REVIEW OF MODERN ANALYSIS II

DESTINE LEE AND ALAN ZHAO

The goal of these notes is to consolidate our knowledge of the material presented in the undergraduate course "Modern Analysis II" offered by Columbia University. The course adheres quite strictly to two books authored by Rudin ([R+64]) and Stein and Shakarchi ([SS09]). We draw from Chapters 9 and 10 of the former, Chapters 1 and 2 of the latter. The idea will be to write about key advancements in the theory of real analysis as presented in these texts. Proofs will be presented with principal focus on main ideas and also motivic ideas behind constructed objects where possible.

1. SS: Measure Theory

In this section we construct the Lebesgue measure in $\mathbb{R}^n$, which may be thought of as a formalization of the intuitive property of volume. Indeed, the theory of the Lebesgue measure will turn out to simultaneously align with and challenge our basic intuitions of volume. For instance, the "fat Cantor set" has a boundary of positive measure!

The theory of the Lebesgue measure starts from a well-known volume: those of rectangles. We now introduce some of the specialized terminology of this section.

Definition 1.1. *A* ***closed rectangle*** *$R \subset \mathbb{R}^n$ is given by*

$$R = \prod_{i=1}^{n} [a_i, b_i], \tag{1.1}$$

where the $a_i, b_i \in \mathbb{R}$ and $a_i \leq b_i$. Its ***volume*** *$|R|$ is given by*

$$|R| = \prod_{i=1}^{n} (b_i - a_i). \tag{1.2}$$

Definition 1.2. *An* ***open rectangle*** *has the definition of a closed rectangle but with the brackets replaced with parentheses.*

From here on we use the word "rectangle" in equivalence with the phrase "closed rectangle".

Definition 1.3. *A family of rectangles $\mathcal{R}$ is said to be* ***almost disjoint*** *if their interiors are pairwise disjoint.*

We will assume henceforth that $\mathcal{R}_{all}$ is the union of all rectangles in $\mathcal{R}$. We state the following two lemmas without proof, as they can be derived from Definition 1.1.

Lemma 1.4. *Let $\mathcal{R}$ be almost disjoint and finite. Assume $\mathcal{R}_{all}$ is a rectangle. Then,*

$$|\mathcal{R}_{all}| = \sum_{R \in \mathcal{R}} |R|. \tag{1.3}$$

Date: September 2020.

Lemma 1.5. *Let R be a rectangle and $\mathcal{R}$ be a finite family of rectangles, and suppose $R \subset \mathcal{R}_{all}$. Then,*

$$|R| \leq \sum_{R' \in \mathcal{R}} |R'|. \tag{1.4}$$

A theme in the volume aspect of real analysis, brought to us by Riemann integration, is to construct complicated regions from rectangles. And indeed, with the technology of infinitesimals afforded to us by Riemann, we reach a natural question: to what extent can we cover an open set $\mathcal{O}$ by rectangles? It turns out that for $\mathbb{R}$, we can find a countable disjoint covering by open intervals. In $\mathbb{R}^n$ for $n \geq 2$, we can find a countable almost disjoint family $\mathcal{R}$ such that $\mathcal{R}_{all} = \mathcal{O}$. The proof of the latter is achieved by finer and finer partitions of $\mathcal{O}$ with respect to the lattice $2^{-N}\mathbb{Z}^n$ as $N \to \infty$. We now prove the former.

Proof. The key question is: how do we generate these open intervals? Each point $x \in \mathcal{O}$ gives rise to an open interval since $\mathcal{O}$ is open. Guided by the cavalier relation $\mathcal{O} = \bigcup_{x \in \mathcal{O}} \{x\}$, we define I_x to be the open interval of minimal length containing x that is contained in $\mathcal{O}$. It turns out that the partition

$$\mathcal{O} = \bigcup_{x \in \mathcal{O}} I_x \tag{1.5}$$

works, and is countable since $I_x \cap \mathbb{Q}$ is non-empty. □

It is a reasonable expectation then that the "volume" of $\mathcal{O}$ should be the sums of the volumes of these component rectangles. So we have the definition of the exterior measure.

Definition 1.6. *Let $E \subset \mathbb{R}^n$. Then, the* ***exterior/outer measure*** *of E is*

$$m_*(E) := inf \sum_{R \in \mathcal{R}} |R| \tag{1.6}$$

where the infimum is taken over all countable families $\mathcal{R}$ such that $E \subset \mathcal{R}_{all}$.

A key property of m_* is that, for any $E \subset \mathbb{R}^n$, we have that $m_*(E) = \inf_{\mathcal{O} \subset E, \mathcal{O} \text{ open}} m_*(\mathcal{O})$. This is an easy proof: consider partitions into cubes.

We now define the Lebesgue measure. A subset E of $\mathbb{R}^n$ is **Lebesgue measurable** (or just **measurable**) if for any $\epsilon > 0$ there exists an open set $\mathcal{O} \supset E$ such that

$$m_*(\mathcal{O} - E) < \epsilon. \tag{1.7}$$

For such a set E we define its **Lebesgue measure** (or **measure**) $m(E)$ as

$$m(E) = m_*(E). \tag{1.8}$$

The set of Lebesgue measurable sets is closed under the familiar operations of set theory. More concretely, it is a σ-algebra. A most useful property of the Lebesgue measure is the following:

Lemma 1.7. *If $m_*(E) = 0$, then E is measurable. In particular, if $F \subset E$ and $m_*(E) = 0$, then F is measurable.*

Proof. The discussion following Definition 1.6 tells us that there exists an open set $\mathcal{O} \supset E$ such that $m_*(\mathcal{O}) < \epsilon$. Since $\mathcal{O} - E \subset \mathcal{O}$, we're done.

Remark. *This proof relies on the fact that $m_*(E_1) \leq m_*(E_2)$ when $E_1 \subset E_2$, which is clear.*

□

Before moving onto the next result, we first prove a property of the exterior measure, which by definition extends to a property of the Lebesgue measure. We omit the proof as once again it is done completely by considering a correct set of cubes.

Lemma 1.8. *If $E = E_1 \cup E_2$ and $d(E_1, E_2) > 0$, then $m_*(E) = m_*(E_1) + m_*(E_2)$.*

We will also need the following lemma, which. The proof falls to basic topology tools (e.g., Heine-Borel, triangle inequality).

Lemma 1.9. *If F is closed and K is compact, and they are disjoint, then $d(F, K) > 0$.*

We are now in a position to state an important theorem.

Theorem 1.10. *Let $\{E_i\}_{i=1}^{\infty}$ be a collection of pairwise disjoint measurable sets, and let $E = \bigcup_{i=1}^{\infty} E_i$. Then,*

$$m(E) = \sum_{i=1}^{\infty} m(E_i). \tag{1.9}$$

Proof. We don't have many tools within the Lebesgue measure thus far that can produce the desired result's form. Up to this point, Lemma 1.8 is all we have. So, with Lemma 1.9 in mind, we need to somehow pass from the E_i into closed/compact sets.

Let us begin by noting that if we prove this Theorem for bounded E_i, then we achieve the general case via compact exhaustion. Let us also note the following fact, which is easily shown by partitions into cubes:

$$m(E) \leq \sum_{i=1}^{\infty} m(E_i). \tag{1.10}$$

So, it remains to prove the reverse inequality. Consideration of E_i^C tells us that for each E_i we can choose a closed $F_i \subset E_i$ such that $m_*(E_i - F_i) < \epsilon/2^i$. For a fixed N, we know $F_1, \ldots, F_N$ are pairwise disjoint, and so we may apply Lemma 1.9 and the fact that $\bigcup_{i=1}^{N} F_j \subset E$ to conclude

$$m(E) \geq \sum_{i=1}^{N} m(F_i) \geq \left(\sum_{i=1}^{N} m(E_i) \right) - \epsilon. \tag{1.11}$$

Taking, $N \to \infty$ and $\epsilon \to 0$, we're done. □

We make a few definitions now to make the following corollary precise. Let $E_1, E_2, \ldots$ be a countable collection of sets in $\mathbb{R}^n$. We write $E_k \nearrow E$ if $E_1 \subset E_2 \ldots$ and $\bigcup_{k=1}^{\infty} E_k = E$. Similarly, we write $E_k \searrow E$ if $E_1 \supset E_2 \ldots$ and $\bigcap_{k=1}^{\infty} E_k = E$.

Corollary 1.10.1. *Suppose $E_1, E_2, \ldots$ is a countable family of measurable sets.*

(1) If $E_k \nearrow E$, then $m(E) = \lim_{N\to\infty} m(E_N)$.
(2) If $E_k \searrow E$ and $m(E_k) < +\infty$ for some k, then $m(E) = \lim_{N\to\infty} m(E_N)$.

Proof. The first follows easily upon consideration of sets of the form $E_k - E_{k-1}$ and applying the Theorem. For the second, we may assume that $m(E_1) < +\infty$ and use the same tactic as in the first. □

We note so far in this section that we have shown for any $\epsilon > 0$ and measurable set E that there exists an open set $O \supset E$ and closed set $F \subset E$ such that $m(O - E), m(E - F) < \epsilon$. But we can also show there is a compact set K satisfying $m(E - K) < \epsilon$ via a compact exhaustion of a suitable closed F combined with the above Corollary.

Before moving on to measurable functions, we pause to place the collection of Lebesgue measurable sets into greater context. We begin with defining a term mentioned earlier: a **σ-algebra** is a collection of subsets that are closed under countable unions, countable intersections, and complements. Indeed, the set of Lebesgue measurable sets satisfy all these conditions. Another important σ-algebra is the **Borel σ-algebra** in $\mathbb{R}^n$, denoted by $\mathcal{B}_{\mathbb{R}^n}$, which is the smallest σ-algebra containing all open sets of $\mathbb{R}^n$. Elements of this set are called **Borel sets**. The **G_δ-sets** are those that are a countable intersection of open sets. Their complements form the set of F_σ **sets**. The following corollary makes precise the notion that the set of Lebesgue measurable sets over $\mathbb{R}^n$ (which we denote $\Lambda(\mathbb{R}^n)$ below) is a completion of the $\mathcal{B}_{\mathbb{R}^n}$.

Corollary 1.10.2. *A set $E \in \mathcal{L}(\mathbb{R}^n)$*

(1) if and only if E differs from a G_δ set by a set of measure zero.
(2) if and only if E differs from a F_σ set by a set of measure zero.

Proof. If any of (1) or (2) hold, then $E \in \mathcal{L}(\mathbb{R}^n)$ since every set involved would be measurable. Conversely, if $E \in \mathcal{L}(\mathbb{R}^n)$, then for (1) we take the intersection of all O_n such that $m(O_n - E) < 1/n$. For (2) we find closed sets $F \subset E$ where $m(E - F) < 1/n$, and take their union. □

As promised, we now turn to the theory of measurable functions. Up to this point, we have seen that dealing with measurable sets is greatly helped by dealing with countable unions of rectangles. In this light, we can already see the forthcoming results that measurable functions will be approximated by the following three functions:

(1) The Characteristic Function:

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases} \tag{1.12}$$

where E is measurable.

(2) The Step Function:

$$f(x) = \sum_{i=1}^{N} a_i \chi_{R_i}(x) \tag{1.13}$$

where the R_i are rectangles.

(3) The Simple Function:

$$f(x) = \sum_{i=1}^{N} a_i \chi_{E_i}(x) \tag{1.14}$$

where the E_i are measurable.

Of course, the second function will be able to approximate the third. We now introduce the definition of a measurable function.

Definition 1.11. *A function $f : E \to \mathbb{R}$ (where $E \subset \mathbb{R}^n$ is measurable) is **measurable** if for all $a \in \mathbb{R}$, the set*

$$\{x \in E : f(x) < a\} \tag{1.15}$$

is measurable.

We now have the following properties:

(1) We can use the discussion following Lemma 1.5, to show that a finite-valued function f is measurable if and only if $f^{-1}(O)$ is measurable for every open set O and $f^{-1}(F)$ is measurable for every closed set F.
(2) If f is continuous on $\mathbb{R}^d$, then f is measurable. If f is measurable and finite-valued and Φ is continuous, then $\Phi \circ f$ is measurable.
(3) Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions. Then,

$$\sup_n f_n(x), \inf_n f_n(x), \limsup_{n\to\infty} f_n(x), \text{ and } \liminf_{n\to\infty} f_n(x) \tag{1.16}$$

are all measurable.
(4) We also have that $\lim_{n\to\infty} f_n(x)$ is measurable.
(5) If f and g are measurable, then the powers f^k, $k \geq 1$, are measurable. Furthermore, if both are finite-valued, fg and $f+g$ are measurable.
(6) If $f = g$ almost everywhere, then f is measurable if and only if g is measurable.

The last property uses the following terminology.

Definition 1.12. *Two functions f and g with the same measurable domain E are equal almost everywhere if the set $\{x \in E : f(x) \neq g(x)\}$ has measure zero.*

With these properties set, we now turn to formalizing the earlier statements saying that measurable functions can be approximated by characteristic, simple, and step functions. This is achieved in the following Theorem.

Theorem 1.13. *Suppose f is measurable on $\mathbb{R}^n$. Then there exists a sequence of simple functions $\{\varphi_k\}_{k=1}^{\infty}$ that satisfies*

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)| \text{ and } \lim_{k\to\infty} \varphi_k(x) = f(x) \tag{1.17}$$

for all x.

Proof. We decompose the function into positive and negative part. It remains to prove that the theorem is true for a non-negative f. In this light, we may assume that the φ_k are also non-negative and reduce to proving the condition $\varphi_k(x) \leq \varphi_{k+1}(x)$.

We want an approximation by simple functions, but we need control on the range and domain of f. This is achieved via truncations F_N defined as follows:

$$F_N(x) = \begin{cases} f(x) & x \in [-N,N]^n, f(x) \leq N \\ N & x \in [-N,N]^n, f(x) > N \\ 0 & \text{otherwise} \end{cases}. \tag{1.18}$$

From here we may produce partitions on the range of F_N which, since we also truncated the domain, allow us to produce simple functions majorized by the F_N. A little algebraic manipulation will then produce the result. □

The proof of the following theorem has difficulty to be resolved in the lemma following the statement. So, we omit the proof of the theorem and prove this lemma.

Theorem 1.14. *Suppose f is measurable on $\mathbb{R}^n$. Then there exists a sequence of step functions ψ_k that converges pointwise to f almost everywhere.*

Lemma 1.15. *For a set E of finite measure, there exists a finite set of rectangles $R_1, \dots, R_k$ such that*

$$m(E \Delta \bigcup_{j=1}^{k} R_j) \leq \epsilon, \tag{1.19}$$

where Δ is the symmetric difference operator $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

Proof. Cover E by a countable family of cubes Q_j such that $\sum |Q_j| \leq m(E) + \epsilon/2$. We can take this finite set to be $Q_1, \dots Q_N$ for some sufficiently large N. □

We conclude this section with Littlewood's three principles, which are as follows.

(1) Every set is nearly a finite union of intervals.
(2) Every function is nearly continuous.
(3) Every convergent sequence is nearly uniformly convergent.

The validity of the first comes from Lemma 1.15. The validity of the third comes from a theorem of Egorov and will be very important in the construction of the Lebesgue integral.

Theorem 1.16. *Suppose $\{f_k\}_{k=1}^{\infty}$ is a sequence of measurable functions defined on a set E of finite measure. Assume further that $f_k \to f$ almost everywhere on E. Given $\epsilon > 0$, we can find a closed set $A_\epsilon \subset E$ such that $m(E - A_\epsilon) \leq \epsilon$ and $f_k \to f$ uniformly on A_ϵ.*

Proof. Without loss of generality, we may assume $f_k \to f$ pointwise everywhere. Furthermore, in the discussion following Corollary 1.10.1 regarding closed set approximations, it remains to show that we can construct a measurable set A'_ϵ where $f_k \to f$ uniformly. To start the construction we first lay out all the possible pieces: for each pair of non-negative integers n and k, define

$$E_k^n = \{x \in E : |f_j(x) - f(x)| < 1/n, \text{ for all } j > k\}. \tag{1.20}$$

One notes that $E_k^n \nearrow E$ as $k \to \infty$, and so we apply Corollary 1.10.1 to find k_n such that $m(E - E_{k_n}^n) < 2^{-n}$. Choose N sufficiently large, and look at the intersection $\bigcap_{n \geq N} E_{k_n}^n$. □

The validity of the second comes from Lusin, who applies the theorem of Egorov.

Theorem 1.17. *Suppose f is measurable and finite valued on a set E of finite measure. Then for every $\epsilon > 0$ there exists a closed set F_ϵ with $F_\epsilon \subset E$ and $m(E - F_\epsilon) \leq \epsilon$ such that $f|_{F_\epsilon}$ is continuous.*

Proof. Let f_n be a sequence of step functions approaching f almost everywhere. Define sets E_n such that $m(E_n) < 2^{-n}$ and f_n is continuous outside the E_n. Now by the previous Theorem, there exists a set $A_{\epsilon/3}$ with $m(E - A_{\epsilon/3}) \leq \epsilon/3$ where the convergence $f_n \to f$ is uniform. Now, we get rid of just enough of these bad sets. Choose N so that $\sum_{n \geq N} 2^{-n} \leq \epsilon/3$ and

$$F' = A_{\epsilon/3} - \bigcup_{n \geq N} E_n. \tag{1.21}$$

It is easy to see continuity on F'. The proof's finish is then just a sufficiently good closed set approximation F_ϵ of F'. □

2. SS: The Lebesgue Integral

In this section we present a working knowledge of the Lebesgue integral. For its construction, we instead discuss the general ideas. Just like with step functions in Riemann integration, we take the Lebesgue integral over $\mathbb{R}^n$ of simple functions to be what we expect: $\int \sum c_i \chi_{E_i}(x) dx = \sum c_i m(E_i)$. Then, to get to bounded functions f of finite support, we approximate by simple functions φ_k as afforded to us by Theorem 1.13 with an application of Egorov's Theorem (Theorem 1.16). Then, define $\int_{\mathbb{R}^n} f = \lim_{k\to\infty} \int_{\mathbb{R}^n} \varphi_k$. For a non-negative, measurable, and not necessarily bounded f, we can define $\int f = \sup_g \int g$ where $0 \leq g \leq f$ and g is measurable, bounded, and supported on a set of finite measure. We say that f is **Lebesgue integrable** if this supremum is finite. To reach the general case, we can decompose a function f into its positive and negative part, and take the integral of both separately.

We now state and prove some important results.

Proposition 2.1. *Let f be a non-negative Lebesgue measurable function. Then,*

i. If f is integrable, then $f(x) < \infty$ for almost every x.

ii. If $\int f = 0$, then $f = 0$ almost everywhere.

Proof. (i): Our hypothesis is that f is integrable, so we try to break it. Corollary 1.10.1 is one way to break it, by considering the sets $E_k = \{x \in \mathbb{R}^n : f(x) \geq k\}$ for k a positive integer.
(ii): We could consider sets of the form $E_k = \{x \in \mathbb{R}^n : f(x) \geq 1/k\}$ for k a positive integer. We could also break up the integral into zero and positive part. □

We are now in a position to state some very important convergence theorems. These theorems will justify many of the integration techniques we have in Riemann integration. Indeed, for f Riemann integrable, its Riemann integral coincides with its Lebesgue integral.

Theorem 2.2 (Bounded Convergence Theorem)**.** *Suppose that $\{f_n\}$ is a sequence of measurable functions that are all bounded by M, are supported on a set E of finite measure, and $f_n \to f$ almost everywhere. Then, $f(x)$ is measurable, bounded, and supported on E for almost all x, and*

$$\int f_n \to \int f. \tag{2.1}$$

Proof. From the assumptions, f is bounded by M and supported almost everywhere by E. Now, by the triangle inequality, it suffices to prove that $\int |f_n - f| \to 0$. This would be near obvious if we had uniform convergence, and so we again apply Theorem 1.16. □

For the next convergence theorem we need the following Lemma, due to Fatou.

Lemma 2.3 (Fatou)**.** *Suppose $\{f_n\}$ is a sequence of measurable functions with $f_n \geq 0$. If $f_n \to f$ almost everywhere, then $\int f \leq \liminf_{n\to\infty} \int f_n$.*

Proof. The only result of this form we have is the Bounded Convergence Theorem (Theorem 2.2). So we use it. Let g be a function bounded on a set E of finite measure. We then recall the definition of the Lebesgue integral for non-negative functions, which uses functions of the form g to approach f. So, if we can show that $\int g \leq \liminf_{n\to\infty} \int f_n$, we can take supremums and we're done.

We then want to approximate g by g_n's that somehow incorporates the f_n. So, set $g_n = \min(g, f_n)$. Then, the g_n are bounded and supported on E, and $g_n \to g$ almost everywhere since $g \leq f$. So, $\int g_n \to \int g$ by the Bounded Convergence Theorem and $\int g_n \leq \int f_n$, and so $\int g \leq \liminf_{n\to\infty} \int f_n$ as requested. □

Its easy corollary is the Monotone Convergence Theorem.

Corollary 2.3.1. *Suppose $\{f_n\}$ is a sequence of non-negative measurable function with $f_n(x) \leq f_{n+1}(x)$ almost everywhere with $f_n \to f$ almost everywhere. Then, $\lim_{n\to\infty} \int f_n \to \int f$.*

Proof. Consider the limsup of $\int f_n$. □

Using the corollary it is easy to prove that

$$\int \sum_{k=1}^{\infty} a_k(x)dx = \sum_{k=1}^{\infty} \int a_k(x)dx, \tag{2.2}$$

where $a_k(x) \geq 0$. Furthermore, the right-hand side is finite, so is the left, and so $\sum a_k(x)$ is finite almost everywhere. This can be used in a proof of the Borel-Cantelli lemma.

Lemma 2.4 (Borel-Cantelli). *Let $E_1, E_2, \ldots$ be a collection of measurable sets with $\sum m(E_k) < +\infty$. Then, the set of points that belong to all E_k has measure zero.*

Proof. In the discussion following the above Corollary, set $a_k(x) = \chi_{E_k}(x)$. Then, our finiteness assumption tells us that $\sum \int a_k < \infty$, which implies that $\sum a_k(x)$ converges almost everywhere. So, for almost every x, only finitely many $a_k(x) = 1$, which implies only finitely many E_k contain x. □

We first introduce two results in the following proposition, the latter of which is known as *absolute continuity*.

Proposition 2.5. *Suppose f is integrable on $\mathbb{R}^n$. Then, for every $\epsilon > 0$:*

i. There exists a set of finite measure B such that

$$\int_{B^C} |f| < \epsilon. \tag{2.3}$$

ii. There is a $\delta > 0$ such that

$$\int_E |f| < \epsilon \tag{2.4}$$

whenever $m(E) < \delta$.

Proof. We have that f is integrable if and only if $|f|$ is. So we assume that $f \geq 0$. For (i), we consider the compact exhaustion sequence $f_N = f(x)\chi_{B_N}(x)$ where B_N is the ball of radius N centered on the origin. We can then apply the Monotone Convergence Theorem. The second can be almost immediately proved by the Monotone Convergence Theorem, with the chosen functions being guided by the idea that $m(\{x \in \mathbb{R}^n : f(x) \geq k\} \to 0$ as $k \to \infty$ introduced in previous proofs. □

We now introduce the very important Dominated Convergence Theorem.

Theorem 2.6 (Dominated Convergence Theorem). *Suppose $\{f_n\}$ is a sequence of measurable functions such that $f_n \to f$ almost everywhere. If $|f_n| \leq g$, where g is integrable, then*

$$\int f_n \to \int f. \tag{2.5}$$

Proof. The triangle inequality reduces the proof to $\int |f_n - f| \to 0$. Then, part (i) of the previous proposition combined with the bounded convergence theorem allows us to find a sufficiently good decomposition of the integral $\int |f_n - f|$. □

We now introduce the theory behind L^1 spaces. We say that f is an L^1 *function* if $\int |f|$ is finite. The *norm of f in L^1*, denoted $\|f\|$, is the quantity $\int |f|$. The set of all L^1 functions over $\mathbb{R}^n$ is denoted $L^1(\mathbb{R}^n)$. it is easy to show that $L^1(\mathbb{R}^n)$ is a $\mathbb{C}$-vector space.

Remark. *The theory of Lebesgue integration developed so far can easily be adapted to complex-valued functions by considering real and imaginary part separately. Naturally, absolute values here are in the complex sense.*

We now state a key theorem of Riesz-Fischer without proof. One possible proof involves the construction of a telescoping series with a follow-up treatment by the convergence theorems.

Theorem 2.7. *The space $L^1(\mathbb{R}^n)$ is complete in its norm.*

The proof of the following uses the standard techniques of convergence theorems and results relating to the approximation of measurable functions by simple functions.

Theorem 2.8. *The following families of functions are dense in $L^1(\mathbb{R}^n)$:*

i. Simple functions
ii. Step functions
iii. Continuous functions of compact support

This theorem is useful in proving the following.

Proposition 2.9. *Suppose $f \in L^1(\mathbb{R}^n)$. Let $f_h = f(x+h)$. Then, $\|f_h - f\| \to 0$ as $h \to 0$.*

We conclude this section with a discussion on Fubini's Theorem and a few of its consequences.

Theorem 2.10 (Fubini)**.** *Suppose $f(x,y)$ is integrable on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} = \mathbb{R}^n$. Then for almost every $y \in \mathbb{R}^{n_2}$:*

i. $f(x,y)$ is integrable on $\mathbb{R}^{n_1}$.
ii. The function $\int_{\mathbb{R}^{n_1}} f(x,y)dx$ is integrable on $\mathbb{R}^{n_2}$.
iii. $\int_{R^{n_2}} \left(\int_{\mathbb{R}^{n_1}} f(x,y)dx\right) dy = \int_{\mathbb{R}^n} f$.

Proof. Uses the approximation of L^1 functions by simple functions. Uses the ability of G_δ sets to approximate characteristic functions. □

Remark. *This theorem gives rise to Tonelli's Theorem, which asserts the same thing as Fubini's theorem except that f is non-negative and measurable instead of integrable, and the integral in (iii) is understood in the extended sense.*

Corollary 2.10.1. *If E is a measurable set in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} = \mathbb{R}^n$, then for almost every $y \in \mathbb{R}^{n_2}$, the set $E_y = \{x \in \mathbb{R}^{n_1} : (x,y) \in E\}$ is a measurable subset of $\mathbb{R}^{n_1}$. Also, $m(E_y)$ is a measurable function in y and*

$$m(E) = \int_{\mathbb{R}^{n_2}} m(E_y)dy. \tag{2.6}$$

Proof. Apply Tonelli's Theorem to a good enough characteristic function. □

We end with a result that our intuition makes us expect to be true.

Proposition 2.11. *Let $E_1 \subset \mathbb{R}^{n_1}$ and $E_2 \subset \mathbb{R}^{n_2}$ be measurable sets. Then, $E = E_1 \times E_2 \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} = \mathbb{R}^n$ is measurable. Moreover,*

$$m(E) = m(E_1)m(E_2) \tag{2.7}$$

with the understanding that if one of the sets E_j has measure zero, then $m(E) = 0$.

Proof. It suffices by the above Corollary to show that E is measurable. Through the proof of Corollary 1.10.2, we can find G_δ sets $G_i \supset E_i$ such that $m(G_i \setminus E_i) = 0$. We do this because we know that $G_1 \times G_2$ is G_δ, and so measurable. We now note that

$$(G_1 \times G_2) \setminus (E_1 \times E_2) \subset (G_1 \setminus E_1) \times G_2) \cup (G_1 \times (G_2 \setminus E_2). \tag{2.8}$$

Then the following lemma, which can be proved directly by consideration of cubes, finishes the job:

Lemma 2.12. *$m_*(E_1 \times E_2) \leq m_*(E_1)m_*(E_2)$.*

□

3. SS: Hilbert Spaces

We will see in this section that Hilbert spaces are useful in that they generalize Euclidean space to infinite dimensions. In fact, we will see in some sense that all infinite-dimensional Hilbert spaces are equivalent!

We start with an example of a Hilbert space: $L^2(\mathbb{R}^n)$, consisting of functions f such that $\int |f|^2 < +\infty$. The norm of $L^2(\mathbb{R}^n)$ is

$$\|f\| = \left(\int |f|^2\right)^{1/2} \tag{3.1}$$

which follows from the natural inner product

$$(f, g) = \int f\overline{g}. \tag{3.2}$$

We list some of the properties of $L^2(\mathbb{R}^n)$ in the proposition below.

Proposition 3.1. *The space $L^2(\mathbb{R}^n)$ has the following properties:*

i. It is a vector space.
ii. $f\overline{g}$ is integrable whenever $f, g \in L^2(\mathbb{R}^n)$. Moreover, Cauchy-Schwarz holds: $|(f, g)| \leq \|f\| \cdot \|g\|$.
iii. If $g \in L^2(\mathbb{R}^n)$ is fixed, the map $f \mapsto (f, g)$ is linear in f, and also $(f, g) = \overline{(g, f)}$.
iv. We have the triangle inequality in $L^2(\mathbb{R}^n)$.

Proof. We must keep algebraic identities with squares in mind as that is the theme here. The proof of (i) and the first assertion of (ii) easily follow from the identity $2AB \leq A^2 + B^2$ when $A, B \geq 0$. The second statement of (ii) in fact also easily follows from this observation. (iii) is obvious by the linearity of the integral. (iv) is just direct computation. □

Just like in L^1, the space $L^2(\mathbb{R}^n)$ is complete in its metric. The proof is almost a copy of Theorem 2.7. We now move to a topic that brings us closer to the theory of general Hilbert spaces: the concept of separability.

Theorem 3.2. *The space $L^2(\mathbb{R}^n)$ is* ***separable****, meaning that there exists a countable collection of elements in $L^2(\mathbb{R}^n)$ such that their finite linear combinations are dense in $L^2(\mathbb{R}^n)$.*

Proof. Abuse the fact that $\mathbb{Q} \subset \mathbb{R}$ is dense and countable. From here consider the natural set of functions. We conduct a first approximation of f by an appropriate truncation. We now apply Theorem 2.8 to approximate these arbitrarily well, which are bounded on a set of finite support. A little more calculation yields the desired result. □

As implied, we now introduce the definition of a Hilbert space.

Definition 3.3. *A set $\mathcal{H}$ is called a **Hilbert space** if it satisfies the following:*

i. *$\mathcal{H}$ is a vector space over $\mathbb{C}$ or $\mathbb{R}$.*
ii. *$\mathcal{H}$ has an inner product $(\cdot, \cdot)$ where: (1) $f \mapsto (f, g)$ is linear for every fixed $g \in \mathcal{H}$, (2) $(f, g) = \overline{(g, f)}$, (3) $(f, f) \geq 0$ for all $f \in \mathcal{H}$.*
iii. *The norm $\|f\| = 0$ if and only if $f = 0$.*
iv. *We have the Cauchy-Schwarz and triangle inequalities.*
v. *$\mathcal{H}$ is complete in the metric $d(f, g) = \|f - g\|$.*
vi. *$\mathcal{H}$ is separable.*

We state some basic properties of an **orthonormal** set $\{e_1, e_2, \ldots\}$, which is an at most countable set of elements of $\mathcal{H}$ such that $(e_i, e_j) = 0$ when $i \neq j$ and $(e_i, e_i) = 1$. Repeated applications of an analog of the Pythagorean theorem yields

Proposition 3.4. *Let $f = \sum a_k e_k \in \mathcal{H}$ where the sum is finite. Then,*

$$\|f\|^2 = \sum |a_k|^2. \tag{3.3}$$

We go further to say these $\{e_i\}$ are an **orthonormal basis** if their finite linear combinations are dense in $\mathcal{H}$. We give some equivalent conditions for an orthonormal set to also be an orthonormal basis.

Theorem 3.5. *Let $\{e_1, e_2, \ldots\}$ be an orthonormal set in $\mathcal{H}$. TFAE:*

i. *The set $\{e_i\}$ is an orthonormal basis.*
ii. *If $f \in \mathcal{H}$ and $(f, e_j) = 0$ for all j, then $f = 0$.*
iii. *If $f \in \mathcal{H}$ and $S_N(f) = \sum_{k=1}^{N} (f, e_k) e_k$, then $S_N(f) \to f$ in the norm.*
iv. *$\|f\|^2 = \sum_{k=1}^{\infty} |(f, e_k)|^2$ (**Parseval's Identity**).*

Proof. (i) $\Longrightarrow$ (ii): Utilize all available hypotheses directly. We know that $\|f\| = 0$ if and only if $f = 0$. Use this to invoke Cauchy-Schwarz.
(ii) $\Longrightarrow$ (iii): We want to show that the sequence $S_N(f)$ is Cauchy. The differences $|S_N(f) - S_M(f)|$ are finite sub-sums of $\sum_{i=1}^{\infty} |a_i|^2$. So, we set out to show that $\sum_{i=1}^{\infty} |a_i|^2$ is finite.

We can naturally get this sum out of $\|S_N(f)\|^2$ and taking $N \to \infty$. Furthermore, the fact that the e_i are pairwise orthogonal leads us to the implication that $(f - S_N(f)) \perp S_N(f)$. Now, applying the Pythagorean Theorem and taking $N \to \infty$ yields **Bessel's inequality**:

$$\sum_{i=1}^{\infty} |a_i|^2 \leq \|f\|^2, \tag{3.4}$$

which implies the desired convergence. Say $S_N(f) \to g$. The rest of the argument takes into consideration the quantity $(f - g, e_j)$ in light of (ii).
(iii) $\Longrightarrow$ (iv), (iv) $\Longrightarrow$ (i): Follows immediately from the proof of (ii) $\Longrightarrow$ (iii) in the discussion before the presentation of Bessel's inequality.

□

Theorem 3.6. *Any Hilbert space has an orthonormal basis.*

Proof. A Hilbert space is by definition separable. Now apply Gram-Schmidt. □

We now move to discuss "isomorphisms" of Hilbert spaces given by unitary mappings. Let $\mathcal{H}$ and $\mathcal{H}'$ be two Hilbert spaces with inner products $(\cdot,\cdot)_{\mathcal{H}}$ and $(\cdot,\cdot)_{\mathcal{H}'}$, respectively, and the corresponding norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}'}$.

Definition 3.7. *A mapping $U : \mathcal{H} \to \mathcal{H}'$ is called **unitary** if*

(1) U is linear.
(2) U is a bijection.
(3) $\|Uf\|_{\mathcal{H}'} = \|f\|_{\mathcal{H}}$ for all $f \in \mathcal{H}$.

*We say that $\mathcal{H}$ and $\mathcal{H}'$ are **unitarily isomorphic** if such a U exists.*

The following is an easy corollary of these definitions.

Corollary 3.7.1. *Any two infinite-dimensional Hilbert spaces are unitarily equivalent.*

Proof. Let $\{e_i\}$ be an orthonormal basis for $\mathcal{H}$ and $\{f_i\}$ be an orthonormal basis for $\mathcal{H}'$. The mapping $U(e_i) = f_i$ suffices. □

Remark. *This is the formalization of the remark at the beginning of the section where we stated that all infinite-dimensional Hilbert spaces are equivalent.*

Corollary 3.7.2. *Any two finite-dimensional Hilbert spaces are unitarily equivalent if and only if they have the same dimension.*

Proof. The "if" direction is easy: just follow the prove of the previous corollary. For the "only if" direction, U is a bijective linear map, and a vector space over a field F cannot surject via a linear map into a higher dimensional vector field over F. □

4. Rudin: Functions $\mathbb{R}^n \to \mathbb{R}^m$

In this section, we will go over some of the most essential properties of functions $\mathbb{R}^n \to \mathbb{R}^m$. We begin by recalling a key theorem on linear operators.

Theorem 4.1. *A linear operator A on a finite-dimensional vector space X is one-to-one if and only if $A(X) = X$.*

Proof. Let $\mathbf{x}_1, \ldots \mathbf{x}_n$ be a basis for X. Then, $A(X)$ is the span of the set $A\mathbf{x}_1, \ldots, A\mathbf{x}_n$. We show that these vectors are linearly independent if and only if A is one-to-one. The "if" is clear, so it remains to show the "only if", which falls to a consideration of linearity. □

The next topic we will go over in this section is the norm of a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ (we denote the set of all these linear transformations as $L(\mathbb{R}^n, \mathbb{R}^m)$).

Definition 4.2. *Let $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Then, the **norm** $\|A\|$ is defined by*

$$\|A\| = \sup_{|\mathbf{x}|\leq 1} |A\mathbf{x}|. \tag{4.1}$$

The following properties of this norm come from (1) consideration of the triangle inequality and (2) the observation that $|A\mathbf{x}| \leq \|A\||\mathbf{x}|$.

Theorem 4.3. *We have the following:*

i. If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|A\| < \infty$ and is uniformly continuous.
ii. Let $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$. Then, $\|A + B\| \leq \|A\| + \|B\|$ and $\|cA\| = c\|A\|$.
iii. $\|BA\| \leq \|B\| \cdot \|A\|$.

Remark. *We can prove that $\|A\|$ is finite in another way by considering the matrix of A and utilizing Cauchy-Schwarz.*

With this norm, we can establish a notion of distance in the set $L(\mathbb{R}^n, \mathbb{R}^m)$. With this we have the notions of open sets, continuity, etc. It is left as an exercise to prove the following, which is doable with basic manipulation of linear maps.

Theorem 4.4. *Let Ω be the set of all invertible linear operators $\mathbb{R}^n \to \mathbb{R}^n$.*

i. If $A \in \Omega$ and $B \in L(\mathbb{R}^n, \mathbb{R}^n)$, and

$$\|B - A\| \cdot \|A^{-1}\| < 1 \tag{4.2}$$

then $B \in \Omega$.

ii. Ω is an open subset of $L(\mathbb{R}^n)$. The mapping $A \to A^{-1}$ is continuous on Ω.

We now move to the definition of a derivative.

Definition 4.5. *Suppose $E \subset \mathbb{R}^n$ is open, $f : E \to \mathbb{R}^m$ is a function, and $\mathbf{x} \in E$. Then, if there exists a linear transformation $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ that satisfies*

$$\lim_{\mathbf{h} \to 0} \frac{|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0, \tag{4.3}$$

*then we say that f is **differentiable** at $\mathbf{x}$ and we write $f'(\mathbf{x}) = A$.*

An easy check using the triangle inequality shows that this definition determines A uniquely. It is also easy to show that where f is differentiable it is also continuous. Finally, we note that for a linear transformation A, $A'(\mathbf{x}) = A$ always. Before moving to partial derivatives, we mention a generalization of the most useful chain rule in real analysis.

Theorem 4.6. *Let $F(\mathbf{x}) = g(f(\mathbf{x}))$, where f, g, and F have appropriate domains and codomains to make this composition make sense. Say that $\mathbf{g}$ is differentiable at $f(\mathbf{x}_0)$ and that f is differentiable at $\mathbf{x}_0$. Then,*

$$F'(\mathbf{x}_0) = g'(f(\mathbf{x}_0))f'(\mathbf{x}_0). \tag{4.4}$$

A closely related concept to that of differentiation is partial differentiation, where we split up $f : \mathbb{R}^n \to \mathbb{R}^m$ into its component parts. Write $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ with respect to a standard basis of $\mathbb{R}^m$. Then, we define the partial derivative $(D_j f_i)(\mathbf{x})$ as

$$(D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t}, \tag{4.5}$$

where $1 \leq j \leq n$ and $1 \leq i \leq m$. By a standard splitting by component argument, we have the following theorem:

Theorem 4.7. *Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and it is differentiable at $\mathbf{x} \in E$. Then, the partial derivatives $(D_j f_i)(\mathbf{x})$ exist, and*

$$f'(\mathbf{x})\mathbf{e}_j = \sum_{i=1}^{m} (D_j f_i)(\mathbf{x})\mathbf{u}_i \tag{4.6}$$

where the $\mathbf{e}_i$ are the standard basis vectors for $\mathbb{R}^n$ and the $\mathbf{u}_i$ are the standard basis vectors for $\mathbb{R}^m$.

By combining this theorem with a differentiable mapping $\gamma : (a, b) \to \mathbb{R}^n$, we may invoke the notion of a gradient to give us some nice expressions. This idea is useful in the following theorem invoking convexity (a natural choice, as this is where lines neatly arise, and lines are easily parameterized in one variable).

Theorem 4.8. *Suppose f maps a convex open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, f is differentiable in E, and there is a real number M such that $\|f'\| \leq M$. Then,*

$$|f(\mathbf{b}) - f(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}| \tag{4.7}$$

for all $\mathbf{a}, \mathbf{b} \in E$.

Proof. For each pair $\mathbf{a}, \mathbf{b} \in E$, consider the line segment γ connecting them. Then consider $g(t) = f(\gamma(t))$. An analog of the Mean Value Theorem for functions $\mathbb{R} \to \mathbb{R}^n$ concludes the proof:

Theorem 4.9 (Rudin, Theorem 5.19)**.** *Suppose $f : [a, b] \to \mathbb{R}^n$ and f is differentiable in (a, b). Then there exists $x \in (a, b)$ such that*

$$|f(b) - f(a)| \leq (b - a)|f'(x)|. \tag{4.8}$$

□

We now introduce a very important class of functions.

Definition 4.10. *A differentiable mapping f of an open set $E \subset \mathbb{R}^n$ to $\mathbb{R}^m$ is called* ***continuously differentiable*** *in E if $f' : E \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous. The set of all continuously differentiable functions on E is denoted $C^1(E)$.*

With this we can provide a converse to Theorem 4.7.

Theorem 4.11. *We have that $f \in C^1(E)$ if and only if the partials $D_j f_i$ exists and are continuous on E.*

Proof. The "only if" direction is an easy bounding argument. For the "if" direction, because we can split into component functions it suffices to consider $m = 1$. The remainder of the proof is left largely to the ordinary Mean Value Theorem, which largely guides the remaining constructions necessary for the proof. □

We now pause to introduce the Contraction Principle, which will be of use in our proof of the Inverse Function Theorem.

Definition 4.12. *Let X be a metric space with metric d. If φ maps X into X and there is a number $c < 1$ such that*

$$d(\varphi(x), \varphi(y)) \leq c d(x, y), \tag{4.9}$$

then φ is said to be a ***contraction*** *of X into X.*

Theorem 4.13. *If X is a complete metric space, and if φ is a contraction of X into X, then there exists precisely one $x \in X$ such that $\varphi(x) = x$.*

Proof. Uniqueness is trivial by the definition of a contraction. Existence is proven by the motivating fact that $c^n \to 0$ as $n \to \infty$. This will naturally lead to the production of a Cauchy sequence which, since X is complete, converges. □

It is now time to state and provide a proof sketch for the useful Inverse Function Theorem, which we now delve into.

Theorem 4.14. *Suppose $f \in C^1(E)$ with codomain $\mathbb{R}^m$ and $E \subset \mathbb{R}^n$ is open, $f'(\mathbf{a})$ is invertible for some $\mathbf{a} \in E$, and $\mathbf{b} = f(\mathbf{a})$. Then,*

i. There exist open sets U and V in $\mathbb{R}^n$ such that $\mathbf{a} \in U$, $\mathbf{b} \in V$, f is 1-1 on U, and $f(U) = V$.
ii. If g is the inverse of f, then $g \in C^1(V)$.

Proof. We start with (i). Set $A = f'(\mathbf{a})$, and choose λ so that $2\lambda\|A^{-1}\| = 1$ (this choice of λ so happens to make the formulas in Rudin's proof look nicer). By the continuity of f' at $\mathbf{a}$, we may select an open ball U centered at $\mathbf{a}$ such that if $\mathbf{x} \in U$, $\|f'(\mathbf{x}) - A\| < \lambda$.

We want to show that f is injective on U. To do so directly is somewhat futile (e.g., we don't have nice properties of linearity), and so we need a workaround. For this we turn to the contraction principle. For $\mathbf{y} \in \mathbb{R}^n$ we associate a function $\varphi(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - f(\mathbf{x}))$. Then, $f(\mathbf{x}) = \mathbf{y}$ if and only if $\mathbf{x}$ is a fixed point of φ. We now have:

$$|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)| \le \frac{1}{2}|\mathbf{x}_1 - \mathbf{x}_2|. \tag{4.10}$$

This identity is deduced by Theorem 4.8. So, we have that f is 1-1 on U. Now we set $V = f(U)$. It remains to show that V is open. This involves fixing $\mathbf{y}_0 \in V$ and showing that $\mathbf{y} \in V$ whenever $|\mathbf{y} - \mathbf{y}_0| < \lambda r$. Let $\mathbf{x}_0 \in U$ map to $\mathbf{y}_0$. Let B be an open ball of radius r around $\mathbf{x}_0$ such that $\overline{B} \subset U$. Then, because of our clever choice of λ, $|\varphi(\mathbf{x}_0) - \mathbf{x}_0| < r/2$, and by the triangle inequality we can deduce that $|\varphi(\mathbf{x}) - \mathbf{x}_0| < r$. So, $\varphi(\mathbf{x}) \in B$ where $\mathbf{x} \in \overline{B}$. The remainder of the first part is shown by the completeness of $\overline{B}$ and Theorem 4.13.

For part (ii), the key ingredients are to (1) use φ to help produce a derivative for g at any $\mathbf{y} \in V$ and (2) use the fact that inversion is a continuous mapping of Ω onto Ω (Theorem 4.4). □

Corollary 4.14.1. *If $f \in C^1(E)$ with codomain $\mathbb{R}^n$ and $E \subset \mathbb{R}^m$ is open, and if f' is always invertible, then $f(W)$ is an open set for every open set $W \subset E$.*

We next introduce the Implicit Function Theorem, which makes use of the Inverse Function Theorem in its proof.

Theorem 4.15. *Let $f \in C^1(E)$ with codomain $\mathbb{R}^n$ and $E \subset \mathbb{R}^n \times \mathbb{R}^m$ is open. Assume that $f(\mathbf{a}, \mathbf{b}) = 0$ for some $(\mathbf{a}, \mathbf{b}) \in E$. Put $A = f'(\mathbf{a}, \mathbf{b})$ and assume that A_x (with $A_x(\mathbf{h}) = A(\mathbf{h}, \mathbf{0})$) is invertible. Then, there exist open sets $U \subset \mathbb{R}^n \times \mathbb{R}^m$ and $W \subset \mathbb{R}^m$ with $(\mathbf{a}, \mathbf{b}) \in U$ and $\mathbf{b} \in W$ having the following property: To every $\mathbf{y} \in W$ corresponds a unique $\mathbf{x}$ such that $(\mathbf{x}, \mathbf{y}) \in U$ and $f(\mathbf{x}, \mathbf{y}) = 0$. If $\mathbf{x} = g(\mathbf{y})$, then g is a C^1-mapping of W into $\mathbb{R}^n$, $g(\mathbf{b}) = \mathbf{a}$, and $f(g(\mathbf{y}), \mathbf{y}) = 0$. Furthermore, $g'(\mathbf{b}) = -(A_x)^{-1}A_y$, where $A_y\mathbf{h} = A(\mathbf{0}, \mathbf{h})$.*

Proof. For the proof of the first part, define F by $F(\mathbf{x}, \mathbf{y}) = (f(\mathbf{x}, \mathbf{y}), \mathbf{y})$. Then, F is a C^1-mapping of $\mathbb{R}^n \times \mathbb{R}^m$ into $\mathbb{R}^n \times \mathbb{R}^m$. It is relatively easy to show that $F'(\mathbf{a}, \mathbf{b})$ is invertible. From here, we may apply the Inverse Function Theorem to F. It is also relatively easy to show that F is 1-1.

For the proof of the second part, let $g(\mathbf{y})$ such that $f(g(\mathbf{y}), \mathbf{y}) = 0$. Then, considering the inverse of F will quickly yield that g is C^1. Proving the formula in the second part has the chain rule as its key ingredient. □

We next introduce the Rank Theorem, which we state without proof, as it was communicated to us by the professor of the course that it isn't quite clear the application of this theorem.

Theorem 4.16. *Suppose m, n, r are non-negative integers with $m, n \ge r$. Let F be a C^1-mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, and F' has rank r. Now fix $\mathbf{a} \in E$ and let $A = F'(\mathbf{a})$. Let Y_1 be the range of A and let P be a **projection** in $\mathbb{R}^m$ (this means that $P \in L(\mathbb{R}^m, \mathbb{R}^m)$ and that $P^2 = P$) whose* *range is Y_1. Let Y_2 be the null space of P. Then, there are open sets U and V in $\mathbb{R}^n$, with $\mathbf{a} \in U$, $U \subset E$, and there is a 1-1 C^1-mapping $H : V \to U$ such that*

$$F(H(\mathbf{x})) = A\mathbf{x} + \varphi(A\mathbf{x}) \tag{4.11}$$

where φ is a C^1-mapping of the open set $A(V) \subset Y_1$ into Y_2.

We end this section with a discussion of higher order derivatives. We say a function $f \in C^2(E)$ if its partials $D_1 f, \ldots D_n f$ are differentiable and each of their partial derivatives are continuous on E. We want to see when we can interchange the order of the derivative in functions of two variables (although the final result can be easily generalized to an arbitrary number of variables by induction). This first preliminary result is easily shown by the ordinary Mean Value Theorem.

Theorem 4.17. *Suppose f is defined in an open set $E \subset \mathbb{R}^2$, and $D_1 f$ and $D_2 D_1 f$ exist at every point of E. Suppose $Q \subset E$ is a closed rectangle with sides parallel to the coordinate axes, having (a, b) and $(a + h, b + k)$ as opposite vertices ($h, k \neq 0$). Put*

$$\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b). \tag{4.12}$$

Then there exists a point (x, y) in the interior of Q such that $\Delta(f, Q) = hk(D_2 D_1 f)(x, y)$.

The proof of the next theorem is a direct application of the above theorem.

Theorem 4.18. *Suppose f is defined in an open set $E \subset \mathbb{R}^2$ and suppose that $D_1 f$, $D_2 D_1 f$, and $D_2 f$ exist at every point of E, and $D_2 D_1 f$ is continuous at some point $(a, b) \in E$. Then,*

$$(D_1 D_2 f)(a, b) = (D_2 D_1 f)(a, b). \tag{4.13}$$

5. Rudin: Integration of Differential Forms

In this section, we first introduce a few objects, prove some of their key properties, and then combine them into a formulation of Stokes' Theorem.

Our next target is an intrinsic notion of *boundary* for surfaces, and to begin, we define the boundary of a k-simplex, with the standard k-simplex as our prototype.

The **standard k-simplex**, which we denote by Q_k, is the subset of $\mathbb{R}^k$ determined by the points

$$\left\{(a_1, \ldots, a_k) : \alpha_i > 0 \text{ and } \sum \alpha_i \leq 1\right\}.$$

Equivalently, the standard k-simplex is a k-simplex embedded in $\mathbb{R}^k$ with vertices $0, e_1, \ldots, e_k$.

Definition 5.1. *An **oriented affine k-simplex** in an open set $E \subseteq \mathbb{R}^n$ is an affine map $\sigma : Q_k \to E$.*

We emphasize that, unlike the standard k-simplex, an oriented affine k-simplex is not a point-set but a function, and in fact a k-surface. This formulation allows us to transfer our theory of differential forms over to these objects.

It is easy to see that an oriented affine k-simplex σ is completely determined by the images $\sigma(0), \sigma(e_1), \ldots, \sigma(e_k)$ of the vertices of Q_k. In fact, the image of σ is precisely their convex hull. Therefore, we may alternatively denote σ by the ordered $(k+1)$-tuple $[p_0, \ldots, p_k]$, where $p_0 = \sigma(0)$ and $p_i = \sigma(e_i)$ for $i \geq 1$.

Our next step is as follows. As our objective is to attain an object with intrinsic geometry, i.e., geometry independent of parametrization, we would like to consider two oriented affine k-simplices to be the same object if they "look" the same, i.e., if their images are equal. However, we would also like to be able to integrate differential forms on these objects, and two k-surfaces with the same image may behave differently with respect to integration, so we will need to track more information.

First of all, the images of $\sigma = [p_0, p_1, \dots, p_k]$ and $\tau = [q_0, q_1, \dots, q_k]$ are equal if and only if their vertices are the same up to reordering. This reordering operation is a permutation on $k+1$ elements, and if this permutation is even, we say σ and τ have the *same orientation*; otherwise, we say σ and τ have *opposite orientation*. The following result reveals that this notion of orientation is precisely what we need to distinguish between two oriented affine k-simplices that appear identical as point-sets but behave differently with respect to integration.

Theorem 5.2. *If σ and τ are oriented affine k-simplices in an open set $E \subseteq \mathbb{R}^n$ with the same image, then for every k-form ω in E,*

$$\int_\sigma \omega = \epsilon \int_\tau \omega,$$

where $\epsilon = 1$ if σ and τ have the same orientation, and $\epsilon = -1$ otherwise.

Before proceeding further, we broaden the class of objects under consideration by introducing the more general notions of *oriented k-simplex* and *oriented k-chain*. This will allow us to define the boundary of more intricate shapes; after all, the boundary of a simplex is not very interesting!

Definition 5.3. *Let T be a twice-continuously differentiable mapping of an open subset $E \subseteq \mathbb{R}^n$ into another open subset $V \subseteq \mathbb{R}^m$. If $\sigma : Q_k \to E$ is an oriented affine k-simplex, we say $\Phi = T \circ \sigma$ is an **oriented k-simplex** (of class C'').*

Definition 5.4. *Fix an open set $E \subseteq \mathbb{R}^n$ and let M_E be the quotient of the free $\mathbb{Z}$-module on the collection of oriented affine k-simplices in E by the submodule of F generated by all sums of the form $\sigma + \tau$, where σ and τ have opposite orientation, and $\sigma - \tau$, where σ and τ have the same orientation. An **affine k-chain** in E is an element of this module M. A **k-chain** (of class C'') in E is defined analogously, with oriented k-simplices in place of oriented affine k-simplices.*

In the sequel, all oriented k-simplices and k-chains will be of class C'', so we omit the qualifier. With these two definitions in hand, we now power our way through to the definition of the boundary of an oriented k-simplex. Begin by recalling from simplicial homology the notion of *boundary* of an oriented affine k-simplex. For the oriented affine k-simplex $\sigma = [p_0, p_1, \dots, p_k]$, the boundary $\partial\sigma$ is defined as follows.

$$\partial\sigma = \sum_{i=0}^{k} (-1)^i [p_0, p_1, \dots, \hat{p}_i, \dots, p_k].$$

In each summand, $\hat{p}_i$ indicates that the point p_i is to be omitted from the tuple, so that $\partial\sigma$ is an affine $(k-1)$-chain. In general, the *boundary of an affine k-chain* is defined by extending linearly.

If V is an open subset of $\mathbb{R}^m$ and $T : E \to V$ is twice continuously differentiable, then we may may consider T as an operator on oriented affine k-simplices such that $T(\sigma)$ transforms an oriented affine k-simplex σ in E into the oriented k-simplex $T \circ \sigma$. As with the boundary operator, we extend linearly to obtain an operator on affine k-chains. Finally, we define the *boundary of an oriented k-simplex $T \circ \sigma$* to be $T(\partial\sigma)$. Finally, we extend linearly to obtain the *boundary of a k-chain*.

Suppose now that E is the image of Q^k under an injective and twice continuously differentiable map $T : Q^k \to \mathbb{R}^k$. We can define the boundary of E as follows. First of all, by taking $\sigma : Q^k \to \mathbb{R}^k$ to be the canonical inclusion, we obtain an oriented affine k-simplex. Apply T to obtain an oriented k-simplex, and apply ∂ to obtain a $(k-1)$-chain. If the Jacobian of T is positive at every point in the interior of Q^k, we say that this $(k-1)$-chain is the *positively oriented boundary* of E. The following result tells us that we are justified in referring to this property as a property of E as opposed to a property of the oriented k-simplex $T \circ \sigma$.

Theorem 5.5. *Suppose $E \subseteq \mathbb{R}^k$ is such that $E = T_1(Q^k) = T_2(Q^k)$ for two injective and twice continuously differentiable maps T_1, T_2 such that the Jacobians of T_1 and T_2 are both positive everywhere in the interior of Q^k. Next, let $\sigma : Q^k \to \mathbb{R}^k$ be the canonical inclusion. Then*

$$\partial T_1(\sigma) = \partial T_2(\sigma).$$

In other words, the notion of positively oriented boundary that we defined above is *intrinsic* to E in that it does not depend on the particular parametrization of E.

By the way, we can now formulate Stokes' Theorem.

Theorem 5.6 (Stokes' Theorem)**.** *If Ψ is a k-chain of class C^2 in an open set $V \subseteq \mathbb{R}^m$ and ω is a $(k-1)$-form of class C^1 in V, then*

$$\int_\Psi d\omega = \int_{\partial\Psi} \omega.$$

References

[R+64] Walter Rudin et al., *Principles of mathematical analysis*, vol. 3, McGraw-hill New York, 1964.

[SS09] Elias M Stein and Rami Shakarchi, *Real analysis: measure theory, integration, and hilbert spaces*, Princeton University Press, 2009.