is customary to denote forms using Greek letters, and functions with Latin letters.

We now define what the dx_i object should be, namely an important 1-Form!

Definition 14.9. Recall the projection function $\pi_i(x_1, ..., x_i, ..., x_n) = x_i$. Then we have for a given unit vector $\mathbf{u} \in \mathbb{R}^n$ that the directional derivative $D\pi_i(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} 0 & ... & 1 & ... & 0 \end{bmatrix} \cdot \mathbf{u} = u_i$. Using this as motivation we define the **i-th Elementary 1-Form** as

$$dx_i(\mathbf{x})(\mathbf{x}, \mathbf{u}) = u_i.$$

The following theorem gives important information about Dual Spaces, and the Corollary explains why the Elementary 1-Forms are so important.

Theorem 14.10. Consider a vector space, V over $\mathbb{R}$ with basis $\{\mathbf{a}_1, ..., \mathbf{a}_n\}$. Then V^* is a vector space if we define for $A, B \in V^*$, $c \in \mathbb{F}$, and $\mathbf{x} \in V$ we

$$(A + B)(\mathbf{x}) = A\mathbf{x} + B\mathbf{x},$$

and

$$(cA)(\mathbf{x}) = c(A\mathbf{x}).$$

Recall that a linear transformation is uniquely determined by how it acts on basis vectors. Therefore, if we define

$$\phi_i(\mathbf{a}_j) = \begin{cases} 1 & i = j \\ \\ 0 & i \neq j \end{cases},$$

these are well defined elements of V^* and form a basis for V^*.

Proof. The first claim that V^* forms a vector space is obvious. The more important claim is the basis claim! Let T be an arbitrary functional in V^*. For each $i = 1, ..., n$ set $c_i = T\mathbf{a}_i \in \mathbb{R}$.

Then we have for any $\mathbf{x} \in V$ that

$$T\mathbf{x} = c_1\phi_1(\mathbf{x}) + ... + c_n\phi_n(\mathbf{x}),$$

so that T can be represented as a linear combination of $\phi_i's$ completing the proof. $\square$

Corollary 14.10.1. If $\mathbf{x} \in \mathbb{R}^n$, then the functions $dx_i(\mathbf{x})$ for $i = 1, ..., n$ for the basis for the space $\mathcal{T}^*_{\mathbf{x}}$. It follows that if we have an arbitrary 1-form ω, at $\mathbf{x}$ we have $\omega(\mathbf{x}) = \sum_{i=1}^n c_{i,\mathbf{x}} dx_i$. It follows that if we set $f_i(\mathbf{x}) = c_{i,\mathbf{x}}$ for each $\mathbf{x}$ then we have $\omega = \sum_{i=1}^n f_i dx_i$, or every 1-form can be uniquely expressed as a sum of elementary 1-forms. The functions f_i are called the **Component Functions of** ω. We will typically require these be differentiable!

14.4 The Wedge Product

Having now defined 1-forms we seek more general k-forms in $\mathbb{R}^n$. Our primary tool in this endeavor will be the Wedge Product, which is essentially a way of combining forms and turning them into higher order forms.

Suppose ω, η are 1-forms, we want some way to combine these forms. The simplest way to combine these forms would be to take the direct product, since the output is a real number, so that for each $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ we define $(\omega \wedge \eta)(\mathbf{x})(\mathbf{x}, \mathbf{v}) = \omega(\mathbf{x})(\mathbf{x}, \mathbf{v}) \cdot (\eta(\mathbf{x})(\mathbf{x}, \mathbf{v})$, but this is not very useful at all. Instead we combine combine them as follows: We've define 1-forms to be linear functional at every point in $\mathbb{R}^n$, they take in a tangent vector and spit out a real number corresponding length of the vector in the i-th direction multiplied by the weight of the i-th component function.

To generalize this to the product of two 1-forms, which will be a 2-form, we will instead take in two tangent vectors and then will output a real number, this time corresponding to the area of the parallelogram formed by these tangent vectors in the ij-th frame. But we know how to compute area of parallelograms, this is via the Determinant! We now make our definition.

Definition 14.11. Suppose ω, η are 1-forms. Then we define the **Wedge Product** of ω and η as a function on $\mathbb{R}^n$ (or some subset) that assigns to each point of $\mathbf{x}$ a function $\omega \wedge \eta(\mathbf{x}) : \mathcal{T}_{\mathbf{x}}^* \times \mathcal{T}_{\mathbf{x}}^* \to \mathbb{R}$ defined by

$$\omega \wedge \eta(\mathbf{x})((\mathbf{x}, \mathbf{v}_1), (\mathbf{x}, \mathbf{v}_2)) = \det \begin{bmatrix} \omega(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) & \eta(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) \\ \omega(\mathbf{x})(\mathbf{x}, \mathbf{v}_2) & \eta(\mathbf{x})(\mathbf{x}, \mathbf{v}_2) \end{bmatrix}.$$

A wedge product of 2 1-forms is called a **Differential 2-Form**, or 2-form.

We now introduce some facts about the Wedge Product and 1/2-forms.

Theorem 14.12. The Wedge Product satisfies the following properties:

(a) We have $(f dx_i) \wedge (g dx_j) = (fg) dx_i \wedge dx_j$,

(b) We have $f dx_i \wedge (g dx_j + h dx_p) = (fg) dx_i \wedge dx_j + (fh) dx_i \wedge dx_p$,

(c) $dx_i \wedge dx_j(\mathbf{x})((\mathbf{x}, \mathbf{e}_p), (\mathbf{x}, \mathbf{e}_q)) = 1$ if $p = i$ and $q = j$, otherwise it is 0,

(d) $dx_i \wedge dx_j = (-1) dx_j \wedge dx_i$,

(e) $dx_i \wedge dx_i = 0$,

(f) For the general case we have

$(f_1 dx_1 + ... + f_n dx_n) \wedge (g_1 dx_1 + ... + g_n dx_n) = \sum_{i,j=1}^n (f_i g_j - f_j g_i) dx_i \wedge dx_j$.

Proof. For (a), its an easy computation that $dx_i \wedge dx_j(\mathbf{x})((\mathbf{x}, \mathbf{v}_1), (\mathbf{x}, \mathbf{v}_2)) = v_{1,i} v_{2,j} - v_{1,j} v_{2,i}$. On the other hand we have at each $\mathbf{x} \in \mathbb{R}^n$ that

$$f dx_i \wedge g dx_j(\mathbf{x})((\mathbf{x}, \mathbf{v}_1), (\mathbf{x}, \mathbf{v}_2)) = \det \begin{bmatrix} f(\mathbf{x}) dx_i(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) & g(\mathbf{x}) dx_j(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) \\ f(\mathbf{x}) dx_i(\mathbf{x})(\mathbf{x}, \mathbf{v}_2) & g(\mathbf{x}) dx_j(\mathbf{x})(\mathbf{x}, \mathbf{v}_2) \end{bmatrix}$$

$$= \det \begin{bmatrix} f(\mathbf{x}) v_{1,i} & g(\mathbf{x}) v_{1,j} \\ f(\mathbf{x}) v_{2,i} & g(\mathbf{x}) v_{2,j} \end{bmatrix} = f(\mathbf{x}) g(\mathbf{x}) v_{1,i} v_{2,j} - f(\mathbf{x}) g(\mathbf{x}) v_{1,j} v_{2,i} = (fg) dx_i \wedge dx_j,$$

completing (a). As a consequence we really need only look at elementary 1-forms and their wedge products!

For (b) we have $(dx_i) \wedge (dx_j + dx_p)$ works, this is an easy computation of determinants, so that (b) holds with (a).

We have (c) follows directly from the formula.

That (d) works is the fact that swapping columns of a matrix changes the determinant by a factor of -1.

We get (e) as $dx_i \wedge dx_i = (-1)dx_i \wedge dx_i$.

Finally, (f) follows from repeated applications of (b) and (e) completing the proof. $\square$

Definition 14.13. Going one step further if we have k 1-forms, $\omega_1, ..., \omega_k$ we can define a **Differential k-Form**, or k-form, as a wedge product of all k 1-forms with the formula

$$\omega_1 \wedge ... \wedge \omega_k(\mathbf{x})((\mathbf{x}, \mathbf{v}_1), ..., (\mathbf{x}, \mathbf{v}_k)) = \det \begin{bmatrix} \omega_1(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) & \dots & \omega_k(\mathbf{x})(\mathbf{x}, \mathbf{v}_1) \\ \vdots & \ddots & \vdots \\ \omega_1(\mathbf{x})(\mathbf{x}, \mathbf{v}_k) & \dots & \omega_k(\mathbf{x})(\mathbf{x}, \mathbf{v}_k) \end{bmatrix}.$$

An interesting fact is that while these are no longer functionals (they don't take in a tangent vector they take in k of them!) if we fix all of the vectors and vary one at a time then these functions are linear with respect to that vector. Functions of this type, that are linear when only varying one input, are called **k-Tensors** and we see that at each point our k-forms return a k-tensor that takes in k vectors and returns a real number.

Definition 14.14. Suppose $I = \{i_1, ..., i_k$ is a collection of integers with in the set $\{1, ..., n\}$ that is increasing, ie $i_1 < ... < i_k$, then we define the ***I*th Elementary k-Form** as the k-form

$$dx_I = dx_{i_1} \wedge ... \wedge dx_{i_k}.$$

Just as in the case of 1-forms, every k-form can be expressed as $\sum_I f_I dx_I$, where I ranges over all increasing collection of indices of length k taking values between 1 and n.

Definition 14.15. We can now extend the Wedge Product to its most general form. If $\omega = \sum_I f_i dx_I$ is a k-form and $\eta = \sum_J g_i dx_J$ is a l form, then we define $\omega \wedge \eta$ as the $k+l$-form by

$$\omega \wedge \eta = \sum_{I,J} f_I g_J dx_{IJ},$$

where $dx_{IJ} = dx_{i_1} \wedge ... \wedge dx_{i_k} \wedge dx_{j_1} \wedge ... \wedge dx_{j_l}$.

Because we can always rearrange the indices to be ordered increasingly, we will typically assume they are increasing, also because if any $i_m = j_n$ then $dx_I \wedge dx_J = 0$, we will assume all i's and j's are different typically.

A computation gives $dx_I \wedge dx_J = (-1)^{kl} dx_J \wedge dx_I$, as it takes kl moves to rearrange.

The indices are very quickly getting a little crazy which is why using I for a collection of indices is much nicer, unfortunately there is no way around this and notation can be one of the bigger hurdles to overcome!

Finally, the last bit of notation before moving onto integration. We denote the set of all differential k-forms over $\mathbb{R}^n$ by $\Omega^k(\mathbb{R}^n)$.

14.5 Integration of Forms

We now would like to define the Integral of a Differential Form over a k-cell. We've put in a lot of work carefully defining what a k-form is and now we can put that to the test. We first define our last two tools needed for integrating forms.

Definition 14.16. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m \in \mathcal{C}^1$ and $\mathbf{p} \in \mathbb{R}^n$. We then create a linear transformation between tangent spaces using the Directional Derivative called the **Pushforward** of f at $\mathbf{p}$, $f_* : \mathcal{T}_{\mathbf{p}}(\mathbb{R}^n) \to \mathcal{T}_{f(\mathbf{p})}(\mathbb{R}^m)$ by

$$f_*(\mathbf{p}, \mathbf{v}) = (f(\mathbf{p}), Df(\mathbf{p}, \mathbf{v})).$$

It is called the pushforward because it takes a tangent vector to $\mathbf{p}$ in the starting space, $\mathbb{R}^n$, and pushes it forward to become a tangent vector to $f(\mathbf{p})$ in the ending space, $\mathbb{R}^m$.
The more important and related concept is the Pullback! We define the **Pullback** induced by f as a function $f^* : \Omega^k(\mathbb{R}^m) \to \Omega^k(\mathbb{R}^n)$ by

$$(f^*(\omega))(\mathbf{x})((\mathbf{x}, \mathbf{v}_1), ..., (\mathbf{x}, \mathbf{v}_k)) = \omega(f(\mathbf{x}))(f_*(\mathbf{x}, \mathbf{v}_1), ..., f_*(\mathbf{x}, \mathbf{v}_k)).$$

The Pullback takes a k-form in $\mathbb{R}^m$ and pulls it back to give a new k-form in $\mathbb{R}^n$ that behaves how the original would based on the function f. It goes in the opposite direction of the function f.

Then next theorem gives some insight into how the Pullback behaves and for computing the integral we will define below!

Theorem 14.17. If f is a differentiable function as in the above definition, ω_1, ω_2 are k-forms in $\mathbb{R}^m$, and $g : \mathbb{R}^m \to \mathbb{R}$, then
(a) $f^* dx_i = D_1 f_i dx_1 + ... + D_n f_i dx_n$,
(b) $f^*(\omega_1 + \omega_2) = f^*\omega_1 + f^*\omega_2$,
(c) $f^*(g\omega_1) = (g \circ f) f^*\omega_1$,
(d) $f^*(\omega_1 \wedge \omega_2) = f^*\omega_1 \wedge f^*\omega_2$.

Finally, we can define the integral of a k-form over a k-chain!

Definition 14.18. For simplicity throughout the rest of the chapter we will assume all k-cells and all component functions of k-forms are at least one time differentiable. When additional assumptions are needed they will be stated.
We start in a very simple case. If $\omega = f dx_1 \wedge ... \wedge dx_k$ is a k-form in $\mathbb{R}^k$ then we define the integral of ω over the set $[0, 1]^k$ as

$$\int_{[0,1]^k} \omega = \int_{[0,1]^k} f \, d\mathbf{x},$$

where the integral on the right is the standard Lebesgue Integral of f in $\mathbb{R}^k$. Since there is only one elementary k-form in $\mathbb{R}^k$ this fully defines the integral for k-forms in $\mathbb{R}^k$ over $[0,1]^k$. This case isn't particularly useful as most commonly we won't have our k-cells living in $\mathbb{R}^k$, but rather they will lie in some larger space $\mathbb{R}^n$, and we would like to integrate over more general k-cells! But here is where the Pullback comes into play! For an arbitrary k-form we pull it back into $\mathbb{R}^k$ with our k-cell! The result will always end up in the first case allowing us to integrate. Explicitly, we extend our original definition to integrate an arbitrary k-cell, ω in $\mathbb{R}^n$ over a k-cell $c : \mathbb{R}^k \to \mathbb{R}^n$ by

$$\int_c \omega = \int_{[0,1]^k} c^* \omega.$$

There are a few special cases, namely we define a 0-cell, c is just a vector in $\mathbb{R}^n$ and a 0-form will be considered simply a function $f : \mathbb{R}^n \to \mathbb{R}$. Then the integral of a 0-form f over a 0-cell c is

$$\int_c f = f(c).$$

Definition 14.19. We expand the above definition to chains via linearity. If $C = \sum a_i c_i$ is a k-chain in $\mathbb{R}^n$ and ω is a k-form in $\mathbb{R}^n$ we define the integral of ω over C as

$$\int_C \omega = \sum a_i \int_{c_i} \omega.$$

Example 14.20. We will do an example to show how to calculate these integrals! Suppose we have the 1-cell $c : [0,1] \to \mathbb{R}^2$ and the 1-form in $\mathbb{R}^2$ $\omega = g(x,y)dx + h(x,y)dy$, then the pullback gives $c^*\omega = g(c(t))Dc_1(t)dx + h(c(t))Dc_2(t)dx$ so that

$$\int_c \omega = \int_0^1 g(c(t))Dc_1(t) + h(c(t))Dc_2(t)\, dx.$$

You may remember, this was the second type of line integral discussed in a typical Calculus

Three course, and this agrees with how it was defined then!

The next theorem will be used in the proof of Stokes' Theorem.

Theorem 14.21. Suppose ω is a k-form in $\mathbb{R}^n$ and c is a k-cell mapping to $\mathbb{R}^n$. Then

(a) We have

$$\int_c \omega = \int_{I^k} c^*\omega,$$

(b) We have

$$\int_{\partial c} \omega = \int_{\partial I^k} c^*\omega.$$

Proof. To start with (a) we instead suppose we have an arbitrary k-form η in $\mathbb{R}^k$, so that $\eta = f dx_1 \wedge ... \wedge dx_k$. Then we have

$$\int_{I^k} \eta = \int_{[0,1]^k} I^{k*}(f dx_1 \wedge ... \wedge dx_k) = \int_{[0,1]^k} f I^{k*}(dx_1) \wedge ... \wedge I^{k*}(dx_k)$$

$$= \int_{[0,1]^k} f dx_1 \wedge ... \wedge dx_k.$$

In particular we have the intuitively obvious fact that

$$\int_{I^k} \eta = \int_{[0,1]^k} \eta.$$

But now $c^*\omega$ is a k-form in $\mathbb{R}^k$ so that via the definition and the above case we have

$$\int_c \omega = \int_{[0,1]^k} c^*\omega = \int_{I^k} c^*\omega,$$

which completes (a).

Result (b) follows from repeat applications of (a). $\square$

To close out this section, we prove a theorem showing that in certain situations we can truly regard the integral of a k-form over a k-cell, as the integral of a function over the range of the cell. Put differently we can think of integrating over a cell as integrating over the range of the cell.

Theorem 14.22. Suppose c is a diffeomorphic k-cell into $\mathbb{R}^k$ with $\det Dc > 0$ for all $\mathbf{x}$ and $\omega = f dx_1 \wedge ... \wedge dx_k$ is a k-form. Then

$$\int_c \omega = \int_{c([0,1]^k)} f \, d\mathbf{x}.$$

Proof. If we look at the definitions we have

$$\int_c \omega = \int_{[0,1]^k} c^*(f dx_1 \wedge ... \wedge dx_k) = \int_{[0,1]^k} f(c)(c^* dx_1) \wedge ... \wedge (c^* dx_k)$$

$$= \int_{[0,1]^k} f(c)(\sum_{i=1}^{k} D_i f_1 dx_i) \wedge ... \wedge (\sum_{i=1}^{k} D_i f_k dx_i).$$

An induction argument gives the above equals

$$\int_{[0,1]^k} f(c(\mathbf{x})) \det Dc(\mathbf{x}) \, d\mathbf{x}.$$

The positivity of the Jacobian and Change of Variables Theorem complete the proof. □

14.6 The Exterior Derivative

We've successfully define the integral of a k-form over a k-cell, our next objective is the Generalized Stokes Theorem, and for that we will need another major operation regarding forms, The Exterior Derivative. We first start with 0-forms and differentials!

Definition 14.23. Suppose we have a function/0-form f. Recall the gradient ∇f. Just as how we defined the elementary 1-forms dx_i as a sort of directional derivative of the functions

π_i, we can do this for arbitrary functions f. We define the **Differential** of f as the 1-form $df(\mathbf{x})(\mathbf{x}, \mathbf{v}) = Df(\mathbf{x}, \mathbf{v})$. Essentially, it takes in a tangent vector $(\mathbf{x}, \mathbf{v})$ and gives the directional derivative of f at $\mathbf{x}$, in the direction of $\mathbf{v}$.
Using how we've defined $dx_i's$ we have $df = D_1 f dx_1 + ... + D_n f dx_n$.

We now extend this notion to general k-forms.

Definition 14.24. Let $\omega = \sum_I f_I dx_I$ be a k-form in $\mathbb{R}^n$. We define the **Exterior Derivative** of ω as the $k+1$ form defined by

$$d\omega = \sum_I (df_I) \wedge dx_I,$$

this is a sum of wedge products of 1 and k forms so it is easy to see it is indeed a $k+1$ form!

The following theorem shows how the exterior derivative interacts with wedge products and pullbacks!

Theorem 14.25. Suppose ω is a k-form (with $\mathcal{C}^2$ component functions), η is an l-form, and g is a differentiable function from $\mathbb{R}^n \to \mathbb{R}^m$. Then
(a) $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge \ d\eta$,
(b) $d(d(\omega)) = 0$, and
(c)$g^*(d\omega) = d(g^*\omega)$.

Proof. It's easy to see that the exterior derivative satisfies a linearity property.
For (a) the result holds for 0-forms, if we define $f \wedge g = fg$, via a modified Product Rule. It now sufficies to prove it in the simpler case $\omega = f dx_I$ and $\eta = g dx_J$, by linearity. We have

$$d(\omega \wedge \eta) = d(f dx_I \wedge g dx_J) = d(fg \wedge dx_{IJ}) = d(fg) \wedge dx_{IJ} = (gdf + fdg) \wedge dx_I \wedge dx_J$$

$$= gdf \wedge dx_I \wedge dx_J + fdg \wedge dx_I \wedge dx_J = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta,$$

which gives (a).

Result (b) is the most important result here. For a 0-form f, we have $df = \sum_{i=1}^n D_i f dx_i$ so that

$$d(df) = d\left(\sum_{i=1}^n D_i f dx_i\right) = \sum_{i=1}^n d(D_i f) dx_i = \sum_{i=1}^n \sum_{j=1}^n D_j(D_i f) dx_j \wedge dx_i$$

$$= \sum_{i,j} (D_i(D_j f) - D_j(D_i f)) dx_i \wedge dx_j.$$

The result for 0-forms then follows from Clairut's Theorem. More generally if $\omega = f dx_I$, then

$$d(d\omega) = d((df) \wedge dx_I) = d(df) \wedge dx_I - df \wedge d(dx_I).$$

We have $d(df) = 0$ by the 0-form case and $d(dx_I) = 0$ giving (b) via linearity.

For (c) given a 0-form f and g as above we have $g^* f = (f \circ g)$ by Theorem XXX (c) above so that

$$d(g^* f)(\mathbf{x})(\mathbf{x}, \mathbf{v}) = Df \circ g(\mathbf{x}) \cdot \mathbf{v} = Df(g(\mathbf{x})) Dg(\mathbf{x}) \cdot \mathbf{v} = Df(g(\mathbf{x})) Dg(\mathbf{x}, \mathbf{v})$$

$$= Df(g(\mathbf{x}), Dg(\mathbf{x}, \mathbf{v})) = df(g(\mathbf{x}))(g(\mathbf{x}), Dg(\mathbf{x}, \mathbf{v})) = df(g(\mathbf{x}))(g_*(\mathbf{x}, \mathbf{v})) = g^*(df).$$

This completes (c) for 0-forms. A computation shows that $d(g^* dx_i) = 0$. Finally, the more general case follows via induction, for if it holds for $k-1$-forms, then we can write $\omega = \eta \wedge dx_i$, where η is a $k-1$-form, so that

$$d(g^*\omega) = d(g^*(\eta \wedge dx_i)) = d(g^*\eta \wedge g^* dx_i) = d(g^*\eta) \wedge g^* + (-1)^{k-1} g^*\eta \wedge d(g^* dx_i)$$

$$= d(g^*\eta) \wedge g^* dx_i = g^*(d\eta) \wedge g^* dx_i = g^*(d\eta \wedge dx_i) = g^*(d\eta \wedge dx_i + (-1)^{k-1} \eta \wedge d(dx_i)) = g^*(d\omega),$$

which with linearity completes the proof. $\square$

14.7 The Generalized Stokes' Theorem

I've just mentioned that statement (b) in the above Theorem was the most important. It gives insight into a deep connection between chains and forms. For a k-chain C we have $\partial^2 C = \partial(\partial C) = 0$ and for a k-form ω we have $d^2\omega = d(d\omega) = 0$.

We also have established a link between chains and forms in the sense we can integrate a k-form over a k-chain.

The last sort of symmetry we have is that the boundary of a k-chain is a $k-1$-chain while the derivative of a k-form is a $k+1$-form.

One might begin to think more about this relation of chains and forms, and eventually one may come to the conclusion that taking the boundary and taking the derivative are sort of "opposites".

Just as in Calculus we found that integration and differentiation were opposites through the Fundamental Theorem of Calculus, we come to the more general conclusion that differentiation and boundaries are opposites via The Generalized Stokes' Theorem! This major result is the reward for all our difficult work in understanding Chains and Forms and it's proof is not too difficult as a result. It will involve little more than a computation and a few manipulations. It will also give an easy tool for understanding all of the classic results of Vector Calculus typically found in Calculus Three, which will close out this chapter, and section of the book around formalizing Multivariable Calculus!

Theorem 14.347 (The Generalized Stokes Theorem) Suppose ω is a $k-1$-form, with $\mathcal{C}^2$ component functions, in $\mathbb{R}^n$ and C is a k-chain in $\mathbb{R}^n$. Then

$$\int_{\partial C} \omega = \int_C d\omega.$$

Proof. The notation $dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k$ indicates that the dx_i term is excluded from the wedge product.

We start in the simple case that $\omega = f dx_I$ is a $k-1$-form on $[0,1]^k$ and C is just I^k.

We see only one dx_i will be excluded so $\omega = f dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k$.

We first look at $I^k_{j,a}$ sides, for a fixed j.

If we evaluate the integral we get

$$\int_{I^k_{j,a}} \omega = \int_{[0,1]^{k-1}} I^{k*}_{j,a}(f dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k) d\mathbf{x}$$

$$= \int_{[0,1]^{k-1}} f(I^k_{j,a})(\sum_{w=1}^{k-1} D_w(I^k_{j,a})_1 dx_w) \wedge ... \wedge \widehat{(\sum_{w=1}^{k-1} D_w(I^k_{j,a})_i dx_w)} \wedge ... \wedge (\sum_{w=1}^{k-1} D_w(I^k_{j,a})_k dx_w)\, d\mathbf{x}.$$

We see that any value of $D_w(I^k_{j,a})_p$ is either 0 or 1 from the definition of $I^k_{j,a}$. If the excluded term $i \neq j$ then we will have $\sum_{w=1}^{k-1} D_w(I^k_{j,a})_j dx_w = 0$ is included and so the whole integral equals 0.

If instead the excluded term $i = j$, then for each $p = 1, ..., k$ (excluding j) we have $\sum_{w=1}^{k-1} D_w(I^k_{j,a})_p dx_w = dx_p$, so the whole integral becomes

$$\int_{[0,1]^{k-1}} f(x_1, ..., x_{i-1}, a, x_i, ..., x_{k-1}) dx_1 \wedge ... \wedge dx_{k-1} = \int_{[0,1]^{k-1}} f(I^k_{i,a})\, d\mathbf{x}.$$

Now if we integrate along the boundary we get

$$\int_{\partial I^k} \omega = \sum_{j=1}^{k} \sum_{a=0}^{1} \int_{[0,1]^{k-1}} I^{k*}_{j,a}(f dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k)$$

$$= (-1)^i \int_{[0,1]^{k-1}} f(I^k_{i,0})\, d\mathbf{x} + (-1)^{i+1} \int_{[0,1]^{k-1}} f(I^k_{i,1})\, d\mathbf{x}.$$

Now we compute the other integral which is easier. We have

$$\int_{I^k} d(\omega) = \int_{I^k} (df) \wedge dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k$$

$$= \int_{I^k} (\sum_{j=1}^{k} D_j f dx_j) \wedge dx_1 \wedge ... \wedge \widehat{dx_i} \wedge ... \wedge dx_k.$$

All of the $D_j f dx_j$'s in the (df) term will cancel except the excluded dx_i term so that with Theorem XXX we have the above equals

$$\int_{I^k} (-1)^{i-1} D_i f dx_1 \wedge ... \wedge dx_k = (-1)^{i-1} \int_{[0,1]^k} (D_i f)\, dx_1 dx_2 ... dx_k.$$

We now use Fubini's Theorem and FTC to get

$$(-1)^{i-1} \int_{[0,1]^k} (D_i f)\, dx_1 dx_2 ... dx_k = (-1)^{i-1} \int_{[0,1]^{k-1}} \left[\int_0^1 (D_i f) dx_i \right] d\mathbf{x}$$

$$= (-1)^{i-1} \int_{[0,1]^{k-1}} f(x_1, ..., x_{i-1}, 1, ..., x_{k-1}) - f(x_1, ..., x_{i-1}, 0, ..., x_{k-1})\, d\mathbf{x}$$

$$= (-1)^{i-1} \int_{[0,1]^{k-1}} f(I^k_{i,1})\, d\mathbf{x} + (-1)^i \int_{[0,1]^{k-1}} f(I^k_{i,0})\, d\mathbf{x}.$$

We finally see that both integrals evaluated to the same thing completing this special case! Now we have for an arbitrary k-cell c and a k-form $\omega = f dx_I$ by Theorem XXX that

$$\int_c d\omega = \int_{I^k} c^*(d\omega) = \int_{I^k} d(c^*\omega)$$

which is now in the special case as above so we have this is equal to

$$\int_{\partial I^k} c^*\omega = \int_{\partial c} \omega.$$

Finally, by the addivitiy of the pullback and definition of k-chains, this result immediately extends to arbitrary k-forms and k-chains completing the proof. $\square$

14.8 Vector Calculus

After proving the Generalized Stokes' Theorem this section will essentially be a list of special cases giving the classic Calculus Three Vector Calculus results.

First, we will give some common notation that agrees with the notation likely encountered in Calculus Three.

Rather than using the basis $x_1, ..., x_n$, because we will likely be working in $\mathbb{R}^3$ or $\mathbb{R}^2$, we will use x, y, and z.

We will call classic Lebesgue Integrals in $\mathbb{R}^2$ Double Integrals and similarly for Triple Integrals, and will use dA and dV.

With each vector valued function $F : \mathbb{R}^3 \to \mathbb{R}^3$ we can write

$F(x, y, z) = F_1(x, y, z)\hat{\mathbf{x}} + F_2(x, y, z)\hat{\mathbf{y}} + F_4(x, y, z)\hat{\mathbf{z}} = F = F_1\hat{\mathbf{x}} + F_2\hat{\mathbf{y}} + F_3\hat{\mathbf{z}}$.

We can think of these as Vector Fields as in the way described earlier in the chapter.

Of course similarly for function $G : \mathbb{R}^2 \to \mathbb{R}^2$ we can write $G = G_1\hat{\mathbf{x}} + G_2\hat{\mathbf{y}}$.

Definition 14.27. We define for a function F above the **1-form associated with F** the 1-form

$$\omega_1(F) = F_1 dx + F_2 dy + F_3 dz.$$

Similarly, G has a 1-form associated with it namely

$$\omega_1(G) = G_1 dx + G_2 dy.$$

The **2-form associated with F** is the 2-form

$$\omega_2(F) = F_1(dy \wedge dz) + F_2(dz \wedge dx) + F_3(dx \wedge dy).$$

Definition 14.28. We define the **Divergence of F**, where F is as above, as a function

$\nabla \cdot F : \mathbb{R}^3 \to \mathbb{R}$ by

$$\nabla \cdot F = D_1F_1 + D_2F_2 + D_3F_3,$$

and the **Curl of F** as the function $\nabla \times F : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$\nabla \times F = (D_2F_3 - D_3F_2)\hat{\mathbf{x}} + (D_3F_1 - D_1F_3)\hat{\mathbf{y}} + (D_1F_2 - D_2F_1)\hat{\mathbf{z}}.$$

We can also associate forms with these objects. The curl already has a 2-form associated with it by above, we will denote this form by λ_{CF}. We associate a 3-form with the Divergence

$$\lambda_{DF} = (D_1F_1 + D_2F_2 + D_3F_3)dx \wedge dy \wedge dz.$$

Finally, the last form association we need is the easiest. The gradient of a function $f : \mathbb{R}^3 \to \mathbb{R}$ was simply its total derivative denoted ∇f, and the 1-form associated with the gradient is the differential df.

Theorem 14.29. Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is as above and $F : \mathbb{R}^3 \to \mathbb{R}^3$ is as above. Then
(a) $d(\omega_1(f)) = \lambda_{CF}$,
(b) $d(\omega_2(f)) = \lambda_{DF}$.
We also have if G is as above, then $d(\omega_1(G)) = (D_1G_2 - D_2G_1)dx \wedge dy$.

Proof. This theorem gives a great picture on how general the exterior derivative is, it can act as both Divergence, Curl, and as we saw above the Gradient of a function, as long as we choose the right form to associate with the function. All three claims here as purely computations.
For (a) we have

$$d(\omega_1(F)) = (dF_1) \wedge dx + (dF_2) \wedge dy + (dF_3) \wedge dz = (D_1F_1dx + D_2F_1dy + D_3F_1dz) \wedge dx+$$

$$(D_1F_2dx + D_2F_2dy + D_3F_2dz) \wedge dy + (D_1F_3dx + D_2F_3dy + D_3F_3dz) \wedge dz$$

$$= (D_2F_1 dy\wedge dx + D_3F_1 dz\wedge dx) + (D_1F_2 dx\wedge dy + D_3F_2 dy\wedge dz) + (D_1F_3 dx\wedge dz + D_2F_3 dy\wedge dz)$$

$$= (D_2F_3 - D_3F_2)dy\wedge dz + (D_3F_1 - D_1F_3)dz\wedge dx + (D_1F_2 - D_2F_1)dx\wedge dy$$

$$= \omega_2(\nabla\times F) = \lambda_{CF}.$$

For (b) we have

$$d(\omega_2(F)) = (dF_1)\wedge dy\wedge dz + (dF_2)\wedge dz\wedge dx + (dF_3)\wedge dx\wedge dy$$

$$= (D_1F_1 dx + D_2F_1 dy + D_3F_1 dz)\wedge dy\wedge dz + (D_1F_2 dx + D_2F_2 dy + D_3F_2 dz)\wedge dz\wedge dx$$

$$+(D_1F_3 dx + D_2F_3 dy + D_3F_3 dz)\wedge dx\wedge dy = (D_1F_1 + D_2F_2 + D_3F_3)dx\wedge dy\wedge dz$$

$$= \lambda_{DF}.$$

Finally, for (c) we have

$$d(\omega_1(G)) = (D_1G_1 dx + D_2G_1 dy)\wedge dx + (D_1G_2 dx + D_2G_2 dy)\wedge dy$$

$$= (D_1G_2 - D_2G_1)dx\wedge dy,$$

completing the proof. □

The next theorem is an easy Corollary!

Theorem 14.30. We have for a function $f:\mathbb{R}^3\to\mathbb{R}$ and $F:\mathbb{R}^3\to\mathbb{R}^3$ that

(a) $\nabla\times(\nabla f)) = 0$ and

(b) $\nabla\cdot(\nabla\times F) = 0$.

Proof. This is easy by looking at associated forms. We have $\lambda_{C\nabla f} = d(\omega_1(\nabla f)) = d(df) = 0$ giving (a).

Then $\lambda_{D\nabla\times F} = d(\lambda_{CF}) = d(d\omega_2(F)) = 0$ giving (b). □

Before moving to the classic results we will also delve a bit deeper into Surface and Line Integrals.

Definition 14.31. Given a vector valued function such as F above and a curve (1-form) c mapping into $\mathbb{R}^3$ we define the **Line Integral** of F over c as

$$\int_c F \cdot d\mathbf{r} = \int_c \omega_1(F).$$

A similar definition can be made for functions like G above with curves in $\mathbb{R}^2$ and the definition is extended to 1-chains via linearity.

If we instead have a surface (2-form) D mapping into $\mathbb{R}^3$ we define the **Surface Integral** of F over D as

$$\iint_D F \cdot dS = \int_D \omega_2(F).$$

Now we can present the classic results! For all of the results we will assume the functions we work with are of class $\mathcal{C}^1$ and all curve's and surface's are diffeomorphic. We will also abuse notation a bit and refer to the image of a surface say $D([0,1]^2)$ simply as D.

Theorem 14.353 (Green's Theorem) Suppose D is a surface in $\mathbb{R}^2$ and $G = G_1\hat{\mathbf{x}} + G_2\hat{\mathbf{y}}$. Then

$$\iint_D (D_1G_2 - D_2G_1)\, dA = \int_{\partial D} G_1 dx + G_2 dy.$$

Proof. The result follows easily from Theorem XXX, Theorem XXX (c), and Stokes' Theorem as

$$\int_{D([0,1]^2)} (D_1G_2 - D_2G_1)\, dA = \int_D (D_1G_2 - D_2G_1)\, dx \wedge dy = \int_{\partial D} G_1 dx + G_2 dy.$$

□

Theorem 14.354 (Fundamental Theorem of Line Integrals) Suppose c is a curve in

$\mathbb{R}^3$, $c(1) = \mathbf{p}$ and $c(0) = \mathbf{q}$, and $f : \mathbb{R}^3 \to \mathbb{R}$. Then

$$\int_c \nabla f \cdot d\mathbf{r} = f(\mathbf{p}) - f(\mathbf{q}).$$

Proof. We have $\partial c = c(1) - c(0)$ so that by Stokes' Theorem

$$\int_c \nabla f \cdot d\mathbf{r} = \int_c df = \int_{\partial c} f = f(\mathbf{p}) - f(\mathbf{q}).$$

□

Theorem 14.355 (Classic Stokes' Theorem) Suppose D is a surface in $\mathbb{R}^3$ and $F : \mathbb{R}^3 \to \mathbb{R}^3$. Then

$$\iint_D \nabla \times F \cdot dS = \int_{\partial D} F \cdot d\mathbf{r}.$$

Proof. Using theorem XXX (a) and The Generalized Stokes' Theorem we get

$$\iint_D \nabla \times F \cdot dS = \int_D \lambda_{CF} = \int_D d(\omega_1(F)) = \int_{\partial D} \omega_1(F) = \int_{\partial D} F \cdot d\mathbf{r}.$$

□

Theorem 14.356 (Divergence Theorem) Suppose E is a 3-cell mapping into $\mathbb{R}^3$ and $F : \mathbb{R}^3 \to \mathbb{R}^3$. Then

$$\iiint_E \nabla \cdot F dV = \iint_{\partial E} F \cdot dS.$$

Proof. Using theorem XXX, theorem xxx (b), and Stokes' Theorem we get

$$\iiint_E \nabla \cdot F dV = \int_E \lambda_{DF} = \int_E d(\omega_2(F)) = \int_{\partial E} \omega_2(F) = \iint_{\partial E} F \cdot dS.$$

□

Chapter 15

Further Topology Major Results

These next two chapters will not be focusing on Analysis, but rather on Topology! A brief recap of Chapter 2 may be helpful. The goal is to delve deeper into Topology and prove some of the major results of the field. This first chapter will focus on some major results namely Urysohn's Lemma, Tychnonoff's Theorem, and The Baire Category Theorem. The rest of the chapter will ultimately be dedicated to building up the machinery for these results, as well as important consequences. Next chapter we will prove more important results, more focused on the Topology of $\mathbb{R}^n$ with the highlights being Brouwer's Fixed Point Theorem and The Jordan Curve Theorem.

15.1 The Axiom of Choice

Throughout this book, we've typically not paid any attention to the underlying axioms that govern set theory, and while this will largely remain, one axiom in particular is worth mentioning for its prevalance in Topology. That axiom of course is the **Axiom of Choice**. Informally, it says that if we have a collection of set we can make a "choice function" that picks one element from each set. We will now formally state the axiom.

Axiom of Choice: Suppose $\mathcal{F}$ is a collection of nonempty sets. Then there exists a function

$f : \mathcal{F} \to \bigcup \mathcal{F}$, satisfying that for $A \in \mathcal{F}$ we have $f(A) \in A$.

The Axiom of Choice at first glance seems quite inconsequential, so what we can choose something from each set. But the power comes from the fact that this holds for any arbitrary collection of sets, finite, infinite, uncountable, it doesn't matter we can always make a choice function. This leads to some very bizarre results. Below we give one of these, The Well-Ordering Theorem, and another important result Zorn's Lemma. It turns out that all three of these are equivalent, and typically the proofs of these claims are given in an early class on Set Theory.

Theorem 15.357 (Zorn's Lemma) If $(P, \leq)$ is a partially ordered set so that every totally ordered subset has an upper bound, then P has a maximal element.

Theorem 15.358 (The Well-Ordering Theorem) For every set X, there exists an order $\leq$, so that $(X, \leq)$ is well-ordered.

15.2 The Baire Category Theorem

We now turn back to our discussion on Topology starting with perhaps the simplest of the Three Main Theorems listed above, The Baire Category Theorem. This theorem has uses all throughout Functional Analysis, which we will explore in those chapters. For that reason

Definition 15.3. Suppose X is a topological space. We defined back in Chapter 2 for some set $A \subseteq X$ the interior of A, denoted A°, as the set of all interior points in A and we proved it was the largest open set contained in A. If $A^\circ = \emptyset$ then we say A has **Empty Interior**. If $\overline{A}$ has an empty interior then we say that A is **Nowhere Dense**.

Finally, a set E is called **Meagre** or **of the First Category** if there exists a countable collection of sets $A_1, \ldots$ with each A_n being nowhere dense and

$$E = \bigcup_{n=1}^{\infty} \overline{A_n}.$$

A set that is not meagre is called **Nonmeagre** or **of the Second Category**. The category naming system, which Baire used, explains why the theorem is called The Baire Category Theorem.

Nowhere Dense sets can be thought of as small, they intuitively contain very little space, and a so a meagre set as a countable union of small spaces is also small. This is not always the case, but if it is we call our space a Baire Space.

Definition 15.4. More formally, a **Baire Space** X is a Topological Space so that every meagre set $E \subseteq X$ is nowhere dense. In a Baire Space combining countably many nowhere dense sets still leaves you with an empty interior.
Using DeMorgan's Laws and our rules regarding open and closed sets this is equivalent to the following: X is a Baire Space if given a collection of open dense setes, $\{A_n\}$, then the intersection $\bigcap_{n=1}^{\infty} A_n$ is dense.

Definition 15.5. We will need one last definition for the theorem. We define the **Diameter** of a set E in a metric space (X, d) as $\text{diam}(E) = \inf\{d(x, y) | x, y \in E\}$.

We can now state and prove the Baire Category theorem.

Theorem 15.362 (The Baire Category Theorem) Every complete metric space is a Baire Space.

Proof. Let (X, d) be our complete metric space. Suppose $A_1, A_2, ...$ is a countable collection of nowhere dense sets and $A = \bigcup_{n=1}^{\infty} A_n$. Our goal is to show that A is nowhere dense, i.e. $A^\circ = \emptyset$. This is done if we can show that every nonempty open set contains a point not in A°, so suppose U is an arbitrary open set. We have $U \not\subseteq A_1$, so there is some point $p_1 \in U$ with $p_1 \notin A_1$. Since A_1 is closed, there is some $r > 0$ so that $\overline{B_r(p_1)} \cap A_1 = \emptyset$, and since U is open $\overline{B_r(p_1)} \subseteq U$ if we choose r small enough, also ensure $r < 1$ if not make it smaller! Label $B_r(p_1) = U_1$. By induction we create a sequence of open balls U_n with $\overline{U_{n+1}} \subseteq U_n$, $\overline{U_n} \cap A_n = \emptyset$, and $\text{diam}(U_n) < \frac{1}{n}$. Finally, the intersection $\bigcap \overline{U_n}$ is nonempty. This can be seen by making a sequence, $\{x_n\}$, where each $x_n \in U_n$. This sequence is Cauchy and so converges by completeness, and this limit must lie in the intersection. This limit point is not in A, but is in U finishing the proof. $\square$

15.3 Separation Axioms

We now turn our attention to the second major theorem: Urysohn's Lemma. However, before we can even discuss the theorem we will need to discuss some of the Separation Axioms. Back in Chapter 2 we gave the most important of these, the Hausdorff Axiom, but now we will introduce more of these axioms we can impose on our spaces. While there are axioms weaker than being Hausdorff, we will ignore these in this section as we are more interested in stronger requirements! Now we can introduce some new axioms!

Definition 15.7. A topological space X is **Regular** if one point sets are closed and if given a point x and a closed set C, with $x \notin C$, then there exist disjoint open sets U and V with $x \in U$ and $C \subseteq V$. This means we can separate with open sets a singleton and a closed set.

This clearly implies being Hausdorff, hence why we included the one point sets are closed criterion. We can strengthen this axiom as follows.

Definition 15.8. A space X is **Completely Regular** if one point sets are closed and if for every closed set C and point x ($x \notin C$), there exists a continuous function $f : X \to [0, 1]$ with $f(x) = 1$ and $f(C) = \{0\}$. The function separates the point and the closed set.

Again, it's easy to see that completely regular implies regular, take $f^{-1}([0, \frac{1}{2}))$ and $f^{-1}((\frac{1}{2}, 1])$. The next properties are the Normal Properties!

Definition 15.9. A space X is **Normal** if one point sets are closed and if given two disjoint closed sets C, D there exist disjoint neighborhoods U, V with $C \subseteq U$ and $D \subseteq V$. Essentially, we can separate disjoint closed sets with neighborhoods. Similarly, X is **C-Normal** if one point sets are closed and if for every closed sets C, D there exists a continuous function $f : X \to [0,1]$ with $f^{-1}(C) = \{0\}$ and $f^{-1}(D) = \{1\}$, we can separate the closed sets with a continuous function.

We immediately have that normal implies regular, C-Normal implies normal, and C-Normal implies completely regular. What we would like to say is that normal implies completely regular to turn this into a true chain of implications. But we will do something much stronger. We will prove Urysohn's Lemma which will tell us that C-Normal is actually equivalent to Normal! After Urysohn's Lemma we can then drop this needless distinction. Before that we will give two theorems giving us a class of normal/ "C-Normal" spaces. The following Lemma will be useful below and in the proof or Urysohns Lemma.

Lemma 15.10. If X is a regular space with U open and $x \in U$, then there exists an open set V such that $x \in V$ and $\overline{V} \subseteq U$.
Similarly, if Y is normal C closed and $C \subseteq W$, then there exists an open set E with $C \subseteq E$ and $\overline{E} \subseteq W$.

Proof. We prove the regular case, the normal one is almost identical. Let $D = U^c$, then by regularity we have disjoint open sets V containing x and B containing D. If $y \in D$, then B is a neighborhood of y disjoint from V, or in other words $\overline{V} \cap D = \emptyset$ which implies that $\overline{V} \subseteq U$. □

Theorem 15.11. Compact Hausdorff spaces are normal.

Proof. Let X be our space and suppose C, D are closed disjoint subsets. Fix $x \in C$. For each $y \in D$ by the Hausdorff property there are disjoint open sets $U_{x,y}$ and $V_{x,y}$ with $x \in U_{x,y}$ and $y \in V_{x,y}$. The collection of sets $V_{x,y}$ form an open cover of D, so by compactness there is a finite sub cover, $V_{x,y_1}, ..., V_{x,y_n}$. Let $U_x = U_{x,y_1} \cap ... \cap U_{x,y_n}$ and let $V_x = V_{x,y_1} \cup ... \cup V_{x,y_n}$. Then $U_x \cap V_x = \emptyset$ so we have separated x from D, i.e. we have proven regularity.
Now to extend to normality, for each $x \in C$ we have open disjoint sets U_x, V_x with $x \in U_x$

and $D \subseteq V_x$. The collection of all U_x forms an open cover for C, so by compactness we have a finite subcover $U_{x_1}, .., U_{x_n}$. Then $U = U_{x_1} \cup ... \cup U_{x_n}$ is an open set containing C, $V = V_{x_1} \cap ... \cap V_{x_n}$ is an open set containing D, and $U \cap V = \emptyset$ completing the proof. $\square$

Corollary 15.11.1. Every Compact Hausdorff space is a Baire Space. Follow the same proof as metric spaces, but use regularity to separate the sets $\{p_n\}$ and A_n, then the Finite Intersection Property guarantees the intersection $\bigcap E_n \neq \emptyset$.

Theorem 15.12. Second countable regular spaces are normal.

Proof. Let X be a secound countable regular space, and let C, D be disjoint closed subsets. If $x \in C$ then D^c is an open set containing x so Lemma XXX gives an open set U_x so that $x \in U_x$ and $\overline{U_x} \subseteq D^c$, i.e. $\overline{U_x} \cap D = \emptyset$. We then find a basis element B_x so that $x \in B_x$ and $B_x \subseteq U_x$. Doing this for each x, we can find a countable collection of basis elements $B_1, B_2, ...$ covering C with each $\overline{B_i} \cap D = \emptyset$.

Do the same thing for D and let $E_1, E_2, ...$ be our countable collection of basis sets covering D all with closures disjoint from C.

Finally, we make these all disjoint by setting

$$B'_n = B_n - \bigcup_{i=1}^{n} \overline{E_i} \quad \text{and} \quad E'_n = E_n - \bigcup_{i=1}^{n} \overline{B_n}.$$

Then the sets

$$B = \bigcup_{n=1}^{\infty} B'_n \quad \text{and} \quad E = \bigcup_{n=1}^{\infty} E'_n,$$

are disjoint open sets containing C and D respectively. $\square$

Theorem 15.13. Metric Spaces are "C-Normal".

Proof. Let C, D be disjoint closed sets. For a set $E \subseteq X$ define a function $f_E : X \to \mathbb{R}$ by

$f_D(x) = \inf_{y \in D} d(x, y)$, and then a new function $G : X \to \mathbb{R}$ by

$$G(x) = \frac{f_C(x)}{f_C(x) + f_D(x)}.$$

Then f_E is continuous and $f_E(x) = 0$ iff $x \in E$, so that G is continuous, $G(C) = \{0\}$, and $G(D) = \{1\}$, so we see G satisfies our requirements. $\square$

15.4 Urysohn's Lemma

Now we turn to Urysohn's Lemma, one of the most important theorems in Topology. As mentioned previously, it essentially tells us our distinction between normal and "C-Normal" spaces is not really there, they are both the same type of space.

Theorem 15.370 (Urysohn's Lemma) A space X is normal if and only if one point sets are closed and if given disjoint closed sets C, D there exists a continuous function $f : X \to [0, 1]$ with $f(C) = \{0\}$ and $f(D) = \{1\}$.

Proof. If such a function exists, then the space is "C-Normal" which as claimed earlier implies normal, this direction is uninteresting.

The other direction is far more exciting, so suppose X is normal and C, D are disjoint closed sets. We cleverly construct a continuous function separating C and D. Let $p_0, p_1, p_2, ...$ be an enumeration of the rational numbers in $[0, 1]$, with $p_0 = 0$ and $p_1 = 1$, this exists by countability of the rationals. Let $U_1 = D^c$, which is open. Lemma XXX above guarantees an open set U_0 satisfying $C \subseteq U_0$ and $\overline{U_0} \subseteq U_1$.

Now we use induction, suppose that for all $p_0, p_1, ..., p_n$ we have a corresponding set U_n with the property that $p_i < p_j$ implies that $U_{p_i} \subseteq U_{p_j}$. We want to construct $U_{p_{n+1}}$. Let $P_n = \{p_1, ..., p_n\}$, and let $p_i = \max(P \cap [0, p_{n+1}])$ and $p_j = \min(P \cap [p_{n+1}, 1])$, these exist as we are working with finite sets. We then have $\overline{U_{p_i}} \subseteq U_{p_j}$ so Lemma XXX again gives an open set which we call $U_{p_{n+1}}$ with $\overline{U_{p_i}} \subseteq U_{p_{n+1}}$ and $\overline{U_{p_{n+1}}} \subseteq U_{p_j}$. By induction, we do this for all $n = 1, 2, 3, ...$ Finally, if p is a rational number with $p > 1$ we define $U_p = X$, and if p is rational with $p < 0$ define $U_p = \emptyset$.

Now for each $x \in X$ we define a function $f(x) = \inf\{p | x \in U_p\}$. Then if $x \in C$ we have $x \in U_p$ for all p so that $f(x) = 0$. If $y \in D$ then for all $p > 1$ $y \in U_p$, but for any $q < 1$ $y \notin U_q$, so that $f(y) = 1$. We only need to prove continuity to complete the proof. Let $x \in X$, set $f(x) = r$, and let $B_\epsilon(r)$ be a neighborhood of r. By the density of $\mathbb{Q}$ there exists rational numbers p_i, p_j with $r - \epsilon < p_i < r < p_j < r + \epsilon$. It follows that $U = U_{p_j} - \overline{U_{p_i}}$ is a neighborhood containing x. We also see that if $y \in U$, then $y \in U_{p_j}$ so that $f(y) \leq p_j$, but $y \notin \overline{U_{p_i}}$ so that $f(y) \geq p_i$. Therefore, $f(U) \subseteq B_\epsilon(r)$ and so f is continuous at x which completes the proof. □

15.5 Consequences of Urysohn's Lemma

We now present some important consequences of Urysohn's Lemma. We see now that we have the following string of implications:

Normal $\Longrightarrow$ Completely Regular $\Longrightarrow$ Regular $\Longrightarrow$ Hausdorff.

We also see that Theorem XXX immediately yields the fact that every metric space is normal. Perhaps surprising, another consequence of Urysohn's Lemma is an extension of the Stone-Weierstrass Theorem. We can extend the theorem to Compact Hausdorff Spaces now, as in the first direction we needed the distance function to separate points. But now with Urysohn's Lemma and the fact that compact Hausdorff spaces are normal, we have a new tool to separate points!

The next theorem is a bigger consequence of Urysohn's Lemma called the Tietze Extension Theorem whcih we now prove!

Theorem 15.371 (Tietze Extension Theorem) Suppose X is a normal space, C is a closed subset of X, and $f : C \to \mathbb{R}$ is continuous. Then we can extend f to all of X and if $|f(A)|$ is bounded, then the extended function can be required to be bounded as well by the same bound.

Proof. We work in the case f is bounded, and we can assume that $|f(x)| \leq 1$. That is $f(A) \subseteq [-1, 1]$. Then the sets $D = f^{-1}([-1, -\frac{1}{3}])$ and $E = f^{-1}([\frac{1}{3}, 1])$ are disjoint closed sets so by Urysohn's Lemma there is a continuous function $g_1 : X \to [-\frac{1}{3}, \frac{1}{3}]$ with $g_1(D) = \{-\frac{1}{3}\}$ and $g(E) = \{\frac{1}{3}\}$. The function g_1 is bounded with $|g_1(x)| \leq \frac{1}{3}$ for all x, and for all x we have $|g_1(x) - f(x)| \leq \frac{2}{3}$ due to how sets C and D were constructed.
Now consider the function $(f - g_1) : C \to [-\frac{2}{3}, \frac{2}{3}]$, we can repeat the process we just did and find a function $g_2 : X \to \mathbb{R}$ with $|g_2(x)| \leq \frac{2}{9}$ and $|g_2(x) - (f - g_1)| \leq (\frac{2}{3})^2$.
In general we can inductively repeat this process so that g_n is a function with $|g_n(x)| \leq \frac{1}{3}(\frac{2}{3})^{n-1}$ an $|g_n(x) - (f - g_1 - ... - g_{n-1})(x)| \leq (\frac{2}{3})^n$.
Finally, set $g(x) = \sum_{n=1}^{\infty} g_n(x)$, the Weierstrass M-Test tells us this series converges uniformly on X, and is therefore continuous. It's also easy to see that for all $x \in A$ we have $g(x) = f(x)$, so that g is our extended function completing this case!
The case of an unbounded function is similar. If $f(A) \subseteq (-\infty, \infty)$, we compose with a homeomorphism to $(-1, 1)$ and then essentially repeat all of our work above, then at the end compose with the inverse of the homeomorphism we chose. An example homeomorphism would be $\frac{2\tan^{-1}(x)}{\pi}$. □

Finally, our last application will be a tool to tell when a topological space is "actually" a metric space. We make this more precise.

Definition 15.16. An **Embedding** of a topological space X into a topological space Y is a function $f : X \to Y$ so that f is a homeomorphism of X and $f(X)$, where $f(X)$ is treated as a subspace of Y. If there exists an embedding of X into Y, we say that f can be embedded in Y, and can essentially treat X as a subspace.
A topological space (X, τ) is **Metrizable** if there exists a metric $d : X \times X \to \mathbb{R}$ so that (X, d) is equivalent to (X, τ), i.e. there is a metric that gives us the same topology. We

certainly have then that X is metrizable if and only if X can be embedded in some metric space.

This answers our question of what we mean by X is "actually" a metric space, we mean it is metrizable. Since metric spaces are very nice, it is obviously advantageous to know if a space is metrizable. Below we state Urysohn's Metrization Theorem, which is what he actually used his Lemma above to prove, which gives a sufficient condition to deduce metrizability. There are stronger metrization theorems that give necessary and sufficient conditions, but we are primarily giving this theorem as a consequence of Urysohn's Lemma.

Theorem 15.373 (Urysohn's Metrization Theorem) Every second countable regular space is metrizable.
The converse is obviously false, a metric space need not be secound countable.

The proof will be delayed until after we've discussed the product topology and will conclude the next section.

15.6 Product Topology Revisited

We now want to generalize the Product Topology to products of infinitely many spaces. To do this we will obviously need a more general notion of Cartesian Products.

Definition 15.18. Suppose I is an index set with $\{X_i\}_{i\in I}$ a collection of sets. Then the **Cartesian Product** of the sets $\{X_i\}$ is the set

$$\prod_{i\in I} X_i = \left\{ f : I \to \bigcup_{i\in I} X_i \middle| f(i) \in X_i \right\}.$$

The functions in the Cartesian Product above are called **i-Tuples** and given a $j \in I$ and $f \in \prod X_i$ the **j-th Coordinate of** f is $f_j = f(j)$. Finally, the function $\pi_j : \prod X_i \to X_j$ by $\pi_j(f) = f_j$ is called the **j-th Projection Function.**

All of these definitions are a bit weirder than our finite case, but they all generalize the definitions we are familiar with. Also, a quick glance at the definitions should tell you that $\prod X_i$ is nonempty due to the Axiom of Choice, as long as all X_i are nonempty.

Now we can define the Product Topology.

Definition 15.19. Let $\{X_i\}$ be a collection of topological spaces indexed by some set I. Then the **Product Topology** on $\prod_{i \in I} X_i$ is the smallest topology so that every projection function π_j is continuous. Such a topology exist (take the intersection of all topologies satisfying this, the discrete topology is always one). Equivalently, the product topology is the topology formed by basis elements

$$\prod U_i,$$

where $U_i \subseteq X_i$ is open and for all but finitely many i we have $U_i = X_i$.

One major theorem regarding the product topology is the following.

Theorem 15.20. Suppose Y is a space and $f : Y \to \prod X_i$. The **i-th Component Function of** f is the function $f_i(a) : Y \to X_i$ by $f_i(y) = \pi_i(f(y))$, so we have $f(y) = (f_i(y))_{i \in I}$. Then f is continuous if and only if each component function is continuous.

Proof. Suppose f is continuous, then $f_i = \pi_i \circ f$ is the composition of continuous functions and is therefore continuous.

For the other direction suppose each f_i is continuous and let $U = \prod U_i$ be a basis element so that for all but finitely many $i \in I$ we have $U_i = X_i$, say $i_1, i_2, ..., i_n$, and for the rest U_i is open in X_i. Then

$$f^{-1}(U) = f^{-1}\left(\prod_{i \in I} U_i\right) = \bigcap_{i \in I} f_i^{-1}(U_i) = \bigcap_{k=1}^{n} f_{i_k}^{-1}(U_{i_k}).$$

As the intersection of finitely many open sets, $f^{-1}(U)$ is therefore open, and so f is continuous. $\square$

An important space we will see soon is the space $\mathbb{R}^{\mathbb{N}} = \prod_{n=1}^{\infty} \mathbb{R}$ with the product topology and similarly $[0,1]^{\mathbb{N}}$. Each element of $\mathbb{R}^{\mathbb{N}}$ is a sequence of real numbers! We will typically write elements in the form $(x_1, x_2, ...)$ rather than $\{x_n\}$ typical sequence form. We can of course extend this and define $\mathbb{R}^I = \prod_{i \in I} \mathbb{R}$, where I can be finite, countable, or uncountable. Perhaps suprisingly, it turns out that $\mathbb{R}^{\mathbb{N}}$ is metrizable! In fact if we set $d_1 : \mathbb{R}^2 \to \mathbb{R}$ by $d_1(x, y) = \min\{|x - y|, 1\}$ and then define $d : \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by

$$d(\mathbf{x}, \mathbf{y}) = \sup\{\frac{d(x_n, y_n)}{n} | n = 1, 2, 3, ...\},$$

this metric agrees with the product topology of $\mathbb{R}^{\mathbb{N}}$! One way to prove this would be to show that all projection maps are continuous and show that every open set induced by this metric is open in the product topology. This gives the result as the product topology is the smallest topology satisfying the first claim.

The next theorem will be a key part of the Urysohn Metrization Theorem.

Theorem 15.21. Let X be a space so that one point sets are closed and suppose $\{f_i\}_{i \in I}$ is a collection of continuous functions, $f_i : X \to \mathbb{R}$, such that for each $x \in X$ and for each neighborhood $N(x)$ there exists a $j \in I$ so that $f_j(x) > 0$ and $f(y) = 0$ for all $y \in U^c$. Then the function $F(x) = (f_i)_{i \in I}$ embeds X in $\mathbb{R}^I$.

Proof. By Theorem XXX above F is continuous. Let $U \subseteq X$ be open. If we prove that $F(U)$ is open in $F(X)$ then we are done. Let $y \in F(U)$ and choose $x \in X$ so that $F(x) = y$. By assumption we choose j so that $f_j(x) > 0$ and $f_j(U^c) = \{0\}$. For this same index we set $W = F(X) \cap \pi_j^{-1}((0, \infty))$ which is open in $F(X)$. We have $f_j(x) = \pi_j(y) > 0$ so $y \in W$. If $p \in W$, then $\pi_j(p) > 0$ so that $f^{-1}(p) \in U$ or $W \subseteq F(U)$. Therefore, $W = F(U) \cap W$ is an open set in the subspace topology so F is an embedding. $\square$

Now we can prove the Urysohn Metrization Theorem!

Proof of Theorem XXX. Let X be a second countable regular space. Our goal is to embed X into $\mathbb{R}^{\mathbb{N}}$ with the use of the above theorem. Since we proved $\mathbb{R}^{\mathbb{N}}$ was metrizable this gives the result.

Let $B_1, B_2, ...$ be our countable basis for X. Let E be the subset of $\mathbb{Z}^2$ with the property $(m, n) \in E$ implies that $\overline{B_m} \cap B_n^c = \emptyset$. By Theorem XXX X is normal, so that Urysohn's Theorem gives us a function corresponding to each $(m, n) \in E$, say $g_{m,n} : X \to \mathbb{R}$, so that $g_{m,n}(\overline{B_m}) = \{1\}$ and $g_{m,n}(B_n^c) = \{0\}$.

Given $x \in X$ and a neighborhood $N(x)$, we have a basis element B_k with $x \in B_k \subseteq N(x)$, so that another application of regularity with Lemma XXX gives another basis element B_w with $x \in B_w$ and $\overline{B_w} \subseteq B_k$ or equivalently $\overline{B_w} \cap B_k^c = \emptyset$. Therefore, for this x and $N(x)$ the function $g_{w,k}$ satisfies $g_{w,k}(x) = 1 > 0$ and $g_{w,k}(N(x)^c) = \{0\}$. Therefore the collection of functions $g_{m,n}$ satisfy Theorem XXX above and so if we reindex them (using the countability of $\mathbb{Z}^2$) and rename the functions $f_1, f_2, ...$, indexing with $\mathbb{N}$, then we can embed X into $\mathbb{R}^{\mathbb{N}}$ which completes the proof. $\square$

15.7 Tychonoff's Theorem

In this section our goal is to prove Tychonoff's Thoerem, which says that the product of any collection of compact spaces is compact using the product topology. In fact, this theorem is the reason the product topology and even compactness itself are defined the way they are! Originally, sequential compactness was taken as the definition but it turns out that the product of sequentially compact spaces need not be compact. Similarly, if one uses the naive definition of the product topology (using a basis of all products of open sets as in the finite case) then again the product of compact sets need not be compact. These definitions have been fine tuned to give useful results that apply broadly and Tychonoff's Theorem is the reward. Our proof will use the notion of a subbase which we now define.

Definition 15.22. Suppose (X, τ) is a topological space and $\mathcal{S} \subseteq \tau$. Then $\mathcal{S}$ is a **Subbase**

for τ if τ is the smallest topology containing $\mathcal{S}$. As usual, such a topology always exists.

A basis is always a subbase for a space, but the other direction need not be true. In a certain sense, subbases are too small to be guaranteed to be a full basis. Consider the following example where $X = \{1, 2, 3\}$ and

$$\mathcal{S} = \{\{1, 2\}, \{2, 3\}\}.$$

The topology generated by $\mathcal{S}$ is $\tau = \{\emptyset, \{1, 2\}, \{2, 3\}, \{2\}, \{1, 2, 3\}\}$, so $\mathcal{S}$ is NOT a basis for τ as there is no element of $\mathcal{S}$ that is a subset of the open set $\{2\}$.

What can be said however, is that a collection $\mathcal{S}$ is a subbase if and only if the set of all finite intersections of members of $\mathcal{S}$ IS a basis.

Subbases are important as (assuming the Axiom of Choice) as they give us yet another tool for dealing with compactness, one that we will use in the proof of Tychonoff's Theorem.

Theorem 15.379 (Alexander's Subbase Theorem) Suppose X is a topological space and $\mathcal{S}$ is a subbase. If every cover of X by members of $\mathcal{S}$ has a finite subcover, then X is compact. In essence, we need only check certain open covers by subbase elements to determine compactness.

Proof. Our proof will use Zorn's Lemma, which as mentioned is equivalent to the Axiom of Choice.

For contradiction suppose X is not compact, but our assumption holds. Let $\mathcal{C}$ be the collection of all open covers of X that do not have a finite subcover, clearly $\mathcal{C}$ is nonempty. Define a partial ordering $\leq$ on $\mathcal{C}$ by $A \leq B$ if $A \subseteq B$. For any totally ordered subset of $\mathcal{C}$, the union of all members is an upper bound so Zorn's Lemma applies. Let M be our maximal element of $\mathcal{C}$.

We see that there is some point $x \in X$ so that x is not in the union of sets in $M \cap \mathcal{S}$, for if the intersection covered X a finite subcover would exist by hypothesis. But for this x there is some $V \in M$ so that $x \in V$, and since the collection of finite intersection of $\mathcal{S}$ forms a basis, we have for some collection $S_1, ..., S_n \in \mathcal{S}$ that $x \in S_1 \cap ... \cap S_n \subseteq V$ and each $S_i \notin M$. Therefore, for each S_i we have that each $M \cup \{S_i\}$ must have a finite subcover by maximality of M, so let M_i be the collection of all sets in the subcover (excluding S_i). Since each M_i is finite for $i = 1, ..., n$ the collection of all these sets say M' is a finite collection of setes in M. We have for each $i = 1, ..., n$ that $S_i \cup (\bigcup M_i) = X$. But then

$$X = \left(S_1 \cup \left(\bigcup M_1\right)\right) \cap ... \cap \left(S_n \cup \left(\bigcup M_n\right)\right) \subseteq \bigcap_{i=1}^{n} \left(V \cup \left(\bigcup M'\right)\right) = V \cup \left(\bigcup M'\right).$$

But since $V \in M$ and M' is a finite collection of sets in M we see that $\{V\} \cup M'$ is a finite subcover of M, contradicting the fact that M had no subcover and completing the proof. $\square$

With the subbase theorem, we can now prove Tychonoff's Theorem.

Theorem 15.380 (Tychonoff's Theorem) Let $\{X_i\}_{i \in I}$ be a collection of compact sets, then $X = \prod_{i \in I} X_i$ with the product topology is compact.

Proof. If any X_i is empty, then the product is empty and so we are done. We therefore assume each X_i is nonempty. Set $\mathcal{S}_i = \{\pi_i^{-1}(U_i) | U_i \text{ is open in } X_i\}$ and then $\mathcal{S} = \bigcup_{i \in I} \mathcal{S}_i$. Then $\mathcal{S}$ is a subbase for X (as the product topology is the smallest topology ensuring all of these sets are open).

To start, it's not too difficult to see that each $\mathcal{S}_i$ forms a cover of X.

We will show that given any cover $\mathcal{C}$ there is some i so that $\mathcal{C} \cap \mathcal{S}_i$ is also a cover for X, for which it will be easy to find a finite subcover.

For contradiction suppose $\mathcal{C} \subseteq \mathcal{S}$ is a cover and suppose for all i that $\mathcal{C} \cap \mathcal{S}_i$ does not cover X. Since each $\mathcal{S}_i$ covers X it follows that for each i there is some nonempty open set $V_i \in X_i$ so that $\pi_i^{-1}(V_i) \notin \mathcal{C}$. For each V_i, we pick some point $x_i \in V$. However, now the point $x \in X$ where $\pi_i(x) = x_i$ cannot lie in $\mathcal{C}$ which contradicts the fact $\mathcal{C}$ was a cover!

So we see that for some $j \in I$ that $\mathcal{C} \cap \mathcal{S}_j$ is a cover for X. To end the proof, by compactness of X_i finitely many of the sets in $\mathcal{C}\cap\mathcal{S}_j$ are needed to cover X and so by Alexander's Subbase Theorem X is compact. $\square$

We conclude this chapter with one consequence of Tychonoff's Theorem, namely the Stone-Čech compactification theorem.

Definition 15.25. We define a **Compactification** of a topological space X as a compact space Y so that X can be embedded as a dense subspace of Y.

Theorem 15.26. Let X be a completely regular space. Then there exists a Hausdorff compactification βX with the property that if $f : X \to \mathbb{R}$ is bounded and continuous, then it has a unique extension to a continuous function $\tilde{f} : Y \to \mathbb{R}$. The space βX is called the **Stone-Čech Compactification of** X.

Proof. Let $\{f_i\}_{i\in I}$ be all the bounded continuous real valued functions on X, indexed by some set I. Each f_i is bounded, and therefore the range lies in some set $[a_i, b_i]$. The set $Y = \prod_{i\in I}[a_i, b_i]$ is compact by Tychonoff's Theorem. Let $x \in X$ by assumption $\{x\}$ is closed. Similarly if $N(x)$ is a neighborhood of x, by complete regularity we find a continuous function $f' : X \to \mathbb{R}$ with $f'(x) = 1$ and $f'(N(x)^c) = \{0\}$, this $f' \in \{f_i\}$. In other words, the collection of functions $\{f_i\}$ satisfies the criterion of Theorem XXX, so the function $g : X \to Y$ by $g(x) = (f_i(x))_{i\in I}$ is an embedding into Y. We can then set $\beta X = \overline{g(X)}$. It follows βX is compact and it is easy to verify it is Hausdorff.
Further, given a function f_i set $\tilde{f}_i : Y \to \mathbb{R}$ by $\tilde{f}_i(x) = \pi_i(x)$. Continuity is immediate and for $x \in X$ if we identify x as $g(x)$ by the embedding, we see that $\tilde{f}_i(g(x)) = \pi_i(g(x)) = f_i(x)$. For contradiction suppose $\tilde{f}_i$ and f'_i are two continuous extensions of f_i then let x be a point so that $\tilde{f}_i(x) \neq f'_i(x)$, and let $\tilde{f}_i(x) \in U$ $f'_i(x) \in V$ be disjoint open sets, which exist as $\mathbb{R}$ is Hausdorff. Then the inverse images of these sets are open sets in Y, so by density of $g(X)$ we reach a contradiction. $\square$

Bibliography

[1] Rudin, Walter. Principles of Mathematical Analysis. McGraw-Hill, 2010.

[2] Rudin, Walter. Real and Complex Analysis: 3rd. Ed. McGraw-Hill, 1987.

[3] Rudin, Walter. Functional Analysis. McGraw-Hill, 1973.

[4] Abott, Stephen. Understanding Analysis: 2nd. Ed. Springer, 2015.

[5] Munkres, James. Topology. Pearson, 1974.

[6] Munkres, James. Analysis on Manifolds. Avalon Publishing, 1997.

[7] Spivak, Michael. Calculus on Manifolds. Addison-Wesley Publishing Company, 1965.

[8] Folland, Gerald. Real Analysis Modern Techniques and Their Applications. John Wiley and Sons, 1999.

[9] Royden, Halsey, Fitzpatrick, Patrick. Real Analysis: 4th. Ed. Pearson, 2010

[10] Ahlfors, Lars. Complex Analysis. McGraw-Hill, 1979.

[11] Tao, Terrence. An Introduction to Measure Theory. American Mathematical Society, 2011.